Songbirds use spectral shape, not pitch, for sound pattern recognition

Micah R. Bregman[a], **Aniruddh D. Patel**[b], **and Timothy Q. Gentner**[c,d,e,f,1]

[a]Department of Cognitive Science, University of California, San Diego, La Jolla, CA 92093; [b]Department of Psychology, Tufts University, Medford, MA 02155; [c]Department of Psychology, University of California, San Diego, La Jolla, CA 92093; [d]Section of Neurobiology, University of California, San Diego, La Jolla, CA 92093; [e]Neurosciences Graduate Program, University of California, San Diego, La Jolla, CA 92093; and [f]Kavli Institute for Brain and Mind, La Jolla, CA 92093

Edited by Dale Purves, Duke University, Durham, NC, and approved December 17, 2015 (received for review August 3, 2015)

Humans easily recognize "transposed" musical melodies shifted up or down in log frequency. Surprisingly, songbirds seem to lack this capacity, although they can learn to recognize human melodies and use complex acoustic sequences for communication. Decades of research have led to the widespread belief that songbirds, unlike humans, are strongly biased to use absolute pitch (AP) in melody recognition. This work relies almost exclusively on acoustically simple stimuli that may belie sensitivities to more complex spectral features. Here, we investigate melody recognition in a species of songbird, the European Starling (*Sturnus vulgaris*), using tone sequences that vary in both pitch and timbre. We find that small manipulations altering either pitch or timbre independently can drive melody recognition to chance, suggesting that both percepts are poor descriptors of the perceptual cues used by birds for this task. Instead we show that melody recognition can generalize even in the absence of pitch, as long as the spectral shapes of the constituent tones are preserved. These results challenge conventional views regarding the use of pitch cues in nonhuman auditory sequence recognition.

absolute pitch | songbirds | comparative cognition | pitch processing | pattern perception

Songbirds are an important animal model for studying the sensorimotor mechanisms of vocal learning and the processing of learned, complex sound sequences (1–4). Although birds lack the six-layered mammalian neocortex (5), the avian auditory system follows the general vertebrate plan (6), including telencephalic circuits organized in a radial columnar pattern that are anatomically (7, 8), genetically (9), and functionally (10) analogous to the mammalian auditory cortical microcircuit. Likewise, songbirds and humans share evolutionarily convergent features of their vocal production biomechanics (11, 12) and of brain circuitry that underlies the rare trait of vocal learning (13, 14). Many aspects of auditory processing are also similar in songbirds and humans. For example, humans and European Starlings (*Sturnus vulgaris*) have similar frequency sensitivity thresholds and auditory filter widths (15–17), perceive the pitch of the missing fundamental (18), and parse multiple pure-tone sequences (separated in frequency) into separate auditory streams (19, 20). At higher levels, the "musical" nature of birdsong has long been appreciated by humans (21), and some songbirds can readily learn to discriminate and imitate human melodic sequences (22–24).

Given these similarities, it is surprising to find a major difference in how humans and songbirds perceive sequences of tones. Humans readily recognize tone sequences that are shifted up or down in log frequency (e.g., the "Happy Birthday" tune played on a piccolo or a tuba), because the pattern of relative pitches (the pitch interval sequence) is maintained. This ability appears effortless to humans: It is present in infancy and is a universal of human music cognition (25–27). In fact, the human ability to use relationships between acoustic cues to recognize sound sequences appears to extend beyond pitch, including loudness and perceptual brightness (28). In contrast, multiple studies over the past three decades indicate that songbirds lack relational pitch processing for tone sequences (22, 29, 30; but see refs. 31, 32). Although songbirds can easily learn to discriminate between sequences of several tones (say, ascending vs. descending sequences of four pitches or between the opening phrases of two different human melodies), even modest generalization to frequency-shifted versions of the same relative pitch patterns requires extensive training (33), and this generalization is restricted to narrow frequency ranges near the training tones (34). However, songbirds can do relational processing for certain aspects of tone sequence structure. Starlings, for example, can learn to discriminate between tone sequences of different tempi and can generalize this discrimination to novel sequences at double the training tempo (35). They can also learn to discriminate between tone sequences that increase versus decrease in loudness and generalize this discrimination to different loudness ranges (36).

Past work has characterized the difference in how humans and songbirds recognize transposed pitch sequences in terms of a reliance on relative versus absolute pitch (AP) in tone sequence perception. Although most humans rely primarily on relative pitch for recognizing tone sequences, songbirds are thought to exhibit a strong bias for relying on AP cues in recognizing tone patterns (37). Here AP does not refer to the human ability to assign a note name or pitch chroma to a tone, such as "G sharp," but the more general ability to recognize tones on the basis of their AP height. This has been amply demonstrated in songbirds (38, 39). (For pure tones, AP height corresponds to frequency, whereas for complex harmonic tones, it corresponds to fundamental frequency.) However, the view that songbirds gravitate to AP cues in recognizing tone sequences is based on studies using fairly simple acoustic stimuli, such as pure tones or harmonic tones that vary in pitch but have a fairly stable spectral shape over the course of the sequence (e.g., sequences of piano tones). More natural sound sequences, including numerous animal

Significance

Past work characterizes songbirds as having a strong bias to rely on absolute pitch for the recognition of tone sequences. In a series of behavioral experiments, we find that the human percepts of both pitch and timbre are poor descriptions of the perceptual cues used by birds for melody recognition. We suggest instead that auditory sequence recognition in some species reflects more direct perception of acoustic spectral shape. Signals that preserve this shape, even in the absence of pitch, allow for generalization of learned patterns.

Author contributions: M.R.B., A.D.P., and T.Q.G. designed research; M.R.B. performed research; M.R.B. and T.Q.G. analyzed data; and M.R.B., A.D.P., and T.Q.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence should be addressed. Email: tgentner@ucsd.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1515380113/-/DCSupplemental.

vocalizations, human speech and song, and multi-instrument music, vary in both pitch and spectral shape over time.

The question of whether the AP bias demonstrated for songbirds with acoustically simple tone sequences holds for sequences that also evolve spectrally over time is particularly salient given the recent finding that starlings are able to recognize frequency-shifted versions of conspecific songs, including songs shifted outside of the frequency range of the training songs (22). Starling songs are spectrotemporally complex, with salient changes in spectral shape over time, and include narrow-band whistles, harmonic warbles, and broader band bursts and rattles that vary in their strength of periodic pitch cues (40). Thus, it is possible that generalization across frequency-shifted songs reflects the birds' ability to detect patterns of spectrotemporal change over time, independent of absolute frequency.

In humans, spectral structure is a critical element used in speech perception and plays an important role in the percept of timbre (41). Starlings are also able to recognize harmonic tone complexes based on spectral structure despite changes in the absolute frequencies of the spectral components (42). Here we investigate how songbirds perceive tone sequences that systematically vary over time in both pitch and timbre. We find that neither pitch nor timbre alone can provide sufficient information to permit accurate tone sequence generalization. Surprisingly, however, generalization is strong for acoustic manipulations that preserve the temporal pattern of spectral shapes and remove pitch (rather than shift it). These results suggest that the absolute spectral envelope (i.e., the overall pattern of spectral amplitudes across particular frequency bands), rather than AP, may be the salient cue for songbirds recognizing sequences of sounds that have both pitch and spectral variation.

Results

To investigate perception of tone sequences that systematically vary in both pitch and timbre, we trained starlings to recognize short four-tone sequences that either ascended or descended in pitch, where each tone had a distinct spectral profile corresponding to a different musical instrument sound (Fig. 1*B*, Fig. S1, Audio Files S1–S3, and *Materials and Methods*). In this training set of six tone sequences, starlings were required to discriminate ascending from descending tone patterns. We analyzed behavioral performance during training by measuring the proportion of correct responses in each block of 200 trials. All five subjects eventually achieved very accurate and stable recognition of ascending versus descending sequences (Fig. 1*C*). Over the last 200 training trials, the mean percentage of correct responses was 91.75% (range, 87.7–96.8%). Likewise, all subjects maintained performance at or above 90% correct in at least two consecutive 200-trial blocks, after 60–142 training blocks (μ = 91.8 blocks). Recognition accuracy measured in the final block of training was uncorrelated with total number of training blocks (r = 0.48, P = 0.41), indicating that all subjects had attained asymptotic proficiency.

