Stroop-like Effects in Pitch Identification Training and Generalization to Untrained Timbres: Evidence from a Contingency Learning Task

Williams Henry & James R. Schmidt
LEAD – CNRS UMR5022, Université de Bourgogne, Dijon, France

Absolute pitch is the ability to automatically identify and name the pitches of tones without the help of a reference tone. Contrary to the common idea that absolute pitch is almost impossible to acquire after a critical period, some research suggests possible improvements in pitch identification in adulthood. Recently, using a simple incidental contingency learning approach, rapid and robust learning of associations between pitches and note names was observed. In the current work, we explored the item specificity of this learning. In our new task, we used three types of instrument tones (i.e., three timbres). For two timbres, contingencies between tones and notes names were directly manipulated. We then tested whether learning transferred (generalized) to tones from a third timbre, for which contingencies were *not* directly manipulated. Our results indicate clear automatic response biases in response times due to the learned contingencies that transferred from trained to untrained tones. Explicit identification of tones also increased at post-test for both trained and untrained tones. These results demonstrate that learning is not purely instrument specific and that learning of the pitch class is observed. Our results also shed light on the possible underlying representations that participants learn in our paradigm.

Received: December 7, 2023, accepted May 31, 2024.

Key words: absolute pitch, contingency learning, timbre, generalization, incidental learning

In the present work, we explore the learnability of pitch classes in a contingency learning task. In particular, we test: 1) whether participants can rapidly learn the note names of pitches played by the instruments used during training, and 2) whether this learning transfers to the same notes played on new instruments. A *pitch class* is a set of all pitches that share the same "chroma," or pitches that are perceived as sharing the same "color" or "qualia." In most occidental music, for instance, musical pitches are divided into 12 pitch classes: do, do♯, ré, ré♯, mi, fa, fa♯, sol, sol♯, la, la♯, and si in French fixed-do solfège (or the letters C though B in North American notation). More technically, increasing the frequency by $2^{1/12}$ (or about 1.059) increases the pitch by one semitone (e.g., from fa to fa♯) if using equal temperament for the musical tuning. Because pitch perception is periodic, doubling (or halving) the fundamental frequency produces a pitch of the same pitch class. For example, a frequency of 220 Hz corresponds to a "la" (A3), as does a frequency of 440 Hz (A4). These two tones correspond to the same pitch class, separated by one octave. Additionally, musical instruments (typically) do not produce pure sine waves, but rather have instrument-specific resonating frequencies, producing the unique *timbre* of the instrument (see Krimphoff et al., 1994), as discussed in more detail later. Thus, two identical notes played on two different instruments belong to the same pitch class but are not identical. These above mentioned caveats aside, learning to name pitches by ear might, superficially, seem rather simple: there are, after all, only 12 pitch classes to learn. As expanded on in the next section, however, the reverse is true: learning to identify pitches by ear (out of context, at least) is considered to be extraordinarily difficult. After discussing these difficulties, we will discuss some evidence for less conscious forms of pitch detection and our contingency learning task aimed at training pitch detection in musically naïve participants.

Absolute Pitch and the Underlying Cognitive Mechanisms

Absolute pitch (AP) is the rare ability to identify and name the pitch chroma of tones in isolation (Bermudez & Zatorre, 2009; Takeuchi & Hulse, 1993), theoretically across octaves and timbres. A major debate remains concerning its genesis (for reviews, see Deutsch, 2013; Loui, 2016). AP may be either mostly explained by genetics (Baharloo et al., 2000; Theusch et al., 2009) or by a critical period, that is, an optimal age range for early music training to acquire AP, generally

Music Perception, volume 42, issue 3, pp. 207–224, issn 0730-7829, electronic issn 1533-8312. DOI: https://doi.org/10.1525/mp.2024.aa002

operationalized as spanning from the age of 3 to 7 years old (Levitin & Rogers, 2005; Wilson et al., 2012). It is generally accepted that AP may be the result of the interaction between genetic predisposition and early music training (Athos et al., 2007; Theusch & Gitschier, 2011; Wilson et al., 2012; Zatorre, 2003).

While the ability to name the pitches of tones is extremely rare, though somewhat less so in tonal language speakers (Deutsch et al., 2006), and seems remarkable (e.g., Deutsch, 2013; for conflicting views arguing that AP is associated with some downsides, see Marvin et al., 2019; Miyazaki, 2004a), some works suggest that *implicit absolute pitch memory* (APM)—that is, non-conscious long-term AP representations—is widespread throughout the population. For instance, non-absolute pitch (NAP) possessors can recognize with reasonable accuracy the phone dial tone from its pitch-shifted (transposed) counterparts (Smith & Schmuckler, 2008). They can also judge above chance-guessing whether well known musical excerpts are played in the correct or incorrect key (e.g., excerpts pitch-shifted by one semitone; Schellenberg & Trehub, 2003; Van Hedger et al., 2018), and this was found to generalize across timbres and octaves (Van Hedger et al., 2023), suggesting that their judgements can rely on the *chroma* dimension (i.e., the pitch class that groups all notes belonging to the same category; e.g., "do" or C in all octaves and timbres). Although many suggest that APM might be fundamentally different from true AP (Kim & Knösche, 2017), Van Hedger et al.'s (2023) findings suggest otherwise (though the pitch height dimension influenced participants' judgments to some extent).

The two-component model (Levitin, 1994; Levitin & Rogers, 2005), a framework explaining the cognitive expression of AP, argues that AP is explained by two processes: 1) implicit APM, shared by AP to NAP possessors; and 2) pitch labeling, only mastered by AP possessors. The first component is a prerequisite for the second. According to this model, AP possessors possess a pitch template—that is, pitch chroma representations—in which solfeggio labels (e.g., do, ré, mi, etc.) are integrated to these labels that they directly access when a to-be-labeled pitch is compared to the template (Levitin & Rogers, 2005). The two-component model thus explains AP as a *labeling ability* (for another account explaining AP as a labeling ability see Itoh et al., 2005; Matsuda et al., 2019). Moreover, some have argued that the second component is not supported by an AP-specific mechanism but might involve a simple conditional associative retrieval process (Bermudez & Zatorre, 2005; Zatorre et al., 1998).

AP may be developed during the critical period via simple environmental factors that trigger sensibility to regularities and may be shaped by implicit or associative learning (Bermudez & Zatorre, 2005; Marvin et al., 2020). For instance, via active music training involving consistent pitch-tone mappings (i.e., fixed-do pedagogy and an instrument with which to practice; Wilson et al., 2012), and through mere exposure (Simpson & Huron, 1994). Consistent with this, evidence for the maintenance or the updating of AP possessors' template suggests a role of implicit reinforcement through similar factors (Hedger et al., 2013; Marvin et al., 2020; Wilson et al., 2012). These latter findings might serve as basis for a learning account of AP in adulthood (see Heald et al., 2017; Van Hedger et al., 2019).

AUTOMATICITY AND THE ABSOLUTE PITCH STROOP EFFECT

Contrary to the traditional view that either one possesses AP or not (Athos et al., 2007), pitch naming ability may be more continuously distributed (Bairnsfather et al., 2022; Bermudez & Zatorre, 2009; Van Hedger et al., 2020). In fact, while AP possessors exhibit high scores on pitch naming tests and NAP possessors score at or near chance guessing, some argue that most musicians fall between these two extremes (i.e., below the threshold for AP but above chance guessing), often termed quasi-absolute pitch (QAP) possessors (Hansen & Reymore, 2023). Depending on their level of automaticity (see Bairnsfather et al., 2022; Hansen & Reymore, 2023; Wilson et al., 2009), some QAP possessors can name a subset of pitches of the chromatic scale absolutely (Bairnsfather et al., 2022; Wilson et al., 2009). QAP possessors are also able to hold a well known pitch in working memory and compare it with a to-be-labeled pitch, then by inferring the interval between them (i.e., relative pitch processing), they may label the tone (Bairnsfather et al., 2022; Wilson et al., 2009). Without time constraints in a pitch naming test, these participants (as any relative pitch possessors) can reach thresholds of correct identification typically used to operationalize AP (e.g., 90% accuracy; see Levitin & Rogers, 2005). Therefore, response times (RTs) are also of importance to better differentiate different classes of pitch naming ability (see Bermudez & Zatorre, 2009; Van Hedger et al., 2020). Intuitively, the mark of the automaticity in identifying the chroma of a tone absolutely is fast RTs (Wilson et al., 2009). This automaticity should have the further effect that AP possessors cannot prevent labeling (e.g., when trying to ignore pitches), which has been evidenced with auditory Stroop procedures, described next.

The so-called Stroop effect or congruency effect (see MacLeod, 1991, for a review), is the observation of impaired (slower and/or less accurate) ink color identification of color words when both dimensions are incongruent (e.g., the color word GREEN printed in red) relative to when they are congruent (e.g., RED printed in red). This effect is taken as evidence for the automaticity of word reading: though the task instructions are to ignore the word, participants cannot help but be influenced by it. Analogous to the color-word Stroop effect, an auditory musical Stroop effect is observed in conceptually similar auditory Stroop paradigms with AP possessors (e.g., Akiva-Kabiri & Henik, 2012; Itoh et al., 2005; Miyazaki, 2000, 2004b). For instance, in Miyazaki (2004b), accurate AP, inaccurate AP, and NAP possessors heard sung pitch tones. The target stimuli were not the pitches themselves but rather the fixed-do solfeggio syllables (e.g., "fa") that were sung. These were either congruent with the pitches of tones (e.g., the syllable "do" sung as a "do"/C) or incongruent (e.g., "do" sung as a "ré"/D). The participants' task was simply to repeat the syllables of sung tones without paying attention to the pitches of the tones (e.g., repeat the syllable "do" regardless of whether the pitch of a tone was "do"/C or another pitch). The congruency effect increased as a function of AP naming ability, with no evidence for such an effect among NAP possessors. This effect was due to differences in RTs on incongruent trials between groups of participants. This suggests that AP possessors automatically "translate" the pitches and find it difficult to avoid naming them (i.e., rather than repeating the sung note name). This is again coherent with the notion that AP is an automatic labeling ability.

