Journal of Comparative Psychology
2013, Vol. 127, No. 2, 142–153

© 2012 American Psychological Association
0735-7036/13/$12.00 DOI: 10.1037/a0029734

Processing Pitch in a Nonhuman Mammal (*Chinchilla laniger*)

William P. Shofner and Megan Chaney
Indiana University

Whether the mechanisms giving rise to pitch reflect spectral or temporal processing has long been debated. Generally, sounds having strong harmonic structures in their spectra have strong periodicities in their temporal structures. We found that when a wideband harmonic tone complex is passed through a noise vocoder, the resulting sound can have a harmonic structure with a large peak-to-valley ratio, but with little or no periodicity in the temporal structure. To test the role of harmonic structure in pitch perception for a nonhuman mammal, we measured behavioral responses to noise-vocoded tone complexes in chinchillas (*Chinchilla laniger*) using a stimulus generalization paradigm. Chinchillas discriminated either a harmonic tone complex or an iterated rippled noise from a 1-channel vocoded version of the tone complex. When tested with vocoded versions generated with 8, 16, 32, 64, and 128 channels, responses were similar to those of the 1-channel version. Behavioral responses could not be accounted for based on harmonic peak-to-valley ratio as the acoustic cue, but could be accounted for based on temporal properties of the autocorrelation functions such as periodicity strength or the height of the first peak. The results suggest that pitch perception does not arise through spectral processing in nonhuman mammals but rather through temporal processing. The conclusion that spectral processing contributes little to pitch in nonhuman mammals may reflect broader cochlear tuning than that described in humans.

Keywords: chinchilla, pitch perception, noise vocoder, stimulus generalization

Pitch is one of the most fundamental auditory perceptions. In speech perception, pitch plays a role in prosody (Ohala, 1983), whereas variations in pitch provide linguistic meaning to words in tonal languages (Cutler & Chen, 1997; Lee, Vakoch, & Wurm, 1996; Ye & Connine, 1999). Pitch also plays a role in gender identification (Gelfer & Mikos, 2005) and may provide attention cues for infant-directed speech (Gauthier & Shi, 2011). In music perception, pitch is essential for melody recognition (Cousineau, Demany, & Pressnitzer, 2009; Dowling & Fujitani, 1971), and impairments in melody recognition in amusic listeners are related to deficits in pitch perception (Ayotte, Peretz, & Hyde, 2002; Peretz et al., 2002). Pitch is also thought to play a role in the ability of human listeners to group and segregate sound sources (Darwin, 2005).

Whether pitch is processed through spectral or temporal mechanisms is an issue still debated today more than 160 years after the original dispute between Seebeck and Ohm (Turner, 1977). Models using either spectral only (Cohen, Grossberg, & Wyse, 1995; Goldstein, 1973; McLachlan, 2011; Shamma & Klein, 2000; Wightman, 1973) or temporal only processing (Balaguer-Ballester, Denham, & Meddis, 2008; de Cheveigné, 1998; Licklider, 1951; Meddis & O'Mard, 1997; Yost & Hill, 1979) can account for a wide variety of pitch percepts, but they have been unable to eliminate conclusively one mechanism over the other. The debate continues largely because it is difficult to alter the spectral structure of a sound without altering its temporal structure. Sounds showing strong harmonic structures in their spectra generally show strong periodicities in their autocorrelation functions (ACFs). For example, compare the spectra and ACFs for a harmonic tone complex and an infinitely iterated rippled noise (IIRN) in Figure 1. Both of these sounds evoke a pitch of 500 Hz in human listeners. Note that the peak-to-valley ratios of the harmonics in the spectrum for the tone complex in Figure 1A are larger than those of the IIRN in Figure 1B, and that the heights and numbers of peaks in the ACF are also larger for the tone complex in Figure 1C than for the IIRN in Figure 1D. An ideal approach would be to study pitch perception using sounds in which spectral and temporal properties can be manipulated independently. However, this ideal approach is unrealistic because the spectrum and ACF are Fourier pairs and thus by definition cannot be manipulated independently.

A more realistic approach is to manipulate various acoustic features of complex sounds and then compare behavioral performance with predictions based on certain assumptions regarding spectral or temporal processing. Examples in the literature include the use of high-pass filtering to eliminate resolved harmonic components (Houtsma & Smurzynski, 1990; Patterson, Handel, Yost, & Datta, 1996), the use of different delay-and-add networks to generate different types of IIRNs (Yost, 1996; Yost, Patterson, & Sheft, 1996), and the use of transposed tones (Oxenham, Bernstein, & Penagos, 2004). Another interesting example is the use of sinusoidally amplitude-modulated noise (SAM-noise). SAM-noise has no harmonic structure in its spectrum and shows no periodicity in its waveform ACF, but it does contain temporal modulations in the envelope, which can evoke pitch percepts (Burns & Viemeis-

This article was published Online First September 17, 2012.

William P. Shofner and Megan Chaney, Department of Speech and Hearing Sciences, Indiana University.

This research was supported in part by Grant R01 DC 005596 from the National Institutes of Health.

Correspondence concerning this article should be addressed to William P. Shofner, Department of Speech and Hearing Sciences, Indiana University, 200 South Jordan Avenue, Bloomington, IN 47405. E-mail: wshofner@indiana.edu

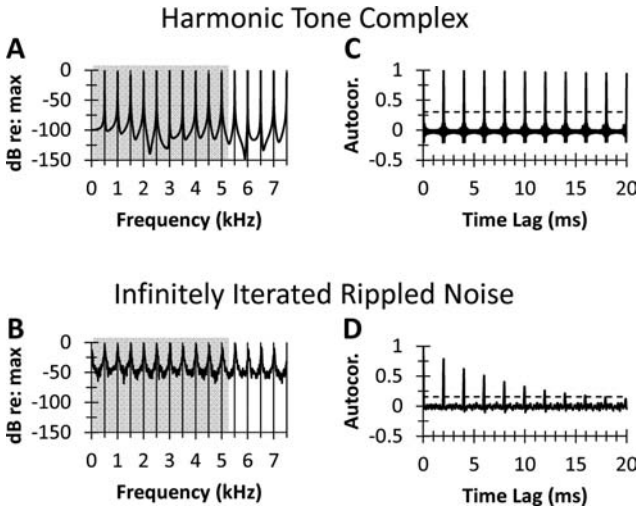


Figure 1. Example spectra (A, B) and autocorrelation functions (C, D) illustrate harmonic structure and periodicity for a harmonic tone complex and infinitely iterated rippled noise, which both evoke a 500-Hz pitch. Vertical lines indicate the frequencies of the harmonics. (A) Spectrum of a wideband harmonic tone complex comprises a fundamental frequency of 500 Hz and successive harmonics up to and including the 20th harmonic. (B) Spectrum of an infinitely iterated rippled noise having a delay of 2 ms and a delayed noise attenuation of –1 dB (see Method section). The spectrum has been smoothed using a seven-bin moving window average. (C) Autocorrelation function for the harmonic tone complex. (D) Autocorrelation function for the infinitely iterated rippled noise. The gray shaded areas indicate the frequencies corresponding to harmonics 1–10. The horizontal dashed lines indicate 2 standard deviations.

ter, 1976, 1981). SAM-noise is a stimulus having a temporal structure but no harmonic structure; thus, it can be argued that any pitch percept evoked by SAM-noise must arise through temporal processing.

In the present study, we show that by passing a wideband harmonic tone complex (wHTC) through a noise vocoder, the resulting sound can have a harmonic structure with little or no temporal structure. The vocoder or voice encoder was originally a device developed at Bell Telephone Laboratories for the analysis and resynthesis of speech (Dudley, 1939). A noise vocoder analyzes a sound, such as speech, by first passing the sound through a fixed number of contiguous bandpass filters and extracting the envelope from each filter or channel. The envelopes from each channel are then used to modulate a fixed number of contiguous bandpass noises. The modulated bandpass noises are then summed to yield the resynthesized sound. Recently, vocoders have been used to degrade the features of sounds for studies in speech perception (Dorman, Loizou, & Rainey, 1997; Friesen, Shannon, Baskent, & Wang, 2001; Loebach & Pisoni, 2008; Loizou, Dorman, & Tu, 1999; Shannon, Zeng, Kamath, & Wygonski, 1998; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995), melody recognition and pitch discrimination (Green, Faulkner, & Rosen, 2002; Qin & Oxenham, 2005; Smith, Delgutte, & Oxenham, 2002), and recognition of environmental sounds (Loebach & Pisoni, 2008; Shafiro, 2008). We found that vocoded versions of a wHTC can have harmonic structures with large peak-to-valley ratios in the spectra but with little or no periodicity strength in the ACFs. In some respects, noise-vocoded wHTCs are the antithesis of SAM-noises: SAM-noises have no harmonic structures but do have temporal structures, whereas noise-vocoded wHTCs have harmonic structures but little or no temporal structures.

