Research Article

Octave Equivalence: Difficult to Perceive, But Improvements Are Possible With Training

James R. Schmidt and Williams Henry

CNRS, LEAD UMR5022, Université Bourgogne Europe, Dijon, France

Abstract: Musical notes separated by exactly one or more octaves share similarities and, in some respects, might be treated as interchangeable. This octave equivalence is sometimes evident, but in many contexts, is very hard to hear. In two large experiments, participants were asked to judge the similarity of tone pairs, presented sequentially, before and after octave equivalence training. Contrary to some prior research on the topic, it was clearly explained what sort of "similarity" they should rate tone pairs on (i.e., octave equivalence). Each pair consisted of either two tones of the same pitch class but separated by one or more octaves, or two tones of adjacent pitch classes also separated by one or more octaves (±1 semitone). Coherent with past work, this task was difficult. However, both musician and nonmusician samples scored above chance in this task at pretest. Also interestingly, performance improved after training. During the training task, participants also heard pairs of tones but were given the correct response to facilitate learning. Pretest performance and improvements for both groups were not substantial, however, again illustrating the difficulty of hearing octave equivalence, depending on the exact context. Potential relationships to relative and absolute pitch are also briefly discussed.

Keywords: octave equivalence, training, pitch classes, absolute pitch, relative pitch

In music, a pitch class is a set of all pitches that are separated by a whole number of octaves. For example, the pitch class "F" (or "fa" in fixed-do solfège) consists of the F notes across all octaves. An F_3 and an F_4, for example, are both F notes, but separated by an octave. In particular, in music, the 12 notes (e.g., C, C♯, D, D♯, E, etc.) are repeated in a circle, as often represented (e.g., Deutsch et al., 2006) with a pitch class circle like that illustrated in Figure 1A. For example, if we start with C_3 and keep increasing the pitch, we will eventually reach another (higher) C (i.e., C_4).

Every musical note therefore has two dimensions that are critical for the current discussion (Demany & Armand, 1984). First, the *pitch height*, which continuously increases from very low notes to very high notes. Second, the *pitch class*, which is cyclic. If we combine these two dimensions into a single graphic, the pitch helix as illustrated in Figure 1B (adapted from Shepard, 1982), we can see that the pitch heights keep increasing as we move upward, but every 12 notes, the pitch classes repeat (Deutsch et al., 2008; Shepard, 1964, 1982). These two dimensions are not equally obvious to the listener. The pitch height is very prominent, while the pitch class is often less obvious, particularly for nonmusicians (Marjieh et al., 2023). For example, if A_3 and A_4 are played in succession, it is very easy to hear that the second of these two tones is higher in pitch than the first, but it is much harder to hear that they both belong to the same pitch class (i.e., "A").

Why do we group different notes in the same pitch class? That is, why do we give the same name to very different notes (e.g., "G" for two notes separated by one or more octaves)? This, of course, is because there are similarities between all notes within the same pitch class. In some cases, they can even be interchangeable, termed *octave equivalence*. Indeed, various findings suggest that, at least in some circumstances, the octave is special. For instance, when young children imitate speech sounds outside of their vocal range, they imitate one octave higher (Peter et al., 2008, 2009), a type of octave equivalence. There is also evidence to suggest that human males and females have voices roughly an octave apart because of musicality (Bannan et al., 2024) and that we have harmonic clarity in our vocalizations to support octave equivalence (Wagner & Hoeschele, 2022), seemingly based on vocal apparatus adaptations that are unique among primates, which might imply that we need more effort to make sounds, but that the sound is clearer as a result (Nishimura et al., 2022).

Octave equivalence does occur in some scenarios. For instance, we can hear it clearly with notes played

Experimental Psychology (2025), *72*(4), 159–174
https://doi.org/10.1027/1618-3169/a000652

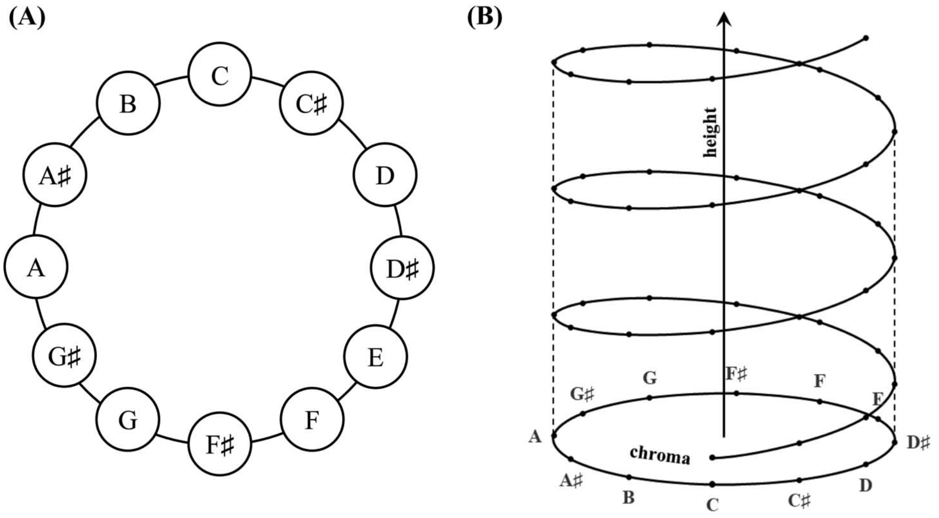


Figure 1. Pitch Class Circle (A) and a helix representation of the relation between Pitch Height and Chroma (B). Pitch height increases continuously. Pitch class/chroma is cyclical, such that pitch classes repeat every 12 semitones (i.e., a new octave).

simultaneously due to fusion effects (e.g., Bonnard et al., 2016; Demany & Semal, 1988; Demany et al., 2021). However, octave equivalence is not always evident to the ear. For example, Deutsch (1972) presented participants with several versions of a familiar song. One of the versions was modified by retaining all the correct pitch classes, but with random variations in octave. That is, some notes were randomly transposed up one octave, other notes down an octave, and others remained correct. Recognition of the song is difficult in this situation (but see, Dowling & Hollombe, 1977; Idson & Massaro, 1976; Kallman & Massaro, 1979; Massaro et al., 1980; see also, Deutsch, 1979). In this sense, "equivalent" notes are not completely interchangeable. After hearing the normal, unmodified version of the song, however, the modified version becomes easily recognizable. In contrast, when notes are randomly modified in scale degrees rather than octaves, it remains difficult to hear the familiar melody. Thus, equivalent notes *can* be perceived as equivalent, but it may require some priming to focus on the pitch classes rather than the pitch heights. In other words, different notes separated by one or more octaves are equivalent in some respects, but this is not always salient.

Other evidence further corroborates the notion that "equivalent" notes are not treated as completely identical. For example, Deutsch and Boulanger (1984) asked participants to notate a random sequence of notes. When all the notes were from a single octave, this task was significantly easier than a task in which the notes alternated between two octaves. Relatedly, detection of musical notes by ear is also influenced not only by differences in pitch class but also differences in pitch height. Specifically, adding one or more octaves between two pitches makes the detection of intervals more difficult (Thurlow & Erchul, 1977).

Still, octave equivalence is detectable by participants and can sometimes have rather automatic effects on performance. For instance, in Hoeschele et al. (2012), participants were trained to respond (press a key) when they heard certain groups of notes, but to omit a response when hearing other groups of notes (i.e., a go/no-go task; Donders, 1868/1969). During this initial training, all notes were from one octave only. In a subsequent nonreinforced transfer phase, notes from the same pitch classes but a new octave were presented to participants. The authors observed a bias to press the key (go) or avoid responding (no-go) to the same pitch classes in the new octave. Coherent with the above discussion, however, further research (Wagner et al., 2022) revealed that such results are only observed when the manipulation discourages focus on pitch height by creating at least three groups of adjacent notes (e.g., responding to a group of the highest and of the lowest pitch notes, but not intermediate pitches). The authors argued that this increases focus on pitch class. Further research has suggested that pitch height is detected more automatically, whereas pitch class detection might require working memory resources (Regev et al., 2019).

Thus, it is clearly the case that octave equivalence is sometimes perceived, although this is usually when a clear context is given. However, the most direct test of octave equivalence perception is to simply present participants two tones and ask them whether they are equivalent or not. More central to the current research, most participants, especially nonmusicians, do not seem to perceive the similarity between two pitches separated by one or more octave, even when presented in immediate succession (Allen, 1967; Kallman, 1982). For instance, Allen (1967) presented participants with pairs of 200 ms tones, the first always being 1,000 Hz, separated by a 1,200 ms delay. Participants were asked to judge the "similarity" of the tone pairs on a seven-point Likert scale. "Similarity" was not explained to participants. Rather, the objective was to see whether participants would be spontaneously sensitive to octave equivalence. Musicians did rate octave equivalent tones as more similar than nonequivalent tones. Nonmusicians did not show the same effect. In both a similar task where participants had to pick between two potential prime-equivalent pitches and in another task where participants needed to adjust the frequency of a second pitch to match a prime, only a minority of musicians where able to achieve this above chance guessing rates (Thurlow & Erchul, 1977).

