Acta Psychologica 191 (2018) 251–260

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Individual differences in absolute pitch performance: Contributions of working memory, musical expertise, and tonal language background

Stephen C. Van Hedger*, Howard C. Nusbaum

Department of Psychology, The University of Chicago, Chicago, IL, USA
Center for Practical Wisdom, The University of Chicago, Chicago, IL, USA

ARTICLE INFO

Keywords:
Absolute pitch
Perception
Working memory
Language
Expertise
Music

ABSTRACT

By definition, individuals with absolute pitch (AP) can categorize with near perfect accuracy without a reference pitch. This definition implies a uniformity of performance across people; however, in reality AP is a complex, multidimensional ability, shaped by both early and recent auditory experiences. In the present study we assess whether AP possessors' accuracy for identifying isolated notes is more distributed when judging more challenging instrumental timbres and octaves, as well as whether variability in note categorization could be explained through individual differences in musical expertise, language background, or working memory. In a standard test of AP, all participants performed virtually perfectly. When tested on the challenging notes, performance was more normally distributed. In exploratory analyses, we found (1) lower accuracy among participants who speak a tonal language, (2) less musical expertise among tonal language participants, and (3) a positive relationship between working memory and note performance among tonal language participants that was not present for non-tonal language participants. Taken together, these results highlight the complexity of AP categorization when considered as an auditory skill rather than a native talent. The observation that working memory may be an important in AP categorization under some challenging circumstances is consistent with recent theoretical accounts of how working memory and expertise relate to auditory recognition more broadly.

1. Introduction

The phenomenon of absolute pitch (AP) – typically defined as the ability to name or produce any musical note without the aid of a reference note – has puzzled researchers for well over a century (Deutsch, 2002). Given the presumed rarity of AP (Bachem, 1955), in conjunction with the musical benefits associated with possessing AP (e.g., Dooley & Deutsch, 2010, 2011), it is perhaps no surprise that the etiology of AP has been of keen interest to researchers and laypersons alike. While there are several theories of how AP develops (see Deutsch, 2013 for a review), an influential hypothesis is that the successful acquisition of AP requires a genetic predisposition, combined with particular early-life experiences that emphasize attention to absolute features of pitch (Zatorre, 2003). This hybrid theory, emphasizing the importance of both nature and nurture in AP acquisition, may explain why an earlier age of beginning musical instruction coincides with a greater odds of possessing AP, yet many individuals who begin early musical instruction do not develop the ability (i.e., early experience may be necessary but not sufficient).

The common use of terms such as “AP possessor” and “non-AP possessor” suggests a kind of fixedness of ability, in which AP can be conceptualized as a clearly delineated and stable talent. Yet, an accumulating body of research has demonstrated that AP may be better conceptualized as a skill (Heald, Van Hedger, & Nusbaum, 2017), with note categorization performance strengthened or weakened depending on a number of short- and long-term experiences. The frequency of hearing particular timbres, octaves, and pitch classes (e.g., C, F#) have been associated with the speed and accuracy of AP note identification, supporting the notion that the process of AP identification is grounded in specific auditory experiences (e.g., Bahr, Christensen, & Bahr, 2005; Miyazaki, 1989). In addition to these effects of frequency of experiencing notes in different timbres, octaves, and classes, there are a number of broad factors in an individual's musical environment that can improve or hinder AP, such as whether a fixed- versus variable-do (e.g., piano versus trumpet) instrument is played (Wilson, Lusher, Martin, Rayner, & McLachlan, 2012), whether one is actively playing a musical instrument at the time of testing (Dohn, Garza-Villarreal, Riisgaard Ribe, Wallentin, & Vuust, 2014), and whether recently heard music adheres to canonical tuning standards (Hedger, Heald, & Nusbaum, 2013; Van Hedger, Heald, Uddin, & Nusbaum, 2018). These findings,

* Corresponding author at: Western Interdisciplinary Research Building, Room 4114, London, ON N6A 5B7, Canada.
E-mail address: svanhedg@uwo.ca (S.C. Van Hedger).

https://doi.org/10.1016/j.actpsy.2018.10.007
Received 6 March 2018; Received in revised form 11 July 2018; Accepted 12 October 2018
Available online 19 October 2018
0001-6918/

taken together, strongly suggest that both short- and long-term musical experiences continually influence AP performance, supporting the framework of AP as an auditory skill.

Beyond influences of specifically musical experiences, AP performance has also been associated with early attention to pitch in the context of language, suggesting a kind of auditory skill transfer. In particular, influential work over the past couple of decades has linked AP accuracy with early tonal language experience (Deutsch, Henthorn, & Dolson, 2004). In this framework, children who grow up hearing and speaking a tonal language, such as Mandarin, attend to pitch in a fundamentally different manner than individuals who grow up hearing and speaking a non-tonal language. This is because a single utterance takes on a different meaning depending on the accompanying pitch pattern. For example, in Mandarin Chinese, the utterance "ma" may mean *mother*, *hemp*, *horse*, or *scold*, depending on whether it is spoken with a high and level pitch (First Tone), a pitch that starts low and ends high (Second Tone), a pitch that starts and ends high but dips in the middle (Third Tone), or a pitch that starts low and end high (Fourth Tone). While these pitch changes can be understood in a relative pitch framework (as absolute pitch processing is not required to understand an utterance that is "rising" or "falling"), empirical work has demonstrated that tonal language speakers display very small pitch variations (often < 50 cents) when repeating utterances, even when tested across days (Deutsch et al., 2004). As such, tonal languages can emphasize absolute cues within a talker, and given that it affects attention to pitch in speech (Lee & Nusbaum, 1993), it may subsequently influence how listeners attend to pitch in musical contexts.

Treating AP as an auditory skill rather than a fixed ability also implies that cognitive constructs, such as working memory (WM), may play a role in the processes underlying note category recognition. Working memory has been conceptualized in different ways, though a common element among many definitions is the use of an executive attention system to maintain and manipulate information in an "online" fashion in service of a cognitive processing goal (e.g., see Baddeley & Hitch, 1974; Cowan, 2008). While there is some debate surrounding whether WM is a fixed cognitive "trait" or can be meaningfully improved through training (e.g., see Morrison & Chein, 2011), WM has been associated with a number of higher-level problem solving and abstract reasoning skills, and appears to be related to the construct of fluid intelligence (e.g., Kane, Hambrick, & Conway, 2005).

There are several related research findings that have positively associated WM with the perception and recognition of musical notes. For example, auditory WM – assessed both musically (through a pitch reproduction task) and non-musically (through an auditory *n*-back task) – has been associated with how well non-AP adults can explicitly learn AP categories (Van Hedger, Heald, Koch, & Nusbaum, 2015). Furthermore, the precision with which an individual can hold an isolated pitch in working memory has been positively associated with absolute pitch memory for well-known music recordings (Van Hedger, Heald, & Nusbaum, 2017). These findings have established an association between WM and pitch representations among individuals who would not qualify as AP possessors; however, there is some evidence that general aspects of auditory memory may differentiate AP and non-AP possessors as well. More specifically, AP possessors have an enhanced auditory digit span (i.e., the ability to remember long strings of numbers in order) relative to non-AP possessors with comparable musical experience (Deutsch & Dooley, 2013), suggesting that AP may develop because of an enhanced auditory memory abilities.

Despite this association between AP and auditory working memory functioning, it is perhaps less clear how variability in note recognition might depend on auditory WM *among AP possessors*, as notes are typically thought to be quickly and effortlessly identified (i.e., all participants would be at ceiling). However, as discussed previously, AP ability may only appear monolithic when it is tested with frequently encountered instruments and octaves. Notes that pose a greater perceptual challenge (e.g., because they are not commonly experienced), in contrast, should reduce performance, though it is unclear whether (1) performance would become more normally distributed (as opposed to simply lowered the same amount across individuals) and (2) whether variance in these more challenging listening situations would relate to auditory WM.