Pitch perception is not a distinctively human characteristic given that pitch-like percepts are common across mammalian species. Chinchillas have a perceptual dimension corresponding to pitch (Shofner, Yost, & Whitmer, 2007), including a perception of the missing fundamental (Shofner, 2011). In addition, the audiogram (R. S. Heffner & Heffner, 1991) and the spectral dominance region for pitch (Shofner & Yost, 1997) of chinchillas are both similar to those of human listeners. We used noise-vocoded wHTCs to investigate the role of spectral processing in pitch perception for the chinchilla. First, we were interested in gaining some insight as to which processing scheme, spectral or temporal, reflects the more primitive state from an evolutionary perspective. Second, because neurophysiological studies related to pitch are based on animal models (e.g., Bendor & Wang, 2005; Langner, Albert, & Briede, 2002; Shofner, 2008; Winter, Wiegrebe, & Patterson, 2001), we were interested in understanding pitch processing for a nonhuman mammal, particularly in light of possible differences in cochlear tuning between humans and nonhuman mammals (Joris et al., 2011; Oxenham & Shera, 2003; Shera, Guinan, & Oxenham, 2002, 2010). Although these differences are controversial (Eustaquio-Martín & Lopez-Poveda, 2011; Ruggero & Temchin, 2005; Siegel et al., 2005), it is predicted that broader cochlear tuning in a nonhuman mammal will degrade spectral processing.

Pitch perception in chinchillas was studied using a stimulus generalization paradigm. In a stimulus generalization paradigm, an animal is presented with a standard stimulus and is trained to respond to a signal stimulus. Stimuli vary systematically along one or more stimulus dimensions (Malott & Malott, 1970), and a systematic change in behavioral response along the physical dimension of the stimulus is known as a generalization gradient. In the present study, standard stimuli were broadband noises, signal stimuli were either wHTCs or IIRNs, and test stimuli consisted of noise-vocoded wHTCs. Test stimuli that evoke similar behavioral responses as either the signal or the standard indicate a similarity (Guttman, 1963) or perceptual equivalence (Hulse, 1995) between the stimuli. If pitch strength or saliency is based on spectral peak-to-valley ratio, then noise-vocoded wHTCs having strong harmonic structures but no temporal structures should evoke similar behavioral responses as the wHTC or IIRN signals, and as the number of analysis/resynthesis channels decreases, a systematic decrease in behavioral responses should occur as the harmonic structure degrades.

Method

The procedures were approved by the Institutional Animal Care and Use Committee for the Bloomington Campus of Indiana University.

Subjects

Four adult, male chinchillas (*Chinchilla laniger*) served as subjects in these experiments. Each chinchilla (c12, c15, c24, c36) had previous experience in the behavioral paradigm and served as subjects in a missing-fundamental study (Shofner, 2011). Chin-

chillas c12, c15, and c24 began testing at the same time and were tested in all four experiments; c36 was added to the study at a later time and was tested in the first two experiments only. Given the small variability in responses between the three chinchillas tested in Experiments 3 and 4, it was not necessary to include c36 in these conditions. Chinchillas received food pellet rewards during behavioral testing, and their body weights were maintained between 80% and 90% of their normal weight. All chinchillas were in good health during the period of data collection.

Acoustic Stimuli

Stimulus presentation and data acquisition were under the control of a Gateway computer and Tucker–Davis Technologies System II modules. Stimuli were played through a D/A converter at conversion rate of 50 kHz and low-pass filtered at 15 kHz. The output of the low-pass filter was amplified, attenuated, and played through a loudspeaker. All stimuli had durations of 500 ms with 10-ms rise/fall times. The sound pressure level (SPL) was determined by placing a condenser microphone at the approximate position of a chinchilla's head and measuring the A-weighted SPL with a sound spectrum analyzer (Ivie IE-33); sound level was fixed at 73 dBA for all sounds. The frequency response of the system varied ± 8 dB from 100 Hz to 10000 Hz (see Figure 2).

The wHTCs were generated on a digital array processor at a sampling rate of 50 kHz and stored as 16-bit integer files. Tone complexes were composed of a fundamental frequency (F0) of either 250 Hz or 500 Hz and each successive harmonic up to 10 kHz. Harmonics were of equal amplitude and added in sine-starting phase. Noise-vocoded versions of the wHTCs were generated using Tiger CIS Version 1.05.02 developed by Qian-Jie Fu (http://tigerspeech.com). The original 16-bit integer files were first converted to wav files with a sampling rate of 44.1 kHz using Cool Edit 2000 and then processed through the vocoder. The vocoder first analyzed the wHTC through a series of bandpass filters from 200 to 7000 Hz. Default filter slopes of 24 dB/octave and center frequencies based on the Greenwood function were used. The number of channels varied from one to 128, and the carrier type of the vocoder was set to white noise. The present study used noise-vocoded wHTCs based on one, eight, 16, 32, 64, and 128 channels. The envelope from each channel was extracted using a low-pass cutoff frequency of 160 Hz for the 500-Hz F0 wHTC and 80 Hz for the 250-Hz F0 wHTC. The vocoded versions of the wHTCs were saved as wav files and reconverted to 16-bit integer files at a sampling rate of 50 kHz using Cool Edit 2000. Details of the generation of IIRNs and cosine noise (CosN) have been described previously (Shofner, Whitmer, & Yost, 2005). CosN was generated by delaying a wideband noise and adding the unattenuated, delayed noise to the original version of the noise. Thus, CosN was generated by applying the delay-and-add process once. For IIRN, the delayed version of the noise was attenuated (−1 dB or −4 dB) and then added (+) to the original noise through a positive feedback loop. Adding the delayed noise through a positive feedback loop is equivalent mathematically to applying the delay-and-add process through an infinite number of iterations. IIRN stimuli are referred to as IIRN[+, *d*, dB atten] where *d* is the delay and dB atten is the delayed noise attenuation. Delays were fixed at 2 ms (500 Hz pitch) or 4 ms (250 Hz pitch). For each rippled noise, 5 s of the waveform were sampled at 50 kHz and stored as 16-bit integer files. A random 500-ms sample was extracted from the 5-s stimulus files to use for presentation in a block of trials during the behavioral testing session.

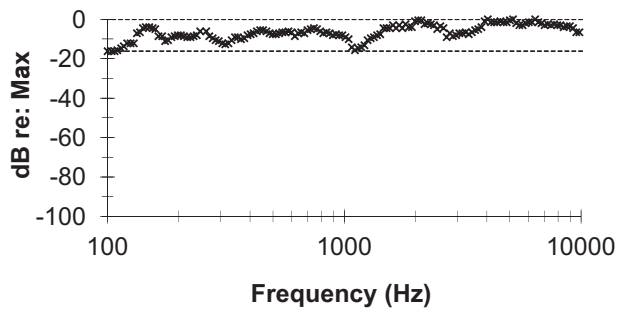


Figure 2. Frequency response of the acoustic system measured using a 40-μs click. The condenser microphone was placed in the approximate position of a chinchilla's head and the output of the loudspeaker was analyzed by measuring the C-weighted sound level for 219 frequencies (×s) between 21 and 19957 Hz with an Ivie IE-33 spectrum analyzer. The horizontal lines show that the relative frequency response varies ±8 dB over a frequency range of 100–10000 Hz.

Spectra for one-, 64-, and 128-channel noise-vocoded 500 Hz F0 wHTCs are illustrated in Figure 3A–3C. Note that the one-channel version is essentially a broadband noise. Both the 64-channel and the 128-channel noise-vocoded wHTCs show strong harmonic structures over harmonics 1–10, but the peak-to-valley ratios appear smaller than that of the wHTC (see Figure 1A). The peak-to valley ratios used in the present study were estimated from the acoustic spectra. A condenser microphone was placed in the approximate position of a chinchilla's head and the output of the loudspeaker was displayed as a spectrum on an Ivie IE-33 spectrum analyzer for 219 frequencies between 21 and 19957 Hz using the C-weight sound level. Harmonic structure was quantified by computing the average peak-to-valley ratio for harmonics 1–10 as

$$Ave.\ \ P\text{-}V(dB) = \frac{\left|dB_1 - dB_{1.5}\right| + \left|dB_{1.5} - dB_2\right| + \cdots + \left|dB_{10} - dB_{10.5}\right|}{19}, \quad (1)$$

where *Ave. P-V* (*dB*) is the average peak-to-valley ratio expressed as a decibel and dB_N are the spectral amplitudes at the specified harmonic number, *N*. The heights and number of peaks in the ACF of the vocoded wHTCs (see Figure 3D–3F) are reduced relative to those of the wHTC (see Figure 1C), indicating that the periodicity strengths of the vocoded wHTCs are less than that of the wHTC. Periodicity strength was quantified from the first 20 ms of the ACF and was based on the sum of crest factors of individual peaks. The standard deviation of the ACF was first computed over the time lag between 0.04 and 20 ms. A peak in the ACF was considered significant if the height was greater than 2 standard deviations. Periodicity strength (PS) was defined as

$$PS = \sum_{i=1}^{n} \frac{AC_i}{\sigma_{ACF}}, \quad (2)$$

where AC_i is the height of a significant peak, σ_{ACF} is the standard deviation of the ACF, and *n* is the number of significant peaks in

