EJN European Journal of Neuroscience

FENS Federation of European Neuroscience Societies

European Journal of Neuroscience, Vol. 47, pp. 321–332, 2018 doi:10.1111/ejn.13826

COGNITIVE NEUROSCIENCE

Neural bases of rhythmic entrainment in humans: critical transformation between cortical and lower-level representations of auditory rhythm

Sylvie Nozaradan,[1,2,3] Marc Schönwiesner,[3,4,5] Peter E. Keller,[1] Tomas Lenc[1] and Alexandre Lehmann[3,4,5,6]
[1]MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Locked Bag 1797, Penrith, Sydney, NSW 2751, Australia
[2]Institute of Neuroscience (IONS), Université catholique de Louvain (UCL), Louvain, Belgium
[3]International Laboratory for Brain, Music and Sound Research (BRAMS), Montreal, QC, Canada
[4]Center for Research on Brain, Language and Music (CRBLM), Montreal, QC, Canada
[5]Faculty of Psychology, Université de Montréal, Montreal, QC, Canada
[6]Otolaryngology Department, Faculty of Medicine, McGill University Hospital, Montreal, QC, Canada

Keywords: EEG, frequency-following responses, frequency-tagging, human auditory brainstem, rhythm perception

Abstract

The spontaneous ability to entrain to meter periodicities is central to music perception and production across cultures. There is increasing evidence that this ability involves selective neural responses to meter-related frequencies. This phenomenon has been observed in the human auditory cortex, yet it could be the product of evolutionarily older lower-level properties of brainstem auditory neurons, as suggested by recent recordings from rodent midbrain. We addressed this question by taking advantage of a new method to simultaneously record human EEG activity originating from cortical and lower-level sources, in the form of slow (< 20 Hz) and fast (> 150 Hz) responses to auditory rhythms. Cortical responses showed increased amplitudes at meter-related frequencies compared to meter-unrelated frequencies, regardless of the prominence of the meter-related frequencies in the modulation spectrum of the rhythmic inputs. In contrast, frequency-following responses showed increased amplitudes at meter-related frequencies only in rhythms with prominent meter-related frequencies in the input but not for a more complex rhythm requiring more endogenous generation of the meter. This interaction with rhythm complexity suggests that the selective enhancement of meter-related frequencies does not fully rely on subcortical auditory properties, but is critically shaped at the cortical level, possibly through functional connections between the auditory cortex and other, movement-related, brain structures. This process of temporal selection would thus enable endogenous and motor entrainment to emerge with substantial flexibility and invariance with respect to the rhythmic input in humans in contrast with non-human animals.

Introduction

Evolution has endowed animals with the ability to interact in real time with external signals that rhythmically stimulate their sensory organs. In humans, this interaction is well illustrated by the common behaviour of moving to a musical rhythm, a form of auditory-motor entrainment shared by all cultures (London, 2004; Repp & Su, 2013). In particular, even when music is not strictly periodic, humans spontaneously organize the constituent complex sound sequences according to a pulse-like periodic meter and experience the urge to move to these pulses (e.g. bobbing the head, tapping the hand or moving the whole body periodically when listening to music; see Burger *et al.*, 2017). This perceptual organization of rhythm into periodic units can thus be considered a cornerstone of the entrainment to music.

There is increasing evidence that perception and motor entrainment to metric periodicities involve neural mechanisms that shape the representation of rhythmic inputs by selectively enhancing neural activity elicited at meter-related frequencies, regardless of the physical prominence of these frequencies in the rhythmic input (Nozaradan *et al.*, 2011, 2012b, 2016c, 2017c; Nozaradan, 2014; Chemin *et al.*, 2014; Celma-Miralles *et al.*, 2016; Stupacher *et al.*, 2017; Tal *et al.*, 2017). Recently, intracranial depth-electrode EEG

Correspondence: Dr Sylvie Nozaradan, [1]MARCS Institute for Brain, Behaviour and Development, as above. E-mail: sylvie.nozaradan@uclouvain.be