In another study by Kallman (1982), two 165-ms sine waves were presented sequentially, separated by 530 ms of silence. Both the prime (first) tone and probe (second) tone were randomly varied, with intervals between 0 and 28 semitones. As in Allen (1967), participants were asked to rate the similarity between the two tones, except that responses were made on a continuous slider. Both nonmusicians and musicians were clearly sensitive to pitch height. Stimuli separated by a few semitones (e.g., 2) were rated as more similar than stimuli separated by more semitones (e.g., 4), and similarity scores decreased rather linearly from 0 to 28. Nonmusicians, however, were not sensitive to pitch class. Specifically, similarity scores were not robustly higher for 12 semitone intervals (i.e., 1 octave) relative to 11 or 13 semitones. The same was true for 24 semitone intervals (i.e., 2 octaves) relative to 23 or 25 semitones. That is, pitches belonging to the same pitch class were not perceived as more similar than nonequivalent tone pairs with similar intervals. Musicians, as a group, did show *some* sensitivity to octave equivalence (see also, Demany & Armand, 1984; Jacoby et al., 2019), with slightly higher similarity scores for octave equivalents, but this was due to a minority of participants with notably higher sensitivity than most of the other musicians (the latter of which showed no evidence of octave equivalence). The results were somewhat more encouraging when the range of stimuli and interval size were more limited, but there was still no large equivalence effect (see also, Borra et al., 2013).

The current research addresses two open questions. First, are participants able to detect octave equivalent tones when this goal is more explicit? As mentioned above, participants were asked to rate the "similarity" of tone pairs in prior work, but this was not defined for participants. It could be suggested, for instance, that participants are perfectly capable of detecting octave equivalence but decided not to make their similarity ratings on this basis. For example, they might have intentionally used pitch height cues, instead. Thus, participants are asked to rate the similarity of tones in the current work, but they are explicitly instructed that they should base their decisions on octave equivalence (which is first explained to them). Second, are participants able to *learn* to detect octave equivalence? After an initial pretest phase, participants were trained in a learning task to distinguish between octave equivalent pairs of tones and pairs of tones separated by some number of octaves plus or minus one semitone. We then measure their equivalence detection abilities again in a posttest to determine to what extent participants improve their ability to detect octave equivalence, if at all.

Experiment 1

Experiment 1 was conducted in a classroom setting as part of a first-year cognitive psychology tutorial. The experiment started with a pretest of octave equivalence in which participants needed to guess whether two tones were similar (equivalent) or dissimilar (nonequivalent). This phase was preceded by roughly 30 min of lecture about what octave equivalence is, with several examples and a discussion of much of the research mentioned in the introduction of the present article. This lecture both served to clarify which types of tone pairs should be considered "similar" (see the introduction) and to clarify the theoretical context of the research for the group project that the students needed to subsequently prepare for tutorial evaluation. In a subsequent learning phase, participants were presented with pairs of tones separated by one or two octaves on half of the trials and one or two octaves plus or minus one semitone on other trails. Critically, participants were explicitly informed which pairs were equivalent and which were not with a visual stimulus presented throughout each trial. A posttest, identical in all respects to the pretest, was then conducted to measure improvements in detection of octave equivalence.

Method

Participants

410 first-year psychology students participated in the experiment as part of a cognitive psychology tutorial class. The tutorial was divided into 14 tutorial groups of between 32 and 37 students and all participants were tested in this group context. Nine participants who indicated auditory problems were excluded. Two others were excluded as they did not respond to the musicianship and auditory problem questions (see Procedure). The remaining sample consisted of 350 nonmusicians

Table 1. Relative trial frequencies of prime–probe pairs in the learning phase of Experiment 1

	Probe																																			
	Octave 3												Octave 4												Octave 5											
Prime	C	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯	B	C	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯	B	C	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯	B
C$_3$												1	2	1										1	2	1										
C♯$_3$													1	2	1										1	2	1									
D$_3$														1	2	1										1	2	1								
D♯$_3$															1	2	1										1	2	1							
E$_3$																1	2	1										1	2	1						
F$_3$																	1	2	1										1	2	1					
F♯$_3$																		1	2	1										1	2	1				
G$_3$																			1	2	1										1	2	1			
G♯$_3$																				1	2	1										1	2	1		
A$_3$																					1	2	1										1	2	1	
A♯$_3$																						1	2	1										1	2	1
B$_3$	1																						1	2	1										1	2
C$_4$	2	1																						1	2	1										1
C♯$_4$	1	2	1																						1	2	1									
D$_4$		1	2	1																						1	2	1								
D♯$_4$			1	2	1																						1	2	1							
E$_4$				1	2	1																						1	2	1						
F$_4$					1	2	1																						1	2	1					
F♯$_4$						1	2	1																						1	2	1				
G$_4$							1	2	1																						1	2	1			
G♯$_4$								1	2	1																						1	2	1		
A$_4$									1	2	1																						1	2	1	
A♯$_4$										1	2	1																						1	2	1
B$_4$	1										1	2	1																						1	2
C$_5$	2	1										1	2	1																						1
C♯$_5$	1	2	1										1	2	1																					
D$_5$		1	2	1										1	2	1																				
D♯$_5$			1	2	1										1	2	1																			
E$_5$				1	2	1										1	2	1																		
F$_5$					1	2	1										1	2	1																	
F♯$_5$						1	2	1										1	2	1																
G$_5$							1	2	1										1	2	1															
G♯$_5$								1	2	1										1	2	1														
A$_5$									1	2	1										1	2	1													
A♯$_5$										1	2	1										1	2	1												
B$_5$											1	2	1										1	2	1											

and 49 musicians. Basic behavioral research is exempt from ethics review by law in France,[1] but all research was in conformance with the relevant ethics principles. Participants were free to refuse to participate or to withdraw from participation at any time without penalty, and they signed an informed consent form before participating. Student names and numbers were not collected to keep the data anonymous.

---

[1] LOI n° 2012-300 du 5 mars 2012 relative aux recherches impliquant la personne humaine (often informally referred to as the “loi Jardé”).

Materials and Apparatus

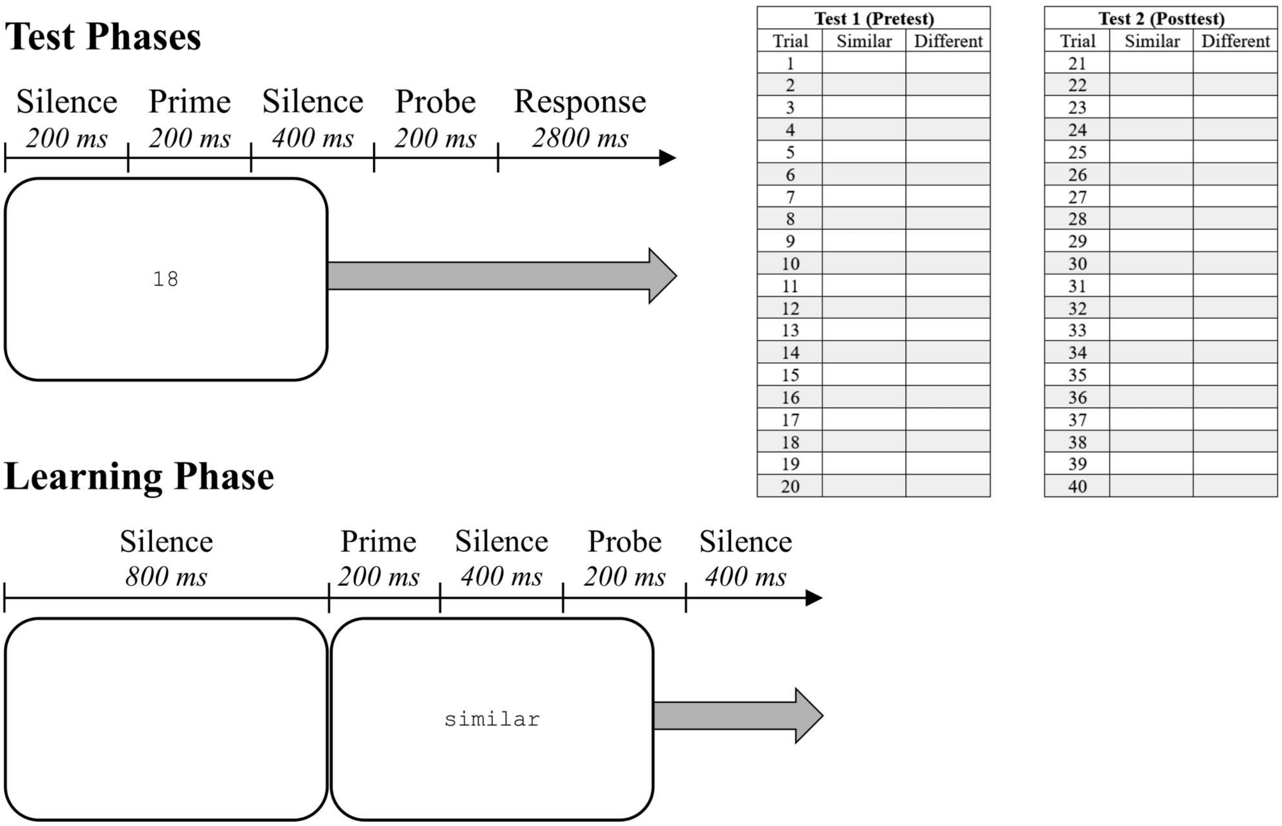
200-ms piano tones spanning the range C_3 to B_5 were generated in Ableton Live 11 Suite using the Grand Piano Single Sample preset, each of which was used as both a prime stimulus (i.e., the first presented tone) and a probe stimulus (i.e., the second presented tone) on different trials. As illustrated in Table 1, each of the 36 prime stimuli had two corresponding octave equivalent stimuli for a total of 72 unique octave equivalent prime–probe pairs (i.e., 36 semitone primes each paired with the remaining 2 equivalent probe tones). The same 36 prime stimuli had four corresponding octave ±1 semitone probes for a total of 144 nonequivalent pairs (i.e., half of which are presented with a probe separated by a whole number of octaves minus 1 semitone and half of which are presented with a probe separated by a whole number of octaves plus 1 semitone). Note that the higher pitched tone of the pair was presented equally often before and after the lower pitched tone. During the learning phase, we presented each octave equivalent pairing twice as often as each nonequivalent pairing such that we had equal numbers of each trial type. This is indicated by the numbers in Table 1, which indicate the relative frequency of the presentation of each pair during the learning phase, discussed further in the Procedure section below. All visual stimuli were presented in 20 pt Courier New font. Visual stimuli were presented on a projector screen in the front of a classroom and auditory stimuli were presented via the loudspeakers. Stimulus presentation and timing were controlled with an offline PsyToolkit (Stoet, 2010, 2017) experiment (see OSF link for scripts). Responses were coded on a paper by participants (see Procedure).

