# Absolute pitch information affects on-line melody recognition in non-AP perceivers

**Sarah C. Creel (creel@cogsci.ucsd.edu)**
University of California, San Diego, Department of Cognitive Science
La Jolla, CA 92093-0515

**Melanie A. Tumlin (mtumlin@cogsci.ucsd.edu)**
University of California, San Diego, Department of Cognitive Science
La Jolla, CA 92093-0515

## Abstract

Perception of absolute pitch (AP) has often been regarded as a qualitatively distinct ability, yet recent work has demonstrated that perceivers unable to label absolute pitches—the hallmark of true AP perception—still possess some knowledge of absolute pitch level. This is sometimes termed "implicit AP." What distinguishes the two types of AP? In two experiments using a melody-learning paradigm and eye tracking, we explore the pervasiveness and automaticity of implicit AP. We argue here that implicit AP reflects a phylogenetically older encoding of pitch information shared with other species, while "true" AP primarily reflects perception of pitch chroma, which may be unique to humans.

**Keywords:** absolute pitch, implicit absolute pitch, melody recognition, eye tracking, music perception

## Introduction

Do all listeners experience sound, music, in the same way? One major divergence from "normal" musical experience seems to be absolute pitch (AP), sometimes called perfect pitch. It consists of the ability to explicitly label particular pitches without reference to an external standard, and is extremely rare (Takeuchi & Hulse, 1993). Due to its rarity and apparently distinct manner of processing sound, there has been much interest in AP perception, as a developmental phenomenon (Miyazaki & Ogawa, 2006), as a correlate of brain morphology (Keenan, Thangaraj, Halpern, & Schlaug, 2001), and as a potentially genetically-specified trait (Gregersen, Kowalsky, Kohn & Marvin, 2000). However, the exact phylogenetic origins of AP perception remain somewhat mysterious, though it is a curiosity that animals tend to default to processing pitch in absolute terms (e.g. MacDougall-Shackleton & Hulse, 1996).

### True AP perception.

Several factors seem to be conducive to acquiring AP perception. One is music education early in life (Takeuchi & Hulse, 1993). However, not all individuals who receive early musical training acquire AP perception, which suggests that other factors must be at work. Another postulated factor is language exposure: Deutsch and colleagues (Deutsch, Henthorn, & Dolson, 2004) have suggested that speakers of tone languages (e.g. Mandarin) are more likely to develop absolute pitch than non-tone-language-speakers, because language forces them to attend to pitch. Other researchers have implicated genetic influences, suggesting that an apparently higher likelihood of AP perception in East Asians is likely hereditary (Gregersen et al., 2000). The ultimate outcome of this interaction of learning and biology is the effortless labeling of pitches according to pitch class—C, D, G$^{\#}$, E-flat, and so forth, with no need to hear an additional reference tone. Studies of memory encoding and interference in AP perceivers suggest that this ability is rapid and automatic: possessors can name individual pitches at much lower latencies than non-AP perceivers can calculate them based on a reference tone.

### Implicit AP.

Despite the rarity of AP perception, there have been numerous recent reports (Levitin, 1994; Schellenberg & Trehub, 2003) of non-AP possessors demonstrating some knowledge of absolute pitch content in their musical memories. This has been termed implicit AP: listeners cannot label individual pitches in the way that AP perceivers can, but perceive and produce music with some degree of absolute pitch accuracy. Levitin (1994) found that individuals without AP can reproduce the absolute pitch of a popular song relatively accurately. Also, individuals without AP are better than chance at discriminating between correct and pitch-shifted (1-2 semitones) versions of familiar songs (Schellenberg & Trehub), and infants can learn predictive AP patterns but not relative pitch patterns (Saffran & Griepentrog, 2001). These studies suggest that under some circumstances, listeners may store and recognize musical material in an absolute, rather than relative, form. This converges with numerous other demonstrations that listeners encode other detailed aspects of musical "surface" in memory, such as timbre (Schellenberg, Iverson & McKinnon, 1999) and articulation (Palmer, Jungers, & Jusczyk, 2001). These studies can be taken more broadly as evidence that listeners store acoustically accurate memories, and can discern whether a new instance does or does not match those memories. On this view, implicit AP perception is one of several consequences of having highly-detailed musical memory.