Received 20 September 2017, revised 12 January 2018, accepted 15 January 2018

Edited by Ali Mazaheri. Reviewed by Takako Fujioka, Stanford University, USA; and Malte Wöstmann, University of Lübeck, Germany

All peer review communications can be found with the online version of the article.

recordings in humans revealed that processes from the earliest auditory cortical areas (Heschl's gyrus) shape the neural representation of rhythmic inputs in favour of increased relative amplitude of the activity corresponding to meter-related frequencies (Nozaradan *et al.*, 2016c). This selective enhancement of meter-related frequencies may require connections between the auditory cortex and other brain structures such as the premotor cortex, the basal ganglia and the supplementary motor area, which are involved in beat-based timing (Grahn, 2012; Merchant & Honing, 2014; Patel & Iversen, 2014; Ross *et al.*, 2016; Nozaradan *et al.*, 2017b). However, this selective enhancement may also be the product of lower-level properties of auditory neurons that process sound input before the cortex. One of the candidates for these processes could be the inferior colliculus, a subcortical structure of the human ascending auditory pathway located in the brainstem. This subcortical auditory structure has been shown to contribute to auditory temporal processing by providing a level of auditory neural precision that is crucial for accurate auditory-motor temporal integration (Tierney & Kraus, 2013, 2016). This feedforward route, with highly accurate temporal precision, modulated itself via corticofugal fibres in a continuous feedforward-feedback loop (Winer, 2006; Bajo & King, 2013), could be critical in auditory-motor synchronization. But these subcortical auditory structures could also be critical for processing rhythmic inputs by enhancing a particular subset of their constituent acoustic events as a basis for meter perception. Relatedly, recent models have been proposed to quantify the extent to which beat perception could be explained by early auditory sensory processing in brainstem auditory nuclei that we share with other animals (Rajendran *et al.*, 2017). This hypothesis is also related to a line of research investigating the extent to which musical conventions and habits are determined by evolutionarily shaped human physiology (Hove *et al.*, 2014; Merchant & Honing, 2014; Wojtczak *et al.*, 2017).

The goal of the current study was to investigate the contribution of these lower-level auditory subcortical stages in the neural representation of rhythm. To this end, we recorded EEG activity while healthy young adults listened to rhythmic sequences. This activity, typically observed at a slow time scale (< 20 Hz) in response to the envelope of acoustic rhythms, is generally compatible with cortical sources (Draganova *et al.*, 2002; Nozaradan *et al.*, 2015). Lower-level neural representations of the rhythms were also captured from the same recordings, in the form of frequency-following responses typically observed at a faster time scale (> 150 Hz) in response to the carrier tones and their modulation over time (see Nozaradan *et al.*, 2016b for a similar method). This faster activity is compatible with a mixture of auditory cortical and brainstem sources, with a primary neural generator in the inferior colliculus (Chandrasekaran & Kraus, 2010; Coffey *et al.*, 2016; Tierney & Kraus, 2016). Activity at these slow and fast time scales was measured in response to a regular rhythm presented at a moderate or fast tempo, and a more irregular, syncopated, rhythm presented only at moderate tempo. After the EEG session, the participants were asked to tap the beat of the rhythms, as a behavioural index of rhythmic motor entrainment. Frequency- and time-domain analyses of these different measures allowed us to investigate cortical and frequency-following responses, their sensitivity to the tempo and rhythmic structure of the inputs and their relationship with motor entrainment behaviour. A selective enhancement of meter-related frequencies in the frequency-following responses would indicate that subcortical auditory properties play a critical role in these processes of temporal selection. Alternatively, such meter-related enhancement could reflect the contribution of cortical activity to these lower-level auditory responses through efferent connections. Conversely, an absence of meter-related enhancement in lower-level auditory responses, especially in the more complex syncopated rhythm, would indicate that this selective process critically involves higher-level cortical areas, enabling endogenous and motor rhythmic entrainment.

