Rhythm judgments reveal a frequency asymmetry in the perception and neural coding of sound synchrony

Magdalena Wojtczak[a,1], Anahita H. Mehta[a], and Andrew J. Oxenham[a]

[a]Department of Psychology, University of Minnesota, Minneapolis, MN 55455

In modern Western music, melody is commonly conveyed by pitch changes in the highest-register voice, whereas meter or rhythm is often carried by instruments with lower pitches. An intriguing and recently suggested possibility is that the custom of assigning rhythmic functions to lower-pitch instruments may have emerged because of fundamental properties of the auditory system that result in superior time encoding for low pitches. Here we compare rhythm and synchrony perception between low- and high-frequency tones, using both behavioral and EEG techniques. Both methods were consistent in showing no superiority in time encoding for low over high frequencies. However, listeners were consistently more sensitive to timing differences between two nearly synchronous tones when the high-frequency tone followed the low-frequency tone than vice versa. The results demonstrate no superiority of low frequencies in timing judgments but reveal a robust asymmetry in the perception and neural coding of synchrony that reflects greater tolerance for delays of low- relative to high-frequency sounds than vice versa. We propose that this asymmetry exists to compensate for inherent and variable time delays in cochlear processing, as well as the acoustical properties of sound sources in the natural environment, thereby providing veridical perceptual experiences of simultaneity.

rhythm perception | time encoding | sound asynchrony perception | auditory perception | mismatch negativity

The perception of music involves the processing of multiple simultaneous sounds, including the perceptual organization of components originating from one instrument into a single "auditory object" (1) and their segregation from components that originate from other instruments. The first stage of this process occurs in the inner ear and involves the decomposition of incoming sound by frequency, resulting in frequency-to-place mapping (tonotopy) along the length of the cochlear partition (2).

A series of recent studies has suggested that these and other basic properties of the auditory system, which are shared by a wide variety of species, may help account for some fundamental aspects of Western music, including the dominance of high-register instruments (i.e., instruments with high fundamental frequencies) in carrying the melody (3–7), as well as the dominance of low-register instruments (such as bass or bass drum) in defining the meter or rhythm (8). In the latter case, a study by Hove et al. (8) involving both behavioral and EEG experiments reported that time encoding was superior for low- vs. high-pitched sounds and suggested that this superior time encoding could explain why low instruments generally lay the rhythm in music.

The suggestion that superior time encoding of low pitches can explain aspects of Western music is intriguing. However, it runs contrary to the results from studies of many other types of auditory temporal processing, including gap detection (9), amplitude-modulation detection (10) and discrimination (11), and duration discrimination (12), which have either revealed no clear effects of frequency or have shown better temporal encoding at higher rather than lower frequencies. Here we reexamine the data and claims of the earlier study (8). By replicating and extending the conditions tested in the original study, we are able to refute the claim of superior time encoding for low pitches. In its place, we present results that reveal an asymmetry in the perception and cortical processing

of simultaneity. Simultaneity, or synchrony, is a critical property of sounds that emanate from the same source; musicians often strive for simultaneity when playing with others or when playing chords on a single instrument to promote the sensation of "fusion" between instruments or notes (13). We find that both the neural coding and perception of simultaneity are more tolerant of asynchrony between low and high sounds when the high sound leads the low sound than vice versa. This finding not only explains the previous results ascribed to rhythm perception but also provides a striking example of a perceptual adaptation to the electromechanical properties of the inner ear and the acoustic regularities in the environment.

Results

MMN Experiment. EEG activity was recorded while listeners were presented with a sequence of pairs of simultaneous low and high tones. The sequence was isochronous, with tones repeated every 500 ms, except for 10% of trials in which the low tone deviated from the expected time of occurrence, and 10% of trials in which the high tone deviated from the expected timing. The EEG evoked response potentials were measured for two test conditions and two control conditions (Fig. 1). In the "early" test condition, the deviants occurred 50 ms before the expected time (low early or high early), as in the study by Hove et al. (8). In the corresponding early control condition, the low or high tones were advanced by 50 ms on 50% of trials each. In the "late" test condition, the deviants were delayed by 50 ms relative to the expected timing (low late and high late), and in the corresponding late control condition, 50% of trials had the low tone delayed and the other 50% had the high tone delayed relative to the counterpart in the other sequence. The mismatch negativity (MMN) responses that occur in the presence of a deviant stimulus, averaged over all brain regions covered by the electrodes (frontal, parietal, central, and temporal), were compared for the low-early and high-early deviants (Fig. 2, *Left*) and for the low-late and high-late deviants (Fig. 2, *Right*). The