INCIDENTAL CONTINGENCY LEARNING AND PITCH NAMING TRAINING

It is well known that the human cognitive system is highly sensitive to statistical regularities or contingencies in the environment, and humans learn them. There are many ways to study the learning of statistical correlations (contingencies) between stimuli and responses (Schmidt, 2012). One example, which inspired the current line of research, is the color-word contingency learning paradigm (Schmidt et al., 2007; for reviews, see MacLeod, 2019; Schmidt, 2021). In this paradigm, participants identify the ink color of a neutral word with a corresponding key while ignoring the (distracting) word, similar to the Stroop task. Critically, each word is presented most often in one color and less often with the others (e.g., the word "TABLE" is presented most often in blue, and rarely in red and green). Although the word is task-irrelevant, participants are faster and more accurate on *high contingency* trials, where the word is presented in the expected color (e.g., TABLE in blue), relative to *low contingency* trials, where the word is presented in an unexpected color (e.g., TABLE in red). This indicates that participants learn the regularities/contingencies between words and colors/responses (Schmidt et al., 2007, 2010; Schmidt & De Houwer, 2016).

With this paradigm, learning is mostly incidental (i.e., learning without intention to do so). Although naïve of the manipulation, participants show robust effects (Schmidt et al., 2007), though explicit instructions to learn the contingencies can boost learning effects, probably due to attentional biases toward to the predictive dimension (Schmidt & De Houwer, 2012a, 2012b). Indeed, the primary advantage of incidental learning procedures may not be due to the non-intentional nature of learning, but rather that participants can see a very large number of stimulus pairings in a very short period of time. That is, tasks that require *deliberate* learning, almost by nature, require more time per trial, such that even if learning is boosted *per trial* (e.g., due to attentional biases), such training might be less efficient for a given *duration* of training. Related to this and most importantly for the present purposes, contingency learning occurs very rapidly and is already observed early in the experiment, for instance, with an effect already present and significant within the first block of 18 to 48 trials (Lin & MacLeod, 2018; Schmidt et al., 2010, Schmidt & De Houwer, 2016). A wide range of similar laboratory tasks exist (e.g., Lewicki et al., 1988), but some work has applied a similar logic to music-related skills, such as note position identification and execution, a component of sight-reading (Iorio et al., 2023; Schmidt et al., 2023). Thus, incidental learning procedures seem well suited for rapidly automatizing complex skills.

Given the rapid learning elicited in incidental contingency learning tasks like those mentioned above, Iorio and colleagues (2024) aimed to train principally nonmusician participants to strengthen their *pitch identification* abilities in a conceptually similar task, as measured by both automatic effects during learning and by explicit pitch naming performance. Their investigation was motivated by recent reports that found significant improvements in AP training in adulthood (e.g., Van Hedger et al., 2019; Wong, Lui, et al., 2020; Wong, Ngan, et al., 2020). Iorio et al. (2024) used the seven tones from a C Major scale (i.e., tones corresponding to the white keys on the piano) with pure sinewaves. These tones were used as distractors and corresponding printed French note names were used as targets. On each trial, the participants' task was to identify the note name appearing on the screen via a corresponding key on a computer keyboard while ignoring the auditory stimulus. Each tone was played most often with its corresponding note name (e.g., the tone for "do"/C was played 54 times out of 60 with the note name "do"), and less often with the other note names (e.g., "do"/C played one time out of 60 with "ré," "mi," "fa," "sol," "la," and "si"). During the learning task, participants responded faster on congruent[1] trials (e.g., the pitch for "do"/C presented with the note name "do") than on incongruent trials (e.g., "do"/C with "mi"). This contingency or congruency effect indicates that participants incidentally learned the association between pitches and their corresponding labels, and that this knowledge was sufficiently automatic that the auditory distracting stimulus influenced note name identification. In a subsequent explicit pitch labeling task, participants improved their accuracy in identifying pitches at post-test relative to pre-test, and the post-training identification was well above chance guessing. This learning is also *not* merely short term. Results revealed robust maintenance of initial learning in a surprise post-test pitch identification task one-week after learning. Furthermore, with a modified version of this paradigm using perfect contingencies (i.e., each pitch *always* played with its correct label), Henry et al. (2024) found that explicit pitch identification is improved when all twelve tones of the chromatic scale (from one octave) are trained. The same one-week retention effect was also observed (see also, Van Hedger et al., 2015).

Perhaps the most interesting detail of the studies of Iorio et al. (2024) and Henry et al. (2024) is that the training phases were very short, around 20 minutes. The idea that some improvement in pitch naming is possible is not necessarily that controversial, but large and robust improvements in such a small timeframe in naïve adults seem counter to the standard narrative surrounding the difficulty of pitch identification capability. Of course, we do not necessarily suggest that this (or any) task would allow an NAP possessor to train themselves to the strict level of identification characteristic of true AP. Rather, we expect only initial improvements, with automatic response biases during learning (e.g., congruent response times faster than incongruent response times) and increased accuracy in explicit pitch identification at post-test relative to pre-test (including above-chance performance at post-test). However, these results demonstrate that such incidental learning tasks produce rapid learning. Further, this learning is potent enough to produce automatic influences on behavior. That is, even though participants are given the explicit goal to *ignore* the tones in some of the tested groups, said tones nevertheless impact note name identification and in a rapid enough way to produce an effect analogous to an AP Stroop effect.

THE CURRENT STUDY

AP is the ability to label the *pitch chroma* of a tone. Note, however, that some AP possessors may rely on other auditory cues (such as timbre) to facilitate pitch identification. For example, many AP possessors identify piano tones faster and more accurately than synthetized complex tones (Miyazaki, 1989) and pure tones (Lockhead & Byrd, 1981; Miyazaki, 1989), sometimes referred to as "absolute piano" (Ward & Burns, 1982). This is probably due to high familiarity with the piano timbre by most music students (Miyazaki, 1989). Similarly, Schlemmer et al. (2005) found that pitches are identified more quickly and with less errors with familiar timbres relative to unfamiliar ones, not necessarily just with piano tones (Levitin, 2004; see Marvin & Brinkman, 2000; Reymore & Hansen, 2020). Detecting the pitch of sung notes is particularly difficult (Vanzella & Schellenberg, 2010). Moreover, facilitative effects of timbre on pitch identification can also be seen with other forms of pitch identification, like QAP (Wilson et al., 2009) and instrument-specific AP (Reymore & Hansen, 2020). The latter form is also characterized by above-chance guessing performance for tones of one's primary instrument relying on mechanisms that could involve timbral cues and motor imagery (Reymore & Hansen, 2020; see Hansen & Reymore, 2023). In brief, these well established results seem to argue that stimulus specificity occupies a significant place within the wide range of pitch identification ability.

Similarly, previous studies that claimed to train participants to acquire (to some extent) contingency knowledge between pitches and corresponding note names only used sinewave (Iorio et al., 2024), or piano tones (Henry et al., 2024). Therefore, although clear learning effects were observed, it might be argued that said learning effects were exclusively due to the learned contingencies between specific auditory stimuli and note names, rather than to learning of actual pitch classes. For example, learning might not generalize to the same notes played in an untrained timbre or octave.

[1] We use the terms "congruent" and "incongruent" throughout the rest of the manuscript. Note that congruent pairings are high contingency and incongruent trials are low contingency. Musically naïve participants, of course, learn the contingencies (i.e., the stimuli are not initially perceived as congruent or incongruent), but we use the labels "congruent" and "incongruent" to facilitate comprehension of the text.

Thus, the underlying representation of the acquired knowledge might be stimulus-specific and might not purely rely on the chroma dimension (the latter of which is the hallmark of true AP). Indeed, pitch is not necessarily processed completely independently of timbre. For example, changes in the brightness of the timbre affect pitch perception (Allen & Oxenham, 2014). It is therefore possible that learning of pitches in one timbre generalizes very poorly to another. A more extreme possibility is that learning with our type of task is entirely instrument specific and does not generalize *at all* to other timbres. If this were the case, of course, it would make our training procedure relatively uninteresting for practical usage in learning AP. To explore this issue, we asked whether automatic response biases due to the associative learning between a subset of pitches and corresponding note names transfers or generalizes to pitches played on another instrument (i.e., timbre), for which participants were not trained. In ongoing work, we are exploring a similar question regarding transfer of learning across octaves (see General Discussion).

In the present report, we focus on the underlying representations learned by participants and evaluate transfer of learning across timbres. We, therefore, used a modified version of the auditory contingency learning paradigm, and used the seven tones from a C Major scale recorded on three instruments (piano, clarinet, and harpsichord). For the tones of two instruments (e.g., piano and clarinet), contingencies between tones and note names were directly manipulated, which we call the *context tones*. For the tones of the third instrument (e.g., harpsichord), no contingencies were directly manipulated between tones and note names, which we call the *transfer tones*. The main goal of the current work is to assess whether learning with context tones generalizes to transfer tones; that is, whether a congruency effect will be found for transfer tones. This would be consistent with learning of a pitch class, rather than purely stimulus-specific learning in our paradigm. Thus, two different major predictions might be proposed. According to a stimulus-specific account of learning, the congruency effect should be restricted to the context tones; that is, observed only for the timbre dimensions for which contingencies between tones and note names are manipulated (i.e., context but not transfer tones). A timbre-independent account of learning, however, would predict not only a congruency effect for the context tones, but also that the congruency effect generalizes to timbres that are not directly trained (i.e., transfer tones).

To evaluate these two conflicting accounts, the experiment was divided into four blocks. During the first block, participants were presented with two types of instrument tones (i.e., two timbres) for which contingencies were manipulated (i.e., context tones only). After the first block, transfer tones were introduced. The first block was included with the thinking that this would allow participants to first learn the pitch class before introducing transfer stimuli. We also used two timbres for the context stimuli as this may induce sufficient variability in the timbral dimension to encourage learning of what all stimuli of the same pitch class share in common (e.g., the fundamental frequency) thereby making it more likely that a congruency effect will emerge for transfer tones (see Wong, Lui et al., 2020).

