Acquiring absolute pitch in adulthood is difficult but possible

Yetta Kwailing Wong[1]

Kelvin F. H. Lui[2]

Ken H.M. Yip[2]

Alan C.-N. Wong[2]

[1]Department of Educational Psychology, Faculty of Education

The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

[2]Department of Psychology

The Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Running head: ABSOLUTE PITCH ACQUISITION

Keywords: Absolute pitch, pitch perception, musical training, perceptual learning, perceptual expertise, critical period

Corresponding author:

(1) Yetta K. Wong
Rm 308, Ho Tim Building,
Department of Educational Psychology, Faculty of Education,
The Chinese University of Hong Kong,
Shatin, N.T., Hong Kong
Phone: +852-3943-3410
Email: yetta.wong@gmail.com

(2) Alan C.-N. Wong
344 Sino Building
Department of Psychology
The Chinese University of Hong Kong, Shatin, Hong Kong
Email: alanwong@cuhk.edu.hk
Phone number: +852-3943-6505

Abstract

Absolute pitch (AP) refers to the rare ability to name the pitch of a tone without external reference. It is widely believed that acquiring AP in adulthood is impossible, since AP is only for the selected few with rare genetic makeup and early musical training. In three experiments, we trained adults to name pitches for 12 to 40 hours. After training, 14% of the participants (6 out of 43) were able to name twelve pitches at 90% accuracy or above, with semitone errors considered incorrect. This performance level was comparable to that of real-world 'AP possessors'. AP training showed classic characteristics of perceptual learning, including performance enhancement, generalization of learning and sustained improvement for at least one to three months. Exploratory extrapolation analyses suggest that 39.5% and 58.1% of the participants would acquire AP if the training lasted for 60 and 180 hours respectively, suggesting the potential for the majority of the participants to acquire AP. We demonstrate that AP continues to be learnable in adulthood. The extent to which one acquires AP may thus be better explained by the amount and type of perceptual experience.

Absolute pitch (AP) refers to the ability to name the pitch of a tone (e.g., naming a tone as "C") or to produce it without external reference tones (Takeuchi & Hulse, 1993; Ward, 1999). While the majority of us can effortlessly identify a countless number of faces, objects, and visual and auditory words, most people find it very difficult to name the twelve pitches, and professional musicians are of no exception (Athos et al., 2007; Levitin & Rogers, 2005; Zatorre, 2003). The most extreme estimate states that, in every 10000 people, there is one 'AP possessor' who can perform AP judgment accurately and effortlessly (Takeuchi & Hulse, 1993). This rare ability is considered a special talent and endowment for gifted musicians (Deutsch, 2002; Takeuchi & Hulse, 1993; Ward, 1999). The genesis of AP has therefore been a perplexing research topic among musicians, psychologists and neuroscientists for more than a century (Deutsch, 2002; Levitin & Rogers, 2005; Takeuchi & Hulse, 1993; Ward, 1999).

In the acquisition of AP, two prerequisites have been widely accepted: the rare genetic disposition to AP, and an early onset of musical training that is within a critical period in childhood similar to that of language development (Baharloo, Johnston, Service, Gitschier, & Freimer, 1998; Chin, 2003; Drayna, 2007; Levitin & Rogers, 2005; Ross, Olson, Marks, & Gore, 2004; Takeuchi & Hulse, 1993; Zatorre, 2003). Accordingly, most professional musicians fail to acquire AP because they do not carry the specific genes and/or because they fail to start their music training within the critical period. Training AP in adulthood, when the critical period of acquiring AP has long passed, should thus be practically impossible.

Past studies showed that AP can improve to some limited extent in adulthood with deliberate practice (Brady, 1970; Cuddy, 1968, 1970; Meyer, 1899; Van Hedger, Heald, Koch, & Nusbaum, 2015), but there is no convincing evidence that adults can attain a performance level comparable to the AP possessors through training (Levitin & Rogers,

2005; Ward, 1999). Nevertheless, all previous training studies had low training intensity (roughly 1-4 hours; Cuddy, 1970; Van Hedger et al., 2015) and/or small sample sizes (3 or less participants per condition; Brady, 1970; Cuddy, 1968; Meyer, 1899). It thus remains unknown if acquisition of AP in adulthood is possible with a more rigorous training protocol.

The current study examined the possibility of AP acquisition in adulthood with perceptual learning paradigms. Perceptual learning has repeatedly demonstrated the human ability to pick up environmental input and fine-tune their perceptual representations in all sensory modalities (Fujioka, Ross, Kakigi, Pantev, & Trainor, 2006; Goldstone, 1998; Kraus & Banai, 2007; Sasaki, Nanez, & Watanabe, 2010; Seitz & Watanabe, 2009; Y. K. Wong, Folstein, & Gauthier, 2011). In three experiments, we trained 43 adults to name the pitch of tones with different combinations of timbres and octaves for 12-40 hours in laboratory and mobile online settings (Table 1). If specific genetic disposition and an early onset of music training are essential for AP acquisition, AP training in adulthood should be largely in vain resulting in very limited improvement in all participants. Alternatively, if AP can be trained in adulthood as a type of perceptual learning, it should be possible, at least for some individuals, to attain a performance level similar to that of real-world 'AP possessors'. Also, we should observe performance enhancement, generalization of learning, and sustained improvement similar to perceptual learning studies (Fahle & Poggio, 2002; Goldstone, 1998; Y. K. Wong et al., 2011).

Table 1. Details of the training protocols in the three experiments. N indicates the number of participants in each experiment. 'Mus/Non-Mus' indicates the number of musicians and non-musicians in each experiment respectively.

	Participants		Training				Test for Generalization	
Experiment	N	Mus/Non-Mus	Mode	Duration(Hr)	Octave	Timbre	Octave	Timbre
1	10	7/3	Lab	12	4,5	sine, piano	3,6	violin
2	22	10/12	Lab	15	4	sine	5	piano, violin
3	11	6/5	Online	40	4,5	Piano, violin, clarinet	6	sine

Experiment 1

In Experiment 1, we used a longer and more intensive perceptual learning protocol than previous studies (Cuddy, 1970; Van Hedger et al., 2015; Brady, 1980; Cuddy, 1968; Meyer, 1899) to test whether learning AP in adulthood is possible, and whether the training-induced improvement follows the classic characteristics of perceptual learning. Since how much training contributes to a sufficiently rigorous training to enable AP acquisition in adulthood is an untested empirical question, we decided to conduct a 12-hour training in this experiment, which was comparable to some previous perceptual learning studies in our laboratory (A. C. Wong, Palmeri, & Gauthier, 2009; Y. K. Wong et al., 2011).

Methods

Participants

Ten adults were recruited at City University of Hong Kong and completed the training. They included 2 males and 8 females, who were 23.1 years old on average (*SD* = 4.50). Seven of them were trained in music for 2-10 years, with the major instrument being

piano (N = 5), violin (N = 1) and flute (N = 1). Three were non-musicians who were not formally trained with music before. One additional participant dropped out in the middle of the training and was excluded from all analyses. All participants filled out a questionnaire about their musical training background, including the musical instruments and the highest ABRSM exam passed, and reported if they regarded themselves as 'AP possessors'. They received monetary compensation for the training and testing. Informed consent was obtained according to the Ethics Committee of City University of Hong Kong.

The sample size was estimated based on a recent AP training study (Van Hedger et al., 2015) using GPower 3.1.9.2. In this study, a large effect size was observed for the training improvement in adults (pretest vs. posttest; f = 1.34). Using the same f, the sample size required to detect any training effect at p = .05 with a power of 0.95 was 5 participants. To be more conservative, we recruited 10 participants. This sample size was also consistent with that used in previous perceptual training studies (Chung & Truong, 2013; Y. K. Wong et al., 2011).

Materials

The experiment was conducted on personal computers using Matlab (Natick, MA) with the PsychToolbox extension (Brainard, 1997; Pelli, 1997) at the Cognition and Neuroscience Laboratory at City University of Hong Kong. Participants were requested to bring their own earphone to the training and testing. They adjusted the volume to a comfortable level before the training or testing started.

In Experiment 1, 120 tones from octaves three to six were used. They were complex sine wave tones and piano tones in octaves three to six, and violin tones in octaves four to

five. The complex sinewave tones were identical to those in prior AP tests, and was generated by summing a series of sinusoidal waveforms including the fundamental frequency and harmonics (Bermudez, Lerch, Evans, & Zatorre, 2009). The piano tones were recorded with an electric keyboard (Yamaha S31). The violin tones were recorded by a volunteer violinist in a soundproof room. The precision of the tones was checked during recording by a tuner. The sound clips were 32-bit with a sampling rate of 44100Hz. They were edited in Audacity such that they lasted for 1 second with a 0.1-second linear onset and 0.1-second linear offset and were matched with similar perceptual magnitude.

Absolute pitch training

The training included 48 tones from two octaves (C4 to B5) in two timbres of complex sine wave and piano. A pitch-naming task was used. During each trial, an isolated tone was presented for 1s. Then, an image that mapped the 12 pitch names to 12 keys of the keyboard (from '1' to '=' at the top row of keys on a standard keyboard) was presented. Participants were required to name the pitch of the presented tone by keypress within 5s.

The training was gamified and structured with different levels. If participants achieved 90% accuracy for a certain level, they would proceed to the next level; otherwise they would stay at the same level. The training was completed by finishing 12 hours of training or by passing all 80 levels with 90% accuracy. Participants finished one hour of training per day. They were trained on at least four days per week and finished the training in three weeks.

The 80-level training protocol was organized into ten 8-level parts with an increasing number of pitches (from 3 pitches in the first 8 levels to 12 pitches in the last 8 levels). Each

eight-level part consisted of four types of levels, which included tones that were progressively richer in timbres and octaves. Each of the four types of levels were repeated twice, once with trial-by-trial feedback provided, and then once without feedback. For example, participants began the training with three pitches (E, F and F#). At levels 1-2, complex sine wave tones in these three pitches in octave 4 were included, with feedback provided at level 1 and then without feedback at level 2. At levels 3-4, complex sine wave tones in both octaves 4 and 5 were included with feedback and then without feedback. At levels 5-6, complex sine wave tones and piano tones in octave 4 were included with feedback and then without feedback. At levels 7-8, complex sine wave tones and piano tones in octaves 4 and 5 were included with feedback and then without feedback. At the no-feedback levels, participants were not provided with any external feedback of the correctness of the tones, so they could not establish any external reference for the AP naming. Instead, they could only generate answers internally in an absolute manner. Therefore, these no-feedback levels served as mini milestones for participants' AP performance at 90% accuracy. If they achieved 90% accuracy at the 8th level, a new pitch was added into the training set, with which they went through the same 8-level part again. Each level included 20 trials, with tones distributed as evenly as possible among the training pitches, octaves and timbres. Semitone errors were considered errors in the training. Before each level, participants were allowed to freely listen to sample tones of the training pitches as many times as they preferred before proceeding to the training. Each training session lasted for an hour, in which individual participants might have finished different numbers of training trials depending on their pace of learning (e.g., the amount of time spent on the training trials or on sample tone listening).

Normally participants earned one point for each correct answer in each trial. To motivate participants, a special trial that was worth three points randomly appeared with a chance of 1/80. Also, participants were given 1, 2 and 3 tokens if they achieved 60%, 75%

and 90% accuracy at a training level respectively. With ten tokens, participants would obtain a chance to initiate the three-point special trial when preferred. At most three chances of initiating these special trials at one level were allowed. The special trials did not appear and could not be initiated during the no-feedback levels. This ensured that participants performed the no-feedback levels without any scoring assistance.