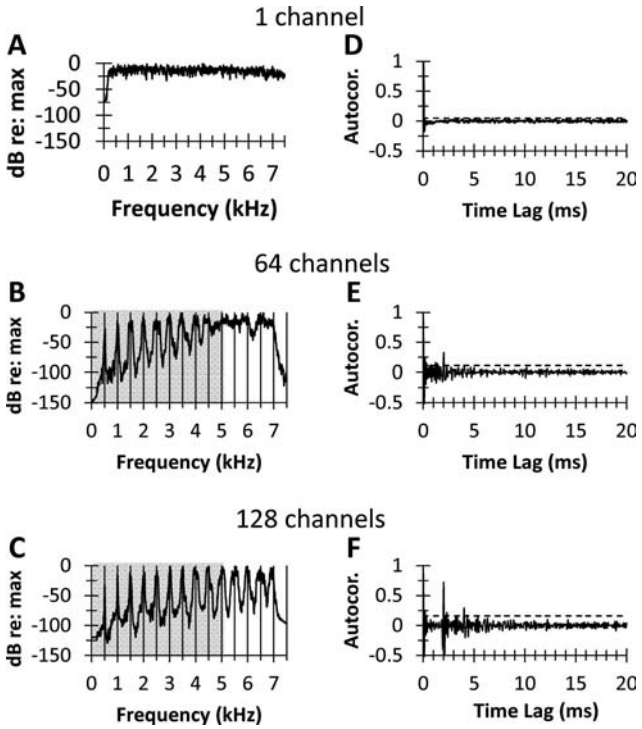


Figure 3. Example spectra (A–C) and autocorrelation functions (D–F) illustrate harmonic structure and periodicity for noise-vocoded versions of the 500-Hz harmonic tone complex. The number of analysis/resynthesis channels used in the vocoding process are one (A, D), 64 (B, E), and 128 (C, F). The gray shaded areas indicate the frequencies corresponding to harmonics 1–10. Vertical lines indicate the frequencies of the harmonics. The horizontal dashed lines indicate 2 standard deviations. Spectra have been smoothed using a seven-bin moving window average.

the ACF between 0.04 and 20 ms. Periodicity strength was defined in this manner in order to have an estimate based on both the heights and number of peaks. If no significant peaks in the ACF were found, then periodicity strength was zero. Table 1 summarizes the average peak-to-valley ratios and periodicity strengths for stimuli used in the present study.

Behavioral Procedure

The testing cage had dimensions of 24 in. width × 24 in. length × 14 in. high; chinchillas were free to roam around the cage and were not restrained in any way. The cage was located on a card table in a single-walled sound-attenuating chamber having internal dimensions of 5 ft 2 in. width × 5 ft 2 in. length × 6 ft 6 in. high. A pellet dispenser was located at one end of the cage with a reward chute attached to a response lever. A loudspeaker was located next to the pellet dispenser approximately 30° to the right of center at a distance of 6 in. in front of the chinchilla.

The training procedures (Shofner, 2002) and stimulus generalization procedures (Shofner, 2002, 2011; Shofner & Whitmer, 2006; Shofner et al., 2005, 2007) have been described previously in detail. Briefly, a standard sound was presented continually in 500-ms bursts at a rate of once per second, regardless of whether or not a trial was initiated. The standard was a one-channel noise-vocoded version of a wHTC. Chinchillas initiated a trial by pressing down on the response lever. The lever had to be depressed for a duration that varied randomly for each trial ranging from 1.15 to 8.15 s for three chinchillas and from 1.15 to 6.15 s for the fourth chinchilla. After the lever was depressed for the required duration, two 500-ms bursts of a selected sound were presented for that trial. The response window was coincident with the duration of the two 500-ms bursts (2,000 ms), except that the response window began 150 ms after the onset of the first burst and lasted until the onset of the next burst of the continual standard stimulus. Thus, the actual duration of the response window was 1,850 ms.

The selected sounds presented during the response window could be signals, test sounds, or standards. A signal trial consisted of two bursts of the signal sound, which were either wHTCs or IIRNs depending on the specific experiment. If the chinchilla released the lever during the response window of a signal trial, then this positive response was treated as a hit and the chinchilla was rewarded with a food pellet. A standard trial consisted of two additional bursts of the one-channel noise-vocoded wHTC. If the chinchilla continued to depress the lever throughout the response window of a standard trial, then this negative response was treated as a correct rejection. Food pellet rewards for correct rejections were generally not necessary to reinforce correct rejections. A lever release during a standard trial was treated as a false alarm. Because false alarm rates were well below 20%, "time outs" following a false alarm were not necessary. A test trial consisted of two bursts of a test sound, which were generally eight- to 128-channel noise-vocoded versions of the wHTCs, although rippled noises were occasionally used as well. Chinchillas did not receive food pellet rewards for responses to test stimuli, regardless of whether the behavioral response was positive or negative.

Chinchillas were tested in blocks consisting of 40 trials. Two different test sounds were presented in a block of trials. In each block, 60% of the trials were signal trials, 20% were standard trials, 10% were Test Sound 1 trials, and 10% were Test Sound 2 trials. Thus, test stimuli were presented infrequently in the block of trials. Behavioral responses were considered to be under stimulus control if the percentage correct for the discrimination of the signal from the standard was at least 81% for each block. Responses were collected for a minimum of 50 blocks (i.e., 2,000 total trials) resulting in at least 200 trials for each test sound.

Results

Experiment 1: wHTCs as Signals

Chinchillas were trained to discriminate a wHTC from a one-channel noise-vocoded version of the wHTC. Behavioral responses were the number of lever releases relative to the number of trials expressed as a percentage (see Figure 4). The responses obtained from four chinchillas to the 500-Hz F0 wHTC signal were high, ranging from 92% to 97%, whereas the responses to the one-channel noise-vocoded version of the wHTC were low, ranging from 2% to 11% (see Figure 4A). That is, chinchillas can discriminate the wHTC from the one-channel noise vocoded version; d's estimated as the differences between z(hits) and z(false alarms) ranged from 2.7 to 4. When tested with vocoded versions of the wHTC using eight to 128 channels, the responses of the chinchillas were low and were similar to those of the one-channel

Table 1
Average Spectral Peak-to-Valley Ratios (P-V) for Harmonics 1–10

	500-Hz F0			250-Hz F0		
Stimulus	P-V (dB)	PS re: PS(wHTC)	AC1	P-V (dB)	PS re: PS(wHTC)	AC1
wHTC	46.2	1.00	0.994	41.6	1.00	0.988
Filtered wHTC	49.4	0.85	0.995	44	0.85	0.989
IIRN (–1 dB)	17.6	0.63	0.785	23.3	0.74	0.825
IIRN (–4 dB)	10.5	0.44	0.544	12.2	0.65	0.571
IIRN + HPN	12.5 (16.2)	0.60	0.384			
CosN	12.8 (15.1)	0.42	0.485			
128 channels	38.7	0.20	0.725	27.3	0.13	0.35
64 channels	25.7	0.09	0.328	20	0.00	0.074
32 channels	15.8 (23.5)	0.04	0.112	11.9	0.00	0.039
16 channels	7.1	0.00	0.042	4.6	0.00	0.03
8 channels	2.9	0.00	0.0007	3.6	0.00	0.026
1 channel	3.1 (1.7)	0.00	0.0013	2.3	0.00	0.023

Note. F0 = fundamental frequency; wHTC = wideband harmonic tone complex; Filtered wHTC = wideband harmonic tone complex bandpass filtered from 200 to 7000 Hz; IIRN (–1 dB) = infinitely iterated ripped noise where delayed noise attenuation was –1 dB; IIRN (–4 dB) = infinitely iterated ripped noise where delayed noise attenuation was –4 dB; IIRN + HPN = IIRN –1 dB added to high-pass noise having a 3-kHz upper cutoff frequency; CosN = cosine noise (i.e. one delay-and-add iteration of rippled noise); 128 channels–1 channel: noise vocoded versions of the wHTCs with specified number of channels. Numbers in parentheses indicate the average peak-to-valley ratios for harmonics 1–5 only. Periodicity strengths normalized to the periodicity strength for the wHTC (PS re: PS(wHTC)) for the first 20 ms of the autocorrelation functions (ACFs) and height of the first peak in the autocorrelation function (AC1) for the 500-Hz F0 and 250-Hz F0 stimuli used in the present study for predictions of behavioral responses.

version (see Figure 4A and 4B). None of the chinchillas tested gave large responses to the 128-channel noise-vocoded wHTC in spite of its large spectral peak-to-valley ratio, which is closer to that of the wHTC than to the peak-to-valley ratio of the one-channel version (see Table 1). A repeated measures analysis of variance (ANOVA) showed a significant effect of stimuli for the 500-Hz condition, $F(7, 21) = 355.2$, $p \ll .0001$. Pairwise comparisons based on Tukey's test showed that there was not a significant difference between the one-channel and the 128-channel versions ($q = 3.22$, $p > .05$), but there was a significant difference between the one-channel version and the unmodified wHTC ($q = 41.8$, $p \ll .001$). In addition to the vocoding process, the software for the noise vocoder also bandpass filters the stimulus from 200 to 7000 Hz. To test whether the bandpass filtering itself affected the behavioral responses, we also tested chinchillas with a filtered version of the wHTC that was not subjected to the vocoding process. Pairwise comparisons based on Tukey's test showed that there was not a significant difference between the filtered wHTC and unmodified wHTC ($q = 0.14$, $p > .05$), but there was a significant difference between the one-channel version and the filtered wHTC ($q = 41.7$, $p \ll .001$). Similar behavioral responses were obtained when the F0 was 250 Hz (see Figure 4C and 4D). A repeated measures ANOVA showed a significant effect of stimuli for the 250-Hz condition, $F(7, 21) = 434.5$, $p \ll .0001$. Pairwise comparisons based on Tukey's test showed that there was not a significant difference between the one-channel and the 128-channel versions ($q = 4.00$, $p > .05$), and there was not a significant difference between the filtered wHTC and unmodified wHTC ($q = -0.14$, $p > .05$). There were significant differences between the one-channel version and the unmodified wHTC ($q = 46.1$, $p \ll .001$) as well as between the one-channel version and the filtered wHTC ($q = 46.3$, $p \ll .001$).