Procedure

There were three main phases in the experiment. Before the experiment started, however, students were given a brief presentation about octave equivalence, both to give a context for the subsequent group report that they needed to prepare and to clarify the goal of the task (i.e., to make it clear what an "equivalent" pairing means). An auditory example of equivalent tones was presented (F_4 and F_5) twice early on in the lecture. Students were also given a "mini experiment," similar to Deutsch (1972), in which they were asked to try to identify a song (we used the popular French children's song *Ah ! Les crocodiles*), first using a version with random octave shifted notes, then the original version, then the random octave shifted version again. No data were collected for this mini experiment. They also responded to two yes-or-no screening questions on their paper-and-pencil response grids (see below). The first was to measure musicianship (translated from French: "Are you a musician?") and the second to screen participants with auditory problems ("Do you have any auditory problems?").

The timing of the main phases of Experiment 1 is illustrated in Figure 2. The first main phase of the experiment was the pretest to evaluate octave equivalence perception prior to training. On each trial, participants were presented with tone pairs. Each trial began with 200 ms of silence, followed by a note for 200 ms, then 400 ms of silence, and then the second note for 200 ms. Between trials, there was another silence for 2,800 ms, the time needed to code responses. There was no feedback about the correct answer in this phase. Responses were coded by participants in a table, as also illustrated in Figure 2. Their task was to guess whether the two notes were "similar" or "different" (clarified in the instructions to mean "belonging to the same pitch class" or "belonging to two different pitch classes," respectively). There were 20 trials, presented in a pseudorandom order. Because participants were tested in a group setting, fixed stimulus orders needed to be generated for the test phases. Seven different pretests and seven different posttests (discussed below) were therefore created, each used for two groups of participants. To generate these 14 test lists, 140 equivalent pairs and 140 nonequivalent pairs were needed. All but 4 of the 144 possible nonequivalent pairs (see Materials and Apparatus) were therefore used exactly once across counterbalancing orders. All 72 of the equivalent pairs were presented at least once, and all but four were presented twice across counterbalancing orders (but never repeated for the same participant). See the OSF repository for the exact orders used and the excluded pairs. For all participants, half of the pairs of notes (10 trials) belonged to the same pitch class, but were separated by one or more octaves (e.g., C_4 and C_5). The other half were not in the same pitch class. About half of these nonequivalent pairs were separated by one or more octaves minus a semitone (e.g., F_4 and E_5), while the other half was separated by one or more octaves plus a semitone (e.g., E_4 and F_5). Whether the higher or lower pitched tone was presented first or second was also randomly varied. The trial number (1–20) was indicated throughout the trial in the middle of the screen (i.e., to help students keep track of the current trial number in their response table). This initial phase was conducted to test octave equivalence abilities prior to learning (excluding the limited number of examples given during the lecture, of course). The instructions for this phase were as follows (translated from French):

> In this phase, you will hear two notes on each trial, one after the other. Your goal is to determine if the two notes sound similar or different. In music, stepwise increases in pitch change the pitch class (e.g., do, do♯, ré, ré♯, etc.). However, if you keep increasing the intervals, you will arrive at another note in the same pitch class. For example: "do, re, mi, fa, sol, la, si, do".



Test 1 (Pretest)		
Trial	Similar	Different
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		

Test 2 (Posttest)		
Trial	Similar	Different
21		
22		
23		
24		
25		
26		
27		
28		
29		
30		
31		
32		
33		
34		
35		
36		
37		
38		
39		
40		

Figure 2. Stimulus and response timing in Experiment 1 and response tables (translated from French). Stimuli not to scale.

> Note that the first and last notes are both called "do". These two notes are not identical, but similar. Your goal is to determine if the two notes you hear sound like they belong to the same pitch class (similar), or if they sound different (different). This might be a difficult task. Guess if you are not sure.

The second phase was the learning task. This task was similar to the prior pretest phase, except that the correct answer ("similar" or "different") was written on the screen (i.e., instead of the trial number). The timing was only slightly different. Each trial started with a white screen for 800 ms, and then the first note was played for 200 ms. The correct answer ("similaire" or "différent" in French) appeared on the screen along with the first note and remained on the screen for the remainder of the trial. Then, there was a 400 ms silence, followed by the second note for 200 ms, followed by a 400 ms silence. The task was simply to read the correct answer. This is an incidental learning phase (for a review, see Schmidt, 2021a, 2021c). In other words, the explicit goal is simply to read the label rather than to try to learn octave equivalence. There were two blocks of 72 trials, presented in a random order, separated by a pause (144 trials in total). Trials were randomly selected (without replacement) from a list of stimuli containing all 144 unique nonequivalent pairs once and each of the 72 unique equivalent pairs twice (see Table 1). The instructions for this phase were (translated from French):

> This phase is that same as the previous phase, except that this time you do not need to guess. You will hear two notes on each trial, but the correct response will be presented on the screen. Your objective is simply to read the correct response ("similar" or "different").

Incidentally, there was no way to measure accuracy or even active engagement in this phase (see Discussion below for more on this).

The third and final phase was identical to the initial pretest phase. This was the posttest phase to determine whether participants improved their ability to detect the similarity (or difference) between two notes (i.e., to detect octave equivalence). Sensitivity (*d′*) was calculated separately for the pretest and posttest phases (see Data Analysis section below). The entire experiment (excluding the lecture before and after the experiment) lasted about 15 min.

Data Analysis

Hits (i.e., responding “similar” for equivalent pairs) and false alarms (i.e., responding “similar” for nonequivalent pairs) were coded for each of the two test phases and transformed into d' scores. Scores of 0% or 100% hits or false alarms were imputed with a $.5/n$ correction (Macmillan & Kaplan, 1985; Stanislaw & Todorov, 1999). For readers unfamiliar with signal detection theory, d' is a measure of sensitivity (in this case, the ability to detect octave equivalence) that is independent from response biases (e.g., the tendency of some participants to respond “different” to almost all stimuli). A d' of zero indicates no sensitivity (i.e., pure guessing), whereas scores higher than zero indicate sensitivity (for more detailed explanations, see Stanislaw & Todorov, 1999). All subsequent data analyses were performed in R. While statistical inferences were made based on *p*-values, Bayes factors were also calculated using the BayesFactor and bayestestR packages to obtain inclusion Bayes factor with matched models for the ANOVA and one-sample Bayesian *t*-tests for the comparisons to chance guessing. Raw data, experiment scripts, and R scripts for this and the following experiment are available on the Open Science Framework at https://osf.io/ndhgv/.

Results

Data for Experiment 1 are presented in Figure 3. An ANOVA with the within-group factor of phase (pretest vs. posttest) and the between-group factor of musicianship (nonmusicians vs. musicians) was first conducted. This revealed a significant main effect of musicianship, $F(1, 397) = 29.548$, $MSE = 0.622$, $p < .001$, $\eta_p^2 = .07$, $\mathrm{BF}_{10} > 10{,}000$, with better overall performance for musicians. There was also a main effect of phase, $F(1, 397) = 6.465$, $MSE = 0.426$, $p = .011$, $\eta_p^2 = .02$, $\mathrm{BF}_{10} = 1.91$, indicating higher scores posttest relative to pretest. There was no interaction between musicianship and phase, $F(1, 397) = 1.261$, $MSE = 0.426$, $p = .262$, $\eta_p^2 < .01$, $\mathrm{BF}_{10} = 0.266$.