Test for generalization

The test for generalization was performed before and within three days after training to examine how well the pitch-naming abilities generalized to untrained octaves and timbres. 120 tones in octaves 3 to 6 were used, in which octaves 4 and 5 were trained, and 3 and 6 untrained. Three timbres were included, with complex sine wave and piano as trained timbres, and violin as an untrained timbre. The tones were presented in three conditions, either with trained octave and timbre, trained octave and untrained timbre, or untrained octave and trained timbre. During each trial, a tone was presented for 1s. Then an image that mapped the 12 pitch names to 12 keys of the keyboard, which was the same as that used in the training, was presented. Participants were required to name the pitch of the presented tone by keypress within 5s. Each tone was presented twice, leading to 240 trials in total. The trials were presented in randomized order. No practice trials were provided in these tests. The dependent measure was the precision of pitch naming, i.e., the average semitone error of participants' responses relative to the correct responses. We adopted this dependent measure instead of the general naming accuracy because measuring the size of the judgment errors additionally informs the precision of pitch naming performance of the individuals, which is more informative than the binary correctness of the responses as measured by general naming

accuracy. An identical test was performed a month later to test whether the AP learning sustained for at least a month.

Results

Acquisition of AP

In general participants made substantial progress in learning to name pitches. At the end of training, they were able to name on average 8.1 pitches (out of 12) at 90% accuracy without any externally provided reference tones or scoring assistance (see Methods), under the stringent scoring criterion of taking semitone errors as incorrect (Figure 1A). Importantly, one of the ten participants passed all levels of training, meaning that he was able to name all of the twelve pitches at 90% accuracy without any externally provided reference tones, suggesting that he has acquired AP through perceptual learning in adulthood.

Is this level of AP performance representative of that of the real-world 'AP possessors'? While the verbal definition of 'AP possessors', i.e., one can name pitches accurately without external references, was widely agreed, there was not a single objective definition of the performance level of 'AP possessors'. We surveyed the literature on Web of Science on 19th April, 2017 with the term 'absolute pitch' in the topic and identified 133 empirical papers. These papers used highly varied definitions of 'AP possessors', including self-report, AP performance measurements, or relative performance on AP tasks (such as 3 SDs higher in AP accuracy than 'non-AP possessors'). We focused on the 66 publications that defined AP objectively based on AP performance instead of self-report, and found that these papers adopted highly varied performance measures to define 'AP possessors', including scoring methods (taking semitone errors as correct, partially correct or incorrect;

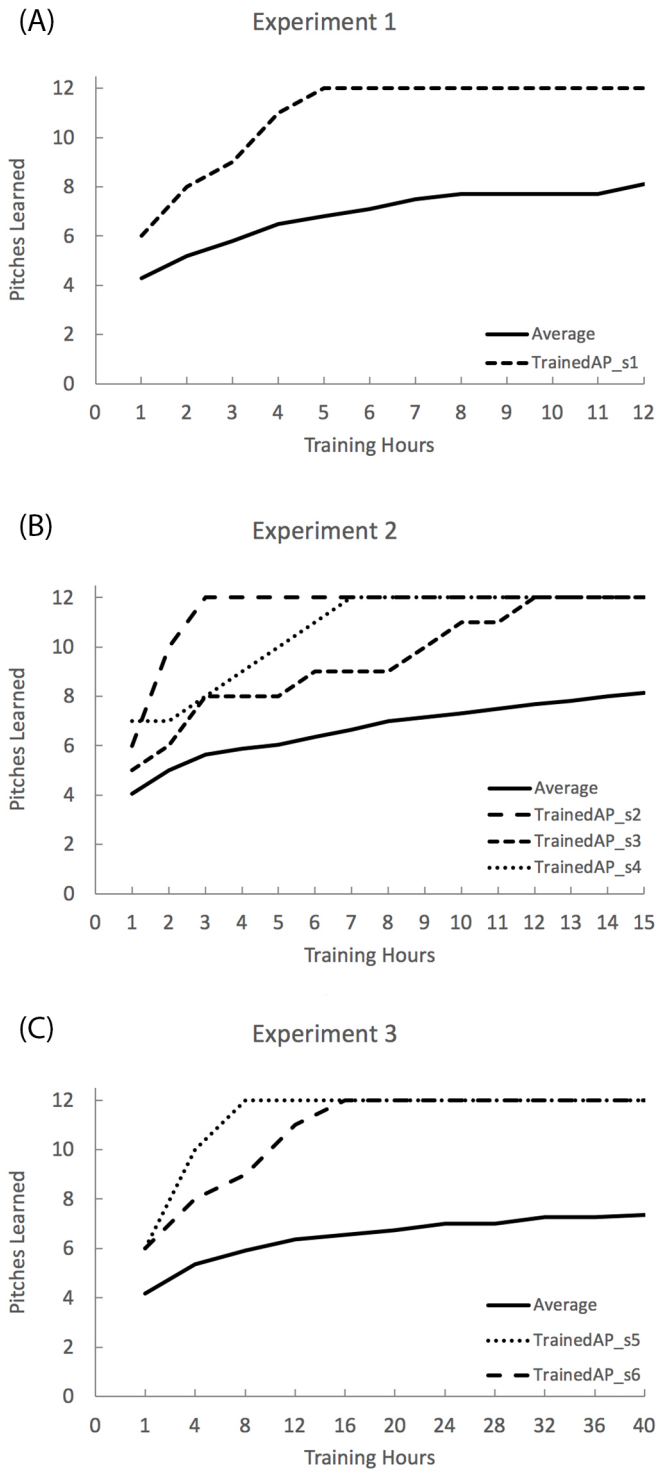

Figure 1. The number of pitches learned in the course of training in the three experiments. Number of pitches learned was defined as the number of pitches included at the highest passed levels at 90% accuracy with no feedback or scoring assistance in each individual. The solid line shows the average number of learned pitches across all individuals in each experiment. The other lines showed the course of training of the individuals who passed all levels of training in each experiment. TrainedAP_s6 in Experiment 3 learned to name 12 pitches at 90% accuracy with a subset of training tones (in one octave and one timbre only) at the end of the 16th hour of training, but she actually passed all levels of training during the 18th hour of training.

using accuracy or the average size of errors, etc.) and cut-off points. Given these variabilities in definition, we did not see any strong reasons to adopt any single definition of 'AP possessors' based on some particular publications. Instead of choosing one single definition, we decided to apply the definition specified in each of the 66 papers to our successfully trained participant to see if this participant would be considered an 'AP possessor' in these papers. We recalculated the participant's performance if needed. We observed that this participant would be considered an 'AP possessors' in 83.3% (55 out of 66) of these papers that adopted an objective AP performance-based definition. In other words, the level of AP performance achieved by this successfully trained participant was representative of and comparable to that of real-world 'AP possessors' defined in the literature.

Generalization & Sustainability of AP learning

The improved pitch-naming performance generalized to untrained octaves and timbres (Figure 2A). A 2 x 3 ANOVA with Prepost (pretest / posttest) and Stimulus Type (octave & timbre trained / octave trained & timbre untrained / octave untrained & timbre trained) as factors revealed a significant main effect of Prepost, $F(1,9) = 38.76$, $p < .001$, $\eta_p^2 = .812$, with a smaller pitch naming error at posttest than pretest. No other main effect or interaction was observed, ps > .19, i.e., we did not observe any difference between the naming performance of tones in trained or untrained timbres and octaves.

To check if the improvement sustained for a month, a one-way ANOVA was performed with Prepost (pretest / posttest / a month later) on pitch naming error with the trained tones[1]. It revealed a significant main effect of Prepost, $F(2,16) = 19.15$, $p < .001$, η_p^2

[1] One participant did not participate in the testing one month later and was excluded from this analysis.

= .705. Post-hoc LSD test showed that pitch naming error reduced after training, $p < .001$, and remained similar a month later, $p > .250$.

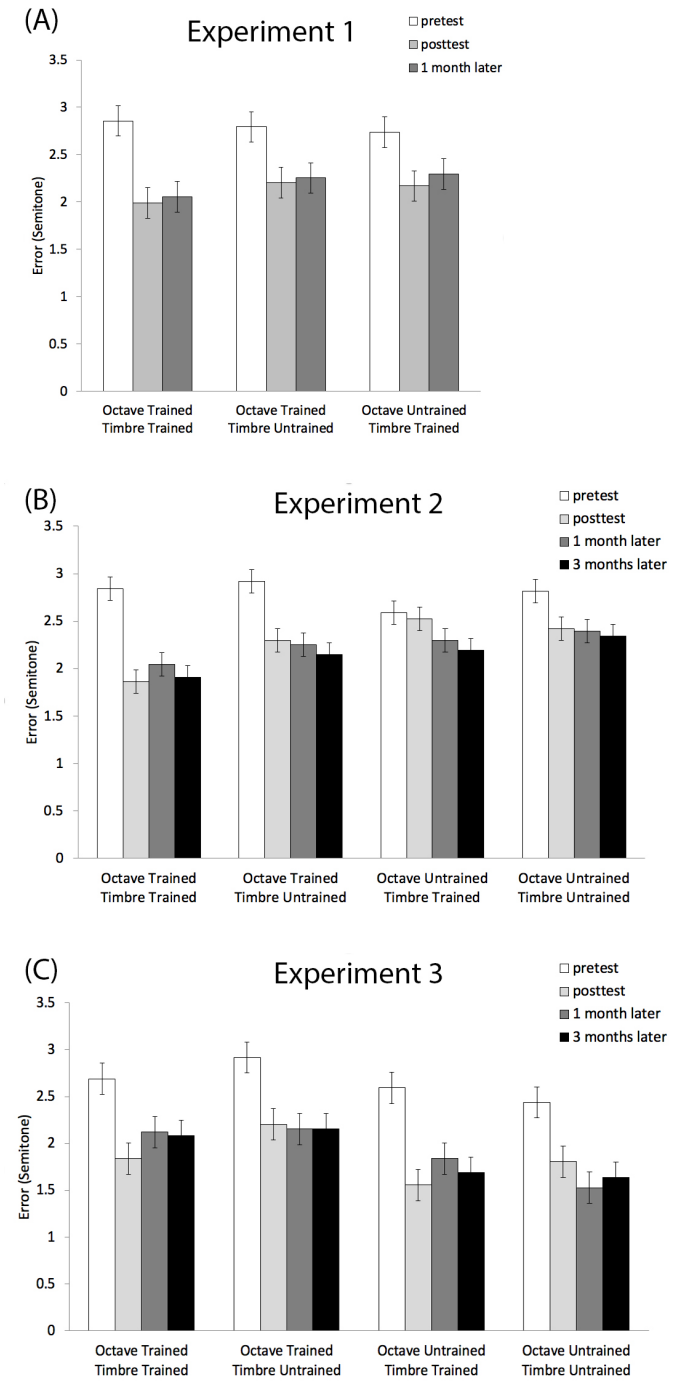

Figure 2. Pitch naming error in the pretest and posttests in the three experiments. Pitch naming error was defined as the number of semitone errors from the correct responses. Error bars represent 95% CIs for the interaction effect between Prepost and Stimulus Type in Experiment 1 (A), for the interaction effect between Prepost, Octave and Timbre in Experiments 2 and 3 (B-C).

Discussion

After the 12-hour AP training, all participants improved their pitch naming performance substantially. On average, they were able to name 8.1 pitches accurately. In particular, one of the participants was able to name all of the twelve pitches at 90% accuracy without externally provided reference tones. This level of AP performance was representative of and comparable to that of real-world 'AP possessors' based on a survey of the literature. This indicates that AP acquisition is possible in adulthood, and a 12-hour training protocol was sufficient for at least one of the participants to acquire AP.

In addition, the characteristics of AP learning matched well with that of perceptual learning (Fahle & Poggio, 2002; Goldstone, 1998). Specifically, AP performance improved after training, and the improved performance did not differ between tones in trained or untrained timbres and octaves, suggesting that the AP learning generalized to untrained octaves and timbres. Also, the AP performance was similar right after training and a month later, suggesting that the AP learning sustained for a month. Overall, the AP learning corresponded well with classic characteristics of perceptual learning in terms of performance enhancement, generalization and sustainability.