These results suggest that the filtered wHTC is generalized to the wHTC, whereas the eight- to 128-channel noise-vocoded versions of the wHTC are generalized to the one-channel version of the wHTC. Because the one-channel version is a broadband noise, then the eight- to 128-channel versions of the wHTC are essentially generalized to wideband noise. If spectral peak-to-valley ratio is the acoustic cue controlling the behavioral responses, then a systematic decrease in behavioral response should be observed as the number of channels decreases (see dashed black line and open circles in Figure 4B and 4D). Clearly, the behavioral and predicted functions differ greatly. It should be noted that for these and all subsequent predictions based on the acoustic features obtained from the stimuli, namely peak-to-valley ratio and periodicity strength (see Table 1), it is assumed that the behavioral responses are proportional to the acoustic measures. However, for the predictions based on the height of the first peak in the autocorrelation function (AC1), it is assumed that the responses are proportional to $10 + 10^{(2AC1)}$ (Yost, 1996). That the responses to the eight- to 128-channel versions were similar to the response for the one-channel version suggests that these stimuli all share a common acoustic feature, such as little or no periodicity strength (see Table 1). The behavioral responses obtained can be accounted for if the chinchillas were using periodicity strength as the acoustic cue (see gray solid line and gray diamonds in Figure 4B and 4 D). Note that the predictions based on AC1 (see gray dashed line and open squares in Figure 4B and 4D) do not account for the behavioral responses as well as periodicity strength. The sum of squares deviation between the data and predictions for the 500-Hz condition is 290.3 when periodicity strength is used, but it is 541.7 when AC1 is used. For the 250-Hz condition, these values are 392.4 and 401.6, respectively. These findings suggest that the behavioral responses are not controlled by the spectral structure but may be controlled by the temporal structure.

Experiment 2: IIRNs as Signals

To further test the hypothesis that the behavioral responses were not controlled by the spectral peak-to-valley ratio, we replaced

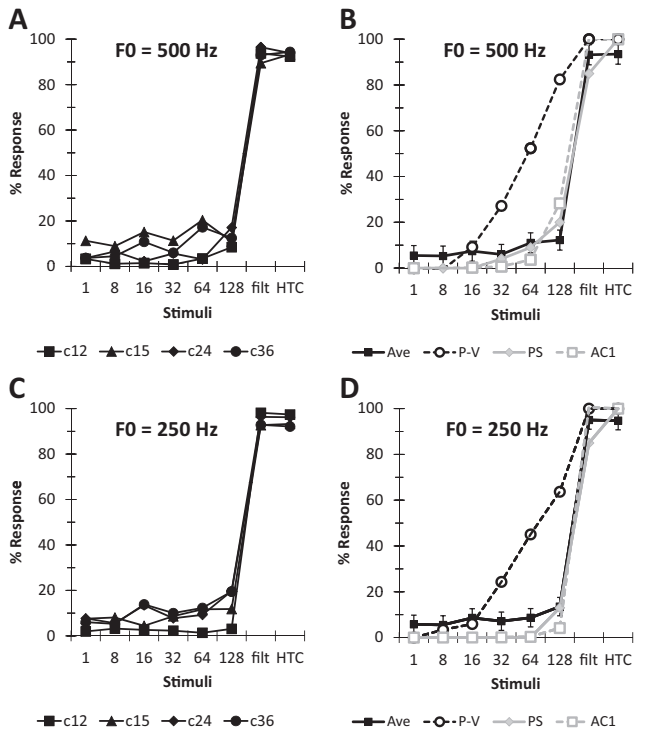


Figure 4. Behavioral responses obtained from the stimulus generalization task for four chinchillas using the unmodified wideband harmonic tone complex (wHTC) as the signal. (A) Responses obtained from individual chinchillas (c12, c15, c24, c36) are shown. Fundamental frequency (F0) of the wHTC is 500 Hz. (B) Averaged behavioral responses from the four chinchillas for the 500-Hz condition (black solid squares and line). Error bars indicate the 95% confidence intervals based on Tukey's standard error. The black dashed line and opened circles show the predictions based on the average peak-to-valley ratios (P-V) from the spectra. The gray solid line and gray diamonds show the predictions based on the periodicity strengths (PS) from the autocorrelation functions (ACFs). The gray dashed line and open squares show the predictions based on AC1. (C) Responses obtained from individual chinchillas (c12, c15, c24, c36) are shown. Fundamental frequency (F0) of the wHTC is 250 Hz. (D) Same as B for the 250-Hz F0 condition. In all panels, numbers along the *x*-axis indicate the number of analysis/resynthesis channels of the vocoded sounds; filt and HTC indicate the filtered only and unmodified harmonic tone complexes, respectively. The form of the predictions is (test – 1 channel)/(HTC – 1 channel).

wHTC signals with IIRN[+, *d*, −1 dB]. These IIRNs have average peak-to-valley ratios that are smaller than the 128-channel vocoded wHTCs (see Table 1) for both 500-Hz and 250-Hz F0s. Figure 5 shows responses obtained from four chinchillas to vocoded wHTCs when IIRN[+, *d*, −1 dB] was used as the signal for delays of 2 ms and 4 ms. The behavioral responses to the IIRN[2 ms, −1 dB] (see Figure 5A and 5B) and IIRN[4 ms, −1 dB] (see Figure 5C and 5D) were high, ranging from 94% to 98%, whereas the responses to the one-channel noise-vocoded version of the wHTC were low, ranging from 1.5% to 11%. That is, chinchillas can discriminate these IIRNs from the one-channel noise-vocoded versions with *d*′s ranging from 2.9 to 3.8. When tested with vocoded versions of the wHTCs using eight to 128 channels, the responses of the chinchillas were low and again were similar to those of the one-channel version. It should be noted that for the 2-ms condition, the responses of two of the four chinchillas were around 30–44% to the 128-channel noise-vocoded wHTC and were above those of the one-channel version (see Figure 5A), whereas the responses of the other two chinchillas to the 128-channel version were essentially identical to those of the one-channel version for the 2-ms condition. For the 4-ms condition, all of the chinchillas gave low responses to the 128-channel vocoded HTC that were similar to the one-channel version (see Figure 5C). A repeated measures ANOVA showed a significant effect of stimuli for the 2-ms delay condition, $F(7, 21) = 125.3$, $p \ll .0001$, and for the 4-ms delay condition, $F(7, 21) = 1112.5$, $p \ll .001$. Pairwise comparisons between the one-channel and 128-channel versions showed no significant difference for the 2-ms condition

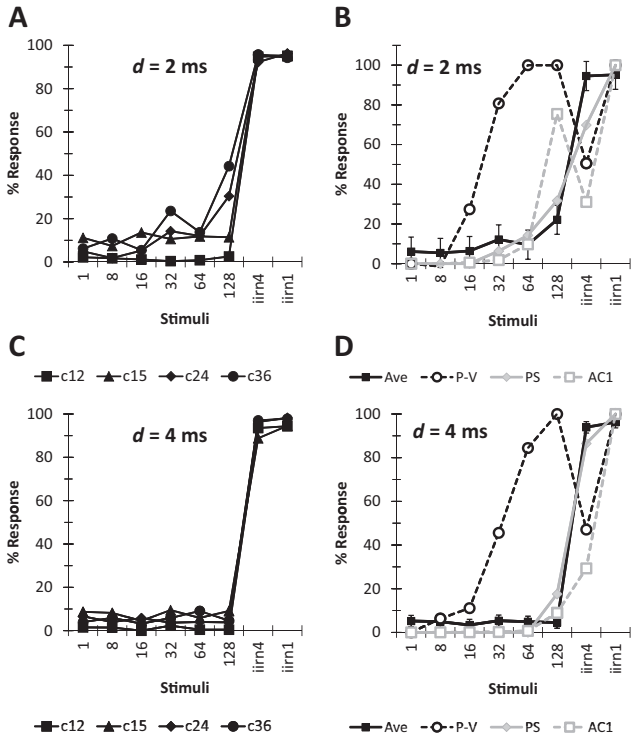


Figure 5. Behavioral responses obtained from the stimulus generalization task for four chinchillas using the infinitely iterated rippled noise (IIRN) [+, *d*, –1 dB] as the signal. (A) Responses obtained from individual chinchillas (c12, c15, c24, c36) are shown. The delay of the IIRN is 2 ms. (B) Averaged behavioral responses from the four chinchillas for the 2-ms condition (black solid squares and line). Error bars indicate the 95% confidence intervals based on Tukey's standard error. The black dashed line and opened circles show the predictions based on the average peak-to-valley ratios (P-V) from the spectra. The gray solid line and gray diamonds show the predictions based on the periodicity strengths (PS) from the autocorrelation functions (ACFs). The gray dashed line and open squares show the predictions based on AC1. (C) Responses obtained from individual chinchillas (c12, c15, c24, c36) are shown. The delay of the IIRN is 4 ms. (D) Same as B for the 4-ms delay condition. In all panels, numbers along the *x*-axis indicate the number of analysis/resynthesis channels of the vocoded sounds; iirn4 and iirn1 indicate the IIRNs having delayed noise attenuations of –4 dB and –1 dB, respectively. The form of the predictions is (test – 1 channel)/(iirn1 – 1 channel).