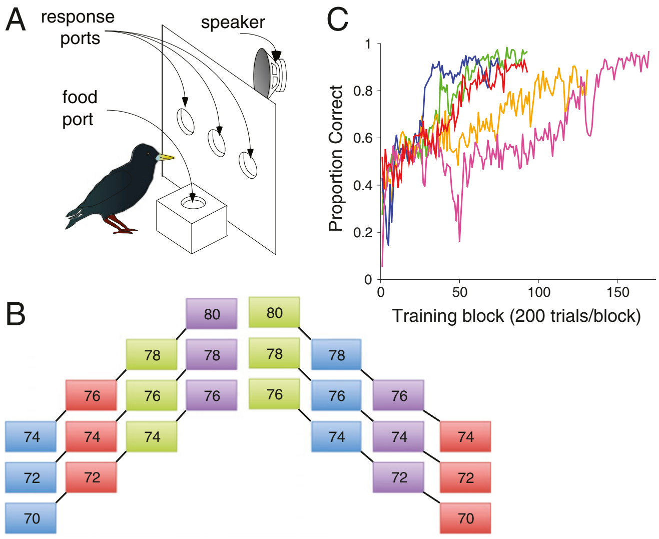

Fig. 1. (*A*) Schematic of the operant panel used for behavioral testing. Three response ports, the food port, and playback speaker are labeled. (*B*) Schematic of the six training stimuli used in experiment 1. Numbers in each box refer to the MIDI note number (e.g., 70, Bb4, 466.16 Hz; 72, C5, 523.25 Hz, etc.; see *Materials and Methods*), and color indicates the instrument timbre used (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 (±SE) for each of the five subjects (one color per subject) over the course of training.

Sequence Recognition at Novel Pitch Levels. Once subjects had reached asymptotic performance on discriminating ascending from descending pitch sequences in the initial training set, we tested their ability to recognize novel stimuli that preserved the relative pitch and timbral pattern of each training sequence (*Materials and Methods*, Fig. S1, and Audio Files S4 and S5). We hypothesized that the distinct patterns of changing spectral shape available in the training stimuli would enable recognition of the tone sequences even when transposed to novel pitch levels. Counter to this hypothesis, although the subjects' performance on training stimuli remained very high during test sessions (range, 87.2–97.0% correct; mean = 93.7% correct), very small shifts in the pitch of test stimuli dramatically reduced recognition performance (Fig. 2*A*). Average performance for each subject on sequences composed of novel pitches ranged from 45.4–58.4% correct (mean = 50.2% correct), which was well below performance on the training sequences and not statistically different from chance based on binomial tests. Even very small pitch changes that placed the test stimuli entirely within the range of the training stimuli rendered the sequence unrecognizable. Specifically, two interleaved stimuli (starting at B4 and C#5, corresponding to shift amounts of one and three semitones; Fig. 2*A*) were one semitone below and one semitone above pairs of training stimuli, respectively, yet recognition of these stimuli was at chance (mean of interleaved stimuli = 49.2% correct) and was significantly below recognition performance on the training sequences interleaved during test sessions [Wilcoxon rank-sum Z (25) = –4.13, P < 0.0001; *Materials and Methods*].

These initial results indicate that starlings do not recognize frequency-shifted versions of spectrotemporally complex tone sequences. That is, although they readily learned to discriminate ascending from descending pitch sequences that were also distinguished by their pattern of spectral shape variation over time, an upward or downward frequency shift of even a single semitone severely disrupted discrimination. Thus, the birds do not appear to take advantage of the redundant pattern of spectral information in the tone sequences to facilitate recognition of transposed sequences. At first glance, these results appear to provide strong support for the prevailing view that songbirds are biased toward using AP cues in tone sequence recognition. Indeed, the results suggest that the bias is particularly strong when tone sequences have spectral shape variation over time (22), even when that spectral shape variation alone could be used to discriminate the sequences.

Sequence Recognition with Novel Timbre. Importantly, confirmation that AP really is the primary cue for tone sequence recognition requires evidence of positive generalization. That is, if our birds are using AP, then they should readily recognize ascending and descending tone sequences that match the pitch of the training stimuli but not their spectral shape. To test this idea, we

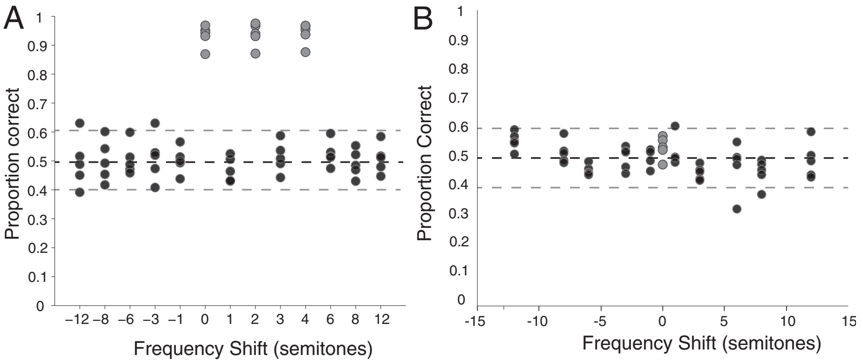

Fig. 2. Pitch and timbre results. (*A*) Mean proportion of correct responses to tone sequences with novel pitches but familiar timbre. The *x* axis shows pitch shift (in semitones) relative to the lowest frequency training stimulus. Gray circles show recognition of the original training sequences during test sessions. Each point shows the data for one subject. (*B*) Performance as in *A* but for sequences composed of sounds with a novel (piano) timbre at the training pitches (gray) or shifted in pitch (black). Dashed lines indicate Bonferroni-corrected 95% CI around chance (50%).

also presented (during the same test sessions already described) sequences composed of tones with a novel piano timbre (Fig. S1 and Audio Files S6–S8) at the same pitch as two of the training sequences (the lowest pitch ascending–descending pair in Fig. 1*B*). If subjects use AP, recognition should generalize to these test sequences. This was not the case. As with alterations in pitch, altering the timbre of the training sequences reduced recognition performance to chance (mean = 53.5% correct; range, 47.7% correct to 57.4% correct; Fig. 2*B*, gray circles). For all subjects, performance was well below recognition of the training sequences (mean = 93.7% correct; binomial test adjusted for multiple comparisons) and within the 95% confidence interval (CI) around chance (~41–59% correct, based on mean of 170.3 trials, corrected for multiple comparisons). Thus, changing the timbre of the tone sequences drove recognition to chance, even when the AP of the training sounds was preserved. Given these initial results, it is not surprising that performance was at chance for tone sequences with novel (piano) timbre at novel pitch levels shifted by ±1, ±3, ±6, ±8, and ±12 semitones relative to the training stimuli (Fig. 2*B*, black circles). Our observation that altering timbre while holding AP constant drives recognition to chance (Fig. 2*B*) challenges the broadly held view that starlings use AP to recognize tone sequences.

Recognition of Noise-Vocoded Sequences. The preceding results demonstrate that independent manipulations of either pitch height or timbre severely disrupt tone sequence recognition (Fig. 2 *A* and *B*). One interpretation of these results is that the percepts of both pitch and timbre provide relatively poor descriptions of the perceptual cues available to starlings. We reasoned instead that starlings may rely on a perception of each tone based on its absolute spectral envelope (i.e., the overall pattern of spectral amplitudes across particular frequency bands) rather than abstracted features derived from the fundamental frequency (AP) or on the relative power in the harmonics (timbre). By this view, disruption to the absolute spectral envelope (and thus recognition) could occur by changing the relative power in the harmonics (which plays a role in human timbre perception) or by changing the absolute frequencies where power is found (which plays a role in human harmonic pitch perception). To test this idea, we conducted a third experiment that directly pitted spectral shape information against AP information, using noise-vocoded versions of the training stimuli (which preserved the absolute spectral envelope of each tone but removed periodic pitch cues; Fig. S2 and Audio Files S9–S11) and piano-tone versions of the training sequences (which preserve the pitch but change the absolute spectral envelope of each tone; Audio Files S6–S8, cf., *Materials and Methods*). If absolute spectral envelope variation is the dominant cue that starlings use to recognize tone sequences, then subjects should more easily recognize noise-vocoded versions of the training stimuli than piano-tone versions. If AP is the dominant cue, then this pattern of results should be reversed.