One reason to expect that AP performance in challenging listening situations may relate to auditory WM comes from recent work primarily situated in the domain of speech. The relatively recent emergence of *cognitive hearing science* (Arlinger, Lunner, Lyxell, & Pichora-Fuller, 2009) as a field demonstrates an emphasis on understanding how cognitive systems interact with signal processing in making sense of how we perceive and respond to our auditory environments. Crucially, work situated within this domain has focused on how cognitive processes such as WM may be recruited for successful auditory recognition when the incoming signal is degraded, ambiguous, or generally mismatched from the listener's internal representation (e.g., Rönnberg et al., 2013). In these circumstances, an auditory trace may be held in WM along with hypotheses about its identity to be evaluated, increasing the demands on WM (see Nusbaum & Morin, 1992). For example, working memory has been positively associated with comprehending speech that is masked energetically with speech-shaped noise (e.g., Parbery-Clark, Skoe, Lam, & Kraus, 2009), masked informationally with competing talkers presented simultaneously (e.g., Zekveld, Rudner, Johnsrude, & Rönnberg, 2013), and produced with an unfamiliar accent or idiolect, such as synthetic speech from early computer systems (e.g., see Pisoni, Nusbaum, & Greene, 1985). These findings, taken together, support a general auditory framework in which perceptually challenging listening environments may require WM for successful comprehension, presumably because listeners must hold onto an ambiguous, weak, or distorted signal in working memory as it is compared to a listener's hypothesized perceptual interpretations. Extending this framework to AP, it is possible that some musical notes (e.g., unfamiliar timbres) that pose a perceptual challenge may similarly rely on WM as listeners compare the incoming auditory signal with internal note representations for interpretation.

In the present experiment, we thus measure how perceptually challenging listening environments influence the accuracy with which AP possessors are able to categorize isolated notes. Moreover, we assess in a more exploratory fashion how individual differences in musical background, tonal language background, and auditory WM ability relate to variability in note classification. We first administered a typical assessment of AP ability, using familiar timbres and octave registers, to confirm our participants' self-reported AP ability. Given that performance of participants with AP typically is near ceiling on standard AP tests, in order to assess the relative contributions of different kinds of individual differences to AP performance we developed a new measure of AP that was intended to increase difficulty. In this new task, AP participants categorized notes presented in timbres and octave ranges that were hypothesized to be harder to identify, increasing the perceptual challenge for AP listeners. Given prior reported AP performance differences based on note familiarity (e.g., Miyazaki, 1990) and timbre differences (e.g., Vanzella & Schellenberg, 2010) the goal was to compare AP performance given these perceptual challenges.

2. Methods

2.1. Participants

Thirty-seven self-identified AP possessors participated in the study (M = 24.89 years, SD = 7.71 years, range: 15–50 years) were able to pass a prescreening test verifying their ability to name isolated musical notes without the aid of a reference note (see "Prescreening Results" for details). Participants were recruited through flyers, emails, and a posting on a social media message board (in a moderated group of self-identified AP possessors) and gave informed consent prior to participation. After completing the experiment, participants were emailed a $20 gift card.

2.2. Materials

The experiment was run online using Inquisit 4 Web (Millisecond Software: Seattle, WA). The prescreening test of AP consisted of 24 notes, taken from piano and guitar timbres, ranging from C4 (middle C) [261.63 Hz] to B5 [987.77 Hz]. Each note was 500 ms in duration. The auditory n-back was adapted from a visual n-back task previously used by Jaeggi et al. (2010), which is available on the Millisecond Test Library. The spoken letters from the n-back task were recorded online from the AT&T Natural Voices TTS Synthesizer (AT&T Labs Inc.).

The Challenging AP Test consisted of sine wave notes, bell notes, mbira notes, and timpani notes. We hypothesized that the sine wave notes would be challenging because of their relative unfamiliarity in musical contexts, lack of harmonics, and selected frequency ranges. We hypothesized that the bell notes would be challenging because of their inharmonic overtones (i.e., energy at frequencies other than simple multiples of the fundamental frequency). Inharmonic overtones create a generally weaker sense of fundamental pitch perception, given that the inharmonic frequencies can be perceived as separate pitches. The mbira is a wooden African instrument (sometimes referred to as a "thumb piano") that is played by plucking metal keys affixed to one end of the instrument. We hypothesized that the mbira notes would be challenging because it is a relatively unfamiliar timbre in Western musical contexts, and additionally, similar to the bell timbre, the mbira contains inharmonic overtones that can create a weaker perception of the fundamental pitch. We hypothesized that the timpani notes would be challenging because of the low octave register and diffuse spectral energy at the fundamental. A sample spectrogram for each of the timbres from the Challenging AP Test is represented in Fig. 1.

The sine wave stimuli were generated in Adobe Audition 3.0, while all other music note stimuli were selected from the instrument database of Reason 4.0 (Propellerhead Software: Stockholm, Sweden). The sine wave notes were 200 ms in duration with 10 ms fade in and fade out. Half of the sine waves were selected from a low octave register (ranging from E2 [82.41 Hz] to D#3 [155.56 Hz]), while the other half of the sine waves were selected from a high octave register (ranging from E7 [2637.02 Hz] to D#8 [4978.03 Hz]). The bell notes were 1000 ms in duration with 10 ms fade out, and the notes ranged from E5 [1318.51 Hz] to D#7 [2489.02 Hz]. The mbira notes were 500 ms in duration with 10 ms fade out. Similar to the sine wave notes, we selected 12 lower-octave notes and 12 higher-octave notes. The lower-octave mbira stimuli ranged from F3 [174.61 Hz] to E4 [329.63 Hz], while the higher-octave mbira stimuli ranged from G#6 [1661.22 Hz] to G7 [3135.96 Hz]. The timpani notes were 1000 ms in duration with 10 ms fade out. The timpani stimuli ranged from F2 [87.31 Hz] to E3 [164.81 Hz]. All musical notes were sampled at 44.1 kHz with 16-bit depth and are available on the Open Science Framework. The questions assessing musical expertise and tonal language experience were administered through Qualtrics (Provo, UT).

2.3. Procedure

After consenting to participate in the study, participants were presented with the prescreening test of AP ability. On each trial, participants heard a single note and then typed a number corresponding to a given note name (with "C" corresponding to 1 and "B" corresponding to 12). A pitch wheel, reminding participants of the note-to-number mapping, was displayed on each trial. Participants were told prior to the prescreening test to respond as quickly and accurately as possible. We played 1000 ms of white noise between trials to minimize the use of relative pitch strategies. Participants needed to correctly identify at least 70% of notes with an average response time of 7 s or faster in order to qualify for the experiment.

After the prescreening, participants completed the n-back task. Participants were first introduced to the general concept of the n-back through a series of instructional slides. These slides told participants that they would hear a series of spoken letters that would occur one after another. The instructions then introduced the goal of the task (e.g., for introducing the 3-back, the text read "you have to press 'A' every time the current letter is the same as the one presented 3 positions back in the sequence. Otherwise, you don't have to respond"). Participants were separately introduced to the 2-back, 3-back, 4-back, and 5-back in this manner prior to beginning. After participants read through the instructional slides, they completed non-scored, practice blocks of the 2-back, 3-back, 4-back, and 5-back (in this order). Participants completed only one block for each "n." Blocks consisted of 20 spoken letters, excluding the first "n" letters in which there could not be a target (e.g., there would be 23 spoken letters for the 3-back block). Each block contained 6 targets (i.e., when the spoken letter matched the one presented *n* positions back) and 14 non-targets (i.e., when the spoken letter did not match the one presented *n* positions back). After these practice blocks, participants completed three scored blocks of each of the 2-back, 3-back, 4-back, and 5-back (in this order). The scored blocks were identical to the practice blocks (consisting of 20 spoken letters, excluding the first "n" letters, with 6 targets and 14 non-targets). Participants completed three blocks for each "n" before moving on (i.e., participants completed three blocks of the 2-back, which were then followed by three blocks of the 3-back etc.). As such, participants heard a total of 60 scored letters for each "n" (18 targets, 42 non-targets). Participants were explicitly told not to use any external aids in the n-back (e.g., writing the letters down on a piece of paper), and the data suggest that this instruction was followed (e.g., no participant achieved a perfect score on the 5-back, and performance decreased as "n" increased).

