

# Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms

Michael J. Hove<sup>a,b,1</sup>, Céline Marie<sup>a,c,1</sup>, Ian C. Bruce<sup>c,d</sup>, and Laurel J. Trainor<sup>a,c,e,2</sup>

<sup>a</sup>Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, ON, Canada L8S 4K1; <sup>b</sup>Martinos Center for Biomedical Imaging, Harvard Medical School/Massachusetts General Hospital, Boston, MA 02129; <sup>c</sup>McMaster Institute for Music and the Mind, Hamilton, ON, Canada L8S 4K1; <sup>d</sup>Department of Electrical and Computer Engineering, McMaster University, Hamilton, ON, Canada L8S 4K1; and <sup>e</sup>Rotman Research Institute, Baycrest Hospital, Toronto, ON, Canada M6A 2E1

Edited by Dale Purves, Duke University, Durham, NC, and approved June 4, 2014 (received for review February 1, 2014)

The auditory environment typically contains several sound sources that overlap in time, and the auditory system parses the complex sound wave into streams or voices that represent the various sound sources. Music is also often polyphonic. Interestingly, the main melody (spectral/pitch information) is most often carried by the highest-pitched voice, and the rhythm (temporal foundation) is most often laid down by the lowest-pitched voice. Previous work using electroencephalography (EEG) demonstrated that the auditory cortex encodes pitch more robustly in the higher of two simultaneous tones or melodies, and modeling work indicated that this high-voice superiority for pitch originates in the sensory periphery. Here, we investigated the neural basis of carrying rhythmic timing information in lower-pitched voices. We presented simultaneous high-pitched and low-pitched tones in an isochronous stream and occasionally presented either the higher or the lower tone 50 ms earlier than expected, while leaving the other tone at the expected time. EEG recordings revealed that mismatch negativity responses were larger for timing deviants of the lower tones, indicating better timing encoding for lower-pitched compared with higher-pitch tones at the level of auditory cortex. A behavioral motor task revealed that tapping synchronization was more influenced by the lower-pitched stream. Results from a biologically plausible model of the auditory periphery suggest that nonlinear cochlear dynamics contribute to the observed effect. The low-voice superiority effect for encoding timing explains the widespread musical practice of carrying rhythm in bass-ranged instruments and complements previously established high-voice superiority effects for pitch and melody.

temporal perception | auditory scene analysis | rhythmic pulse | beat

The auditory system evolved to identify and locate streams of sounds in the environment emanating from sources such as vocalizations and environmental sounds. Typically, the sound sources overlap in time and frequency content. In a process termed auditory-scene analysis, the auditory system uses spectral, intensity, location, and timing cues to separate the incoming signal into streams or voices that represent each sounding object (1). Human music, whether vocal or instrumental, also often consists of several voices or streams that overlap in time, and our perception of such music depends on the basic processes of auditory-scene analysis. Thus, one might hypothesize that certain widespread music compositional practices might originate in basic properties of the auditory system that evolved for auditory-scene analysis. Two such musical practices are that the main melody line (spectral/pitch information) is most often placed in the highest-pitched voice and that the rhythmic pulse or beat (temporal information) is most often carried in the lowest-pitched voice. For pitch encoding, a high-voice superiority effect has been demonstrated previously and originates in the sensory periphery. In the present paper, we investigate the basis of carrying rhythmic beat information in lower-pitched voices, as manifested

in musical conventions such as the walking bass lines in jazz, the left-hand (low-pitched) rhythms in ragtime piano, and the isochronous pulse of the bass drum in some electronic, pop, and dance music. Specifically, we examine responses from the auditory cortex to timing deviations in higher- versus lower-pitched voices, as well as the relative influence of these timing deviations on motor synchronization to the beat in a tapping task.

In previous research, we showed a high-voice superiority effect for pitch information processing. Specifically, we presented either two simultaneous melodies or two simultaneous repeating tones, one higher and the other lower in pitch (2–6). In each case, on 25% of trials, a pitch deviant was introduced in the higher voice and, on 25% of trials, a pitch deviant was introduced in the lower voice. We recorded electroencephalography (EEG) while participants watched a silent movie and analyzed the mismatch negativity (MMN) response to the pitch deviants. MMN is generated primarily in the auditory cortex and represents the brain's automatic detection of unexpected changes in a stream of sounds (for reviews, see refs. 7–9). MMN has been measured in response to changes in features such as pitch, timbre, intensity, location, and timing. It occurs between 120 and 250 ms after onset of the deviance. That MMN represents a violation of expectation is supported by the fact that MMN amplitude increases as changes become less frequent and occurs for both decreases and increases in intensity. Because MMN is generated primarily in the auditory cortex, it manifests at the

## Significance

To what extent are musical conventions determined by evolutionarily-shaped human physiology? Across cultures, polyphonic music most often conveys melody in higher-pitched sounds and rhythm in lower-pitched sounds. Here, we show that, when two streams of tones are presented simultaneously, the brain better detects timing deviations in the lower-pitched than in the higher-pitched stream and that tapping synchronization to the tones is more influenced by the lower-pitched stream. Furthermore, our modeling reveals that, with simultaneous sounds, superior encoding of timing for lower sounds and of pitch for higher sounds arises early in the auditory pathway in the cochlea of the inner ear. Thus, these musical conventions likely arise from very basic auditory physiology.

