

COMPARATIVE COGNITION & BEHAVIOR REVIEWS

Why Doesn't a Songbird (the European Starling) Use Pitch to Recognize Tone Sequences? The Informational Independence Hypothesis

Aniruddh D. Patel

Tufts University

*Azrieli Program in Brain, Mind, & Consciousness,
Canadian Institute for Advanced Research (CIFAR)*

It has recently been shown that the European starling (*Sturnus vulgaris*), a species of songbird, does not use pitch to recognize tone sequences. Instead, recognition relies on the pattern of spectral shapes created by successive tones. In this article I suggest that rather than being an unusual case, starlings may be representative of the way in which many animal species process tone sequences. Specifically, I suggest that recognition of tone sequences based on pitch patterns occurs only in certain species, namely, those that modulate the pitch and spectral shape of sounds independently in their own communication system to convey distinct types of information. This informational independence hypothesis makes testable predictions and suggests that a basic feature of human music perception relies on neural specializations, which are likely to be uncommon in cognitive evolution.

Keywords: *music, pitch, songbirds, speech, evolution*

Introduction and Questions Addressed

Cross-species studies of music perception allow one to investigate the evolutionary history of specific components of music cognition (Fitch, 2015; Hoeschele, Merchant, Kikuchi, Hattori, & ten Cate, 2015; Honing, ten Cate, Peretz, & Trehub, 2015; Patel, in press; Patel & Demorest, 2013). For example, a fundamental aspect of human music cognition is the ability to perceive a beat in rhythmic auditory patterns and to synchronize body movements to this beat in a predictive manner across a wide range of tempi. It has recently been shown that some nonhuman animals have this capacity, whereas others (including, surprisingly, nonhuman primates) may lack it (see Patel, 2014, for a review). This suggests

that the capacity may reflect specialized neural mechanisms present only in certain animal lineages. Studying which animals do versus do not have this capacity can give us clues to what these mechanisms are and why they evolved.

Turning from rhythm to melody, how do other species process instrumental (nonvocal) human melodies? Darwin (1871, p. 333) believed that basic aspects of melody (and rhythm) perception reflected ancient brain mechanisms widely shared among animals, writing that “the perception, if not the enjoyment, of musical cadences [i.e., melodies] and of rhythm, is probably common to all animals, and no doubt depends on the

common physiological nature of their nervous systems.” One way to test this idea is to ask if other animals rely on the same perceptual attributes as humans when recognizing tone sequences (cf. Crespo-Bojorque & Toro, 2016).

In human music, melodic sequences typically employ tones that have complex spectral structure. Often these are complex harmonic tones, consisting of a fundamental frequency and harmonics that are integer multiples of the fundamental. This acoustic structure is characteristic of many instruments (e.g., the clarinet, trumpet, etc.) and of the human speaking or singing voice when producing vowels (Stevens, 2000; Sundberg, 1987). Melodic sequences have many perceptual attributes, such as patterns of pitch, duration, timbre, and loudness. Yet humans regard a tone sequence played on different instruments (i.e., with distinct timbres) as “the same melody” if the pattern of pitches is the same. Of course, humans also recognize that the instrument has changed (e.g., a clarinet vs. a trumpet), but we gravitate to pitch patterns as the basis for tone sequence recognition. This is a fundamental aspect of human music cognition.

Is using pitch as the primary cue for recognizing tone sequences a basic and widespread feature of auditory processing? Songbirds, long considered among nature’s most musical creatures (Marler & Slabbekoorn, 2004), are an excellent animal model for addressing this question. Consistent with the view that melodic processing taps into ancient brain mechanisms, it has long been thought that songbirds, like humans, rely on pitch for tone sequence recognition (albeit absolute pitch rather than relative pitch, as discussed below). However, a recent study has challenged this view, because it shows that European starlings (*Sturnus vulgaris*) do not use pitch for recognizing sequences of complex harmonic tones (Bregman, Patel, & Gentner, 2016). Instead, recognition appears to be based on the pattern of spectral

shapes created by the successive tones. These results raise several questions, three of which are addressed in the current article:

1. Why would starlings gravitate to spectral shape, not pitch, for tone sequence recognition?
2. Why do humans show the converse pattern, spontaneously gravitating to pitch patterns for tone sequence recognition?
3. Which way of processing tone sequences, starling or human, is more evolutionarily ancient and widespread among animals?

The purpose of this article is to offer answers to the first two questions based on considering the role of pitch versus spectral shape in starling versus human communication systems. The answer to the third question must of course await further cross-species work. However, at the end of this article I suggest that the human way of perceiving tone sequences may be relatively rare. This suggestion is framed as a hypothesis about the evolutionary prerequisites for perceiving tone sequences based on pitch patterns.

Before addressing these three questions, it is first necessary to clarify the relationship between pitch, timbre, and spectral shape. *Pitch* is the “highness or lowness” of a sound. In a complex harmonic tone, the pitch typically corresponds to the fundamental frequency (F0), although the neural mechanisms that derive pitch are not simple “F0 detectors” and instead involve analysis of the tone’s spectral and temporal structure (Oxenham, 2013). Thus the fundamental frequency can be physically absent from a complex harmonic tone and yet still be perceived as the pitch of the tone (the “missing fundamental”). Pitch is therefore a perceptual (vs. purely acoustic) attribute of sound. Further evidence that pitch is a perceptual attribute is the fact that for certain harmonic complexes, individuals can show salient differences in the pitch they perceive (Ladd et al., 2013). *Timbre*, or sound quality, is what distinguishes two sounds when they have the same pitch, intensity, and duration, yet remain discriminably different (e.g., a clarinet vs. trumpet playing the same note). Timbre is a perceptual attribute derived from many aspects of a sound’s acoustic properties including its spectral structure, amplitude envelope, and how both of these change over time. (For a brief introduction to timbre, see Patel, 2008, Chapter 2; for more detail, see McAdams, 2013). *Spectral shape* is one aspect of a sound’s structure that contributes to timbre, and is less detailed than its full