Experiment 2

In Experiment 2, we aimed to replicate the feasibility of acquiring AP in adulthood through perceptual learning and further characterize AP learning in adulthood. First, we tested the robustness of AP acquisition in adulthood by using a different set of training protocol, including a different set of training tones, training tasks, training duration and design. Second, we asked whether training with a smaller set of stimuli, i.e., tones in one

octave and one timbre only, would lead to higher specificity in AP learning, as one would expect based on the perceptual learning literature (Fahle & Poggio, 2002; Goldstone, 1998; Wong et al., 2011). Third, we also explored whether musicians benefit from the training more than non-musicians due to their prior musical training.

Methods

Participants

Twenty-two participants, including ten musicians and twelve non-musicians, were recruited at City University of Hong Kong and completed the training. The musicians included 4 males and 6 females and were 23.5 years old on average (*SD* = 3.34). They were trained in music for 6-14 years, with the major instrument being piano (*N* = 9) and guitar (*N* = 1). The non-musicians included 5 males and 7 females and were 21.8 years old on average (*SD* = 1.27). Two additional musicians dropped out of the training soon after participating because they could not commit to the whole training. The sample size per group was determined by matching that of Experiment 1. All participants filled out a similar questionnaire on musical training background similar to that of Experiment 1. Participants received monetary compensation for the training and testing, with additional bonuses for passing each level of training. Informed consent was obtained according to the Ethics Committee of City University of Hong Kong.

Materials

In Experiment 2, 72 tones in timbres of complex sine wave, piano and violin from octaves four to five were used. The tones were identical to that used in Experiment 1. An additional glissando clip, which was often perceived as an endless glissando travelling

continuously from a high pitch to a low pitch (Deutsch, 1995; Deutsch, Hamaoui, & Henthorn, 2007), was used to further interfere any existing auditory memory of tones before AP testing (see below).

Absolute pitch training

The training included 12 complex sine wave tones from octave 4. Similar to Experiment 1, the training was gamified and structured with different levels. If participants achieved 90% accuracy for a certain level, they would proceed to the next level; otherwise they would stay at the same level. The training was completed by finishing 15 hours of training or by passing all 30 levels with 90% accuracy. Participants finished one hour of training per day. They were trained four days per week and finished the training in four weeks.

The 30-level training protocol was organized into ten 3-level parts with an increasing number of pitches (from 3 pitches in the first 3 levels to 12 pitches in the last 3 levels). Within each 3-level part, the first level involved a verification task, in which participants judged whether the pitch of the presented tone matched with a pitch name shown on the screen with trial-by-trial feedback. The second level was a naming task, in which participants named the pitch of the tones with trial-by-trial feedback. The third level was a similar naming task without trial-by-trial feedback. If participants achieved 90% accuracy with a specific set of pitches at the third level, a new pitch was introduced into the training set, and they went through the three-level structure again. If participants failed to pass the third level in three attempts, they would be moved back to the second level. This allowed participants to re-learn the materials with the assistance of trial-by-trial feedback in case they were not ready for no-feedback training. The number of trials increased from 12 trials per level to 30 trials per level

gradually, with two trials added every time a new pitch was introduced. This allowed the blocks of trials to better represent and capture the increasing number of pitches included in the training set. At each level, the tones were distributed as evenly as possible among the training pitches, octaves and timbres. Semitone errors were considered errors in the training. Before each level, participants were allowed to freely listen to sample tones of the training pitches as many times as they preferred before proceeding to the training. Each training session lasted for an hour, in which individual participants might have finished different numbers of training trials depending on their pace of learning (e.g., the amount of time spent on the training trials or on sample tone listening). A similar system of three-point special trials and token accumulation applied as in Experiment 1.

Similar to Experiment 1, these no-feedback levels were designed to serve as mini milestones for participants' AP performance at 90% accuracy without external references. To further destroy any existing auditory memory trace of tones before testing, a 15s glissando clip was played every time after listening to the sample tones and before the start of the no-feedback levels. This should have effectively interfered with any auditory memory trace of tones since a 15-s latency with ineffective rehearsal (because of interfering tasks or stimuli) has been shown to result in very low accuracy during recall (10% or lower; Peterson & Peterson, 1959; see also the discussion in Takeuchi & Hulse, 1993; and Wengenroth et al., 2013). This further minimized the possibility that participants performed the no-feedback levels with the assistance of external reference tones.

Test for generalization

Similar to Experiment 1, the test for generalization was performed before, within three days after training, a month later and three months later to examine how well the pitch-

naming abilities generalized to untrained octaves and timbres. Seventy-two tones were used. The tones were either in octaves 4 to 5, in which octave 4 was trained and octave 5 was untrained. Also, the tones were either in three timbres, with complex sine wave as a trained timbre, and piano and violin as untrained timbres. The test was administered with four conditions, either with trained octave and timbre, trained octave and untrained timbre, untrained octave and trained timbre, or untrained octave and timbre. A similar pitch-naming task was used as in Experiment 1. There were 144 trials presented in randomized order, with each tone presented twice. Ten practice trials were provided with feedback before testing. The 15s glissando tone was presented after the practice with feedback before the testing to minimize any auditory working memory of tones that were previously heard (Peterson & Peterson, 1959; Takeuchi & Hulse, 1993; Wengenroth et al., 2013). Pitch-naming error, i.e., the average semitone error of participants' response in comparison with the correct response, was used as the dependent measure.

Results

Acquisition of AP

In general participants made substantial progress in learning to name pitches. At the end of training, the musicians and non-musicians were able to name on average 9.4 and 7.2 pitches (out of 12) at 90% accuracy respectively (Figure 1B). Importantly, three out of the twenty-two participants (13.6%) passed all levels of training, meaning that they were able to name all of the twelve pitches at 90% accuracy without externally provided reference tones, suggesting that they have acquired AP through perceptual learning in adulthood.

Generalization & Sustainability of AP learning

The improvement in pitch-naming performance for participants, was larger for the trained than untrained tones (Figure 2B). We did not observe significant differences in the degree of improvement between musicians and non-musicians. A 2 x 2 x 2 x 2 ANOVA with Group (musician / non-musician), Prepost (pretest / posttest), Octave (trained / untrained) and Timbre (trained / untrained) as factors on pitch naming error revealed a significant main effect of Prepost, $F(1,20) = 38.84$, $p < .001$, $\eta_p^2 = .660$, as the pitch naming error reduced after training. There were significant interactions between Prepost and Octave, $F(1,20) = 11.3$, $\eta_p^2 = .361$, and between Prepost, Octave and Timbre, $F(1,20) = 17.8$, $p < .001$, $\eta_p^2 = .471$. Post-hoc Scheffé test ($p < .05$) showed that the precision of pitch naming was higher for tones with the trained octave and timbre than other tones only after training. The interaction between Prepost and Group was not significant, $F(1,20) = 2.74$, $p = .113$, $\eta_p^2 = .121$, meaning that we did not observe any difference in the degree of AP learning between musicians ($M_{pretest} = 2.64$, $SD = .70$; $M_{posttest} = 1.98$, $SD = .80$) and non-musicians ($M_{pretest} = 2.91$, $SD = .47$; $M_{posttest} = 2.53$, $SD = .56$). The four-way interaction did not reach significance ($F < 1$).

To check if the improvement sustained for a month and for three months, a one-way ANOVA with Prepost (pretest / posttest / 1 month later / 3 months later) was performed on pitch naming error with the trained tones only[2]. It revealed a significant main effect of Prepost, $F(3,57) = 20.3$, $p < .001$, $\eta_p^2 = .517$. Post-hoc Scheffé tests ($p < .05$) showed that pitch naming error was smaller at posttest than pretest, and stayed similar one and three months later, suggesting that the AP improvement sustained for at least three months.

[2] One musician and one non-musician did not participate in the testing three months later and were excluded from this analysis.

Discussion

In Experiment 2, we replicated the feasibility of acquiring AP in adulthood and further characterized AP learning in adulthood. First, three adults successfully acquired AP using a different set of training protocols, suggesting that AP acquisition in adulthood is not a result of a certain specific type of training protocol used in Experiment 1, but can be generally observed with different perceptual learning paradigms.

Second, we replicated the perceptual learning nature of AP learning in adulthood in terms of performance enhancement and sustainable improvement for at least three months. The findings regarding generalization of learning was interesting. We observed that the AP learning was more specific (or less generalizable to untrained tones) when we used a smaller set of training tones in Experiment 2. It is in contrast to the findings in Experiment 1, in which using training tones in multiple octaves and timbres resulted in AP learning that was fully generalized to untrained tones. These findings fit well with the perceptual learning literature that the degree of learning specificity and generalization depends on the psychological multidimensional space of the training and testing stimuli (Nosofsky, 1986, 1987; Palmeri & Gauthier, 2004; Y. K. Wong et al., 2011).

Third, as an exploratory analysis we did not observe a higher degree of AP learning in musicians than non-musicians, though there was a numerical trend that musicians improved more than non-musicians at posttest compared with pretest. It is possible that AP training, which involves explicit naming of isolated tones, is sufficiently different from the daily musical training experience that musicians do not learn AP more efficiently than non-musicians. Alternatively, it is possible that the group difference might be more delicate and we did not have sufficient statistical power to reveal it. The type of prior music training may

also influence the efficiency of AP learning. Further studies may consider using larger group sizes with better specified music training experience to clarify this question.

Experiment 3

In Experiment 3, we further asked whether AP training in adulthood is feasible outside of the laboratory. Participants performed an online AP training anywhere with a stable Internet connection. They each finished 40 hours of training in their own pace within an eight-week period.

Methods

Participants

Eleven participants were recruited at City University of Hong Kong and completed the training. They included 4 males and 7 females and were 20.1 years old on average (*SD* = 0.83). Six of them have received musical training for 3 to 15 years, with the major instrument being piano (N = 3), violin (N = 2) and flute (N = 1). Five were non-musicians who were not formally trained with music or had brief training that lasted for less than a year. They performed the Distorted Tunes Test for tone-deaf screening at pretest (available at https://www.nidcd.nih.gov/tunestest/test-your-sense-pitch) so as to ensure that none of the participants would encounter difficulty in acquiring AP simply because of deficits in perceiving pitch in general. All participants performed the task satisfactorily with no suspected deficit with detecting distorted tunes (accuracy ranged from 73.1% to 100%, *M* = 88.2%, *SD* = 0.088). One additional participant was excluded from the training with highly accurate pitch naming precision during the pretest (with an average of 0.72 semitones from correct responses), leaving the training unnecessary. One additional musician was excluded

from data analyses because she showed highly inconsistent and uninterpretable training performance[3]. The sample size was determined to match that of Experiments 1 and 2. All participants filled out a similar questionnaire on musical training background similar to that of Experiment 1. Participants received monetary compensation for the training and testing, with additional bonuses for passing each level of training. Informed consent was obtained according to the Ethics Committee of City University of Hong Kong.

Materials

In Experiment 3, 120 tones were used, including that in timbres of complex sine wave and piano in octaves four to six, and that in timbres of violin and clarinet in octaves four to five. The complex sine wave, piano and violin tones were identical to that used in Experiments 1 and 2. The clarinet tones were downloaded from online sound library (http://newt.phys.unsw.edu.au/music/clarinet/index.html) and edited in an identical way as the other tones. The same glissando clip in Experiment 2 was also used in this experiment.

[3] The performance pattern of this particular participant was highly inconsistent and uninterpretable. First, most participants tended to progress and improve more slowly as the training proceeded. However, this particular participant showed an opposite pattern. In particular, the progress of this participant was very slow in the first half of the training when only a few pitches were included. However, her progress suddenly and dramatically improved in the second half of the study when most of the pitches were included in the training. Also, she showed a large standard deviation in the number of attempts required to pass the training levels (*SD* = 55.4), which was more than two times higher than that of the rest of the group, showing that her performance was highly unstable and deviant of the learning patterns of others. Since the training was performed in an uncontrolled environment (unlike the previous experiments that were performed in the laboratory), the highly deviant learning pattern could be caused by many uninteresting reasons that are irrelevant to AP learning. Therefore, we decided to exclude the data of this participant for further data analyses.