Author contributions: M.J.H., C.M., and L.J.T. designed research; M.J.H. and C.M. performed research; M.J.H., C.M., I.C.B., and L.J.T. analyzed data; and M.J.H., C.M., I.C.B., and L.J.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>M.J.H. and C.M. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. E-mail: ljtmcmaster.ca.

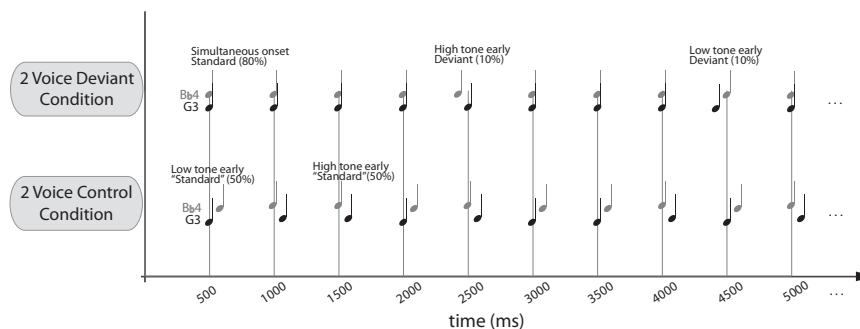
This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1402039111/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1402039111/-DCSupplemental).

surface of the head as a frontal negativity accompanied by a posterior polarity reversal. In our studies of the high-voice superiority effect, we found that MMN was generated to pitch deviants in both the high-pitch and low-pitch voices, but that it was larger to deviants in the higher-pitch voice in both musician and nonmusician adults (2–4) and in infants (5, 6). Furthermore, modeling results suggest that this effect originates early in the auditory pathway in nonlinear cochlear dynamics (10). Thus, we concluded that pitch encoding is more robust for the higher compared with lower of two simultaneous voices.

Here, we investigate whether a low-voice superiority effect holds for timing information. In addition to the widespread musical practice of using bass-range instruments to lay down the rhythmic foundation of music (11–13), a few behavioral studies suggest that lower-pitched voices dominate time processing, both in terms of perception (14) and in determining to which voice people will align body movements (15, 16). In the current study, we present the isochronous stream of two simultaneous piano tones used previously in studies of the high-voice superiority effect for pitch (3–5), but here we occasionally present the lower tone or the higher tone 50 ms too early. In an EEG experiment, we compared the MMN response from the auditory cortex to timing deviants of the lower- versus higher-pitched tone (Fig. 1). In a finger-tapping study, participants tapped in synchrony with a stimulus sequence similar to that used in the EEG study, and we measured the extent of tap-time adjustment following a timing shift in which either the higher or lower tone occurred 50 ms too early.

## Results

**MMN Experiment. Amplitude.** EEG was recorded while participants listened to two simultaneous 300-ms piano tones [G<sub>3</sub> (196.0 Hz) and B-flat<sub>4</sub> (466.2 Hz)] that repeated with sound onsets every 500 ms. On 10% of trials, the lower tone occurred 50 ms too early and, on another 10% of trials, the higher tone occurred 50 ms too early (Fig. 1). MMN amplitude was examined in a three-way ANOVA with factors voice (high-tone early, low-tone early), hemisphere (left, right), and region [frontal left (FL) and frontal right (FR); central left (CL) and central right (CR); temporal left (TL) and temporal right (TR); and occipital left (OL) and occipital right (OR)]. The significant main effect of Voice [ $F(1,16) = 16.56$ ,  $P < 0.001$ ] showed larger MMN amplitude for the low-tone early than high-tone early stimuli (Fig. 2). Moreover, the voice  $\times$  region interaction was significant [ $F(3,48) = 18.88$ ,  $P < 0.001$ ]. Post hoc tests showed a larger low-voice superiority effect ( $d$  = amplitude difference between the low- and the high-tone early deviant) over frontal ( $|d| = 0.52 \mu\text{V}$ ;  $P < 0.001$ ), central ( $|d| = 0.36 \mu\text{V}$ ;  $P = 0.03$ ), and occipital regions ( $|d| = 0.35 \mu\text{V}$ ;  $P = 0.02$ ), but no difference over the temporal regions ( $|d| = 0.24 \mu\text{V}$ ;  $P = 0.24$ ). No other effects or interactions were significant.



**Fig. 1.** Depiction of the two EEG experimental conditions. In the two-voice deviant condition, standards (80% presentation rate) consisted of simultaneous high- and low-pitched piano tones [fundamental frequencies of 196 Hz (G<sub>3</sub>) and 466.2 Hz (B-flat<sub>4</sub>)]; deviants consisted of randomly presenting one of the tones 50 ms earlier than expected (10% high-tone early and 10% low-tone early). In the two-voice control condition, only the high-tone early and the low-tone early stimuli were presented, with each occurring on 50% of trials.

**Latency.** No difference in latency was observed in the frontal right ( $P = 0.41$ ) and frontal left ( $P = 0.86$ ) regions. Moreover, no latency differences were apparent in any other regions, as can be seen in Fig. 2.