Method

PARTICIPANTS

One hundred and twenty participants took part in this experiment, 108 (44 males, 64 females; mean age = 38.17 years, $SD = 11.64$) of which were retained for the final sample (see exclusion criteria below). They were recruited on prolific.co and performed the experiment online. This site is headquartered in the UK but has participants from all over the world. A participation requirement was English as native language. They were paid 4.50£ for their participation. Prior to the beginning of the experiment, each participant provided informed consent, and their anonymization was guaranteed. According to our a priori established criteria, five participants were removed from the analysis because they responded with less than 80% correct responses during the main contingency learning task, suggesting poor focus on the task (e.g., Schmidt & De Houwer, 2016). Two additional participants with hearing impairments were removed. Three participants did not complete the experiment and did not appear in our final sample. Recruitment advertisements mentioned that to take part in the current study, participants should not be AP possessors. However, some self-reported AP possessors took part in the experiment (14.55% of the sample). One of them was removed because their pre-test accuracy (see Design and Procedure section below) was 66.67%. One additional participant was removed, because, although not a self-reported AP possessor, their accuracy was 71.43% at both pre-test and post-test. In the sample as a whole, pre-test accuracy did not differ between self-reported AP and self-reported NAP possessors, $t(108) = -0.77$, $p = .441$, $d_p = -0.21$, 95% CI [-0.74, 0.32], $BF_{10} = 0.35$,[2] seemingly indicating

[2] The Bayes factor calculation is explained in the Data Analysis section below.

that most of those indicating AP either grossly overestimated their pitch naming ability or that they misunderstood the question. After removing the two participants that were accurate at pre-test, the difference between groups remained nonsignificant, $t(106) = -1.64$, $p = .104$, $d_p = -0.46$, 95% CI [-1.00, 0.09], $BF_{10} = 0.84$, though the NAP group exhibited numerically higher scores ($M_{AP} = 18.41\%$, $SE = 2.22$; $M_{NAP} = 23.66\%$, $SE = 1.23$). With these consideration in mind, we considered that the remaining sample were NAP possessors.

Among the remaining participants, some who self-reported as having had past or current musical activities were kept in our final sample ($n = 5$). Although we wanted to test participants that were maximally naïve in pitch detection (thus the pre-screening question), these participants were kept for two reasons. First, while we did not collect any information about the age of onset in music training, it was previously found that this (continuous) variable does not explain any variance in the contingency effect when musicians are trained with the above-mentioned AP contingency learning task (Iorio et al., 2024). Second, the more critical screening was to exclude participants from the sample that already show evidence of AP, which was not the case for these five participants. In any case, removing these participants does not change the significance or interpretation of the main results.

MATERIALS

The experiment was programmed and run with Psytoolkit (Stoet, 2010, 2017). Some of the auditory stimuli from Van Hedger and colleagues (2019) were used, comprising the seven tones from a C Major scale spanning from C4 to B4, which were recorded on a piano, a clarinet, and a harpsichord. A total of 21 auditory stimuli were used, each having a file duration (played in full during the experiment) of 1,000 ms. Visual stimuli were the seven solfeggio labels used in Romance languages to indicate the currently used tones (i.e., do, ré, mi, fa, sol, la, and si), and were presented in lowercase 30 pts MS Reference Sans Serif font. Responses were made on a QWERTY keyboard. The note names from "do" to "si" were mapped to the W to I keys, respectively.

DESIGN AND PROCEDURE

All phases involved in this experiment are presented in Figure 1. Before performing the experiment, participants completed a brief survey regarding their past and current musician activities, AP status, and hearing impairments. Next, the experiment began with two note name identification practice phases. In the first practice phase, each note name was presented 10 times in a random order, for a total of 70 practice trials. Each trial began with a fixation cross in the middle of the screen for 500 ms followed by a note name. Correct note name identifications triggered the next trial. In case of incorrect identifications, note names changed from black to red, and the next trial began once participants identified them correctly. There was no time limit to respond. During the first practice phase, note reminders (i.e., do, ré, mi, fa, sol, la, si) were presented at the bottom of the screen to help participants remember the key mappings. These note reminders corresponded spatially to the keys for each note name (i.e., "do" on the left through to "si" on the right). The second practice phase, also 70 trials, was almost identical to the first. The only difference was that the reminders were no longer presented in the second practice phase and participants were encouraged to respond from memory. It is important to note that these two phases were designed to automatize the note name to key correspondences, therefore, no musical tones were played.

The subsequent phase was an explicit pitch identification task. In this phase, on each trial, participants were presented with a tone and were required to identify it with a corresponding key. The seven tones from the C Major scale were presented. Each of the 21 tones (i.e., the 7 tones for each of the 3 timbres) was presented one at a time in a randomized order. There was no time limit, but participants were invited to respond as accurately and rapidly as possible (we encouraged speed in order to measure both improvements in pitch identification accuracy after learning, in addition to faster and automatic identification). The response was followed by 500 ms of silence, followed by white noise for 1,000 ms, and again 500 ms of silence before the next trial began.

The fourth phase was the main contingency learning task. Contingencies or congruencies between the tones and printed note names were manipulated (congruent vs. incongruent). The type of item was also manipulated (context tones vs. transfer tones). Context tones comprised two types of instrument tones (e.g., piano and clarinet tones), representing 14 tone stimuli, and transfer tones comprised one type of instrument tones (e.g., harpsichord tones), representing 7 tone stimuli. Which of the three timbres served as the transfer tones was counterbalanced across participants via random assignment. More precisely, piano and clarinet tones were used as context tones in Group 1, and harpsichord tones as transfer tones. In Group 2, piano and harpsichord tones were used as context tones, and clarinet tones were used as transfer tones. In Group 3, clarinet and

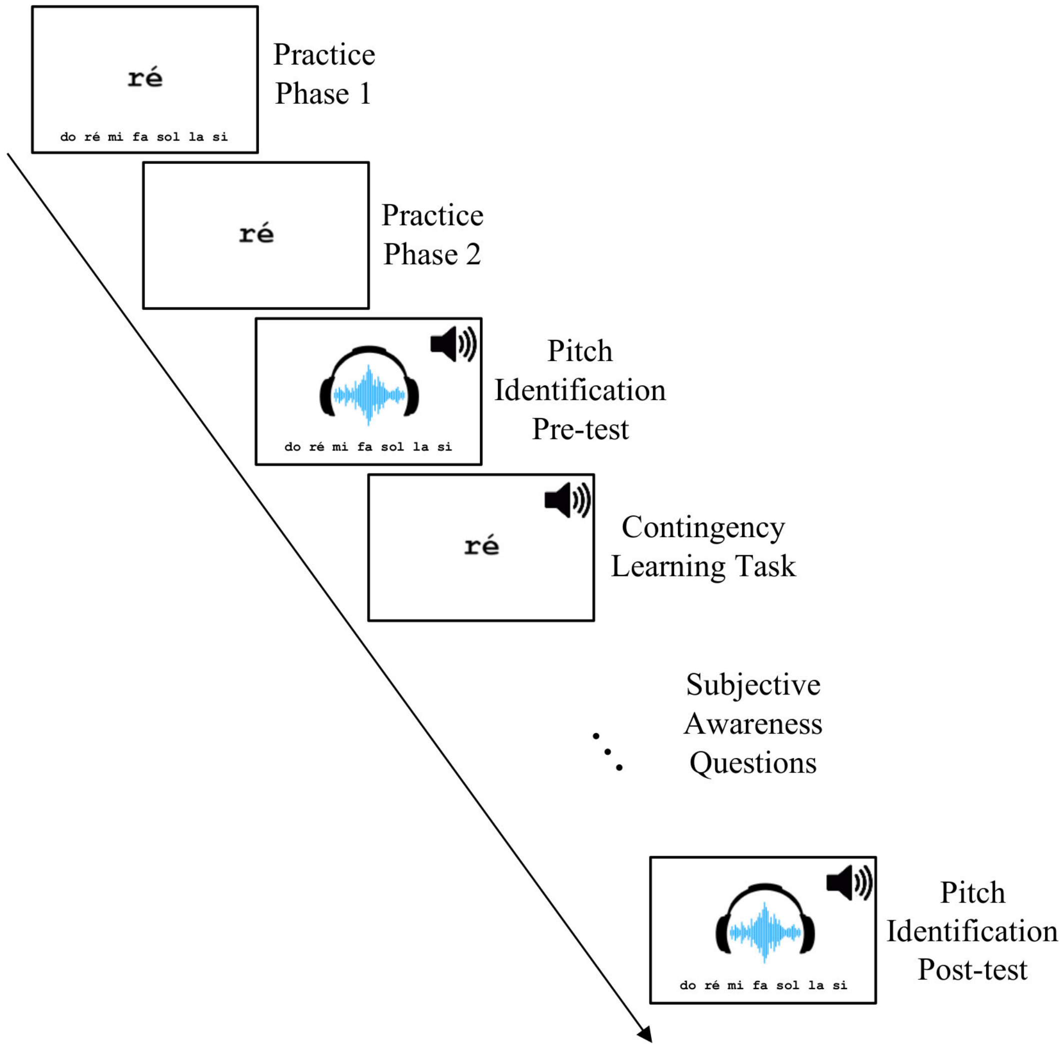

FIGURE 1. Phases of the current experiment. *Note.* Stimuli not to scale.

harpsichord tones were used as context tones, and piano tones were used as transfer tones.

We note that, for brevity, we do not include this counterbalancing factor in the analyses reported below, but we note that including this factor in our analyses did not modify the results in any way (see Supplementary Materials accompanying this article at online.ucpress.edu/mp). For context tones, congruent pairings between each instrument tone and their respective note names (e.g., the note for "do"/C with the note name "do") were presented more frequently than incongruent pairings (e.g., the note for "do"/C with the note name "fa"). Therefore, for context tones, contingencies were directly manipulated, and each tone was predictive of its corresponding note name/correct response. Given past reports demonstrating larger and more robust contingency effects with a higher proportion of high contingency (in this context: congruent) pairings (Forrin & MacLeod, 2018; Schmidt et al., 2023), tones were presented 90% of the time with the congruent note name for context tones. Note that tones cannot be presented 100% of the time with the congruent note name if one wishes to measure learning effects in the learning phase (e.g., in response times or error rates), as it is necessary to include some meaningful number of incongruent pairings in order to measure learning (e.g., incongruent – congruent response times). Conversely, for transfer tones, each tone-name pairing was presented equally often. That is, each tone was presented equally often with all note names, corresponding to a rate of around 14.29% congruent trials. As such, contingencies between tones and note names were *not*

TABLE 1. ***Relative Pairings Between Tones and Note Names for Context and Transfer Tones***

	Auditory stimuli													
	Context tones							Transfer tones						
Displayed note names	do	ré	mi	fa	sol	la	si	do	ré	mi	fa	sol	la	si
Do	**54**	1	1	1	1	1	1	**3**	3	3	3	3	3	3
Ré	1	**54**	1	1	1	1	1	3	**3**	3	3	3	3	3
Mi	1	1	**54**	1	1	1	1	3	3	**3**	3	3	3	3
Fa	1	1	1	**54**	1	1	1	3	3	3	**3**	3	3	3
Sol	1	1	1	1	**54**	1	1	3	3	3	3	**3**	3	3
La	1	1	1	1	1	**54**	1	3	3	3	3	3	**3**	3
Si	1	1	1	1	1	1	**54**	3	3	3	3	3	3	**3**

Note. Congruent pairings are presented in bold.

directly manipulated for transfer tones, and each tone was not predictive of any response.

