Absolute Memory of Learned Melodies

Children Trained by the Suzuki Violin Method

Victoria Saah, 2003

Advised by Elizabeth Marvin, Ph.D. and Jennifer Williams Brown, Ph.D.

RC Department of Music; Eastman School of Music

The field of music cognition examines many interactions the brain has with music, from interpreting to producing to memorizing music. Within this field, a subset of studies looks at the acquisition and manifestation of an ability called absolute pitch. Absolute pitch (AP) is defined as “the ability to identify the frequency or musical name of a specific tone, or, conversely, the ability to produce some designated frequency, frequency level, or musical pitch without comparing the note with any objective reference tone (i.e., without using relative pitch [RP])”.9 In other words, an AP possessor can either name a given pitch (called pitch labeling) or produce a named pitch without comparing it to any other pitch (called pitch memory).9 Relative pitch (RP) possessors can also label and produce pitches, but only if they are given a named pitch with which to compare, called a reference tone. However, many adult RP listeners exhibit pitch memory without this reference tone, but only in the context of a learned melody. For example, many adults can begin a familiar song on the correct pitch without a reference tone.9 This study examines pitch memory for songs in RP listeners who are young, and who are near the age at which AP is hypothesized to develop.

Current theories explaining the acquisition of AP disagree over whether AP is inherited or learned. Much empirical evidence supports the theory that AP is learned,19 and some researchers suggest that the learning takes place during a critical period in childhood. As early as 1916, Copp introduced this theory and claimed that 80% of children could learn to produce and recognize middle C on command, but only if they began learning when they were young. Sergeant found a strong negative correlation between the age of beginning music lessons and the incidence of AP later.14 Miyazaki also suggests an early training period for acquisition of AP as compared to RP, pointing out that the development of AP at a young age may impede the subsequent development of RP. Finally, Takeuchi and Hulse proposed a theory that identifies a period of learning in childhood, during which the acquisition of AP is possible.15 They found that, compared to non-AP listeners, AP possessors were much slower and less accurate at identifying black-key pitches than white-key pitches. Since the white-key pitches are typically learned first in childhood, and since songs learned in early music lessons are usually in keys with mostly white notes (for simplicity), the researchers hypothesized that the children acquire AP before they have integrated the black keys into their vocabulary. This hypothesis suggests that a window of time exists until approximately age 6 or 7 during which acquiring AP for familiar notes is possible, and after which acquiring AP for newly learned notes is more difficult and less stable.

There is also evidence that even among RP possessors at adulthood, one aspect of AP may be present: the ability to produce a learned melody at a consistent pitch level, without a reference tone. Several studies demonstrate that adult RP possessors are able to access a stable pitch memory similar to that of AP listeners when singing familiar songs. For example, Halpern found that when adults with RP were asked to sing familiar songs multiple times in multiple sessions, each adult started on almost the same pitch within each song (with a mean standard deviation of 0.28 semitones).9 Bergeson and Trehub found similar results in a study of mothers who sang to their infants: the mothers began each song within a mean of 0.82 semitones of where they started the previous time.4 Clearly, there exists a pitch memory in RP possessors that allows each adult to reproduce a pitch within a melodic context consistently over time.

There is also evidence that this pitch memory is stable and accurate between (as opposed to within) adults as well. Levitin asked forty-six undergraduate and graduate psychology students, both with and without musical backgrounds, to hear two favorite popular songs (chosen by each participant from a selection of CDs) in their minds during two trials and then to sing them into a microphone.9 Levitin hypothesized that these songs would have been heard in the same key every time the listener heard them, and that “repeated exposure to a song creates a memory representation that preserves the actual pitches of the song.”9 This preservation of pitch and key across all exposures to the song is what he called an objective standard. This standard allowed Levitin to identify a pitch memory
(an “absolute” memory because the correct pitch was often preserved) across the adult population. He found that 40% of these participants sang the correct pitch on at least one of two trials. Furthermore, 81% of the participants began the song within two semitones of the correct pitch on at least one of two trials. These results suggest that the general adult population, which includes primarily RP listeners, can produce pitches from a pitch memory that is relatively accurate and stable over time.