---

Significance

Why music developed in the way it did is of fundamental interest across the sciences, arts, and humanities. A recent study reported that rhythm perception was enhanced for low-pitch sounds, and suggested that this enhancement may provide a biological explanation for why low-pitch instruments generally provide the beat in music. Our results refute this hypothesis, but reveal a robust asymmetry in the perception and cortical coding of synchrony between low and high tones. Synchrony is an important and ubiquitous property of sounds emanating from the same source. Our findings suggest a way in which the veridical perception of simultaneous events is maintained in the face of delays produced both by cochlear processing and by the inherent properties of acoustic sources.
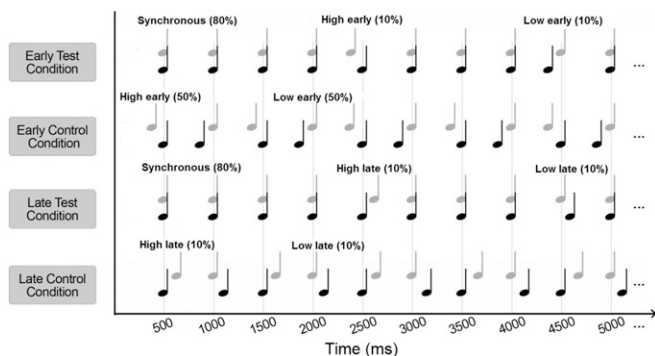

Fig. 1. Illustration of the four stimulus conditions in the EEG MMN experiment. In each condition, a sequence of pairs of harmonic tones with fundamental frequencies of 196 Hz (low tone denoted by black notes) and 466 Hz (high tone denoted by gray notes) was used. In the test conditions, the tones were synchronous on 80% of trials. One test condition (early test condition) had low tones or high tones presented early, on 10% of trials each. The other test condition (late test condition) had the low or high tones presented late on 10% of trials each. The two corresponding control conditions consisted of low and high tones presented early (50% of trials each) and low and high tones presented late (50% of trials each).

grand average responses were compared, rather than MMN responses averaged over separate brain regions, because the number of electrodes in each region was relatively small (a total of 64 vs. 128 electrodes used by Hove et al.) and because we did not intend to identify regions that contributed most to the different responses to the low-early and high-early deviants.

A two-way repeated-measures analysis of variance (ANOVA) was conducted on the MMN values, with the main factors of tone order and pitch of the tone carrying the rhythmic irregularity. The ANOVA showed that the MMN amplitude was significantly larger when the low tone was leading; that is, for the low-early and high-late deviants than for the high-early and low-late deviants ($F_{1,12} = 16.5$; $P = 0.002$), but there was no significant dependence of MMN amplitudes on whether the low or high tone deviated from the rhythmic pattern ($F_{1,12} = 0.68$; $P = 0.43$). There was also a significant interaction between the two main factors ($F_{1,12} = 38.5$; $P = 0.004$), which reflected the overall larger MMN for the early vs. late deviants. This interaction was expected, as the late deviants were likely to produce a smaller MMN because the onset of the stimulus coincided with the onset of the rhythmically regular tone.

Behavioral Experiment.

Two-pitch sequences. Participants were asked to detect a deviation from an otherwise isochronous presentation of simultaneous low and high tones. Within a trial, the pairs of simultaneous tones were presented five times with onsets separated by 500 ms, with the exception of the deviant. The timing of the onset of the deviant relative to the expected time (in an asynchronous pair of the low and high tones) was adaptively varied to measure the just-detectable asynchrony in the stimulus. To induce some initial entrainment, the deviant could only occur in the third or fourth tone burst. Thresholds for detecting a deviation introduced by presenting the low tone early, low tone late, high tone early, or high tone late were measured in separate blocks. To examine the generality of the findings for a given pitch difference between the low and high tones, pure tones and complex tones with equal-amplitude components (uniform spectrum) were used in this experiment, in addition to the spectrally shaped complex tones that were used in the MMN experiment. The use of pure tones reduced the potential overlap and interaction of the simultaneous tones within the auditory periphery; the use of tones with a uniform spectrum was to produce substantial spectral overlap between the two tones within the auditory periphery, thereby

partially dissociating pitch (fundamental frequency) from spectral content. The results plotted in Fig. 3 show the mean thresholds for the three types of stimuli.

For the pure tones (Fig. 3, *Left*) and spectrally shaped complex tones (Fig. 3, *Middle*), thresholds for detecting a deviation from the regular pattern were lower in conditions in which the onset of the low tone preceded the onset of the high tone in the deviant (i.e., for low lead) than vice versa (i.e., high lead), irrespective of whether the low or high tone occurred at the expected time. For the complex tones with the uniform, or flat, spectrum, no clear differences between the conditions were observed.