Subsequent one-sample *t*-tests were conducted to evaluate whether participants were able to identify octave equivalences above chance (i.e., above a d' of 0). Pretest performance was significantly above chance for both the nonmusicians ($d' = 0.172$), $t(349) = 4.637$, $SE = 0.037$, $p < .001$, $\eta^2 = .06$, $\mathrm{BF}_{10} = 1875$, and musicians ($d' = 0.555$), $t(48) = 4.179$, $SE = 0.133$, $p < .001$, $\eta^2 = .27$, $\mathrm{BF}_{10} = 191$. Posttest performance was also significantly above chance for both the nonmusicians ($d' = 0.270$), $t(349) = 7.817$, $SE = 0.035$, $p < .001$, $\eta^2 = .15$, $\mathrm{BF}_{10} > 10{,}000$, and musicians ($d' = 0.811$), $t(48) = 5.010$, $SE = 0.162$, $p < .001$, $\eta^2 = .34$, $\mathrm{BF}_{10} = 2{,}425$.

Another ANOVA was run with the factors phase (pretest vs. posttest) and counterbalancing order (7 orders). This revealed a significant main effect of phase, $F(1,392) = 7.275$, $MSE = 0.378$, $p = .007$, $\eta_p^2 = .02$, $\mathrm{BF}_{10} = 2.01$, indicating improvements between pre- and posttest. There was also a main effect of counterbalancing order, $F(6,392) = 4.060$, $MSE = 0.637$, $p < .001$, $\eta_p^2 = .06$, $\mathrm{BF}_{10} = 15.48$, and an interaction, $F(6,392) = 9.363$, $MSE = 0.378$, $p < .001$, $\eta_p^2 = .13$, $\mathrm{BF}_{10} > 10{,}000$, seemingly indicating that some counterbalancing orders were easier than others.

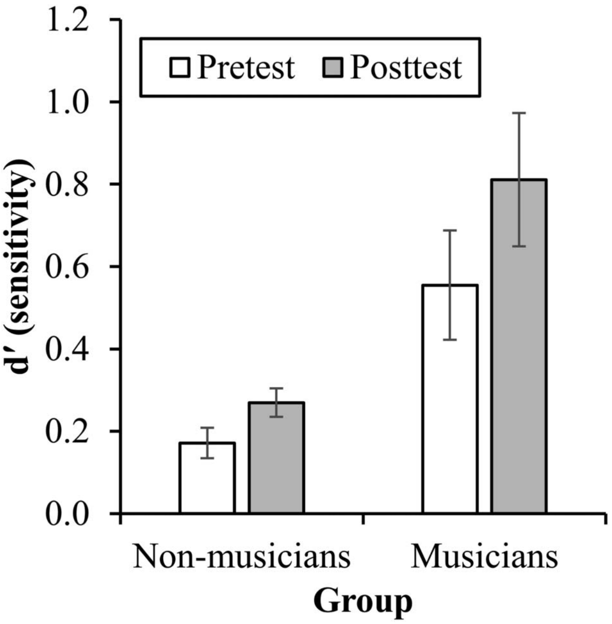


Figure 3. Experiment 1 pre- and posttest octave equivalence scores for musicians and nonmusicians with standard error bars.

As a final exploratory analysis, we assessed the number of participants with particularly high performance at pre- and posttest. Only 6 of 49 musicians (12.2%) and 5 of 350 nonmusicians (1.4%) had a d' score above 1.68 at pretest (corresponding to 80% correct responses with an even number of hits and false alarms). The corresponding numbers were 12 of 49 (24.5%) and 8 of 350 (2.3%) at posttest. The former finding clearly indicates that detection of octave equivalences is not easy, and the latter finding indicates that acquisition of such a skill is similarly difficult. Furthermore, only three musicians (6.1%) and one nonmusician (0.3%) attained the higher criterion of 2.96 (90% accuracy) at posttest. No participant was this accurate at pretest.

Discussion

Interestingly, both nonmusicians and musicians scored, on average, significantly above chance in the octave equivalence pretest. However, musicians were notably more accurate, coherent with past work on the question. Importantly, both groups improved with practice, as indicated by higher posttest scores relative to pretest scores. It might be further noted, however, that the vast majority of participants were far from 100% accuracy and the statistically significant effect should be taken with a grain of salt. Pretest scores were still, relatively speaking, poor. Indeed, posttest scores were also far from 100% accurate, indicating the difficulty of learning to identify octave equivalency by ear. There are a few caveats with Experiment 1, however. First, the group testing conditions were far from ideal. Indeed, given that participants did not even need to respond during the learning phase, inattention by some less-motivated students is perhaps inevitable (e.g., the first author, responsible for some of the tutorial groups, noted some students not even looking at the similar/dissimilar labels on the screen). These considerations could potentially lead to an underestimation of typical improvements with learning. More problematically, however, exploratory analyses revealed significant counterbalancing effects, with some of the orders showing robust pre–post improvements and others showing *reduced* accuracy at posttest. Replication was therefore deemed necessary.

Experiment 2

To mitigate against some of the shortcomings of Experiment 1, Experiment 2 was conducted as a conceptual replication of Experiment 1 with several improvements. First, participants were tested individually online to avoid the distraction of group testing conditions in the classroom. Second, the learning phase required active responding, which we hypothesized might increase engagement and therefore learning. Measuring accuracy during learning also allowed us to screen out participants giving low effort responses (perhaps particularly a concern for an online study). Third, stimuli in the testing phases were randomized for each participant individually to better control rank and order effects. Also, to maximize detection of an effect with a (somewhat) smaller sample, only one-octave differences were used in the range of C_3 to B_4. Greater than an octave differences are harder to detect, even for expert musicians (Russo & Thompson, 2005b). A detailed lecture about octave equivalence prior to the study was no longer feasible, but the task instructions still indicated to participants on what basis they should make similarity judgments.

Method

Participants

Participants were recruited online via https://www.prolific.com/. Participants signed an online consent form before beginning the experiment and were free to withdraw participation at any time without penalty. Participants were asked to verify that their sound was working before starting the experiment, but there was no built-in test to verify that participants heard sounds all throughout the experiment.[2] There was a total of 418 submissions, but this included returned submissions with low effort responses (see below). Prescreening questions were used to restrict the sample to fluent French speakers and those without auditory problems. Unlike in Experiment 1, accuracy was recorded during the learning phase. In the recruitment, participants were warned that they needed to maintain at least 80% accuracy during the (very easy) learning phase. Participants meeting this criterion (and several somewhat below this) were paid £2 for participation. For actual data analyses, only participants with five or fewer errors (of 96 trials)[3] were retained, leaving a total of 294 participants for analyses. The remaining sample consisted of 191 nonmusicians and 103 musicians.

Materials, Apparatus, Procedure, and Data Analysis

Experiment 2 was identical in all respects to Experiment 1 with the following exceptions. The experiment was conducted online. A detailed lecture about octave equivalence could not be given, but instructions were lengthened to clarify the task (see below). The same piano tones were used but restricted to the range of C_3 to B_4 and only pairs of 11 to 13 semitone differences were used (24 unique equivalent and 48 unique nonequivalent prime–probe pairs), as illustrated in Table 2. The numbers in the table again indicate the relative frequency of presentation of the different stimulus pairs during the learning phase. Responses were also coded online with a computer keyboard, using the F key for a "similar" response and the J key for a

---

[2] Of course, any participants who completed the experiment without sound would have only diluted any of our observed effects, as they would necessarily be guessing randomly the identity of the auditory stimuli that they could not hear.

[3] Although this a priori trim might seem strict and did reduce the sample noticeably (almost 30%), this accuracy criterion is quite a low bar considering how easy the task was. As described below and in the Methods of Experiment 1, the task was only two-choice, the response labels were presented on the screen continuously, and the correct response (matching one of the response labels) was printed in the middle of the screen a full 800 ms before a response could even be initiated. Participants then had a further 2,000 ms to respond (i.e., 2,800 ms total). Error rates were very non-normally distributed, with an otherwise normal distribution near ceiling (above our cutoff) and a very long left tail, with accuracy rates as low as 0% (e.g., some participants did not respond at all during the learning phase).

Table 2. Relative trial frequencies of prime–probe pairs in the learning phase of Experiment 2

	Probe																						
	Octave 3												Octave 4										
Prime	C	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯	B	C	C♯	D	D♯	E	F	F♯	G	G♯	A	A♯	B
C$_3$												1	2	1										
C♯$_3$													1	2	1									
D$_3$														1	2	1								
D♯$_3$															1	2	1							
E$_3$																1	2	1						
F$_3$																	1	2	1					
F♯$_3$																		1	2	1				
G$_3$																			1	2	1			
G♯$_3$																				1	2	1		
A$_3$																					1	2	1	
A♯$_3$																						1	2	1
B$_3$	1																						1	2
C$_4$	2	1																						1
C♯$_4$	1	2	1																					
D$_4$		1	2	1																				
D♯$_4$			1	2	1																			
E$_4$				1	2	1																		
F$_4$					1	2	1																	
F♯$_4$						1	2	1																
G$_4$							1	2	1															
G♯$_4$								1	2	1														
A$_4$									1	2	1													
A♯$_4$										1	2	1												
B$_4$											1	2	1											

"different" response. During all phases, a response key was presented on the bottom of the screen to remind participants of the key for each response. Specifically, an uppercase F above a lowercase "similaire" was presented on the bottom left and an uppercase J above a lowercase "différent" was presented on the bottom right.