Nonetheless, there is much that is not understood about implicit AP perception and how it differs from true AP perception. First, how automatic is implicit AP perception—is it something listeners only attend to effortfully during recognition? If implicit AP perception is instead relatively automatic, then effects of AP match to memory should be evident fairly rapidly. Second, how obligatory is implicit AP perception? Is it something that listeners can ignore when in a more relative-pitch processing mode? If AP recognition is obligatory, listeners should experience interference when AP provides bad information for recognition.

In the current pair of experiments, we delve into the nature and pervasiveness of implicit AP perception. Using non-AP-perceivers, we ask whether absolute pitch information is an obligatory part of musical recognition, and how rapidly it is computed. For experimental control, we trained listeners to recognize brief (5-note, 1-second) novel melodies as "words" for unfamiliar pictures. After training, we tracked listeners' eye movements to correct and incorrect pictures as they heard a melody. Importantly, eye movements, which have been used for measuring word recognition for a number of years (e.g. Allopenna, Magnuson, & Tanenhaus, 1998), are a relatively implicit index of recognition. Thus listeners' eye movements should be minimally susceptible to conscious strategies.

Results from eye tracking as words are spoken have demonstrated that recognition is rapid and incremental (see Allopenna et al. 1998). That is, during a spoken word, listeners are updating a set of guesses as to what word they are hearing. This is reflected in eye movements. If two words share sounds initially, such as mask and mast, a listener will be equally likely to look at either a displayed picture of a mask or one of a mast until the end of hearing "mask" spoken. However, if the two words are dissimilar, such as mask and flute, the listener hearing "mask" will look to the mask around the beginning of mask. The time point where looks to two similarly-named pictures diverge suggests what sound information listeners are able to use in the speech signal to identify words.

The rationale in the current studies is similar. We taught listeners melodies with certain properties, and then examined how rapidly they fixated the correct picture (of two) when the melody "labels" did or did not overlap in absolute pitch. Among the melodies learned (Figure 1), certain pairs of melodies matched each other until the end, with either identical absolute pitch (AP-same; CDEFG: CDEFE) or with absolute pitch level differing by 6 semitones (AP-different; GFAGC; C$^{\#}$BD$^{\#}$C$^{\#}$F$^{\#}$). If listeners can use AP information to recognize melodies, they should look sooner to the correct object on AP mismatch trials than AP match trials.

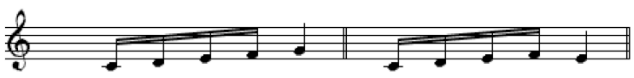
## Experiment 1

In this experiment, we trained listeners to associate melodies with pictures. We then measured looks to the pictures while listeners heard a melody "label" in real time to determine what cues listeners used to distinguish paired melodies. Some paired melodies matched in AP content, while the rest only matched in relative pitch terms. Importantly, all paired melodies were discriminable based on their final tone (in both relative and absolute terms), so that AP perception was not necessary to achieve perfect accuracy in the task.

### Method

**Participants.** *N*=17 members of the UCSD community, with varied musical backgrounds, received course credit for experimental participation. One participant was excluded for possessing AP perception, and was replaced. The final sample comprised 16 participants without AP perception.

**Stimuli.** Participants learned 16 melodies (Figure 1) as labels for 16 black-and-white pictures (examples in Figure 2). Melodies were all drawn from the diatonic major set, and were recorded in BarFly 1.73 software (Taylor, 1997; available at http://barfly.dial.pipex.com/) using the QuickTime instruments flute timbre. Melodies were distributed across 4 pitch ranges: C4-G4, F$^{\#}$4-C$^{\#}$5, C5-G5, F$^{\#}$5-C6. There were 8 pairs of melodies, and each pair began identically and diverged at the last note. The final interval differed in direction between the two members of a pair (one rose, one fell), to make melodies maximally discriminable. The onset of the last note in all melodies was 500 milliseconds (ms).

(a) Sample AP-same melodies

(b) Sample AP-different melodies

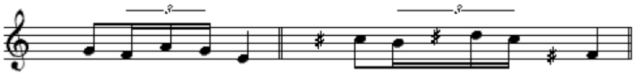


Figure 1. Sample melodies from Experiment 1. (a) AP-same pair; (b) AP-different pair.