Absolute pitch training

This training was administered online such that participants could be trained anywhere with stable Internet connection. It included piano, violin and clarinet tones in octaves 4 and 5. A pitch-naming task was used. During each trial, an isolated tone was presented for 1s. Then, the 12 pitch names were presented. Participants were required to name the pitch of the presented tone by mouse clicking within 5s.

Similar to Experiments 1 and 2, the training was gamified and structured with different levels. If participants achieved 90% accuracy for a certain level, they would proceed to the next level; otherwise they would stay at the same level. The training was completed by finishing 40 hours of training or by passing all 80 levels with 90% accuracy. The training time of the participants was counted into the total training time in units of 15 minutes (i.e., 17 minutes training was taken as 15 minutes, and 14 minutes training as 0 minute). To minimize any idling time, the session for listening to sample tones before each level would be ended with 10s of inactivity. The program would automatically log out if there was inactivity for 3 minutes. They were trained for an average of five hours per week and finished the training in about eight weeks.

Similar to Experiment 1, the training protocol involved 80 levels and was organized into ten 8-level parts with an increasing number of pitches (from 3 pitches in the first 8 levels to 12 pitches in the last 8 levels). Each eight-level part consisted of four types of levels, which included tones that were progressively richer in timbres and octaves. Each of the four types of levels was repeated twice, once with feedback provided, and then once without feedback. For example, participants began the training with three pitches (E, F and F#). At levels 1-2, piano tones in these three pitches in octave 4 were included, with feedback provided at level 1 and then without feedback at level 2. At levels 3-4, piano tones in both

octave 4 and 5 were included with feedback and then without feedback. At levels 5-6, piano, violin and clarinet tones in octave 4 were included with feedback and then without feedback. At levels 7-8, piano, violin and clarinet tones in octave 4 and 5 were included with feedback and then without feedback. At the no-feedback levels, participants were not provided with any external feedback of the correctness of the tones, so they could not establish any external reference for the AP naming. Instead, they could only generate answers internally in an absolute manner. If they achieved 90% accuracy at the 8th level, a new pitch was added into the training set, with which they went through the same eight-level structure again. The number of trials increased from 12 trials per level to 30 trials per level gradually, with two trials added every time a new pitch was introduced. This allowed the training to better represent and capture the increasing number of pitches included in the training set. At each level, the tones were distributed as evenly as possible among the training pitches, octaves and timbres. Semitone errors were considered errors in the training. Similar to Experiment 2, participants were allowed to freely listen to sample tones of the training pitches as many times as they preferred before proceeding to each level of the training. The same glissando clip was played after listening to the sample tones and before the start of the no-feedback levels so as to destroy any existing auditory memory trace of previously heard tones and to minimize the possibility that participants performed the no-feedback levels with external reference tones. Also, the same system of tokens and three-point special trials were adopted in this study. Importantly, the special trials did not appear and could not be initiated during the no-feedback levels, similar to Experiments 1 and 2. This ensured that participants performed the no-feedback levels without any scoring assistance.

To better learning, a summary table of the participant's performance was provided at the end of each level. The table showed the accuracy for each pitch and listed out the wrong answers provided for each pitch (if any). This information enabled participants to better

evaluate the type of errors they made. Also, when the training set included 5 pitches or more, an additional compulsory exercise was introduced after every 10 attempts of passing the levels. The exercise would include 20 trials that centered around the worst-performed pitch in the past 10 blocks of trials. The timbre and octave of the tones would follow the level right before this exercise. This was designed to further help participants focus on the pitches that needed most training.

Test for generalization

The test for generalization was performed in the laboratory before and within three days after training to examine how well the pitch-naming abilities acquired online generalized to trained and untrained octaves and timbres in the laboratory setting. In this test, 72 tones in octaves 4 to 6 (C4 to B6) were used, in which octaves 4 and 5 were trained, and octave 6 untrained. Two timbres were included, with piano as the trained timbre, and complex sine wave as the untrained timbre. A similar pitch naming task as in Experiments 1 and 2 was used. During each trial, a tone was presented for 1s. Then an image that mapped the 12 pitch names to 12 keys of the keyboard appeared. Participants were required to name the pitch of the presented tone by keypress within 5s. The trials were blocked in four conditions, either with a trained or untrained octave crossed with a trained or untrained timbre. There were 192 trials in total, with 48 trials in each condition. The trials within each block were randomized. Similar to the test in Experiment 2, ten practice trials were provided with feedback before testing. A 15s glissando tone was presented after the practice but before the testing to destroy any existing auditory working memory of tones that were previously heard (Peterson & Peterson, 1959; Wengenroth et al., 2013). Pitch-naming error, i.e., the

average semitone error of participants' response in comparison with the correct response, was used as the dependent measure.

Results

Acquisition of AP

In general, the participants made substantial progress in learning to name pitches. At the end of training, they were able to name on average 7.4 pitches (out of 12) at 90% accuracy (Figure 1C). Importantly, two participants (out of eleven; 18.2%) passed all levels of training, meaning that they were able to name all of the twelve pitches at 90% accuracy without any externally provided reference tones. These suggest that they have acquired AP through perceptual learning in adulthood.

Generalization & Sustainability of AP learning

The online AP training improved pitch-naming performance in the laboratory, and the improvement generalized well to untrained octaves and less so to untrained timbres (Figure 2C). A 2 x 2 x 2 ANOVA with Prepost (pretest / posttest), Octave (trained / untrained) and Timbre (trained / untrained) as factors on pitch naming error revealed a significant main effect of Prepost, $F(1,10) = 15.1$, $p = .003$, $\eta_p^2 = .602$, as pitch naming error reduced after training. The interaction between Prepost and Timbre was significant, $F(1,10) = 8.55$, $p = .015$, $\eta_p^2 = .461$. Post-hoc Scheffé tests ($p < .05$) showed that the pitch naming error was similar for trained and untrained timbres before training. While the error of both types of timbres reduced at posttest, the error was smaller for trained timbres than untrained timbres after training. The interaction between Prepost and Octave did not reach significance ($F < 1$),

meaning that we did not observe any difference in the degree of AP learning for trained and untrained octaves. The three-way interaction did not reach significance (*ps* > .14).

For the sustainability of the improvement, a one-way ANOVA with Prepost (pretest / posttest / 1 month later / 3 months later) on pitch naming error with the trained tones revealed a main effect of Prepost, $F(3,27) = 9.538$, $p < .001$, $\eta_p^2 = .515$[4]. Post-hoc Scheffé tests ($p < .05$) showed that pitch naming error reduced after training, and stayed similar at all subsequent posttests, suggesting that the improvement sustained for at least three months.

Discussion

In Experiment 3, we further replicated the finding that AP can be acquired in adulthood. Using a 40-hour training on the Internet at anywhere at one's own pace, two adults were able to name tones in all twelve pitches at 90% accuracy without external reference tones, and thus have successfully acquired AP through perceptual learning in adulthood. This demonstrates that AP training is feasible outside of the laboratory.

The characteristics of AP learning matched with that of perceptual learning well. Using tones in multiple timbres and octaves, naming accuracy improved for both trained and untrained tones, suggesting that AP learning generalized to untrained tones similar to findings of Experiment 1. The improvement for trained timbres was larger than that of untrained timbres, while the improvement for trained and untrained octaves was similar. This indicated that the generalization of learning was more complete for octave than timbre. The AP improvement sustained for at least three months. Overall, the performance enhancement,

[4] One non-musician did not participate in the posttest one month later and was excluded from this analysis.

generalization and sustainability correspond well with that of the perceptual learning literature (Fahle & Poggio, 2002; Goldstone, 1998; Wong et al., 2011).

Overall Analyses of Experiments 1-3

Characterizing the Trained AP

In three experiments, we identified six out of 43 (14.0%) participants who acquired AP in adulthood after training. Below we further characterize their AP ability by asking three questions. To begin with, were these successfully trained participants simply 'AP possessors' before training? Two aspects of our data indicate that they were not 'AP possessors' before training. First, their pretest performance of AP naming for the training tones was low (M_{error} = 1.97 semitones; SD_{error} = 0.47), which would be rarely considered real-world 'AP possessors' in the literature, if at all, among the publications that defined 'AP possessors' based on AP performance measures. Second, they spent much longer time to pass all of the training levels than one would expect if they were already 'AP possessors' before training. Specifically, the AP training should be highly straightforward for 'AP possessors', and thus passing each of the levels should be easy with one or two attempts. If so, they should be able to pass the training well within the first hour of training. However, they required 7.83 hours on average to pass the training (SD = 5.23; range: 3 to 18; Figure 1), which was unreasonably long if they were 'AP possessors'. Therefore, the pretest performance and training progress of these successfully trained participants did not support the idea that there were 'AP possessors' before training.

In the literature, it is well known that 'AP possessors' suffer from performance decrement with unfamiliar timbres and octaves (Levitin & Rogers, 2005; Takeuchi & Hulse,

1993; Y. K. Wong & Wong, 2014). In our test for generalization, we included a large proportion of untrained tones with varied timbres and octaves, ranging from 66.7% to 75% of the trials, and the untrained tones were often intermixed with the trained tones. Was the trained AP ability also affected by the exposure to a large proportion of untrained tones? To address this question, we examined the AP performance of the trained tones during the posttest right after training. A one-way ANOVA with Prepost (pretest / posttest) on pitch naming error with the trained tones revealed a main effect of Prepost, $F(1,5) = 19.6$, $p = .0068$, $\eta_p^2 = .797$, indicating that the pitch naming error was smaller at posttest (M_{error} = 1.07 semitones; SD_{error} = 0.51) than pretest. While the average error met the defining criterion of the publications that used pitch naming error as the performance-based criterion for defining real-world 'AP possessors' (1 semitone error or below; Bermudez & Zatorre, 2009; Hutchins, Hutka, & Moreno, 2015; Loui, Zamm, & Schlaug, 2012), there was a wide range of pitch naming errors across individuals (range: .38 – 1.83 semitones). It suggests that the trained AP of some individuals were highly susceptible to the exposure to a large proportion of unfamiliar tones.

Finally, did the trained AP sustain for at least a month? A one-way ANOVA with Prepost (pretest / posttest / 1 month later)[5] on pitch naming error with the trained tones revealed a main effect of Prepost, $F(1,5) = 19.7$, $p < .001$, $\eta_p^2 = .798$. Post-hoc LSD test ($p < .05$) showed that the pitch naming error was smaller at posttest than pretest and remained similar a month later (M_{error} = 1.03 semitones; SD_{error} = 0.52). We did not observe any difference in AP performance between posttest and that a month later, suggesting that the trained AP sustained for at least a month.

[5] We did not include the posttest at three months later because this test was not included in Experiment 1.

Effectiveness of Different Training Paradigms

The three training paradigms used in the three experiments differed in many ways, including the training stimuli, duration, the venue of training, and the design of the training progression. While it is difficult to pinpoint the contributions of specific training parameters, we can at least view these training paradigms as different packages and ask one question: Was any of the training paradigms more effective than the others in general?

A one-way ANOVA with Experiment (1 / 2 / 3) as a between-subject factor was performed on the number of learned pitches at the end of the training, with the number of learned pitches defined by the number of pitches included at the highest passed level without feedback. The main effect of Experiment did not reach significance ($F < 1$), suggesting that the AP learning was not significantly different between training paradigms.