The timing of the main phases of Experiment 2 are illustrated in Figure 4. During each of the test phases, 12 equivalent and 12 nonequivalent pairs were randomly selected for each participant. Unlike in Experiment 1, the trial number was not presented, as this was not needed (i.e., as participants responded directly with the keyboard rather than on a paper-and-pencil response grid). In addition, the 2,800 ms silence at the end of the trial was removed. Instead, a question mark appeared in the middle of the screen 500 ms after the second note and remained on the screen until the participant responded (no time limit). The question mark disappeared after a response. The next trial always started 2,300 ms after question mark onset (i.e., regardless of response time) to discourage low effort rapid responses (i.e., because rapid responses did not accelerate the trial advancement), except when the response time was already longer than 2,300 ms (in which case, the next trial started immediately). Instructions were only slightly modified (translated from French):

> In music, stepwise increases in pitch change the pitch class (e.g., do, do♯, ré, ré♯, etc.). However, if you keep increasing the intervals, you will arrive at another note in the same pitch class. For example: "do, re, mi, fa, sol, la, si, do." Note that the first and last notes are both called "do." These two notes are not identical, but similar. You will be tested for your ability to hear this "equivalence" before and after a period of training. In this phase, you will hear two notes on each trial, one after the other. Your goal is to determine if the two notes sound like they belong to the same pitch

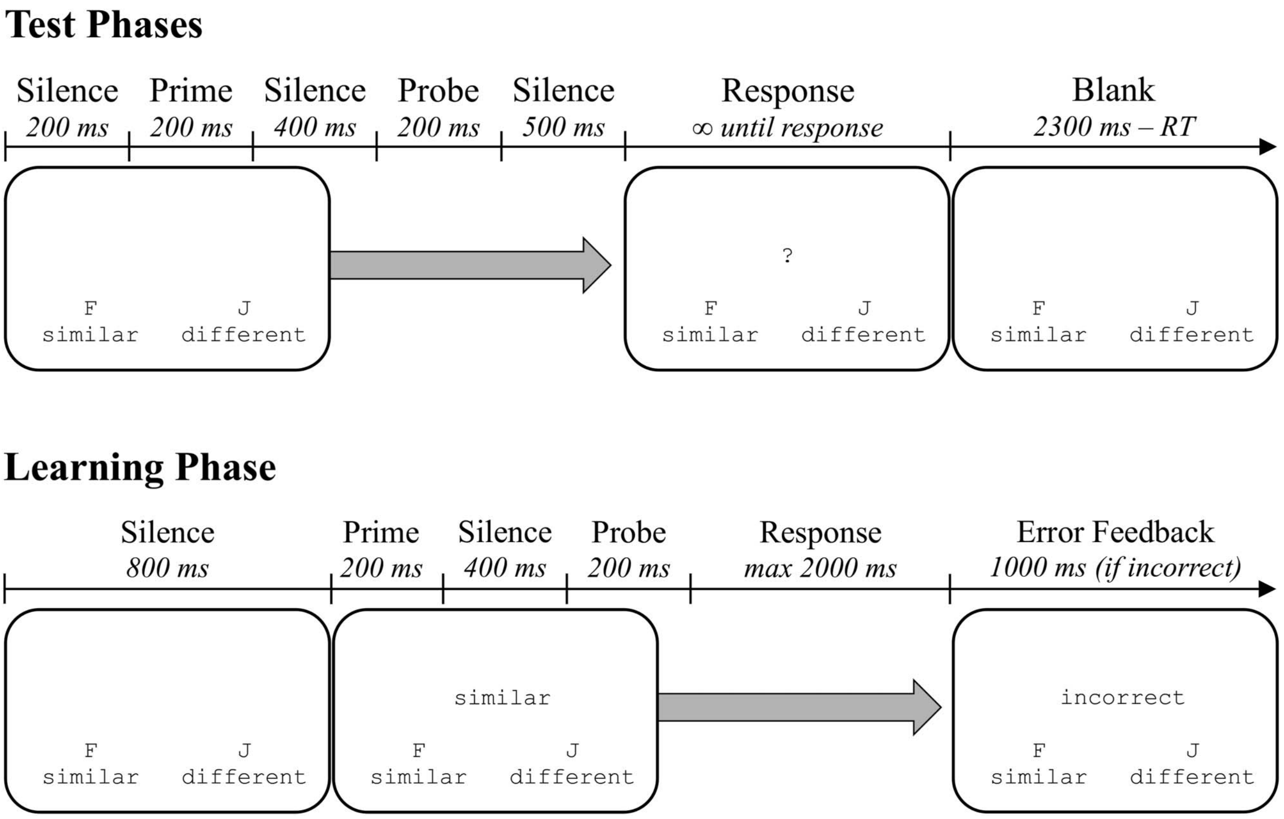


Figure 4. Stimulus and response timing in Experiment 2. Stimuli not to scale.

> class (F key), or if they sound different (J key). This might be a difficult task. Guess if you are not sure. Try to respond as accurately as possible. Rapid responses will not speed up the procedure.

The learning phase was similar to the prior experiment, with the following exceptions. There were two blocks of 48 trials. The stimulus timing was identical to Experiment 1. Participants had to wait to respond until the end of the second stimulus. After correct responses, the next trial started immediately. If participants made an error or did not respond within 2,000 ms, “erreur” (incorrect/error) was printed in red in the middle of the screen for 1,000 ms before the next trial started. Instructions were only slightly modified to refer to the response keys (translated from French):

> This phase is the same as the previous phase, except this time you do not need to guess. You will hear two notes on each trial, but the correct response will be presented on the screen. Your objective is to press the F key when you see the word “similar” and the J key when you see the word “different.” You must wait until the second note is finished to respond.

The entire experiment lasted about 10 min. Data analyses were also identical to those in Experiment 1, except that we further removed participants with too many errors during the learning phase (see Participants). This, of course, could not be done in Experiment 1, as no responses were recorded in the classroom setting. Individual data files were merged with CSVDataMerge (Schmidt, 2021b).

Results

Data for Experiment 2 are presented in Figure 5. An ANOVA with the within-group factor of phase (pretest vs. posttest) and the between-group factor of musicianship (nonmusicians vs. musicians) was first conducted. There was a main effect of phase, $F(1,292) = 5.380$, $MSE = 0.313$, $p = .021$, $\eta_p^2 = .02$, $\mathrm{BF}_{10} = 1.17$, indicating higher scores

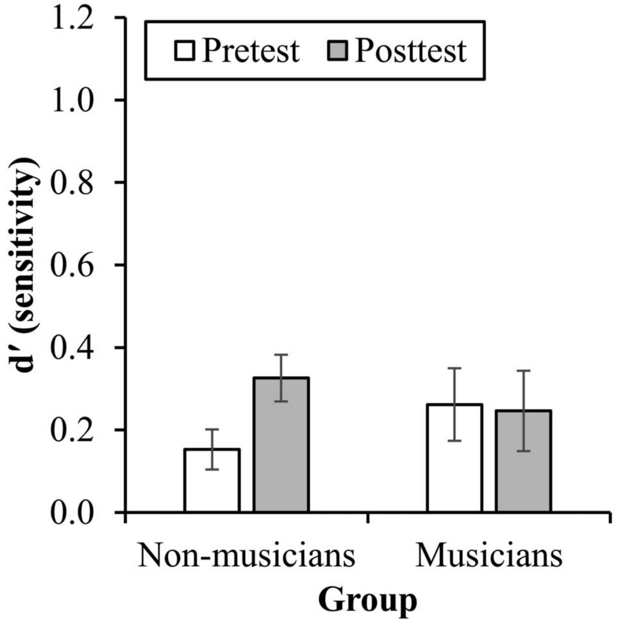


Figure 5. Experiment 2 pre- and posttest octave equivalence scores for musicians and nonmusicians with standard error bars.

posttest relative to pretest. The main effect of musicianship was not significant this time, $F(1, 292) = 0.029$, $MSE = 0.997$, $p = .866$, $\eta_p^2 < .01$, $BF_{10} = 0.17$. There was a marginal interaction between musicianship and phase, $F(1, 292) = 3.802$, $MSE = 0.313$, $p = .052$, $\eta_p^2 = .01$, $BF_{10} = 0.823$. Interestingly, the improvement was only significant for nonmusicians, $t(190) = 2.990$, $SE = 0.058$, $p = .003$, $\eta^2 = .04$, $BF_{10} = 5.98$, but not for musicians, $t(102) = 0.203$, $SE = 0.076$, $p = .839$, $\eta^2 < .01$, $BF_{10} = 0.11$.

Subsequent one-sample *t*-tests were conducted to evaluate whether participants were able to identify octave equivalences above chance. Pretest performance was significantly above chance for both the nonmusicians ($d' = 0.153$), $t(190) = 3.145$, $SE = 0.049$, $p = .002$, $\eta^2 = .05$, $BF_{10} = 9.35$, and the musicians ($d' = 0.263$), $t(102) = 2.976$, $SE = 0.088$, $p = .003$, $\eta^2 = .08$, $BF_{10} = 6.84$. Posttest performance was also significantly above chance for both the nonmusicians ($d' = 0.326$), $t(190) = 5.763$, $SE = 0.057$, $p < .001$, $\eta^2 = .15$, $BF_{10} > 10{,}000$, and the musicians ($d' = 0.246$), $t(102) = 2.528$, $SE = 0.097$, $p = .013$, $\eta^2 = .06$, $BF_{10} = 2.14$.