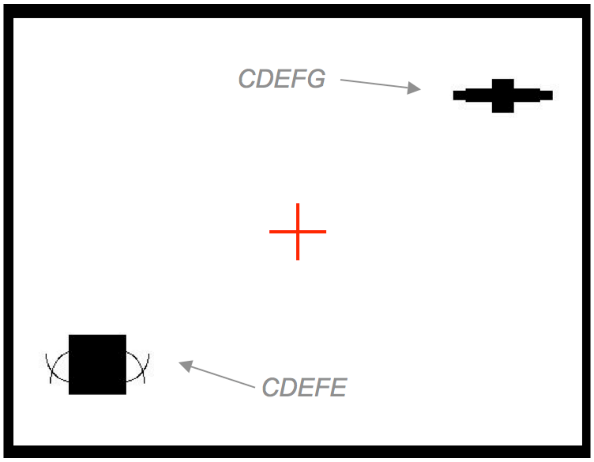


Figure 2. A sample test trial, with examples of two nonsense pictures. The pictures here are labeled with AP-same melodies.

For each pair, all intervals up to the final one were identical (same ratios between subsequent pitches).

However, for half the pairs, the pitches as well as the relative pitch intervals were the same (AP-same), while for the other half, the actual pitches were separated by a tritone and only the intervals were the same (AP-different). The tritone separation was selected to be comparable to Saffran and Griepentrog's (2001) AP experiment in which adults failed to learn to distinguish tone groups in a segmentation task. This also served to minimize confusion of the key area from melody to melody, as closely-related pitch areas tend to be parsed according to the preceding context (Bartlett & Dowling, 1980). AP match/mismatch was counterbalanced across melody pairs and participants.

Four different quasirandom melody-to-shape assignments were used to control against spurious cross-modal similarities between particular melodies and particular pictures. Each trial (see Procedure) showed pictures in two of four locations (upper left, upper right, lower left, lower right of screen); one of the two pictures was the target. The other picture was either the picture for the paired melody, or the picture for a particular dissimilar melody. The two types of "other" pictures occurred equally often, and each target appeared equally often in each of the four screen locations. This circumvented potential strategies that learners could use to avoid having to learn the melodies themselves (e.g., when picture X appears in the upper left, it is the target).

**Procedure.** During training, participants were instructed that they would see two pictures, would hear a melody, and would be asked to select the picture that went with the melody. After each trial, the correct picture stayed on screen, providing feedback as to correctness. Correctness was assessed after each 128-trial block. When a participant scored 90% correct in one block, they proceeded to the test phase. Testing was identical to training, except that no feedback was provided.

**Equipment.** All testing took place in a quiet room. Participants were seated in front of an Eyelink Remote eye tracker (SR Research, Mississauga, ON), as experimental stimuli were presented via headphones on a Mac Mini running OS 10.4 and Matlab 7.6. Matlab software was written by the first author using the PsychToolbox 3 (Brainard, 1997; Pelli, 1997) and Eyelink Toolbox (Cornelissen, Peters & Palmer, 2002). PsychToolbox also provided calibration routines. The eye tracker itself was controlled by a networked PC running Eyelink software in DOS. Data were processed off-line using custom scripts in Python written by the first author.

## Results

**Accuracy.** During the first three blocks of testing (Figure 3), a small but significant difference in error rates occurred between AP-matched trials and AP-mismatched trials (*p*=.002). Restricted just to paired trials, the effect did not reach significance (*p*=.1). This is an important result because it suggests that participants are not strategically using pitch height as a cue to discern between melodies (or if they are, they are not very successful). There was an effect of trial type (unpaired > paired) on error rates, *p* = .002, indicating that listeners found trials showing pictures with similar melodies to be more difficult.