**Correlation with music training.** The duration of music training did not significantly correlate with MMN amplitude, with the size of the low-voice superiority effect, or with MMN latency (all  $P > 0.1$ ).

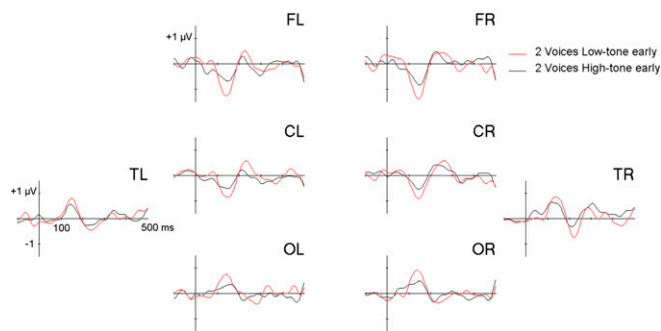
**Finger-Tapping Experiment. Motor response to timing deviations.** Participants tapped in synchrony with a pacing sequence similar to that used in the EEG experiment, wherein the higher or lower tone occasionally occurred 50 ms early (deviants). We measured the compensatory phase correction response to both deviant types: that is, the timing of the participants' taps following a timing shift of the low- versus high-tone. Measuring phase-correction effects in tapping is an established method to probe the timing of sensorimotor integration (17). The timing of the following tap derives from a weighted average of the preceding tap interval produced by a participant and the timing of the preceding tone (17).

Participants' taps were shifted in time significantly more following a lower tone that was 50 ms too early (mean tap shift = 14 ms early, SEM = 1.6) compared with a higher tone that was 50 ms too early (mean tap shift = 11 ms early, SEM = 1.4),  $t(17) = 2.72$ ,  $P = 0.014$ . This effect indicates that motor synchronization to a polyphonic auditory stimulus is more influenced by the lower-pitched stream.

**Correlation with music training.** The duration of music training did not significantly correlate with tapping variability, the response to timing deviants, or the relative difference between high and low responses (all  $P > 0.2$ ).

## Discussion

The results show a low-voice superiority effect for timing. We presented simultaneous high-pitched and low-pitched tones in an isochronous stream that set up temporal expectations about the onset of the next presentation, and occasionally presented either the higher or the lower tone 50 ms earlier than expected, while leaving the other tone at the expected time. MMN was larger in response to timing deviants for the lower than the higher tone, indicating better encoding for the timing of lower-pitched compared with higher-pitch tones at the level of the auditory cortex. A separate behavioral study showed that tapping responses were more influenced by timing deviants to the lower- than higher-pitched tone, indicating that auditory-motor synchronization is also more influenced by the lower of two simultaneous pitch streams. Together, these results indicate that the lower tone has greater influence than the high tone on determining both the



**Fig. 2.** Difference waveforms showing the MMN elicited by early presentation of the low tone (red line) and the high tone (black line). Results are plotted by region (frontal, central, temporal, and occipital) and hemisphere (L, R). Difference waveforms for high-tone early and low-tone early stimuli were calculated by subtracting each stimulus's average standard wave in the two-voice control condition from its average deviant wave in the two-voice deviant condition. Using acoustically identical stimuli as deviants in one condition and as standards in a separate condition isolates the effects of timing deviations without confounding acoustic differences.

perception of timing and the entrainment of motor movements to a beat. To our knowledge, these results provide the first evidence showing a general timing advantage for low relative pitch, and the results are consistent with the widespread musical practice of most often carrying the rhythm or pulse in bass-ranged instruments.

The results are also consistent with previous behavioral and musical evidence. For example, when people are asked to tap along with two-tone chords that contain consistent asynchronies between the voices (e.g., when one voice consistently leads by 25 ms), tap timing is more influenced by the lower tone (16). Similarly, when presented with two-voice polyrhythms in which, for example, one isochronous voice pulses twice while the other voice pulses three times (or three vs. four, or four vs. five), listeners' rhythmic interpretations are more influenced by whichever pulse train is produced by the lower-pitch voice (14). There is also evidence that, when presented with music, people align their movements to low-frequency sounds (18) and that increased loudness of low frequencies is associated with increased body movements (19). Finally, a transcranial magnetic stimulation (TMS) study indicates that "high groove" songs that highly activate the motor system also have more spectral flux in low frequencies (20).

In music, rhythm can, of course, be complex. Typically, a regular pulse or beat can be derived from a complex musical surface, and beats are perceptually grouped through accents to convey different meters, such as marches (groups of two beats) or waltzes (groups of three beats). In addition, musical rhythms can contain events whose onsets fall between perceived strong beats, and, when these onsets are salient, they are said to produce syncopation. In music, bass onsets tend to mark strong beat positions (11, 12). For example, in stride and ragtime piano, bass notes lay down the pulse on strong beats whereas higher-pitched notes play syncopated aspects of the rhythm (13, 21). Similarly, popular and dance music commonly contains an isochronous "four-on-the-floor" bass drum pulse. This pulse is critical for inducing a regular sense of beat to synchronize with (22) and provides a strong rhythmic foundation on which other rhythmic elements such as syncopation can be overlaid (23).