Materials and methods

Participants

Twenty healthy volunteers (11 females, one left-handed, mean age 22.9 ± 3.2 years, aged between 18 and 30) took part in the study after providing written informed consent. These volunteers were either musical amateurs, with only informal experience as listeners or dancers, or were considered as musicians according to their number of years of formal music training (13 participants with 7.6 ± 5.4 years). None of the participants had prior experience with the tapping task used in the present study. All had no history of hearing, neurological or psychiatric disorder and were not taking any drug at the time of the experiment. The study was approved by the Research Ethics Committee of the Faculty for Arts and Sciences of the University of Montreal and conducted at that university.

Auditory stimulation

Stimuli were generated using MATLAB (The MathWorks, Natick, USA) and presented binaurally through electromagnetically shielded insert earphones at 75 dB SPL (ER 2, Etymotic Research, Elk Grove Village, IL, USA).

The auditory stimulus consisted of 40.8-s sequences of a repeated non-harmonic chord with three partials (f1 = 243 Hz, f2 = 398 Hz and f3 = 566 Hz). These frequencies were chosen because frequency-following responses at these tone frequencies recorded in humans with scalp EEG are most likely to originate from brainstem nuclei of the ascending auditory pathway due to the low-pass limitation in the production of sustained frequency-locked responses to sounds in the human auditory cortex (Skoe & Kraus, 2010; but see Coffey *et al.*, 2016 for recent data on the contribution of cortical neural populations to frequency-following responses recorded from the human scalp in a frequency range below 150 Hz). A non-harmonic relationship between the three tones was chosen to avoid overlap of the harmonics of each of these frequencies across partials. This non-harmonic chord was also expected to elicit frequency-locked responses in the scalp EEG at frequencies that were not present in the stimulus spectrum. These responses, corresponding to distortion-product otoacoustic emissions generated at the cochlear level and transmitted over the ascending auditory pathway (Skoe & Kraus, 2010), were expected at frequencies corresponding to the cross-modulation across the three partials, that is f3-f2, f2-f1 and f3-f1 (168, 155 and 323 Hz, respectively).

The non-harmonic chord was modulated in amplitude in such a way as to generate three different non-isochronous sequences. These amplitude modulations consisted in three different rhythmic patterns continuously looped throughout the 40.8-s sequences. Two of these rhythmic patterns (henceforth referred to as 'rhythm #1' and 'rhythm #2') lasted 2.4 s and were composed of 12 short time intervals of 200-ms duration. These intervals were either 'acoustic' (linear onset/offset ramps lasting 10% of the interval duration, with a modulation depth of 90%, from 0.1 to 1) or 'silent' (amplitude maintained at 0.1; Fig. 1A). The rhythms used in the current study were selected based on previous evidence that they induce a meter perceived at a consistent periodicity across individuals based on grouping by four intervals (4 × 200 ms) and related metrical periodicities (beat subdivision by 2 and 4, beat integer grouping by 3; Povel & Essens,