It is not known, however, whether this type of memory for pitch exists in children. One recent study showed that infants tended to track absolute pitch patterns (and are disinterested in transposed pitch patterns) in nonmusical contexts, but little focus has been directed toward pitch memory in children. The purpose of our study was to investigate whether a stable pitch memory might exist in children who are repeatedly exposed to an objective standard. Since children hear and sing many folk tunes and nursery rhymes in many different keys, the task of locating an objective standard is more difficult than it appears. Thus, to find a group of songs consistently in the same keys, we chose to study the specific population of children learning violin under the Suzuki method. This method of teaching an instrument emphasizes the repeated listening to and practice of beginner-level songs, as well as the singing of these songs to auralize (“hear” in their minds) the melodies. As a result, Suzuki students are repeatedly exposed to songs that are consistently in the same keys. We hypothesized that, if they were asked to reproduce these learned songs, they would be able to access a pitch memory (similar to AP possessors and the general adult population) and sing the songs in or close to the correct keys.

In two sessions, participants were asked to sing the same six songs from Suzuki’s Violin Book Volume I, which began on varied pitches: C#5, A4, D4, F#4, D5, G4. We recorded the sung performances to measure participants’ accuracy against the objective standard. The design of this study was influenced strongly by that of previous researchers. Like Levitin, we asked the students first to hear the songs in their heads before singing them. Finally, we used distractor tones between each song and in the AP post-test so that no RP could be used between trials.

Within this population of young Suzuki students, we also looked for an effect of age on the accuracy of pitch production. To test the early-learning theory of AP acquisition, we solicited two groups of children who began learning violin using the Suzuki method at different ages: one began at 3-5 years, the other 6-8 years. At the time we solicited participants, we did not know whether or not the children had AP. We hypothesized that if there were a period in early childhood during which children acquired AP, the group who began at a younger age might contain some AP listeners and would perform better. As for the question of whether any of these children were developmentally ready to sing in a stable key, Davidson, McKernon, and Gardner showed that by the age of 5 years children could begin a song and maintain its key all the way through. To control for differences in vocal development between the age groups, we gave each participant a pitch-matching test to ensure that both young and old students could physically hear, internalize, and repeat a given pitch pattern accurately.

Finally, in addition to our pitch memory hypothesis and testing for the effect of age, we examined the effect of viewing the musical score on their performance. We hypothesized that if the children saw the score before singing the song, that the visual stimulus might trigger the hearing of the starting pitch in their minds and thus improve their accuracy. Previous studies involving the score’s effect on auralization (this inner hearing of the pitch) in adults were done by Terhardt and Seewann in 1983, showing that non-AP participants could discriminate between correct and transposed excerpts when the score was visible (though the study did not compare performances without the score). Ward, on the other hand, had two conditions: one in which participants viewed the score while listening and one where they listened without any score visible. He investigated whether having the notation visible (i.e., having the opportunity to auralize the pitches from the score) helped participants recognize these familiar pieces when they were transposed. His results revealed that listeners could identify small transpositions only with the score in front of them and not without it. Larger transpositions could be identified with or without the score. Thus, the score is not required for auralization, but it helps. In this same vein, the children learning the Suzuki songs may not need the score in order to begin on the correct pitch, but auralizing the starting pitch from the score might improve their accuracy.

In summary, we tested one primary prediction and two secondary predictions: our primary prediction was that the majority of these children will begin learned songs within a semitone of the correct pitch. Secondarily, the children who began musical training at a younger age will achieve higher accuracy. Finally, also secondary to the first hypothesis, we predicted that on the trials where the score is visible, the children will achieve higher accuracy.

The participants were 10 students (4 male, 6 female) at the Kanack School in Rochester, NY. They ranged in age from 7 to 10 years (mean, 8.30; SD, 1.25) and were all right-handed. All of the students studied violin under the Suzuki method and had been playing for approximately 3 years, within a range of 2.5 to 3.5 years. Since they had played for the same amount of time, they all learned the songs from Volume 1 and had played them consistently since their first year. These students had one of two teachers, both of whom taught a standardized version of the Suzuki method. This method requires the children to sing the songs they learn on the violin and continue to sing and play those songs for the duration of those lessons.