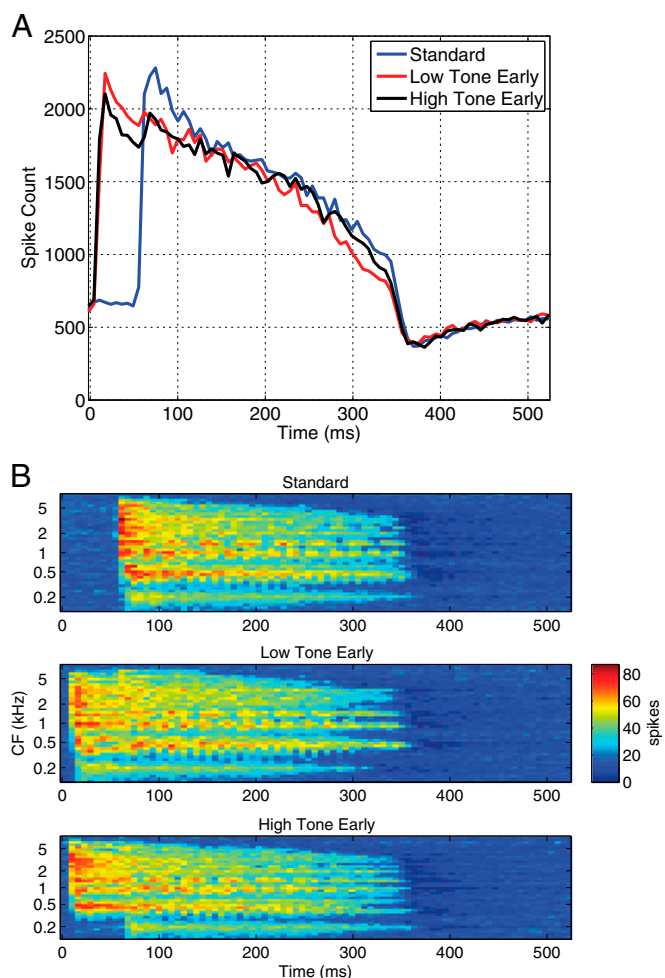
This low-voice superiority effect for timing appears to be the reverse of the high-voice superiority effect demonstrated previously for pitch (reviewed in ref. 10). In the case of pitch, MMN responses to pitch deviants were larger for the higher-pitched compared with lower-pitched of two simultaneous tones, a finding that is consistent with behavioral observations (24–26) and

the widespread musical practice of putting the main melody line in the highest-pitched voice. Interestingly, evidence indicates that the high-voice superiority effect arises in the auditory periphery and is relatively unaffected by experience. First, it is found in both musicians and nonmusicians (2–4), in 3-mo-olds (6), and in 7-mo-olds (5). Second, despite general changes in the size of MMN responses between 3 and 7 mo of age, differences in MMN amplitude between high- and low-voice deviants remain constant. Third, this difference is not correlated with musical experience at either age. Fourth, even in adult musicians who play bass-range instruments and have for years attended to the lowest-pitch part in musical ensembles, the high-voice superiority effect for pitch is not reversed, although it is reduced. Finally, modeling results show a peripheral mechanism for the origin of the high-voice superiority effect for pitch. Using the biologically plausible model of the auditory periphery of Bruce, Carney, and coworkers (27, 28), Trainor et al. (10) showed that, due to nonlinearities in the cochlea in the inner ear, harmonics of the higher tone suppress harmonics of the lower tone, resulting in superior pitch salience for the higher tone.

Keeping the same logic, one could ask whether peripheral mechanisms can explain the low-voice superiority effect for processing the timing of tones. First, however, it is important to note that the findings of the present study cannot be explained by loudness differences across frequency. Using the Glasberg and Moore (29) loudness model, we found a very similar level of loudness across stimuli with means of 86.2, 85.7, and 86.2 phons for low-tone early, high-tone early, and simultaneous-onset stimuli, respectively.

Because the deviant tones occurred sooner than expected, one obvious candidate to explain the low-voice superiority effect for timing is forward masking. In forward masking, the presentation of one sound makes a subsequent sound more difficult to detect. The well-known phenomenon of the upward spread of masking, in which lower tones mask higher tones more than the reverse (30), suggests that, in the present experiments, when the lower tone occurred early, it might have masked the subsequent higher tone more than the higher tone masked the lower tone when the higher tone occurred early. We examined evidence for such a peripheral mechanism by inputting the stimuli used in the present experiment into the biologically plausible model of the auditory periphery of Bruce, Carney, and coworkers (27, 28). Because timing precision is likely reflected by spike counts in the auditory nerve, we used spike counts as the output measure rather than the pitch salience measure used in Trainor et al. (10). As can be seen in Fig. 3A, when the lower tone began 50 ms earlier than the higher tone, the spike count at the onset of the lower tone was similar to the spike count when both tones began simultaneously in the standard stimulus, suggesting that the onsets of the low-tone early and standard (simultaneous onsets) stimuli are similarly salient at the level of the auditory nerve. Furthermore, in this low-tone early stimulus, when the higher tone began 50 ms after the lower tone, Fig. 3A shows that there was no accompanying increase in the spike count at the onset of the higher tone, because of forward masking. Thus, the model indicates that the timing of the low-tone early stimulus is unambiguously represented as the onset of the lower tone. However, the situation was different for the high-tone early stimulus. When the higher tone began before the lower tone, Fig. 3A shows that the spike count at the onset of the higher tone was a little lower than for the standard stimulus where both tones began simultaneously. Furthermore, in the high-tone early stimulus, when the low tone entered 50 ms later, the spike count increased at the onset of the low tone. Thus, in the high-tone early stimulus, the spike count increased at both the onset of the high tone and at the onset of the low tone. The timing onset of this stimulus is thereby more ambiguous compared with the case where the low tone began first.