Exploratory analyses – Extrapolation of AP Learning

While our training studies had a designated duration (12-40 hours), it is well possible that participants may continue to improve with more training. For example, some perceptual learning studies lasted for more than a hundred hours and showed improvement throughout training (Watanabe, 2002). To explore this possibility, we estimated how long it would take participants to learn to accurately identify all of the twelve pitches. First, we extracted the number of learned pitches after each hour of training, defined by the number of pitches included at the highest passed level without feedback. In Experiment 3, since the training time per session was highly irregular, we collapsed the training time into 11 data points, including the 1st hour and every 4th hour afterwards, and extracted the number of learned pitches similarly. Then, the individual training progress was fitted with a power function. The

power function was selected because it captures the general characteristics of learning curves, e.g., one's learning rate may slow down during the learning process, and has been widely used to describe the effect of practice and learning, such as that of reaction time (Logan, 1992) and forgetting (Wixted & Carpenter, 2007). Finally, we extrapolated the power function to estimate the number of hours required for each individual to learn twelve pitches, based on the assumption that the learning function would stay similar in the subsequent training.

The extrapolation analyses indicated that 39.5% and 58.1% of the participants would be able to learn all of the twelve pitches if the training lasted for 60 and 180 hours respectively (Table 2; Supplementary Materials). There are a few interesting observations. First, about 44.0% of these extrapolated successful cases were non-musicians, suggesting that non-musicians could also reach the level similar to AP possessors if the AP training continues. Second, the extrapolation of learning was constrained by individual learning progress during training, and the extrapolation per se would not be sufficient to bring everyone to acquire all of the pitches. For example, 14.0% of the participants (6 out of 43) will not be able to learn all of the twelve pitches even with 3000 training hours, indicating that there were limitations in the training effect related to the individuals or the training protocol. Third, we tested whether the findings would still hold if we fit the training progress with another function. We used logarithmic function that can also capture the potential change of learning rate with time and observed qualitatively similar results (Supplementary Materials). These suggest that the extrapolation results were robust and the choice of the mathematical function was not critical. In sum, the extrapolation analyses suggest that with more time and commitment, the majority of the participants would be able to acquire AP.

Table 2. Number of participants who have successfully acquired AP in the three experiments by passing the training, or by extrapolating their individual training progress to 60 or 180 hours of training. The bracket shows the number of musicians / non-musicians for each item respectively.

Experiment	N	Training Passed	Training Passed By Extrapolation	
			≤60 hours	≤180 hours
1	10	1 (1/0)	6 (3/3)	6 (3/3)
2	22	3 (3/0)	9 (5/4)	17 (9/8)
3	11	2 (2/0)	2 (2/0)	2 (2/0)
Total	43	6 (6/0)	17 (10/7)	25 (14/11)
Percentage		14.0%	39.5%	58.1%

General Discussion

The observations that AP rarely occurs in the population and is difficult to learn in adulthood have led to the belief that AP is only possible for few selected individuals - those with particular genetic makeup and experience in early childhood. Our study provides the first empirical evidence that AP can be acquired by a considerable proportion of the population even during adulthood in both laboratory and online settings. Importantly, 14% of the participants successfully acquired AP with performance levels comparable to that of existing AP possessors within our short training period. The trained AP ability cannot be explained by pre-existing AP abilities. This proportion of training success is unprecedented in the literature of AP, suggesting that AP continues to be learnable in adulthood.

The pattern of AP learning is consistent with that of the perceptual learning literature. In particular, the learning effects sustained for at least one to three months and generalized to untrained tones to an extent depending on the specificity of the training set. When the training set included tones from more octaves and timbres (e.g., Experiments 1 & 3), AP

learning generalized better to untrained tones, whereas the learning was more specific to the trained tones when the training targeted on a smaller set of tones (e.g., Experiment 2). The specificity of learning as a function of the psychological multidimensional space during training and testing is consistent with findings in perceptual learning studies (Fahle & Poggio, 2002; Goldstone, 1998; Y. K. Wong et al., 2011). Similar to perceptual learning, AP can be trained using different training protocols, tasks, stimuli and durations (Experiments 1-3), and we did not observe any difference in the effectiveness of these training paradigms. These suggest that AP acquisition can be considered a type of perceptual learning.

Our exploratory extrapolation analyses suggest that 39.5% and 58% of the participants could learn to name all of the twelve pitches within 60 and 180 hours of AP training. Although the power function is widely used to model different types of learning and practice effects (Wixted, 2007; Logan, 1992), this extrapolation analysis was based on the assumption that one's learning curve, estimated using the power function, would stay largely similar in subsequent training. The validity of this assumption is an empirical question that can only be addressed by actually conducting the training. Given this caveat, these results do suggest the potential that the majority of the population may be able to acquire AP in adulthood with more time and commitment. Overall, AP acquisition may be regarded as a type of perceptual learning that is difficult but possible in adulthood.

These findings challenge the strong form of the genetic and critical period explanations of the genesis of AP. Given that 14% of the participants acquired AP within our training duration, and up to 58.1% could acquire AP with longer training, any relevant genetic dispositions that *enable* AP acquisition should be a lot more prevalent in the population than previously assumed (Baharloo et al., 1998; Drayna, 2007). Our results also suggest that there is not a critical period or rigid cut-off point after which AP acquisition would be impossible. In addition, the current findings are more consistent with the

continuous view (Bermudez et al., 2009; Levitin & Rogers, 2005) instead of the dichotomous view (Athos et al., 2007; Ward, 1999; Zatorre, 2003) of the distribution of AP, since participants acquired various degree of AP ability through the training.

The participants in the current study were native speakers of Cantonese, which is a tonal language. In tonal languages, words that are different only in tones (pitch heights or contour) can have entirely different semantic meanings. For example, in Cantonese, the word 'ma' means 'mother' in tone one, 'grandmother' in tone four, 'horse' in tone five, and 'to scold' in tone six. It has been proposed that tonal language speakers learn to associate words with tonal templates that are absolute, precise and stable during the critical period of language development, which could later facilitate the development of absolute pitch ability (Deutsch, Dooley, Henthorn, & Head, 2009; Deutsch, Henthorn, & Dolson, 2004; Deutsch, Henthorn, Marvin, & Xu, 2006). However, this hypothesis also assumes a speech-based critical period of AP acquisition for tonal language speakers (p.2399, Deutsch et al., 2009), and therefore it predicts that the acquisition of AP is impossible in adulthood even for tonal language speakers. Our findings challenge such speech-based critical period of AP acquisition. Future studies should compare AP acquisition in tonal and non-tonal language speakers with various factors (musical background and experience, social-economic status, intelligence, etc.) controlled to understand the contribution of tonal language background on AP development.

How are the trained AP compared with that in real-world 'AP possessors'? While we know that the trained AP and the real-world 'AP possessors' have comparable abilities in AP naming, we do not know much about whether they represent the AP information similarly in terms of underlying cognitive and neural mechanisms. For example, there could be different forms of pitch representations in the trained AP, such as representing each pitch in an absolute manner, or representing some of the pitches in an absolute manner while performing

AP naming for other pitches by relative comparison of the tones to the internally generated references. The latter form of trained AP would be similar to the 'Absolute A', which is considered a subtype of 'AP possessors' that only possess AP memory with the pitch 'A', but are able to perform accurate AP judgment by applying relative pitch judgment based on the internal memory of 'A' (Levitin & Rogers, 2005). Our AP measures could not differentiate between these different forms of AP representation. At the neural level, current evidence suggests that AP is supported by increased activity in the superior temporal gyrus (Ohnishi et al., 2001; Schulze, Gaab, & Schlaug, 2009; Wengenroth et al., 2013; Wilson, Lusher, Wan, Dudgeon, & Reutens, 2009) and the left dorsal lateral prefrontal cortex (Bermudez & Zatorre, 2005; Zatorre, Perry, Beckett, Westbury, & Evans, 1998). AP is also associated with neural activities within the first 300ms after hearing the tones (Itoh, Suwazono, Arao, Miyazaki, & Nakada, 2005; Pantev et al., 1998; Rogenmoser, Elmer, & Jäncke, 2015; Wengenroth et al., 2013; Wu, Kirk, Hamm, & Lim, 2008), and various structural and functional connectivity differences (Jäncke, Langer, & Hänggi, 2012; Keenan, Thangaraj, Halpern, & Schlaug, 2001; Loui, Li, Hohmann, & Schlaug, 2011; Loui et al., 2012; Oechslin, Meyer, & Jäncke, 2010; Schlaug, Jäncke, Huang, & Steinmetz, 1995; Wengenroth et al., 2013; Wilson et al., 2009). It is possible that the trained AP recruits similar neural mechanisms to the real-world 'AP possessors'. Alternatively, they may engage a different set of neural computations and/or structure because they learn this skill later in life in adulthood. Further studies may compare the neural mechanisms of trained AP and the naturally acquired AP in the real-world to address this question.

The role of experience in AP acquisition

The training effects of the current study are consistent with those observed in other perceptual learning studies in terms of performance enhancement, generalization, and sustained improvement (Fahle & Poggio, 2002; Goldstone, 1998; Y. K. Wong et al., 2011),

demonstrating that AP learning is a type of perceptual learning. We propose that the genesis of AP may be better understood from a learning or experience perspective.

The role of experience in shaping pitch naming performance has been repeatedly demonstrated in the literature. For example, better pitch naming is often observed in musicians with testing conditions associated with more prior experience, including better performance with the timbre of one's own instrument (Takeuchi & Hulse, 1993), with a highly used pitch like 'A4' as the tuning tone in orchestras (Levitin & Rogers, 2005; Takeuchi & Hulse, 1993), with a multisensory testing context similar to one's musical training (Y. K. Wong & Wong, 2014), and with the more frequently used white keys than black keys (Athos et al., 2007; Miyazaki, 1989, 1990; Takeuchi & Hulse, 1993). Experience also explains why non-musicians can encode the AP information for specific songs or melodies that are highly familiar (Levitin, 1994; Schellenberg & Trehub, 2003). Even performance in AP possessors can be disrupted with recent listening experience with detuned music (Hedger, Heald, & Nusbaum, 2013).

Experience can provide a parsimonious account of the differential manifestations of AP ability for different individuals. Various subtypes of AP have been introduced to explain AP ability that is well above chance but not comparable with that of AP possessors. For example, 'pseudo AP', 'quasi-AP', 'implicit AP', 'latent AP' and 'residual AP' have been used to refer to non-discarded AP information during development, unexpressed potential to develop AP limited by the critical period, incomplete forms of AP, etc. (Deutsch, 2013; Levitin & Rogers, 2005; Schellenberg & Trehub, 2003; Takeuchi & Hulse, 1993; Ward, 1999). Instead of being qualitatively different forms of AP, they may simply reflect different degrees and types of experience, and/or the attention and motivation in learning to process pitch information in an absolute manner. For instance, the ability to label the 'A4' tone only

(referred as ‘quasi-AP’ or the ‘absolute A’) can be best explained by their habitual use of A4 as the tuning tone (Levitin & Rogers, 2005).

Our results cannot conclude whether general musical training experience is sufficient to make AP training more efficient. All of the six individuals that successfully acquired AP were musically trained (five were pianists, and one was a violinist). However, there was merely a numerical trend that musicians improved more than non-musicians at posttest compared with pretest, and the difference did not reach significance (Experiment 2). While it is possible that we simply did not have sufficient statistical power to reveal the learning advantage in musicians, it is also possible that AP, the explicit naming of isolated tones, is sufficiently different from the daily musical training experience such that it does not lead to more efficient AP learning in musicians than non-musicians. And the type of music training experience may also be a factor in the success of AP acquisition. Therefore, it is difficult to interpret the role of general musical training in AP training based on the current data. Future studies may consider using larger group sizes with better specified music training experience to clarify this question.

Why is AP difficult to learn?

From the perceptual learning perspective, we should all have the potential to acquire AP regardless of age or prior musical training. While the extrapolation results do suggest that the majority of participants in the current study might acquire AP given enough time and commitment, for some participants learning was time consuming or even seemed impossible. While successful learning can be affected by many factors such as motivation and persistence (Vansteenkiste, Simons, Lens, Sheldon, & Deci, 2004), the difficulty to acquire AP is in clear contrast with the relative ease of other types of learning. For example, children can learn to

name familiar objects in new languages fairly easily (Gathercole & Baddeley, 1990), and adults can learn to name tens of novel objects with nonwords in several hours (A. C.-N. Wong, Palmeri, & Gauthier, 2009; Y. K. Wong et al., 2011). What makes naming the twelve pitches so difficult to learn?