Next, participants completed the Challenging AP Test, consisting of sine wave notes, bell notes, mbira notes, and timpani notes (blocked, in this order). Each block contained 24 trials each for a grand total of 96 trials. Given that we only sampled one octave from the timpani (to preserve ecological validity of a typically encountered timpani octave register), participants heard each note twice, randomly presented, to generate 24 trials. Prior to each block, participants were told which timbre they would be classifying. Similar to the Prescreening Test, participants made their responses by typing in a corresponding number to the musical note. Also similar to the prescreening test, we played 1000 ms of white noise between trials.

After the Challenging AP Test, participants filled out a brief questionnaire. The questionnaire specifically collected information about participants' musical training – years of instruction on their primary instrument, the names of instruments they reported playing (as well as number of years playing each instrument, binned into categories of *1–5 years*, *6–10 years*, *11–15 years*, and *> 15 years*), the age at which they first began musical instruction, their self-reported AP ability (scale of 0 to 100), their self-reported musical proficiency (scale of 0 to 100), their self-reported knowledge of music theory (scale of 0 to 100), and whether they had a college degree in music. We also asked participants if they spoke a tonal language. After confirmation of completing the questionnaire, participants were sent a gift card and an email detailing the purpose of the experiment. During this email exchange, participants had the opportunity to provide comments and report any problems they encountered while completing the experiment. No participant reported any problems with completing the experiment (e.g., audio presentation problems).

2.4. Statistical approach and calculation of measures

For each analysis, we report a *p*-value in addition to a Bayes factor (BF), calculated using JASP 0.8.2 (JASP Team, 2018). We kept the default priors provided by the program, as recommended by Wagenmakers et al. (2017). The reported BF (BF_{10}) represents the relative evidence in favor of the alternative versus the null hypothesis. For example, a BF_{10} of 4 would mean that the observed data is four times more likely to occur under the alternative hypothesis than the

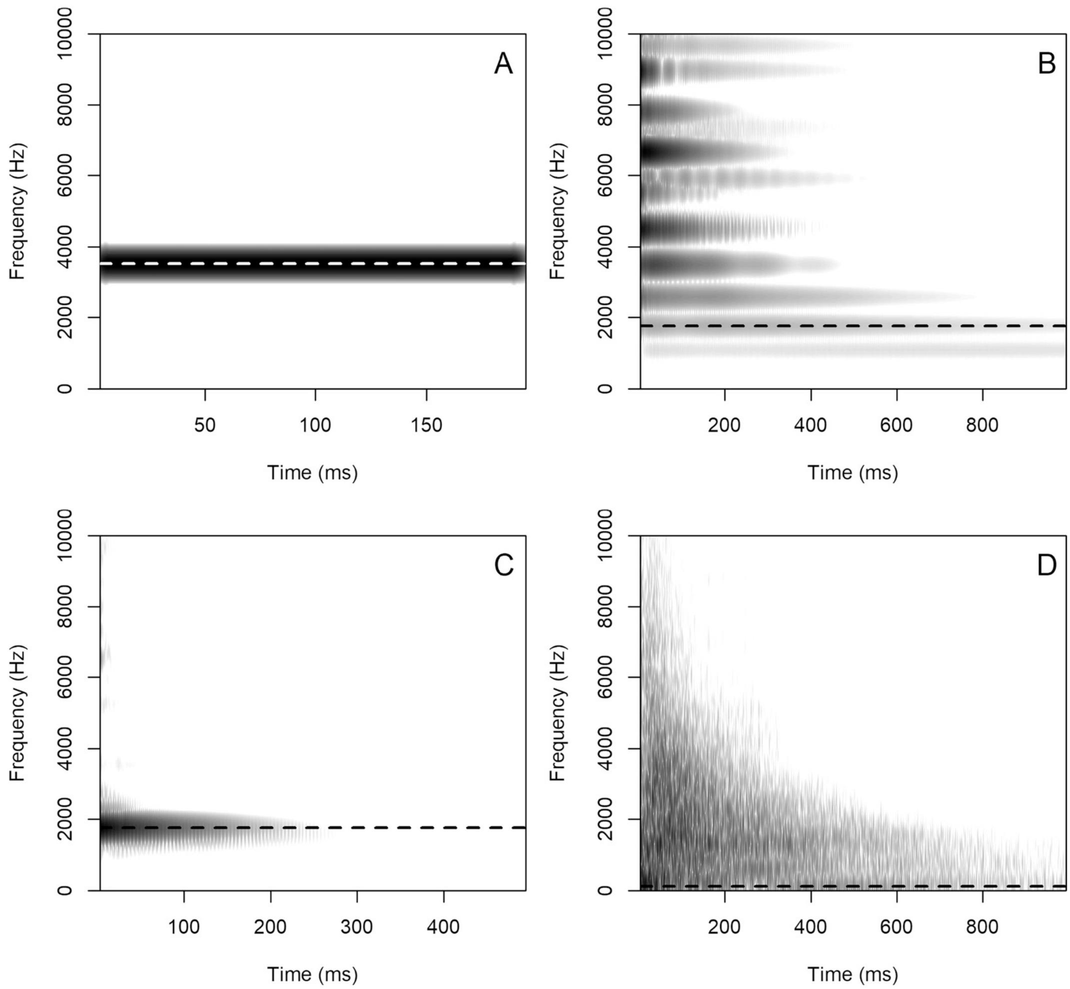

Fig. 1. Sample spectrograms for the sine wave notes (A), bell notes (B), mbira notes (C), and timpani notes (D). The dashed lines represent the fundamental frequency (A7/3520 Hz for sine; A6/1760 Hz for bell; A6/1760 Hz for mbira; A2/110 Hz for timpani).

null hypothesis given the priors of the model, whereas a BF_{10} of 0.25 would mean that the observed data is four times more likely to occur under the null hypothesis than the alternative hypothesis given the priors of the model. For our multiple regression analyses, the reported Bayes factor ($BF_{Inclusion}$) can be interpreted in a similar manner, but it represents the relative evidence in favor of including a particular term in a model (as opposed to removing the term from the model). We interpret our BF_{10} using the categories outlined by Wagenmakers et al. (2017), in which evidence for the alternative hypothesis is described as anecdotal (BF of 1–3), moderate (BF of 3–10), strong (BF of 10–30), very strong (BF of 30–100), or extreme (BF > 100).

In assessing AP performance, we operationalize note categorization accuracy in terms of both mean accuracy (e.g., 75% correct) and Mean Absolute Deviation (MAD). MAD has been previously used in assessments of AP ability (e.g., Bermudez & Zatorre, 2009) and provides a more graded measure of performance. The MAD is calculated by taking the absolute value of the difference between the correct note and the reported note (in semitones). Given that pitch classes repeat every 12 notes in Western music, MAD values can range from 0 (correct) to 6. Thus, random guessing would yield a MAD of 3 (a uniform distribution ranging from 0 to 6), while perfect performance would yield a MAD of 0. All mean response times reported in the paper have been culled such that values > 3 standard deviations from the mean have been removed.

We calculated participants' performance on the auditory n-back using signal detection theory (Macmillan & Creelman, 2005). Specifically, we calculated d-prime (d′) scores for each “n” (2-back, 3-back, 4-back, and 5-back). If participants achieved a perfect score (either correctly identifying 18 out of 18 targets or false alarming to 0 out of 42 non-targets), then we doubled the denominator and added one error (e.g., 35/36 hits). As such, the maximum d′ score a participant could obtain was 4.17.

We created a composite musical expertise score from eight musical measures (see Table 3 for summary statistics). The first three questions were self-assessments of musical ability (rating one's AP ability, musical ability, and knowledge of music theory on a scale of 0–100). The remaining five measures were more objective assessments of musical training, incorporating (1) the number of years an individual reported playing their primary instrument, (2) the number of years an individual reported playing their primary instrument divided by their age, (3) whether a participant reported a college degree in music, (4) the reported number of played instruments, and (5) the number of played instruments for which proficiency had been gained (defined as > 5 years of experience). This music composite was calculated with z-normalized values in R using the *multicon* package.

3. Results

The results are arranged in the following manner. First, we describe general aspects of performance (mean accuracy, MAD, and mean response time) for the Prescreening and the Challenging AP Tests. We then assess how these performance measures differ as a function of tonal language background, musical background, and working memory.