**Fig. 3.** Spike-count results from a biologically plausible model of the auditory periphery for the three auditory stimuli (standard with simultaneous onset of high and low tones; high tone 50 ms early; and low tone 50 ms early). (A) Spike counts across time. (B) The spike counts by characteristic frequency (CF) and time.

This pattern of results can also be seen in the time-frequency representation shown in Fig. 3B. The *Top* plot shows the spike count across frequency for the simultaneous (standard) case. A clear onset is shown across frequency channels. The *Middle* plot shows the spike count across frequency for the low-tone early stimulus. Here, a clear onset is shown across frequency channels 50 ms earlier (at the onset of the lower tone) and no subsequent increase when the second higher tone begins. The lack of subsequent increase is likely because the harmonics of both tones extend from their fundamental frequencies to beyond 4 kHz so the frequency channels excited by the high tone are already excited by the lower tone. Thus, a change in the exact pattern of neural activity is observed at the onset of the high tone but the spatial extent of excitation in the nerve does not change. Finally, the *Bottom* plot shows the spike count across frequency for the high-tone early stimulus. Note that, at the onset of the higher tone, spikes occur at and above its fundamental; however, when the lower tone enters, an additional increase in spikes occurs at the lower frequencies. Thus, in the case where the higher tone begins first, two onset responses can be seen, making the timing of the stimulus more ambiguous. Greater ambiguity in the onset of the high-tone early stimulus compared with the low-tone early stimulus may have contributed to the low-voice superiority effect for timing as seen in larger MMN responses and greater tapping

adjustment for the low-tone early compared with high-tone early stimuli.

Although this modeling work suggests that nonlinear dynamics of sound processing in the cochlea of the inner ear contribute to the low-voice superiority effect observed in our data, there may be additional contributions from brainstem mechanisms. Sound masking effects can be seen in the auditory nerve (31), but they occur largely near hearing thresholds (i.e., only with very quiet sound input) whereas behaviorally measured masking (i.e., the inability to consciously perceive one sound in the presence of another sound) continues to increase at louder sound levels (32–34). Furthermore, animal models indicate that masking effects are seen in the brainstem at the level of the inferior colliculus that more closely follow behaviorally measured masking at suprathreshold levels (35). These mechanisms are not yet entirely understood, but a model involving more narrowly tuned (i.e., responding only to a specific range of sound frequencies) excitatory neural connections and more broadly tuned (i.e., responding to a wide range of sound frequencies) inhibitory connections can explain masking effects such as contextual enhancement and suppression at the level of the inferior colliculus (36). Furthermore, there is evidence that low-frequency masking sounds drive these contextual effects to a greater extent than high-frequency masking sounds (36). The effects of different frequency separations between the higher and lower tones, different fundamental frequency ranges, different intensity fall-off of harmonics, and different sizes of timing deviation could all be explored further using the model of Bruce and coworkers (27, 28). Finally, although we know of no studies to date, corticofugal neural feedback from auditory cortical areas, and possibly the cerebellum, to peripheral auditory structures might also contribute to forward masking.

In the case of the high-voice superiority effect for pitch, some evidence suggests that effects of extensive experience in that bass-range musicians showed less of an effect (although not a reversal), but innate contributions appeared to be more powerful (4). Thus, composers likely place the melody most commonly in the highest voice because, all else being equal, it will be easiest to perceive in that position. The possible role of experience-driven plasticity in the low-voice superiority effect for timing remains to be tested. Particularly if the brainstem strongly contributes to the low-voice superiority effect, it might be affected by experience more than the high-voice superiority effect, as musical training affects the precision of temporal encoding in the brainstem (37). These possibilities could be investigated further by examining infants and comparing percussionists and bass-range instrument players to soprano-range instrument players.

### Conclusion

A low-voice superiority effect for encoding timing can be measured behaviorally and at the level of the auditory cortex. We observed a larger MMN response to timing deviations in the lower-pitched compared with higher-pitched of two simultaneous sound streams, as well as stronger auditory–motor synchronization in the form of larger tapping adjustments in response to unexpected timing shifts for lower- than higher-pitched tones. This effect complements previous findings of a high-voice superiority effect for pitch, as measured by a larger MMN response to pitch changes in the higher compared with lower of two simultaneous streams. In both cases, models of the auditory periphery suggest that nonlinear cochlear dynamics contribute to the observed effects. It remains for future research to explore these mechanisms further and to examine the effects of experience on their manifestation. Together, these studies suggest that widespread musical practices of placing the most important melodic information in the highest-pitched voices, and carrying the most important rhythmic information in the lowest-

pitched voices, might have their roots in basic properties of the auditory system that evolved for auditory-scene analysis.