($q = 4.55$, $p > .05$) and for the 4-ms condition ($q = 0.54$, $p > .05$), but there were significant differences between the one-channel version and the signal IIRN with −1 dB for both the 2-ms ($q = 25.2$, $p < .001$) and 4-ms conditions ($q = 72.6$, $p \ll .001$). Chinchillas were also tested with IIRNs having −4 dB delayed noise attenuation. Pairwise comparisons based on Tukey's test showed that there was no significant difference between the responses to the test IIRNs with −4 dB and the responses to the signal IIRNs with −1 dB for both the 2-ms ($q = 0.18$, $p > .05$) and 4-ms delay conditions ($q = 1.85$, $p > .05$). Pairwise comparisons showed that there was a significant difference between responses to the test IIRNs with −4 dB and those obtained for the 128-channel vocoded versions for the 2-ms ($q = 20.5$, $p < .001$) and 4-ms conditions ($q = 71.3$, $p \ll .001$).

These results suggest that the eight- to 128-channel noise-vocoded versions of the wHTCs were not generalized to the IIRN[+, *d*, −1 dB] for both 2-ms and 4-ms delays, but rather are generalized to the one-channel versions of the wHTCs. For the 2-ms condition, the 128-channel version was generalized to be intermediate between the IIRN signal and the one-channel version of the wHTC (see Figure 5A) for two of the four chinchillas, which may be related to the weak periodicity strength of 0.2 for the 128-channel version. For the 4-ms condition, the 128-channel version was generalized to the one-channel version by all four chinchillas. The predicted responses that would be obtained if the chinchillas were using the average peak-to-valley ratio over harmonics 1–10 for the 2-ms and 4-ms conditions do not account for the obtained behavioral responses (see black dashed line and open circles in Figure 5). Also, the responses to the test IIRNs with −4 dB were larger than those obtained for either the 128-channel or 64-channel vocoded wHTCs in spite of the larger peak-to-valley ratios of these two vocoded wHTCs (see Table 1). The behavioral responses can be accounted for if the chinchillas were using periodicity strength as the acoustic cue (see gray solid line and gray diamonds in Figure 5B and 5D); again, it should be noted that the predictions based on AC1 do not account for the behavioral responses as well as periodicity strength. The sum of squares deviation between the data and predictions for the 2-ms delay condition is 889.7 when periodicity strength is used, but is 7095.3 when AC1 is used. For the 4-ms delay condition, these values are 354.4 and 4312.3, respectively. These results again suggest that the behavioral responses are controlled by temporal structure rather than spectral structure.

Experiment 3: Testing Spectral Shape

As the number of analysis/resynthesis channels decreases, there is a change in the overall spectral shape of the vocoded wHTC such that there can be a strong harmonic structure at low frequencies, but at higher frequencies, the spectrum shows characteristics of a flat-spectrum noise (compare Figure 6A with Figure 3C). To test the influence that overall spectral shape may have on controlling the behavioral responses, we added IIRN[+, 2 ms, −1 dB] to a high-pass noise (HPN) having a low cutoff frequency of 3 kHz. This produced a sound having a harmonic structure at low frequencies and a flatter spectrum at higher frequencies (see Figure 6B) similar to the 32-channel vocoded wHTC (see Figure 6A). Three chinchillas discriminated the IIRN + HPN signal from the one-channel noise-vocoded wHTC standard. Chinchillas were tested with the 32-channel noise-vocoded version as well as with CosN (see Figure 6C). For this experiment, the signal and both test sounds all have similar average peak-to-valley ratios (see Table 1). Figure 7 shows responses obtained from three chinchillas to the 32-channel vocoded wHTC and CosN when IIRN + HPN was the signal. A repeated measures ANOVA showed a significant effect of stimuli, $F(3, 6) = 320.5$, $p \ll .0001$. Pairwise comparisons based on Tukey's test showed that there was not a significant difference between responses for the one-channel and 32-channel versions ($q = 0.32$, $p > .05$) or between CosN and IIRN + HPN ($q = 0.40$, $p > .05$). There was a significant difference between the responses for the 32-channel version and CosN ($q = 31.4$, $p < .001$).

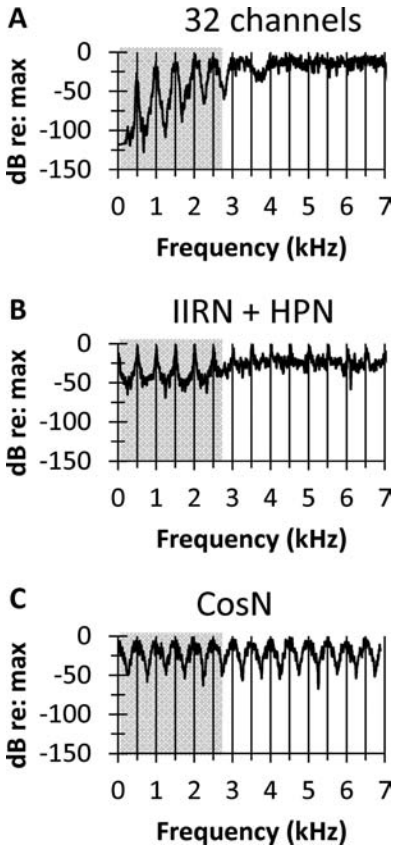


Figure 6. Spectra for (A) the 32-channel noise-vocoded wideband harmonic tone complex (wHTC), (B) infinitely iterated rippled noise (IIRN) [+, 2 ms, –1 dB] added to a high-pass noise (HPN) having a lower cutoff frequency of 3000 Hz, and (C) cosine noise (CosN). The gray shaded areas indicate the frequencies corresponding to harmonics 1–5. Spectra have been smoothed using a seven-bin moving window average. Vertical lines indicate the frequencies of the harmonics. The delays used to generate the IIRN and CosN was 2 ms.

If average peak-to-valley ratio is the acoustic cue that controls the behavioral responses, then it would be expected that the chinchillas would generalize across all three of these stimuli (see black dashed line and open circles in Figure 7B). Clearly, Figure 7B and the above statistical analysis show that the chinchillas did not respond equally to the test sounds and the signal. The responses obtained suggest that average peak-to-valley ratio does not control

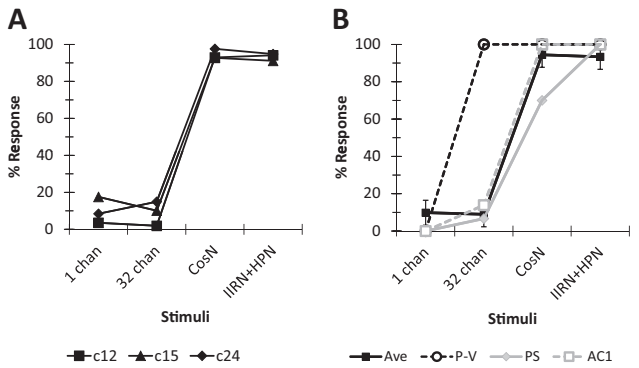


Figure 7. Behavioral responses from three chinchillas using the stimuli illustrated in Figure 6. (A) Responses obtained from individual chinchillas (c12, c15, c24, c36) are shown. (B) Averaged behavioral responses from the four chinchillas for the 2-ms condition (black solid squares and line). Error bars indicate the 95% confidence intervals based on Tukey's standard error. The black dashed line and opened circles show the predictions based on the average peak-to-valley ratios (P-V) from the spectra. The gray solid line and gray diamonds show the predictions based on the periodicity strengths (PS) from the autocorrelation functions (ACFs). The gray dashed line and open squares show the predictions based on AC1. The form of the predictions is (test – 1 channel)/(IIRN+HPN – 1 channel). CosN = cosine noise; IIRN = infinitely iterated rippled noise; HPN = high-pass noise.

the behavioral response (see Figure 7B). It is interesting to note that if behavioral performance reflected the average peak-to-valley ratios for only harmonics 1–5 where clear peaks in the spectra are observed (see Figure 6A and 6B), then the responses to the 32-channel noise-vocoded wHTC should be high given that the peak-to-valley ratio for this stimulus is larger than either the IIRN + HPN or CosN stimuli (see values in parentheses in Table 1). If overall spectral shape is the acoustic cue, then it would be expected that the 32-channel vocoded wHTC would be generalized to the IIRN + HPN, but the CosN would not be generalized to the IIRN + HPN because the spectrum of CosN shows ripples at all harmonics and does not show the characteristics of a flat-spectrum noise above 3 kHz. In contrast to these predictions, the behavioral responses show that the responses to the 32-channel noise-vocoded wHTC were generalized to the one-channel standard and not to the IIRN + HPN signal even though the 32-channel version and the IIRN + HPN had similar overall spectral shapes. Also, the responses to the CosN were generalized to the IIRN + HPN signal, again a result not predicted by spectral shape.