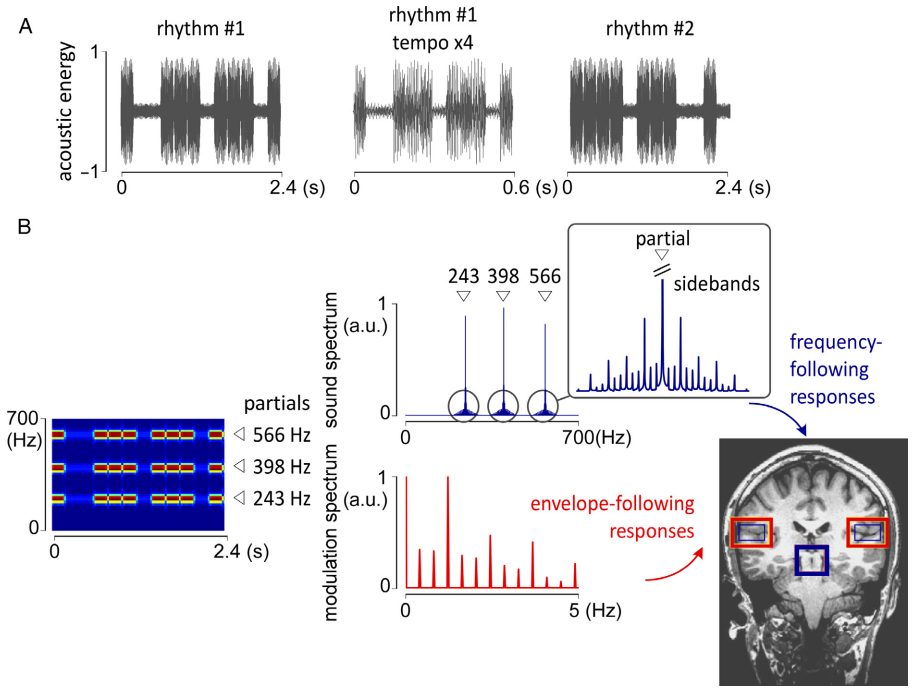

FIG. 1. Experimental design and stimulus. (A) Three rhythmic patterns, consisting of non-isochronous amplitude modulation between 0.1 and 1 (rhythm #1, unsyncopated, played at moderate and fast tempo, and rhythm #2, syncopated, played at moderate tempo). (B) Time–frequency map of rhythm #1 (left panel). The amplitude modulation was carried by three non-harmonic partial sine tones (243, 398 and 566 Hz). When repeated in a 40.8-s sequence, the spectrum of this rhythmic sequence (in blue, upper middle panels) exhibits three peaks at the partials' frequencies, each flanked by sidebands at frequencies corresponding to each partial [+/-]12 frequencies constituting the modulation spectrum. EEG activities at these frequencies are expected in the form of frequency-following responses. The envelope spectrum (in red, lower middle panel) presents 12 peaks ranging from the frequency corresponding to the period of the entire rhythmic pattern to the frequency corresponding to the period of the shortest interval in the rhythm. EEG envelope-following responses corresponding to cortical activity are expected at these frequencies. Neuroimaging scan (bottom right) used here to illustrate the most probable sources of these activities (brainstem and auditory cortex). [Colour figure can be viewed at wileyonlinelibrary.com].

1985; Nozaradan *et al.*, 2012b, 2016a). Rhythm #1 was considered 'unsyncopated', as the meter coincided with the structural arrangement of onsets of 'acoustic' intervals. The internal representation of beat thus matches physical cues in this rhythm. In contrast, rhythm #2 was considered syncopated, as the meter does not systematically correspond to the onset structures. This rhythm can thus be considered more complex, as the internal representation of meter matches external cues to a lesser extent.

The third rhythmic sequence was identical in all respects to rhythm #1, except for the fact that the duration of the shortest interval was of 50 ms instead of 200 ms (Fig. 1A). This rhythm was thus presented four times faster than the first one, hence further labelled 'rhythm #1 tempo x4', enabling us to test the sensitivity of neural activity to the tempo of presentation of rhythm. Critically, this tempo was set above the ecological range at which musical meter is usually induced (> 200 ms periodicity; e.g. van Noorden & Moelants, 1999; McAuley, 2010). At this speeded-up tempo, the perceived meter should no longer be driven by a grouping by four individual intervals, as such grouping would correspond to 200 ms (4 × 50 ms), which is too fast to spontaneously generate comfortable perceptual and motor entrainment. Instead, the perceived meter should coincide with an integer multiple grouping of intervals, namely, a grouping by 12 intervals, thus corresponding to 600 ms (Nozaradan *et al.*, 2012b).

To determine the frequencies at which the cortical frequency-tagged responses to the acoustic envelope were expected in the EEG, the temporal envelope of the three 40.8-s auditory stimuli (rhythm #1, rhythm #1 tempo x4 and rhythm #2) was extracted using the Hilbert transform as implemented in MATLAB and then transformed into the frequency domain using a discrete Fourier transform, yielding a frequency spectrum of acoustic energy (Cirelli *et al.*, 2016; Nozaradan *et al.*, 2016a,b,c, 2017a,b). The frequencies of interest appeared in the obtained spectrum as 12 peaks of acoustic energy ranging from the frequency corresponding to the period of an entire rhythm cycle to the period of the shortest interval for each rhythm and tempo (Fig. 1B). In this set of 12 frequencies, the meter-related frequencies thus coincide with the first, third, sixth and twelfth frequencies (as defined above, i.e. corresponding to grouping by four intervals, subdivision of this group by 2 and 4, multiple of this group by 3), whereas the other frequencies are considered meter-unrelated.