A possible reason concerns the interference from speech. Speech contains rich and fine-grained tonal information, such as pitch levels and pitch contours, that conveys different meanings and emotions (Saffran & Griepentrog, 2001). With years of experience in perceiving speech, tones and speech words may have formed intricate mappings through experience that the same words could have formed associations with a range of tones or tonal patterns depending on meaning and context, which could be specific to each individual. It is possible that retrieving verbal labels of tones during a pitch-naming task activates the many-to-many mappings between tones and speech words, and that interferes with one's performance and learning. This proposal is consistent with the observation that musicians performed worse with tones presented with human voice than other timbres (Vanzella & Schellenberg, 2010). The individual difference in speech interference could drive the individual variability in pitch naming performance and AP acquisition. Future work can further explore this possibility in explaining pitch-naming performance in the general public.

Conclusion

Overall, the current study shows that AP continues to be learnable in adulthood. The results challenge the notion that AP is only possible for few individuals with particular genes and training within the critical period, in that AP acquisition does not seem to be a rigid developmental process limited to a specific time period in childhood. The extent to which one acquires AP may be better explained in terms of the amount and type of experience.

Acknowledgments

The authors declare no conflict of interest. We thank Gabriel Chan Pak Hong and Michael Lai Wei Chun for their help in data collection, Mandy Chu Yan Ting for the technical support, Helen Wong Hoi Shan for her help in violin tone production, and Patrick Bermudez for providing the complex sine wave tones.

Authors Contributions

Y. Wong and A. Wong developed the study concept and designed the study. Y. Wong collected the data. Y. Wong, K. Lui and K. Yip analyzed the data. Y. Wong and A. Wong drafted the manuscript. All authors approved the final version of the manuscript for submission.

References

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

Baharloo, S., Johnston, P. a, Service, S. K., Gitschier, J., & Freimer, N. B. (1998). Absolute pitch: an approach for identification of genetic and nongenetic components. *American Journal of Human Genetics*, *62*(2), 224–231. https://doi.org/10.1086/301704

Bermudez, P., Lerch, J. P., Evans, A. C., & Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cerebral Cortex*, *19*, 1583–1596.

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

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

Brady, P. T. (1970). Fixed-scale mechanism of absolute pitch. *The Journal of the Acoustical Society of America*, *48*, 883–887. https://doi.org/10.1121/1.1912227

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Chin, C. S. (2003). The Development of Absolute Pitch: A Theory Concerning the Roles of Music Training at an Early Developmental Age and Individual Cognitive Style. *Psychology of Music*, *31*(2), 155–171. https://doi.org/10.1177/0305735603031002292

Chung, S. T. L., & Truong, S. R. (2013). Learning to identify crowded letters: Does the learning depend on the frequency of training? *Vision Research*, *77*, 41–50. https://doi.org/10.1016/j.visres.2012.11.009

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

Cuddy, L. L. (1970). Training the absolute identification of pitch. *Perception & Psychophysics*, *8*(5A), 265–269.

Deutsch, D. (1995). Musical illusions and paradoxes. La Jolla, CA: Philomel Records. (compact disc).

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

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

Deutsch, D., Dooley, K., Henthorn, T., & Head, B. (2009). Absolute pitch among students in

an American music conservatory: Association with tone language fluency. *The Journal of the Acoustical Society of America*, *125*(4), 2398–2403. https://doi.org/10.1121/1.3081389

Deutsch, D., Hamaoui, K., & Henthorn, T. (2007). The glissando illusion and handedness. *Neuropsychologia*, *45*, 2981–2988. https://doi.org/10.1016/j.neuropsychologia.2007.05.015

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute Pitch, Speech, and Tone Language: Some Experiments and a Proposed Framework. *Music Perception*, *21*(3), 339–356. https://doi.org/10.1525/mp.2004.21.3.339

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. https://doi.org/10.1121/1.2151799

Drayna, D. T. (2007). Absolute pitch: a special group of ears. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14549–14550. https://doi.org/10.1073/pnas.0707287104

Fahle, M., & Poggio, T. (2002). *Perceptual Learning*. Massachusetts: MIT Press.

Fujioka, T., Ross, B., Kakigi, R., Pantev, C., & Trainor, L. J. (2006). One year of musical training affects development of auditory cortical-evoked fields in young children. *Brain*, *129*, 2593–2608. https://doi.org/10.1093/brain/awl247

Gathercole, S. E., & Baddeley, D. A. (1990). Phonological memory deficits in language-disordered children: Is there a causal connection? *Journal of Memory and Language*, *29*, 336–360.

Goldstone, R. (1998). Perceptual learning. *Annual Review of Psychology*, *49*, 585–612.

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

Hutchins, S., Hutka, S., & Moreno, S. (2015). Symbolic and Motor Contributions to Vocal Imitation in Absolute Pitch. *Music Perception: An Interdisciplinary Journal*, *32*(3), 254–265. https://doi.org/10.1525/mp.2015.32.3.254

Itoh, K., Suwazono, S., Arao, H., Miyazaki, K., & Nakada, T. (2005). Electrophysiological correlates of absolute pitch and relative pitch. *Cerebral Cortex (New York, N.Y. : 1991)*, *15*(6), 760–769. https://doi.org/10.1093/cercor/bhh177

Jäncke, L., Langer, N., & Hänggi, J. (2012). Diminished Whole-brain but Enhanced Peri-sylvian Connectivity in Absolute Pitch Musicians. *Journal of Cognitive Neuroscience*, *24*(6), 1447–1461.

Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *NeuroImage*, *14*(6), 1402–1408. https://doi.org/10.1006/nimg.2001.0925

Kraus, N., & Banai, K. (2007). Auditory-Processing Malleability Focus on Language and

Music, *16*(2), 105–110.

Levitin, D. J. (1994). Absolute memory for musical pitch: evidence from the production of learned melodies. *Perception & Psychophysics*, *56*, 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

Logan, G. D. (1992). Shapes of Reaction-Time Distributions and Shapes of Learning Curves: A Test of the Instance Theory of Automaticity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(5), 883–914. https://doi.org/10.1037/0278-7393.18.5.883

Loui, P., Li, H. C., Hohmann, A., & Schlaug, G. (2011). Enhanced cortical connectivity in absolute pitch musicians: A model for local hyperconnectivity. *Journal of Cognitive Neuroscience*, *23*(4), 1015–1026. https://doi.org/10.1162/jocn.2010.21500

Loui, P., Zamm, A., & Schlaug, G. (2012). Enhanced functional networks in absolute pitch. *NeuroImage*, *63*(2), 632–640. https://doi.org/10.1016/j.neuroimage.2012.07.030

Meyer, M. (1899). Is the Memory of Absolute Pitch Capable of Development by Training? *Psychological Review*. https://doi.org/10.1037/h0069034

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

Miyazaki, K. (1990). The Speed of Musical Pitch Identification Possessors by Absolute-Pitch Possessors. *Music Perception*, *8*(2), 177–188. https://doi.org/10.2307/40285495

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen*, *115*(1), 39–61. Retrieved from http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=2937873

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *J Exp Psychol Learn Mem Cogn*, *13*(1), 87–108. Retrieved from http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=2949055

Oechslin, M. S., Meyer, M., & Jäncke, L. (2010). Absolute pitch--functional evidence of speech-relevant auditory acuity. *Cerebral Cortex (New York, N.Y. : 1991)*, *20*(2), 447–455. https://doi.org/10.1093/cercor/bhp113

Ohnishi, T., Matsuda, H., Asada, T., Aruga, M., Hirakata, M., Nishikawa, M., … Imabayashi, E. (2001). Functional anatomy of musical perception in musicians. *Cereb Cortex*, *11*(8), 754–760. Retrieved from http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=11459765

Palmeri, T., & Gauthier, I. (2004). Visual object understanding. *Nat Rev Neurosci*, *5*(4), 291–303. https://doi.org/10.1038/nrn1364

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998).

Increased auditory cortical representation in musicians. *Nature*, *392*(6678), 811–814. https://doi.org/10.1038/33918

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

Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual items. *Journal of Experimental Psychology*, *61*, 12–21.

Rogenmoser, L., Elmer, S., & Jäncke, L. (2015). Absolute pitch: Evidence for early cognitive facilitation during passive listening as revealed by reduced P3a amplitudes. *Journal of Cognitive Neuroscience*, *27*(3), 623–637. https://doi.org/10.1162/jocn_a_00708

Ross, D. a., Olson, I. R., Marks, L. E., & Gore, J. C. (2004). A nonmusical paradigm for identifying absolute pitch possessors. *The Journal of the Acoustical Society of America*, *116*(3), 1793. https://doi.org/10.1121/1.1758973

Saffran, J. R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: evidence for developmental reorganization. *Developmental Psychology*, *37*(1), 74–85. https://doi.org/10.1037/0012-1649.37.1.74

Sasaki, Y., Nanez, J. E., & Watanabe, T. (2010). Advances in visual perceptual learning and plasticity. *Nature Reviews. Neuroscience*, *11*, 53–60. https://doi.org/10.1038/nrn2737

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science*, *14*(3), 262–266. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12741751

Schlaug, G., Jäncke, L., Huang, Y., & Steinmetz, H. (1995). In vivo evidence of structural brain asymmetry in musicians. *Science*, *267*(5198), 699–701.

Schulze, K., Gaab, N., & Schlaug, G. (2009). Perceiving pitch absolutely: comparing absolute and relative pitch possessors in a pitch memory task. *BMC Neuroscience*, *10*, 106. https://doi.org/10.1186/1471-2202-10-106

Seitz, A. R., & Watanabe, T. (2009). The phenomenon of task-irrelevant perceptual learning. *Vision Research*, *49*(21), 2604–2610. https://doi.org/10.1016/j.visres.2009.08.003

Takeuchi, a H., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin*, *113*(2), 345–361. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8451339

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

Vansteenkiste, M., Simons, J., Lens, W., Sheldon, K. M., & Deci, E. L. (2004). Motivating learning, performance, and persistence: the synergistic effects of intrinsic goal contents and autonomy-supportive contexts. *Journal of Personality and Social Psychology*, *87*(2), 246–260. https://doi.org/10.1037/0022-3514.87.2.246

Vanzella, P., & Schellenberg, E. G. (2010). Absolute pitch: effects of timbre on note-naming ability. *PloS One*, *5*(11), e15449. https://doi.org/10.1371/journal.pone.0015449

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

298). Academic Press.

Wengenroth, M., Blatow, M., Heinecke, A., Reinhardt, J., Stippich, C., Hofmann, E., & Schneider, P. (2013). Increased Volume and Function of Right Auditory Cortex as a Marker for Absolute Pitch. *Cerebral Cortex*, *24*(May), 1127–1137. https://doi.org/10.1093/cercor/bhs391

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 (New York, N.Y. : 1991)*, *19*(3), 724–732. https://doi.org/10.1093/cercor/bhn121

Wixted, J. T., & Carpenter, S. K. (2007). The Wickelgren power law and the Ebbinghaus savings function. *Psychological Science*. https://doi.org/10.1111/j.1467-9280.2007.01862.x

Wong, A. C.-N., Palmeri, T., & Gauthier, I. (2009). Conditions for face-like expertise with objects: Becoming a Ziggerin expert - but which type? *Psychological Science*, *20*(9), 1108–1117.

Wong, A. C., Palmeri, T. J., & Gauthier, I. (2009). Conditions for Facelike Expertise With Objects Becoming a Ziggerin Expert — but Which Type ? *Psychological Science*, *20*(9), 1108–1117.