Two-way repeated-measures ANOVAs with factors of tone order and pitch of the on-time tone were run separately for the three types of stimuli (pure tones, spectrally shaped complex tones, and spectrally uniform complex tones). The ANOVAs for the pure tones and spectrally shaped complex tones revealed a significant difference between thresholds from blocks with the low tone leading and the high tone leading [pure tones: $F_{1,10} = 10.55$ ($P = 0.01$); shaped complex tones: $F_{1,6} = 11.4$ ($P = 0.02$)], but no significant dependence on whether the low or the high tone occurred at the expected time [pure tones: $F_{1,10} = 1.85$ ($P = 0.15$); shaped complex tones: $F_{1,6} = 0.61$ ($P = 0.47$)]. There were no significant interactions between the main factors for either type of stimulus [pure tones: $F_{1,10} = 1.85$ ($P = 0.20$); shaped complex tones: $F_{1,6} = 4.88$ ($P = 0.07$)]. For complex tones with flat spectra (i.e., with equally intense components), there was no significant effect of either main factor and no significant interaction [tone order: $F_{1,7} = 0.03$ ($P = 0.88$); on-time pitch: $F_{1,7} = 1.60$ ($P = 0.25$); interaction of tone order and pitch of the on-time tone: $F_{1,7} = 0.04$ ($P = 0.94$)].

Single-pitch sequence. For the sequences of tone pairs presented simultaneously, listeners could perform the task by either detecting a change in the rhythmic pattern or detecting an asynchrony between two tones without regard to the violation of the rhythmic pattern. To determine which strategy was used, participants were tested in a task in which they detected a rhythmic irregularity in an otherwise isochronous sequence of single tones. In separate conditions, a deviation from a rhythmic pattern was introduced by advancing or delaying a tone with respect to the expected time. This design led to four conditions: two for the low tone (low early and low late) and two for the high tone (high early and high late).
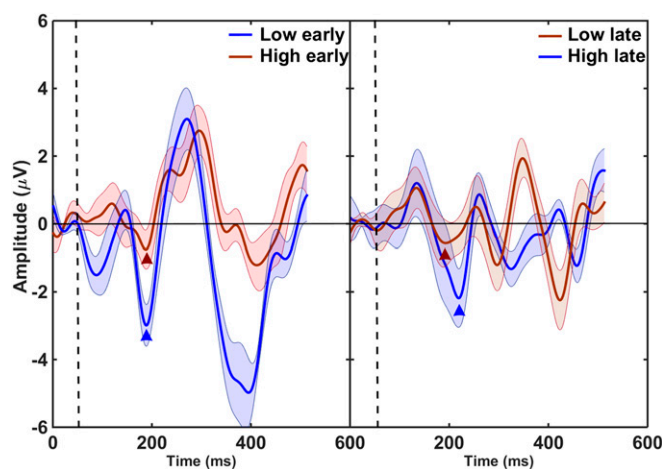

Fig. 2. Difference waveforms showing grand average MMN responses to deviants with one tone presented 50 ms before the expected time (*Left*) and deviants with one tone presented 50 ms after the expected time (*Right*). The MMN amplitudes were larger for deviants with the low tone leading (*Left*, low early; *Right*, high late). The vertical dashed lines denote the onset of the first tone in the deviant, and the waveforms before the dashed line represent the baseline for each difference waveform. The triangles denote center positions of 50-ms windows used to calculate the MMN amplitude for each condition.
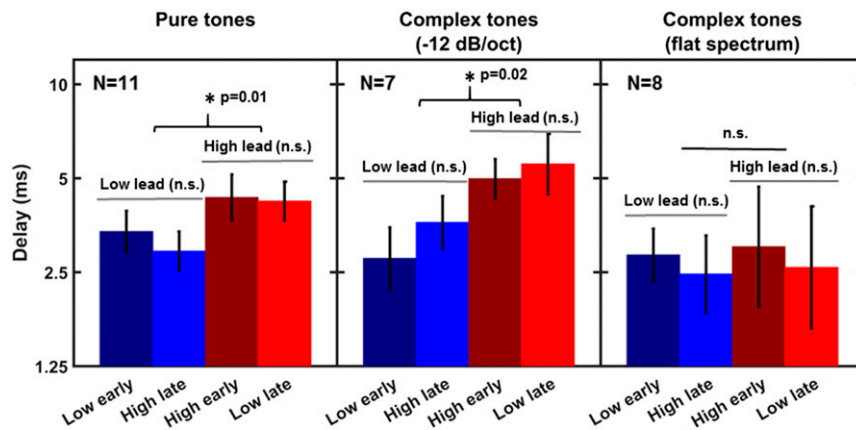
Fig. 3. Mean thresholds for detecting rhythmic irregularity created by advancing (early) or delaying (late) one of the tones, low or high, relative to the expected time. The error bars represent ±1 SEM. The three panels show data for three different types of stimuli with the same pitch difference: pure tones (*Left*), spectrally shaped complex tones (*Middle*), and complex tones with flat spectra (*Right*).

Mean thresholds for detecting rhythmic irregularity are shown in Fig. 4.

