Cognition 122 (2012) 51–60

Contents lists available at SciVerse ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/COGNIT

Stimulus-dependent flexibility in non-human auditory pitch processing

Micah R. Bregman [a,b], Aniruddh D. Patel [c], Timothy Q. Gentner [b,d,*]

[a] Department of Cognitive Science, UC San Diego, La Jolla, CA, United States
[b] Kavli Institute for Brain and Mind, UC San Diego, La Jolla, CA, United States
[c] The Neurosciences Institute, San Diego, CA, United States
[d] Department of Psychology, UC San Diego, La Jolla, CA, United States

ARTICLE INFO

Article history:
Received 9 February 2011
Revised 9 June 2011
Accepted 8 August 2011
Available online 10 September 2011

Keywords:
Absolute pitch
Relative pitch
Pitch perception
Songbird
Auditory cognition

ABSTRACT

Songbirds and humans share many parallels in vocal learning and auditory sequence processing. However, the two groups differ notably in their abilities to recognize acoustic sequences shifted in absolute pitch (pitch height). Whereas humans maintain accurate recognition of words or melodies over large pitch height changes, songbirds are comparatively much poorer at recognizing pitch-shifted tone sequences. This apparent disparity may reflect fundamental differences in the neural mechanisms underlying the representation of sound in songbirds. Alternatively, because non-human studies have used sine-tone stimuli almost exclusively, tolerance to pitch height changes in the context of natural signals may be underestimated. Here, we show that European starlings, a species of songbird, can maintain accurate recognition of the songs of other starlings when the pitch of those songs is shifted by as much as ±40%. We observed accurate recognition even for songs pitch-shifted well outside the range of frequencies used during training, and even though much smaller pitch shifts in conspecific songs are easily detected. With similar training using human piano melodies, recognition of the pitch-shifted melodies is very limited. These results demonstrate that non-human pitch processing is more flexible than previously thought and that the flexibility in pitch processing strategy is stimulus dependent.

1. Introduction

Periodicity is a fundamental property of many natural sounds, and its perception, “pitch”, plays a central role in auditory processing in both humans and animals. Although humans are capable of using absolute pitch height in some auditory tasks, the intervals between sounds, referred to as “relative pitch”, are generally more important in sound recognition (Attneave & Olson, 1971). In contrast, most animals are thought to rely primarily on absolute pitch for auditory recognition (Hulse & Cynx, 1985; Hulse, Cynx, & Humpal, 1984; Page, Hulse, & Cynx, 1989). These comparative studies have focused almost exclusively, however, on stimuli constructed from sine-wave tones. Here, using songbirds as a model, we contradict this prevailing view by showing in a series of studies that non-human pitch processing strategies are flexible and stimulus-dependent. Over similar ranges of shifts in absolute pitch, recognition of conspecific songs is maintained, but recognition of tonal melodies is quite poor.

Among normal adult humans, the ability to reliably recognize pitches without an external reference is rare. Even musically trained individuals with excellent relative pitch often cannot provide an accurate note label (i.e. C, F#) for individual pitches (Ward, 1999). Among the general population approximately 1 in 10,000 (Ward, 1999) have this ability, known as musical absolute pitch (musical AP). In this paper, we use the term absolute pitch not to refer to this music-specific form of note recognition, but to memory for pitch height, which corresponds to the fundamental frequency of a sound.

---

* Corresponding author at: Department of Psychology, UC San Diego, 9500 Gilman Dr., MC 0109, La Jolla, CA, United States. Tel./fax: +1 858 822 6763.

E-mail address: tgentner@ucsd.edu (T.Q. Gentner).

0010-0277/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.cognition.2011.08.008

Long-term memory for pitch height appears to be common, even among individuals without musical AP (Schellenberg & Trehub, 2003). For example, when asked to hum or sing a very familiar song, the median of the distribution of people's starting pitches is broad, but centered close to the original source (Levitin, 1994). Likewise, many individuals without musical AP can correctly identify when versions of familiar television theme songs are transposed from their original pitch (Schellenberg & Trehub, 2003). Furthermore, musically untrained listeners (without AP) can identify whether a dial tone is at the correct pitch, too high or too low, even though they do not possess musical AP and are unable to attach specific pitch labels to the notes they recognize or produce (Smith & Schmuckler, 2008).

A widespread feature of human pitch perception is that humans easily recognize two sequences of notes as the same melody if the relationship between pitches is preserved, even if all of the pitches are different (Attneave & Olson, 1971; Dowling & Harwood, 1986). This ability is known as relative pitch, and is observed even in infants (Plantinga & Trainor, 2005). But the ability to recognize sounds independent of their absolute frequency content is more general in human cognition. Normal listeners have no difficulty understanding the same word spoken by two different individuals, even in a tonal language, because it is the pitch contour and relationship between pitches rather than the absolute frequency of words that are used to convey lexical and intonational meaning (Ladd, 2008).

Examples of relative pitch perception among non-human animals are rare. In fact, outside of human music perception, the only documented cases of relative pitch perception involve recognizing a single interval (Hurley, Ratcliffe, & Weisman, 1992; Shackleton, Ratcliffe, & Weary, 1992; Weisman, Ratcliffe, Johnsrude, & Hurly, 1990; Yin, Fritz, & Shamma, 2010). Relative pitch plays a role in the calls and song of some avian species, but again only a single interval between two notes has been implicated (Christie, Mennill, & Ratcliffe, 2004; Weisman et al., 1990). For example, chickadees shift the absolute pitch of their song but preserve the relative pitch between two notes (Shackleton et al., 1992). Recognizing a sequence of three or more pitches by the configuration of their intervals has not been reported in any of these cases. Rhesus monkeys do appear to recognize transposed tonal melodies, but only when transposed by octaves, not by 0.5 or 1.5 octaves (Wright, Rivera, Hulse, Shyan, & Neiworth, 2000).

Previous laboratory studies support a diminished role for relative pitch processing in songbirds and suggest instead that absolute pitch height processing is the dominant strategy (Cynx, 1995; Hulse & Cynx, 1985; Hulse & MacDougall-Shackleton, 1996; Nagel, McLendon, & Doupe, 2010; Page et al., 1989; Weisman et al., 1998, 2010). In a striking example of their limitation using relative pitch cues, European starlings never successfully learned to recognize ascending and descending tone sequences within a large stimulus set even after 10,000–30,000 training trials (Page et al., 1989). Failure to learn this recognition suggests an inability to use relative pitch when absolute pitch cues were not informative. Recognizing whether a tone sequences is ascending or descending would be trivial for most human subjects.