The IIRN + HPN and CosN both have larger periodicity strengths than either the one-channel or 32-channel versions of the wHTC (see Table 1). The behavioral responses obtained can be accounted for if the chinchillas are using periodicity strength as the acoustic cue (see gray solid line and gray diamonds in Figure 7B). It is interesting to note that in this case behavioral responses can be better accounted for if the chinchillas are using AC1 as the acoustic cue (see gray dashed line and open squares in Figure 7B). The sum of squares deviation between the data and predictions is 742.6 when periodicity strength is used, but it is 196.4 when AC1 is used. These results again suggest that the behavioral responses are controlled by temporal structure rather than spectral structure.

Experiment 4: Discrimination of the 128-Channel Version From the One-Channel Version

The results of Experiments 1 and 2 indicate that the 128-channel noise-vocoded versions of wHTCs are generalized to broadband noise in the chinchilla. Stimuli are generalized (i.e., perceptually equivalent) when chinchillas cannot discriminate between the stimuli (Guttman & Kalish, 1956). Given the large spectral differences between the 128-channel and one-channel noise-vocoded wHTCs (see Table 1), can chinchillas discriminate between these two stimuli? Three chinchillas were tested in a discrimination task in which they now received food pellet rewards for positive responses to the 128-channel version (i.e., hits). In blocks of 40 trials, the 128-channel version of the 500-Hz F0 wHTC was presented on 32 trials and the one-channel version on eight trials. Hits and false alarms were measured for each block, and the d's were estimated for each block and over a total of 1,200 trials (30 blocks) for two chinchillas and 2,120 trials (53 blocks) for a third chinchilla (see Figure 8). If the p(hits) or p(false alarms) were 1.0

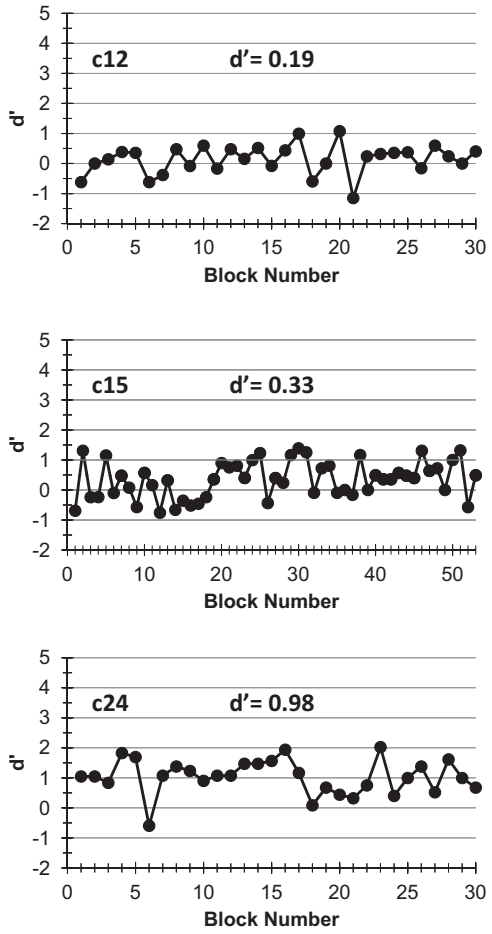


Figure 8. Performance of three chinchillas (c12, c15, c24) obtained for discriminating the one-channel noise-vocoded wideband harmonic tone complex (wHTC) from the 128-channel version. The fundamental frequency of the unmodified wHTC used for the vocoding process was 500 Hz. The d' for each block of 40 trials was measured. The d' shown in each panel is based on the hits and false alarms computed over all blocks.

or 0.0 for a block, then the probability was adjusted as $1 - 1/(2N)$ or $1/(2N)$, respectively, where N is the number of signal or blank trials (Macmillan & Creelman, 2005). One chinchilla could discriminate the stimuli at a threshold level of performance. giving a d' of 0.98 (see c24 in Figure 8), whereas two chinchillas could not discriminate between the stimuli giving d's of 0.19 and 0.33 (see c12 and c15, respectively, in Figure 8). There was no evidence of an improvement in performance over the number of trials completed.

Discussion

The simplest spectral processing scheme for pitch perception is that the pitch percepts of HTCs are determined by the fundamental frequency component, either by its physical presence in the stimulus or by its reintroduction to the peripheral representation through cochlear distortion products. Similar to human listeners, nonhuman mammals appear to have pitch-like perceptions of the "missing fundamental" (Chung & Colavita, 1976; H. Heffner & Whitfield, 1976; Shofner, 2011; Tomlinson & Schwarz, 1988) that do not arise through the reintroduction of cochlear distortion products (Chung & Colavita, 1976; Shofner, 2011). Thus, this simplest form of spectral processing does not appear to exist in the auditory system of nonhuman mammals. An alternative spectral processing scheme would be that the pitch percepts are based on representations of the harmonic spectra within the auditory system (i.e., harmonic template matching). The results of the present study argue against the harmonic template model for the chinchilla. In the present study, chinchillas discriminated either a wHTC or IIRN (i.e., salient "pitch" percept) from a one-channel noise-vocoded wHTC (i.e., noise percept). When tested with eight- to 128-channel noise-vocoded wHTCs, the behavioral responses obtained were not related to the peak-to-valley ratios of the harmonic components of the test stimuli. The results suggest that an auditory representation of the harmonic structure is not controlling the behavioral responses and that spectral processing contributes little to the perception of pitch in chinchillas.

Why is harmonic structure apparently not being processed by the chinchilla auditory system? Consider the representation of the harmonic components along the basilar membrane. Figure 9 illustrates the positions of two harmonics, H1 and H2, along the basilar membrane; each component falls within an idealized auditory filter that is centered at the frequency of the harmonic. The auditory filter bandwidth is given as equivalent rectangular spread (ERS), which expresses frequency in terms of position along the basilar membrane (Shera et al., 2010). Resolvability of the harmonics is expressed in terms of position along the basilar membrane and is defined as the difference between the upper cutoff frequency for the filter centered at H1 (X_{UH1}) and the lower cutoff frequency for the filter centered at H2 (X_{LH2}). If this difference is greater than zero, then the two auditory filters do not overlap in position along the basilar membrane and the harmonics are resolved. If the difference is less than zero, then there is overlap of the filters and the harmonics are not resolved. The ERS for various center frequencies are available (see Figure 16A of Shera et al., 2010) and position along the basilar membrane can be estimated from the Greenwood frequency–position functions (Greenwood, 1990). Resolvability decreases as harmonic number increases for both chinchilla and human cochleae (see Figure 9). For humans, harmonics

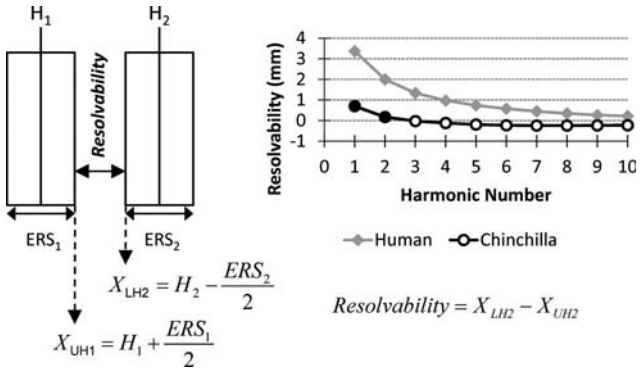


Figure 9. Illustration of auditory filters defining resolvability of harmonic components, H_1 and H_2, along the basilar membrane. The harmonics each fall within separate auditory filters having bandwidths expressed as equivalent rectangular spread (ERS). The upper (X_{UH1}) and lower (X_{LH2}) cutoff frequencies of the two filters are shown by the equations. Resolvability is defined as the difference between these cutoff frequencies. The graph shows resolvability of harmonics 1–10 of a wideband harmonic tone complex (wHTC) for chinchillas (black line and circles) and humans (gray line and diamonds). Filled symbols indicate harmonics for which resolvability > 0.

1–10 are resolved, whereas only harmonics 1 and 2 are resolved in chinchillas. In addition, harmonics 1 and 2 are clearly farther apart in the human cochlea than in the chinchilla cochlea; that is, resolvability of harmonics 1 and 2 is better in humans than in chinchillas. The difference between resolvability in humans and chinchillas reflects the sharper cochlear tuning described in humans (Joris et al., 2011; Oxenham & Shera, 2003; Shera et al., 2002, 2010). Thus, although there can be a large peak-to-valley ratio between harmonics for noise-vocoded wHTCs, the representation of that harmonic structure is highly degraded in the chinchilla cochlea.