We first tested for transfer from the original three ascending versus descending training sequences to noise-vocoded versions of these stimuli and then to piano-tone versions of the same stimuli (*Materials and Methods*). Generalization from the original training is revealed in both the strength of the initial transfer and in the subsequent acquisition rate. Surprisingly, even in the first 100 trials after transfer, four of five subjects performed better than expected by chance on the noise-vocoded sequences (mean = 70.0% correct; range, 55–86%; upper bound of 99% CI, 63.1%). Over the first five 100-trial blocks with the noise-vocoded stimuli, performance was similarly strong (mean = 78.2% correct; range, 67–90.6%) but significantly below the performance just before transfer [paired $t(4) = 4.05$, $P = 0.015$; Fig. 3]. Response accuracy continued to improve with additional training, reaching a mean of 86.5% correct in blocks 6–10 (Fig. 3) and continuing to improve thereafter until it was statistically indistinguishable from pretransfer levels in blocks 11–14 (mean = 87.2% correct). In contrast, transfer from the original training sequences to the piano-tone stimuli was much poorer. For all subjects, performance in the first 100-trial block after transfer to piano-tone sequences did not differ significantly from chance (mean = 55.4% correct; range, 49–61%; upper bound of 99% CI, 63.1%). Moreover, although performance did improve with further training, it did so only slowly with the mean performance for one of five and four of five subjects exceeding the upper 99% CI around chance in blocks 1–5 and 6–10, respectively (Fig. 3).

Direct comparisons support the strong differences in responding to the noise-vocoded and piano-tone sequences (Fig. 3). Response accuracy varied significantly as a function of test condition [pretransfer, posttransfer blocks 1–5, posttransfer blocks 6–10; $F(2, 8) = 97.08$, $P < 0.0001$, linear mixed effects model (LMM)]. More importantly, the mean accuracy of response to the noise-vocoded stimuli was significantly higher than that for the piano tones, $F(1, 4) = 29.4$, $P = 0.0056$, LMM, and the acquisition rate was significantly faster, $F(2, 8) = 26.8$, $P = 0.0003$, LMM Condition × Stimulus Interaction. Post hoc tests (Tukey's honestly significant difference) show no difference between performance on the two stimulus sets pretransfer ($P = 0.99$) but statistically significant differences in mean performance in blocks 1–5 posttransfer ($P = 0.0014$) and blocks 6–10 posttransfer ($P = 0.0008$). This provides a strong demonstration that starlings can generalize recognition of ascending versus descending tone sequences but do so using the absolute spectral envelope rather than periodic pitch cues.

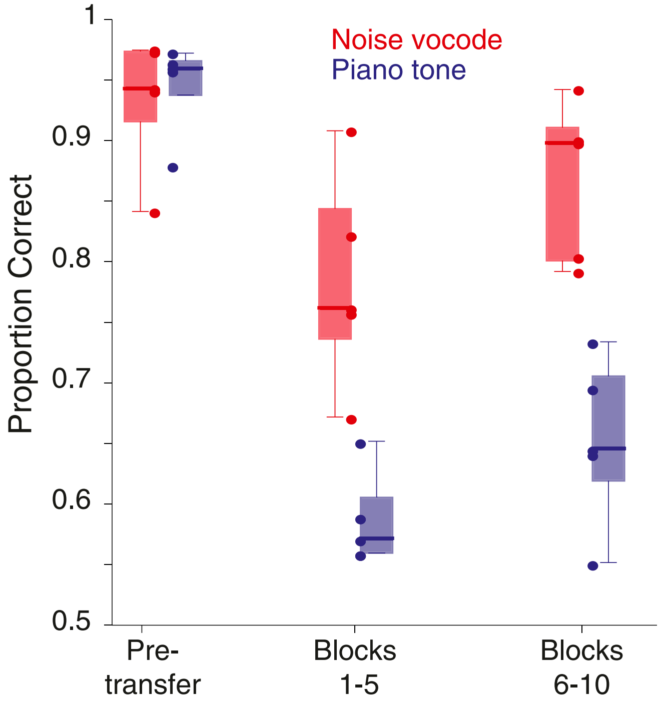

Fig. 3. Comparison of transfer to noise-vocoded versus piano-tone sequences. Box plots showing the proportions of correct responses averaged over the first and second sets of 100-trial blocks during the transfer to the noise-vocoded (red) and piano-tone (blue) stimuli. Performance over the 500 trials just before each transfer is also given. Circles show the data for each subject in each condition. A similar analysis based on 50-trial blocks shows the same effects.

Discussion

Our results challenge the long-held view that songbirds, unlike humans, rely primarily on AP cues for the recognition of tone sequences. By using sound sequences that simultaneously vary in pitch and timbre (as many natural sound sequences do) and by using an acoustic resynthesis technique from speech science that removes pitch cues (noise vocoding) but preserves overall spectral shape, we find that the absolute spectral envelope of each sound (i.e., overall spectral shape across particular frequencies) drives the recognition of tone patterns rather than AP.

These results are surprising given the similarities between starlings and humans in basic psychoacoustic abilities (16) and given that these birds do perceive the pitch of tones with complex harmonic structure, including the pitch of the missing fundamental (18). Thus, although the pitch of a sound may be salient to songbirds, unlike humans they do not seem to use pitch to generalize across sound patterns. This was particularly notable in experiment 3, where songbirds transferred much more readily to noise-vocoded versions of the ascending versus descending training sequences than to piano-tone versions that preserved the pattern of AP of the training stimuli. To human ears, the piano-tone versions sound quite similar to the training sequences because of their identical pitch patterns, whereas the noise-vocoded versions sound strikingly different from the training stimuli (compare Audio Files S1–S3 to Audio Files S6–S11). However, for the birds, this pattern of perceptual similarity seems to be reversed: The noise-vocoded stimuli are treated as more similar to the training stimuli than are the piano-tone versions that preserve AP. In humans, speech recognition is famously robust to the pitch-degrading manipulations introduced by noise vocoders (43), whereas similar manipulations have severe impacts on music perception (44). Our observation that birds rely on spectral shape features to recognize sound sequences suggests a similarity to human speech recognition.

An additional implication of our results is that unlike humans, songbirds may not have largely independent percepts of pitch and timbre (a possibility also suggested by ref. 45). Although hearing scientists have often considered pitch and timbre as distinct dimensions of auditory perception, this distinction may not be an automatic consequence of having a complex auditory system. Indeed, research shows that even humans do not always fully separate these percepts (46–49). For example, in a four-alternative choice task with two tones presented (no change, pitch change, timbre change, both change), nonmusicians reported that both pitch and timbre had changed 26% of the time when in fact the pitch had remained constant and only the timbre changed (50). Musicians, however, made this error just 2% of the time, even though the two timbres (piano vs. trumpet) were easily discriminable by nonmusicians. This raises the idea that the perceptual separability of pitch and timbre is experience-dependent (presumably musicians are better at perceiving pitch and timbre independently because they have more exposure listening to the same instruments playing at different pitches). Likewise, the perception of pitch itself may be more plastic than traditionally appreciated. Individuals considered to have AP can show considerable variability in the range of frequencies they label as the same pitch (51), and recent work shows that exposure to subtly detuned music can significantly alter the note categories of adults with AP (52).