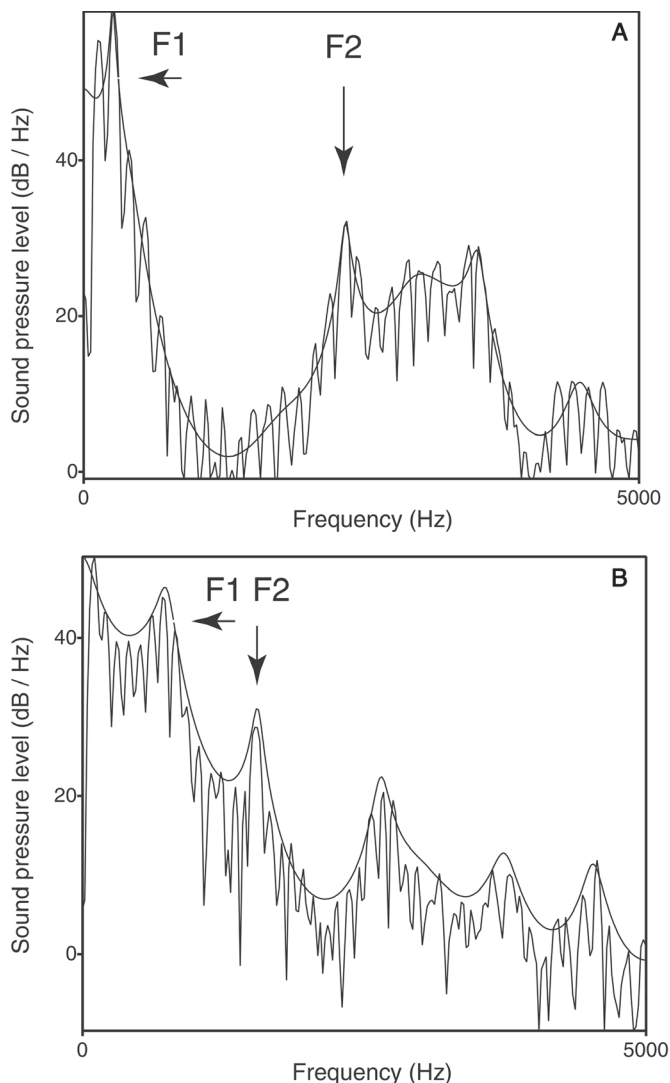
Author Note: Aniruddh D. Patel, Department of Psychology, Psychology Building, Room 118, Medford, MA 02155.

Correspondence concerning this article should be addressed to Aniruddh D. Patel at a.patel@tufts.edu.

Acknowledgments: I thank Frederic Theunissen for information about the modulation of spectral shape by songbirds; Tim Gentner for providing the starling song spectrogram in Figure 1; Tecumseh Fitch, Bob Ladd, Shihab Shamma, and Chris Sturdy for helpful feedback on the ideas in this article; and Marisa Hoesechele for careful and thoughtful editing.

spectral structure because it refers to the *overall* distribution of energy across frequency bands. Spectral shape is also distinct from pitch, because the same spectral shape can be realized with different fundamental frequencies (e.g., when a person utters the same vowel with low vs. high F0) or with the presence or absence of pitch (e.g., when a person produces a voiced vs. whispered version

Figure 1. Frequency spectra for the vowels /i/ (in “beat”) and /æ/ (in “bat”) are shown in panels A and B, respectively. In these spectra the jagged lines show the harmonics of the voice and the smooth curves show the spectral shape or “spectral envelope.” The formants are the peaks in the spectral envelope. The first two formant peaks (F1, F2) are indicated by arrows. The vowel /i/ has a low F1 and high F2, and /æ/ has a high F1 and low F2. These resonances result from differing positions of the tongue in the vocal tract. From *Music, Language, and the Brain* (p. 57), by A. D. Patel, 2008, New York, NY: Oxford University Press. Copyright 2008 by Oxford University Press. Reprinted with permission.



of the same vowel). As an illustration of spectral shape, Figure 1 shows the spectra of two vowels spoken at the same F0.

The vertical jagged lines show the harmonics of the voice, and the spectral shape of each vowel is traced by a thin black line draped along the top of the spectra. Note how this thin line does not preserve the fine details of spectral amplitudes across frequencies: Instead it shows the overall distribution of energy, that is, its spectral shape or “spectral envelope” (I use “spectral shape” and “spectral envelope” interchangeably in the remainder of this article). Research on speech perception has shown that spectral envelopes (and their pattern of change over time) carry a good deal of information regarding the identity of phonemes, independent of pitch. Specifically, if spectral envelope patterns are retained, speech remains highly intelligible even if all pitch information is removed (i.e., if noise is used as the carrier signal). The technique used to demonstrate this, called “noise vocoding” (Davis, Johnsrude, Hervais-Adelman, Taylor, & McGettigan, 2005; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), was used by Bregman et al. (2016) to show that spectral shape, not pitch, was the primary cue starlings use to recognize sequences of complex harmonic tones.

I now turn to addressing the three questions listed above, beginning with a review of the research that led to these questions, including the findings of Bregman et al. (2016).

Previous Findings on Tone Sequence Recognition by Starlings

Research with European starlings (henceforth, starlings) has played a key role in cross-species studies of music cognition. Starlings are excellent candidates for such research because they rely on complex auditory processing in their own communication system. These vocal-learning songbirds produce long, acoustically rich songs in nature (Gentner & Hulse, 2000) and are vocal mimics, capable of imitating nonconspecific sounds including the calls and songs of other birds (Hindmarsh, 1984). In terms of auditory psychophysics, starlings are among the best studied nonhuman species and show several broad similarities to humans, including their audiograms and auditory filter widths (Dooling, Okanoya, Downing, & Hulse, 1986; Klump, Langemann, & Gleich, 2000). Furthermore, although birds lack the six-layered neocortex found in mammals, neuroanatomical research has revealed that the songbird auditory pallium

is functionally analogous to the mammalian auditory cortical microcircuit (Calabrese & Woolley, 2015; Karten, 2013). Consistent with this finding, experiments have shown that starlings, like humans, perceive the pitch of the missing fundamental of harmonically structured tones (Cynx & Shapiro, 1986). These observations, combined with the fact that the avian auditory system follows the general vertebrate plan (Carr, 1992), might naturally lead one to expect that starlings and humans would share basic features of melody perception.

One basic feature of human melody perception is the ability to recognize a familiar melody when it is “transposed” in pitch (shifted up or down in log frequency). For example, when the “Happy Birthday” tune is played on a piccolo versus a tuba, a human listener effortlessly recognizes the tune in both cases, even if that person has never previously heard it played in such a high or low pitch register. This shows that human melodic recognition does not depend on the absolute pitches of notes but on the pattern of relations between pitches, or “relative pitch” (the pattern of pitch intervals between notes, which remains constant across different transpositions). Although humans do show some memory for the absolute pitch of familiar tunes (Creel & Tumlin, 2012; Levitin, 1994), and a small percentage of people develop the ability to recognize individual tones based on their absolute pitch (Levitin & Rogers, 2005), most humans strongly rely on relative pitch for melody recognition, beginning in infancy (Plantinga & Trainor, 2005).

At first glance, melodic recognition based on relative pitch seems a basic ability, likely to be common among animals. Indeed, early Gestalt psychologists used the recognition of transposed melodies as an example of holistic perception whereby objects retain their identity when relations between their parts are maintained even if the identity of individual parts is changed (Rock & Palmer, 1990). Gestalt principles are generally not assumed to be uniquely human, because many animals need to recognize objects based on relational rather than absolute features. Thus it is reasonable to expect that starlings, like humans, would readily recognize familiar melodies when they were transposed.