If a smaller stimulus set is used, starlings eventually learn to recognize ascending and descending tone sequences after many trials. However, when presented with stimuli outside of the pitch range used during training, subjects took as long to learn to accurately recognize the new stimuli as they did to learn the original stimuli (Cynx, 1995; Hulse & Cynx, 1985). Indeed, in all cases where pitch generalization has been observed, it is constrained by the absolute pitch range of the training examples (Cynx, Hulse, & Polyzois, 1986; Hulse & Cynx, 1985). Hulse termed this property the "frequency range constraint".

In contrast to their deficit recognizing stimuli on the basis of relative pitch, many species accurately use absolute pitch height cues. Several bird species, including zebra finches and chickadees perform well in pitch range discrimination studies (e.g. Lee, Charrier, Bloomfield, Weisman, & Sturdy, 2006; Weisman et al., 1998). In fact, these studies have repeatedly observed better encoding of pitch height than is observed in most humans (Weisman et al., 2010). Zebra finches (*Taeniopygia guttata*) are also able to detect mistunings in harmonic complexes with lower thresholds than humans, suggesting that they have very fine-grained pitch perception abilities (Lohr & Dooling, 1998). Thus, although songbirds can use both relative and absolute pitch cues for recognition, their ability to use relative pitch appears to be very limited, secondary to absolute pitch cues, and to require extensive training.

Does the observed bias toward absolute pitch cues in tone sequence recognition reflect general perceptual limitations in the ability to recognize stimuli shifted in pitch? To investigate this question, we studied pitch flexibility in a series of recognition tasks by European starlings, where a detailed set of prior experiments has established a bias toward absolute pitch cues in tone sequence recognition. We examined whether starlings continue to recognize their own species' song in the context of changing pitch. Are they committed to absolute pitch as a recognition cue, or are they more flexible?

We conducted operant training experiments using stimuli derived from conspecific song. In the first experiment, we trained two groups (of four subjects each) to recognize conspecific song excerpts. The first group learned to recognize songs that were always played at the same pitch, so that absolute frequency cues were useful for song recognition. The second group was trained with the same set of songs shifted to cover a range of absolute pitch levels. After training, we tested their ability to maintain recognition performance with these songs at novel pitch levels. In a second experiment, we tested whether two subjects could explicitly use pitch height cues in the context of song recognition by training subjects to recognize high pitch and low pitch versions of the same song, and then measuring recognition of songs at intermediate pitches. As a control, we also trained three starlings to recognize two tonal melodies, and then tested the generalization of this learned recognition to pitch shifted versions of the same melodies.

2. Methods

2.1. Subjects

Eleven European starlings (*Sturnus vulgaris*) of unknown sex served as subjects in these experiments. Previous operant song recognition studies (e.g. Gentner, Hulse, Ball, & Bentley, 2000) have not revealed any measurable behavioral differences in performance between male and female subjects. All subjects were captured as adults in Southern California between 2008 and 2010. After capture, subjects were housed in a mixed-sex aviary. The photoperiod in both the aviary and experimental apparatus corresponds to local sunrise and sunset times. All procedures were conducted as part of a protocol approved by the UCSD Institutional Animal Care and Use Committee.

Eight subjects participated in experiment 1 (recognizing pitch shifted song). Two of these subjects, as well as an additional naïve animal participated in experiment 3 (transposed melody recognition). Two naïve animals participated in experiment 2 (absolute pitch song recognition).

2.2. Apparatus

Subjects were trained using a custom-built operant response panel (Fig. 1a) housed inside a wire cage. This cage was placed inside a sound isolation chamber (Acoustic Systems). One side of the wire cage was modified to provide access to the operant panel. Each operant panel included three response ports spaced 6 cm apart. An infrared beam was used to detect pecks to the response ports, and a hopper accessible through the floor of the cage provided food reward. Stimuli were presented through a single full-range Radio Shack speaker mounted approximately 30 cm behind the response panel. A full-spectrum fluorescent bulb provided the lighting inside the apparatus. Custom software running on the Linux operating system presented stimuli, controlled lighting and food access, and recorded responses.

2.3. Stimuli and experiment summary

2.3.1. Experiment 1: pitch-shifted conspecific song

Eight subjects were trained to recognize excerpts of starling song. These subjects were divided into two training groups with four birds in each group. The first group was trained using four exemplars of starling song at original pitch (which we refer to as unshifted) while the second group was trained with 64 exemplars of song derived from the same four exemplars used in the first group, by pitch-shifting each exemplar up and down to 16 different pitch levels. Specifically, these 64 stimuli were constructed by phase vocoding each of the four song excerpts up to ±30% pitch shifts in increments of 4% (training pitch levels were at ±2%, ±6%, ±10%, ±14%, ±18%, ±22%, ±26%, ±30%). After training, subjects (in both groups) were tested on 32 novel pitch shifted song exemplars. These exemplars were created by pitch-shifting each of the original four song exemplars up to ±40% (test pitch levels were at ±4%, ±20%, ±34% and ±38%). This testing allowed us to evaluate the degree to which shifts in pitch disrupted song recognition. Complete details regarding training procedure are described in Section 2.4.

To construct pitch-shifted stimuli, we used a phase vocoding algorithm, implemented in Matlab (e.g. Dolson, 1986) to shift the pitch of starling song while keeping the temporal patterning and the relative pitch between song notes unaltered (Fig. 1b). Pitch shifting can be performed in two ways: additive (i.e. shifting the entire spectrogram in frequency) or multiplicative (multiplying spectral frequency by a fixed amount). Human pitch perception is logarithmic such that each doubling of frequency is perceived as the same interval (an octave). Hence multiplicative (or logarithmic) shifting preserves the perceived intervals between song notes. In this study we used logarithmic pitch shifting. Further details regarding the effects of phase vocoding are described in Section 2.3.4.