To determine the frequencies at which the frequency-following responses were expected in the EEG, the auditory stimulus was transformed in the frequency domain using a discrete Fourier transform. The frequencies of interest appeared in the obtained frequency spectrum as three peaks at 243, 398 and 566 Hz, flanked by sidebands at [+/-] the 12 frequencies constituting the sound envelope spectrum (Fig. 1B). These sidebands thus corresponded to each partial amplitude-modulated by the rhythmic envelope. In addition, frequency-following responses were also expected at frequencies not present in the stimulus spectrum but elicited due to distortion-product otoacoustic emissions, that is at f3-f2, f2-f1 and f3-f1, flanked by sidebands at [+/-] the 12 frequencies corresponding to the amplitude modulation of these tones.

Finally, two additional sequences were presented to the participants for control purposes. One sequence was identical in all respects to rhythm #1, except for the fact that the rhythmic amplitude modulation was not carried by the non-harmonic chord, but by continuous white noise. This 'white noise sequence' was thus

expected to elicit cortical responses at frequencies corresponding to the modulation spectrum, but no frequency-following responses at 243, 398 and 566 Hz and sidebands or at distortion-product frequencies and sidebands. The other control sequence consisted in the same non-harmonic chord as in rhythms #1 and 2, but with no amplitude modulation over time (amplitude maintained at 1 during 40.8 s). This 'continuous tone sequence' was thus expected to elicit frequency-following responses at 243, 398 and 566 Hz or at distortion-product frequencies, but no sidebands and no cortical responses at frequencies corresponding to the modulation spectrum.

Experimental procedure

Participants were comfortably seated in a chair with their head resting on a support. They were instructed to relax, avoid any unnecessary head or body movement and keep their eyes fixed on a point displayed on the wall in front of them. The five different sequences (rhythm #1, rhythm #1 tempo x4, rhythm #2, white noise and continuous tone) were presented in a block design, with the order of the blocks counter-balanced across participants. Each block consisted of 11 trials, each starting with 3 s of silence, followed by the 40.8-s sound stimulus. The polarity of the sound waveform was inverted on every other trial to prevent potential electrical artefacts at the frequencies of the sound input (Skoe & Kraus, 2010). The sound sequences were presented with electrically shielded insert earphones to further avoid electrical artefacts due to the sound transmission device. Each trial onset was self-initiated. To ensure that participants focused their attention on the sound in the auditory task, they were asked to carefully listen to the sound and report any irregularity in the sound and specifically in the sound duration at the end of each trial. There were no actual irregularities in any of the sequences (and hence there was no extensive assessment of the detection performance), but probably due to the long-lasting repetition of the sound participants reported hearing subtle changes throughout the trial in half of the trials on average (Nozaradan *et al.*, 2016b).

In the eleventh trial of each block, the participants were asked to perform a tapping task with their right hand as consistently and accurately as possible on the beat they felt in the sequence (Nozaradan *et al.*, 2012b, 2016a). Participants were asked to start tapping as soon as they could feel a beat in the sequence, and to maintain the synchronized movement as well as possible throughout the entire trial. Tapping was performed with small up and down movements of the hand from the wrist joint, maintaining the forearm and elbow in fixed position on an armrest cushion. When performing the tapping movement, the fingers of the tapping hand came transiently in contact with a sensor pad which continuously recorded finger pressure. The experimenter regularly went to the recording room, to monitor compliance to these instructions. The tapping trial was not used for the EEG analysis, but was used as a behavioural index of motor entrainment to the beat, allowing us to confirm that the periodicity selected across participants for this motor entrainment task corresponded to the theoretical assumption of a frequency corresponding to grouping by four intervals or related harmonics.