In this light, a series of experiments by Stuart Hulse and colleagues (commencing with Hulse, Cynx, & Humpal, 1984) produced surprising results. These studies showed that starlings could easily be trained to discriminate between different tone sequences (e.g., ascending vs. descending in pitch), but they did not generalize this discrimination when the sequences were transposed outside of the training range. (When transpositions

remained within the training range, they did show some generalization, a finding termed the “frequency range constraint”; Hulse, Page, & Braaten, 1990). Subsequent work replicated this finding and revealed that it was not unique to starlings (reviewed in Hulse, Takeuchi, & Braaten, 1992). Nonrecognition of transposed tone sequences was also observed in several other avian species; in rats; and even in a primate, the capuchin monkey (D’Amato, 1988; though see Wright, Rivera, Hulse, Shyan, & Neiworth, 2000, for different results with Rhesus macaques). These findings led to the widespread belief that songbirds recognize tone sequences on the basis of absolute (rather than relative) pitch, that is, based on the specific frequencies used in tone sequences (Weisman, Williams, Cohen, Njegovan, & Sturdy, 2006; though see Hoeschele, Guillette, & Sturdy, 2012). This view was strengthened by the finding that songbirds can readily learn to categorize a large set of pure tones (spanning more than two octaves) into eight alternating “go” and “no go” frequency bands based on their absolute frequency, a task on which most humans (i.e., those without musical absolute pitch) do poorly (Weisman, Njegovan, Williams, Cohen, & Sturdy, 2004).

This lack of relational pitch processing in starlings was even more surprising given that these birds are capable of relational processing for other aspects of tone sequences. For example, starlings can learn to discriminate between tone sequences that increase versus decrease in loudness and can generalize this discrimination to different loudness ranges (Bernard & Hulse, 1992). They can also discriminate isochronous from arrhythmic tone sequences and generalize this discrimination to new tempi (Hulse, Humpal, & Cynx, 1984).

Thus it seems that recognition of frequency-shifted tone sequences (i.e., melodic recognition based on relative pitch) may rely on specialized (vs. ancient and widespread) neural mechanisms. Consistent with this idea, neuroimaging work with humans has revealed that recognition of transposed melodies involves a complex network of regions including interactions between auditory and parietal cortex (Foster & Zatorre, 2010).

Testing With More Complex Acoustic Stimuli

When an animal seems to lack a seemingly “basic” perceptual or cognitive capacity seen in humans, such as recognition of transposed tone sequences, it is important to determine if testing with different methods or stimuli would lead to different results. In particular, it is important to consider whether more naturalistic

stimuli or tasks would reveal capacities that might otherwise remain hidden (de Waal, 2016). Inspired by these concerns, and by evidence for songbird brain regions sensitive to conspecific vocalizations (Doupe, 1997), Bregman, Patel, and Gentner (2012) tested whether starlings could recognize conspecific songs that were shifted up or down in frequency. Specifically, starlings were trained to discriminate between different conspecific songs and tested for their ability to recognize these songs when they were shifted up or down in frequency (by up to 40%). Unlike previous findings with melodies, the starlings readily recognized the shifted songs for shifts both within and outside the training range.

The findings of Bregman et al. (2012) inspired us to think about how stimulus characteristics might have influenced the results of past studies on starling

recognition of transposed melodies. Most such studies used pure tones, which are quite unlike the sounds used by starlings in their own communication system. Starling songs are spectrotemporally complex, with a

Figure 2. An excerpt of European starling song, illustrating its spectrotemporal complexity.

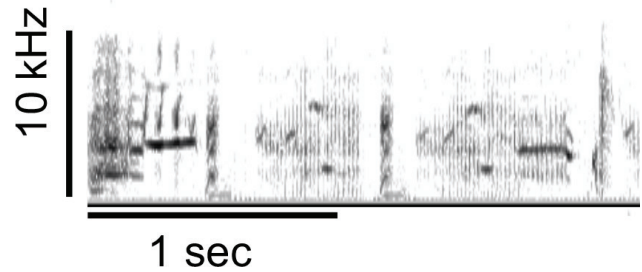
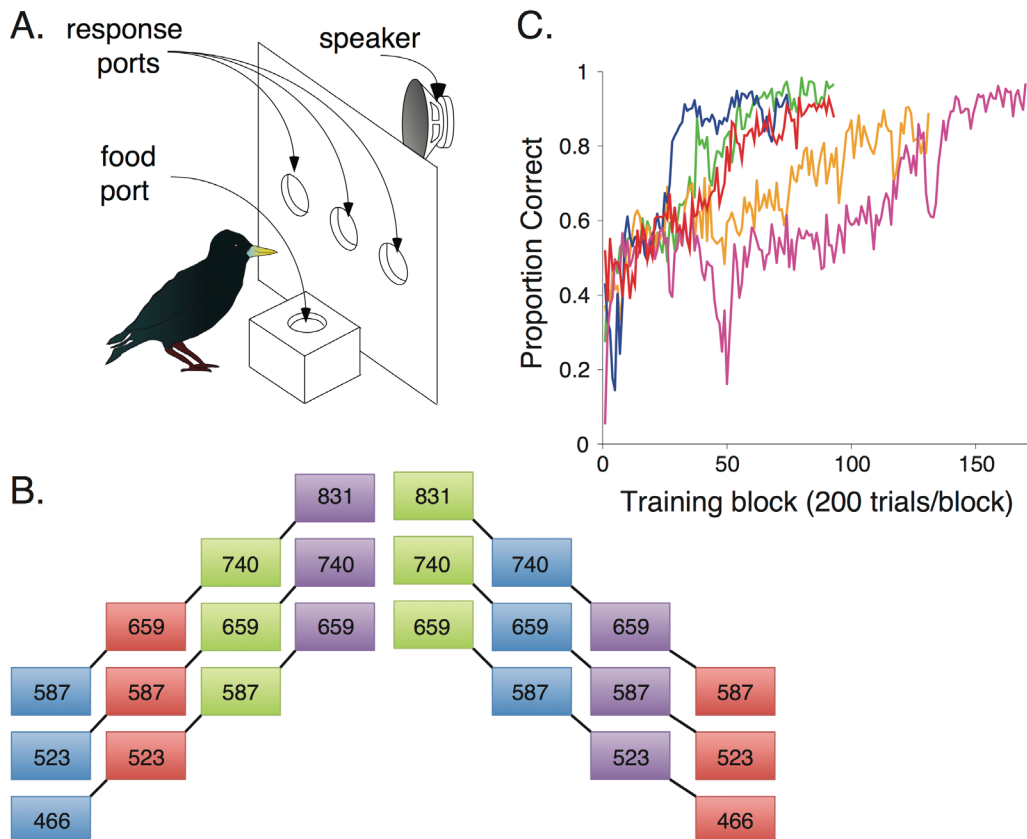


Figure 3. (A) Schematic of the operant panel used for behavioral testing in Bregman et al. (2016). Three response ports, the food port, and playback speaker are labeled. (B) Schematic of the six training stimuli used in Experiment 1. Colored boxes indicate complex harmonic tones, and numbers in each box are the fundamental frequencies of each tone. (These numbers have been rounded to the nearest integer value for display purposes: actual values in Hz [and corresponding Western musical note names] were: 466.16 [Bb4], 523.25 [C5], 578.33 [D5], 659.25 [E5], 739.99 [F#5], 830.61 [G#5]). Colors indicate musical instrument timbre (blue, oboe; red, choir “aah”; green, muted trumpet; purple, synthesizer). Each of the three ascending and three descending tone sequences are connected with black lines. (C) Mean proportion of correct responses for each of the five subjects (one color per subject) over the course of training. From “Songbirds Use Spectral Shape, not Pitch, for Sound Pattern Recognition,” by M. R. Bregman, A. D. Patel, and T. Q. Gentner, 2016, *Proceedings of the National Academy of Sciences*, 113, p. 1667. Copyright 2016 by National Academy of Sciences. Adapted with permission.



rich mix of tonal, noisy, broadband, and narrowband elements (see Figure 2 for an example).