On their surface, our results may seem to be at odds with earlier work showing that starlings can recognize similar spectral structures at different APs (42). In that study, however, pitch did not change within a tone sequence during discrimination training, unlike the current work. The distinct spectral structures used by ref. 42 may have also differed in other perceptual properties, such as degree of consonance or dissonance, that can drive generalization (53). Similarly, if variation in spectral shape (rather than AP) drives tone sequence recognition in songbirds, then one might ask why starlings were unable to recognize frequency-shifted versions of the training sequences (Fig. 2). Although frequency-shifting these sequences preserves the relative relationship among spectral components of each tone (and across the tone sequence), it nonetheless alters the absolute frequencies (Figs. S3–S5). It thus seems that the absolute spectral envelope governs avian tone sequence recognition. For pure tones, the spectral band envelope corresponds directly to pitch; for complex tones, the spectral band envelope contributes to both pitch and timbre percepts.

Importantly, absolute spectral envelope is not likely to be the only perceptual feature that songbirds can use for auditory recognition. Previously, we showed that starlings can maintain the learned recognition of conspecific songs even when those songs are shifted in frequency by large amounts (22). Although the precise cues that starlings use to recognize frequency-shifted conspecific songs require future study, such manipulations alter the absolute spectral envelope of the signals. However, the spectrotemporal complexity of songs and other natural stimuli provides additional perceptual cues (e.g., rhythm and amplitude envelope) that are invariant across frequency shifts. Moreover, the importance of any given cue can vary depending on listening task. Sensitivity to these features in specific tasks may help to explain prior evidence suggesting specialized song-processing mechanisms in birds (15).

Our results indicate that behavioral effects tied classically to changes in the frequencies of pure tones (29, 30, 37–39) should not be strictly interpreted as changes to the percept of pitch. Instead, we suggest a revised perspective on melody recognition by songbirds. We propose that unlike humans, for whom pitch plays a dominant role in the perception of melodic sequences, songbirds rely on a perceptual representation that appears more closely tied to absolute spectral envelope. This surprising difference has implications both for research in the cognitive psychology of auditory perception and for neuroscientists investigating the physiological and computational processes underlying auditory recognition. A promising avenue of research lies in linking cross-species differences in the physiological organization of the auditory system to observed differences in the use of auditory cues. Further research manipulating spectral shape and pitch salience of tone sequences, for example using noise-vocoding or sinusoidal-vocoding, which allow control over spectral resolution while removing versus preserving pitch cues, respectively (54, 55), will help researchers understand species differences in auditory sequence recognition.

Materials and Methods

Subjects. Five adult wild-caught European Starlings (*S. vulgaris*) of unknown sex were tested. No subjects had previously been used in other tasks or had prior exposure to experimental stimuli. Before beginning experimental training, subjects were housed in a large mixed-sex aviary.

Stimuli. Stimuli were sequences of four complex tones with no intervening pauses. Each tone was 368 ms in duration so that the full sequences lasted about 1.5 s. We created the sequences by MIDI synthesis as 16 bit, 44.1 KHz wave files using the built-in Quicktime MIDI synthesizer in Mac OS 10.6. General MIDI instrument codes for the sounds were 69 (oboe), 53 (sung "aah" formant), 60 (muted trumpet), and 81 (synthesizer square wave). Our criterion in selecting these sounds was to choose instruments with a sustained amplitude envelope, so that pitch and spectral shape (and not amplitude envelope) were the primary cues to distinguish the tones.

Training stimuli. From the set of synthesized tones, we created six training stimuli, each a sequence of four tones. Each tone within a sequence was distinct in both pitch and timbre from the other tones in that sequence. For three of the sequences, the pitch of each tone increased systematically with intervals of two semitones between each note from start to end, and for the three remaining sequences, the pitch of the tones decreased by the same intervals. The order of timbres in the ascending and descending pitch sequences also differed systematically so that the serial pattern structure between the two types of sequences was redundant across pitch and timbre (Fig. 1*B*). The lowest ascending training stimulus started on Bb4 (466.16 Hz) and continued to C5 (523.25 Hz), D5 (587.33 Hz), and E5 (659.26 Hz) (corresponding to MIDI notes 70, 72, 74, and 76, respectively). The corresponding descending stimulus used the same pitches in reverse order, starting at E5 and ending at Bb4. The other two ascending stimuli started on C5 and D5, ending on F#5 and G#5, respectively, whereas descending stimuli used the same pitches in reverse order. Thus, the two other ascending and descending stimuli represented upward shifts of two and four semitones relative to the original Bb4–C5–D5–E5 or E5–D5–C5–Bb4 sequence. All stimuli were normalized to a mean power of 65 dB.

Pitch-shifted stimuli. We synthesized test stimuli with the same interval spacing and timbre sequences as the training stimuli but starting at pitches not heard during training. The novel ascending stimuli started at Bb3, D3, F#3, A4, B4, C#5, F5, G5, and Bb5. Relative to the lowest ascending training sequence starting on Bb4, these sequence represent shifts of –12, –8, –6, –3, –1, 1, 3, 6, 8, and 12 semitones, respectively. Novel sequences starting at B4 and C#5 lie between two training stimuli but were never heard during training, whereas the other test stimuli lie partly or entirely outside of the training frequency range.

Piano-tone stimuli. We also constructed novel timbre versions of the training stimuli—that is, three ascending and three descending sequences—matched in AP and duration pattern to the sequences in Fig. 1 but using only piano tones. In addition, we synthesized versions of these novel timbre sequences that were shifted by ±1, ±3, ±6, ±8, and ±12 semitones relative to lowest frequency ascending/descending pair of training stimuli.

Noise-vocoded stimuli. To disrupt pitch cues while retaining the frequency-specific spectral shape of each tone, we created noise-vocoded versions of the training stimuli. Noise vocoding is accomplished by dividing an acoustic signal into a fixed number of frequency bands, extracting the amplitude envelope within each band, and then using this envelope to modulate band-pass filtered white noise. These amplitude-modulated noise bands are then recombined to create the noise-vocoded signal (see ref. 56). Noise vocoding has been used for many years in speech research to investigate the role of detailed spectral structure (independent of pitch) in speech perception (43). Noise-vocoded speech sounds somewhat like whispered speech and is highly intelligible to humans if the number of frequency bands is sufficiently large (e.g., 15 bands between 50 and 8,000 Hz, as in ref. 57), because this preserves the overall time-varying shape of the speech spectrum.

We constructed noise-vocoded versions of our training stimuli by dividing each original training stimulus into 16 logarithmically spaced frequency bands, with the first band spanning 50 Hz to 193 Hz and the 16th band spanning 8,865 Hz to 11,000 Hz. We then computed the amplitude envelope for each of these bands and applied it to band-limited white noise. Vocoding was done using a custom-written Praat script (see computer script found in Dataset S1).

Procedures. Each subject was trained and tested in four phases: shaping, recognition training, recognition testing, and novel stimulus transfer using a two-alternative choice operant training procedure. Further details on the operant training are provided in previous publications (e.g., refs. 22, 58). Subjects were housed individually and trained inside a sound isolation chamber with access to an operant panel (Fig. 1*A*). During training, subjects initiated trials when the house lights were on (matched to local daylight). Water was freely available, and animals were not fed except when earning a food reward after completing an experimental trial. All procedures were completed as part of a protocol approved by the University of California, San Diego Institutional Animal Care and Use Committee.

Shaping. During shaping, each subject was trained to obtain food from a hopper underneath the food port (Fig. 1*A*) by pecking the center response port. After pecking the center port 100 times, they were trained to peck the center port and then either the left or right response port (cued randomly with a flashing light) for a food reward. Each subject completed several hundred trials pecking the left and right response ports.