During the study, nine out of ten participants returned surveys filled out by their parents. These data showed that two (~20%) of the participants study another instrument besides the violin, and five (~50%) have immediate family members who play at least one instrument. About half of the students showed an interest in music before they started playing the violin, some around ages 1-2 years and even younger. Seven of the students (~80%) were sung to by their parents when they were younger; interestingly, the two without the experience of parental singing were the two who performed the least well on the pitch-matching test.

The 10 students were split into 2 age groups of 5 students

[...]

jur.rochester.edu
In Session 1, the participant heard the first set of distractor tones to prevent the use of RP to find the starting pitch of the first song. The student was then asked to sing song #1, and only the first few notes were recorded. The investigator stopped the student after these few notes and repeated this process twice for two additional songs (songs 1, 2, 3). Then this process was repeated for three different songs (songs 4, 5, 6). During songs 4 through 6, students viewed the score for each song individually.

At the end of Session 1, the student was given the pitch-matching test; it was given at the end so that no tones from this test interfered with participants’ recall ability or choice of starting pitch when singing the songs. The participant sang back each pair of notes after hearing them, and the investigator digitally recorded the singing. The purpose of this post-test was to control for vocal development and ensure that the students were able to sing the notes they auralized in their minds.

Session 2 began with the same process as the first session, but with the last three songs first (songs 4, 5, 6), this time without the score. Then the student was asked to sing the first three songs from Session 1 (songs 1, 2, 3) with the scores visible, and those responses were recorded.

At the end of Session 2, the AP post-test was performed. As before, it was given at the end of the session so that no bias was created by hearing other experimental tones. Before each pitch sounded, one of the three sets of distractor tones was played. After hearing the distractor tone sequence and then the pitch, each participant named the pitch aloud and the investigator wrote down each answer (in order to eliminate any motor skill interference by the children and also to avoid any answering bias caused by having participants see previous responses). The purpose of this post-test was to identify the cognitive mechanism used to find the starting pitches. Since we were examining pitch memory in RP listeners, this test identified any AP listeners who presumably would have used a different cognitive process.

The distractor tones, the pitch-matching post-test and the AP post-test were composed once and used for all participants. The distractor tones were made up of 12 equally tempered pitches between A3 and G#4 in a sequence that lasted a total of 1 second. We randomized their order using a random-number table and their timbre was that of a synthesized violin in the musical notation program Finale. No pitch was repeated and there were no rests in the sequence. We used quarter, eighth, and sixteenth note values, the order of which was also determined using a random-number table. A total of three distinct distractor tone sequences were derived with the same rhythmic values and pitches, but randomized differently so that the students did not grow accustomed to the timbre of any one sequence.

The pitch-matching post-test was made up of 12 pairs of equally tempered pitches, whose starting notes were drawn from D4, E4, F4, G4 or A4; this is a comfortable singing range for children and encompasses most of the starting notes of their violin pieces. These starting pitches were ordered using a random-number table. Participants heard a two-note pattern (an interval). Each note lasted 1 second and had a synthesized violin timbre. The 6 intervals used were minor seconds, major seconds, minor thirds, major thirds, perfect fourths and perfect fifths. Each was heard twice, once ascending and once descending. The order of these intervals was also determined using a random-number table.

Finally, the AP post-test was made of the 18 equally tempered pitches between G3 and C5; their order was determined by a random-number table, and their timbre was a synthesized violin timbre in Finale. Each pitch was heard only once and lasted 1 second in duration. One of the three distractor tone sequences was played before each pitch sounded.