3.1. Prescreening AP test

All participants passed the prescreening test verifying their self-reported AP ability. Participants averaged 95.4% correct (*SD*: 7.5%, range: 75% to 100%) using the most conservative scoring criterion (only giving credit for correct answers), which was above the 70% cutoff for inclusion in the experiment. Given this high level of note categorization accuracy, it is not surprising that participants' MAD was, on average, close to 0 (*M*: 0.06, *SD*: 0.09, range: 0 to 0.33). Furthermore, all participants' average response times were within our cutoff of seven seconds. Participants made their response in an average of 3.99 s (*SD*: 0.92 s), with a range of 2.34 s to 6.46 s. We thus interpret the results of the Prescreening AP Test as strong evidence that the self-reported AP possessors in the current experiment indeed possessed genuine AP.

3.2. Challenging AP test

Participants' performance was more distributed in the Challenging AP test, which is notable considering the same participants performed at or near ceiling in the Prescreening Test. Participants' mean accuracy for sine notes was 56.4% (*SD*: 16.3%, range: 25% to 100%), for bell notes was 49.3% (*SD*: 19.7%, range: 20.8% to 87.5%), for mbira notes was 65.7% (*SD*: 23.3%, range: 12.5% to 100%), and for timpani notes was 60.8% (*SD*: 24.3%, range: 16.7% to 100%). Aggregated across all timbres, mean accuracy was 58.1% (*SD*: 17.3%, range: 22.9% to 93.8%). While these performance ranges may appear surprising given participants' near-ceiling performance on the Prescreening AP Test, it should be emphasized that simply calculating correct and incorrect responses may be inappropriate given the difficult nature of the Challenging AP Test (as such an approach treats all misclassifications as equivalent). Thus, we view MAD as a more appropriate measure for reflecting performance variability.

Participants' MAD for sine notes was 0.72 (*SD*: 0.47, range: 0 to 2), for bell notes was 1.66 (*SD*: 0.60, range: 0.17 to 2.63), for mbira notes was 0.81 (*SD*: 0.68, range: 0 to 2.79), and for timpani notes was 0.61 (*SD*: 0.52, range: 0 to 2.23). Aggregated across all timbres, MAD was 0.95 (*SD*: 0.44, range: 0.19 to 1.94), which was significantly worse than pretest performance (t (36) = 13.03, $p < .001$, $BF_{10} = 1.90e + 12$), but also significantly better than chance performance, represented by a MAD of 3 (t (36) = −28.47, $p < .001$, $BF_{10} = 9.93e + 22$).

Participants' response time (in seconds) for sine notes was 4.01 (*SD*: 1.38, range: 2.17 to 7.51), for bell notes was 4.80 (*SD*: 1.50, range: 2.33 to 8.44), for mbira notes was 3.78 (*SD*: 1.00, range: 2.00 to 6.22), and for timpani notes was 3.96 (*SD*: 1.36, range: 2.13 to 8.72). Aggregated across all timbres, average response time was 4.14 (*SD*: 1.18, range: 2.25 to 7.02), which was nominally slower than but not significantly different from pretest performance (t (36) = −1.01, p = .319, $BF_{10} = 0.283$). Taken together, the results from the Challenging AP Test support the notion that AP performance, even among self-identified AP possessors, can display significant note categorization variability under more perceptually challenging listening conditions However, it remains unclear whether this increased variability relates to our measured factors of tonal language, musical background, and working memory. In the following sections we thus assess how performance in the AP Tests relates to these measured factors.

3.3. Tonal language background

We did not specifically recruit AP participants with tonal language experience; however, 13 of our 37 participants reported speaking a

Table 1
Note categorization accuracy, mean absolute deviation (MAD), and response time (RT) in seconds for the Prescreening AP Test (top) and Challenging AP Test (bottom) parsed by tonal language, musical expertise, and working memory ability.

A. Prescreening AP test

	Tonal$^+$ (n = 13)	Tonal$^-$ (n = 24)	Music$^+$ (n = 18)	Music$^-$ (n = 19)	WM$^+$ (n = 18)	WM$^-$ (n = 19)
ACC (%)	93.6 (9.4)	96.4 (6.3)	97.2 (5.7)	93.6 (8.7)	93.8 (8.7)	96.9 (6.0)
MAD	0.07 (0.10)	0.05 (0.09)	0.03 (0.07)	0.08 (0.10)	0.08 (0.11)	0.04 (0.07)
RT (s)	3.97 (0.93)	4.00 (0.94)	3.76 (0.89)	4.21 (0.91)	4.02 (0.83)	3.96 (1.02)

B. Challenging AP test

	Tonal$^+$	Tonal$^-$	Music$^+$	Music$^-$	WM$^+$	WM$^-$
Overall						
ACC (%)	48.7 (17.6)	63.1 (15.2)*	59.6 (15.0)	56.6 (10.5)	56.4 (15.2)	59.6 (19.3)
MAD	1.19 (0.48)	0.82 (0.36)*	0.98 (0.49)	0.92 (0.38)	0.93 (0.39)	0.97 (0.49)
RT (s)	4.29 (1.03)	4.05 (1.27)	3.68 (1.14)	4.57 (1.01)*	4.09 (0.94)	4.18 (1.40)
Sine						
ACC (%)	47.1 (15.0)	61.5 (15.0)*	57.4 (16.5)	55.5 (16.6)	53.7 (13.5)	59.0 (18.5)
MAD	0.94 (0.56)	0.59 (0.37)+	0.69 (0.48)	0.74 (0.47)	0.67 (0.35)	0.76 (0.57)
RT (s)	4.14 (1.32)	3.95 (1.43)	3.58 (1.27)	4.43 (1.38)+	3.90 (1.08)	4.13 (1.64)
Bell						
ACC (%)	44.6 (17.9)	51.9 (20.4)	50.9 (19.4)	47.8 (20.3)	49.1 (20.0)	49.6 (19.8)
MAD	1.74 (0.58)	1.61 (0.62)	1.65 (0.58)	1.66 (0.64)	1.66 (0.59)	1.65 (0.63)
RT (s)	5.13 (1.18)	4.61 (1.65)	4.07 (1.38)	5.49 (1.38)*	4.70 (1.24)	4.88 (1.75)
Mbira						
ACC (%)	53.8 (27.7)	72.0 (18.1)*	68.5 (18.7)	62.9 (27.2)	64.1 (22.0)	67.1 (25.0)
MAD	1.13 (0.81)	0.64 (0.55)+	0.76 (0.59)	0.86 (0.77)	0.85 (0.72)	0.77 (0.66)
RT (s)	3.81 (0.78)	3.76 (1.11)	3.37 (0.96)	4.17 (0.89)*	3.76 (0.80)	3.80 (1.18)
Timpani						
ACC (%)	49.4 (23.6)	67.0 (22.9)*	61.6 (24.7)	60.1 (24.6)	58.6 (22.2)	62.9 (26.6)
MAD	0.93 (0.69)	0.45 (0.32)*	0.56 (0.38)	0.67 (0.64)	0.54 (0.32)	0.68 (0.66)
RT (s)	4.09 (1.55)	3.88 (1.28)	3.69 (1.30)	4.21 (1.41)	4.01 (1.14)	3.91 (1.58)

Note: Results are split based on whether participants spoke a tonal language (Tonal$^+$/Tonal$^-$), whether participants were above or below a median split on musical expertise (Music$^+$/Music$^-$), and whether participants were above or below a median split on working memory (WM$^+$/WM$^-$).

* $p < .05$.

\+ $.05 \le p < .10$.