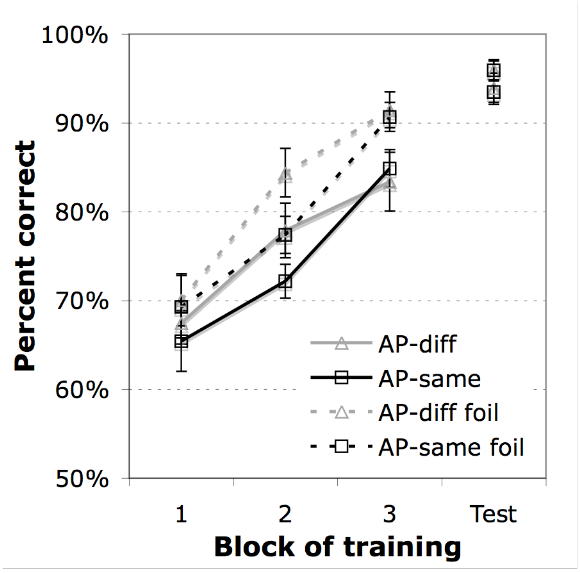


Figure 3. Accuracy during the first three blocks of training and test, Experiment 1. Error bars are standard errors.

**Gaze fixation patterns.** As is done in word recognition tasks, we defined a set of windows over which early effects should be visible, from 200 ms to 1000ms, and analyzed each 100ms window for a divergence in looks to the target (the correct object for that melody) or the other object onscreen (Figure 4). For AP-matched trials, the target-other difference did not reach significance until 700-800 ms (*p* = .007), the first conceivable time point at which listeners should be able to discern these melodies (onset of last note plus the 200 ms delay that it takes to plan and carry out an eye movement; see Hallett, 1986). However, for RP-matched trials, this divergence point was somewhat sooner, at 600-700 ms (*p* = .0008). This means that eye movements on RP-matched trials must have been planned prior to the point that final-interval information was available (between 400-500 ms).

## Discussion

AP rapidly and implicitly aids listeners in melody recognition. While we cannot rule out deliberate strategy use, if such strategies were in play, listeners did not seem to benefit: there was no significant reduction of errors for AP-different trials either before or during the test. That is, listeners were not significantly more accurate with AP-different melodies than with AP-same melodies. However, eye movements, which are difficult to consciously control, reflected more rapid recognition when an AP mismatch was present. This result supports the notion that non-AP-possessors both represent and use absolute pitch information in recognizing melodies. Further, storage of this information is consistent with a body of work demonstrating a high level of acoustic detail in listeners' musical representations, rather than representations that abstract over qualities such as musical prosody or absolute pitch content.

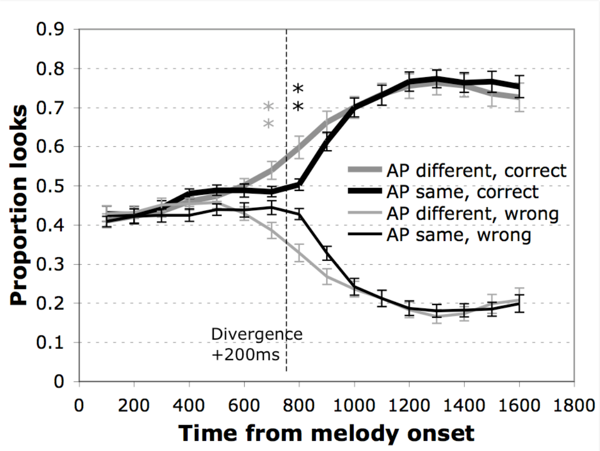


Figure 4. Looks to correct (thick lines) and incorrect (thin lines) pictures during test, Experiment 1. Error bars are standard errors. ***p*<.01

One potential counterexplanation of the above result is that listeners were not using an absolute pitch frame of reference, but a frame of reference relative to the pitch range of the entire set of stimuli (a "relative range" strategy). Recall that four pitch ranges were used in Experiment 1. That is, instead of encoding the absolute pitches of the stimuli, perhaps they encoded the pitch range, for instance, as low, mid-low, mid-high, and high. This is difficult to discriminate from absolute pitch even with a strongly delayed test phase, because as soon as the test phase begins the pitch range is reestablished.

We addressed this in Experiment 2 issue by requiring listeners to use relative pitch information, and to look for interference from absolute pitch processing. We trained participants on melodies at one set of absolute pitch levels (around C4, around F$^{\#}$4, around C5, around F$^{\#}$5) and then tested them at a different pitch level (F$^{\#}$4, C5, F$^{\#}$5, C6). We created a set of melodies where not two but three melodies overlapped until a final note. Two of the melodies were in one pitch range at training (such as F$^{\#}$4), while the third was pitched a tritone below at training (such as around C4). The first test block continued this pattern. The second and third test blocks, however, shifted all melodies up by exactly a tritone. If listeners are encoding pitch relative to the range of the experimental stimuli, then performance after the shift to the new pitch range should be equivalent to performance before the shift. If, instead, listeners are implicitly activating absolute-pitch matches, then trials which had not been AP-same during training should show interference at test (see Figure 5).