Source material consisted of four excerpts of starling song (14.8–15.0 s in length). Song excerpts were recorded in sound isolation chambers in the laboratory from two adult males. Each male contributed two excerpts, and were not experimental subjects in the study. These excerpts were used to create two stimulus sets. The first set contained the four exemplars at natural recorded pitch. The second set included 64 song exemplars, as described above. In each case, the songs from one male were used as "go" stimuli, and songs from the other were used as "no-go" stimuli. The test stimuli consisted of 32 novel pitch-shifted song exemplars, as discussed above, and were the same for both training groups. For subjects trained using four exemplars, all test stimuli were from

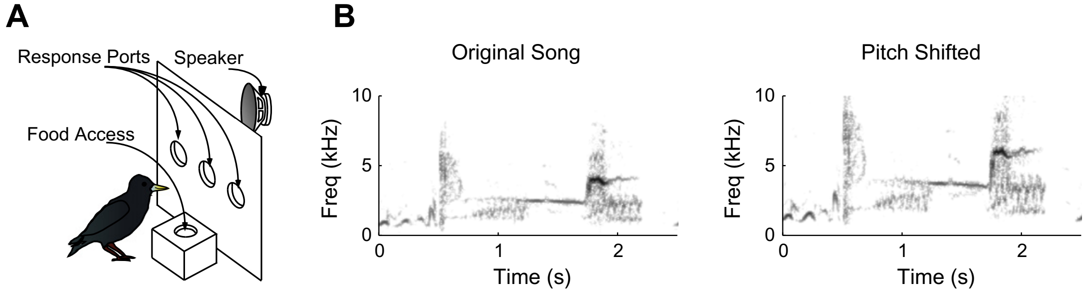

Fig. 1. (a) Schematic of operant panel used for behavioral training. (b) Spectrograms of two sample stimuli showing an example of logarithmic song shifting using phase vocoding. The spectrogram on the right represents a sound that has been pitch shifted up by 34% relative to the sound depicted in the left spectrogram.

outside the training pitch range. For subjects trained using shifted song, test stimuli were both from within the training range (within ±30% of the source recordings) and outside the training set (between ±30% and ±40% relative to the source recordings).

2.3.2. Experiment 2: absolute pitch recognition stimuli

To investigate whether subjects can attend to the absolute pitch of conspecific song in the context of a recognition task, we trained two subjects to recognize song when absolute pitch is the only available cue. In a 2-alternative-choice (2AC) procedure, subjects were trained to make a left response to a downward pitch-shifted song exemplar and a right response to upward shifted versions of the same exemplar. Four training stimuli were constructed from a single excerpt of starling song. Two corresponded to −24% and −28% pitch shifts relative to the original recording; the other two were shifted by +24% and +28%. After learning to recognize upward and downward shifted songs, the subjects were tested using 11 exemplars equally spaced between −20% and 20% in increments of 4% relative to the source recording.

2.3.3. Experiment 3: pitch-shifted melodies

To investigate the effect of shifting pitch on melody recognition, we trained three subjects to recognize the opening phrases of two human melodies, “Oh! Susanna” (3.97 s, 14 notes, mean 463 Hz, no-go response) and “London Bridge” (3.78 s, 11 notes, mean 368 Hz, go response) synthesized using a midi piano timbre. Each song was altered so that all notes were the same length to remove rhythmic cues that could otherwise be used for recognition. After learning to recognize these melodies, subjects were tested using versions of the learned melodies transposed up and down by 1–3 semitones as well as up by six semitones and 12 semitones (one octave).

2.3.4. Phase vocoding

We used phase vocoding (implemented in Matlab) to create stimuli for experiments 1 and 2. Phase vocoding is an analysis-synthesis technique that can alter stimulus length while preserving spectral features, or can shift spectral features while preserving stimulus length (Dolson, 1986). Phase vocoding does slightly alter other stimulus features as well as pitch, as do all resynthesis techniques. The most common artifacts introduced by phase vocoding are transient smearing, where the attack envelope of rapid temporal features is smoothed, and “phasiness,” which imparts a slight percept of reverberation to the altered sound. Both of these artifacts are most prominent at large shift amounts greater than ±30% (Laroche & Dolson, 1999). As one way of measuring possible signal changes due to phase vocoding, we measured differences in signal entropy (Tchernichovski, Nottebohm, Ho, Pesaran, & Mitra, 2000) and found no systematic increase in entropy when pitch was shifted away from natural pitch.[1] A priori, we would expect any artifacts introduced by phase vocoding to make song more difficult to recognize as it is shifted away from natural pitch. Because our main result demonstrates robust recognition of pitch-shifted song, any artifacts introduced by phase vocoding would likely weaken, not strengthen this result.

2.4. Behavioral training procedure

Each subject was trained and tested using one of two procedures: go-nogo (GNG, experiments 1 and 3) or 2-alternative-choice (2AC, experiment 2). We used a 2AC procedure for experiment 2 because, although it is slower for subjects to learn, it allows robust interpretation of responses to ambiguous stimuli when response rates tend to be low. For all subjects, the operant training procedure was separated into four sessions: shaping, stimulus recognition training, transfer, and stimulus recognition testing.

2.4.1. Shaping

To become familiar with the operant device, subjects are first cued with a flashing LED to peck the center response port to receive 4 s of access to the food hopper. After reliably pecking and obtaining food reward on 100 trials, subjects complete several hundred trials where they must peck to obtain reward without a visual cue. Complete details of the shaping procedure are described in other studies (Gentner, 2008).

2.4.2. Stimulus recognition and transfer training

During stimulus recognition training, a peck to the center response port initiated a trial and a stimulus was played from a speaker located behind the operant panel. After stimulus playback, a 4 s response window began, and pecks to the response ports during this window were recorded. For GNG training, a peck to the center response port following a “go” stimulus was immediately rewarded by 2 s of food access. A peck to the center response port following a “no-go” stimulus was immediately punished by a period (5–30 s) during which the cage lights were extinguished and a trial could not be initiated.

For 2AC training, after stimulus playback subjects pecked either the left or right response ports. Each response was rewarded if the correct response port for a given stimulus was pecked and was punished when the incorrect port was pecked.

No responses were recorded or reinforced after the 4-s response window ended and once the response window elapsed, the trial ended with neither food reward nor lights out punishment. After each trial, there was a 2 s inter-trial-interval during which a trial could not be started. When stimulus recognition performance reached a d-prime of 2 and did not increase in each of three consecutive blocks (1 block = 100 trials), subjects were transferred to the next training session.

During the transfer training session, experimental trials and stimuli were identical, but food reinforcement and punishment were reduced to occur randomly on 60% (instead of 100%) of trials. On the remaining 40% of trials where responses were made, the subjects were not reinforced with either food or a lights out period.

---

[1] Measured entropy was 3.20 ± 1.59 for original pitch vs. 2.86 ± 1.55 for −20% shifted song and 3.59 ± 1.47 for +20% shifted song (mean ± standard deviation).