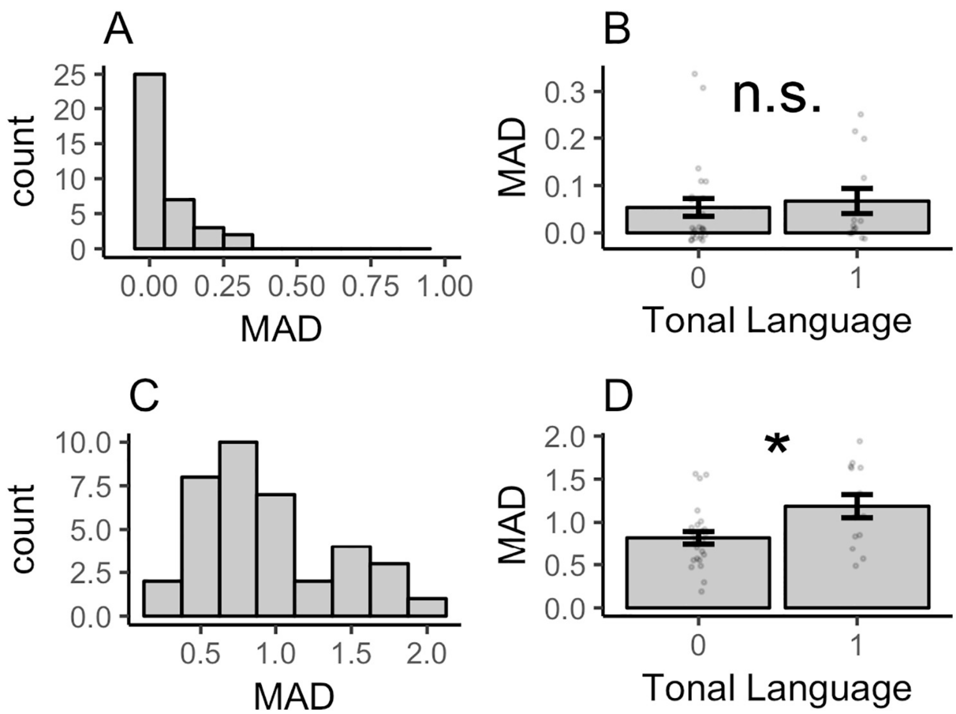

Fig. 2. For the Prescreening Test, MAD displayed virtually no variability (A), and there was no significant difference between tonal and non-tonal language participants with respect to performance (B). For the Challenging AP Test, MAD displayed more variability (C), and tonal language participants had a significantly higher MAD compared to non-tonal participants (D). $^{*}p < .05$.

tonal language. In the Prescreening AP Test (Table 1A), we did not observe any evidence that mean accuracy, MAD, or response time differed between tonal and non-tonal language participants. Given the overall high levels of performance across all participants, this is perhaps not surprising. In the Challenging AP Test (Table 1B), however, tonal language participants performed significantly worse overall compared to non-tonal language participants, at least in terms of mean accuracy (t (21.76) = 2.49, p = .021, BF_{10} = 4.02) and MAD (t (19.30) = −2.40, p = .027; BF_{10} = 4.16). The performance difference between tonal and non-tonal participants for MAD is represented in Fig. 2D. The lower performance for tonal versus non-tonal language participants, moreover, was generally present for the sine notes, mbira notes, and timpani notes, but not bell notes. Response time did not significantly differ between tonal and non-tonal language participants in the Challenging AP Test.

3.4. Musical background

Musical expertise, as assessed through our composite measure (see Method for details), had a mean of 0 (SD: 0.53, range: −0.98 to 0.98), which was expected given that it consisted of an average of eight z score-normalized items. The median value was 0.062, and we assessed how musical expertise related to AP performance by performing a median split and binning participants into either high or low expertise groups. Given the odd number of participants (37), this resulted in 18 participants with high musical expertise and 19 participants with low musical expertise (though it should be noted that performing the median split to create 19 participants with high musical expertise and 18 participants with low musical expertise does not change the interpretation of the results).

Musical expertise did not significantly relate to mean accuracy, MAD, or response time in the Prescreening AP Test (Table 1A), likely due to the relatively homogeneous performance in prescreening. For the Challenging AP Test (Table 1B), however, we observed that participants who were high on the musical expertise composite responded more quickly overall than participants who were low on the musical expertise composite (t (34.56) = 2.45, p = .019, BF_{10} = 3.06). This expertise-related difference in response time was observed for sine, bell, and mbira notes; it was not significantly different for timpani notes.

In addition to the musical expertise composite measure, we assessed whether participants could be meaningfully grouped together based on their instrument expertise. While there was considerable diversity with respect to the number of instruments participants reported (as participants reported experience playing an average of 3.75 different instruments), instrument expertise - defined as over ten years of reported playing experience - generally could be classified in one of three categories. The first category was *piano*. The second category was *string* (including violin, viola, cello, and bass). The third category was *wind* (including brass and woodwind instruments, as well as the voice). Approximately two-thirds of the participants (25 of 37) reported sufficient expertise to be included in the *piano* group, while just under half of the participants reported sufficient expertise to be included in the *string* and *wind* groups (16 and 18 of 37, respectively).

Given the number of participants in each of these three categories, these instrument groupings were clearly not mutually exclusive. Thus, we assessed group interrelations with contingency tables and their Bayesian equivalent. In this analysis, the *piano* group appeared to be independent from both the *string* (X^2 (1) = 0.02, p = .893, BF_{10} = 0.371) and *wind* (X^2 (1) = 0.01, p = .909, BF_{10} = 0.371) groups. The *string* and *wind* groups, however, were dependent (X^2 (1) = 10.09, p = .001, BF_{10} = 59.09), in that participants reported either playing a string or wind instrument (i.e., there was a negative relationship between group membership). Put another way, these results suggest that piano expertise is unrelated to expertise in either string or wind instruments; however, expertise in a string instrument means that expertise in a wind instrument is less likely (and vice versa).

We specifically assessed how expertise within these instrument groups related to performance in both the Prescreening and Challenging AP Tests. Instrument grouping was not significantly related to note categorization performance on the Prescreening AP Test, either in terms of mean accuracy or MAD (all ps > .203 and BF_{10} < 0.61). Piano participants, however, responded marginally faster than non-piano participants in prescreening (t (24.37) = 1.76, p = .092; BF_{10} = 0.972), though the Bayesian analysis did not support this difference. Descriptive statistics of performance in the Prescreening AP Test, separated by instrument group, are provided in Table 2A.

Table 2
Note categorization accuracy, mean absolute deviation (MAD), and response time (RT) in seconds for the Prescreening AP Test (A) and Challenging AP Test (B) parsed by instrument expertise.

A. Prescreening AP test

	Piano^{+} (n = 25)	Piano^{-} (n = 12)	String^{+} (n = 16)	String^{-} (n = 21)	Wind^{+} (n = 18)	Wind^{-} (n = 19)
ACC (%)	94.5 (8.7)	97.2 (4.1)	96.4 (6.6)	94.6 (8.2)	94.4 (7.8)	96.3 (7.3)
MAD	0.07 (0.11)	0.05 (0.06)	0.04 (0.07)	0.07 (0.11)	0.08 (0.11)	0.04 (0.08)
RT (s)	3.82 (0.93)	4.35 (0.82)$^{+}$	3.96 (0.98)	4.02 (0.89)	4.14 (0.90)	3.85 (0.94)

B. Challenging AP test

	Piano^{+}	Piano^{-}	String^{+}	String^{-}	Wind^{+}	Wind^{-}
Overall						
ACC (%)	53.7 (17.7)	67.1 (12.7)*	56.3 (16.7)	59.4 (18.0)	60.5 (17.7)	55.7 (17.0)
MAD	1.05 (0.46)	0.75 (0.33)*	1.04 (0.46)	0.88 (0.42)	0.87 (0.39)	1.03 (0.47)
RT (s)	3.97 (1.17)	4.45 (1.19)	4.06 (1.15)	4.19 (1.23)	4.03 (1.15)	4.23 (1.24)
Sine						
ACC (%)	52.3 (16.7)	64.9 (12.0)*	56.3 (14.3)	56.5 (18.0)	56.9 (18.8)	55.9 (14.0)
MAD	0.79 (0.50)	0.57 (0.38)	0.77 (0.50)	0.68 (0.45)	0.69 (0.48)	0.74 (0.47)
RT (s)	3.88 (1.31)	4.28 (1.55)	3.84 (1.31)	4.15 (1.45)	4.08 (1.48)	3.95 (1.32)
Bell						
ACC (%)	45.0 (16.8)	58.3 (22.7)$^{+}$	44.5 (19.9)	53.0 (19.1)	53.9 (18.4)	45.0 (20.3)
MAD	1.77 (0.49)	1.41 (0.75)	1.80 (0.62)	1.54 (0.58)	1.57 (0.57)	1.74 (0.64)
RT (s)	4.60 (1.49)	5.19 (1.51)	4.67 (1.35)	4.89 (1.64)	4.48 (1.43)	5.09 (1.53)
Mbira						
ACC (%)	60.8 (25.1)	75.7 (15.4)*	64.1 (24.1)	66.9 (23.1)	69.7 (20.0)	61.8 (25.3)
MAD	0.93 (0.76)	0.57 (0.39)$^{+}$	0.87 (0.64)	0.76 (0.72)	0.68 (0.60)	0.93 (0.75)
RT (s)	3.55 (0.91)	4.26 (1.05)$^{+}$	3.83 (0.99)	3.74 (1.03)*	3.61 (0.95)	3.94 (1.04)
Timpani						
ACC (%)	56.7 (26.4)	69.4 (17.2)$^{+}$	60.2 (21.5)	61.3 (26.8)	61.6 (26.1)	60.1 (23.3)
MAD	0.70 (0.59)	0.43 (0.29)$^{+}$	0.73 (0.67)	0.52 (0.38)	0.52 (0.38)	0.70 (0.63)
RT (s)	3.84 (1.44)	4.21 (1.21)	3.91 (1.45)	3.99 (1.33)	3.96 (1.24)	3.95 (1.51)