## Experiment 2

### Method

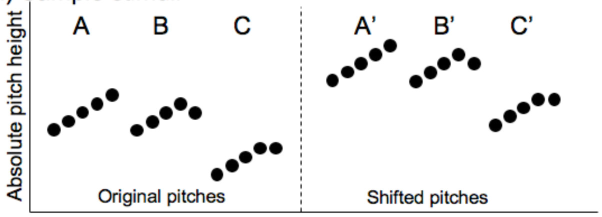
**Participants.** N=16 participants from the same pool as Experiment 1 completed the training and test phases.

**Stimuli.** There were 18 different melodies consisting of 6 sets of three (Figure 5), distinguished only at the final tone. Two of each set were identical in both RP and AP, while the third melody was a tritone lower and matched only in relative terms. All possible pairings of the melodies in a set of three yielded 1/3 AP-match trials and 2/3 AP-mismatch trials. The onset of the final tone in each melody occurred at 667 ms. Which melody in a triple was the low one was counterbalanced across participants.

**Procedure.** Training and testing proceeded similarly to Experiment 1, except that after one 72-trial block of testing, all melodies were shifted up in pitch by 6 semitones. There was a brief break before the shift during which participants conversed with experimenters. The effect of this shift was to set up the potential for interference from AP memory. That is, if memories of melodies were encoded in AP terms, then certain shifted melodies would now be competing with AP-identical traces of other melodies. In Figure 5b, for instance, if participants are comparing shifted melodies to AP memory traces, then shifted melody C' is now an AP match to (unshifted) melody A. Thus, interference for C' trials with A or B objects as competitors was expected to increase after the shift. This could manifest itself in terms of errors, fixation proportions, or both.

**Equipment.** This was identical to Experiment 1.

(a) Sample stimuli



(b) Pitch-shifted trials

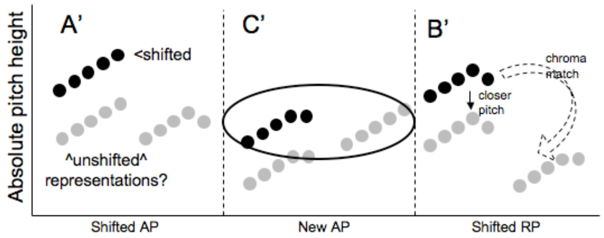


Figure 5. (a) Sample stimuli from Experiment 2. (b) Depiction of post-shift test trials. Gray indicates AP memories and black indicates the (shifted) melody presented on a trial. Circled area shows a new AP competitor.

### Results

**Accuracy.** We measured accuracy both during and after training. In training, AP pairs showed numerically lower accuracy than the two RP pair types, which did not differ. In the first test block, AP pairs were nonsignificantly less accurate than the two RP pair types combined, which again did not differ (original AP: 85% correct; new AP: 92%; shifted RP: 91%). In post-shift block 1 (Figure 6), there was a decided alteration in performance: while shifted-AP trial error rates and shifted-RP error rates stayed the same, new-AP trial performance declined (*p*=.005). One explanation might be that these errors occurred primarily in the trials immediately after the shift, during which listeners might be experiencing some confusion before adopting a RP perspective. Discounting this explanation, new-AP trials were still below the unshifted baseline in the shifted block 2 (*p*<.05), which presumably was ample time for recovery from the pitch shift. Note that this is not a general increase in all errors, only the errors for trials with an AP competitor in memory.