After both sessions of the experiment were complete for each subject, we scored their responses. In scoring the participants’ singing, only the starting pitch was evaluated. Octave errors were considered to be correct, consistent with Levitin’s analysis, since singers often change the octave to accommodate their vocal tessitura. Since octave errors were permitted, only pitch class (chroma) was examined (not pitch height), so that the furthest a response could be from the correct pitch was 6 semitones. Errors by one semitone were also considered correct in both the violin songs and the pitch-matching test, since even AP possessors sometimes produce discrepant responses by one semitone. This method of scoring also provided some leniency for premature vocal control. The actual spectral analysis provided the frequency (in Hz) and decibel (dB) level of each frequency within the sung tone. Only the steady state portion of the tone was analyzed. From the dB level, we discerned the intended sung pitch from its overtone series because the fundamental frequency was the loudest and most prevalent.

The analysis of the pitch-matching post-test was identical to that of the experiment proper: only the first note was analyzed. A score of 10/12 (83%) had to be obtained to pass the pitch-matching post-test. Semitone errors were considered correct for the same reasons as they were in the singing of the violin songs.

Scoring the AP post-test was simpler, as it only involved comparing the pitches named by the participants (and written down by the investigator) with the actual pitches of the notes heard. Participants were determined to have AP if they scored 16/18 (approx. 90%) or higher on the test. No semitone errors were permitted on the AP post-test. Two participants were honest enough to state that when they heard each given pitch, they were not
familiar enough with the letter names of the notes to assign a letter to that pitch. Instead, they offered to name the fingering on the violin of where that note would be. This unfamiliarity was probably the case with some of the other students as well, who did not protest the naming task; many of them were hesitant with the letter names, and some only used the letters E, A, D, B and G as responses, instead of all twelve possible note names.

In the two cases where the participants named fingering instead, there was an ambiguity in the scoring: when one says “finger 2 on the D string,” that note can either be F or F#, depending on where one places finger 2 on the string. Only once did this become an issue (all other times, neither note was correct), and in this case the experimenter assumed that the participant meant the correct note and not the incorrect one. Even with this assumption, the participant scored well below 90%, so ultimately this ambiguity proved to have no effect on our results: none of the participants had absolute pitch, according to our post-test. Their mean percent correct was only 10.00%, with a standard deviation of 6.56%. Their range of accuracy was 0% to 22% of the pitches correctly identified.

The results of the pitch-matching post-test showed that most of the children could hear and correctly sing back a two-note pitch pattern. Nine of the ten participants passed the pitch-matching post-test with a mean percentage correct of 98.11% and a standard deviation of 5.67%. The tenth subject sang only 25% of the pitches back correctly. His data was therefore omitted from all data analyses. Since most of the students passed the test, we confirmed that they were not hearing correct pitches and singing incorrect ones; they were actually singing the notes they auralized in their minds.

The results of the experiment overall were not what we anticipated. Contrary to our primary hypothesis, the children did not perform significantly better than chance in singing learned songs in the correct key. Chance was 0.25 (25%) because semitone errors were permitted; three possible correct answers of 12 possibilities yield a 3/12 or 0.25 probability of a correct answer by chance. The participants’ mean percentage correct was 31.44%, with SD of 17.12%. The results of a two-tailed, one-sample t-test of their performance against 0.25 revealed that they did not perform significantly better than chance would have predicted [t(9) = 1.129, p<0.292]. In fact, the participants sang the correct pitch almost as often as all of the other pitch classes, so the correct pitch or the semitones either side of it were not favored.

The hypothesized age effect yielded similar findings: the younger group of students did not perform significantly better than the older group. Levene’s test for equal variances found that the variances of the two age groups did not significantly differ [F(9)=0.329, p<0.584]. Five participants were in the younger group (mean % accuracy=0.33, SD=0.157) and four participants were in the older group (mean % accuracy=0.29, SD=0.210), since one participant’s data was omitted due to the pitch-matching problem. A two-tailed, independent samples t-test found that the younger group did not identify the pitches more accurately than the older group [t(9)=0.337, p<0.767].

Comparing the means of the two groups (M=2.85 semitones away without notation and M=2.74 with notation and their performance without notation [F=0.471, p<0.494]). Comparing the means of the two groups (M=2.85 semitones away without the score and M=2.74 with the score) revealed that having the score in front of them had no effect on what note they chose to sing [t(108)=0.297, p(2-tailed)<0.767].