Recognition training. After shaping, subjects learned to associate ascending stimuli (with the characteristic timbre sequence; see Fig. S1) with the left response port and descending stimuli with the right response port. On each trial, a peck to the center response port started playback of a randomly selected training stimulus from a speaker behind the operant panel. Pecks to the left or right response port within 2 s after the stimulus playback ended led to reinforcement. Incorrect responses were punished with 10–20 s of lights out; correct responses resulted in 2 s of food access paired with a secondary visual reinforcement (blinking LEDs in all three response ports). To improve performance and increase the number of trials performed, we transitioned subjects over multiple sessions to a fixed-ratio reinforcement schedule where they were fed only if they responded correctly to a fixed number of consecutive trials, otherwise receiving only the secondary visual reinforcer for correct responses. Incorrect responses or nonresponses reset the count of correct trials. Eventually the fixed ratio was set at six trials for each subject.

Recognition testing. To test generalization of the pitch-shifted and novel timbre tone sequences, we used a probe procedure. On 66% of trials, we presented the training stimuli (randomly selected as in the initial training), and on the remaining 33% of trials, we presented one of the pitch-shifted or novel timbre stimuli. To keep response rates high, we required subjects to respond correctly to six consecutive training stimulus trials. Responses to the test stimuli were never immediately followed by reinforcement and did not affect the consecutive correct response counter. Failure to respond to any stimulus reset the response counter so that six correct responses to training stimuli were again required to receive a food reward.

Transfer procedure. We a used a transfer procedure rather than a probe-recognition procedure in experiment 3. In the transfer procedure, subjects were switched immediately from sessions in which the training stimuli are presented on all trials to sessions in which the test stimuli were presented on all trials. During these test sessions, subjects were reinforced (or punished) for correct (or incorrect) responses to test stimuli, as during the training sessions. Above chance performance during initial transfer trials and/or rapid acquisition indicates the generalization, or transfer, of learning from the training to the test stimuli. Initial transfer performance that falls to chance and takes longer to recover indicates weaker generalization and the need to relearn the recognition task anew. The primary difference compared with the probe procedure is that the transfer procedure differentially reinforces responses and thus affords subjects an opportunity to learn stimulus response associations, providing access to a more subtle behavioral measure in acquisition rate. We exposed subjects to a mean of 2,636 trials with the noise-vocoded stimuli (range, 1,389–3,308) before returning them to the original training stimuli. After ensuring stable, accurate recognition (mean = 94.4% correct; range, 87.6–97.0% correct), we then transferred them to piano-tone versions of the training stimuli.

ACKNOWLEDGMENTS. The research was partially funded by National Institutes of Health Grant R01DC008358. A.D.P. was supported by Neurosciences Research Foundation as part of its program on music and the brain at The Neurosciences Institute.

1. Comins JA, Gentner TQ (2014) Temporal pattern processing in songbirds. *Curr Opin Neurobiol* 28:179–187.
2. Doupe AJ, Kuhl PK (1999) Birdsong and human speech: Common themes and mechanisms. *Annu Rev Neurosci* 22:567–631.
3. Fee MS, Scharff C (2010) The songbird as a model for the generation and learning of complex sequential behaviors. *ILAR J* 51(4):362–377.
4. Rohrmeier M, Zuidema W, Wiggins GA, Scharff C (2015) Principles of structure building in music, language and animal song. *Philos Trans R Soc Lond B Biol Sci* 370(1664):20140097.
5. Karten HJ (2013) Neocortical evolution: Neuronal circuits arise independently of lamination. *Curr Biol* 23(1):R12–R15.
6. Carr CE (1992) Evolution of the central auditory system in reptiles and birds. *The Evolutionary Biology of Hearing*, eds Webster DB, Fay RR, Popper AN (Springer-Verlag, New York), pp 511–544.
7. Wang Y, Brzozowska-Prechtl A, Karten HJ (2010) Laminar and columnar auditory cortex in avian brain. *Proc Natl Acad Sci USA* 107(28):12676–12681.
8. Jarvis ED, et al.; Avian Brain Nomenclature Consortium (2005) Avian brains and a new understanding of vertebrate brain evolution. *Nat Rev Neurosci* 6(2):151–159.
9. Dugas-Ford J, Rowell JJ, Ragsdale CW (2012) Cell-type homologies and the origins of the neocortex. *Proc Natl Acad Sci USA* 109(42):16974–16979.
10. Calabrese A, Woolley SM (2015) Coding principles of the canonical cortical microcircuit in the avian brain. *Proc Natl Acad Sci USA* 112(11):3517–3522.
11. Riede T, Goller F (2010) Peripheral mechanisms for vocal production in birds—Differences and similarities to human speech and singing. *Brain Lang* 115(1):69–80.
12. Tierney AT, Russo FA, Patel AD (2011) The motor origins of human and avian song structure. *Proc Natl Acad Sci USA* 108(37):15510–15515.
13. Petkov CI, Jarvis ED (2012) Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Front Evol Neurosci* 4:12.
14. Reiner A, et al.; Avian Brain Nomenclature Forum (2004) Revised nomenclature for avian telencephalon and some related brainstem nuclei. *J Comp Neurol* 473(3):377–414.
15. Dooling RJ (1986) Perception of vocal signals by budgerigars (Melopsittacus undulatus). *Exp Biol* 45(3):195–218.
16. Klump GM, Langemann U, Gleich O (2000) The European starling as a model for understanding perceptual mechanisms. *Auditory Worlds: Sensory Analysis and Perception in Animals and Man*, eds Manley GA, Oeckinghaus H, Koessl M, Klump GM, Fastl H (Verlag Chemie, Weinheim, Germany), pp 193–211.
17. Langemann U, Klump GM, Dooling RJ (1995) Critical bands and critical-ratio bandwidth in the European starling. *Hear Res* 84(1-2):167–176.
18. Cynx J, Shapiro M (1986) Perception of missing fundamental by a species of songbird (Sturnus vulgaris). *J Comp Psychol* 100(4):356–360.
19. Itatani N, Klump GM (2014) Neural correlates of auditory streaming in an objective behavioral task. *Proc Natl Acad Sci USA* 111(29):10738–10743.
20. MacDougall-Shackleton SA, Hulse SH, Gentner TQ, White W (1998) Auditory scene analysis by European starlings (Sturnus vulgaris): Perceptual segregation of tone sequences. *J Acoust Soc Am* 103(6):3581–3587.
21. Baptista LF, Keister RA (2005) Why birdsong is sometimes like music. *Perspect Biol Med* 48(3):426–443.
22. Bregman MR, Patel AD, Gentner TQ (2012) Stimulus-dependent flexibility in non-human auditory pitch processing. *Cognition* 122(1):51–60.
23. Nicolai J, Gundacker C, Teeselink K, Güttinger HR (2014) Human melody singing by bullfinches (Pyrrhula pyrrula) gives hints about a cognitive note sequence processing. *Anim Cogn* 17(1):143–155.
24. West MJ, King AP (1990) Mozart's starling. *Am Sci* 78(2):106–114.
25. Brown S, Jordania J (2013) Universals in the world's musics. *Psychol Music* 41(2): 229–248.
26. Patel AD, Demorest SM (2013) Comparative music cognition: Cross-species and cross-cultural studies. *The Psychology of Music*, ed Deutsch D (Elsevier Academic Press, San Diego), 3rd Ed, pp 647–681.
27. Plantinga J, Trainor LJ (2005) Memory for melody: Infants use a relative pitch code. *Cognition* 98(1):1–11.
28. McDermott JH, Lehr AJ, Oxenham AJ (2008) Is relative pitch specific to pitch? *Psychol Sci* 19(12):1263–1271.
29. Hulse SH, Cynx J, Humpal J (1984) Absolute and relative pitch discrimination in serial pitch perception by birds. *J Exp Psychol Gen* 113(1):38–54.
30. Page SC, Hulse SH, Cynx J (1989) Relative pitch perception in the European starling (Sturnus vulgaris): Further evidence for an elusive phenomenon. *J Exp Psychol Anim Behav Process* 15(2):137–146.
31. Hoeschele M, Guillette LM, Sturdy CB (2012) Biological relevance of acoustic signal affects discrimination performance in a songbird. *Anim Cogn* 15(4):677–688.
32. Njegovan M, Weisman R (1997) Pitch discrimination in field- and isolation-reared black-capped chickadees (Parus atricapillus). *J Comp Psychol* 111(3):294–301.
33. Hulse SH, Cynx J (1985) Relative pitch perception is constrained by absolute pitch in songbirds (Mimus, Molothrus, and Sturnus). *J Comp Psychol* 99(2):176–196.
34. Cynx J, Hulse SH, Polyzois S (1986) A psychophysical measure of pitch discrimination loss resulting from a frequency range constraint in European starlings (Sturnus vulgaris). *J Exp Psychol Anim Behav Process* 12(4):394–402.
35. Hulse SH, Kline CL (1993) The perception of time relations in auditory tempo discrimination. *Anim Learn Behav* 21(3):281–288.
36. Bernard DJ, Hulse SH (1992) Transfer of serial stimulus relations by European starlings (Sturnus vulgaris): Loudness. *J Exp Psychol Anim Behav Process* 18(4):323–334.
37. Hulse SH, Takeuchi AH, Braaten RF (1992) Perceptual invariances in the comparative psychology of music. *Music Percept* 10(2):151–184.
38. Weisman RG, Njegovan MG, Williams MT, Cohen JS, Sturdy CB (2004) A behavior analysis of absolute pitch: Sex, experience, and species. *Behav Processes* 66(3):289–307.
39. Weisman RG, Williams MT, Cohen JS, Njegovan MG, Sturdy CB (2006) The comparative psychology of absolute pitch. *Comparative Cognition: Experimental Explorations of Animal Intelligence*, eds Wasserman EA, Zentall TR (Oxford University Press, New York), pp 71–88.
40. Adret-Hausberger M, Jenkins PF (1988) Complex organization of the warbling song in starlings. *Behaviour* 107(3):138–156.
41. McAdams S (2013) Musical timbre perception. *The Psychology of Music*, ed Deutsch D (Elsevier Academic Press, San Diego, CA), pp 765–767.
42. Braaten RF, Hulse SH (1991) A songbird, the European starling (Sturnus vulgaris), shows perceptual constancy for acoustic spectral structure. *J Comp Psychol* 105(3): 222–231.
43. Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M (1995) Speech recognition with primarily temporal cues. *Science* 270(5234):303–304.
44. Kong YY, Cruz R, Jones JA, Zeng FG (2004) Music perception with temporal cues in acoustic and electric hearing. *Ear Hear* 25(2):173–185.
45. Hoeschele M, Cook RG, Guillette LM, Hahn AH, Sturdy CB (2014) Timbre influences chord discrimination in black-capped chickadees (Poecile atricapillus) but not humans (Homo sapiens). *J Comp Psychol* 128(4):387–401.
46. Krumhansl CL, Iverson P (1992) Perceptual interactions between musical pitch and timbre. *J Exp Psychol Hum Percept Perform* 18(3):739–751.
47. Melara RD, Marks LE (1990) Perceptual primacy of dimensions: Support for a model of dimensional interaction. *J Exp Psychol Hum Percept Perform* 16(2):398–414.
48. Pitt MA, Crowder RG (1992) The role of spectral and dynamic cues in imagery for musical timbre. *J Exp Psychol Hum Percept Perform* 18(3):728–738.
49. Caruso VC, Balaban E (2014) Pitch and timbre interfere when both are parametrically varied. *PLoS One* 9(1):e87065.
50. Pitt MA (1994) Perception of pitch and timbre by musically trained and untrained listeners. *J Exp Psychol Hum Percept Perform* 20(5):976–986.
51. Bahr N, Christensen CA, Bahr M (2005) Diversity of accuracy profiles for absolute pitch recognition. *Psychol Music* 33(1):58–93.
52. Hedger SC, Heald SL, Nusbaum HC (2013) Absolute pitch may not be so absolute. *Psychol Sci* 24(8):1496–1502.
53. Hulse SH, Bernard DJ, Braaten RF (1995) Auditory discrimination of chord-based spectral structures by European starlings (Sturnus vulgaris). *J Exp Psychol Gen* 124(4): 409–423.
54. Hervais-Adelman AG, Davis MH, Johnsrude IS, Taylor KJ, Carlyon RP (2011) Generalization of perceptual learning of vocoded speech. *J Exp Psychol Hum Percept Perform* 37(1):283–295.
55. Shannon RV, Fu QJ, Galvin J, 3rd (2004) The number of spectral channels required for speech recognition depends on the difficulty of the listening situation. *Acta Otolaryngol Suppl* (552):50–54.
56. Davis MH, Johnsrude IS, Hervais-Adelman A, Taylor K, McGettigan C (2005) Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *J Exp Psychol Gen* 134(2):222–241.
57. Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. *J Neurosci* 23(8):3423–3431.
58. Gentner TQ (2008) Temporal scales of auditory objects underlying birdsong vocal recognition. *J Acoust Soc Am* 124(2):1350–1359.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