## Materials and Methods

**MMN Experiment. Participants.** The EEG study consisted of 17 participants (7 males, 10 females; mean age = 19.9 y, SD = 2.2). Eight additional participants were excluded due to excessive artifacts in the EEG data. After providing informed written consent to participate, subjects completed a questionnaire for auditory screening purposes and to assess linguistic and musical background (years experience and instrument). Subjects were not selected with respect to musical training, but 11 of them were amateur musicians (mean training = 4.8 y, SD = 3.5); musical training did not have a significant effect on the results.

**Stimuli.** The stimuli consisted of 300-ms computer-synthesized piano tones (Creative Sound Blaster), with fundamental frequencies of 196.0 Hz ( $G_3$ , international standard notation) and 466.2 Hz ( $B\text{-flat}_4$ ). These tones were used previously in Fujioka et al. (3) and Marie and Trainor (5).  $G_3$  and  $B\text{-flat}_4$  are 15 semitones apart (frequency ratio = 2.3), creating an interval that is neither highly consonant nor dissonant. The individual tones were equalized for loudness using the equal-loudness function (group waveforms normalized) from Cool Edit Pro software and combined to create wave files with the two tones. Stimuli were presented at a baseline tempo of 500 ms at  $\sim 60$  dB (A) measured at the location of the participant's head. Each condition was 8 min long, containing 1,088 total stimuli (trials). Fig. 1 shows the two conditions of the experiment: Two-voice deviants and two-voice control. In the two-voice deviants condition, the standards were composed of the low and high tones presented simultaneously (same onset time) and were presented on 80% of trials. On 10% of trials, high-tone early deviants were presented, in which the onset of the higher tone was shifted 50 ms earlier than onset of the lower tone. On the remaining 10% of trials, low-tone early deviants were presented, in which the lower tone was shifted 50 ms earlier than the high tone. Standards and deviants were presented in a pseudo-random order, with the constraint that two identical deviants could not occur in a row. The two-voice control condition presented only the high-tone early and low-tone early stimuli, with each occurring on 50% of trials in random order. Thus, the high-tone early and low-tone early stimuli served as deviants in the two-voice deviant condition, but as standards in the two-voice control condition. Using acoustically identical stimuli as deviants in one condition and as "standards" in a separate condition allowed us to calculate difference waveforms that isolated the effects of timing, without confounding acoustic differences (38).

**Procedure.** Participants were tested individually. The procedures were approved by the McMaster Research Ethics Board. Each participant sat in a sound-treated room facing a loudspeaker placed one meter in front of his or her head. During the experiment, participants watched a silent movie with subtitles and were instructed to pay attention to the movie and not to the sounds coming from the loudspeaker. They were also asked to minimize their movement, including blinking and facial movements, so as to decrease movement artifacts and obtain the best signal-to-noise ratio in the EEG data.

**EEG recording and processing.** EEG data were recorded at a sampling rate of 1,000 Hz from 128-channel HydroCel GSN nets (Electrical Geodesics) referenced to Cz. The impedance of all electrodes was below 50 k $\Omega$  during the recording. EEG data were bandpass filtered between 0.5 and 20 Hz (roll-off = 12 dB/octave) using EEProbe software. Recordings were referenced off-line using an average reference and then segmented into 600-ms epochs ( $-100$  to 500 ms relative to the onset of the sound). EEG responses exceeding  $\pm 70$   $\mu$ V in any epoch were considered artifacts and were excluded from the averaging. For seven subjects, we used a high-pass filter of 1.6 Hz to eliminate some very slow wave activity.

**Event-related potential data analysis.** The high-tone early and low-tone early stimuli were considered deviants in the two-voice deviant condition, but standards in the two-voice control condition. For both high-tone early and low-tone early stimuli, standards and deviants were separately averaged, and difference waveforms were computed for each participant by subtracting the average standard wave of condition two-voice control from the average deviant wave of condition two-voice deviant. This procedure captures the mismatch negativity elicited by the timing deviance, while using the same acoustic stimuli as standards and deviants. To quantify MMN amplitude, the grand average difference waveform was computed for each electrode for each deviant type (high-tone early, low-tone early). Subsequently, for statistical analysis, 88 electrodes were selected and divided into five groups for each hemisphere (left and right) representing frontal, central, parietal, occipital, and temporal regions [FL, FR, CL, CR, parietal left (PL), parietal right (PR), OR, OL, TL, and TR] (Fig. S1). Forty electrodes were not included in the groupings due to the following considerations: electrodes on the forehead

near the eyes were excluded to reduce eye movement artifacts; electrodes at the edge of the net were excluded to reduce contamination from face and neck muscle movement; and electrodes in the midline were excluded to enable comparison of the EEG response across hemispheres.

Initially, the presence of MMN was tested with  $t$  tests to determine where the difference waves were significantly different from zero. As expected, there were no significant effects at parietal sites (PL, PR) ( $P > 0.5$ ) so these regions were eliminated from further analysis. All other regions showed a clear MMN (all  $P < 0.05$ ) (Fig. 2). To analyze MMN amplitude, first the most negative peak in the right frontal (FR) region between 100 and 200 ms post-stimulus onset was determined from the grand average difference waves for both conditions, and a 50-ms time window was constructed centered at this latency. For each subject and each region, the average amplitude in this 50-ms time window for each condition was used as the measure of MMN amplitude. Finally, for each condition for each subject, the latency of the MMN was measured as the time of the most negative peak between 100 and 200 ms at the frontal regions because visual inspection showed the largest MMN amplitude in these regions. ANOVAs were conducted on amplitude and latency data. Greenhouse–Geisser corrections were applied where appropriate, and Tukey post hoc tests were conducted to determine the source of significant interactions.