The behavioral responses to the vocoded wHTCs do appear to be related to the temporal structure of the ACFs suggesting that pitch perception in chinchillas is largely based on temporal processing. Periodicity strength was a better predictor of the behavioral responses for conditions in which the F0 of the wHTC was 500 Hz, but it was about equal to the AC1 predictions when the F0 of the wHTC was 250 Hz (Experiment 1). For Experiment 2, periodicity strength was the better predictor for conditions using IIRNs for both 2-ms and 4-ms delays, but AC1 was the better predictor in Experiment 3 when the delay was 2 ms and the IIRN signal was combined with the HPN. Thus, although the predictions based on periodicity strength or AC1 are mixed, they argue that an auditory representation based on temporal structure is controlling the behavioral responses in chinchillas. This conclusion is consistent with those of recent studies in gerbils (Klinge & Klump, 2009, 2010) in which the authors argue that the detection of mistuned harmonics arises largely through temporal processing as a consequence of a reduction in spatial selectivity along the shorter cochlea of the gerbil (Klinge, Itatani, & Klump, 2010). However, as previously noted, the spectrum and ACF are Fourier pairs, and consequently, any temporal processing scheme has a potentially viable, complementary spectral processing scheme. Although we argue in favor of temporal processing, we acknowledge that the

behavioral responses obtained from the chinchillas could be based on an auditory representation of the harmonic structure that is highly degraded because of broader cochlear tuning. Moreover, any temporal representation is likely to be converted into a place code in the central auditory system (e.g., Bendor & Wang, 2005; Langner et al., 2002).

The analysis described above for resolvability suggests that there may indeed be a greater role for spectral processing in humans than in nonhuman mammals, and it will be important to obtain behavioral responses to vocoded wHTCs in human listeners to clarify this issue. For example, because different complex sounds can evoke differences in pitch strength or saliency (Fastl & Stoll, 1979; Shofner & Selas, 2002; Yost, 1996), then the pitch strength of vocoded wHTCs should decrease systematically as the number of analysis/synthesis channels used by the vocoder is reduced if pitch is extracted through spectral processing. Moreover, it is interesting to note that Experiment 4 demonstrated poor performance in chinchillas for discriminating the 128-channel noise-vocoded wHTC from the one-channel version (i.e., $d' \leq 1.0$), but casual listening suggests that these stimuli are easily discriminated. Performance was measured for 200 trials by the first author and p(hits) and p(false alarms) obtained were 1.0 and 0.0, respectively (i.e., infinite d').

In conclusion, we argue that temporal processing can account for behavioral responses in chinchillas, suggesting that temporal processing of pitch is the more primitive processing mechanism common across mammalian species, including humans. To account for the behavioral responses based on spectral processing, then it must be assumed that cochlear tuning in nonhuman mammals is broader than tuning in humans as suggested by the resolvability analysis previously described. Our results along with those of Klinge et al. (2010) imply that spectral processing of pitch in nonhuman mammals is presumably less developed because of broader cochlear tuning and shorter cochleae; thus, pitch perception in nonhuman mammals must arise largely through temporal processing. As the modern human cochlea lengthened and tuning sharpened through evolution, the foundation was laid for additional neural mechanisms of pitch extraction in the spectral domain to evolve.

References

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125,* 238–251. doi:10.1093/brain/awf028

Balaguer-Ballester, E., Denham, S. L., & Meddis, R. (2008). A cascade autocorrelation model of pitch perception. *Journal of the Acoustical Society of America, 124,* 2186–2195. doi:10.1121/1.2967829

Bendor, D., & Wang, X. (2005, August 25). The neuronal representation of pitch in primate auditory cortex. *Nature, 436,* 1161–1165. doi:10.1038/nature03867

Burns. E. M., & Viemeister, N. F. (1976). Nonspectral pitch. *Journal of the Acoustical Society of America, 60,* 863–869. doi:10.1121/1.381166

Burns. E. M., & Viemeister, N. F. (1981). Played-again SAM: Further observations on the pitch of amplitude-modulated noise. *Journal of the Acoustical Society of America, 70,* 1655–1660. doi:10.1121/1.387220

Chung, D. Y., & Colavita, F. B. (1976). Periodicity pitch perception and its upper frequency limit in cats. *Perception & Psychophysics, 20,* 433–437. doi:10.3758/BF03208278

Cohen, M. A., Grossberg, S., & Wyse, L. L. (1995). A spectral network model of pitch perception. *Journal of the Acoustical Society of America, 98,* 862–879. doi:10.1121/1.413512

Cousineau, M., Demany, L., & Pressnitzer, D. (2009). What makes a melody: The perceptual singularity of pitch sequences. *Journal Acoustical Society of America, 126,* 3179–3187. doi:10.1121/1.3257206

Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics, 59,* 165–179. doi:10.3758/BF03211886

Darwin, C. J. (2005). Pitch and auditory grouping. In C. J. Plack, A. J. Oxenham, R. R. Fay, & A. N. Popper (Eds.), *Pitch: Neural coding and perception* (pp. 278–305). New York, NY: Springer.

de Cheveigné, A. (1998). Cancellation model of pitch perception. *Journal of the Acoustical Society of America, 103,* 1261–1271. doi:10.1121/1.423232

Dorman, M. F., Loizou, P. C., & Rainey, D. (1997). Speech intelligibility as a function of the number of channels of stimulation for signal processors using sine-wave and noise-band outputs. *Journal of the Acoustical Society of America, 102,* 2403–2411. doi:10.1121/1.419603

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49,* 524–531. doi:10.1121/1.1912382

Dudley, H. (1939). Remaking speech. *Journal of the Acoustical Society of America, 11,* 169–177. doi:10.1121/1.1916020

Eustaquio-Martín, A., & Lopez-Poveda, E. (2011). Isoresponse versus isoinput estimates of cochlear tuning. *Journal of the Association for Research in Otolaryngology, 12,* 281–299. doi:10.1007/s10162-010-0252-1

Fastl, H., & Stoll, G. (1979). Scaling of pitch strength. *Hearing Research, 1,* 293–301. doi:10.1016/0378-5955(79)90002-9

Friesen, L., Shannon, R. V., Baskent, D., & Wang, X. (2001). Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants. *Journal of the Acoustical Society of America, 110,* 1150–1163. doi:10.1121/1.1381538

Gauthier, B., & Shi, R. (2011). A connectionist study on the role of pitch in infant-directed speech. *Journal of the Acoustical Society of America, 130,* EL380–EL386. doi:10.1121/1.3653546

Gelfer, M. P., & Mikos, V. A. (2005). The relative contributions of speaking fundamental frequency and formant frequencies to gender identification based on isolated vowels. *Journal of Voice, 19,* 544–554. doi:10.1016/j.jvoice.2004.10.006

Goldstein, J. L. (1973). An optimum processor theory for the central formation of the pitch of complex tones. *Journal of the Acoustical Society of America, 54,* 1496–1516. doi:10.1121/1.1914448

Green, T., Faulkner, A., & Rosen, S. (2002). Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants. *Journal of the Acoustical Society of America, 112,* 2155–2164. doi:10.1121/1.1506688

Greenwood, D. D. (1990). A cochlear frequency–position function for several species—29 years later. *Journal of the Acoustical Society of America, 87,* 2592–2605. doi:10.1121/1.399052

Guttman, N. (1963). Laws of behavior and facts of perception. In S. Koch (Ed.), *Psychology: A study of a science: Vol. 5. The process areas, the person, and some applied fields: Their place in psychology and in science* (pp. 114–178). New York, NY: McGraw-Hill.

Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology, 51,* 79–88. doi:10.1037/h0046219

Heffner, H., & Whitfield, I. C. (1976). Perception of the missing fundamental by cats. *Journal of the Acoustical Society of America, 59,* 915–919. doi:10.1121/1.380951

Heffner, R. S., & Heffner, H. E. (1991). Behavioral hearing range of the chinchilla. *Hearing Research, 52,* 13–16. doi:10.1016/0378-5955(91)90183-A

Houtsma, A. J. M., & Smurzynski, J. (1990). Pitch identification and discrimination for complex tones with many harmonics. *Journal of the Acoustical Society of America, 87,* 304–310. doi:10.1121/1.399297

Hulse, S. H. (1995). The discrimination-transfer procedure for studying auditory perception and perceptual invariance in animals. In G. M. Klump, R. J. Dooling, R. R. Fay, & W. C. Stebbins (Eds.), *Methods in comparative psychoacoustics* (pp. 319–330). Basel, Switzerland: Birkhauser-Verlag.

Joris, P. X., Bergevin, C., Kalluri, R., McLaughlin, M., Michelet, P., van der Heijden, M., & Shera, C. A. (2011). Frequency selectivity in Old-World monkeys corroborates sharp cochlear tuning in humans. *Proceedings of the National Academy of Sciences USA, 108,* 17516–17520.