Wong, Y. K., Folstein, J. R., & Gauthier, I. (2011). Task-irrelevant perceptual expertise. *Journal of Vision*. https://doi.org/10.1167/11.14.3

Wong, Y. K., & Wong, A. C.-N. (2014). Absolute pitch memory: Its prevalence among musicians and dependence on the testing context. *Psychonomic Bulletin & Review*, *21*, 534–542. https://doi.org/10.3758/s13423-013-0487-z

Wu, C., Kirk, I. J., Hamm, J. P., & Lim, V. K. (2008). The neural networks involved in pitch labeling of absolute pitch musicians. *Neuroreport*, *19*(8), 851–854.

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 of the United States of America*, *95*(6), 3172–3177.

Supplementary Materials

Extrapolation of training progress with power functions

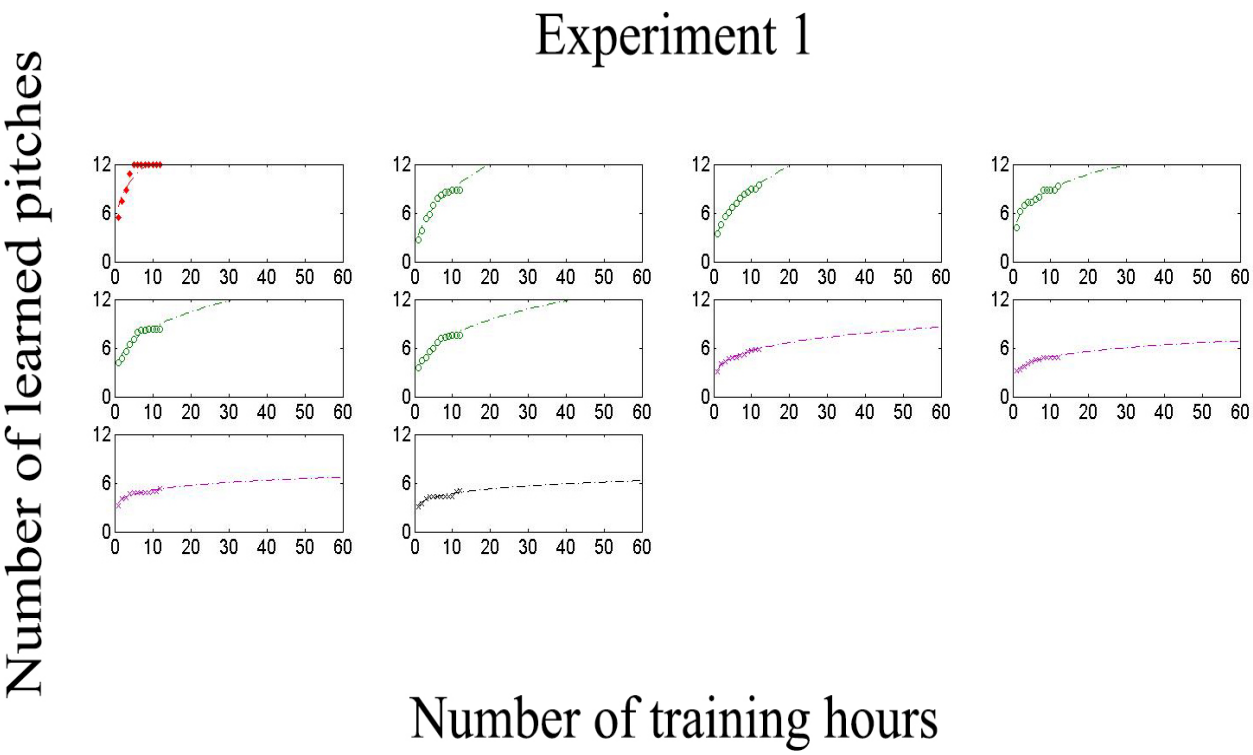

Figure S1. Individual curve fitting results of the number of learned pitches and training duration using a power function in Experiment 1. The markers indicate the original data points and the dashed line indicates the fitted function. Red markers and lines indicate participants who passed the training. Green markers and lines indicate participants with extrapolated passing of the training within 60 hours of training. Purple markers and lines indicate participants with extrapolated passing of the training within 3000 hours of training. Black markers and lines indicate participants who would not pass the training within 3000 hours of training in the extrapolation.

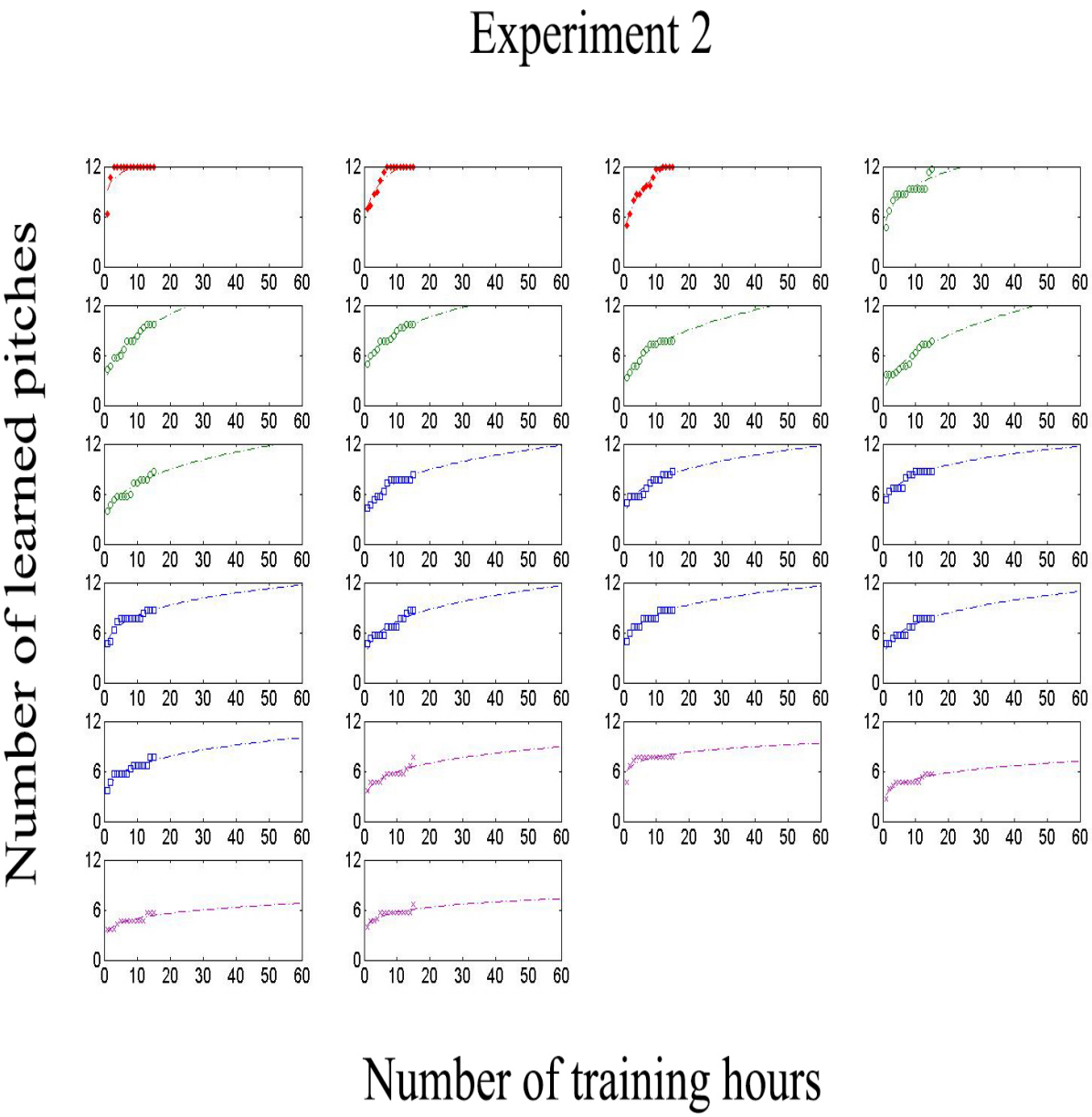

Figure S2. Individual curve fitting results of the number of learned pitches and training duration using a power function in Experiment 2. The markers indicate the original data points and the dashed line indicates the fitted function. Red markers and lines indicate participants who passed the training. Green markers and lines indicate participants with extrapolated passing of the training within 60 hours of training. Blue markers and lines indicate participants with extrapolated passing of the training within 180 hours of training. Purple markers and lines indicate participants with extrapolated passing of the training within 3000 hours of training.

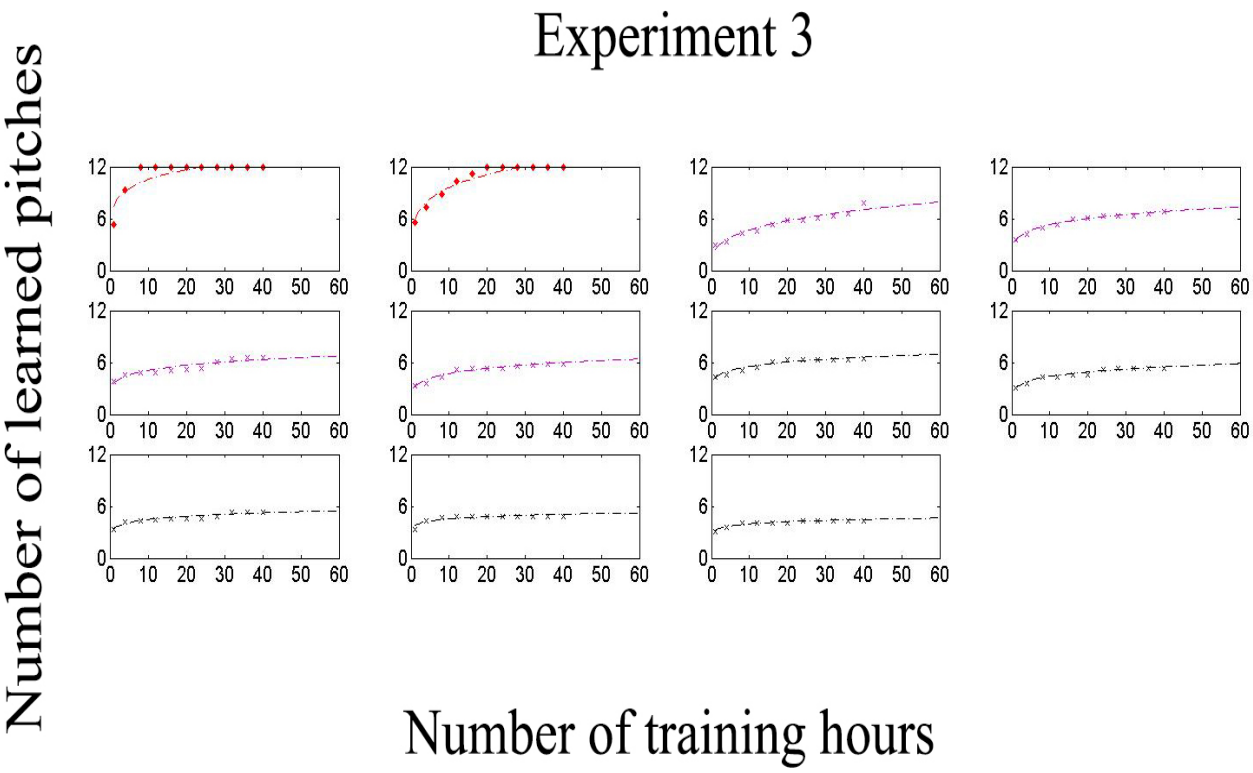

Figure S3. Individual curve fitting results of the number of learned pitches and training duration using a power function in Experiment 3. The markers indicate the original data points and the dashed line indicates the fitted function. Red markers and lines indicate participants who passed the training. Purple markers and lines indicate participants with extrapolated passing of the training within 3000 hours of training. Black markers and lines indicate participants who would not pass the training within 3000 hours of training in the extrapolation.