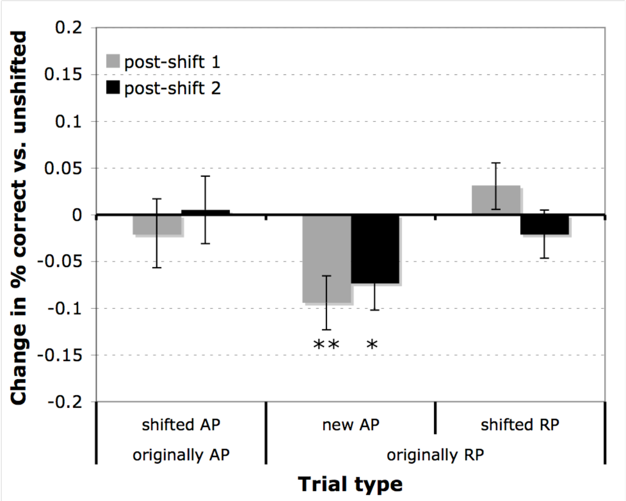


Figure 6. Accuracy changes in post-shift test blocks, Experiment 2. Error bars are standard errors. ***p*<.01, **p*<.05

**Gaze fixation patterns.** For the first (unshifted) block of test trials, correct looks on RP trials (that is, AP-mismatched trials) reached significance at 800-900 ms (*p*=.002), while correct looks on AP trials did not reach significance until 1100-1200 ms (*p*=.0009). This generally resembles the pattern in the first experiment, where AP-mismatched melodies were also recognized sooner. Fixations for the two shifted test blocks in general patterned with error rates, but were extremely noisy, presumably due to increased uncertainty on the part of participants.

## Discussion

In the current experiment, we tested whether participants were able to make an AP shift without any cost to recognition, and found that they could not. While the shift to RP processing was overall quite good—performance was well above chance (86%, $p < .0001$) after all melodies underwent a pitch shift of six semitones—participants were still hindered when a shifted melody occurred at the absolute pitch level of a previously-learned competitor melody, making more errors when a shifted melody overlapped in AP with an unshifted melody. This suggests that listeners were unable to ignore the AP content of the originally-learned stimulus. Such a result is consistent with the notion of obligatory use of acoustically accurate representations.

# General discussion

Implicit AP perception—access to accurate absolute pitch information in memory—appears to be rapid and obligatory in non-AP perceivers. In Experiment 1, listeners' ease of learning was not strongly affected by AP match or mismatch between melodies, yet listeners' eye movements reflected faster recognition of AP-different melodies (Experiment 1). Furthermore, listeners seemed unable to tune out AP information in a context where relative pitch processing would be advantageous (Experiment 2), suggesting that accessing musical memory obligatorily references absolute pitch content. Thus, both fixation latencies (Experiment 1) and pitch-shift errors (Experiment 2) reflect recognition costs associated with AP overlap. All of this implies that absolute pitch content is a necessary and relevant part of musical memory and the recognition of musical material.

## Comparison to true AP.

True AP is automatic, obligatory, and involves labeling of pitch chroma. Implicit AP seems to share some of these properties. It is automatic in that listeners use it rapidly for on-line recognition of melodies (Experiment 1), and is obligatory in that listeners cannot ignore AP content in an RP task (Experiment 2). Only labeling seems to be absent in implicit AP.

Recall that one aspect of true AP perception is that listeners identify certain pitches—those related by integer multiples that are powers of 2—as the same pitch class or "chroma." For instance, 220, 440, and 880 Hz are all perceived as the note A. This is salient enough to AP possessors that they occasionally make "octave errors," such as identifying an 880 Hz A as a 440 Hz A. There is no evidence that implicit AP contains chroma information. In fact, in Experiment 2, the RP-to-RP shifted trials were such that the melody closer in absolute pitch was correct, while the melody closer in chroma was incorrect. This did not lead to any increase in errors after the pitch shift. Thus, implicit AP may be more about pitch height than about pitch chroma.