**Finger-Tapping Experiment. Participants.** The finger-tapping study consisted of 18 participants (13 males, 5 females; mean age = 26.4 y, SD = 6.6). One additional participant was excluded due to highly variable and unsynchronized tapping. After providing informed written consent, participants were asked about their musical background (years experience and instrument). Subjects were not selected based on musical training, but 16 had some musical training (mean = 8.7 y, SD = 5.2). Musical training did not significantly affect results. Participants in the tapping experiment had somewhat more musical training than those in the MMN experiment, but none were professional musicians and musical training did not affect performance in either experiment.

**Phase-correction response.** The perception of timing is often studied by having participants tap a finger in synchrony with an auditory pacing sequence that contains occasional timing shifts. Occasionally shifting a sound onset (in an otherwise isochronous sequence) introduces a large tap-to-target error (i.e., participants cannot anticipate the timing shift and therefore do not tap in time with the shifted sound); participants react to this error by adjusting the timing of their following tap (for a review, see ref. 17). This adjustment, known as the phase-correction response, is largely automatic and pre-attentive (39) and starts to emerge  $\sim 100$ – $150$  ms after a perturbation (i.e., precedes the MMN) (40). It is not clear from our EEG experiment whether the increased neural response to the lower-pitch compared with higher-pitch timing deviants represents a purely perceptual phenomenon, or whether it might also impact motor synchronization. In the present finger-tapping experiment, participants tapped along to the stimulus used in the MMN study; we measured their phase-correction responses when either the high or the low tones occurred earlier than expected.

**Stimuli and procedure.** The stimuli in the finger-tapping study were the same as in the EEG study and consisted of piano tones at  $G_3$  and  $B\text{-flat}_4$ . Low and high tones were presented simultaneously (standards) or with either the high or the low tone 50 ms earlier than expected (deviants) at a baseline interonset interval (IOI) of 500 ms. Previous studies indicate that changes of this order are detectable (17). Runs lasted 60 s and contained 99 standards and 22 deviants. The order of standards and deviants was pseudo-random, with the constraint that two deviants could not directly follow each other. The experiment consisted of 16 runs and 176 timing deviants in each condition (high-tone early and low-tone early).

The pacing sequence was presented over headphones (Sony MDR-7506), and participants tapped along on a cardboard box that contained a microphone. Participants were instructed to "tap as accurately as possible with the tones, and do not try to predict the slight deviations that may occur." Participants started each run with the space bar. The entire experiment lasted  $\sim 30$  min.

**Data collection and analysis.** Stimulus presentation was controlled via a laptop running a MAX/MSP program. The taps were recorded from a microphone onto the right channel of a stereo audio file (Audacity program at 8,000 Hz sampling rate) on a separate computer. The left channel of the same audio file recorded audio triggers sent from the presentation computer that signaled trial onset, offset, and condition. The audio recordings containing taps and trial information were analyzed using Matlab. Individual taps were extracted; intertap intervals (ITIs) were calculated and aligned with the pacing sequence and the corresponding timing deviants. Intertap intervals after a deviant that were less than 400 ms or greater than 600 ms were excluded ( $\sim 1.7\%$  of all ITIs).

The response to timing deviations was calculated by subtracting the baseline interonset interval (500 ms) from the intertap interval after a timing deviant (17). For example, after a sound onset that was earlier than expected, if the following intertap interval was 490 ms, the phase correction response would be  $-10$  ms (i.e., 10 ms earlier than expected). Because all deviants were earlier than expected, for simplicity, we report the early phase-correction responses as positive values, so that larger values represent larger phase-correction responses. The magnitude of

response for the high- versus low-tone early conditions was compared with a paired samples  $t$  test.

**ACKNOWLEDGMENTS.** We thank Elaine Whiskin, Cathy Chen, and Iqra Ashfaq for help with data acquisition and processing. This research was supported by grants to L.J.T. from the Canadian Institutes of Health Research and grants to L.J.T. and I.C.B. from the Natural Sciences and Engineering Research Council of Canada. M.J.H. was supported by National Institute of Mental Health Grant T32MH016259.