Supporting Information

Bregman et al. 10.1073/pnas.1515380113

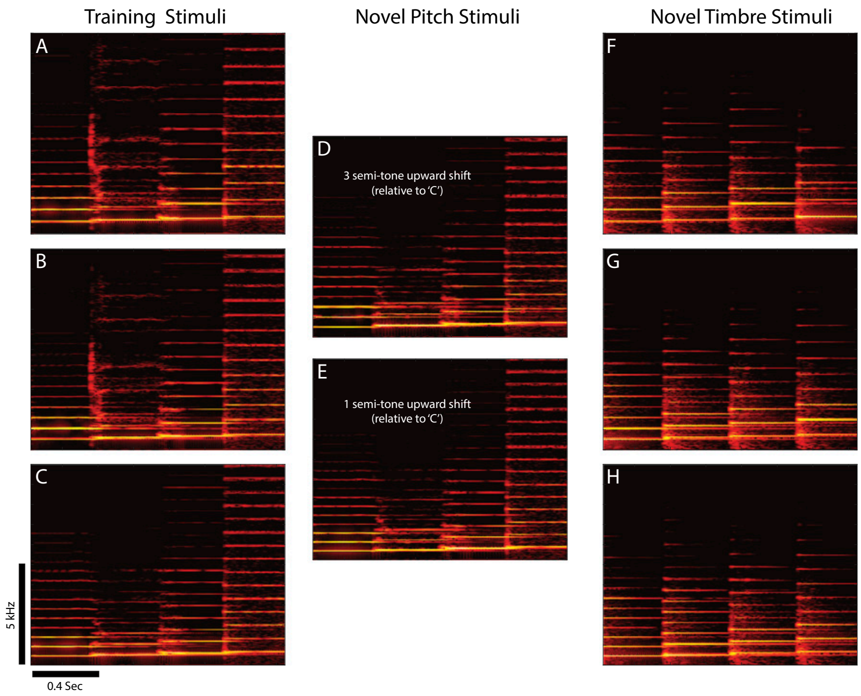

Fig. S1. Example training and test stimuli. Spectrograms showing the three ascending tone sequence stimuli used for initial training (*A*–*C*), examples of the novel pitch test stimuli (*D* and *E*), and examples of the novel timbre test stimuli (*F*–*H*). Each panel shows a single sequence of four complex tones ascending in pitch (see Fig. 1*B*). The stimuli in *D* and *E* are shifted three and one semitone above the sequence in *C*, respectively. The pitches for each four-tone sequence in *A*–*C* match those in *F*–*H*, but all of the tones in *F*–*H* have a novel timbre (piano) that was not heard in any of the training stimuli (*A*–*C*). Scale bars show time and frequency. Apparent bleed at the boundaries between individual tones is an artifact of the spectrogram windowing.