A repeated-measures ANOVA showed no significant effect of pitch (low vs. high) ($F_{1,11} = 0.14$; $P = 0.72$), no effect of whether the irregular tone was leading vs. lagging ($F_{1,11} = 0.37$; $P = 0.56$), and no interaction ($F_{1,11} = 0.03$; $P = 0.87$). The mean thresholds for single-pitch sequences were nearly an order of magnitude higher than those for the two-pitch sequences (replotted from Fig. 3 for comparison), indicating that sensitivity to a local asynchrony between two tones presented in an isochronous context is much greater than sensitivity to a violation of a rhythmic pattern.

Discussion

No Evidence for Superior Temporal Encoding at Low Frequencies. The results from both the EEG MMN experiment and the psychophysical experiments do not support the hypothesis that time encoding and perception is superior for low-pitch tones (8). The amplitude of the MMN response was larger in response to the low-early than to the high-early deviant, consistent with the original findings of Hove et al. (8). However, the MMN response was also larger for the high-late than for the low-late deviant, contrary to the predictions based on superior rhythm encoding of low-pitch sounds. The same conclusion can be drawn from the behavioral results: when tone pairs were presented, discrimination thresholds were the same whether the low or high tones were irregular. A difference in thresholds only emerged when conditions in which the low tones led the high tones were compared with conditions in which high tones led the low tones, regardless of which tones were rhythmically irregular. When only a single tone sequence was presented, discrimination thresholds were much poorer than for the tone pairs and were the same for both low and high tones alone.

Taken together, our results suggest there is no advantage of low over high tones in temporal or rhythm processing. Instead, the differences in performance reflect a perceptual and neural asymmetry in the processing of temporal asynchronies between tones of different frequencies, reflecting greater tolerance for delays of low- relative to high-frequency sounds than vice versa.

This asymmetry could also account for the results of the sensorimotor task reported by Hove et al. (8). They found that a low-leading irregularity in their sequences of tone pairs had a greater effect on their subjects' tapping behavior than did a high-leading irregularity. Although they interpreted the outcome in terms of better time encoding of the low tones, these results could also be explained in terms of an asymmetry in the processing of temporal asynchrony: Because the low-leading asynchrony was more salient

than the high-leading asynchrony, it resulted in a greater change in the subjects' motor behavior.

Dissociating Pitch from Frequency. In our behavioral experiments, we tested not only piano-like complex tones but also more synthetic pure tones and complex tones with equal-amplitude components. These different conditions allowed us to dissociate pitch (linked to the fundamental frequency or repetition rate of a waveform) from brightness (associated with the spectral centroid of a sound). If the asymmetry we observed had been a result of differences in perceived pitch, then all three conditions should have shown the asymmetry, as all three involved low- and high-pitch tones. Instead, we found that the effect was only observed in the two conditions in which the low and high tones did not substantially overlap in spectrum. Thus, the perceptual and neural asymmetry we observed appears to be a result of tonotopic separation, as established in the cochlea, rather than perceived pitch.
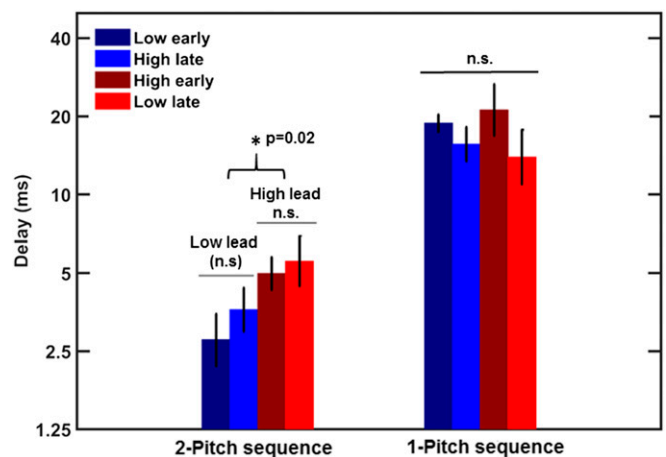

Fig. 4. Mean thresholds for detecting rhythmic irregularity in a sequence of a one-pitch complex tone spectrally shaped with a slope of −12 dB/oct are shown by the right set of bars. The left set of bars shows thresholds for detecting a rhythmic irregularity created by manipulating the timing of one tone in a pair of low and high tones, replotted from the middle panel of Fig. 3. The error bars represent ±1 SEM. The much higher thresholds for the one-pitch sequences indicate that detecting rhythmic irregularity is more difficult than detecting a local asynchrony between two tones in the two-pitch sequence, making asynchrony a much more salient cue for detecting a temporal change in the two-pitch sequence.