The main contingency learning task was divided into four blocks. During the first block, only context tones were played. Participants completed 189 congruent trials (selected randomly from the list of 14 congruent stimuli) and 21 incongruent trials (selected randomly from the list of 84 incongruent stimuli), randomly intermixed. Transfer tones were added to the last three blocks. In each of these last three blocks, participants completed 63 congruent and 7 incongruent context trials in addition to 7 congruent and 42 incongruent transfer trials (i.e., each possible stimulus pairing once). In all, participants completed 567 trials. The relative trial frequencies for context and transfer tones are presented in Table 1.

On each trial, participants had to identify a displayed note name with the corresponding key and were required to respond as accurately and rapidly as possible. Each trial began with the tone. A note name appeared on the screen 250 ms after the tone onset and participants had 3,000 ms to respond. We note that the auditory stimulus was presented slightly before the target note name because previous work with conceptually related (but nonmusical) tasks have indicated that this advanced presentation boosts learning, probably because it gives the predictive stimulus a processing head start (Forrin & MacLeod, 2017; Schmidt, 2018; Schmidt & De Houwer, 2016), allowing it to bias responding to the target to a greater degree. Correct responses were followed by a blank screen for 500 ms, followed by the next trial. If participants responded incorrectly or failed to respond in 3,000 ms, the note name was replaced by "XXX" printed in red on the middle of the screen for 500 ms, followed by a blank screen for 500 ms before the next trial began.

After the learning phase, participants responded to two subjective awareness questions. Precisely, participants were first asked whether they noticed the regularities between tones and note names with the following question:

> *During the fourth part of the current experiment, note names were presented with tones. More precisely, each tone was presented most often with a note name and less often with the others. That is to say that one sound was presented most often with the note name "do", another sound most often with the note name "ré", etc . . . Did you notice those regularities?*

Next, participants were asked whether they noticed that some tones, albeit different regarding their timbral quality, sounded similar (i.e., shared the same qualia/pitch) with the following question:

> *During the task, tones were musical notes. You heard seven different musical notes (do, ré, mi, fa, sol, la, and si), played on three different instruments (piano, clarinet, and harpsichord). When two different instruments play the same note (e.g., "do"), they sound similar but not quite the same. Did you notice that some tones were the same (i.e., sounded similar) even though they were played on different instruments?*

The last phase consisted of an explicit pitch identification task that was identical in all respects to the pre-test in the third phase, described earlier. This last phase allows us to assess improvements in explicit pitch identification ability relative to pre-test (e.g., rather than purely automatic biases in RT measures). It can also be described as a measure of objective awareness of the acquired knowledge during the main contingency learning task.

DATA ANALYSES

Analyses were performed in RStudio (4.2.2). For the learning phase, analyses were conducted on response times and error rates. For RT analyses, only trials where the participant responded correctly were considered. Response times faster than 150 ms were excluded as anticipations and response times slower than 3,000 ms were not possible (i.e., due to the response deadline). Further analyses were performed on accuracy scores in the explicit pitch identification task, in addition to response times.

Though not the primary aim of the present work, it is relevant to determine to what extent implicit learning contributed to the congruency effect during the main contingency learning task. To test this, linear models were run on each tone type (context and transfer) separately, with the RT congruency effect (incongruent – congruent) as the dependent variable and rates of objective awareness at post-test as a predictor variable. Objective awareness was re-centered at chance guessing (i.e., $\approx$ 14.29%), such that the intercept indicates the size of the RT congruency effect when participants are guessing at chance levels at post-test (i.e., no conscious awareness).

Effect sizes for ANOVA results are reported as η_p^2 (Cohen, 1973), and 90% confidence intervals are also provided (Kelley, 2007). Complementarily, we ran Bayesian ANOVAs (Morey & Rouder, 2022) and estimated the "Bayes factor inclusion" corresponding to each factor and interaction assessed in each analysis (Keysers et al., 2020). This estimate reveals the relative evidence of models including the factor of interest to better explain the data compared to models that do not include it (Makowski, Ben-Shachar et al., 2019; Makowski, Ben-Shachar & Lüdecke, 2019). For *t*-tests, we report Cohen's *d* (Cohen, 1988; Goulet-Pelletier & Cousineau, 2018) and 95% confidence intervals (Cousineau & Goulet-Pelletier, 2021; Fitts, 2021; Steiger & Fouladi, 2016). Bayes *t*-tests are also reported as BF_{10} indicating by how much the data are more likely under the alternative hypothesis relative to the null. Raw data, data analysis scripts, and experiment files are available in the Open Science Framework repository: https://osf.io/h85gd/.

Results

CONTINGENCY LEARNING PHASE

Response Times

Response time data are depicted in Figure 2. An initial 2 × 2 within-subjects ANOVA on RTs with congruency (congruent vs. incongruent) and item type

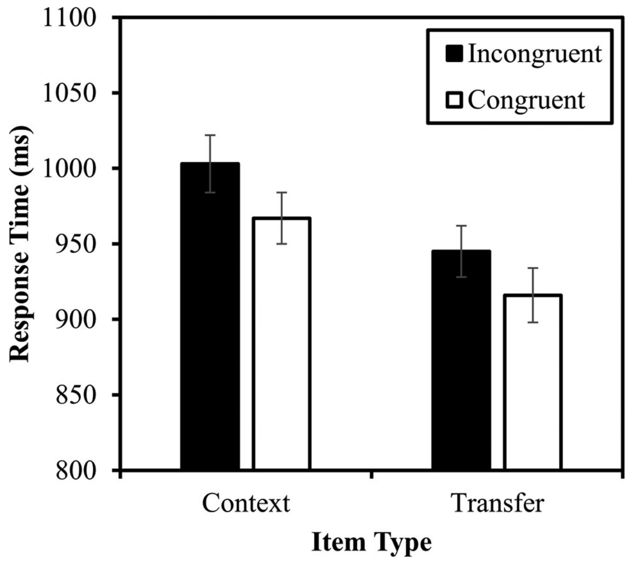

FIGURE 2. Mean response times in the learning phase for congruent and incongruent trials by item type with standard error bars.

(context vs. transfer) as independent variables was conducted. The analysis revealed a significant main effect of congruency, $F(1, 107) = 32.54$, $p < .001$, $\eta_p^2 = .233$, 90% CI [.124, .338], $BF_{incl} > 1000$, indicating that participants were faster on congruent trials ($M = 941$, $SE = 17$) relative to incongruent trials ($M = 974$, $SE = 17$). The main effect of item type was also significant, $F(1, 107) = 64.25$, $p < .001$, $\eta_p^2 = .375$, 90% CI [.256, .472], $BF_{incl} > 1000$, with faster responses to transfer tones ($M = 930$, $SE = 17$) relative to context tones ($M = 985$, $SE = 18$).[3] The interaction between the two factors was not significant, $F(1, 107) = 0.69$, $p = .409$, $\eta_p^2 = .006$, 90% CI [.000, .053], $BF_{incl} = 0.029$. Despite the lack of an interaction, we are particularly interested in the individual effect for context and transfer tones. The congruency effect was significant both for context tones, $t(107) = 5.14$, $p < .001$, $d_D = 0.49$, 95% CI [0.29, 0.69], $BF_{10} > 1000$, ($M = 37$, $SE = 7$) and for transfer tones, $t(107) = 3.91$, $p < .001$, $d_D = 0.38$, 95% CI [0.18, 0.57], $BF_{10} = 114.25$, ($M = 29$, $SE = 7$), though the latter effect was numerically (albeit not significantly) smaller.

[3] As a minor aside, the main effect of item type is simply explained by the fact that context tones were presented throughout the entire learning phase (including the initial blocks where responding is slow to a novel task), whereas transfer tones were only added in later blocks. Indeed, the main effect of item type is no longer significant when the first block of learning is removed from the analysis, $F(1, 107) = 2.43$, $p = .122$, $\eta_p^2 = .022$, 90% CI [.000, .086], $BF_{incl} = 0.050$.

Error Rates

The same ANOVA on the mean percentage of errors revealed no main effect of congruency, $F(1, 107) = 3.85$, $p = .052$, $\eta_p^2 = .035$, 90% CI [.000, .107], $BF_{incl} = 0.388$, though the means were numerically in the expected direction ($M_{high} = 5.59\%$, $SE = 0.40$; $M_{low} = 6.23\%$, $SE = 0.39$). The main effect of item type was significant, $F(1, 107) = 28.59$, $p < .001$, $\eta_p^2 = .211$, 90% CI [.105, .315], $BF_{incl} > 1000$, with transfer tones responded to more accurately ($M = 4.81\%$, $SE = 0.38$) relative to context tones ($M = 7.01\%$, $SE = 0.46$).[4] The interaction was not significant, $F(1, 105) = 1.55$, $p = .215$, $\eta_p^2 = .014$, 90% CI [.000, .071], $BF_{incl} = 0.050$.