Note: Results are split based on whether participants reported expertise with piano (Piano^{+}/Piano^{-}), expertise with a stringed instrument (String^{+}/String^{-}), or expertise with a wind instrument (Wind^{+}/Wind^{-}).

* $p < .05$

$^{+}$ $.05 \le p < .10$.

For the Challenging AP Test, we did not find any evidence that performance differed in the *wind* or *string* groups, operationalized in terms of mean accuracy, MAD, and response time (all *ps* > .167 and $BF_{10} < 0.687$). Participants in the *piano* group, in contrast, displayed some evidence of lower performance, at least in terms of mean accuracy and MAD. Specifically, *piano* participants had a lower mean accuracy compared to non-*piano* participants. This difference was statistically significant but was only anecdotally supported by the Bayes Factor ($t(29.39) = 2.63$, $p = .013$, $BF_{10} = 2.54$). Similarly, *piano* participants had a higher overall MAD compared to non-*piano* participants. This difference was once again statistically significant but only anecdotally supported by the Bayes Factor ($t(29.43) = -2.31$, $p = .028$, $BF_{10} = 1.61$). Descriptive statistics of performance in the Challenging AP Test, separated by instrument group, are provided in Table 2B.

3.5. Auditory N-back

Participants' average d-primes were 3.27 (*SD*: 0.95) for the 2-back, 2.17 (*SD*: 0.88) for the 3-back, 1.57 (*SD*: 0.85) for the 4-back, and 1.39 (*SD*: 0.92) for the 5-back. Given the ceiling or near-ceiling performance across a majority of the participants for the 2-back, we treated this level as additional practice. For the remaining levels, we averaged d′ scores to create a composite n-back measure for each participant (*M*: 1.71, *SD*: 0.73, range: −0.07 to 3.46). Only one participant obtained a composite score below zero; the next lowest participant's composite was 0.75. However, upon inspection of the results it did not appear that this low composite score was due to task noncompliance (e.g., never responding to any stimulus). As such, we included this participant in our analyses, though it should be noted that removing this participant does not change the pattern of our results. Similar to the music composite measure, we performed a median split and binned participants into either a high ($n = 18$) or low ($n = 19$) working memory group. We did not observe any working memory-related performance differences for the Prescreening AP Test (Table 1A). For the Challenging AP Test (Table 1B), we also did not observe any working memory-related performance differences for the Challenging AP Test (all *ps* > .327 and $BF_{10} < 0.47$).

3.6. Relationship among individual differences

Our individual difference analyses can be summarized into two main points. First, performance measures in the Prescreening AP Test did not meaningfully vary based on tonal language, musical background, or working memory. This is not surprising given that prescreening performance was very close to ceiling and displayed very little variability across individuals. Second, for the Challenging AP Test, tonal language background and piano expertise were associated with lower performance (mean accuracy and MAD), whereas overall musical expertise, represented by the composite measure, was associated with faster response times.

These analyses provide some evidence that AP performance variability may relate to individual differences in previous auditory experiences such as tonal language background and musical expertise, yet it is possible that these factors may not be independent. We thus assessed whether tonal language speakers differed from non-tonal language speakers on the musical expertise composite, instrument expertise, or working memory. In this analysis, tonal language participants displayed *lower* musical experience composite scores (M ± SD: −0.24 ± 0.49)

Table 3
Summary of questions assessing musical background, split by tonal and non-tonal participants.

	Tonal			Non-tonal		
	M	*SD*	Range:	M	*SD*	Range:
Measure						
Musical training (years)	15.31	6.17	[9, 32]	18.94	9.73	[2, 47]
Age of music Onset (years)	5.15	1.57	[3, 9]	6.21	2.41	[2, 11]
Age of AP realization	9.92	3.69	[3, 16]	10.17	5.27	[2, 27.5]
Years of music training/age	0.69	0.18	[0.26, 0.91]	0.70	0.18	[0.13, 0.94]
Number of total instruments	3.54	1.98	[2, 9]	3.75	1.62	[1, 7]
Number of proficient instruments	1.62	0.77	[0, 3]	1.83	0.92	[1, 4]
College degree in music (proportion)	0.15	0.38	[0, 1]	0.58	0.50	[0, 1]
Self-assessed AP (0–100)	80.69	18.15	[50, 100]	88.22	14.76	[30, 100]
Self-assessed music (0–100)	79.62	17.86	[39, 100]	87.91	9.36	[70, 100]
Self-assessed theory (0–100)	73.85	15.34	[40, 100]	76.39	16.58	[42, 98]
Music experience composite	−0.24	0.49	[−0.98, 0.61]	0.13	0.51	[−0.92, 0.98]

Note: Descriptively, non-tonal language participants reported greater musical expertise compared to tonal language participants on a majority of the measures. The musical experience composite, which consisted of all measures with the exception of age of music onset and the age at which participants realized they possessed AP, was significantly different between non-tonal and tonal language participants.

compared to non-tonal participants (M ± SD: 0.13 ± 0.51; t(25.47) = 2.21, p = .037, BF_{10} = 1.96). Table 3 provides a description of how tonal and non-tonal language participants differed on the items that comprised the musical expertise composite. Tonal language participants also were less likely to be members of the *wind* expertise category (X^2 (1) = 5.25, p = .022, BF_{10} = 5.13). Given these two findings, it is perhaps not surprising that we additionally found that members of the *wind* expertise category were higher on the music composite measure (M ± SD: 0.21 ± 0.50) compared to individuals without expertise on a *wind* instrument (M ± SD: −0.20 ± 0.48; t(34.68) = −2.47, p = .019, BF_{10} = 3.14). Tonal language participants did not significantly differ from non-tonal language participants on working memory, *piano* category membership, or *string* category membership (all ps > .103 and BF_{10} < 1.32).

These results highlight that the tonal language participants in the present experiment may have faced particular difficulties in the Challenging AP Test. Not only did the tonal language participants possess linguistic pitch categories that could potentially interfere with AP note categories, but they also displayed lower amounts of musical expertise. These findings led us to assess whether these participants would display a different relationship between WM and AP performance, particularly if listeners must rely on WM as the challenges of the listening environment are increased (cf. Heald & Nusbaum, 2014).

3.7. Interactions of working memory and tonal language

Given the unanticipated finding that tonal language was associated with lower musical expertise, combined with the observation that tonal language participants had higher MAD for the Challenging AP Test, we constructed an exploratory multiple linear regression model to assess how the relationship between working memory and MAD interacted with tonal language. The reasoning behind this analysis was that tonal language participants – who may have had difficulty in the Challenging AP Test because of both lower musical expertise and competing pitch representations from their tonal language – might rely on WM for successful note recognition more so than non-tonal language participants. Similarly, tonal language participants' AP performance may particularly benefit from musical expertise, as this kind of experience might help guard against possible interference from pitch representations in a tonal language context.