2.4.3. Stimulus recognition testing

During test sessions, two stimulus sets were presented simultaneously. Presentation of learned stimuli occurred on 80% of trials, and responses to these stimuli continued to be reinforced on 60% of these trials (as during the transfer block). On the other 20% of trials, novel test stimuli were presented. Reinforcement of test stimuli was random: responses resulted in a food reward on 10% of trials, a time-out on 10% of trials, and no reinforcement on the remaining 80%. Test trials with no response were never rewarded or punished. Test stimuli were used to observe the subject's classification of novel stimuli within the context of the classification task learned during training. Consistent recognition of the novel test stimuli provides good evidence of generalization.

2.4.4. Data analysis

We evaluated subject's performance using d-prime (d') as a measurement of recognition, calculated as the difference in z-scores between the hit rate and false alarm rate (Green & Swets, 1966). D-prime is appropriate when response rate varies, and reflects both the correct responses to go stimuli and the false responses to no-go stimuli. Although statistically the threshold for above chance performance depends on response rate, d-prime values greater than one can be used as a heuristic to reflect good recognition performance. Where mean d' values are reported, we also report the standard error of the mean. All statistical tests were conducted with an alpha level of 0.05.

To evaluate whether an individual subject's performance was above chance, we calculated the 95% confidence interval for a chance d-prime value. D-prime at chance is always zero, but its variance at chance depends on both response rate and block size. Using a Monte Carlo simulation with 10,000 samples, we generated a simulated distribution of d-prime values for an animal that responds randomly with the observed response rate. The reported 95% confidence interval for chance performance was estimated using this distribution. Where recognition performance for groups of subjects is reported, we estimated the 95% confidence interval using the mean response rate of the group.

To characterize pitch generalization functions, we fit Gaussian functions to the group mean recognition performance. These functions are of the form:

$$Y = Ae^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

These estimates are descriptive, are used for comparison, and do not imply that the generalization is Gaussian in nature. We used these Gaussian models to estimate when pitch recognition performance would fall to chance, although we recognize this is only one of many possible analytic approaches.

To evaluate whether pitch generalization depended on training, we used the σ parameter from the Gaussian functions fit to each subject's individual generalization curve. We used a Wilcoxon rank sum test to evaluate the differences in curve width between the training groups. Larger σ values correspond to wider generalization curve widths and hence better recognition of pitch-shifted songs.

3. Results and discussion

3.1. Starlings rapidly learn to recognize pitch-shifted song

In experiment 1, we trained eight subjects to recognize excerpts of starling song. Four subjects learned four unshifted training exemplars, and the other four subjects learned 64 shifted exemplars. Recognition accuracy improved rapidly with training, with subjects reaching a d-prime of 1 in an average of 6 100-trial blocks when trained on the unshifted stimulus set, and 10 blocks when trained on the shifted stimulus set (Fig. 2). After 10 blocks of training, recognition performance reached a mean d-prime of 1.87 ± 0.33 for unshifted training subjects and 1.07 ± 0.47 for shifted training subjects and this difference was not statistically significant (ranksum = 13, $n = 8$, $p = 0.200$). Likewise, recognition accuracy increased over the course of training at similar rates for both groups ($F = 11.08$, $p < 0.001$ main effect of learning; $F = 0.91$, $p = 0.5523$ learning × group interaction, rmANOVA). Both groups received similar amounts of training (unshifted training subjects: mean 37 blocks, range 15–71; shifted training subjects: mean 36 blocks, range 18–69).

In operant song recognition tasks it is common to observe a dramatic decrease in acquisition rate when the size of the stimulus set is increased (Page et al., 1989; Wasserman & Bhatt, 1992). The small decrease in acquisition rate accompanying the 16-fold increase in stimulus set size suggests that the different pitch shifted song exemplars were not memorized as separate auditory tokens. Instead, subjects may have generalized across pitch-shifted versions of each song. Generalization to novel song exemplars may reflect the flexible use of pitch as one of several available cues. Alternatively, starlings may not use pitch at all during song recognition, in which case our stimulus manipulation would have been irrelevant. If pitch is not relevant, song may be instead be recognized through a combination of acoustic features, including those that are not disrupted by pitch shifting such as timbre and temporal modulations. In experiment 2, we directly address the issue of sensitivity to pitch cues in song recognition.

Even if pitch plays little role in song recognition, rapid learning of many stimuli is inconsistent with a model of

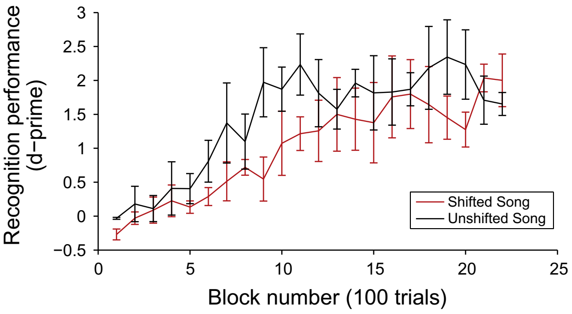

Fig. 2. Acquisition performance for two groups of subjects. Recognition performance (d') is plotted over time, measured in blocks of 100 trials. Each group has four subjects and performance is shown as group mean ± group standard error.

pitch perception that relies primarily on memorizing explicit, absolute pitch based, representations of song.

Subjects trained with both shifted and unshifted song recognized novel exemplars shifted by amounts up to ±38%. Recognition performance for song stimuli shifted to novel pitches (Fig. 3a) remained above chance for at least one subject at all tested shift amounts for those in the 4-exemplar group. Recognition performance for the 64-exemplar group was higher (mean d-prime of 1.92 ± 0.08) than for the 4-exemplar group (mean of 0.89 ± 0.12). An ANOVA revealed statistically significant main effects of pitch level ($F = 5.05$, df = 7, $p = 0.00023$) and training group ($F = 71.31$, df = 1, $p < 10^{-10}$). There was not a statistically significant interaction between pitch level and training group.

In general, recognition performance decreased as stimuli were shifted further from natural pitch. This decrease, however, was more rapid for subjects trained with unshifted song than those trained with shifted song. In addition to containing more exemplars, the shifted stimulus set contains songs with more variation in pitch height than the unshifted set. Prior work in recall of human speech suggests that acoustic variability can enhance memory for words (Nygaard, Sommers, & Pisoni, 1995). However, in word recognition tasks, performance decreased when trial by trial variability in the acoustic features of speech increased (Mullennix, Pisoni, & Martin, 1989).