Extrapolation of training progress with logarithmic functions

		Training Passed By Extrapolation	
Experiment	N	≤60 hours	≤180 hours
1	10	4 (2/2)	6 (3/3)
2	22	6 (5/1)	12 (7/5)
3	11	2 (2/0)	2 (2/0)
Total	43	12 (9/3)	20 (12/8)
Percentage		27.9%	46.5%

Table S1. Number of participants who could pass the training if the training lasted for 60 or 180 hours. The estimation was performed by extrapolating the individual training progress using a logarithmic function. The bracket shows the number of musicians / non-musicians for each item respectively.

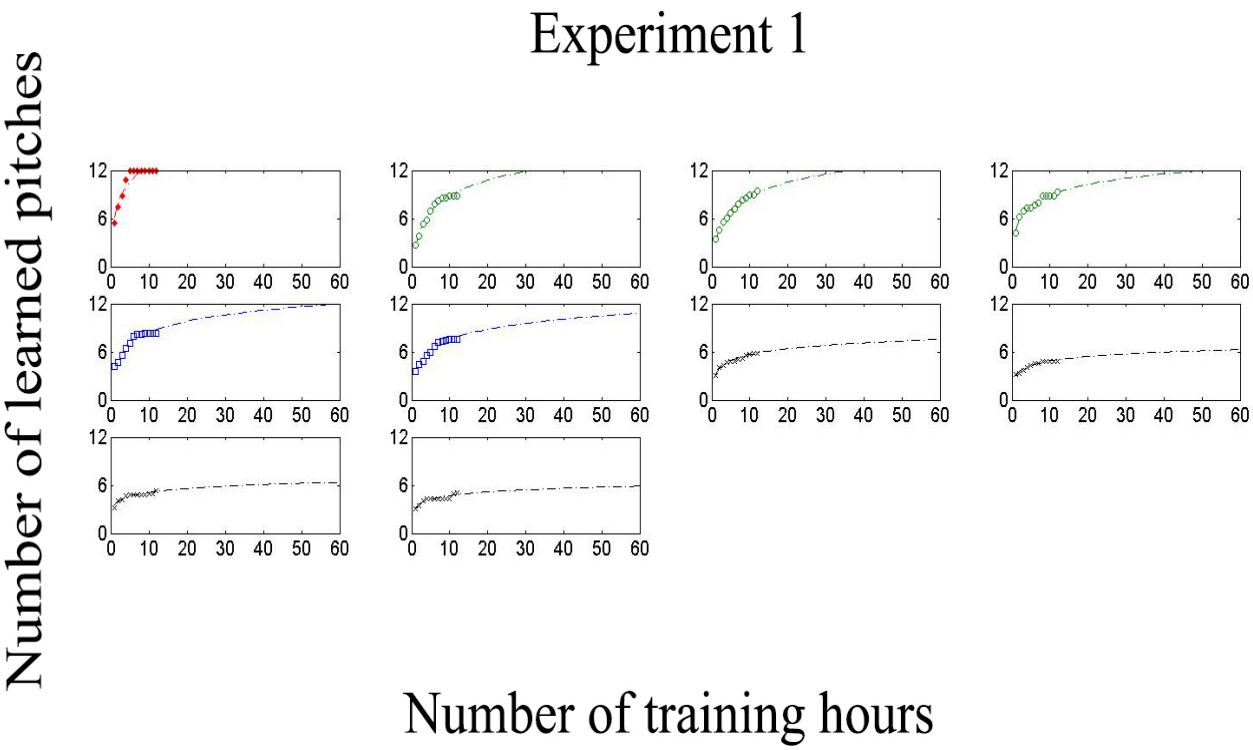

Figure S4. Individual curve fitting results of the number of learned pitches and training duration using a logarithmic function in Experiment 1. The markers indicate the original data points and the dashed line indicates the fitted function. Red markers and lines indicate participants who passed the training. Green markers and lines indicate participants with extrapolated passing of the training within 60 hours of training. Purple markers and lines indicate participants with extrapolated passing of the training within 3000 hours of training. Black markers and lines indicate participants who would not pass the training within 3000 hours of training in the extrapolation.

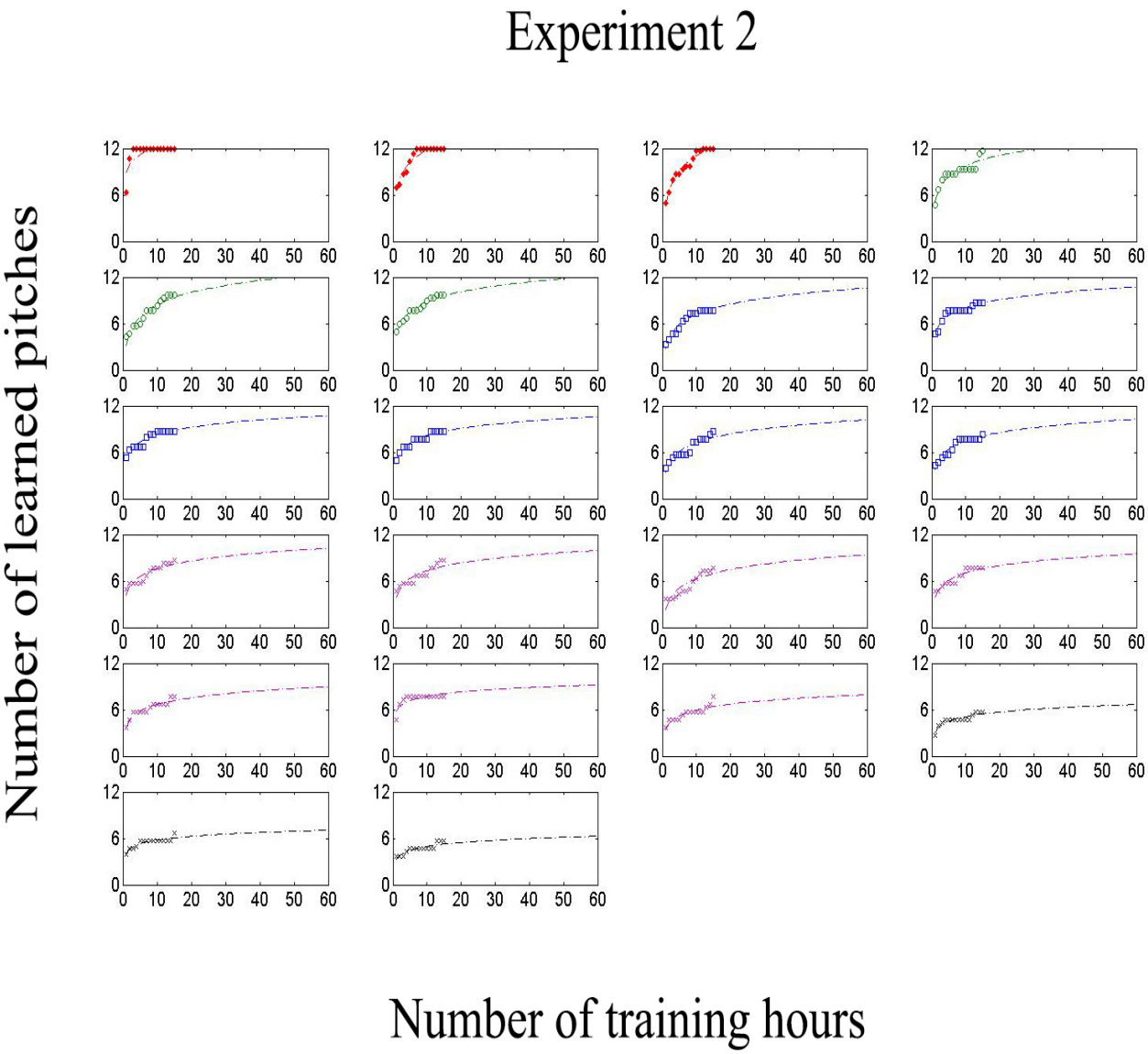

Figure S5. Individual curve fitting results of the number of learned pitches and training duration using a logarithmic function in Experiment 2. The markers indicate the original data points and the dashed line indicates the fitted function. Red markers and lines indicate participants who passed the training. Green markers and lines indicate participants with extrapolated passing of the training within 60 hours of training. Blue markers and lines indicate participants with extrapolated passing of the training within 180 hours of training. Purple markers and lines indicate participants with extrapolated passing of the training within 3000 hours of training. Black markers and lines indicate participants who would not pass the training within 3000 hours of training in the extrapolation.

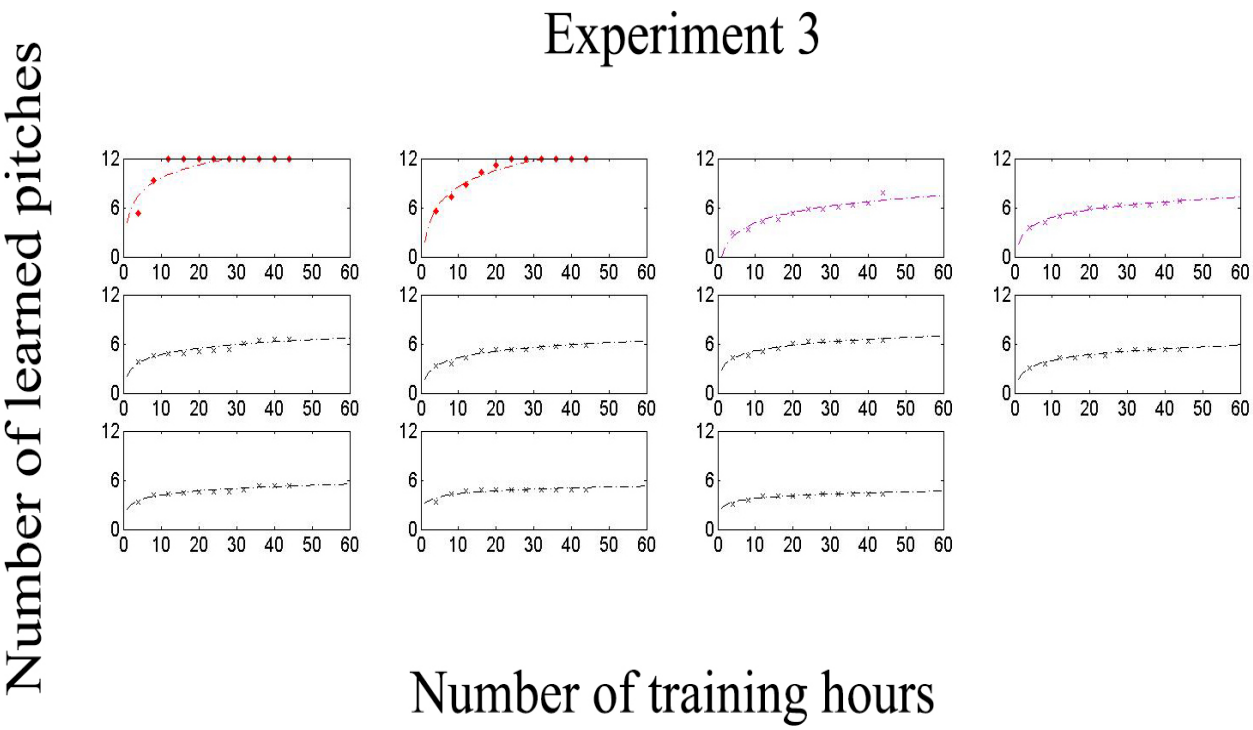

Figure S6. Individual curve fitting results of the number of learned pitches and training duration using a logarithmic function in Experiment 3. The markers indicate the original data points and the dashed line indicates the fitted function. Red markers and lines indicate participants who passed the training. Purple markers and lines indicate participants with extrapolated passing of the training within 3000 hours of training. Black markers and lines indicate participants who would not pass the training within 3000 hours of training in the extrapolation.