We included the previously described auditory n-back and musical expertise composite measures in our model as both main effects and as interactions with tonal language. In this model, we found a significant main effect of tonal language (B = 1.26, SE = 0.36, p = .001; $BF_{Inclusion}$ = 9.89), which was expected given the reported differences between tonal and non-tonal language participants in the prior section. We also found a significant interaction between tonal language and auditory n-back (B = −0.61, SE = 0.21, p = .006; $BF_{Inclusion}$ = 4.04), as well as between tonal language and musical expertise (B = −0.57, SE = 0.28, p = .048; $BF_{Inclusion}$ = 2.10). The interaction between tonal language and auditory n-back arose because better n-back performance was related to better AP performance within the tonal-language participants, whereas there was a slight trend in the opposite direction for non-tonal language participants (showing poorer AP performance with increased n-back performance). Similarly, the interaction between tonal language and musical expertise was characterized by a relationship showing that increased musical expertise was related to better AP performance for challenging notes within the tonal language participants, whereas there was a much weaker positive relationship between musical expertise and AP performance among non-tonal language participants. No other main effect or interaction was significant. The overall model, which was significant (F (5, 31) = 3.99, p = .007), explained 39.2% of variance in MAD (29.3% when adjusted for multiple predictors). The interaction between tonal language and auditory n-back is plotted in Fig. 3.

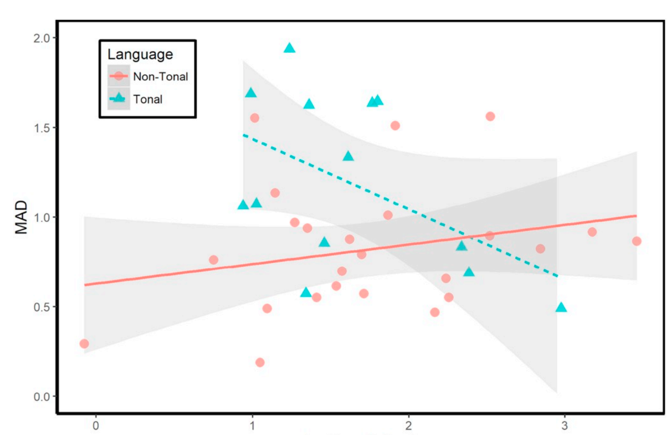

Fig. 3. Interaction between tonal language background and auditory n-back in explaining mean absolute deviation (MAD) among AP possessors. Among the tonal language participants, higher auditory n-back scores corresponded to better AP performance (lower MAD). These two factors did not appear to be related among non-tonal participants.

4. Discussion

Theories of AP note categorization are often limited by the treatment of AP as a clearly delineated ability (e.g., Athos et al., 2007) – displaying minimal and ultimately inconsequential variability in note performance. Although individual differences in AP possessors has been observed for some time, there is a tendency to treat AP listeners as homogeneous largely due to ceiling effects on standard measures of note identification. However, the differences among listeners, in respect of a history of auditory experience, appears to be informative about the basic cognitive mechanisms that are important for AP as a complex, distributed ability (e.g., Bahr et al., 2005; Bermudez & Zatorre, 2009).

The present results demonstrate influences of tonal language and musical expertise on aspects of AP performance. However, these are very different kinds of listening experiences. On the one hand, extended musical training directs listeners' attention to pitches and to the function of those pitches in a formal musical system. Thus, the kind of listening experience represented by musical expertise is directly related to the skill of identifying notes from their sounds. By contrast, tonal language experience is the application of a formal system for categorization to a different kind of pitch information—pitch patterns rather than absolute pitch values. In this respect, tonal-language listeners have experience focusing attention on the lowest frequency (the fundamental) of a signal in both language and music domains. However, rather than have a single categorization system for pitches, AP individuals who speak a tonal language have two (tonal phonology and musical notes). This difference may be important in understanding the differential reliance on auditory working memory in the two groups of listeners.

More specifically, if AP is conceptualized as a kind of "second language" for tonal language speakers (Deutsch et al., 2004), then working memory measures may relate to AP performance because they index an individual's executive control of attention, which may be important for efficiently switching between speech and music pitch representations. Indeed, bilingual speakers often display improvements in attentional control relative to monolinguals, presumably due to the experience of flexibly shifting between two language systems (e.g., Zhou & Krott, 2016). While we did not observe any tonal language advantages in our working memory task, it is nevertheless possible that differences in working memory among tonal language participants related to the efficacy of controlling attention to prevent tonal language categories from interfering with the task.

With prior research establishing a positive association between explicit, active musical training and better AP performance (Dohn et al., 2014), the relationship between working memory and tonal language may have been particularly pronounced in the present experiment given that our tonal language participants were also lower on measures of musical expertise compared to our non-tonal language participants. In the present sample, it is unfortunately not possible to disentangle these factors. However, tonal language and musical expertise are clearly dissociable in theory, and thus future research could assess whether the observed "tonal language effects" on working memory can be conceptualized as "musical expertise effects" on working memory by testing a sample of tonal and non-tonal AP possessors who are matched on both high and low amounts of musical experience. If working memory influences note category judgments solely as a function of musical expertise, then we would predict no interaction of tonal language and working memory when tonal and non-tonal participants are matched on musical expertise. If tonal language participants still display a stronger relationship between working memory and AP performance when matched on musical expertise, this suggests that working memory may be important among tonal language speakers in the control of attention between speech and music representations of pitch. Yet, given the rarity of AP and the fact that it is often not easily dissociated with musical training (see Baharloo, Service, Risch, Gitschier, & Freimer, 2000), this approach for future research may be difficult to execute.

The present results may also help inform the relationship between auditory working memory and perceptual categorization more generally. Working memory has been previously demonstrated to be an important predictor of perceptual category learning (Lewandowsky, Yang, Newell, & Kalish, 2012), including the explicit training of AP (Van Hedger et al., 2015). For well-established categories, working memory may relate to successful recognition only in more challenging listening situations, such as detecting speech in noise (Zekveld et al., 2013). This kind of framework is also compatible with the idea that certain kinds of experiences may reduce the demands on WM for successful recognition. The notion that increasing expertise with relevant perceptual categories subsequently reduces demands on perceptual processing has been well established in the domain of skill acquisition, extending back to the examination of telegraphers learning to parse incoming Morse code (Bryan & Harter, 1899). In a broader auditory context, active processing models of speech perception (Heald & Nusbaum, 2014; Nusbaum & Schwab, 1986) have argued that the recognition of an incoming sound may be a more controlled process when the relevant perceptual categories are less practiced, in that multiple interpretations must be held in working memory as different sensory attributes are tested against possible representations. This explanation aligns well with the present results, as less music training may require maintaining an incoming sound in working memory for longer as it is compared to multiple potential categories, resulting in a stronger, positive relationship between WM and auditory recognition, as well as slower responses – both of which were observed in the present study.

There are some notable limitations of the present experiment that should be addressed in future work. First, future research would benefit from a larger sample, particularly with respect to AP possessors who are tonal language speakers. This would help address the question of whether tonal language relates to working memory independently of musical expertise in making an AP note category judgment. Second, our selection of stimuli for the Challenging AP Test varied on instrumental familiarity, extremity of octave register, and presumed strength of the fundamental, which makes it hard to disentangle specific attributes which make a note more or less difficult to identify (as well as whether individual factors relate more strongly to particular manipulations, such as familiarity of instrument). Third, the online nature of our experiment allowed broader participant recruitment at the expense of some experimental control. While no participant reported difficulties completing the experiment (e.g., with auditory presentation), future studies in this area would benefit from greater control of the auditory environment (e.g., in a laboratory environment where auditory presentation can be consistent across individuals and where background noise can be minimized).

To conclude, the present experiment demonstrates that note recognition among AP possessors is sensitive to inter-individual differences in musical expertise, tonal language background, and working memory ability, though these factors do not relate to AP in a simple manner. Rather, we found that an individual's tonal language background, in addition to their musical expertise, moderated the relationship between AP recognition and working memory, such that tonal language speakers displayed a stronger relationship between AP performance and working memory compared to non-tonal language speakers. We interpret these findings in a broader framework of auditory recognition, in which cognitive constructs such as working memory may differentially relate to comprehension when the *listening effort* of a given situation is high (Rönnberg et al., 2013). Lower amounts of musical expertise and tonal language experience may increase listening effort through less precise note representations and competing pitch representations, respectively, thus increasing the reliance on auditory working memory for successful categorization of perceptually challenging notes. Overall, the present results highlight how individual factors relate to the variance in note categorization among AP possessors, suggesting that note recognition among AP possessors can be understood as a dynamic auditory skill.