We fit a Gaussian function to each subject's generalization performance and recorded the σ parameter (see Section 2). The group trained with shifted song showed wider generalization ($\sigma = 26.39 \pm 0.18$) than the group trained with unshifted song ($\sigma = 21.12 \pm 0.77$) and this difference was statistically significant (ranksum = 10, $n = 8$, $p = 0.0286$) (Fig. 3c). This difference in generalization performance provides evidence that pitch is a relevant cue for song recognition. If song pitch were an irrelevant cue for recognition, we would not expect to find a difference in generalization between the two training groups. Generalization performance was not correlated with duration of training and thus uncorrelated with the number of exposures to each stimulus ($r = -0.44$, $p = 0.2730$).

Improved generalization for the group trained with shifted song is consistent with research demonstrating improved generalization for subjects trained with varied stimuli that more fully characterize the test stimulus set (Greenspan, Nusbaum, & Pisoni, 1988). Furthermore, each of the test stimuli were closer in pitch to a training stimulus for the subjects trained with shifted song, so if starlings can tolerate a fixed shift amount before failing to recognize song, we would expect the generalization curve for those trained with shifted song to be wider.

The worst performance was observed for stimuli shifted by −38%, where only 1 of the 4 tested subjects trained with unshifted song showed above chance recognition performance. It is interesting to note that although there was little difference in acquisition rate between the subject groups, we do find a difference in the shape of the generalization curves for downward shifted song.

Although recognition performance decreased as song was shifted further from natural pitch, overall response rates (Fig. 3b) remained nearly constant across the test stimuli. Mean response rates were higher for subjects trained with the shifted stimulus set (0.59 ± 0.02) than those trained with the unshifted set (0.43 ± 0.04). A 2-way ANOVA revealed a significant main effect of training group (64 vs. 4-exemplar, $F = 9.73$, df = 1, $p = 0.0030$). There was no main effect of subject ($F = 0.72$, df = 7, $p = 0.6560$) and no significant interaction between training group and shift amount ($F = 0.12$, df = 7, $p = 0.9969$).

Previous studies using songbirds (including starlings) have observed dramatic decreases in recognition performance as pitch is shifted. In particular, they demonstrated that if pitch is shifted outside of the pitch range of the training stimulus, recognition performance falls to chance. In contrast, even in the case where subjects were only exposed to songs at a single pitch level during training, we observed good generalization performance for songs shifted up to ±20%. The subjects' use of pitch cues appears

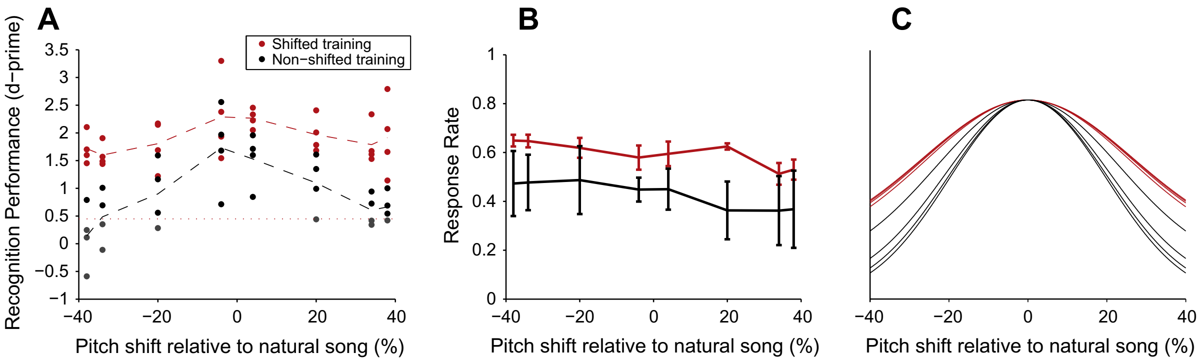

Fig. 3. (a) Generalization performance for novel shifted song exemplars. Each subject's recognition performance is plotted separately at each tested pitch shift. Subjects trained with four unshifted song exemplars are marked in black, subjects trained with 64 exemplars of pitch shifted song are marked in red. Mean recognition performance is indicated with the dashed line for each group. Points falling below the dotted horizontal line (all of which are part of the 4-exemplar group) correspond to those below the chance d-prime level of 0.46 and are marked in gray. (b) Response rates do not vary as pitch levels are shifted. Colors correspond to (a). (c) Gaussian functions depicting pitch generalization in each subject. Colors correspond to (a). Note how the generalization curves for subjects trained with pitch-shifted song (red curves) show a greater width (σ) than the curves for subjects trained with unshifted song. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

more flexible than has been observed in previous studies that have investigated pitch perception using tone sequence stimuli.

3.2. Subjects are sensitive to pitch differences in natural song

How does learned recognition generalize to novel pitch-shifted stimuli? Perhaps subjects hear and attend to pitch differences, but are able to flexibly use pitch as one of many song features for recognition when absolute pitch differences are not meaningful predictors of song identity. An alternative hypothesis is that starling subjects do not (or cannot) use pitch as a recognition cue for natural song stimuli.

In experiment 2, we tested whether two subjects could use absolute pitch height cues for song recognition. We trained each subject in a 2-alternative-choice task to respond left to “low” pitch stimuli (−24% and −28% relative to recorded pitch) and right to “high” pitch (+24% and +28%) manipulations of the same song excerpt. Both subjects learned this recognition task, reaching good performance (mean 89.1 ± 1.67% correct, last three blocks) demonstrating that they can use absolute pitch height to classify song stimuli. This observation is consistent with previous studies that report excellent absolute pitch perception in songbirds using tone sequence stimuli (e.g. Cynx, 1995; Weisman et al., 1998).

Testing two subjects using stimuli presented at intermediate pitches (Fig. 4) produced responses consistent with subjects using pitch to recognize these songs. As stimulus pitch increases, there is a general trend toward a decreasing proportion of left responses. This suggests that under such circumstances, subjects are able to make use of absolute pitch cues to recognize natural song excerpts.

The response rate for both subjects was high to stimuli near the pitches of the training stimuli, but decreased dramatically in the middle of the pitch range. This low response rate is unsurprising because it is these stimuli that are most ambiguous with respect to the training stimuli (neither high nor low in pitch). A low response for intermediate stimuli is exactly what we would expect when song pitch is a strong perceptual cue.