Thus we hypothesized that if starlings were trained to recognize tone sequences with spectrotemporal variation, they would recognize the sequences if transposed. We tested this hypothesis in Experiment 1 of Bregman et al. (2016), in which starlings were trained to discriminate between sequences of complex harmonic tones distinguished by both pitch patterns and spectral patterns (Figure 3). Variation in spectral patterns was achieved by having each tone produced with a distinct musical timbre. As previously noted, spectral structure is an important acoustic attribute contributing to timbre, thus by varying timbre from note to note we varied spectral structure in a controlled way.

Specifically, the birds were trained to peck one key if they heard a sequence of four tones rising in pitch (with tonal timbres in the order O, C, M, S, where O = oboe, C = choir “aah,” M = muted trumpet, S = synthesizer) and another key if they heard a sequence of four tones falling in pitch (with timbres in the order M, O, S, C). Thus the pattern of pitches and timbres provided redundant cues for perceptual discrimination. Once the birds learned this discrimination (which they did readily; cf. Figure 3C), we tested their ability to recognize (i.e., generalize their discrimination to) transposed versions of these sequences, both within and outside the training range.

Contrary to our hypothesis, the birds showed no generalization, even for small transpositions within the training range. This finding seemed consistent with the idea that the birds use absolute pitch (AP) to recognize tone sequences. Indeed, it seemed that their commitment to AP might even be *strengthened* when sequences contain spectrotemporal variation versus when sequences are made from pure tones. Recall that when tested with pure-tone sequences in prior research, starlings recognize sequences transposed within the frequency range of the training stimuli (the frequency range constraint). In contrast, the starlings in our study failed to recognize transpositions that stayed entirely within the training range, for example, an upward shift of just one semitone relative to the lowest ascending–descending sequence pair in Figure 3B.

However, before concluding that the starlings were using AP as the primary cue for recognizing tone sequences in our study, we felt it was necessary to obtain positive evidence that this was the case. We reasoned that if AP was the primary cue used to recognize melodies, then starlings should generalize to sequences matched

in AP to the training sequences but differing in timbre. Thus in Experiment 2 of Bregman et al. (2016), we tested if starlings would generalize their discrimination to a rising versus falling melodic pair identical in AP to one of the training pairs (the lowest ascending–descending pair in Figure 3B) but differing in timbre (made from piano tones). We found that the birds did not generalize to these AP-matched sequences, indicating that that AP was *not* the essential cue for tone sequence recognition.

Evidence for the Use of Spectral Shape, not Pitch, in Starling Tone Sequence Recognition

Experiments 1 and 2 of Bregman et al. (2016) showed that when starlings learn to recognize a tone sequence, they use neither relative pitch (Experiment 1) nor absolute pitch (Experiment 2) as the primary cue for recognition. Another way to put this is that starlings do not recognize a familiar melody if it is transposed in pitch when its sequence of timbres is preserved (Experiment 1), nor (conversely) if its sequence of pitches is preserved but its timbre is changed (Experiment 2). This suggests that pitch and timbre are not good descriptors of the cues used by starlings to recognize tone sequences and/or that these perceptual attributes are not independent for the birds (a possibility also suggested by Hoeschele, Cook, Guillette, Hahn, & Sturdy, 2014).

If starlings do not use pitch for tone sequence recognition, what perceptual cues are they using? We reasoned that they might be using a cue more directly related to the acoustic structure of sounds, namely, the spectral shape of sounds. As noted above, spectral shape is known to be an important cue in speech recognition in humans.

To test the spectral shape hypothesis, we created noise-vocoded (NV) versions of our training sequences, using 16 frequency bands spanning 50–11,000 Hz. (In noise vocoding, the number of frequency bands determines how faithfully the spectral envelope traces the underlying spectral structure: more bands result in more detailed tracing. 16 bands has been shown to result in high intelligibility in speech perception research; Shannon, Fu, & Galvin, 2004). Experiment 3 of Bregman et al. (2016) tested whether starlings would generalize their discrimination of the melodic training sequences (cf. Figure 3B) to NV versions of these sequences. This experiment used a transfer paradigm, whereby the three pairs of ascending versus descending training sequences (which were discriminated with a high degree of accuracy by the starlings) were replaced with their NV counterparts. Success of transfer is indicated by the strength of the initial transfer and by the subsequent acquisition

rate. By both measures, the starlings showed strong transfer to the NV sequences. To directly compare the strength of this transfer to the ability to recognize melodies on the basis of AP, the second part of Experiment 3 tested transfer from the original training sequences to piano-tone versions of these sequences matched in AP to the training sequences. In this case, the starlings showed poor generalization, even though these birds had prior experience with the transfer task (i.e., via the NV experiment).

These results indicated that spectral shape, not pitch, was a key cue for tone sequence recognition by starlings. We know that spectral shape rather than detailed spectral structure (or pitch) was important for recognition because NV preserved the overall spectral structure (i.e., the spectral envelope) while eliminating pitch information. It is important to note, however, that simply retaining spectral envelope is not enough to guarantee tone sequence recognition by starlings. In Experiment 1, the training sequences retained their spectral envelopes when transposed (because the sequences were simply shifted up or down in log frequency), yet the birds did not recognize the transposed tone sequences. This suggests that it is not just spectral envelope that is important for starling tone sequence recognition but “absolute spectral envelope,” that is, the overall pattern of spectral amplitudes *across particular frequency bands*.

Given these findings, we can now reinterpret the results of prior experiments on avian tone sequence recognition that employed pure tones. In pure tones (which have just one frequency), the absolute spectral envelope corresponds directly to pitch, which can lead one to interpret the results of such studies in terms of absolute pitch as a recognition cue. It is only when one uses more spectrally complex sounds that one can dissociate spectral envelope and pitch, and test whether pitch patterns are truly the basis for tone sequence recognition.

Challenges to the Idea of Absolute Spectral Envelope as a Recognition Cue

The idea that starlings use absolute spectral envelope for tone sequence recognition faces challenges from four findings that seem to contradict this idea. First, in Bregman et al. (2012), starlings readily recognized conspecific songs when they were shifted up or down in frequency. Such shifts do not preserve the absolute spectral envelopes of songs, because the original songs are shifted into different frequency bands. Thus we suspect that

absolute spectral envelope is not the only cue starlings use for recognizing conspecific songs. Starling songs are spectrotemporally complex and provide a number of nonspectral cues that might be used for recognition (and that would not be affected by frequency shifting), including amplitude modulation patterns, the timing of syllable onsets, and the timing (and/or serial order) of other acoustic landmarks. Thus the importance of different cues for sound pattern recognition may depend on the stimuli used and the listening task. Further research is needed to determine what cues starlings use to recognize frequency-shifted songs. I suspect (contra our original interpretation in Bregman et al., 2012, but in line with Bregman et al., 2016) that the relevant cues do not concern pitch patterns within the songs.