Acknowledgements

This research was supported by the Multidisciplinary University Research Initiatives (MURI) Program of the Office of Naval Research though grant, DOD/ONR N00014-13-1-0205. The stimuli and data can be accessed through the Open Science Framework (10.17605/OSF.IO/WRFBE).

References

Arlinger, S., Lunner, T., Lyxell, B., & Pichora-Fuller, M. K. (2009). The emergence of cognitive hearing science. *Scandinavian Journal of Psychology, 50*(5), 371–384. https://doi.org/10.1111/j.1467-9450.2009.00753.x.

Athos, E. A., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., Freimer, N., & Gitschier, J. (2007). Dichotomy and perceptual distortions in absolute pitch ability. *Proceedings of the National Academy of Sciences of the United States of America, 104*(37), 14795–14800. https://doi.org/10.1073/pnas.0703868104.

Bachem, A. (1955). Absolute pitch. *The Journal of the Acoustical Society of America, 27*(6), 1180–1185.

Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation, 8*, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1.

Baharloo, S., Service, S. K., Risch, N., Gitschier, J., & Freimer, N. B. (2000). Familial aggregation of absolute pitch. *American Journal of Human Genetics, 67*(3), 755–758. https://doi.org/10.1086/303057.

Bahr, N., Christensen, C. A., & Bahr, M. (2005). Diversity of accuracy profiles for absolute pitch recognition. *Psychology of Music, 33*(1), 58–93. https://doi.org/10.1177/0305735605048014.

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

Bryan, W. L., & Harter, N. (1899). Studies on the telegraphic language. The acquisition of a hierarchy of habits. *The Psychological Review, 6*(4), 345–375.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research. Vol. 169*. *Progress in brain research* (pp. 323–338). Elsevier. https://doi.org/10.1016/S0079-6123(07)00020-9.

Deutsch, D. (2002). The puzzle of absolute pitch. *Current Directions in Psychological Science, 11*(6), 200–204.

Deutsch, D. (2013). Absolute Pitch. In D. Deutsch (Ed.). *The psychology of music* (pp. 141–182). (3rd ed.). San Diego, CA: Academic Press. https://doi.org/10.1016/B978-0-12-381460-9.00005-5.

Deutsch, D., & Dooley, K. (2013). Absolute pitch is associated with a large auditory digit span: A clue to its genesis. *The Journal of the Acoustical Society of America, 133*(4), 1859–1861. https://doi.org/10.1121/1.4754917.

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

Dohn, A., Garza-Villarreal, E. A., Riisgaard Ribe, L., Wallentin, M., & Vuust, P. (2014). Musical activity tunes up absolute pitch ability. *Music Perception, 31*(4), 359–371. https://doi.org/10.1525/rep.2008.104.1.92.

Dooley, K., & Deutsch, D. (2010). Absolute pitch correlates with high performance on musical dictation. *The Journal of the Acoustical Society of America, 128*(2), 890–893. https://doi.org/10.1121/1.3458848.

Dooley, K., & Deutsch, D. (2011). Absolute pitch correlates with high performance on interval naming tasks. *The Journal of the Acoustical Society of America, 130*(6), 4097. https://doi.org/10.1121/1.3652861.

Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience, 8*(March), 35. https://doi.org/10.3389/fnsys.2014.00035.

Heald, S. L. M., Van Hedger, S. C., & Nusbaum, H. C. (2017). Understanding sound: Auditory skill acquisition. *Psychology of learning and motivation. Vol. 67*Elsevier Ltd.https://doi.org/10.1016/bs.plm.2017.03.003.

Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2013). Absolute pitch may not be so absolute. *Psychological Science, 24*(8), 1496–1502. https://doi.org/10.1177/0956797612473310.

Jaeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y. F., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning - Implications for training and transfer. *Intelligence, 38*(6), 625–635. https://doi.org/10.1016/j.intell.2010.09.001.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66–71. https://doi.org/10.1037/0033-2909.131.1.66.

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics, 53*(2), 157–165. https://doi.org/10.3758/BF03211726.

Lewandowsky, S., Yang, L.-X., Newell, B. R., & Kalish, M. L. (2012). Working memory does not dissociate between different perceptual categorization tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(4), 881–904. https://doi.org/10.1037/a0027298.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Miyazaki, K. (1989). Absolute pitch identification: Effects of timbre and pitch region. *Music Perception, 7*(1), 1–14.

Miyazaki, K. (1990). The speed of musical pitch identification possessors by absolute-pitch possessors. *Music Perception, 8*(2), 177–188.

Morrison, A. B., & Chein, J. M. (2011). Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychonomic Bulletin and Review, 18*(1), 46–60. https://doi.org/10.3758/s13423-010-0034-0.

Nusbaum, H. C., & Morin, T. M. (1992). Paying attention to differences among talkers. In Y. Tohkura, Y. Sagisaka, & E. Vatikiotis-Bateson (Eds.). *Speech perception, speech production, and linguistic structure* (pp. 113–134). Tokyo: OHM.

Nusbaum, H. C., & Schwab, E. C. (1986). The role of attention and active processing in speech perception. In E. C. Schwab, & H. C. Nusbaum (Eds.). *Pattern recognition by humans and machines: Volume 1, speech perception* (pp. 321). New York, NY: Academic Press.

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

Pisoni, D. B., Nusbaum, H. C., & Greene, B. G. (1985). Perception of synthetic speech generated by rule. *Proceedings of the IEEE, 73*(11), 1665–1676. https://doi.org/10.1109/PROC.1985.13346.

Rönnberg, J., Lunner, T., Zekveld, A., Sörqvist, P., Danielsson, H., Lyxell, B., ... Rudner, M. (2013). The ease of language understanding (ELU) model: Theoretical, empirical, and clinical advances. *Frontiers in Systems Neuroscience, 7*(July), 1–17. https://doi.org/10.3389/fnsys.2013.00031.

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

Van Hedger, S. C., Heald, S. L. M., & Nusbaum, H. C. (2017). Long-term pitch memory for music recordings is related to auditory working memory precision. *The Quarterly Journal of Experimental Psychology,* 1–13. https://doi.org/10.1080/17470218.2017.1307427.

Van Hedger, S. C., Heald, S. L. M., Uddin, S., & Nusbaum, H. C. (2018). A note by any other name: Intonation context rapidly changes absolute note judgments. *Journal of Experimental Psychology. Human Perception and Performance*. https://doi.org/10.1037/xhp0000536.

Vanzella, P., & Schellenberg, E. G. (2010). Absolute pitch: Effects of timbre on note-naming ability. *PLoS One, 5*(11), https://doi.org/10.1371/journal.pone.0015449.

Wagenmakers, E. J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., ... Morey, R. D. (2017). Bayesian inference for psychology. Part II: Example applications with JASP. *Psychonomic Bulletin and Review,* 1–19. https://doi.org/10.3758/s13423-017-1323-7.

Wilson, S. J., Lusher, D., Martin, C. L., Rayner, G., & McLachlan, N. (2012). Intersecting factors lead to absolute pitch acquisition that is maintained in a "fixed do" environment. *Music Perception, 29*(3), 285–296.

Zatorre, R. J. (2003). Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. *Nature Neuroscience, 6*(7), 692–695. https://doi.org/10.1038/nn1085.

Zekveld, A. A., Rudner, M., Johnsrude, I. S., & Rönnberg, J. (2013). The effects of working memory capacity and semantic cues on the intelligibility of speech in noise. *The Journal of the Acoustical Society of America, 134*(3), 2225–2234. https://doi.org/10.1121/1.4817926.

Zhou, B., & Krott, A. (2016). Bilingualism enhances attentional control in non-verbal conflict tasks - Evidence from ex-Gaussian analyses. *Bilingualism: Language and Cognition, 21*(1), 162–180. https://doi.org/10.1017/S1366728916000869.