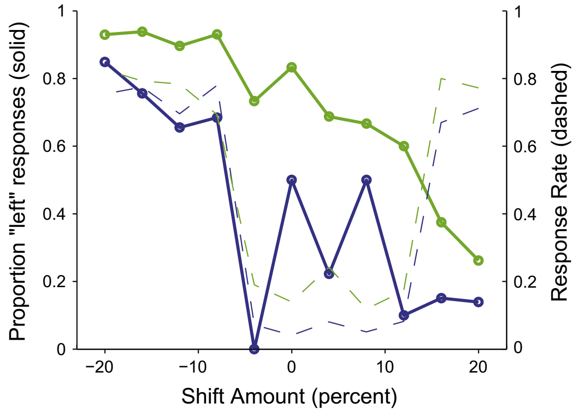

Fig. 4. Generalization performance to novel stimuli after subjects were trained to recognize high absolute pitch and low absolute pitch versions of the same conspecific song stimulus. For each of two subjects (subject 1 in green, subject 2 in blue), solid lines indicate the proportion of responses that were made to the left response port (associated during training with low pitch stimuli). Dashed lines indicate the response rate for each test stimulus. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

When absolute pitch is explicitly reinforced (as in experiment 2) subjects may use pitch as a primary recognition cue, but when pitch is uninformative and timbre is highly informative (as in experiment 1), they may use timbre exclusively. Although this possibility exists, it is unlikely that pitch is ignored completely, since generalization performance differed between groups trained with 64 vs. four song exemplars. If starlings were truly not using pitch at all in generalization, we might expect both groups to perform similarly when presented with novel pitch-shifted exemplars.

3.3. Melody recognition is impaired by transposition

Since previous studies found dramatically different results using tone sequences, in experiment 3 we tested three starling subject’s ability to recognize pitch-shifted human melodies synthesized using a piano timbre. Subjects learned to recognize two melody excerpts (isochronous versions of “Oh! Susanna” and “London Bridge”) at fixed pitch, although acquisition was slower than for the previous birdsong recognition tasks. One subject required 74 100-trial blocks to reach $d' > 1$ for two consecutive blocks. All subjects did, however, reach accurate performance (mean d' last 10 blocks: 2.52 ± 0.12). Subjects were then tested with melodies transposed 1–3 semitones, six semitones and 12 semitones (one octave). Transposition resulted in dramatically decreased performance (Fig. 5) with no subjects able to recognize the two melodies when shifted upward by three semitones (corresponding to an 18.92% pitch difference).

When compared directly with generalization performance for subjects trained with unshifted song exemplars, we observed worse generalization for melody stimuli. For example, whereas 3 out of 4 subjects trained with fixed pitch song stimuli accurately recognized song when shifted upward by 40%, no subjects recognized the melodies when shifted upward by six semitones (a similar shift magnitude of 41.42%). Moreover, all subjects failed to recognize melodies when they were shifted upward by three semitones (18.92%).

Qualitatively, we characterized the shape of the generalization by fitting a Gaussian distribution to the mean performance of all subjects as we did for pitch shifted song. For performing this fit, we used data points at −3, −2, −1, 1, 2 and 3 semitones, converted to their corresponding values in percent. The best fit distribution had parameters $A = 0.81$, $\mu = -1.11$, $\sigma = 11.65$.

For subjects trained with shifted song, the best Gaussian had parameters $A = 2.16$, $\mu = 4.98$, $\sigma = 55.75$. For subjects trained with unshifted song, the best function had parameters $A = 1.53$, $\mu = 3.30$, $\sigma = 24.55$.

Using these distributions we estimated the pitch shift amount at which subjects would fall to chance when recognizing melodies. We used the upper 95% confidence

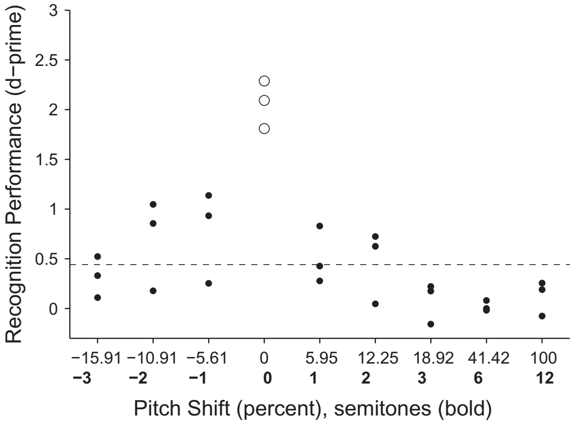

Fig. 5. Generalization performance to novel pitch shifted human melodies after subjects were trained to recognize two fixed-pitch human melodies. Horizontal axis is pitch shift (in percent, not to scale) and semitones (in bold) and vertical axis indicates recognition performance in d'. Performance for training melodies is shown using open circles. Each subject's performance for novel melody generalization is plotted separately (filled circles) and the upper 95% confidence interval for chance performance is shown as a dashed line. Subjects falling below this line performed at chance. Recognition performance is not different than chance for any subject at 3, 6 and 12 semitones (one octave), and at −3 semitones only one subject's performance was above chance.

interval for chance d' of .4477 based on a response rate of 0.5 and a block size of 100 trials. Based on these estimates, subjects trained with unshifted song could tolerate a downward shift in conspecific song to −35% before reaching chance while subjects trained to recognize melodies could tolerate a shift of −13% before falling to chance. This model also indicates that on average subjects could tolerate an upward shift of natural song of +42% before reaching chance, while only a +12% shift in melodic pitch sequences.

4. General discussion

While humans maintain accurate recognition of words or melodies over large pitch height changes, songbirds and other non-human animals have often been characterized as having difficulty performing such tasks. Perhaps this reflects fundamental differences in the structure or flexibility of songbird auditory representations. Alternatively, because non-human studies have used sine-wave stimuli almost exclusively, tolerance to pitch height changes in the context of more natural signals may be underestimated. We have shown that European starlings, a species of songbird, can maintain accurate recognition of otherwise familiar conspecifics songs that have been shifted in pitch by as much as ±40%, even though the pitch-shifts within this range of generalization are easily discriminable. In contrast, control birds trained on a similar task using human piano melodies, showed very poor generalization to pitch-shifted versions of the same melodies. Together, these results demonstrate that non-human pitch processing is more flexible than previously thought and that this flexibility in pitch processing is stimulus-dependent.