TEST PHASES

Subjective Awareness

In response to the first of two subjective awareness questions, 43% of the participants indicated that they noticed the regularities in the experiment. We submitted our data to a 2 × 2 × 2 mixed ANOVA on RTs with congruency (congruent vs. incongruent), item type (context vs. transfer), and awareness (aware vs. unaware) as independent variables. Unsurprisingly, the main effect of congruency was significant, $F(1, 106) = 30.68$, $p < .001$, $\eta_p^2 = .224$, 90% CI [.116, .330], $BF_{incl} > 1000$, as was the main effect of item type, $F(1, 106) = 62.50$, $p < .001$, $\eta_p^2 = .371$, 90% CI [.251, .469], $BF_{incl} > 1000$, but the main effect of awareness was not significant, $F(1, 106) = 2.97$, $p = .088$, $\eta_p^2 = .027$, 90% CI [.000, .095], $BF_{incl} = 1.10$. Consistent with the previous analysis, the interaction between congruency and item type was not significant, $F(1, 106) = 0.99$, $p = .321$, $\eta_p^2 = .009$, 90% CI [.000, .061], $BF_{incl} = 0.189$. Neither the interaction between congruency and awareness, $F(1, 106) = 0.33$, $p = .567$, $\eta_p^2 = .003$, 90% CI [.000, .043], $BF_{incl} = 0.167$, nor the interaction between item type and awareness, $F(1, 106) = 0.01$, $p = .919$, $\eta_p^2 = .000$, 90% CI [.000, .009], $BF_{incl} = 0.142$, nor the three-way interaction, $F(1, 106) = 1.39$, $p = .241$, $\eta_p^2 = .013$, 90% CI [.000, .069], $BF_{incl} = 0.239$, were significant.

For those who were subjectively aware, the congruency effect was positive ($M = 39$, $SE = 11$) and significant for context tones, $t(45) = 3.59$, $p < .001$, $d_D = 0.53$, 95% CI [0.22, 0.83], $BF_{10} = 34.93$. For transfer tones, the congruency effect was positive ($M = 19$, $SE = 10$), but failed to reach significance, $t(45) = 1.98$, $p = .054$, $d_D = 0.29$, 95% CI [-0.005, 0.58], $BF_{10} = 0.94$. For those who were unaware of the contingencies, the congruency effect was positive and significant for context tones ($M = 35$, $SE = 10$), $t(61) = 3.67$, $p < .001$, $d_D = 0.47$, 95% CI [0.20, 0.73], $BF_{10} = 49.14$, and also for transfer tones ($M = 37$, $SE = 11$), $t(61) = 3.39$, $p = .001$, $d_D = 0.43$, 95% CI [0.17, 0.69], $BF_{10} = 22.06$.

For the second subject awareness question, 72% of the participants indicated that they noticed the similarity between notes belonging to the same pitch class. Again, we conducted a 2 × 2 × 2 mixed ANOVA on RTs with congruency (congruent vs. incongruent), item type (context vs. transfer), and awareness (aware vs. unaware) as independent variables was conducted. The main effect of congruency was significant, $F(1, 106) = 22.09$, $p < .001$, $\eta_p^2 = .172$, 90% CI [.075, .276], $BF_{incl} > 1000$, as was the main effect of item type, $F(1, 106) = 52.19$, $p < .001$, $\eta_p^2 = .330$, 90% CI [.211, .431], $BF_{incl} > 1000$, the main effect of awareness was not significant, $F(1, 106) = 0.08$, $p = .774$, $\eta_p^2 = .001$, 90% CI [.000, .028], $BF_{incl} = 0.503$. The interaction between congruency and item type was not significant, $F(1, 106) = 0.08$, $p = .782$, $\eta_p^2 = .001$, 90% CI [.000, .027], $BF_{incl} = 0.182$. Regarding interactions involving the awareness factor, neither the interaction between congruency and awareness, $F(1, 106) = 0.84$, $p = .362$, $\eta_p^2 = .008$, 90% CI [.000, .057], $BF_{incl} = 0.255$, nor the interaction between item type and awareness, $F(1, 106) = 0.03$, $p = .864$, $\eta_p^2 = .000$, 90% CI [.000, .018], $BF_{incl} = 0.166$, nor the three-way interaction, $F(1, 106) = 1.10$, $p = .296$, $\eta_p^2 = .010$, 90% CI [.000, .063], $BF_{incl} = 0.302$, were significant.

For those who were subjectively aware of these similarities, the size of the congruency effect was positive ($M = 43$, $SE = 8$) and significant for context tones, $t(77) = 5.09$, $p < .001$, $d_D = 0.58$, 95% CI [0.34, 0.82], $BF_{10} > 1000$, and for transfer tones ($M = 30$, $SE = 9$), $t(77) = 3.45$, $p < .001$, $d_D = 0.39$, 95% CI [0.16, 0.62], $BF_{10} = 26.20$. For those who were unaware of the similarity between tones of the same pitch chroma, the congruency effect was positive, but not significant for both context ($M = 21$, $SE = 13$), $t(29) = 1.56$, $p = .130$, $d_D = 0.28$, 95% CI [-0.08, 0.65], $BF_{10} = 0.57$, and transfer tones ($M = 28$, $SE = 15$), $t(29) = 1.85$, $p = .074$, $d_D = 0.34$, 95% CI [-0.03, 0.70], $BF_{10} = 0.88$.

Objective Awareness

Pre- and Post-Test Explicit identifications. A 2 × 2 mixed ANOVA was conducted on mean percentage of correct responses, with testing moment (pre-test vs. post-test) and item type (context tones vs. transfer tones) as independent variables. The analysis revealed

[4] As for the main effect of item type in response times, this main effect is entirely due to the inclusion of the initial learning trials with context tones only. Again, the main effect of item type is nonsignificant when the first block of learning is removed from the analysis, $F(1, 107) = 1.19$, $p = .278$, $\eta_p^2 = .011$, 90% CI [.000, .064], $BF_{incl} = 0.014$.

a main effect of testing moment, $F(1, 107) = 12.86$, $p = .001$, $\eta_p^2 = .107$, 90% CI [.032, .203], $BF_{incl} = 387.45$, with the mean percentage of correct responses being higher at post-test ($M = 28.14\%$, $SE = 1.80$) relative to pre-test ($M = 22.49\%$, $SE = 1.14$), indicating an improvement in pitch identification after learning. The mean percentage of correct responses was above chance guessing at pre-test, $t(107) = 7.19$, $p < .001$, $d_1 = 0.69$, 95% CI [0.48, 0.90], $BF_{10} > 1000$, perhaps indicating some prior knowledge by some participants and a too liberal exclusion criterion for pre-test performance (for further discussion of this finding, see the General Discussion). However, higher pre-test scores did not modify any of the key effects reported in this paper (see Supplementary Materials accompanying this article at online.ucpress.edu/mp). The mean percentage of correct responses was also significant at post-test, $t(107) = 7.71$, $p < .001$, $d_1 = 0.74$, 95% CI [0.53, 0.95], $BF_{10} > 1000$. Neither the main effect of item type, $F(1, 107) = 1.01$, $p = .316$, $\eta_p^2 = .009$, 90% CI [.000, .061], $BF_{incl} = 0.164$, nor the interaction between the two factors were significant, $F(1, 107) = 1.30$, $p = .257$, $\eta_p^2 = .012$, 90% CI [.000, .066], $BF_{incl} = 0.237$, indicating that improvements were not robustly better for the trained context stimuli relative to the untrained transfer tones.

Despite the lack of interaction, we assessed improvements at post-test relative to pre-test per item type, separately. These results are depicted in Figure 3. For context tones, participants were significantly more accurate at post-test ($M = 28.11\%$, $SE = 1.85$) compared to pre-test ($M = 23.81\%$, $SE = 1.24$), $t(107) = 2.50$, $p = .014$, $d_D = 0.24$, 95% CI [0.05, 0.43], $BF_{10} = 2.08$. Also, for transfer tones, participants were significantly more accurate at post-test ($M = 28.17\%$, $SE = 2.21$) relative to pre-test ($M = 21.16\%$, $SE = 1.51$), $t(107) = 3.18$, $p = .002$, $d_D = 0.31$, 95% CI [0.11, 0.50], $BF_{10} = 11.95$.

The same analysis was conducted on RTs in the test phases and revealed a main effect of testing moment, $F(1, 107) = 55.09$, $p < .001$, $\eta_p^2 = .340$, 90% CI [.221, .440], $BF_{incl} > 1000$, with faster RTs at post-test ($M = 2024$, $SE = 101$) relative to pre-test ($M = 2974$, $SE = 174$). Again, neither the main effect of item type, $F(1, 107) = 1.49$, $p = .225$, $\eta_p^2 = .014$, 90% CI [.000, .070], $BF_{incl} = 0.161$, nor the interaction between the two factors were significant, $F(1, 109) = 2.09$, $p = .151$, $\eta_p^2 = .019$, 90% CI [.000, .081], $BF_{incl} = 0.280$. As for accuracy rates, we assessed improvements at post-test relative to pre-test per item type. The data are depicted in Figure 4. For context tones, participants responded significantly faster at post-test ($M = 2036$, $SE = 104$) relative to pre-test ($M = 2871$, $SE = 148$), $t(107) = 8.00$, $p < .001$, $d_D = 0.77$, 95% CI [0.55, 0.98], $BF_{10} > 1000$. Also for transfer tones, participants responded significantly faster at post-test ($M = 2013$, $SE = 102$) relative to pre-test ($M = 3077$, $SE = 222$), $t(107) = 5.74$, $p < .001$, $d_D = 0.55$, 95% CI [0.35, 0.75], $BF_{10} > 1000$.

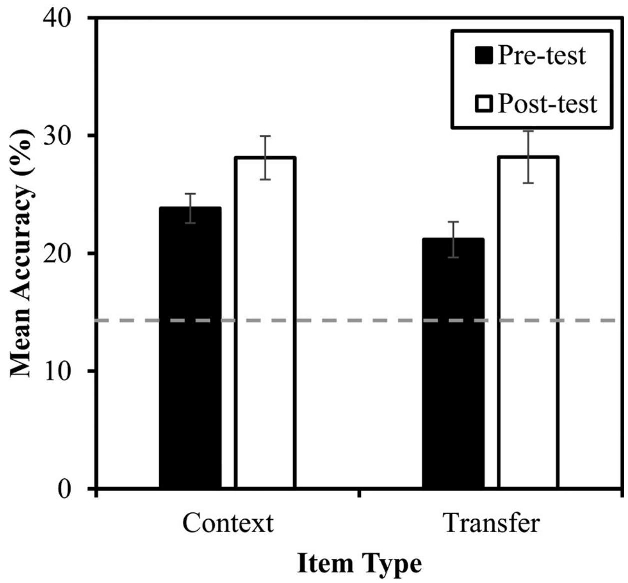

FIGURE 3. Mean accuracy (in %) at pre-test and post-test as a function of item type with standard error bars. *Note*. The dashed line represents the objective chance guessing threshold (i.e., 14.29%).