- Bregman AS (1990) *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA).
- 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.
- Fujioka T, Trainor LJ, Ross B (2008) Simultaneous pitches are encoded separately in auditory cortex: an MMNm study. *Neuroreport* 19(3):361–366.
- Marie C, Fujioka T, Herrington L, Trainor LJ (2012) The high-voice superiority effect in polyphonic music is influenced by experience. *Psychomusicology* 22:97–104.
- Marie C, Trainor LJ (2013) Development of simultaneous pitch encoding: Infants show a high voice superiority effect. *Cereb Cortex* 23(3):660–669.
- Marie C, Trainor LJ (2014) Early development of polyphonic sound encoding and the high voice superiority effect. *Neuropsychologia* 57:50–58.
- Picton TW, Alain C, Otten L, Ritter W, Achim A (2000) Mismatch negativity: Different water in the same river. *Audiol Neurootol* 5(3-4):111–139.
- Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin Neurophysiol* 118(12):2544–2590.
- Schröger E (2007) Mismatch negativity: A microphone into auditory memory. *J Psychophysiol* 21:138–146.
- 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.
- Lerdahl F, Jackendoff RS (1983) *A Generative Theory of Tonal Music* (MIT Press, Cambridge, MA).
- Large EW (2000) On synchronizing movements to music. *Hum Mov Sci* 19:527–566.
- Snyder J, Krumhansl CL (2001) Tapping to ragtime: Cues to pulse finding. *Music Percept* 18:455–489.
- Handel S, Lawson GR (1983) The contextual nature of rhythmic interpretation. *Percept Psychophys* 34(2):103–120.
- Repp BH (2003) Phase attraction in sensorimotor synchronization with auditory sequences: Effects of single and periodic distractors on synchronization accuracy. *J Exp Psychol Hum Percept Perform* 29(2):290–309.
- Hove MJ, Keller PE, Krumhansl CL (2007) Sensorimotor synchronization with chords containing tone-onset asynchronies. *Percept Psychophys* 69(5):699–708.
- Repp BH (2005) Sensorimotor synchronization: A review of the tapping literature. *Psychon Bull Rev* 12(6):969–992.
- Burger B, Thompson MR, Luck G, Saarikallio S, Toiviainen P (2013) Influences of rhythm- and timbre-related musical features on characteristics of music-induced movement. *Front Psychol* 4:183.
- Van Dyck E, et al. (2013) The impact of the bass drum on human dance movement. *Music Percept* 30:349–359.
- Stupacher J, Hove MJ, Novembre G, Schütz-Bosbach S, Keller PE (2013) Musical groove modulates motor cortex excitability: A TMS investigation. *Brain Cogn* 82(2):127–136.
- Iyer V (2003) Embodied mind, situated cognition, and expressive microtiming in African-American music. *Music Percept* 19:387–414.
- Honing H (2012) Without it no music: Beat induction as a fundamental musical trait. *Ann N Y Acad Sci* 1252:85–91.
- Pressing J (2002) Black Atlantic rhythm: Its computational and transcultural foundations. *Music Percept* 19:285–310.
- Zenatti A (1969) *Le Développement Génétique de la Perception Musicale* (CNRS, Paris).
- Palmer C, Holleran S (1994) Harmonic, melodic, and frequency height influences in the perception of multivoiced music. *Percept Psychophys* 56(3):301–312.
- Crawley EJ, Acker-Mills BE, Pastore RE, Weil S (2002) Change detection in multi-voice music: The role of musical structure, musical training, and task demands. *J Exp Psychol Hum Percept Perform* 28(2):367–378.
- Zilany MS, Bruce IC, Nelson PC, Carney LH (2009) A phenomenological model of the synapse between the inner hair cell and auditory nerve: Long-term adaptation with power-law dynamics. *J Acoust Soc Am* 126(5):2390–2412.
- Ibrahim RA, Bruce IC (2010) Effects of peripheral tuning on the auditory nerve's representation of speech envelope and temporal fine structure cues. *The Neurophysiological Bases of Auditory Perception*, eds Lopez-Poveda EA, Palmer AR, Meddis R (Springer, New York), pp 429–438.
- Glasberg BR, Moore BJC (2002) A model of loudness applicable to time-varying sounds. *J Audio Eng Soc* 50:331–342.
- Moore BJC (2013) *An Introduction to the Psychology of Hearing* (Elsevier Academic Press, London), 6th Ed.
- Relkin EM, Turner CW (1988) A reexamination of forward masking in the auditory nerve. *J Acoust Soc Am* 84(2):584–591.
- Lüscher E, Zwislocki J (1949) Adaptation of the ear to sound stimuli. *J Acoust Soc Am* 21:135–139.
- Jesteadt W, Bacon SP, Lehman JR (1982) Forward masking as a function of frequency, masker level, and signal delay. *J Acoust Soc Am* 71(4):950–962.
- Plack CJ, Oxenham AJ (1998) Basilar-membrane nonlinearity and the growth of forward masking. *J Acoust Soc Am* 103(3):1598–1608.
- Nelson PC, Smith ZM, Young ED (2009) Wide-dynamic-range forward suppression in marmoset inferior colliculus neurons is generated centrally and accounts for perceptual masking. *J Neurosci* 29(8):2553–2562.
- Nelson PC, Young ED (2010) Neural correlates of context-dependent perceptual enhancement in the inferior colliculus. *J Neurosci* 30(19):6577–6587.
- Kraus N, Chandrasekaran B (2010) Music training for the development of auditory skills. *Nat Rev Neurosci* 11(8):599–605.
- 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.
- Repp BH, Keller PE (2004) Adaptation to tempo changes in sensorimotor synchronization: Effects of intention, attention, and awareness. *Q J Exp Psychol A* 57(3):499–521.
- Repp BH (2011) Temporal evolution of the phase correction response in synchronization of taps with perturbed two-interval rhythms. *Exp Brain Res* 208(1):89–101.