Our demonstration that starlings are able to generalize accurate song recognition across a wide range of pitch-shifts may reflect a latent capacity for relative pitch processing that has simply been unexplored. Alternatively, the songbirds may be able to easily ignore salient changes in pitch height in the presence of other invariant acoustic cues to recognition. Regardless, and although there may be major differences between species, it is important not to oversimplify these differences by characterizing non-humans as “absolute pitch processors” and humans as “relative pitch processors”. It is only in humanly-constructed tone sequences that we observe dramatic differences between songbirds and humans. In more natural contexts, in which birds are tested with their species-specific song, our understanding of how pitch guides recognition is much more limited.

Recognition of natural complex sounds is probably not tied exclusively to any single acoustic feature. Rather, pitch is one of many cues that can be integrated and weighted in the recognition process. Recent studies show that humans simultaneously integrate relative and absolute cues, even among listeners without AP (Schellenberg & Trehub, 2003; Smith & Schmuckler, 2008 Creel & Tumlin, in press). Prior research in songbirds has also characterized relative and absolute pitch cues' simultaneous availability, though strong evidence for relative pitch use has been more elusive (Hulse & MacDougall-Shackleton, 1996; Hulse et al., 1984; Page et al., 1989). Although there remains little evidence for facile relative pitch processing in songbirds, this limitation should not be generalized to conclude that songbird auditory memories are based on representations that are inflexible in pitch. Indeed, lack of observation of a particular ability in non-human animals should not be interpreted as strong evidence of inability.

Despite their difficulty recognizing pitch-shifted tone sequences, our studies demonstrate that starlings are able to recognize stimuli when their absolute pitch height varies widely. This raises two important points. First, starlings (and perhaps other songbird species) are not committed to absolute-pitch based representations for sound recognition. Second, as in humans, the processes underlying auditory recognition in other species may *fundamentally* depend on the stimulus. Therefore, it is problematic to move from studies using a single stimulus type (e.g. tone sequences) to general principles underlying sound recognition. This has important implications for neurophysiologists, who must use caution when studying higher-level auditory mechanisms using well controlled, but unnatural sine tone stimuli.

4.1. Topics for future research

Much research remains to be done, particularly to investigate how pitch is used as a cue in natural sound recognition. It is also important to continue to investigate the flexibility of pitch cues in songbirds and humans. Human pitch perception varies dramatically among individuals (AP possessors vs. non-AP individuals; musicians vs. non-musicians), varies with stimulus familiarity, and may vary over development. Studies in infants and young children suggest a bias toward absolute pitch cues, with a shift toward relative pitch cues in older children and adults (Saffran & Griepentrog, 2001; Stalinski & Schellenberg, 2010, but see Plantinga & Trainor, 2005). None of these topics have been extensively investigated in a non-human species.

Several studies suggest a developmental transition in children that allows relative pitch information to act as a primary cue for recognition. This transition may be due to exposure to sounds (i.e. speech and music) where relative pitch plays a crucial role during development. There is no evidence for a similar developmental shift in songbirds, although, as we have mentioned, there are no known cases where relative pitch (beyond the interval between two notes) plays a role in natural sounds. To test this hypothesis empirically, juvenile songbirds in the laboratory could be exposed to structured tone sequence stimuli where relative pitch was a relevant feature. Perhaps this exposure would improve generalization ability on transposed tone sequence recognition tasks.

A recent operant study examined recognition of pitch-shifted song in zebra finches (*T. guttata*), another songbird species. Zebra finches appear to only recognize pitch-shifted conspecific song within a narrow range of pitch shifts, and their performance decreases dramatically outside of the range of natural variation in song production, approximately 3–5% above or below natural pitch (Nagel et al., 2010). Our results, in contrast, show that European starlings recognize species-specific songs that have been shifted to a wide range of novel pitches without explicit training. The difference between these results may be due to species differences in auditory perception. Another possibility is that starling songs are more easily identified using cues other than pitch. In this case, we might expect zebra finches to perform similarly to starlings if the task were repeated with starling song excerpts. Exploring the range of natural sounds for which starlings can generalize across pitch could provide insight to the cues being used.

Previous studies have demonstrated that humans and songbirds use a multitude of acoustic features for recognition that are individually variable, flexible and species or language specific. Field sparrows, for example, seem to use many acoustic features in conspecific recognition, but tend to weight invariant features rather than highly variable ones (Nelson, 1988). This hypothesis is consistent with starlings weighting pitch less strongly as a cue when it becomes highly variable. Human second language learners may have difficulty recognizing non-native phonemes because they lack sensitivity to important acoustic cues that are uninformative in their native language (Iverson, Kuhl, Akahane-Yamada, & Diesch, 2003). Even within a native language, attention to different cues can be modulated by the informativeness of features in different tasks. Greenspan et al. (1988) showed that human listeners trained with speech synthesizers to recognize either words or sentences show no generalization across these domains. Likewise, Nygaard and Pisoni (1998) trained humans to extract relevant speech cues from either words or sentences of natural speech. When trained to attend to sentence-level cues, identification generalized only to words within sentences, not to isolated words. Cues used by learners predict some aspects of their learning success (Chandrasekaran, Sampath, & Wong, 2010). However, the most informative features do not entirely drive cue weighting—in some circumstances human listeners use certain cues even if others are more informative (Holt & Lotto, 2006). Thus, processing flexibility is not limitless.

An important goal of our work is to develop the songbird as a neurobiological model of auditory perception. To do so, we need a rich understanding of how humans and songbirds are similar and how they differ in their perceptual capabilities. Although prior work suggests that humans and songbirds exhibit large differences in pitch perception, our study encourages a more nuanced view. By altering their perceptual strategy to make use of the most salient and stable cues starlings have demonstrated a degree of flexibility in pitch processing that has not been previously observed in non-human animals. This opens the door to comparative studies that more accurately capture the full range of avian pitch processing mechanisms.

Acknowledgments

Research supported by NIH/NIDCD R01DC008358 to Timothy Q. Gentner. Aniruddh D. Patel was supported by Neurosciences Research Foundation as part of its program on music and the brain at The Neurosciences Institute, where he is the Esther J. Burnham Senior Fellow. The authors thank three anonymous reviewers for helpful suggestions.

References

Attneave, F., & Olson, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. *American Journal of Psychology, 84*(2), 147–166.

Chandrasekaran, B., Sampath, P. D., & Wong, P. C. M. (2010). Individual variability in cue-weighting and lexical tone learning. *Journal of the Acoustical Society of America, 128*, 456–465.