Explaining the Asymmetry in the Perception and Neural Processing of Acoustic Asynchrony. The neural and behavioral asymmetries observed in the present study are consistent with some earlier behavioral findings using just pure tones, which were separated by two octaves or more (14, 15). The fact that these effects occur with such large separations (and when potential interference is eliminated with a masking noise spectrally located between the tones) rules out the explanations of mutual masking provided by Hove et al. using an auditory model (8). Why, then, are delays of low-frequency sounds tolerated more readily than delays of high-frequency sounds when judging simultaneity? We offer one explanation based on the electromechanical properties of the cochlea, and one based on the natural properties of some sound sources.

Compensating for cochlear delays. Sound entering the inner ear forms traveling waves along the cochlear partition. The cochlea's response to sound produces frequency-dependent delays, whereby the response to low frequencies is slower than the response to high frequencies (16). The temporal dispersion along the cochlea has been shown in animals by analyzing auditory nerve responses (17–20), and in humans by estimating response latencies using noninvasive physiological measurements of compound action potentials (21–23), auditory-brainstem responses (24–28), and otoacoustic emissions (2, 23, 25, 28, 29). Based on a model of the cochlea proposed by de Boer (30), the difference in latency between the response to a 100-Hz tone at the apical end of the cochlea and the response to a 10-kHz tone at the basal end has been estimated to be about 10 ms, with the greatest gradient occurring for frequencies up to about 1.5–2 kHz (16, 31). The model predictions are illustrated by the black curve in Fig. 5. Interestingly, psychophysical studies in humans have shown that these physiologically documented delays are not perceived, with physically simultaneous sources generally being judged as simultaneous, despite their different delays in traveling through the auditory periphery (14, 15). Moreover, the veridical perception of cross-frequency timing has been demonstrated at both very low [25 dB sound pressure level (SPL)] and high (85 dB SPL) sound levels (14, 15), despite evidence that differences in cross-frequency cochlear delays may decrease with increasing level (23, 25, 29). Although the relative cochlear delays between low- and high-frequency sounds can vary depending on their intensities, the response to the low-frequency sound will always be slower, as illustrated by the three red lines in Fig. 5 (25). Thus, higher levels of the auditory pathways may have developed tolerance to low-frequency delays, but not to high-frequency delays (which could not occur in the cochlea). In summary, the perceptual asymmetry in perceiving asynchronies between low and high frequencies may reflect an adaptation of the auditory system to the intrinsic filtering properties of the inner ear, resulting in the veridical perception of simultaneity.

Compensating for acoustic delays. Another possible explanation for humans' greater sensitivity to low-leading asynchronies relates to the behavior of some sound sources and acoustic enclosures. As with the cochlea, low-frequency sounds can take longer to reach their maximum amplitude than high-frequency sounds that begin at the same time. To allow for such potential differences in delays between simultaneous sounds, the auditory system may have adapted to tolerate low-frequency delays more readily than high-frequency delays. A similar argument has been used to explain in part why humans are often more tolerant of lags in audio relative to video than vice versa (33–35), based on the different velocities of sound and light. Finally, it should be noted that these two explanations, one based on inner-ear mechanics and the other based on sound source properties, are not mutually exclusive and may both play a role in our asymmetric perception of auditory asynchrony.

Methods

MMN Experiment.

Participants. Seventeen participants (ten women and seven men, aged 18–30 y) were recruited for this study. The participants had normal hearing, as
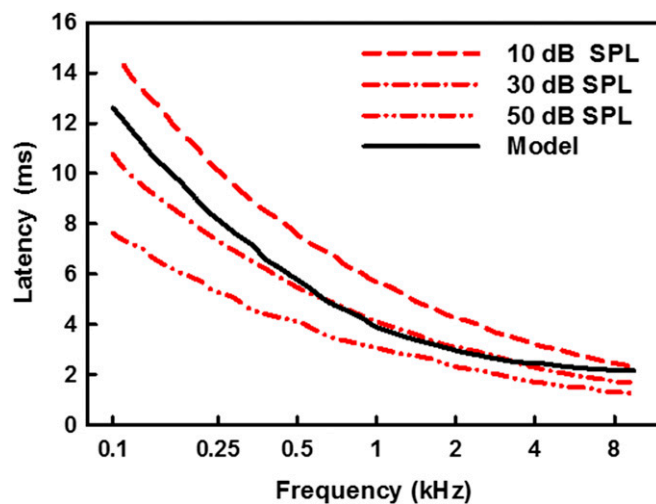

Fig. 5. Latency of cochlear responses as a function of frequency predicted by a transmission line model: the black curve shows latencies from a model of the cochlea by de Boer (32), and the red curves show trends based on sample auditory-brainstem data from Neely et al. (25).

indicated by audiometric thresholds at or below 15 dB hearing level (HL) for octave frequencies between 250 Hz and 8 kHz, and had no history of hearing or neurological disorders. Participants in all the experiments in this study provided informed written consent, and the protocol for the study was approved by the Institutional Review Board of the University of Minnesota.