Second, prior work has shown that songbirds can recognize similar spectral structures at different absolute frequencies (Braaten & Hulse, 1991; cf. Hoeschele, Cook, Guillette, Brooks, & Sturdy, 2012). Such structures would have different absolute spectral envelopes, again challenging the idea of absolute spectral envelope as a recognition cue. However, the distinct spectral structures used by Braaten and Hulse (1991) may have also differed in other perceptual properties, such as degree of consonance or dissonance, which can drive generalization (Hulse, Bernard, & Braaten, 1995).

Third, starlings perceive the missing fundamental of harmonic complexes. This has been shown by training the birds to discriminate between two pure tones and testing their ability to generalize this discrimination to harmonic complexes containing four consecutive higher harmonics of these tones. Starlings show immediate and accurate performance on this generalization task (Cynx & Shapiro, 1986), which implies that they perceive the missing fundamental. For the current purposes, the relevant point is that this generalization occurred even though the absolute spectral envelopes of the training and test stimuli were quite different (e.g., one training stimulus was a 200 Hz pure tone, whereas the associated harmonic complex had frequencies of 800, 1000, 1200, and 1400 Hz). Although this generalization cannot be based on absolute spectral envelope, it is important to note that Cynx and Shapiro (1986) studied the perception of individual tones that lacked any spectrotemporal variation, whereas Bregman et al. (2016) studied the perception of tone sequences with spectrotemporal variation. Thus the use of absolute spectral envelope as a key cue for tone sequence recognition may apply when tones are acoustically complex and are in sequences where spectral structure varies.

Fourth, as noted earlier, when starlings are trained to discriminate ascending from descending pure tone sequences, they can generalize to transposed versions of these sequences within the frequency range of the training stimuli (the frequency range constraint). Such transpositions do not preserve the pattern of absolute spectral envelopes of tone sequences. However, pure tones have no variation in spectral structure from one sound to the next, because each tone consists of single frequency. As noted earlier, when Bregman et al. (2016) used sequences of complex harmonic tones with variation in spectral structure, they found no generalization to transpositions that lay entirely within the training range (including transpositions of just one semitone). This points to the need for further studies that vary the acoustic complexity of tones and the amount of tone-to-tone variation in spectral structure to determine at what point spectral envelope becomes a primary cue for tone sequence recognition. It may be, for example, that pitch plays a role in starling tone sequence recognition if complex harmonic tones are less acoustically complex than real musical instrument sounds, and/or if the spectral structure of complex harmonic tones does not vary from tone to tone. Consistent with this idea, Bregman et al. (2012) found that starlings trained to discriminate short melodic sequences made from piano tones could recognize transpositions of such sequences by one or two semitones. Because these sequences were made entirely of piano tones, there was little variation in spectral structure from tone to tone. Thus pitch may have been given more weight as a cue for tone sequence recognition than in the study of Bregman et al. (2016), where spectral structure showed considerably more variation from tone to tone.

From these considerations it is clear that future work on avian tone sequence recognition should vary the acoustic complexity (and degree of spectral modulation) of complex harmonic tones in order to determine how the hierarchy of cues used by birds for tone sequence recognition depends on stimulus structure. It will also be important to know whether this hierarchy differs between species. As a comparison to starlings, zebra finches (*Taeniopygia guttata*) would be interesting to study because their songs have many harmonically structured elements and because these finches have been studied in terms of their sensitivity to aspects of spectral shape and pitch as recognition cues for natural sounds (e.g., Uno, Maekawa, & Kaneko, 1997; Vignal & Mathevon, 2011).

Having reviewed the background and findings of Bregman et al. (2016), we can now turn to the three questions raised at the opening of this paper.

Why Would Starlings Use Spectral Shape, not Pitch, for Tone Sequence Recognition?

In seeking to understand why starlings use spectral shape rather than pitch to recognize tone sequences, a logical starting point is to think about the roles these two factors play in the bird's own communication system. As just noted, starling song is spectrotemporally complex, and to human ears it often sounds as if more than one pitch is being produced simultaneously (which could reflect the two sides of the avian syrinx producing different sounds). Starling songs also contain frequent inharmonic sounds that do not yield a clear pitch. In other words, unlike in "tonal" birdsongs, which are dominated by just one time-varying frequency, starling songs do not project a clear, unitary pitch sequence. This may be one reason why these birds don't rely on pitch patterns when recognizing tone sequences.

It is important to note that the non-use of pitch for tone sequence recognition by starlings is not due to an inability to perceive pitch. As previously stated, starlings perceive the pitch of harmonic complexes, as shown by research on perception of the missing fundamental. Yet even though they can perceive the pitch of harmonically structured tones, the results of Bregman et al. (2016) show that they do not *use* pitch patterns for tone sequence recognition. In other words, humanlike pitch perception of individual tones (e.g., as recently documented in the common marmoset *Callithrix jacchus* by Song, Osmanski, Guo, & Wang, 2016) does not automatically lead to humanlike processing of tone sequences.

If starlings do not use pitch for tone sequence recognition, why do they rely on spectral shape for this task? One idea is that attending to the spectral shape of sound sequences is beneficial in the animal's own communication system. A recent acoustic study of the full vocal repertoire of another songbird that produces spectrotemporally complex sounds (the zebra finch) revealed that their 10 call types were primarily distinguished by spectral shape (Elie & Theunissen, 2016). These variations in spectral shape are driven by changes in the shape of the vocal tract (Riede, Schilling, & Goller, 2013). Thus starlings may use spectral shape as a primary cue in distinguishing between different conspecific call types. Spectral shape patterns may also be important in starlings'

ability to recognize other individual starlings based on their songs (Gentner & Hulse, 1998).

The exact spectral structure of a sound is a very detailed acoustic property. The results of Bregman et al.'s (2016) Experiment 3, and of the study of Elie and Theunissen (2016), suggest that the full details of spectral structure are not necessary for sound pattern recognition/discrimination by songbirds. Instead, the spectral envelope (which conveys the *overall* distribution of energy across frequency bands) appears to be sufficient. This is an interesting parallel to speech perception in humans (Shannon, 2016).

Returning to pitch perception, if starlings perceive the pitch of sounds but don't use pitch patterns for sound pattern recognition, what function does pitch perception serve for them? A possible answer is suggested by the findings of Elie and Theunissen (2016). They examined a large number of acoustic features to see which best discriminated between different types of zebra finch vocalizations. As noted previously, spectral shape was a key feature, but an important secondary feature was pitch saliency, which distinguishes noisy sounds from tonal or harmonic sounds. Thus, pitch perception may play an important role for starlings as a cue in distinguishing call types based on pitch saliency.

Why Do Humans Use Pitch for Tone Sequence Recognition?