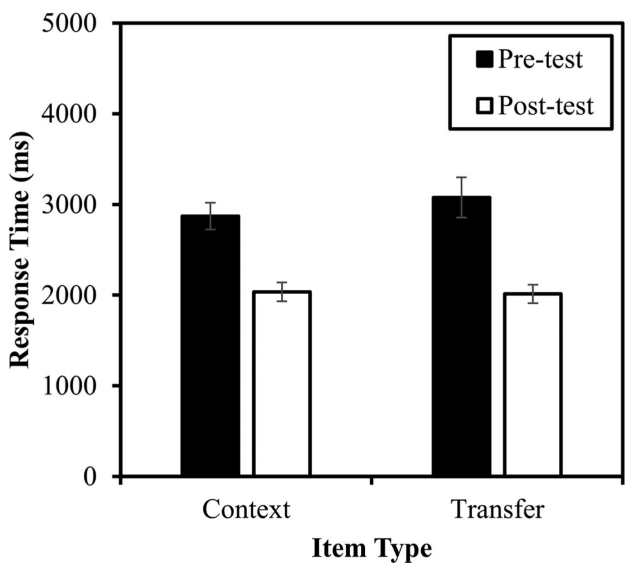

FIGURE 4. Mean response times (in ms) at pre-test and post-test as a function of item type with standard error bars.

Intercept Analyses. For context tones, the intercept was significant, $\beta = 23.26$, $SE = 8.72$, $t = 2.67$,

p = .009, indicating that the congruency effect was about 23 ms when participants responded at chance-guessing at post-test, arguing that implicit learning contributed to the main congruency effect. The slope was also significant, $\beta = 0.96$, $SE = 0.38$, $t = 2.52$, $p = .013$, indicating that increased awareness boosted learning. For transfer tones, the intercept was similarly significant, $\beta = 19.46$, $SE = 9.29$, $t = 2.09$, $p = .039$, but the slope was not, $\beta = 0.70$, $SE = 0.41$, $t = 1.73$, $p = .087$. Given the absence of a main effect of congruency in the learning phase for error rates, we did not conduct the same analysis on errors.

Discussion

In the present paper, we evaluated whether our auditory contingency learning task can be used to train participants to learn the associations between the pitches from a C Major scale and their corresponding labels (i.e., note names). Tones were presented much more frequently with their corresponding note names than with each non-corresponding note name. Importantly, we further assessed whether contingency learning with this kind of task facilitates the learning of pitch classes or whether, perhaps less interestingly, learning might be purely item- or instrument-specific. That is, it could be proposed that learning does occur, but that this learning is entirely restricted to the associations between a note name and a very specific auditory stimulus, a pitch incorporating its timbral features. Most critically, learning from context tones from two timbres *did* transfer/generalize to untrained tones of another timbre, both in more automatic RT measures during learning and in explicit note identification accuracy.

During the learning phase, for context tones, we found that participants responded more quickly on congruent relative to incongruent trials, replicating prior reports (Iorio et al., 2024). This again indicates learning of the associations between tones and corresponding note names. Moreover, this effect indicates automaticity. In other words, due to their newly acquired contingency knowledge, tones had automatic influences on note name identification. Analogous to nonmusical Stroop tasks (Stroop, 1935; see MacLeod, 1991, for a review) and auditory musical Stroop tasks (e.g., Akiva-Kabiri & Henik, 2012; Itoh et al., 2005; Miyazaki, 2000, 2004b), non-target auditory note stimuli automatically bias identification of the target note name information. Learning tasks such as our own similarly produce this type of learning in musical (e.g., Iorio et al., 2023, 2024; Schmidt et al., 2023) and nonmusical contexts (e.g., MacLeod & Dunbar, 1988; Schmidt et al., 2007). As discussed in the introduction, automatic processing of pitches is a key characteristic of absolute identification of said pitches. For instance, AP possessors not only identify pitches correctly, but also very rapidly and involuntarily. The fact that similarly automatic effects are observed in naïve (NAP) participants after brief training is thus fascinating given the standard narrative about the inherent difficulty of learning to identify pitches in adulthood.

Most importantly for the present report, automatic effects assessed in RTs generalized to transfer tones, for which no contingencies were directly manipulated. These results are in line with what a timbre-independent account of learning would predict. That is, the automatic (Stroop-like) effects we observe cannot be attributed to purely item- or instrument-specific learning. Instead, our results seem more coherent with the notion that participants learned the association between pitch classes/chromas and corresponding note names, and this had automatic effects on performance regardless of the timbral features of the presented auditory stimulus. Therefore, the underlying representations of the acquired knowledge may not be so different than those of true AP possessors (i.e., who rely on pitch chroma; cf. Van Hedger et al., 2023), but with some caveats discussed below.

Our work is also consistent with results of experiments of Wong, Lui, and colleagues (2020), in which they trained participants to *explicitly* identify pitches. They hypothesized that the degree of overlap between the psychological space for trained and untrained tones would affect the degree of generalization toward untrained tones on an explicit identification task. That is, the more the variability between trained tones along irrelevant dimensions for true AP (i.e., timbre and pitch height), the more the generalization to untrained tones. Consistent with their account, in our work, we used two different timbres as context tones for which contingencies were directly manipulated, and we presented these tones first. The idea was that this manipulation would allow context tones to become closer in psychological space because participants can learn the common qualia between the different instrument tones for a given pitch class (i.e., to discover the common point between the two). In turn, later presentations of untrained (i.e., transfer) tones could activate these shared representations (i.e., pitches) in the psychological space, thereby resulting in generalization effects across the timbre dimension. We also continued presenting context tones along with transfer tones, with the idea that continued presentation of predictive context tones would help to maintain the learned pitch class representations, though the degree to which this design decision mattered could be explored in future research. More precisely, the current experiment did not explore to what extent generalization does depend on (a) the use of more than one context timbre, or (b) initial learning with context tones only. Though previous work, albeit with a different learning approach, has revealed the importance of these factors (Wong, Lui et al., 2020), future research might investigate the extent to which these factors also influence performance in the current task.

In addition to automatic effects on responses times, participants also improved on explicit identification of pitches after training, with mean percentage of correct identification being higher at post-test relative to pre-test. Lack of evidence for a difference in mean accuracy rate between context and transfer tones also adds credence to the notion that participants are learning pitch classes rather than stimulus-specific associations. That is, learning generalized across timbres. Although incidental learning procedures are particularly useful for automatization via more implicit forms of learning, these improvements in explicit identification suggest that our procedure also allows for acquired knowledge between pitches and corresponding note names to become to some extent verbalizable. Moreover, improvements on explicit identification were accompanied by faster response times at post-test relative to pre-test. This again indicates automaticity with learning (i.e., characteristic of a change in strategies from slow algorithmic processing toward a fast and automatic one; Logan, 1988), which is the key feature of AP processing (e.g., Levitin & Rogers, 2005).

Incidentally, some evidence for both implicit and explicit learning was observed. For instance, our intercept analyses on the objective awareness data indicate that a congruency effect is still present for context and transfer tones when participants had chance-level guessing accuracy in the objective awareness phase. This is consistent with implicit learning (including implicit generalization across timbres). We also found some influences of conscious knowledge on learning effects. For instance, the congruency effect increased with increasing levels of objective awareness (slope analysis), albeit only for context stimuli. Moreover, when asking participants about their subjective awareness of the regularities or the recurrence of pitch classes across timbres, we did not find evidence that awareness modulated the congruency effect, neither with context tones, nor with transfer tones (i.e., we did not find evidence for interactions involving awareness). Overall, these results do not suggest that awareness modified learning and generalization effects. Combined with our intercept and slope analyses, our results suggest that incidental learning shapes, to some extent, both conscious and unconscious knowledge.

These results also clearly show that contingency knowledge did not become consciously accessible (i.e., verbalizable) in a mandatory fashion, arguing that some participants learn the contingencies in a non-conscious fashion (see Greenwald et al., 1995; Iorio et al., 2023). Of course, consciously accessible knowledge is most relevant for AP, which we also observed. But our results add further credence to the research on implicit AP. However, our work takes work on implicit AP even further. Past work has demonstrated that participants seem to be able to identify pitches absolutely in certain contexts (e.g., determining whether a song is played in the correct or a transposed key; Schellenberg & Trehub, 2003; Van Hedger et al., 2018, 2023), but our results further indicate evidence of implicit learning of associations between pitches and *pitch names*. Therefore, as highlighted in other work on AP training, our results fit well with a continuous view of pitch identification ability (Van Hedger et al., 2019; Wong, Lui et al., 2020).

At least one other pair of studies have also observed robust improvements after a single training session and generalization effects (Van Hedger et al., 2015). During the training phase, participants completed 180 forced-choice pitch naming trials with feedback for the 12 piano notes of one octave. Post-test performance on the same 12 piano tones robustly increased from pre-test performance in both experiments. In a further generalization test with pitches from different timbres and octaves, performance was just barely significantly better than pre-test performance in Experiment 1 and nonsignificant (but marginal) in Experiment 2. A small subsample of participants from Experiment 1 ($n = 6$) also showed retention between five to seven months later on the initially trained piano tones, but not in the generalization test (where performance was not robustly above chance level guessing). As one potential critique of the generalization test, however, a full one quarter of the test tones in this phase were the initially trained piano tones. Another quarter were piano tones from the immediately higher octave, another quarter guitar tones from the trained octave, and the final quarter guitar tones from the immediately lower octave. These four types of "generalization" tones were not tested separately. If we assume that participants only improved with the initially trained piano tones, generalization performance was very close to what we would expect based on pre-test and post-test scores alone.[5] To what extent

[5] In particular, the percentage correct responses for Experiment 1 were 13.7% for pre-test, 36.2% for post-test, and 21.7% for the generalization

transfer across timbres (or octaves) was actually observed is therefore unclear. On the other hand, the authors rightly point out that the generalization test was probably harder (i.e., as notes spanning three octaves and two timbres were randomly intermixed) relative to the pre- and post-tests (i.e., where notes from only one octave and one timbre were randomly intermixed), so these calculations might underestimate true generalization in the studies of Van Hedger and colleagues. In any case, these results are coherent with data of the present work showing improvements in explicit pitch identification, and the present results also illustrate automatic response biases in our novel reaction time measures, a key feature of AP.