EEG recording

The EEG was recorded using 64 active sintered Ag-AgCl electrodes placed on the scalp according to the International 10/10 system (ActiveTwo, BioSemi, The Netherlands). Active electrodes contain the first amplifier stage within the electrode cover and provide impedance transformation on the electrode to prevent interference currents from generating significant impedance-dependent nuisance voltages. We therefore did not control electrode impedances but rather kept direct-current offset close to zero during electrode placement. Vertical and horizontal eye movements were monitored using three additional electrodes placed on the outer canthus of each eye and on the inferior area of the left orbit. The recorded electrode signals, referenced to the common mode sense (CMS), were amplified, sampled at 8192 Hz (ActiveTwo amplifier, BioSemi, The Netherlands) and stored for offline analysis using BIOSEMI ACTIVIEW software. To facilitate computer processing of the recorded data, the signal of three EEG channels (Fz, FCz and Cz) was exported at 8192 Hz for further analysis, as these electrodes are known to capture the major part of auditory frequency-following and cortical responses (Skoe & Kraus, 2010). Those signals were then rereferenced to average mastoids, a standard montage to measure these auditory responses (Skoe & Kraus, 2010).

Frequency-domain analysis of EEG responses

EEG processing was carried out using LETSWAVE6 (www.nocions.org/letswave6) running under MATLAB (The MathWorks, Natick, USA). Continuous EEG recordings were filtered using a 0.1-Hz high-pass fast-Fourier transform (FFT) filter to remove slow drifts in the recorded signals. Epochs lasting 40.8 s were obtained by segmenting the recordings from +0 to +40.8 s relative to the onset of the auditory stimulus, thus yielding 10 epochs for each participant and condition. Baseline correction of these epochs was computed by subtracting the amplitude of each epoch by the average amplitude measured from −0.5 to −0.2 s relative to the onset of the auditory stimulus. A notch FFT filter centred on 60 Hz with a width of 1 Hz was used to remove the artefact due to the electric line noise, and each epoch was also submitted to a DC removal, to set the average amplitude to be aligned to zero.

For each participant and condition, EEG epochs were averaged across trials. The time-domain averaging procedure was used to enhance the signal-to-noise ratio of the EEG activity elicited in response to the stimulus by attenuating the contribution of activity that was not phase-locked to the sound stimulus across trials. The obtained averaged waveforms were then transformed into the frequency domain using a discrete Fourier transform, thus yielding spectra with 0.02 Hz resolution. These spectra were normalized by subtracting the average amplitude measured at −3 to −13 and +3 to +13 frequency bins relative to each frequency bin across the *x*-axis of the spectra. This procedure is based on the assumption that in the absence of a response the signal amplitude at a given frequency should be similar to the signal amplitude of the mean of the surrounding frequency bins (Mouraux *et al.*, 2011; Retter & Rossion, 2016; Nozaradan *et al.*, 2017a). For each condition and participant, the obtained spectra were then averaged across the three scalp electrodes. The magnitude of the responses was then estimated at the expected response frequencies, based on the spectrum of the sound and its amplitude modulation.

The magnitude of the cortical responses (henceforth referred to as 'envelope-following responses') was estimated at the 12 frequencies constituting the envelope modulation spectrum of each rhythmic sequence. The magnitude of the frequency-following responses was estimated at the frequency of the three partials of the chord and their distortion products, and at the sideband frequencies.

Statistical evaluation

Statistical analyses were performed using SPSS Statistics 21.0 (IBM, Armonk, NY, USA). Significance level was set at $P < 0.05$. When

relevant, the Greenhouse–Geisser correction was used to correct for violations of sphericity in the performed ANOVAs.

We first examined whether the auditory stimulus elicited significant envelope-following and frequency-following responses at the expected frequencies in the EEG spectra for the five conditions (i.e. the five sequences: rhythm #1, rhythm #1 tempo x4, rhythm #2, white noise and continuous tone sequences). For each condition, a one-sample *t*-test was used to determine whether the amplitudes of the responses (microVolts) were significantly different from zero across participants. Indeed, in the absence of significant response to the input, the averaged noise-subtracted signal amplitude may be expected to tend towards zero (Mouraux *et al.*, 2011; Nozaradan *et al.*, 2017a). This test was performed by taking the amplitude averaged separately across the 12 frequencies corresponding to the modulation spectrum of the sequences, across the three partials of the chord, across the three distortion products and across their sidebands, respectively.