Following the between-subjects analysis, we examined consistencies in pitch production within subjects, from the first session to the second. We found that participants were remarkably consistent in choosing a particular starting pitch for each song, regardless of whether that pitch was near or far from the objective standard. In other words, even if the participant chose to start a song in Session 1 on a pitch that was 6 semitones away from the correct pitch, the participant usually sang that song in Session 2 near that incorrect pitch, still about 6 semitones away. On average on the second day, the participants never chose a pitch more than an average of 1.56 semitones away from the pitch they chose on the first day.

To further test the relationship between the two sessions, we averaged each participant’s performance in each session, resulting in two scores per participant (one per session). A two-tailed Spearman correlation between these averages showed significant covariance in pitch production from
A group of children perform “Andantino” at an outdoor concert from memory.

the first to the second session [\(r=.803, p<.01\)]. In other words, most of the variance between sessions could be explained by which participant was singing, as each participant was quite consistent.

Thus the pitch representations that these children had in their minds were, whether correct or incorrect, relatively stable over time. The very high correlation between sessions within each student’s performance is consistent with the results obtained by Halpern and Bergeson and Trehub, both of whom examined repeated pitch production within participants. Halpern attributed her participants’ consistent performance to an aspect of absolute pitch in adults. This conclusion cannot apply to the current study because the children’s responses varied so widely from the objective standard. They sang both correct and various incorrect pitches consistently, so their pitch representations were stable but did not preserve the actual key of each piece.

On the other hand, Bergeson and Trehub attributed the consistent performances to a few factors that could apply to the current study. The first two are memory enhancers termed “positive affect” and “congruent mood.” These two factors set up a context of learned melodies. One would intuitively think that children with musical training would exhibit a similar, if not more secure, absolute memory for pitch. Since these children passed the pitch-matching test, we know they could produce the pitches they auralized, just as the adults could. Since they could sing the pitches they internalized, it is clear that they must have had the correct pitches of the objective standard internalized (i.e., they did not exhibit an accurate memory for pitch like the adults did).

These children may not have developed an absolute memory by this point in their childhood, but it is likely that at least some of these children will acquire one by young adulthood like Levitin’s participants. It is therefore possible that the explanation is one of cognitive development; perhaps the ability to form stable pitch representations for melodies at their objective standard develops during puberty. Another possibility is that these children have pitch memories, but not for these songs. Perhaps there exists a threshold number of times that one (adult or child) must listen to or sing a song within a certain time period before the correct pitch level is retained in and retrievable from the mind. The adults in Levitin’s study sang a favorite song by a pop artist and most likely sang this song almost daily, while these Suzuki students sang these violin songs while learning them, but ultimately more often played the songs on the violin.

The result that the younger children did not perform significantly better than the older children carries implications for research on AP acquisition. Most research done to support the AP early-learning hypothesis has examined adult participants with AP and compared their early musical training to that of participants without AP. In contrast, this study examined children and did not find that the children who started earlier showed any more instance of AP. Since this correlation between AP and early training was not supported by this study, one of three reasons may be responsible. First, perhaps the age around which the early-learning window begins to close is not 6 years as this study’s design specifies, but rather is age 7 or older. Second, perhaps there is a threshold age at which AP is expressed. The children who began music training early may indeed develop AP in the future (while the ones who did not may not develop AP), but they will not exhibit signs of AP or pass AP tests until a certain age or ability level. Third, the early-learning hypothesis may be relevant only for those children who do eventually exhibit AP. In other words, early training (vs. later training) may have no effect on children who do not have a genetic predisposition toward AP acquisition.1,2,7

Finally, the reason we found no effect of reading musical notation may lie in the particular pedagogy of the Suzuki method. One subject, when the score was placed before her, replied calmly that she did not need the score since she could not read music; she stated that she only looks at the music for dynamic markings. Critics of the Suzuki method have long stressed
this point: since Suzuki students learn through repetition and imitation, a few of them manage to learn to play an instrument using their ear and not ever reading the notes. Most of the other participants glanced cursorily at the score before turning away to think of how the song’s melody sounded. Only one participant studied the first note on the score to try and remember it. While one cannot deduce the effect notation had on the subjects by their physical reactions alone, it is clear by their equal performance with and without notation and by their nonchalant attitudes that having the score in front of them did not help them to remember the starting pitches because the participants did not use the score or retrieve any visual cues from it.