Why do humans (unlike starlings) gravitate to pitch as a key attribute when recognizing tone sequences? This tendency may have its roots in human tendency to modulate pitch independently of spectral shape for communicative purposes. For example, human singing often contains phrases in which the same pitch pattern is used with different words (which are largely cued by spectral shape patterns), as in the verses of certain popular songs, where the same melodic patterns are paired with different words. Conversely, in speech it is not uncommon to hear the same sequence of words spoken with different pitch patterns (e.g., "It's your birthday!" vs. "It's your birthday?"). Thus pitch and spectral shape exhibit a significant degree of "informational independence" in human auditory communication. This may be why humans automatically perceptually separate pitch and spectral shape when processing tone sequences and (because spectral shape patterns in tone sequences convey no lexical information) attend to pitch as a key feature for tone sequence recognition.

Is Using Pitch to Recognize Tone Sequences Evolutionarily Ancient or Recent? The Informational Independence Hypothesis

Starling's use of spectral shape rather than pitch to recognize tone sequences raises the question of how widespread this tendency is among nonhuman animals. If using pitch to recognize tone sequences reflects ancient, widespread brain mechanisms of sound pattern recognition, then this trait should be common among animals, and starlings should be an exception to the rule. If, on the other hand, tone sequence recognition based on pitch is a recent evolutionary trait, then humans may share this trait with few other species.

I suspect that the use of pitch to recognize tone sequences may be a rare trait, present only in animals with certain types of communication systems. These are systems in which pitch and spectral shape patterns are modulated independently in acoustic sequences to convey distinct types of information (cf. the previous section). I call this the informational independence hypothesis. Underlying this hypothesis is the idea that the perceptual separation of pitch and spectral shape in sound sequences reflects auditory neural specializations driven by specific communicative needs. In this light, it is interesting to note that human auditory cortical brain regions involved in pitch perception appear to be partly distinct from those involved in the analysis of spectral shape (Norman-Haignere, Kanwisher, & McDermott, 2013; Warren, Jennings, & Griffiths, 2005).

What other species exhibit informational independence of pitch and spectral shape in their communication systems? A signature of such systems is the use of similar pitch patterns with different sequences of spectral shapes and/or the use of different pitch patterns with similar sequences of spectral shapes. From a comparative perspective, such systems may be quite rare in animal communication. To be sure, many bird and mammal species modulate spectral shape and pitch for communicative purposes (e.g., Elie & Theunissen, 2016; Fitch, 2000; Pisanski, Cartei, McGettigan, Raine, & Reby, 2016), for example, to distinguish affiliative from agonistic calls (Morton, 1977). However, this does not prove that they are modulated independently in sequences to convey distinct types of information. In searching for other species that exhibit informational independence of spectral shape and pitch, it may be useful to examine animals that produce communicative sequences employing harmonic sounds with clear pitch,

and in which these sounds can be ordered in different ways for communicative purposes. In this regard, Campbell's monkeys (*Cercopithecus campbelli*) and Bengalese finches (*Lonchura striata domestica*) would be promising species to examine (Abe & Watanabe, 2011; Okanoya, 2004; Ouattara, Lemasson, & Zuberbühler, 2009; Schlenker et al., 2014). Common marmosets also merit study because they produce numerous tonal calls (Pistorio, Vintch, & Wang, 2006) and are well studied in the laboratory (e.g., Miller, Mandel, & Wang, 2010).

The informational independence hypothesis makes a testable prediction: If an animal does not exhibit informational independence of pitch and spectral shape in its own communicative sequences, it will not use pitch to recognize tone sequences. Note that this prediction pertains to the recognition of sequences of complex harmonic tones that vary in spectral structure and could be tested using the stimuli and methods from the noise-vocoding experiment of Bregman et al. (2016). Thus the hypothesis makes the counterintuitive prediction that birds that sing "tonal" songs dominated by just one time-varying frequency, such as the Common Yellowthroat (*Geothlypis trichas*), will not use pitch to recognize tone sequences. Such bird songs sound very musical to human ears, yet because they largely consist of a single time-varying frequency, spectral shape, and pitch are not dissociable and thus are not modulated independently during song production. (In searching for species to study in order to test this prediction, it will be important to examine a species' songs as well as its calls. One would want to test animals where there is no evidence of informational independence of pitch and spectral shape in either the songs or the calls.)

If the informational independence hypothesis is supported by future work, an interesting question will be the extent to which other species can learn to use pitch to recognize tone sequences. Ferrets (*Mustela putorius furo*) would be an interesting species to study in this regard. Neural research suggests that ferret auditory cortical neurons sensitive to pitch are also typically sensitive to timbre (Bizley, Walker, Silverman, King, & Schnupp, 2009), an auditory attribute in which spectral shape plays a key role (Caclin, McAdams, Smith, & Winsberg, 2005). Thus pitch and spectral shape may not be well separated in the brains of these animals. If these animals (like starlings) tend to use spectral shape for tone sequence recognition, then one could address the learning question just alluded to. Specifically, if young ferrets were raised in an acoustic environment where pitch and spectral shape patterns were varied independently in sequences

of complex harmonic tones to convey distinct "meanings" (e.g., related to food availability or other meaningful environmental variables), would the animals (as adults) use pitch to recognize tone sequences? If so, this would suggest that the human tendency to recognize tone sequences based on pitch patterns need not reflect evolved neural specializations and could emerge through experience-dependent neural plasticity (Fritz, Shamma, Elhilali, & Klein, 2003) driven by the informational independence of pitch and spectral shape in human auditory communication.

Conclusion

Cross-species studies of music perception have recently begun to grow in number and scope. Such studies provide an empirical approach to studying the evolutionary history of music cognition (Patel, in press). In this article I have focused on a major difference in how songbirds (European starlings) versus humans recognize sequences of complex harmonic tones that vary in spectral structure. For humans, the pitch pattern of the tones is a key cue for recognition, whereas for starlings the pattern of spectral shapes, not pitch, is key for recognition. I suggest that this difference has its roots in the way pitch and spectral shape are used in the natural communication systems of starlings versus humans and propose that recognizing such tone sequences based on pitch patterns is likely to be an unusual trait, reflecting auditory neural specializations that are rare in cognitive evolution.

References

- Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience*, *14*, 1067–1074. doi:10.1038/nn.2869
- Bernard, D. J., & Hulse, S. H. (1992). Transfer of serial stimulus relations by European starlings (*Sturnus vulgaris*): Loudness. *Journal of Experimental Psychology: Animal Behavior Processes*, *18*, 323–344. doi:10.1037/0097-7403.18.4.323
- Bizley, J. K., Walker, K. M., Silverman, B. W., King, A. J., & Schnupp, J. W. (2009). Interdependent encoding of pitch, timbre, and spatial location in auditory cortex. *The Journal of Neuroscience*, *29*, 2064–2075. doi:10.1523/JNEUROSCI.4755-08.2009