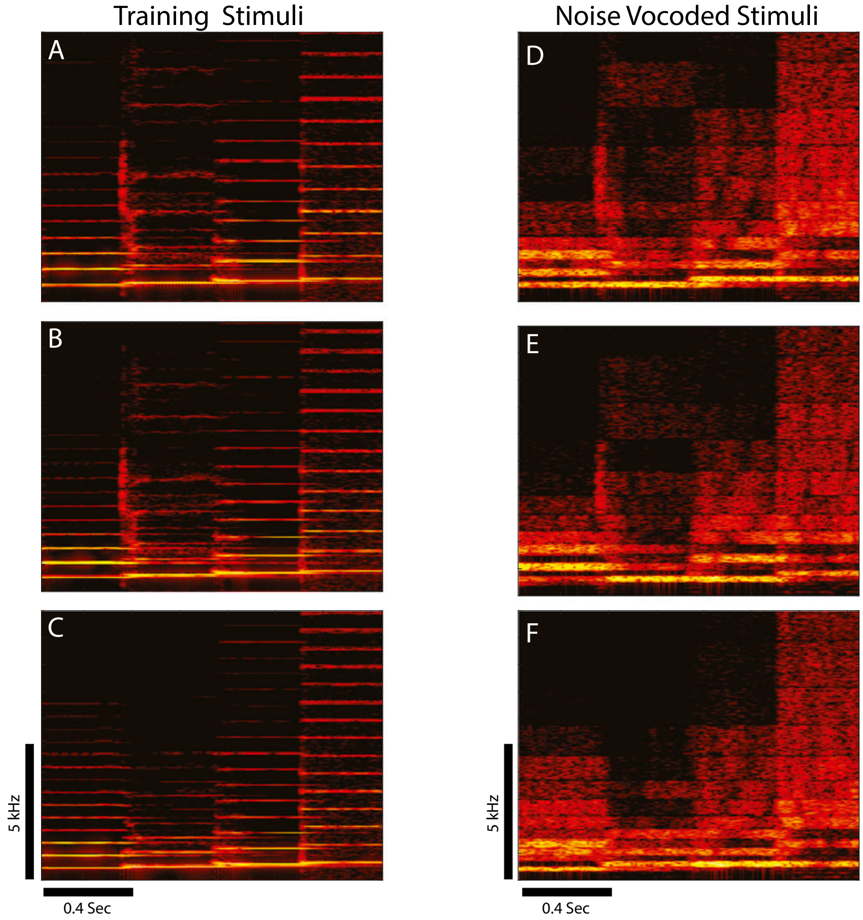

Fig. S2. Example noise-vocoded stimuli. Spectrograms showing the three ascending tone sequence stimuli used for initial training (*A*–*C*) as in Fig. S1 and the three corresponding noise-vocoded versions of these stimuli (*D*–*F*). Scale bars show time and frequency.

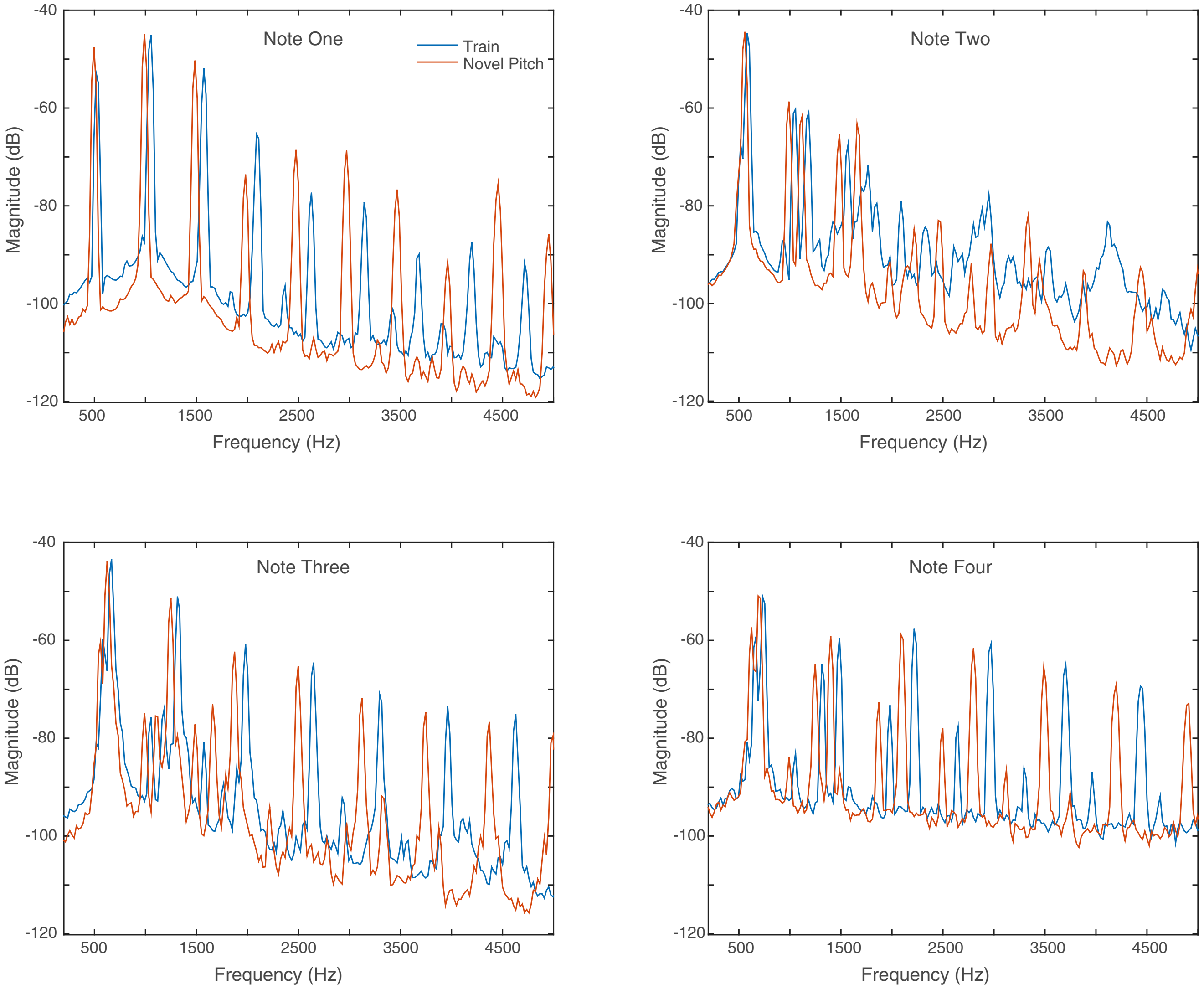

Fig. S3. Example comparisons of the overall absolute spectral shape for notes in the training and novel pitch test sounds. Each panel shows the frequency spectra for one of the four notes (blue lines) in a training stimulus (Fig. S1*B*) and the corresponding notes (red lines) in a novel pitch stimulus (Fig. S1*E*) that is one semitone lower.