Although the current work, unlike much prior work on AP, did not aim to train participants to a high level of accuracy on the pitch identification task (and we only trained 7 pitches), our data do add further support for the notion that pitch identification performance is learnable. Indeed, there is a growing body of evidence that AP knowledge can be developed in adulthood (e.g., Van Hedger et al., 2019; Wong, Lui, et al., 2020; Wong, Ngan, et al., 2020; see Heald et al., 2017). This research is less consistent with the typical dogma that absolute pitch naming ability is exclusively reserved for the genetically gifted and/or those that began music training very early in childhood (see Van Hedger et al., 2019; Wong, Lui et al., 2020). Of course, the current results (and results of prior works) do not argue against a *role* of early learning or genetics, both of which play significant roles in a range of human skills and competencies, but the standard narrative is that AP is not learnable at all. Of course, more work is still needed to further explore this debate. For instance, we observed robust *improvements* in pitch identification (which generalized across timbres), but mean performance still fell well short of the strict criteria usually used to evaluate AP (e.g., over 90% accurate across several octaves and all semitones). However, improvements were obtained in a small timeframe, so further improvements seem plausible.

Future works might add further learning sessions to determine, for instance, whether 1) performance continues to improve with further practice, and 2) continued training with our procedure might converge toward genuine AP behavior (e.g., Van Hedger et al., 2019; Wong, Lui, et al., 2020). That is, does learning continue to improve following standard laws of practice (e.g., see Schmidt et al., 2023) and do these improvements occur but fail to reach a level as precise as a true AP possessor? However, to explore these questions, future works should use a wider range of stimuli. For instance, our learning task using all twelve semitones of the chromatic scale (Henry et al., 2024) is particularly pertinent, since the use of the twelve semitones is important for differentiating different classes of pitch naming ability (e.g., Wilson et al., 2009). Globally, future work like this could further evaluate the plausibility of an account of automaticity that conceives AP as a skill that is developed throughout constant practice as any sensory-motor skill at any age (see Van Hedger et al., 2019, for a similar account and for a discussion).

As a potential limit of our work, however, we did not assess the long-term retention of explicit verbalizable knowledge. Although we have found robust retention of pitch identification knowledge after a week delay in other work involving a contingency learning task (e.g., Henry et al., 2024; Iorio et al., 2024; see also, Van Hedger et al., 2015), generalization to untrained stimuli was not tested in these past studies. Future work on generalization might therefore integrate a delayed explicit pitch identification phase into the procedure to assess long-term retention of the entire pitch class. Other work might further assess how persistent the automatic biases we observed in response times are with delay. To this end, automatic response biases could be tested in a contingency-free test phase (e.g., Schmidt & De Houwer, 2016) after a delay (Schmidt et al., 2020).

To what extent the underlying knowledge acquired in AP training is similar to representations in genuine AP possessors requires further investigation. Clearly, generalization across timbres is possible, but this does not necessarily indicate learning of the entire pitch class of a given note or a subset of notes. For instance, are participants also able to generalize across octaves? In our study, we used the seven natural notes from a C Major scale spanning only one octave, leaving open the possibility that learning is restricted to the used octave (e.g., learning specific to pitch height; e.g., Bongiovanni et al., 2023). For example, it could be proposed that participants are not learning pitch classes at all but are rather using pitch height to estimate note names (e.g., lowest pitches with the leftmost keys and higher pitches with rightmost keys). Related, it could be proposed that the higher than chance pre-test performance observed in the present experiment could be explained by use of such a strategy.[6] Further research might

test. We would expect a score of 19.325% in the generalization test based on the pre-test and post-test scores: (36.2 × .25) + (13.7 × .75). Similarly, in Experiment 2 these percentages are: 10.9% for pre-test, 25.4% for post-test, and 15.4% (14.635% expected) for the generalization test.

[6] Partially consistent with this, we did observe higher pre-test accuracy for the lowest note. Interestingly, accuracy was not increased for the

therefore explore whether generalization across octaves is similarly possible with our incidental contingency learning approach. In some currently ongoing research with a conceptually similar design as the present one but using two context octaves and one transfer octave, we have already found some positive results (i.e., generalization to untrained octaves), consistent with the longer-term learning studies of Wong, Lui, and colleagues (2020).

As another interesting observation, we did not find clear influences of the timbre on learning, with both automatic and explicit learning effects being similar for the trained (context) and untrained (transfer) stimuli. This is interesting because timbre does often play a role in AP. For instance, many AP possessors show a higher dependency on timbral cues, with better performances for tones from familiar instruments, whether on their more (Wilson et al., 2009) or less automatic (Hansen & Reymore, 2023; Reymore & Hansen, 2020) pitch identification. This echoes the research conducted on the interaction effects between the pitch and the timbre dimensions. For example, it was shown that changes in the brightness of the timbre affect pitch perception, and vice versa (Allen & Oxenham, 2014). This latter research focuses on the independence (or not) of pitch and timbre processing, with possible perceptual difficulties in segregating both dimensions (see Allen et al., 2017). It is therefore clearly not the case that all pitch learning is timbre neutral. It is interesting (and encouraging) that our task seemingly promotes more general timbre-independent learning. This might be considered consistent with other related recent research demonstrating generalization effects across the timbre dimension in APM (Van Hedger et al., 2023) and other AP training paradigms (e.g., Wong, Lui et al., 2020). In our view, this calls for further research to understand which factors or learning scenarios promote general learning of pitch classes rather than timbre-dependent (or even item specific or pitch height specific) learning.

In brief, we hope that our work will inspire more investigations into whether our approach (or other similar approaches) allows participants to acquire knowledge and underlying representations comparable to that of genuine AP possessors (i.e., of entire pitch classes), or whether the type of learning explored in the current report relies on more superficial representations (e.g., stimulus specific or perhaps pitch height specific). Clearly, transfer across timbres is possible, but there may be elements of pitch class learning that are unobtainable by adult NAP possessors, making true AP impossible to train. In any case, the present results indicate clear generalization effects in indirect measures of learning (e.g., response times) and on explicit identifications, providing important insights for the AP training framework regarding underlying knowledge learned with our paradigm, seemingly closer to genuine AP possessors' representations than previously assumed.

Author Note

Correspondence concerning this article should be addressed to Williams Henry (Williams.Henry@u-bourgogne.fr).

References

Akiva-Kabiri, L., & Henik, A. (2012). A unique asymmetrical Stroop effect in absolute pitch possessors. *Experimental Psychology*, *59*(5), 272–278. https://doi.org/10.1027/1618-3169/a000153

Allen, E. J., Burton, P. C., Olman, C. A., & Oxenham, A. J. (2017). Representations of pitch and timbre variation in human auditory cortex. *Journal of Neuroscience*, *37*(5), 1284–1293. https://doi.org/10.1523/JNEUROSCI.2336-16.2016

Allen, E. J., & Oxenham, A. J. (2014). Symmetric interactions and interference between pitch and timbre. *Journal of the Acoustical Society of America*, *135*(3), 1371–1379. https://doi.org/10.1121/1.4863269

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

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

Bairnsfather, J. E., Osborne, M. S., Martin, C., Mosing, M. A., & Wilson, S. J. (2022). Use of explicit priming to phenotype absolute pitch ability. *PLOS ONE*, *17*(9), e0273828. https://doi.org/10.1371/journal.pone.0273828

Bermudez, P., & Zatorre, R. J. (2005). Conditional associative memory for musical stimuli in nonmusicians: Implications for absolute pitch. *The Journal of Neuroscience*, *25*(34), 7718–7723. https://doi.org/10.1523/JNEUROSCI.1560-05.2005

highest notes. Post-test accuracy was higher for the extremes, though accuracy (and response times) improved for all seven notes. Mean scores per note are available in the Supplementary Materials accompanying this article at online.ucpress.edu/mp.

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

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

Cohen, J. (1973). Eta-squared and partial eta-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, *33*(1), 107–112. https://doi.org/10.1177/001316447303300111

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Cousineau, D., & Goulet-Pelletier, J. C. (2021). A study of confidence intervals for Cohen's dp in within-subject designs with new proposals. *The Quantitative Methods for Psychology*, *17*(1), 51–75. https://doi.org/10.20982/tqmp.17.1.p051

Deutsch, D. (2013). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 141–182). Elsevier Academic Press. https://doi.org/10.1016/B978-0-12-381460-9.00005-5

Deutsch, D., Henthorn, T., Marvin, E., & Xu, H. (2006). Absolute pitch among American and Chinese conservatory students: Prevalence differences, and evidence for a speech-related critical period. *Journal of the Acoustical Society of America*, *119*(2), 719–722. https://doi.org/10.1121/1.2151799

Fitts, D. A. (2021). Expected and empirical coverages of different methods for generating noncentral t confidence intervals for a standardized mean difference. *Behavior Research Methods*, *53*, 2412–2429. https://doi.org/10.3758/s13428-021-01550-4

Forrin, N. D., & MacLeod, C. M. (2017). Relative speed of processing determines color–word contingency learning. *Memory and Cognition*, *45*, 1206–1222. https://doi.org/10.3758/s13421-017-0721-4

Forrin, N. D., & MacLeod, C. M. (2018). Contingency proportion systematically influences contingency learning. *Attention, Perception, and Psychophysics*, *80*(1), 155–165. https://doi.org/10.3758/s13414-017-1424-4

Goulet-Pelletier, J.-C., & Cousineau, D. (2018). A review of effect sizes and their confidence intervals, Part 1: The Cohen's *d* family. *The Quantitative Methods for Psychology*, *14*(4), 242–265. https://doi.org/10.20982/tqmp.14.4.p242

Greenwald, A. G., Klinger, M. R., & Schuh, E. S. (1995). Activation by marginally perceptible ("subliminal") stimuli: Dissociation of unconscious from conscious cognition. *Journal of Experimental Psychology: General*, *124*(1), 22–42. https://doi.org/10.1037/0096-3445.124.1.22

Hansen, N. C., & Reymore, L. (2024). Timbral cues underlie instrument-specific absolute pitch in expert oboists. *PLOS ONE*, *19*(10), e0306974. https://doi.org/10.1371/journal.pone.0306974

Heald, S. L. M., Van Hedger, S. C., & Nusbaum, H. C. (2017). Understanding sound: Auditory skill acquisition. In B. H. Ross (Ed.), *The psychology of learning and motivation* (pp. 53–93). Elsevier Academic Press. https://doi.org/10.1016/bs.plm.2017.03.003

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

Henry, W., Iorio, C., & Schmidt, J. R. (2024). *Acquiring pitch-label associations using the twelve semitones and perfect contingencies: An auditory contingency learning task* [Manuscript in preparation].