Again, we performed exploratory analyses to evaluate how many participants had high levels of performance. Only 7 of 103 musicians (6.8%) and 7 of 191 nonmusicians (3.7%) had a *d′* score above 1.68 at pretest, again indicating the difficulty of the task. The corresponding numbers were 9 of 103 (8.7%) and 10 of 191 (5.2%) at posttest, again indicating that learning this skill is difficult. Even fewer attained the higher criterion of 2.96, with three musicians (2.9%) and one nonmusician (0.5%) attaining this score at pretest and four musicians (3.9%) and four nonmusicians (2.1%) attaining this score at posttest.

Discussion

Again, both nonmusicians and musicians scored, on average, significantly above chance in the octave equivalence pretest. Interestingly, Experiment 2 did not reveal the same overall higher scores for musicians relative to nonmusicians that we observed in Experiment 1, although this seemed to be primarily due to the unexplained lack of improvement posttest for musicians. Furthermore, there was an overall improvement in equivalence detection with practice, although this was only apparent in nonmusicians in Experiment 2. Again, most participants were far from 100% accuracy. In fact, it is perhaps also worth noting that while some improvements were again observed, these improvements were not larger than in Experiment 1 as initially predicted. Of course, there were several differences between the two experiments (e.g., grouped vs. individual testing, classroom vs. online, length of training phase), but we had initially posited that the restriction of the tone pairs to 1-octave (±1 semitone) differences and active responding during learning would boost learning. This did not seem to be the case.

General Discussion

In the present work, we investigated the extent to which participants are able to detect the equivalence of octave-separated tones and to what extent this detection ability can be improved with practice. On average, both nonmusicians and musicians did correctly classify octave-equivalent tones as similar and nonequivalent tones as different more often than one would expect from pure guessing. In prior work on this question, instructions given to participants were more ambiguous (Allen, 1967; Kallman, 1982; Thurlow & Erchul, 1977), with participants being asked to judge the “similarity” of tones and sensitivity to octave equivalence seemed largely limited to a minority of musicians. Sample sizes were also much smaller in this prior research. It is also important to note, however, that pretest performance, although above chance-level guessing statistically, was still rather poor. Thus, the present results demonstrate that even when instructions are clearer, the task is not easy. Similarly, some improvements are possible with training, but it still remains challenging to easily distinguish between octave equivalent tone pairs and tones from two adjacent pitch classes. Indeed, improvements were statistically significant, but also rather small practically speaking.

Future research might explore the extent to which more extended training leads to better performance. Another multiday training study (Litke & Olsen, 1979) did already observe notable improvements in the ability to set a dial to adjust a second tone to exactly one octave higher than a cue tone. This was done with only three cue tones and there were also some questionable data analysis choices,[4] but mean performance in all age groups tested was not semitone precise (which would be needed for the task in the current work). Similarly, further research might be conducted to determine what type of learning environment might be most conducive to learning octave equivalence.

Although not the focus of the current series of experiments, the relationship between octave equivalence perception and absolute and/or relative pitch might be interesting to explore in future research. *Absolute pitch* (AP), or more informally *perfect pitch*, is the ability to name isolated tones by ear (for reviews, see Bachem, 1955; Deutsch, 2013; Di Stefano & Spence, 2024; Levitin & Rogers, 2005; Loui, 2016; Moulton, 2014; Schmidt, in press; Takeuchi & Hulse, 1993; Ward, 1999). This contrasts with *relative pitch* (*RP*). An RP possessor can identify intervals by ear. They can, for instance, identify pitches by ear correctly after being given a pitch of known identity as a context to "calculate" the pitch name of a subsequent tone, although this is much less rapid and automatic than AP perception (Levitin, 1994; Levitin & Rogers, 2005; Takeuchi & Hulse, 1993). Of course, the pitch pair comparison task used in the current studies explicitly requires a relative interval comparison of two notes (specifically, the unison) and there is some debate as to whether AP helps (e.g., Dooley & Deutsch, 2011) or hurts (e.g., Miyazaki et al., 2018) with RP perception. However, there is confusability between adjacent intervals (Killam et al., 1975; Rakowski, 1990) and interval perception becomes difficult with larger intervals (Russo & Thompson, 2005b), especially in the case of multiple octave differences. In contrast, AP possessors perceive the chroma of notes as automatically as someone with normal color vision labels the color of an object, such that F_3 and F_4, for example, are perceived as the *same pitch class* and not merely as being separated by a specific interval (12 semitones). In this sense, AP and octave equivalence may be similar to the extent that they both rely on pitch chroma perception. Indeed, AP possessors are seemingly so biased by pitch class that they often make octave errors (i.e., indicating the correct note but in the wrong octave) in a task that requires pitch height detection (Weisman et al., 2010). As hinted at above, however, another way of resolving the octave equivalence test used here is to rely on the perceived interval (rather than the perceived similarity as such). Thus, the extent to which octave equivalence perception is more related to AP or RP abilities could be studied in future work.

Relatedly, it might be proposed that some form of octave equivalence training might aid in the learning of AP. It is well-known that AP is difficult to learn as an adult (e.g., Gough, 1922; Heller & Auerbach, 1972; Meyer, 1899; Vianello & Evans, 1968; Wedell, 1934; for reviews, see Takeuchi & Hulse, 1993; Ward, 1999; cf., Mull, 1925), although some results suggest that improvements are possible (e.g., Cuddy, 1968; Henry et al., 2025; Henry & Schmidt, 2024, 2025; Iorio et al., 2024; Lundin & Allen, 1962; Terman, 1965; Van Hedger et al., 2015, 2019; Y. K. Wong, Lui, et al., 2020; Y. K. Wong, Ngan, et al., 2020). With an incidental learning procedure that we developed, for instance, rapid improvements in pitch identification were observed after about 15 min of training (Henry et al., 2025; Iorio et al., 2024). However, in more recent research, we have found that improvements, although not absent, are much less impressive when learning pitches from multiple octaves at once (Henry & Schmidt, 2024; see also, Bongiovanni et al., 2023). The same was not true when timbres were varied (Henry & Schmidt, 2025), although timbre can influence interval perception (Russo & Thompson, 2005a). Together, these results suggest that participants have a difficulty in focusing on pitch class, instead being more attracted to pitch height cues (see also, Wagner et al., 2022). Indeed, Henry and Schmidt (2024) did observe robust improvements in pitch identification of Shepard tones (see Shepard, 1964, 1982), for which pitch height is ambiguous, but this learning did not transfer well to piano tones in a subsequent test phase (i.e., where the pitch height cues are reintroduced). Speculatively, it might be supposed that octave equivalence training might boost the ease of AP learning if such equivalence training increases focus on pitch class rather than pitch height.

Future research might also explore to what extent the delay between pitches influences learning. For instance, detection of octave equivalence of *simultaneously* presented tones is notably easier than detection of equivalence between sequential tones (e.g., Bonnard et al., 2016; Demany et al., 2021; Demany & Semal, 1988). To what extent it would be helpful or harmful to overlap tones while learning to identify octave equivalence is less clear,

---

[4] For instance, second block error rates were compared with whichever subsequent block of trials showed the best performance on a participant-by-participant basis as a measure of learning, which will necessarily maximize on random error and exaggerate true learning. Indeed, final block performance was less impressive.

however. Similarly, the extent to which different types of octave equivalence training might help directing attention to pitch chroma rather than pitch height to facilitate subsequent AP training could be explored. For example, we briefly mentioned that our task is an incidental learning task because the main goal of the task is to identify/read the correct response label rather than to intentionally learn (e.g., with trial-and-error guessing with reinforcement). *Implicit learning* tasks imply incidental learning and nonconscious acquired knowledge (Berry & Dienes, 1993; Cleeremans et al., 1998; Perruchet, 2019; Perruchet & Pacteau, 1990; Reber, 1967, 1989; Shanks, 2005). *Explicit learning*, the exact opposite, is sometimes advantageous, particularly when a regularity is easy to learn (e.g., Destrebecqz, 2004; Schmidt & De Houwer, 2012), but can also be disadvantageous, particularly when a regularity is difficult to learn (Berry & Broadbent, 1988; Fletcher et al., 2005; Howard & Howard, 2001; Reber, 1976; Reber et al., 1980; Wulf et al., 1998). Implicit learning is generally also much faster than explicit learning (for a review, see Schmidt, 2021a, 2021c). Still, it might be pertinent in future research to compare octave equivalence training with intentional versus unintentional learning and across participants with different levels of awareness of what was learned.

There are also some caveats with the present work. For instance, in the current work, we argue that octave equivalence perception is difficult to learn, but an even stronger case might be that it is not learnable at all. Related to the discussion of relative pitch above, it may be that participants in our experiment learned to identify the interval between each pair of tones rather than the similarity between said pairs. Improvements at posttest may therefore represent interval learning rather than an improvement in octave equivalence detection. Indeed, musical intervals can be trained (e.g., Litke & Olsen, 1979; Little et al., 2019; S. S. H. Wong et al., 2021) and generally are trained with a musical education. Whether there is anything "special" or "extra" about the unison is less certain from the present results alone. This might prove difficult to disambiguate in future research given that the improvement effect was already quite small, and one would also need to take into consideration the fact that larger intervals are harder to identify than smaller ones (Russo & Thompson, 2005b). For instance, we might anticipate that it would be easier to learn to identify a perfect fifth (7 semitones) than a unison (12 semitones) even if there *is* something special about the octave.