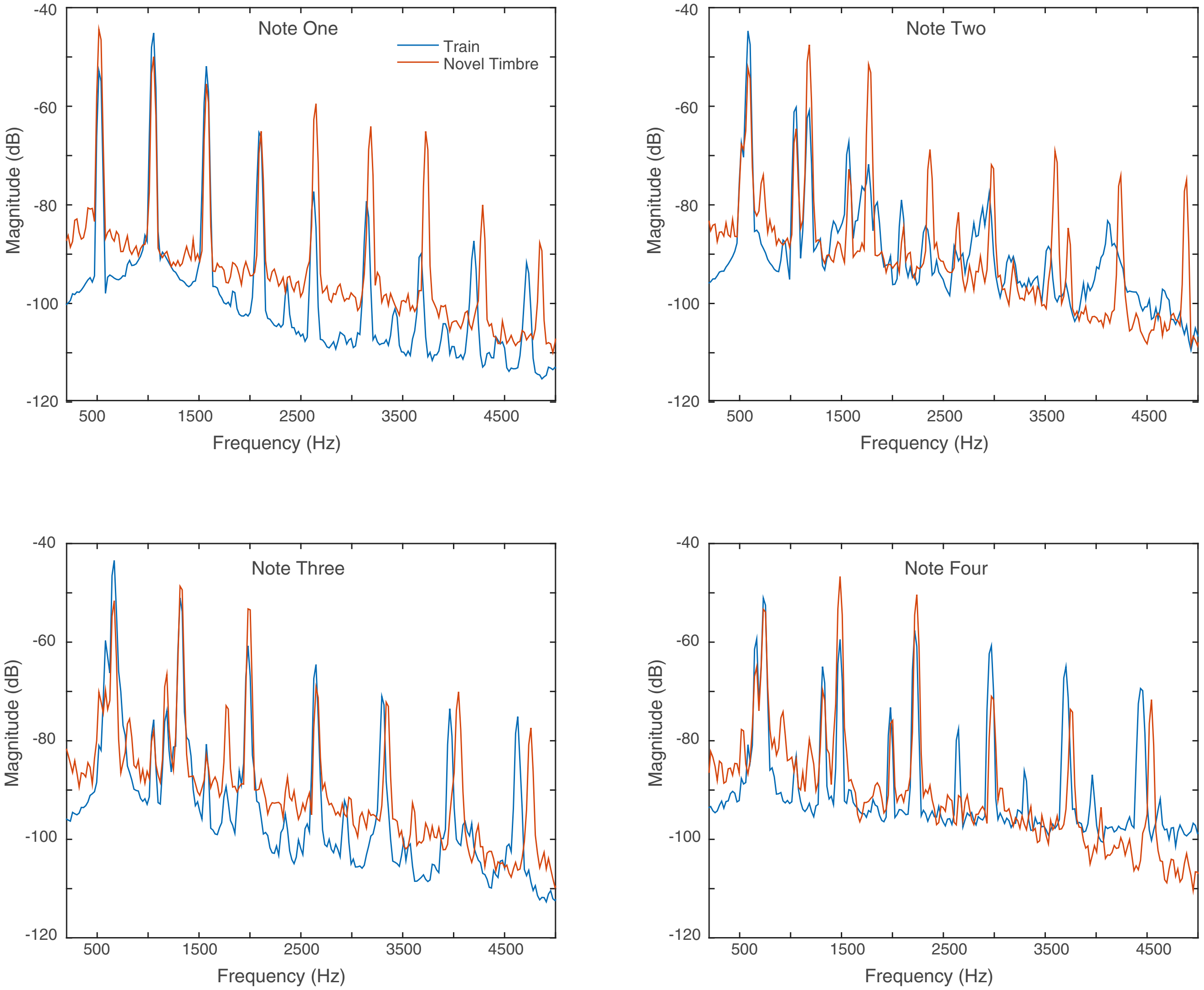

Fig. S4. As in Fig. S3 but comparing the spectra of notes in a training stimulus (blue lines; Fig. S1*B*) to spectra from notes in a novel timbre stimulus (red lines; Fig. S1*G*).

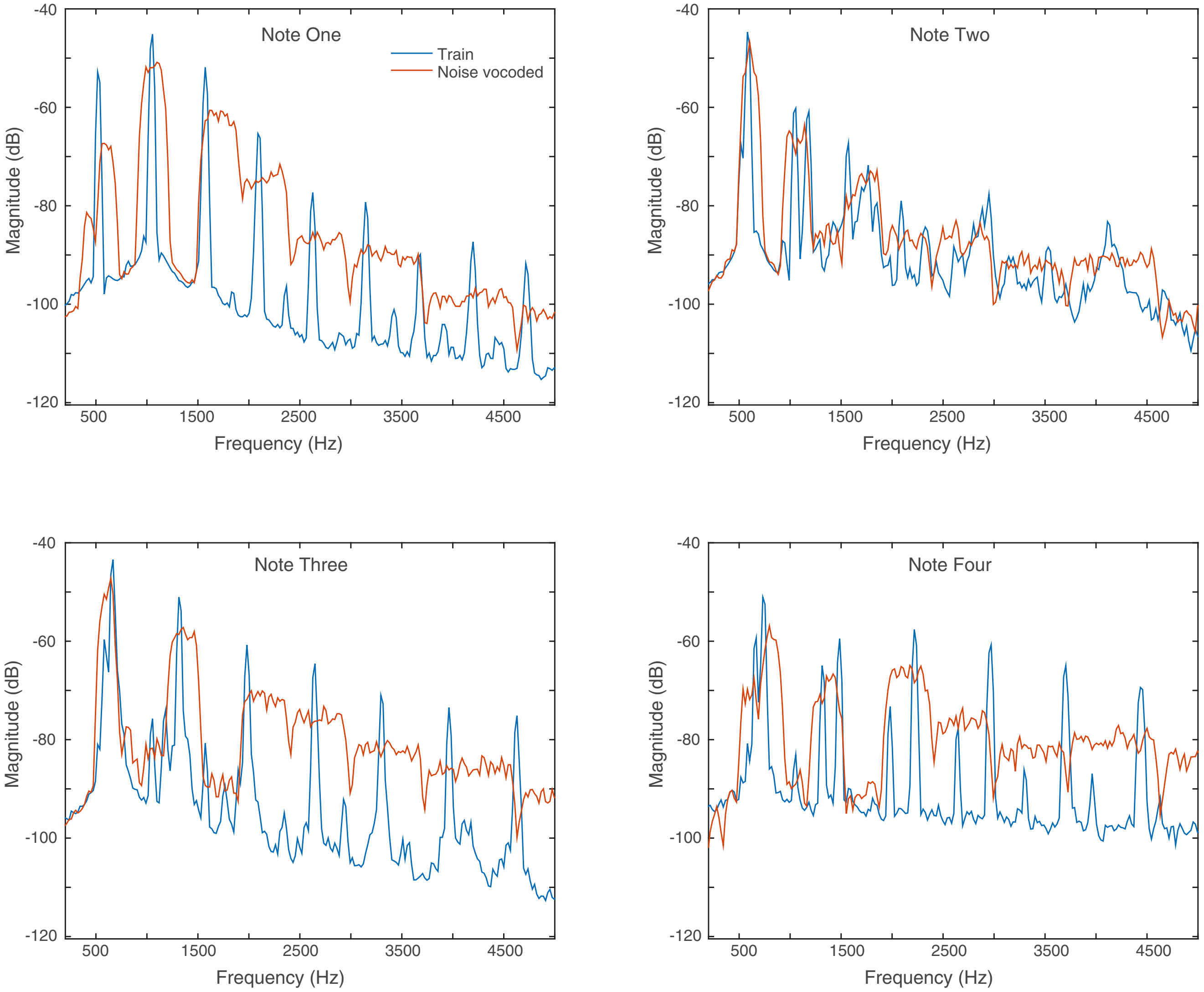

Fig. S5. As in Fig. S3 but comparing the spectra of notes in a training stimulus (blue lines; Fig. S2*B*) to spectra from notes in a noise-vocoded version of the same stimulus (red lines; Fig. S2*E*).

Audio File S1. Sound file (wav format) for the stimulus shown in Fig. S1*A*.

Audio File S1

Audio File S2. Sound file (wav format) for the stimulus shown in Fig. S1*B*.

Audio File S2

Audio File S3. Sound file (wav format) for the stimulus shown in Fig. S1*C*.

Audio File S3

Audio File S4. Sound file (wav format) for the stimulus shown in Fig. S1*D*.

Audio File S4

Audio File S5. Sound file (wav format) for the stimulus shown in Fig. S1*E*.

Audio File S5

Audio File S6. Sound file (wav format) for the stimulus shown in Fig. S1*F*.

Audio File S6

Audio File S7. Sound file (wav format) for the stimulus shown in Fig. S1*G*.

Audio File S7

Audio File S8. Sound file (wav format) for the stimulus shown in Fig. S1*H*.

Audio File S8

Audio File S9. Sound file (wav format) for the stimulus shown in Fig. S2*D*.

Audio File S9

Audio File S10. Sound file (wav format) for the stimulus shown in Fig. S2*E*.

Audio File S10

Audio File S11. Sound file (wav format) for the stimulus shown in Fig. S2*F*.

Audio File S11

Other Supporting Information Files

Dataset S1 (DOCX)