- Braaten, R. F., & Hulse, S. H. (1991). A songbird, the European starling (*Sturnus vulgaris*), shows perceptual constancy for acoustic spectral structure. *Journal of Comparative Psychology*, *105*, 222–231. doi:10.1037/0735-7036.105.3.222
- Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2012). Stimulus-dependent flexibility in non-human auditory pitch processing. *Cognition*, *122*, 51–60. doi:10.1016/j.cognition.2011.08.008
- Bregman, M. R., Patel, A. D., & Gentner, T. Q. (2016). Songbirds use spectral shape, not pitch, for sound pattern recognition. *Proceedings of the National Academy of Sciences*, *113*, 1666–1671. doi:10.1073/pnas.1515380113
- Caclin, A., McAdams, S., Smith, B. K., & Winsberg, S. (2005). Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones. *The Journal of the Acoustical Society of America*, *118*, 471–482. doi:10.1121/1.1929229
- Calabrese, A., & Woolley, S. M. (2015). Coding principles of the canonical cortical microcircuit in the avian brain. *Proceedings of the National Academy of Sciences USA*, *112*, 3517–3522. doi:10.1073/pnas.1408545112
- Carr, C. E. (1992). Evolution of the central auditory system in reptiles and birds. In D. B. Webster, R. R. Fay, & A. N. Popper (Eds.), *The evolutionary biology of hearing* (pp. 511–544). New York, NY: Springer-Verlag. doi:10.1007/978-1-4612-2784-7_32
- Creel, S. C., & Tumlin, M. A. (2012). Online recognition of music is influenced by relative and absolute pitch information. *Cognitive Science*, *36*, 224–260. doi:10.1111/j.1551-6709.2011.01206.x
- Crespo-Bojorque, P., & Toro, J.M. (2016). Processing advantages for consonance: A comparison between rats (*Rattus Norvegicus*) and humans (*Homo Sapiens*). *Journal of Comparative Psychology*, *130*, 97–108. doi:10.1037/com0000027
- Cynx, J., & Shapiro, M. (1986). Perception of missing fundamental by a species of songbird (*Sturnus vulgaris*). *Journal of Comparative Psychology*, *100*, 356–360. doi:10.1037/0735-7036.100.4.356
- D'Amato, M. R. (1988). A search for tonal pattern perception in cebus monkeys: Why monkeys can't hum a tune. *Music Perception: An Interdisciplinary Journal*, *5*, 453–480. doi:10.2307/40285410
- Darwin, C. (1871). *The descent of man and selection in relation to sex*. London, UK: John Murray.
- Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General*, *134*, 222–241. doi:10.1037/0096-3445.134.2.222
- de Waal, F. (2016). *Are we smart enough to know how smart animals are?* New York, NY: Norton.
- Dooling, R. J., Okanoya, K., Downing, J., & Hulse, S. (1986). Hearing in the starling (*Sturnus vulgaris*): Absolute thresholds and critical ratios. *Bulletin of the Psychonomic Society*, *24*, 462–464. doi:10.3758/BF03330584
- Doupe, A. J. (1997). Song-and order-selective neurons in the songbird anterior forebrain and their emergence during vocal development. *The Journal of Neuroscience*, *17*, 1147–1167.
- Elie, J. E., & Theunissen, F. E. (2016). The vocal repertoire of the domesticated zebra finch: A data-driven approach to decipher the information-bearing acoustic features of communication signals. *Animal Cognition*, *19*, 285–315. doi:10.1007/s10071-015-0933-6
- Fitch, W. T. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*, 258–267. doi:10.1016/S1364-6613(00)01494-7
- Fitch, W. T. (2015). Four principles of bio-musicology. *Philosophical Transactions of the Royal Society B*, *370*(1664), 20140091. doi:10.1098/rstb.2014.0091
- Foster, N. E., & Zatorre, R. J. (2010). A role for the intraparietal sulcus in transforming musical pitch information. *Cerebral Cortex*, *20*, 1350–1359. doi:10.1093/cercor/bhp199

- Fritz, J., Shamma, S., Elhilali, M., & Klein, D. (2003). Rapid task-related plasticity of spectrotemporal receptive fields in primary auditory cortex. *Nature Neuroscience*, 6, 1216–1223. doi:10.1038/nn1141
- Gentner, T. Q., & Hulse, S. H. (1998). Perceptual mechanisms for individual vocal recognition in European starlings, *Sturnus vulgaris*. *Animal Behaviour*, 56, 579–594. doi:10.1006/anbe.1998.0810
- Gentner, T. Q., & Hulse, S. H. (2000). Female European starling preference and choice for variation in conspecific male song. *Animal Behaviour*, 59, 443–458. doi:10.1006/anbe.1999.1313
- Hindmarsh, A. M. (1984). Vocal mimicry in starlings. *Behaviour*, 90, 302–324. doi:10.1163/156853984X00182
- Hoeschele, M., Cook, R. G., Guillette, L. M., Brooks, D. I., & Sturdy, C. B. (2012). Black-capped chickadee (*Poecile atricapillus*) and human (*Homo sapiens*) chord discrimination. *Journal of Comparative Psychology*, 126, 57–67. doi:10.1037/a0024627
- Hoeschele, M., Cook, R. G., Guillette, L. M., Hahn, A. H., & Sturdy, C. B. (2014). Timbre influences chord discrimination in black-capped chickadees (*Poecile atricapillus*) but not humans (*Homo sapiens*). *Journal of Comparative Psychology*, 128, 387–401. doi:10.1037/a0037159
- Hoeschele, M., Guillette, L. M., & Sturdy, C. B. (2012). Biological relevance of acoustic signal affects discrimination performance in a songbird. *Animal Cognition*, 15, 677–688. doi:10.1007/s10071-012-0496-8
- Hoeschele, M., Merchant, H., Kikuchi, Y., Hattori, Y., & ten Cate, C. (2015). Searching for the origins of musicality across species. *Philosophical Transactions of the Royal Society B*, 370(1664), 20140094. doi:10.1098/rstb.2014.0094
- Honing, H., ten Cate, C., Peretz, I., & Trehub, S. E. (2015). Without it no music: Cognition, biology and evolution of musicality. *Philosophical Transactions of the Royal Society B*, 370(1664), 20140088. doi:10.1098/rstb.2014.0088
- Hulse, S. H., Bernard, D. J., & Braaten, R. F. (1995). Auditory discrimination of chord-based spectral structures by European starlings (*Sturnus vulgaris*). *Journal of Experimental Psychology: General*, 124, 409–423. doi:10.1037/0096-3445.124.4.409
- 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, 38–54. doi:10.1037/0096-3445.113.1.38
- Hulse, S. H., Humpal, J., & Cynx, J. (1984). Discrimination and generalization of rhythmic and arrhythmic sound patterns by European starlings (*Sturnus vulgaris*). *Music Perception: An Interdisciplinary Journal*, 1, 442–464. doi:10.2307/40285272
- Hulse, S. H., Page, S. C., & Braaten, R. F. (1990). Frequency range size and the frequency range constraint in auditory perception by European starlings (*Sturnus vulgaris*). *Animal Learning & Behavior*, 18, 238–245. doi:10.3758/BF03205281
- Hulse, S. H., Takeuchi, A. H., & Braaten, R. F. (1992). Perceptual invariances in the comparative psychology of music. *Music Perception: An Interdisciplinary Journal*, 10, 151–184. doi:10.2307/40285605
- Karten, H. (2013). Neocortical evolution: Neuronal circuits arise independently of lamination. *Current Biology*, 23, R12–R15. doi:10.1016/j.cub.2012.11.013
- Klump, G. M., Langemann, U., & Gleich, O. (2000). The European starling as a model for understanding perceptual mechanisms. In G. A. Manley, H. Oeckinghaus, M. Koessler, G. M. Klump, & H. Fastl (Eds.), *Auditory worlds: Sensory analysis and perception in animals and man* (pp. 193–211). Weinheim, Germany: Verlag Chemie.
- Ladd, D. R., Turnbull, R., Browne, C., Caldwell-Harris, C., Ganushchak, Swoboda, K., ... Dedi, D. (2013). Patterns of individual differences in the perception of missing-fundamental tones. *Journal of Experimental Psychology: Human Perception and Performance*, 39, 1386–1397. doi:10.1037/a0031261