Stimuli. The stimuli were two harmonic tones, each consisting of 20 consecutive harmonics. The fundamental frequencies were 196 Hz for the low tone and 466 Hz for the high tone, and thus their pitches were comparable to those of the synthesized G_3 and B-flat$_4$ piano notes used by Hove et al. (8), respectively. The tones had a duration of 300 ms, including 10-ms onset and 250-ms offset raised-cosine ramps. The stimuli were spectrally shaped by imposing a slope of −12 dB/oct on the components. The stimuli were presented at a nominal tempo of 120 beats/min (500-ms onset-to-onset interval) at an overall (root-mean-squared) level of 65 dB SPL. Two test runs and two control runs were used to obtain the MMN amplitudes (Fig. 1). Each run consisted of 1,100 presentations and lasted 15 min. In one test run (early test condition), 10% of trials consisted of deviants with the low tone presented on time and the high tone presented 50-ms before the expected time, and another 10% of trials consisted of the high tone presented on time and the low tone presented 50 ms earlier than expected. The remaining 80% of trials consisted of simultaneous pairs of the two tones presented isochronously every 500 ms. The deviants occurred in random trials with a restriction that the same deviant could not occur in two consecutive presentations. In the corresponding control condition (early control condition), 50% of trials consisted of the low-tone early and the other 50% high-tone early trials. In another test run (late test condition), 10% of deviants consisted of pairs with low-tone late and high-tone on time and another 10% consisted of high-tone late and low-tone on time, whereas in the corresponding control run (late control condition), each of the two types of pairs occurred on 50% of trials. For each type of test run, the stimuli from the corresponding control runs were used as "standards" to calculate the reference response that was subtracted from that for the corresponding deviant. This approach was used in the calculation of MMN responses to eliminate confounding effects of acoustic differences between the synchronous and asynchronous stimuli (36).

Procedure. The EEG data were acquired continuously throughout a run. During data acquisition, participants were seated in a double-walled, electrically shielded, sound-attenuating booth. Participants were fitted with a cap (Easy Cap; Falk Minow Services) containing 64 silver/silver-chloride scalp electrodes. Two additional reference electrodes, one placed on each mastoid, and two ocular electrodes were used. The impedance of all electrodes was monitored and maintained below 10 kΩ. The data were recorded at a sampling rate of 1,024 Hz, using a 64-channel Brain-Vision system consisting of a Brain-Vision recorder (Version 1.01b) and a BrainAmp integrated amplifier (Brain Products GmbH). The sounds were presented via ER-2 insert phones (Etymotic Research), and participants watched a silent movie with subtitles during data acquisition.

EEG data analysis. EEG analysis was carried out using Matlab (MathWorks). Waveforms from each electrode were referenced to the average waveform from the two mastoid electrodes and bandpass filtered between 1.6 and 20 Hz with a

zero phase shift. The EEG activity was epoched into 600-ms segments starting 100 ms before the onset of the on-time sound, and epochs with artifacts resulting from eye blinking detected by the ocular electrodes were discarded. Epochs for each type of deviant (low early, high early, low late, and high late) were averaged separately across all electrodes. Similarly, epochs for each corresponding type of control were averaged across all electrodes separately. The MMN was then calculated by subtracting each deviant waveform from the respective control waveform. To quantify the MMN amplitude, a subset of 32 electrodes was selected that were uniformly distributed across both hemispheres and the frontal, central, parietal, and temporal regions. The average difference waveforms from these electrodes were then used to quantify MMN amplitude for each condition. The MMN amplitude was calculated as the average amplitude over a 50-ms window centered on the most negative peak between 100 and 200 ms after stimulus onset. The MMN amplitudes calculated for each subject and each condition were then used to perform statistical analyses.

Behavioral Detection of Rhythmic Irregularities in Sequences of Tone Pairs.
Participants. Thirteen listeners (seven women and six men, aged 18–27 y) were recruited for the study. All the listeners had normal hearing thresholds (≤15 dB HL) for audiometric frequencies between 250 Hz and 8 kHz and had no history of hearing or neurological disorders. Not all participants completed the experiment for all of the types of stimuli used. From the thirteen subjects who participated in the two-tone conditions, subsets of 11 completed the pure-tone conditions, 8 completed the spectrally shaped complex-tone conditions, and 7 completed the flat-spectrum complex-tone conditions. Participants were given two practice runs for each condition before data collection commenced.
Stimuli. In addition to the stimuli used in the EEG MMN experiment, two other stimulus conditions were used: a pair of pure tones with frequencies of 196 and 466 Hz and a pair of complex tones that were identical to those used in the EEG study except that all 20 harmonics in each tone were presented at the same intensity. The duration, the gating, and the level of presentation of the stimuli were the same as in the EEG study.
Procedure. A one-interval two-alternative forced-choice procedure was used to measure thresholds for detecting rhythmic irregularity in a sequence of five bursts of simultaneous low and high tone pairs. Four conditions were tested. In separate runs, the irregularity was introduced by advancing or delaying either the low tone (conditions "low early" and "low late") or the high tone (conditions "high early" and "high late") relative to the expected time, as determined by the regular period of 500 ms (with the nominal silent interval of 200 ms between 300-ms tone bursts). On a given trial, either the third or the fourth burst, chosen at random, contained the asynchronous tone pair, and the listeners were asked to indicate during which burst the irregularity occurred. Visual feedback