Another caveat with the present work is the somewhat differing results in the two experiments. Our main goal, of course, was to determine to what extent octave equivalence perception improves with training, but the results for musicians were notably different in the two experiments, despite rather large sample sizes. In particular, musicians improved notably at posttest in Experiment 1, but this was not the case in Experiment 2. Whether this represents a Type 1 and/or Type 2 error or some particularity of the methodology is unclear. For instance, we already had to exclude numerous online participants for clearly low effort responses (a common problem for online studies). It could be that the musicians in our Experiment 2 found our study particularly boring and disengaged more than the nonmusicians in our sample. There was also a notable difference in the percentage of musicians in the two samples (12% in Experiment 1 and 35% in Experiment 2), although this perhaps indicates a self-selection bias in the second experiment. It is perhaps also worth noting that musicianship was established with a single question ("Are you a musician?"), which may have been interpreted differently by our student and general population samples. Relatedly, it may be that the nonmusicians scored above-chance in the pretest because some of these participants did, in fact, have some practice with octaves. Future research might address this problem by studying more quantifiable measures of musicianship (e.g., years of musical study).

References

Allen, D. (1967). Octave discriminability of musical and non-musical subjects. *Psychonomic Science*, *7*(12), 421–422. https://doi.org/10.3758/BF03331154

Bachem, A. (1955). Absolute pitch. *Journal of the Acoustical Society of America*, *27*(6), 1180–1185. https://doi.org/10.1121/1.1908155

Bannan, N., Dunbar, R., & Bamford, J. (2024). The evolution of gender dimorphism in the human voice: The role of octave equivalence. *Current Anthropology*, *65*(3), 503–527. https://doi.org/10.1086/730147

Berry, D. C., & Broadbent, D. E. (1988). Interactive tasks and the implicit-explicit distinction. *British Journal of Psychology*, *79*(2), 251–272. https://doi.org/10.1111/j.2044-8295.1988.tb02286.x

Berry, D. C., & Dienes, Z. (1993). *Implicit learning: Theoretical and empirical issues*. Lawrence Erlbaum Associates.

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

Bonnard, D., Dauman, R., Semal, C., & Demany, L. (2016). Harmonic fusion and pitch affinity: Is there a direct link? *Hearing Research*, *333*, 247–254. https://doi.org/10.1016/j.heares.2015.08.015

Borra, T., Versnel, H., Kemner, C., van Opstal, A. J., & van Ee, R. (2013). Octave effect in auditory attention. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(38), 15225–15230. https://doi.org/10.1073/pnas.1213756110

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: News from the front. *Trends in Cognitive Sciences*, *2*(10), 406–416. https://doi.org/10.1016/S1364-6613(98)01232-7

Cuddy, L. L. (1968). Practice effects in the absolute judgment of pitch. *The Journal of the Acoustical Society of America*, *43*(5), 1069–1076. https://doi.org/10.1121/1.1910941

Demany, L., & Armand, F. (1984). The perceptual reality of tone chroma in early infancy. *The Journal of the Acoustical Society of America*, *76*(1), 57–66. https://doi.org/10.1121/1.391006

Demany, L., Monteiro, G., Semal, C., Shamma, S., & Carlyon, R. P. (2021). The perception of octave pitch affinity and harmonic fusion have a common origin. *Hearing Research*, *404*, Article 108213. https://doi.org/10.1016/j.heares.2021.108213

Demany, L., & Semal, C. (1988). Dichotic fusion of two tones one octave apart: Evidence for internal octave templates. *The Journal of the Acoustical Society of America*, *83*(2), 687–695. https://doi.org/10.1121/1.396164

Destrebecqz, A. (2004). The effect of explicit knowledge on sequence learning: A graded account. *Psychologica Belgica*, *44*(4), 217–247. https://doi.org/10.5334/pb-44-4-217

Deutsch, D. (1972). Octave generalization and tune recognition. *Perception & Psychophysics*, *11*(6), 411–412. https://doi.org/10.3758/BF03206280

Deutsch, D. (1979). Octave generalization and the consolidation of melodic information. *Canadian Journal of Psychology*, *33*(3), 201–205. https://doi.org/10.1037/h0081718

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

Deutsch, D., & Boulanger, R. C. (1984). Octave equivalence and the immediate recall of pitch sequences. *Music Perception*, *2*(1), 40–51. https://doi.org/10.2307/40285281

Deutsch, D., Dooley, K., & Henthorn, T. (2008). Pitch circularity from tones comprising full harmonic series. *The Journal of the Acoustical Society of America*, *124*(1), 589–597. https://doi.org/10.1121/1.2931957

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

Di Stefano, N., & Spence, C. (2024). Should absolute pitch be considered as a unique kind of absolute sensory judgment in humans? A systematic and theoretical review of the literature. *Cognition*, *249*, Article 105805. https://doi.org/10.1016/j.cognition.2024.105805

Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, *30*, 412–431. https://doi.org/10.1016/0001-6918(69)90065-1 (Original work published 1868)

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), 4097–4104. https://doi.org/10.1121/1.3652861

Dowling, W. J., & Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception & Psychophysics*, *21*(1), 60–64. https://doi.org/10.3758/BF03199469

Fletcher, P. C., Zafiris, O., Frith, C. D., Honey, R. A. E., Corlett, P. R., Zilles, K., & Fink, G. R. (2005). On the benefits of not trying: Brain activity and connectivity reflecting the interactions of explicit and implicit sequence learning. *Cerebral Cortex*, *15*(7), 1002–1015. https://doi.org/10.1093/cercor/bhh201

Gough, E. (1922). The effects of practice on judgments of absolute pitch. *Archives of Psychology*, *7*(47), i–93.

Heller, M. A., & Auerbach, C. (1972). Practice effects in the absolute judgment of frequency. *Psychonomic Science*, *26*(4), 222–224. https://doi.org/10.3758/BF03328601

Henry, W., Iorio, C., & Schmidt, J. R. (2025). Acquiring pitch-label associations using the twelve semitones and perfect contingencies: An auditory contingency learning task [Manuscript in preparation].

Henry, W., & Schmidt, J. R. (2024). Octave equivalence and the incidental learning of absolute pitch [Manuscript submitted for publication].

Henry, W., & Schmidt, J. R. (2025). Stroop-like effects in pitch identification training and generalization to untrained timbres: Evidence from a contingency learning task. *Music Perception*, *43*(3), 207–224. https://doi.org/10.1525/mp.2024.aa002

Hoeschele, M., Weisman, R. G., & Sturdy, C. B. (2012). Pitch chroma discrimination, generalization, and transfer tests of octave equivalence in humans. *Attention, Perception, & Psychophysics*, *74*(8), 1742–1760. https://doi.org/10.3758/s13414-012-0364-2

Howard, D. V., & Howard, J. H. (2001). When it does hurt to try: Adult age differences in the effects of instructions on implicit pattern learning. *Psychonomic Bulletin & Review*, *8*(4), 798–805. https://doi.org/10.3758/bf03196220

Idson, W. L., & Massaro, D. W. (1976). Octave convergence in melodic perception. *The Journal of the Acoustical Society of America*, *60*(S1), Article S41. https://doi.org/10.1121/1.2003334

Iorio, C., Bigand, E., & Schmidt, J. R. (2024). Incidentally acquiring label-pitch associations with a musical contingency learning task. *Collabra: Psychology*, *10*(1), Article 118806. https://doi.org/10.1525/collabra.118806

Jacoby, N., Undurraga, E. A., McPherson, M. J., Valdés, J., Ossandón, T., & McDermott, J. H. (2019). Universal and non-universal features of musical pitch perception revealed by singing. *Current biology: CB*, *29*(19), 3229–3243.e12. https://doi.org/10.1016/j.cub.2019.08.020

Kallman, H. J. (1982). Octave equivalence as measured by similarity ratings. *Perception & Psychophysics*, *32*(1), 37–49. https://doi.org/10.3758/BF03204867

Kallman, H. J., & Massaro, D. W. (1979). Tone chroma is functional in melody recognition. *Perception & Psychophysics*, *26*(1), 32–36. https://doi.org/10.3758/BF03199859

Killam, R. N., Lorton, P. V., & Schubert, E. D. (1975). Interval recognition: Identification of harmonic and melodic intervals. *Journal of Music Theory*, *19*(2), 212–234. https://doi.org/10.2307/843589

Levitin, D. J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Perception & Psychophysics*, *56*(4), 414–423. https://doi.org/10.3758/bf03206733

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

Litke, E. M., & Olsen, C. C. (1979). Acquiring relative pitch perception as a function of age. *Journal of Research in Music Education*, *27*(4), 243–254. https://doi.org/10.2307/3344711

Little, D., Cheng, H. H., & Wright, B. (2019). Inducing musical-interval learning by combining task practice with periods of stimulus exposure alone. *Attention, Perception & Psychophysics*, *81*(1), 344–357. https://doi.org/10.3758/s13414-018-1584-x

Loui, P. (2016). Absolute pitch. In S. Hallam, I. Cross, & M. H. Thaut (Eds.), *The Oxford Handbook of Music Psychology* (pp. 81–94). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198722946.013.10