The general goal of the study, however, was to compare cortical and lower-level representations of the rhythms. The cortical representation consisted in the relative amplitude across the 12 frequencies corresponding to the modulation spectrum of each rhythm (standardized as *z* score values; Nozaradan *et al.*, 2016a; Cirelli *et al.*, 2016). The frequency-following representation consisted in the relative amplitude across the 12 sideband frequencies corresponding to the modulation spectrum of each rhythm. These 12 standardized amplitudes were obtained for each participant and rhythm by first averaging the amplitudes across corresponding positive and negative symmetrical sidebands of each partial and distortion product, then by standardizing the obtained values as *z* score values. These envelope-following and frequency-following responses were then compared using a 2 × 12 repeated-measures ANOVA for each rhythm, with a factor 'envelope- vs. frequency-following responses' (2 levels) and a factor 'Frequency' (12 levels for the 12 frequencies elicited by each rhythm). An interaction between the two factors would indicate an overall significantly different representation of the rhythm between brainstem and cortex, reflected in a significantly different distribution of the amplitudes across frequencies.

Furthermore, we specifically tested the hypothesis of a relative enhancement of responses at meter-related frequencies and relative decrease of responses elicited at meter-unrelated frequencies (Nozaradan *et al.*, 2017c). This hypothesis was tested with a 2 ('envelope-following vs. frequency-following responses') × 2 ('Meter-related vs. unrelated frequencies') repeated-measures ANOVA. An interaction between the two factors would indicate that the contrast between meter-related and unrelated frequencies is significantly different at cortical and lower levels. Finally, we calculated the distance of EEG responses from the acoustic input, to test the hypothesis of a greater distance from the sound (i.e. greater degree of transformation of the input) for cortical responses compared to frequency-following responses. This distance was calculated for each rhythm and participant as the difference between the standardized values of the sound and the cortical vs. frequency-following responses at meter-related frequencies. These measures were then compared using paired *t*-tests.

Time-domain analysis of EEG responses

The time-domain analysis aimed to visualize the waveform of the responses over time and estimate the amplitude over individual intervals composing the rhythms. The time course of the cortical responses was obtained using a low-pass filter at 30 Hz (FFT filter, 1-Hz width). Epochs averaged across the three channels and across trials were then segmented into chunks corresponding to the time window between the onsets of the repeated rhythmic pattern in the sequences (starting at + 2.4 s relative to the onset of the entire sound sequences, to discard the evoked potentials elicited by the onset of the sound sequence). The time course of frequency-following responses was obtained the same way, except that the EEG signal was first filtered between 134 and 354 Hz (band-pass FFT filter, 1-Hz width) to isolate the frequency range containing the significant frequency-following responses, and the amplitude of this signal was then extracted using the Hilbert transform (Nozaradan *et al.*, 2016b).

From these 2.4-s epochs, the amplitudes were measured as the root mean square amplitude of the frequency-following responses, and as the amplitude of the peak of maximum prominence of the cortical responses, for each of the twelve 200-ms individual intervals composing the rhythms (Lehmann *et al.*, 2016). This latter measure was used for cortical responses as it allowed estimation of the waveform's 'peakiness' regardless of the mean amplitude within individual intervals. Three different comparisons were tested on these measures. First, the amplitudes of responses were compared across 'acoustic' and 'silent' intervals (repeated-measures ANOVAs), with the hypothesis that the frequency-following responses would show reduced amplitudes to 'silent' intervals as compared to cortical responses (Lehmann *et al.*, 2016).

Furthermore, the amplitudes of responses to 'acoustic' and 'silent' intervals were merged across rhythms and sorted according to the number of immediately preceding 'acoustic' intervals in the rhythm. These categories were then compared to measure possible effects of neural adaptation due to repetitions of immediately preceding 'acoustic' intervals, that is to test whether the responses to 'acoustic' and 'silent' intervals were reduced or enhanced when immediately preceded by a cluster of 'acoustic' intervals. Because the number of measures was not identical across categories (the number of preceding 'acoustic' intervals was not exactly the same across rhythms), these categories were compared using Kruskal–Wallis ANOVAs.