Klinge, A., Itatani, N., & Klump, G. M. (2010). A comparative view on the perception of mistuning: Constraints of the auditory periphery. In E. A. Lopez-Poveda, A. R. Palmer, & R. Meddis (Eds.), *The neurophysiological basis of auditory perception* (pp. 465–475). New York, NY: Springer. doi:10.1007/978-1-4419-5686-6_43

Klinge, A., & Klump, G. M. (2009). Frequency difference limens of pure tones and harmonics within complex stimuli in Mongolian gerbils and humans. *Journal of the Acoustical Society of America, 125,* 304–314. doi:10.1121/1.3021315

Klinge, A., & Klump, G. M. (2010). Mistuning detection and onset asynchrony in harmonic complexes in Mongolian gerbils. *Journal of the Acoustical Society of America, 128,* 280–290. doi:10.1121/1.3436552

Langner, G., Albert, M., & Briede, T. (2002). Temporal and spatial coding of periodicity information in the inferior colliculus of awake chinchilla (Chinchilla laniger). *Hearing Research, 168,* 110–130. doi:10.1016/S0378-5955(02)00367-2

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research, 25,* 527–542. doi:10.1007/BF01758181

Licklider, J. C. R. (1951). A duplex theory of pitch perception. *Experientia, 7,* 128–134. doi:10.1007/BF02156143

Loebach, J. L., & Pisoni, D. B. (2008). Perceptual learning of spectrally degraded speech and environmental sounds. *Journal of the Acoustical Society of America, 123,* 1126–1139. doi:10.1121/1.2823453

Loizou, P. C., Dorman, M., & Tu, Z. (1999). On the number of channels needed to understand speech. *Journal of the Acoustical Society of America, 106,* 2097–2103. doi:10.1121/1.427954

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.

Malott, R. W., & Malott, M. K. (1970). Perception and stimulus generalization. In W. C. Stebbins (Ed.), *Animal psychophysics: The design and conduct of sensory experiments* (pp. 363–400). New York, NY: Appleton-Century-Crofts.

McLachlan, N. (2011). A neurocognitive model of recognition and pitch segregation. *Journal of the Acoustical Society of America, 130,* 2845–2854. doi:10.1121/1.3643082

Meddis, R., & O'Mard, L. (1997). A unitary model of pitch perception. *Journal of the Acoustical Society of America, 102,* 1811–1820. doi:10.1121/1.420088

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica, 40,* 1–18. doi:10.1159/000261678

Oxenham, A. J., Bernstein, J. G. W., & Penagos, H. (2004). Correct tonotopic representation is necessary for complex pitch perception. *Proceedings of the National Academy of Sciences USA, 101,* 1421–1425. doi:10.1073/pnas.0306958101

Oxenham, A. J., & Shera, C. A. (2003). Estimates of human cochlear tuning at low levels using forward and simultaneous masking. *Journal of the Association for Research in Otolaryngology, 4,* 541–554. doi:10.1007/s10162-002-3058-y

Patterson, R. D., Handel, S., Yost, W. A., & Datta, A. J. (1996). The relative strength of the tone and noise components in iterated rippled noise. *Journal of the Acoustical Society of America, 100,* 3286–3294. doi:10.1121/1.417212

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grain pitch discrimination. *Neuron, 33,* 185–191. doi:10.1016/S0896-6273(01)00580-3

Qin, M. K., & Oxenham, A. J. (2005). Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification. *Ear & Hearing, 26,* 451–460. doi:10.1097/01.aud.0000179689.79868.06

Ruggero, M. A., & Temchin, A. N. (2005). Unexceptional sharpness of frequency tuning in the human cochlea. *Proceedings of the National Academy of Sciences USA, 102,* 18614–18619. doi:10.1073/pnas.0509323102

Shafiro, V. (2008). Identification of environmental sounds with varying spectral resolution. *Ear & Hearing, 29,* 401–420. doi:10.1097/AUD.0b013e31816a0cf1

Shamma, S., & Klein, D. (2000). The case of the missing pitch templates: How harmonic templates emerge in the early auditory system. *Journal of the Acoustical Society of America, 107,* 2631–2644. doi:10.1121/1.428649

Shannon, R. V., Zeng, F.-G., Kamath, V., & Wygonski, J. (1998). Speech recognition with altered spectral distribution of envelope cues. *Journal of the Acoustical Society of America, 104,* 2467–2476. doi:10.1121/1.423774

Shannon, R. V., Zeng, F.-G., Kamath, V., Wygonski, J., & Ekelid, M. (1995, October 13). Speech recognition with primarily temporal cues. *Science, 270,* 303–304. doi:10.1126/science.270.5234.303

Shera, C. A., Guinan, J. J., Jr., & Oxenham, A. J. (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proceedings of the National Academy of Sciences USA, 99,* 3318–3323.

Shera, C. A., Guinan, J. J., Jr., & Oxenham, A. J. (2010). Otoacoustic estimation of cochlear tuning: Validation in the chinchilla. *Journal of the Association for Research in Otolaryngology, 11,* 343–365. doi:10.1007/s10162-010-0217-4

Shofner, W. P. (2002). Perception of the periodicity strength of complex sounds by the chinchilla. *Hearing Research, 173,* 69–81. doi:10.1016/S0378-5955(02)00612-3

Shofner, W. P. (2008). Representation of the spectral dominance region of pitch in the steady-state temporal discharge patterns of cochlear nucleus units. *Journal of the Acoustical Society of America, 124,* 3038–3052. doi:10.1121/1.2981637

Shofner, W. P. (2011). Perception of the missing fundamental by chinchillas in the presence of low-pass masking noise. *Journal of the Association for Research in Otolaryngology, 12,* 101–112. doi:10.1007/s10162-010-0237-0

Shofner, W. P., & Selas, G. (2002). Pitch strength and Stevens's power law. *Perception & Psychophysics, 64,* 437–450. doi:10.3758/BF03194716

Shofner, W. P., & Whitmer, W. M. (2006). Pitch cue learning in chinchillas: The role of spectral region in the training stimulus. *Journal of the Acoustical Society of America, 120,* 1706–1712. doi:10.1121/1.2225969

Shofner, W. P., Whitmer, W. M., & Yost, W. A. (2005). Listening experience with iterated rippled noise alters the perception of "pitch" strength of complex sounds in the chinchilla. *Journal of the Acoustical Society of America, 118,* 3187–3197. doi:10.1121/1.2049107

Shofner, W. P., & Yost, W. A. (1997). Discrimination of rippled-spectrum noise from flat-spectrum noise by chinchillas: Evidence for a spectral dominance region. *Hearing Research, 110,* 15–24. doi:10.1016/S0378-5955(97)00063-4

Shofner, W. P., Yost, W. A., & Whitmer, W. M. (2007). Pitch perception in chinchillas (Chinchilla laniger): Generalization using rippled noise.

Journal of Comparative Psychology, 121, 428–439. doi:10.1037/0735-7036.121.4.428

Siegel, J. H., Cerka, A. J., Recio-Spinoso, A., Temchin, A. N., van Dijk, P., & Ruggero, M. A. (2005). Delays of stimulus-frequency otoacoustic emissions and cochlear vibrations contradict the theory of coherent reflection filtering. *Journal of the Acoustical Society of America, 118,* 2434–2443.

Smith, Z. M., Delgutte, B., & Oxenham, A. J. (2002, March 7). Chimaeric sounds reveal dichotomies in auditory perception. *Nature, 416,* 87–90. doi:10.1038/416087a

Tomlinson, R. W. W., & Schwarz, D. W. F. (1988). Perception of the missing fundamental in nonhuman primates. *Journal of the Acoustical Society of America, 84,* 560–565. doi:10.1121/1.396833

Turner, R. S. (1977). The Ohm–Seebeck dispute, Hermann von Helmholtz, and the origins of physiological acoustics. *British Journal for the History of Science, 10,* 1–24. doi:10.1017/S0007087400015089

Wightman, F. L. (1973). The pattern-transformation model of pitch. *Journal of the Acoustical Society of America, 54,* 407–416. doi:10.1121/1.1913592

Winter, I. M., Wiegrebe, L., & Patterson, R. D. (2001). The temporal representation of the delay of iterated rippled noise in the ventral cochlear nucleus of the guinea-pig. *Journal of Physiology, 537,* 553–566. doi:10.1111/j.1469-7793.2001.00553.x

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14,* 609–630. doi:10.1080/016909699386202

Yost, W. A. (1996). Pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America, 100,* 3329–3335. doi:10.1121/1.416973

Yost, W. A., & Hill, R. (1979). Models of the pitch and pitch strength of ripple noise. *Journal of the Acoustical Society of America, 66,* 400–410. doi:10.1121/1.382942

Yost, W. A., Patterson, R., & Sheft, S. (1996). A time domain description for the pitch strength of iterated rippled noise. *Journal of the Acoustical Society of America, 99,* 1066–1078. doi:10.1121/1.414593

Received March 1, 2012
Revision received July 10, 2012
Accepted July 11, 2012 ■

Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write APA Journals at Reviewers@apa.org. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.
- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.
- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, "social psychology" is not sufficient—you would need to specify "social cognition" or "attitude change" as well.
- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

APA now has an online video course that provides guidance in reviewing manuscripts. To learn more about the course and to access the video, visit http://www.apa.org/pubs/authors/review-manuscript-ce-video.aspx.