- Levitin, D. J. (1994). Absolute memory for musical pitch: Evidence from the production of learned melodies. *Perception & Psychophysics*, *56*, 414–423. doi:10.3758/BF03206733
- Levitin, D. J., & Rogers, S. E. (2005). Absolute pitch: Perception, coding, and controversies. *Trends in Cognitive Sciences*, *9*, 26–33. doi:10.1016/j.tics.2004.11.007
- Marler, P. R., & Slabbekoorn, H. (2004). *Nature's music: The science of birdsong*. San Diego, CA: Academic Press. doi:10.1016/B978-012473070-0/50000-1
- McAdams, S. J. (2013). Musical timbre perception. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 35–67). London, UK: Academic Press/Elsevier. doi:10.1016/B978-0-12-381460-9.00002-X
- Miller, C. T., Mandel, K., & Wang, X. (2010). The communicative content of the common marmoset phee call during antiphonal calling. *American Journal of Primatology*, *72*, 974–980. doi:10.1002/ajp.20854
- Morton, E. S. (1977). On the occurrence and significance of motivation-structural rules in some bird and mammal sounds. *American Naturalist*, *111*, 855–869. doi:10.1086/283219
- Norman-Haignere, S., Kanwisher, N., & McDermott, J. H. (2013). Cortical pitch regions in humans respond primarily to resolved harmonics and are located in specific tonotopic regions of anterior auditory cortex. *The Journal of Neuroscience*, *33*, 19451–19469. doi:10.1523/JNEUROSCI.2880-13.2013
- Okanoya, K. (2004). The Bengalese finch: A window on the behavioral neurobiology of birdsong syntax. *Annals of the New York Academy of Sciences*, *1016*, 724–735. doi:10.1196/annals.1298.026
- Ouattara, K., Lemasson, A., & Zuberbühler, K. (2009). Campbell's monkeys concatenate vocalizations into context-specific call sequences. *Proceedings of the National Academy of Sciences*, *106*, 22026–22031. doi:10.1073/pnas.0908118106
- Oxenham, A. J. (2013). The perception of musical tones. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 1–33). London, UK: Academic Press/Elsevier. doi:10.1016/B978-0-12-381460-9.00001-8
- Patel, A. D. (2008). *Music, language, and the brain*. New York, NY: Oxford University Press.
- Patel, A. D. (2014). The evolutionary biology of musical rhythm: Was Darwin wrong? *PLoS Biology*, *12*(3), e1001821. doi:10.1371/journal.pbio.1001821
- Patel, A. D. (in press). Evolutionary music cognition: Cross-species studies. In D. Levitin & J. Rentfrow (Eds.), *Foundations in music psychology: Theory and research*. Cambridge, MA: MIT Press.
- Patel, A. D., & Demorest, S. (2013). Comparative music cognition: Cross-species and cross-cultural studies. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 647–681). London, UK: Academic Press/Elsevier. doi:10.1016/B978-0-12-381460-9.00016-X
- Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice modulation: A window into the origins of human vocal control? *Trends in Cognitive Sciences*, *20*, 304–318. doi:10.1016/j.tics.2016.01.002
- Pistorio, A. L., Vintch, B., & Wang, X. (2006). Acoustic analysis of vocal development in a New World primate, the common marmoset (*Callithrix jacchus*). *The Journal of the Acoustical Society of America*, *120*, 1655–1670. doi:10.1121/1.2225899
- Plantinga, J., & Trainor, L. J. (2005). Memory for melody: Infants use a relative pitch code. *Cognition*, *98*, 1–11. doi:10.1016/j.cognition.2004.09.008
- Riede, T., Schilling, N., & Goller, F. (2013). The acoustic effect of vocal tract adjustments in zebra finches. *Journal of Comparative Physiology A*, *199*, 57–69. doi:10.1007/s00359-012-0768-4
- Rock, I., & Palmer, S. (1990). The legacy of Gestalt psychology. *Scientific American*, *263*, 84–90. doi:10.1038/scientificamerican1290-84

- Schlenker, P., Chemla, E., Arnold, K., Lemasson, A., Ouattara, K., Keenan, S., ... Zuberbühler, K. (2014). Monkey semantics: Two “dialects” of Campbell’s monkey alarm calls. *Linguistics and Philosophy* 37, 439–501. doi:10.1007/s10988-014-9155-7
- Shannon, R. V. (2016). Is birdsong more like speech or music? *Trends in Cognitive Sciences*, 20, 245–247. doi:10.1016/j.tics.2016.02.004
- Shannon, R. V., Fu, Q. J., & Galvin, J. (2004). The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Oto-Laryngologica*, 124, 50–54. doi:10.1080/03655230410017562
- Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, 270, 303–304. doi:10.1126/science.270.5234.303
- Song, X., Osmanski, M. S., Guo, Y., & Wang, X. (2016). Complex pitch perception mechanisms are shared by humans and a New World monkey. *Proceedings of the National Academy of Sciences*, 113, 781–786. doi:10.1073/pnas.1516120113
- Stevens, K. N. (2000). *Acoustic phonetics*. Cambridge, MA: MIT Press.
- Sundberg, J. (1987). *The science of the singing voice*. DeKalb: University of Illinois Press.
- Uno, H., Maekawa, M., & Kaneko, H. (1997). Strategies for harmonic structure discrimination by zebra finches. *Behavioural Brain Research*, 89, 225–228. doi:10.1016/S0166-4328(97)00064-8
- Vignal, C., & Mathevon, N. (2011). Effect of acoustic cue modifications on evoked vocal response to calls in zebra finches (*Taeniopygia guttata*). *Journal of Comparative Psychology*, 125, 150–161. doi:10.1037/a0020865
- Warren, J. D., Jennings, A. R., & Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage*, 24, 1052–1057. doi:10.1016/j.neuroimage.2004.10.031
- Weisman, R. G., Njegovan, M. G., Williams, M. T., Cohen, J. S., & Sturdy, C. B. (2004). A behavior analysis of absolute pitch: Sex, experience, and species. *Behavioural Processes*, 66, 289–307. doi:10.1016/j.beproc.2004.03.010
- Weisman, R. G., Williams, M. T., Cohen, J. S., Njegovan, M. G., & Sturdy, C. B. (2006). The comparative psychology of absolute pitch. In T. Zentall & E. Wasserman (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 71–86). Oxford, UK: Oxford University Press.
- 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. doi:10.1037/0096-3445.129.3.291