Lundin, R. W., & Allen, J. D. (1962). A technique for training perfect pitch. *Psychological Record*, *12*(2), 139–146. https://doi.org/10.1007/BF03393451

Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*(1), 185–199. https://doi.org/10.1037/0033-2909.98.1.185

Marjieh, R., Griffiths, T. L., & Jacoby, N. (2023). *Musical pitch has multiple psychological geometries* (pp. 2023.06.13.544763). bioRxiv. https://doi.org/10.1101/2023.06.13.544763

Massaro, D. W., Kallman, H. J., & Kelly, J. L. (1980). The role of tone height, melodic contour, and tone chroma in melody recognition. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(1), 77–90. https://doi.org/10.1037/0278-7393.6.1.77

Meyer, M. (1899). Is the memory of absolute pitch capable of development by training? *Psychological Review*, *6*(5), 514–516. https://doi.org/10.1037/h0069034

Miyazaki, K., Rakowski, A., Makomaska, S., Jiang, C., Tsuzaki, M., Oxenham, A. J., Ellis, G., & Lipscomb, S. D. (2018). Absolute pitch and relative pitch in music students in the east and the west: Implications for aural-skills education. *Music Perception*, *36*(2), 135–155. https://doi.org/10.1525/mp.2018.36.2.135

Moulton, C. (2014). Perfect pitch reconsidered. *Clinical Medicine*, *14*(5), 517–519. https://doi.org/10.7861/clinmedicine.14-5-517

Mull, H. K. (1925). The acquisition of absolute pitch. *The American Journal of Psychology*, *36*(4), 469–493. https://doi.org/10.2307/1413906

Nishimura, T., Tokuda, I. T., Miyachi, S., Dunn, J. C., Herbst, C. T., Ishimura, K., Kaneko, A., Kinoshita, Y., Koda, H., Saers, J. P. P., Imai, H., Matsuda, T., Larsen, O. N., Jürgens, U., Hirabayashi, H., Kojima, S., & Fitch, W. T. (2022). Evolutionary loss of complexity in human vocal anatomy as an adaptation for speech. *Science*, *377*(6607), 760–763. https://doi.org/10.1126/science.abm1574

Perruchet, P. (2019). What mechanisms underlie implicit statistical learning? Transitional probabilities versus chunks in language learning. *Topics in Cognitive Science*, *11*(3), 520–535. https://doi.org/10.1111/tops.12403

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: Implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General*, *119*(3), 264–275. https://doi.org/10.1037/0096-3445.119.3.264

Peter, B., Larkin, T., & Stoel-Gammon, C. (2009). Octave-shifted pitch matching in nonword imitations: the effects of lexical stress and speech sound disorder. *The Journal of the Acoustical Society of America*, *126*(4), 1663–1666. https://doi.org/10.1121/1.3203993

Peter, B., Stoel-Gammon, C., & Kim, D. (2008). Octave equivalence as an aspect of stimulus-response similarity during nonword and sentence imitations in young children. *Proceedings of the 4th International Conference on Speech Prosody, SP 2008* (pp. 731–734), Campinas, Brazil, May 6–9, 2008.

Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. *Psychology of Music*, *18*(1), 60–72. https://doi.org/10.1177/0305735690181005

Reber, A. S. (1967). Implicit learning of artificial grammars. *Journal of Verbal Learning and Verbal Behavior*, *6*(6), 855–863. https://doi.org/10.1016/S0022-5371(67)80149-X

Reber, A. S. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(1), 88–94. https://doi.org/10.1037/0278-7393.2.1.88

Reber, A. S. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*(3), 219–235. https://doi.org/10.1037/0096-3445.118.3.219

Reber, A. S., Kassin, S. M., Lewis, S., & Cantor, G. (1980). On the relationship between implicit and explicit modes in the learning of a complex rule structure. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 492–502. https://doi.org/10.1037/0278-7393.6.5.492

Regev, T. I., Nelken, I., & Deouell, L. Y. (2019). Evidence for linear but not helical automatic representation of pitch in the human auditory system. *Journal of Cognitive Neuroscience*, *31*(5), 669–685. https://doi.org/10.1162/jocn_a_01374

Russo, F. A., & Thompson, W. F. (2005a). An interval size illusion: The influence of timbre on the perceived size of melodic intervals. *Perception & Psychophysics*, *67*(4), 559–568. https://doi.org/10.3758/BF03193514

Russo, F. A., & Thompson, W. F. (2005b). The subjective size of melodic intervals over a two-octave range. *Psychonomic Bulletin & Review*, *12*(6), 1068–1075. https://doi.org/10.3758/BF03206445

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

Schmidt, J. R. (2021b). CSVDataMerge: A simple and free program for concatenating experimental data files. *Journal of Open Research Software*, *9*(1), Article 34. https://doi.org/10.5334/jors.368

Schmidt, J. R. (2021c). Incidental learning of simple stimulus-response associations: A review of colour-word contingency learning research. *Année Psychologique*, *121*(2), i–xlvi. https://doi.org/10.3917/anpsy1.212.0077

Schmidt, J. R. (2025, May 6). *Octave equivalence: Difficult to perceive, but improvements are possible with training* [Data, Analysis Scripts]. https://osf.io/ndhgv/

Schmidt, J. R. (in press). Is absolute pitch learnable? Implicit and explicit absolute pitch. *Music Perception*.

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

Shanks, D. R. (2005). Implicit learning. In K. Lamberts, & R. L. Goldstone (Eds.), *Handbook of Cognition* (pp. 202–220). Sage Publications. https://doi.org/10.4135/9781848608177

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

Shepard, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, *89*(4), 305–333. https://doi.org/10.1037/0033-295X.89.4.305

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments, & Computers*, *31*(1), 137–149. https://doi.org/10.3758/BF03207704

Stoet, G. (2010). PsyToolkit: A software package for programming psychological experiments using Linux. *Behavior Research Methods*, *42*(4), 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

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

Terman, M. (1965). Improvement of absolute pitch naming. *Psychonomic Science*, *3*(6), 243–244. https://doi.org/10.3758/BF03343117

Thurlow, W. R., & Erchul, W. P. (1977). Judged similarity in pitch of octave multiples. *Perception & Psychophysics*, *22*(2) 177–182. https://doi.org/10.3758/BF03198752

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. (2019). Absolute pitch can be learned by some adults. *PLoS ONE*, *14*(9), Article e0223047. https://doi.org/10.1371/journal.pone.0223047

Vianello, M. A., & Evans, S. H. (1968). Note on pitch discrimination learning. *Perceptual and Motor Skills*, *26*(2), Article 576. https://doi.org/10.2466/pms.1968.26.2.576

Wagner, B., & Hoeschele, M. (2022). The links between pitch, timbre, musicality, and social bonding from cross-species research. *Comparative Cognition & Behavior Reviews*, *17*, 13–32. https://doi.org/10.3819/CCBR.2022.170002

Wagner, B., Sturdy, C. B., Weisman, R. G., & Hoeschele, M. (2022). Pitch chroma information is processed in addition to pitch height information with more than two pitch-range categories. *Attention, Perception, & Psychophysics*, *84*(5), 1757–1771. https://doi.org/10.3758/s13414-022-02496-1

Ward, W. D. (1999). Absolute pitch. In D. Deutsch (Ed.), *The Psychology of Music* (pp. 265–298). Academic Press.

Wedell, C. H. (1934). The nature of the absolute judgment of pitch. *Journal of Experimental Psychology*, *17*(4), 485–503. https://doi.org/10.1037/h0070761

Weisman, R. G., Balkwill, L.-L., Hoeschele, M., Moscicki, M. K., Bloomfield, L. L., & Sturdy, C. B. (2010). Absolute pitch in boreal chickadees and humans: Exceptions that test a phylogenetic rule. *Learning and Motivation*, *41*(3), 156–173. https://doi.org/10.1016/j.lmot.2010.04.002

Wong, S. S. H., Chen, S., & Lim, S. W. H. (2021). Learning melodic musical intervals: To block or to interleave? *Psychology of Music*, *49*(4), 1027–1046. https://doi.org/10.1177/0305735620922595

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

Wong, Y. K., Ngan, V. S., Cheung, L. Y., & Wong, A. C.-N. (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

Wulf, G., Höß, M., & Prinz, W. (1998). Instructions for motor learning: Differential effects of internal versus external focus of attention. *Journal of Motor Behavior*, *30*(2), 169–179. https://doi.org/10.1080/00222899809601334

History
Received May 6, 2025
Revision received September 12, 2025
Accepted October 24, 2025

Acknowledgments

I would like to thank Léa Entzmann and Florent Zecchini for their help collecting the data for Experiment 1.

Open Science

The manuscript includes a DOI generated by a public data repository pointing to the raw data underlying the findings reported in the article.

Raw data, data analysis scripts, and PsyToolkit code for both experiments along with the response grids and the counterbalancing orders for Experiment 1 (not applicable to Experiment 2) are available at https://osf.io/ndhgv/ (Schmidt, 2025).

Pre-registration: My manuscript contains no experiment with a completely executed pre-registration.

ORCID
James R. Schmidt
https://orcid.org/0000-0002-0412-396X

James R. Schmidt
CNRS UMR5022 – LEAD, Pôle AAFE
Université Bourgogne Europe
11 Esplanade Erasme
21000 Dijon
France
james.schmidt@ube.fr