The results of the pitch-matching post-test and the AP post-test also deserve some attention. On the pitch-matching post-test, participants either sang almost all of the starting pitches correctly or they performed at chance level. Most of these children demonstrated the ability to internalize and sing back pitches. We therefore know that the children could sing each violin song in its correct key if given the starting pitch (i.e., if given a reference tone). However, they have not acquired the absolute memory for pitch that Levitin has shown many adults possess. Since the children performed at chance level in singing the violin songs in the correct key, perhaps the pitch memory skill is similarly dichotomous; like the ability to match pitch, they either have the ability to remember pitches or they do not (at least at this age).

The AP post-test results also hovered around chance level, which confirmed that no participants used AP to find a starting pitch. However, the fact that some of the students did not know the note names makes these results more difficult to interpret. By our definition of AP, if a student cannot match a note with its appropriate letter name, then the student does not have AP. But perhaps the student could have an idea of what that note is and just not have the pitch labeling capability because the notes have not yet been paired with their names in the students’ minds. It is difficult to determine the extent to which children have cognitive representations of pitches in their minds because many of them have not yet acquired the tools with which to express and identify them. Since the conventional definition of AP specifies both pitch labeling and pitch memory abilities, then these children cannot be said to have AP at this time if the cognitive link between a pitch’s sound and its name has not formed. As mentioned previously, however, these children may not yet have reached the developmental threshold at which students with early musical training begin to show signs of having AP.

It became clear from interpreting these results that the small sample size made a significant impact on the findings. Future research should sample a larger number of participants so that a single particularly good or particularly poor performance does not strongly affect the outcome. For example, one of the younger participants began zero out of twelve songs within a semitone of the correct pitch. This performance brought down the entire accuracy average for the younger group, and because the sample size is so small (n = 5), the average changed dramatically due to one participant’s performance. A larger sample size may allow significant effects of age or notation to surface and not be so drastically altered by individual performances.

It is difficult, however, to increase the sample size of children and still maintain an objective standard. Children are exposed to many melodies in many different keys; finding one or more pieces heard or sung consistently in the same key is difficult, and it is extremely difficult to find more than a few children exposed to that objective standard enough that they can reproduce it. Nonetheless, future studies attempting to find an effect of age on pitch production and absolute memory should sample more participants. Perhaps a design like Levitin’s would allow the children to choose a favorite popular song from a rack of CD's, which would ensure that they knew the song quite well.

Future studies should also examine the assumptions of the pitch-matching test. It is altogether possible that in short-term memory, pitch is internalized and reproduced by different cognitive processes than in long-term memory (for example, for a familiar song). Therefore, it could be erroneous to assume that if participants can create an accurate short-term representation of the pitch, they can also create an accurate long-term one. In fact, this point may provide an alternate explanation for the current findings: perhaps these children have not developed the cognitive ability to create long-term pitch memories or, if they have the ability, have not developed a way to access these memories. Overall, these results raise new questions about the cognitive development of AP and RP in children. Is there perhaps a chronological or developmental threshold past which these skills are exhibited? If so, will the aspects of absolute memory for learned melodies that exist in the adult population begin to exist in the younger population after a certain developmental threshold is reached? Also, will the stable pitch memory within each participant become more accurate (i.e., closer to the correct pitch) if the participant sings the song in the correct key more often? These findings encourage more rigorous examinations of aspects of absolute pitch in children and more empirical tests of the early-learning theory. Only by learning more about how we develop cognitively through childhood can we begin to foster that growth.

Victoria Saah received a B.A. in Music from the University of Rochester in May, 2003. This article is an abridged version of her senior honors thesis. Victoria is currently working in Erie, Pennsylvania, while applying to medical school. She plans to begin medical school in the fall of 2004.