indicating the correct response was provided after each trial. A run started with an easily detectable delay between the onsets that was adaptively varied using a two-down one-up tracking technique converging on 70.7% correct detections (37). The delay was initially varied by a factor of 2. After the first two reversal points (defined as delays at which the direction was reversed from an increase to a decrease in delay and vice versa), the delay was varied by a factor of $2^{1/2}$, and after two more reversals it was varied by a factor of $2^{1/4}$ for the remaining eight reversals of each run. A run terminated after a total of 12 reversals, and the threshold estimate was calculated as the geometric mean of the delays at the last eight reversal points. Three runs were completed for each type of stimulus and each condition, and the final threshold was calculated by geometrically averaging the three threshold estimates. The order of conditions was randomized for each subject.

Behavioral Detection of Rhythmic Irregularities in Sequences of Single Tones.
Participants. Twelve listeners (nine women and three men, aged 19–25 y) with hearing thresholds at or below 15 dB HL at audiometric frequencies between 250 and 8 kHz and with no history of hearing and neurological disorders participated in this experiment. One of the 12 listeners had also performed the experiment with the sequences of tone pairs. For each condition, listeners performed two practice runs before data collection commenced.
Stimuli and procedure. Thresholds for detecting rhythmic irregularity for a repeated presentation of a single harmonic tone were measured for the spectrally shaped low and high harmonic tones that were used in the EEG study. An adaptive two-interval two-alternative forced-choice procedure with a two-down one-up tracking technique was used. Each of the two observation intervals contained six bursts of 300-ms harmonic tone separated by 200-ms silent intervals, except for the intervals right before and right after the target (signal). The signal was the penultimate burst in the signal interval. In separate runs, the signal was advanced or delayed relative to the expected time determined by the rhythm. Listeners were asked to indicate whether the irregular tone occurred in the first or the second interval, and they were given visual feedback indicating the correct response after each trial. The adaptive-tracking procedure and the stepping rule were the same as that for the sequences of tone pairs. Four conditions were tested: two for the low tone (low early and low late) and two for the high tone (high early and high late). Three threshold estimates were geometrically averaged to obtain the final estimate.