Christie, P. J., Mennill, D. J., & Ratcliffe, L. M. (2004). Pitch shifts and song structure indicate male quality in the dawn chorus of black-capped chickadees. *Behavioral Ecology and Sociobiology, 55*(4), 341–348.

Creel, S. C., & Tumlin, M. A. (in press). On-line recognition of music is influenced by relative and absolute pitch information. *Cognitive Science*.

Cynx, J. (1995). Similarities in absolute and relative pitch perception in songbirds (starling and zebra finch) and a nonsongbird (pigeon). *Journal of Comparative Psychology, 109*(3), 261–267.

Cynx, J., Hulse, S. H., & Polyzois, S. (1986). A psychophysical measure of pitch discrimination loss resulting from a frequency range constraint in European starlings (*Sturnus vulgaris*). *Journal of Experimental Psychology: Animal Behavior Processes, 12*(4), 394–402.

Dolson, M. (1986). The phase vocoder: A tutorial. *Computer Music Journal, 10*, 14–27.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. Orlando: Academic Press.

Gentner, T. Q. (2008). Temporal scales of auditory objects underlying birdsong vocal recognition. *Journal of the Acoustical Society of America, 124*, 1350–1359.

Gentner, T. Q., Hulse, S. H., Ball, G. E., & Bentley, G. E. (2000). Individual vocal recognition and the effect of partial lesions to HVc on discrimination, learning, and categorization of conspecific song in adult songbirds. *Journal of Neurobiology, 42*, 117–133.

Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.

Greenspan, S. L., Nusbaum, H. C., & Pisoni, D. B. (1988). Perceptual learning of synthetic speech produced by rule. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 421–433.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America, 119*, 3059–3071.

Hulse, S. H., & Cynx, J. (1985). Relative pitch perception is constrained by absolute pitch in songbirds (Mimus, Molothrus, and Sturnus). *Journal of Comparative Psychology, 99*(2), 176–196.

Hulse, S. H., Cynx, J., & Humpal, J. (1984). Absolute and relative pitch discrimination in serial pitch perception by birds. *Journal of Experimental Psychology: General, 113*(1), 38–54.

Hulse, S. H., & MacDougall-Shackleton, S. A. (1996). Concurrent absolute and relative pitch processing by European starlings (*Sturnus vulgaris*). *Journal of Comparative Psychology, 110*(2), 139–146.

Hurley, T., Ratcliffe, L. M., & Weisman, R. G. (1992). Relative pitch recognition in white-throated sparrows. *Animal Behaviour, 40*, 176–181.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., & Diesch, E. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*, 47–57.

Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge, UK: Cambridge University Press.

Laroche, J., & Dolson, M. (1999). Improved phase vocoder time-scale modification of audio. *IEEE Transactions on Speech and Audio Processing, 29*, 323–332.

Lee, T. T., Charrier, I., Bloomfield, L. L., Weisman, R. G., & Sturdy, C. B. (2006). Frequency-range discriminations and absolute pitch in black-capped chickadees (*Poecile atricapillus*), mountain chickadees (*Poecile gambeli*), and zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology, 120*, 217–228.

Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Percept Psychophysics, 56*(4), 414–423.

Lohr, B., & Dooling, R. J. (1998). Detection of changes in timbre and harmonicity in complex sounds by zebra finches (*Taeniopygia guttata*) and budgerigars (*Melopsittacus undulatus*). *Journal of Comparative Psychology, 112*, 36–47.

Mullennix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365–378.

Nagel, K. I., McLendon, H. M., & Doupe, A. J. (2010). Differential influence of frequency, timing, and intensity cues in a complex acoustic categorization task. *Journal of Neurophysiology*, 415.

Nelson, D. A. (1988). Feature weighting in species song recognition by the field sparrow. *Behaviour, 106*, 158–182.

Nygaard, L. C., & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Percept Psychophysics, 60*, 355–376.

Nygaard, L. C., Sommers, M. S., & Pisoni, D. B. (1995). Effects of stimulus variability on perception and representation of spoken words in memory. *Attention, Perception & Psychophysics, 57*(7), 989–1001.

Page, S. C., Hulse, S. H., & Cynx, J. (1989). Relative pitch perception in the European starling (*Sturnus vulgaris*): Further evidence for an elusive phenomenon. *Journal of Experimental Psychology: Animal Behavior Processes, 15*(2), 137–146.

Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition, 98*(1), 1–11.

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., & Trehub, S. (2003). Good pitch memory is widespread. *Psychological Science, 14*(3), 262–266.

Shackleton, S., Ratcliffe, L., & Weary, D. M. (1992). Relative frequency parameters and song recognition in black-capped chickadees. *The Condor, 94*(3), 782–785.

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

Stalinski, S. M., & Schellenberg, E. G. (2010). Shifting perceptions: Developmental changes in judgments of melodic similarity. *Developmental Psychology, 46*, 1799–1803.

Tchernichovski, O., Nottebohm, F., Ho, C., Pesaran, B., & Mitra, P. (2000). A procedure for an automated measurement of song similarity. *Animal Behaviour, 59*, 1167–1176.

Ward, W. D. (1999). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music*. Academic Press.

Wasserman, E. A., & Bhatt, R. S. (1992). Conceptualization of natural and artificial stimuli by pigeons. In W. K. Honig & J. G. Fetterman (Eds.), *Cognitive aspects of stimulus control* (pp. 203–223). Hillsdale, NJ: Erlbaum.

Weisman, R. G., Balkwill, L., Hoeschele, M., Moscicki, M. K., Bloomfield, L. L., Sturdy, C. B., et al. (2010). Absolute pitch in boreal chickadees and humans: Exceptions that test a phylogenetic rule. *Learning and Motivation, 41*, 156–173.

Weisman, R. G., Njegovan, M., Sturdy, C., Phillmore, L., Coyle, J., & Mewhort, D. (1998). Frequency-range discriminations: Special and general abilities in zebra finches (*Taeniopygia guttata*) and humans (*Homo sapiens*). *Journal of Comparative Psychology, 112*(3), 244–258.

Weisman, R. G., Ratcliffe, L., Johnsrude, I., & Hurly, T. A. (1990). Absolute and relative pitch production in the song of the black-capped chickadee. *The Condor, 92*(1), 118–124.

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.

Yin, P., Fritz, J. B., & Shamma, S. A. (2010). Do ferrets perceive relative pitch? *Journal of the Acoustical Society of America, 127*(3), 1673–1680.