Iorio, C., Bigand, E., Schmidt, J. R. (2024). Incidentally acquiring pitch-label associations with a musical contingency learning task. *Collabra: Psychology*, *10*, 118806. https://doi.org/10.1525/collabra.118806

Iorio, C., Šaban, I., Poulin-Charronnat, B., & Schmidt, J. R. (2023). Incidental learning in music reading: The music contingency learning task. *Quarterly Journal of Experimental Psychology*, *76*(2), 429–449. https://doi.org/10.1177/17470218221092779

Itoh, K., Suwazono, S., Arao, H., Miyazaki, K. I., & Nakada, T. (2005). Electrophysiological correlates of absolute pitch and relative pitch. *Cerebral Cortex*, *15*(6), 760–769. https://doi.org/10.1093/cercor/bhh177

Kelley, K. (2007). Methods for the behavioral, educational, and social sciences: An R package. *Behavior Research Methods*, *39*, 979–984. https://doi.org/10.3758/BF03192993

Keysers, C., Gazzola, V. & Wagenmakers, EJ. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature Neuroscience*, *23*, 788–799. https://doi.org/10.1038/s41593-020-0660-4

Kim, S. G., & Knösche, T. R. (2017). On the perceptual subprocess of absolute pitch. *Frontiers in Neuroscience*, *11*, 557. https://doi.org/10.3389/fnins.2017.00557

Krimphoff, J., McAdams, S., & Winsberg, S. (1994). Caractérisation du timbre des sons complexes. II. Analyses acoustiques et quantification psychophysique. *Le Journal de Physique IV*, *4*(C5), 625–628. https://doi.org/10.1051/jp4:19945134

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

Levitin, D. J. (2004). L'oreille absolue: Autoréférencement et mémoire. *Année psychologique*, *104*(1), 103–120. https://doi.org/10.3406/psy.2004.3930

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

Lewicki, P., Hill, T., & Bizot, E. (1988). Acquisition of procedural knowledge about a pattern of stimuli that cannot be articulated. *Cognitive Psychology*, *20*(1), 24–37. https://doi.org/10.1016/0010-0285(88)90023-0

Lin, O. Y.-H., & MacLeod, C. M. (2018). The acquisition of simple associations as observed in color–word contingency learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(1), 99–106. https://doi.org/10.1037/xlm0000436

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

Loui, P. (2016). Absolute pitch. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 81–94). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.001.0001

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review*, *95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, *109*(2), 163–203. https://doi.org/10.1037/0033-2909.109.2.163

MacLeod, C. M. (2019). Learning simple associations. *Canadian Psychology / Psychologie canadienne*, *60*(1), 3–13. https://doi.org/10.1037/cap0000170

MacLeod, C. M., & Dunbar, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(1), 126–135. https://doi.org/10.1037/0278-7393.14.1.126

Makowski, D., Ben-Shachar, M. S., Chen, S. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in Psychology*, *10*, 2767. https://doi.org/10.3389/fpsyg.2019.02767

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Marvin, E. W., & Brinkman, A. R. (2000). The effect of key color and timbre on absolute pitch recognition in musical contexts. *Music Perception*, *18*(2), 111–137. https://doi.org/10.2307/40285905

Marvin, E. W., VanderStel, J., & Siu, J. C. S. (2020). In their own words: Analyzing the extents and origins of absolute pitch. *Psychology of Music*, *48*(6), 808–823. https://doi.org/10.1177/0305735619832959

Matsuda, M., Igarashi, H., & Itoh, K. (2019). Auditory T-complex reveals reduced neural activities in the right auditory cortex in musicians with absolute pitch. *Frontiers in Neuroscience*, *13*, 809. https://doi.org/10.3389/fnins.2019.00809

Miyazaki, K. I. (1989). Absolute pitch identification: Effects of timbre and pitch region. *Music Perception*, *7*(1), 1–14. https://doi.org/10.2307/40285445

Miyazaki, K. (2000). Interaction in musical-pitch naming and syllable naming: an experiment on a Stroop-like effect in hearing. In T. Nakada (Ed.), *Integrated human brain science: Theory, method, application (music)* (pp. 415–423). Elsevier.

Miyazaki, K. I. (2004a). How well do we understand absolute pitch? *Acoustical Science and Technology*, *25*(6), 426–432. https://doi.org/10.1250/ast.25.426

Miyazaki, K. I. (2004b). The auditory Stroop interference and the irrelevant speech/pitch effect: Absolute-pitch listeners can't suppress pitch labeling. In *Proceedings of the 18th International Congress on Acoustics*. Kyoto, Japan.

Morey, R. D., Rouder, J. N. (2022). *BayesFactor: Computation of Bayes Factors for Common Designs_* (Version 0.9.12-4.4). R package. https://CRAN.R-project.org/package=BayesFactor

Reymore, L., & Hansen, N. C. (2020). A theory of instrument-specific absolute pitch. *Frontiers in Psychology*, *11*, 560877. https://doi.org/10.3389/fpsyg.2020.560877

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

Schlemmer, K. B., Kulke, F., Kuchinke, L., & Van Der Meer, E. (2005). Absolute pitch and pupillary response: Effects of timbre and key color. *Psychophysiology*, *42*(4), 465–472. https://doi.org/10.1111/j.1469-8986.2005.00306.x

Schmidt, J. R. (2021). Apprentissage incident des associations simples de stimulus-réponse: Revue de la recherche avec la tâche d'apprentissage de contingences couleur-mot: *L'Année Psychologique*, *121*(2), 77–127. https://doi.org/10.3917/anpsy1.212.0077

Schmidt, J. R. (2012) Human contingency learning. In N. M. Seel (Ed.), *Encyclopedia of the sciences of learning*. Springer. https://doi.org/10.1007/978-1-4419-1428-6_646

Schmidt, J. R. (2018). Best not to bet on the horserace: A comment on Forrin and MacLeod (2017) and a relevant stimulus-response compatibility view of colour-word contingency learning asymmetries. *Memory and Cognition*, *46*, 326–335. https://doi.org/10.3758/s13421-017-0755-7

Schmidt, J. R., Crump, M. J. C., Cheesman, J., & Besner, D. (2007). Contingency learning without awareness: Evidence for implicit control. *Consciousness and Cognition*, *16*(2), 421–435. https://doi.org/10.1016/j.concog.2006.06.010

Schmidt, J. R., & De Houwer, J. (2012a). Adding the goal to learn strengthens learning in an unintentional learning task. *Psychonomic Bulletin and Review*, *19*(4), 723–728. https://doi.org/10.3758/s13423-012-0255-5

Schmidt, J. R., & De Houwer, J. (2012b). Learning, awareness, and instruction: Subjective contingency awareness does matter in the colour-word contingency learning paradigm. *Consciousness and Cognition*, *21*(4), 1754–1768. https://doi.org/10.1016/j.concog.2012.10.006

Schmidt, J. R., & De Houwer, J. (2016). Time course of colour-word contingency learning: Practice curves, pre-exposure benefits, unlearning, and relearning. *Learning and Motivation*, *56*, 15–30. https://doi.org/10.1016/j.lmot.2016.09.002

Schmidt, J. R., De Houwer, J., & Besner, D. (2010). Contingency learning and unlearning in the blink of an eye: A resource dependent process. *Consciousness and Cognition*, *19*(1), 235–250. https://doi.org/10.1016/j.concog.2009.12.016

Schmidt, J. R., De Houwer, J., & Moors, A. (2020). Learning habits: Does overtraining lead to resistance to new learning? *Collabra: Psychology*, *6*(1), 21. https://doi.org/10.1525/collabra.320

Schmidt, J. R., Iorio, C., & Poulin-Charronnat, B. (2023). Automatizing sight reading: Contingency proportion and task relevance in the music contingency learning procedure. *Collabra: Psychology*, *9*(1), 89743. https://doi.org/10.1525/collabra.89743

Simpson, J., & Huron, D. (1994). Absolute pitch as a learned phenomenon: Evidence consistent with the Hick-Hyman Law. *Music Perception*, *12*(2), 267–270. https://doi.org/10.2307/40285656

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

Steiger, J. H., & Fouladi, R. T. (2016). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 197–229). Routledge. https://doi.org/10.4324/9781315629049

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*, 1096–1104. https://doi.org/10.3758/BRM.42.4.1096

Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, *44*(1), 24–31. https://doi.org/10.1177/0098628316677643

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643–662. https://doi.org/10.1037/h0054651

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

Theusch, E., Basu, A., & Gitschier, J. (2009). Genome-wide study of families with absolute pitch reveals linkage to 8q24. 21 and locus heterogeneity. *The American Journal of Human Genetics*, *85*(1), 112–119. https://doi.org/10.1016/j.ajhg.2009.06.010

Theusch, E., & Gitschier, J. (2011). Absolute pitch twin study and segregation analysis. *Twin Research and Human Genetics*, *14*(2), 173–178. https://doi.org/10.1375/twin.14.2.173

Van Hedger, S. C., Bongiovanni, N. R., Heald, S. L., & Nusbaum, H. C. (2023). Absolute pitch judgments of familiar melodies generalize across timbre and octave. *Memory and Cognition*, *51*, 1898–1910. https://doi.org/10.3758/s13421-023-01429-z

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

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

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

Vanzella, P., & Schellenberg, E. G. (2010). Absolute pitch: Effects of timbre on note-naming ability. *PLOS ONE*, *5*, e15449.

Ward, W. D., & Burns, E. M. (1982). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 431–451). Academic Press.

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. https://doi.org/10.1525/mp.2012.29.3.285

Wilson, S. J., Lusher, D., Wan, C. Y., Dudgeon, P., & Reutens, D. C. (2009). The neurocognitive components of pitch processing: insights from absolute pitch. *Cerebral Cortex*, *19*(3), 724–732. https://doi.org/10.1093/cercor/bhn121

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

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

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

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., & Evans, A. C. (1998). Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. *Proceedings of the National Academy of Sciences*, *95*(6), 3172–3177. https://doi.org/10.1073/pnas.95.6.3172