1. Griffiths TD, Warren JD (2004) What is an auditory object? *Nat Rev Neurosci* 5(11): 887–892.
2. Shera CA, Guinan JJ, Jr, Oxenham AJ (2002) Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. *Proc Natl Acad Sci USA* 99(5): 3318–3323.
3. Fujioka T, Trainor LJ, Ross B, Kakigi R, Pantev C (2005) Automatic encoding of polyphonic melodies in musicians and nonmusicians. *J Cogn Neurosci* 17(10): 1578–1592.
4. Fujioka T, Trainor LJ, Ross B (2008) Simultaneous pitches are encoded separately in auditory cortex: An MMNm study. *Neuroreport* 19(3):361–366.
5. Trainor LJ, Marie C, Bruce IC, Bidelman GM (2014) Explaining the high voice superiority effect in polyphonic music: Evidence from cortical evoked potentials and peripheral auditory models. *Hear Res* 308:60–70.
6. Marie C, Trainor LJ (2013) Development of simultaneous pitch encoding: Infants show a high voice superiority effect. *Cereb Cortex* 23(3):660–669.
7. Marie C, Trainor LJ (2014) Early development of polyphonic sound encoding and the high voice superiority effect. *Neuropsychologia* 57:50–58.
8. Hove MJ, Marie C, Bruce IC, Trainor LJ (2014) Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proc Natl Acad Sci USA* 111(28):10383–10388.
9. Shailer MJ, Moore BCJ (1983) Gap detection as a function of frequency, bandwidth, and level. *J Acoust Soc Am* 74(2):467–473.
10. Kohlrausch A, Fassel R, Dau T (2000) The influence of carrier level and frequency on modulation and beat-detection thresholds for sinusoidal carriers. *J Acoust Soc Am* 108(2): 723–734.
11. Lee J (1994) Amplitude modulation rate discrimination with sinusoidal carriers. *J Acoust Soc Am* 96(4):2140–2147.
12. Abel SM (1972) Duration discrimination of noise and tone bursts. *J Acoust Soc Am* 51(4):1219–1223.
13. Rasch RA (1978) The perception of simultaneous notes such as in polyphonic music. *Acustica* 40:21–33.
14. Wojtczak M, Beim JA, Micheyl C, Oxenham AJ (2012) Perception of across-frequency asynchrony and the role of cochlear delays. *J Acoust Soc Am* 131(1):363–377.
15. Wojtczak M, Beim JA, Micheyl C, Oxenham AJ (2013) Effects of temporal stimulus properties on the perception of across-frequency asynchrony. *J Acoust Soc Am* 133(2):982–997.
16. Dau T, Wegner O, Mellert V, Kollmeier B (2000) Auditory brainstem responses with optimized chirp signals compensating basilar-membrane dispersion. *J Acoust Soc Am* 107(3):1530–1540.
17. Recio-Spinoso A, Temchin AN, van Dijk P, Fan Y-H, Ruggero MA (2005) Wiener-kernel analysis of responses to noise of chinchilla auditory-nerve fibers. *J Neurophysiol* 93(6):3615–3634.
18. Temchin AN, Recio-Spinoso A, van Dijk P, Ruggero MA (2005) Wiener kernels of chinchilla auditory-nerve fibers: Verification using responses to tones, clicks, and noise and comparison with basilar-membrane vibrations. *J Neurophysiol* 93(6):3635–3648.
19. Temchin AN, Recio-Spinoso A, Ruggero MA (2011) Timing of cochlear responses inferred from frequency-threshold tuning curves of auditory-nerve fibers. *Hear Res* 272(1-2):178–186.
20. Palmer AR, Shackleton TM (2009) Variation in the phase of response to low-frequency pure tones in the guinea pig auditory nerve as functions of stimulus level and frequency. *J Assoc Res Otolaryngol* 10(2):233–250.
21. Elberling C (1974) Action potentials along the cochlear partition recorded from the ear canal in man. *Scand Audiol* 3:13–19.
22. Eggermont JJ (1979) Narrow-band AP latencies in normal and recruiting human ears. *J Acoust Soc Am* 65(2):463–470.
23. Schoonhoven R, Prijs VF, Schneider S (2001) DPOAE group delays versus electrophysiological measures of cochlear delay in normal human ears. *J Acoust Soc Am* 109(4):1503–1512.
24. Eggermont JJ, Don M (1980) Analysis of the click-evoked brainstem potentials in humans using high-pass noise masking. II. Effect of click intensity. *J Acoust Soc Am* 68(6):1671–1675.
25. Neely ST, Norton SJ, Gorga MP, Jesteadt W (1988) Latency of auditory brain-stem responses and otoacoustic emissions using tone-burst stimuli. *J Acoust Soc Am* 83(2):652–656.
26. Don M, Ponton CW, Eggermont JJ, Masuda A (1993) Gender differences in cochlear response time: An explanation for gender amplitude differences in the unmasked auditory brain-stem response. *J Acoust Soc Am* 94(4):2135–2148.
27. Donaldson GS, Ruth RA (1993) Derived band auditory brain-stem response estimates of traveling wave velocity in humans. I: Normal-hearing subjects. *J Acoust Soc Am* 93(2):940–951.
28. Harte JM, Pigasse G, Dau T (2009) Comparison of cochlear delay estimates using otoacoustic emissions and auditory brainstem responses. *J Acoust Soc Am* 126(3):1291–1301.
29. Sisto R, Moleti A (2007) Transient evoked otoacoustic emission latency and cochlear tuning at different stimulus levels. *J Acoust Soc Am* 122(4):2183–2190.

PSYCHOLOGICAL AND COGNITIVE SCIENCES

30. de Boer E (1997) Classical and non-classical models of the cochlea. *J Acoust Soc Am* 101(4):2148–2150.

31. Fobel O, Dau T (2004) Searching for the optimal stimulus eliciting auditory brainstem responses in humans. *J Acoust Soc Am* 116(4 Pt 1):2213–2222.

32. de Boer E (1980) Auditory physics. Physical principles in hearing theory I. *Phys Rep* 62: 87–174.

33. van Eijk RLJ, Kohlrausch A, Juola JF, van de Par S (2008) Audiovisual synchrony and temporal order judgments: Effects of experimental method and stimulus type. *Percept Psychophys* 70(6):955–968.

34. Conrey B, Pisoni DB (2006) Auditory-visual speech perception and synchrony detection for speech and nonspeech signals. *J Acoust Soc Am* 119(6):4065–4073.

35. Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. *Perception* 9(6): 719–721.

36. Peter V, McArthur G, Thompson WF (2010) Effect of deviance direction and calculation method on duration and frequency mismatch negativity (MMN). *Neurosci Lett* 482(1):71–75.

37. Levitt H (1971) Transformed up-down methods in psychoacoustics. *J Acoust Soc Am* 49(2):2–467.