## Origins of implicit AP perception.

One account of this pattern of results is that humans begin with the same pitch-processing abilities animals do—and that what animals possess is essentially implicit AP perception. For instance, animals generally do not display knowledge of chroma (though see Wright et al., 2000 for chroma use in a task tapping short-term memory). Animals instead show normally-distributed response distributions to learned AP cues, without spikes at octave doublings (e.g. Cynx, 1993). Animals also show interference from AP information when relative pitch processing becomes irrelevant (e.g. MacDougall-Shackleton & Hulse, 1996), as did humans in Experiment 2. Whether animals process pitch explicitly, as in true AP, or implicitly, as in implicit AP, is not clear. While animals are typically taken to possess absolute pitch processing akin to human true AP perception, proving true AP perception requires labeling. Given that it is extremely difficult to ask animals to react differentially to (i.e., label) more than two or three alternatives, it is hard to demonstrate much more than that animals possess somewhat better acuity in AP perception than do non-AP-perceiving humans (Njegovan, Ito, Mewhort, & Weisman, 1995). On the whole, it seems plausible that humans begin with essentially the same pitch perception abilities as many other animals, with attention to relative pitch increasing over development (see Saffran & Griepentrog, 2001). Thus, the initial state of humans AP perception may be homologous to animal AP. The apparent developmental shift away from absolute pitch processing is perhaps due to exposure to speech, which is typically consistent at the level of contour but not the level of absolute pitch (though there are exceptions to this; Deutsch, personal communication). Exposure to biologically significant, AP-inconsistent stimuli (speech) may explain why humans display poorer AP resolution than animals, who have had less opportunity to "unlearn" AP.

One open question about this experimental demonstration of AP storage is whether this information is maintained for longer periods of time (days, weeks, months). In a study asking a similar question, Marvin and Brinkman (1999) showed that even expert listeners could not determine whether short musical pieces began and ended in the same key. It is possible that consolidation processes in memory might remove even more AP variability at longer delays, though previous research (Levitin, 1994; Schellenberg & Trehub, 2003) suggests fairly accurate long-term maintenance. It remains for future investigation to determine whether the evident storage of AP information in memory remains or weakens over time.

## References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language, 38,* 419–439.

Bartlett, J. C., & Dowling, W. J. (1980). Recognition of transposed melodies: a key-distance effect in developmental perspective. *Journal of Experimental Psychology: Human Perception & Performance, 6,* 501-515.

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

Cornelissen, F.W., Peters. E., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments &Computers, 34,* 613-617.

Cynx, J. (1993). Auditory frequency generalization and a failure to find octave generalization in a songbird, the European starling (*Sturnus vulgaris*). *Journal of Comparative Psychology, 107,* 140-146.

Deutsch, D., Henthorn, T. & Dolson, M. (2004). Absolute pitch, speech, and tone language: Some experiments and a proposed framework. *Music Perception, 21,* 339-356.

Gregersen, P., Kowalsky, E., Kohn, N., & Marvin, E. (2000). Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment. *American Journal of Medical Genetics, 98,* 280-282.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance.* New York: Wiley.

Keenan, J. P., Thangaraj, V., Halpern, A. R., & Schlaug, G. (2001). Absolute pitch and planum temporale. *NeuroImage, 14,* 1402-1408.

Levitin, D. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics, 56,* 414-423.

MacDougall-Shackleton, S. A., & Hulse, S. H. (1996). Concurrent absolute and relative pitch processing by European starlings (*Sturnus vulgaris*). Journal of *Comparative Psychology, 110,* 139-146.

Marvin, E. W. & Brinkman, A. (1999). The effect of modulation and formal manipulation on perception of tonic closure by expert listeners. *Music Perception, 16,* 389-408.

Miyazaki, K., & Ogawa, Y. (2006). Learning absolute pitch by children: A cross-sectional study. *Music Perception, 24,* 63-78.

Njegovan, M., Ito, S., Mewhort, D., & Weisman, R. (1995). Classification of frequencies into ranges by songbirds and humans. *Journal of Experimental Psychology: Animal Behavior Processes, 21,* 33-42.

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory & Language, 45,* 526-545.

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

Saffran, J .R., & Griepentrog, G. J. (2001). Absolute pitch in infant auditory learning: Evidence for developmental reorganization. *Developmental Psychology, 37,* 74-85.

Schellenberg, E. G., Iverson, P., & McKinnon, M. C. (1999). Name that tune: Identifying popular recordings from brief excerpts. *Psychonomic Bulletin & Review, 6,* 641-646.

Schellenberg, E. G., & Trehub, S. E. (2003). Good pitch memory is widespread. *Psychological Science, 14,* 263-266.

Takeuchi, A. A., & Hulse, S. H. (1993). Absolute pitch. *Psychological Bulletin, 113,* 345-361.

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General, 129,* 291-307.