Finally, the amplitudes of responses were also compared between on-beat and off-beat intervals, that is, to test the difference in amplitude between responses coinciding with a beat and responses to intervals not coinciding with a beat (Rajendran *et al.*, 2017). This on-beat vs. off-beat difference was calculated at the four different beat locations possible for a three-beat meter (as theoretically expected here; see Nozaradan *et al.*, 2012b, 2016a). In other words, if the meter is expected to correspond to a three-beat meter, 12-interval rhythms offer four different possible sets of beat position or phase (either at intervals #1-5-9, or #2-6-10, or #3-7-11, or #4-8-12). One-way ANOVAs comparing across these possible beat positions were thus used to test whether the expected relative enhancement at meter-related frequencies observed in the frequency domain was reflected in a consistent increase at specific equally spaced intervals across participants.

Tapping analysis

As for the EEG signals, the tap sensor signals were segmented into 40.8-s epochs aligned to the onset of the sound sequences and transformed into the frequency domain using a discrete Fourier transform. The obtained spectra were then normalized by subtracting the average amplitude measured at −3 to −13 and +3 to +13 frequency bins relative to each frequency bin across the *x*-axis of the spectra.

Mean intertap intervals and the coefficient of variation of these intervals were also calculated for each sequence and participant, by taking the latency of the onset of finger pressure of each tap (i.e. the moment at which the fingers contact the sensor pad at each tap). As tapping was expected to be close to isochronous based on the

instruction given to the participants, we also examined the phase of this tapping, that is, the position of the target intervals within the rhythms. These target intervals were defined as the intervals coinciding with the highest percentage of taps (i.e. interval onset closest to the tap).

Results and statistical analyses

Sound analysis

The modulation spectrum of rhythm #1 (unsyncopated) presented at moderate and fast tempo is characterized by prominent meter-related frequencies as compared to meter-unrelated frequencies (Fig. 2, upper graph, dashed lines), especially a prominent frequency corresponding to the grouping by four intervals. In contrast, rhythm #2 (syncopated) shows virtually no prominence of meter-related frequencies compared to meter-unrelated frequencies.

Validity of the method to capture simultaneous cortical and frequency-following responses to rhythm

We first tested whether EEG responses were elicited at the frequencies of interest. As expected, significant envelope-following responses were elicited in all conditions ($P < 0.0001$; one-sample *t*-test against zero after noise subtraction), except in the 'continuous tone sequence' with no amplitude modulation (control condition; $P = 0.266$; see Table 1 for details). In contrast, no significant frequency-following responses to the three partials (f1, f2, f3) were elicited ($P > 0.053$ for all sequences). This is probably due to the fact that the partials were not harmonically related, which has been shown to yield frequency-following responses with a substantially reduced signal-to-noise ratio as compared to responses elicited by a harmonic chord in the same frequency range (see, for instance, Bidelman & Krishnan, 2011). However, significant responses at frequencies corresponding to distortion products (otoacoustic emissions at f2-f1, f3-f1 and f3-f2) were elicited in all sequences ($P < 0.0001$) except the 'white noise sequence' with no partials and thus no expected distortion product (control condition; $P = 0.510$). Moreover, these cross-modulation responses were surrounded by significant sideband responses in all sequences ($P < 0.0003$), except in the 'white noise sequence' with no expected cross-modulation product and in the 'continuous tone sequence' with no amplitude modulation (control conditions; $P = 0.315$ and $P = 0.544$, respectively). To summarize, these results corroborate previous evidence that valid estimates of concurrent cortical and frequency-following responses to rhythmic sequences can be recorded simultaneously with this method.

Comparing envelope-following and frequency-following responses to rhythm

As shown at the group-level average of the frequency spectra (Fig. 2), the obtained spectra were highly similar across the sound signals, the cortical and frequency-following responses and the tapping signal, for rhythm #1 (unsyncopated) in which meter-related

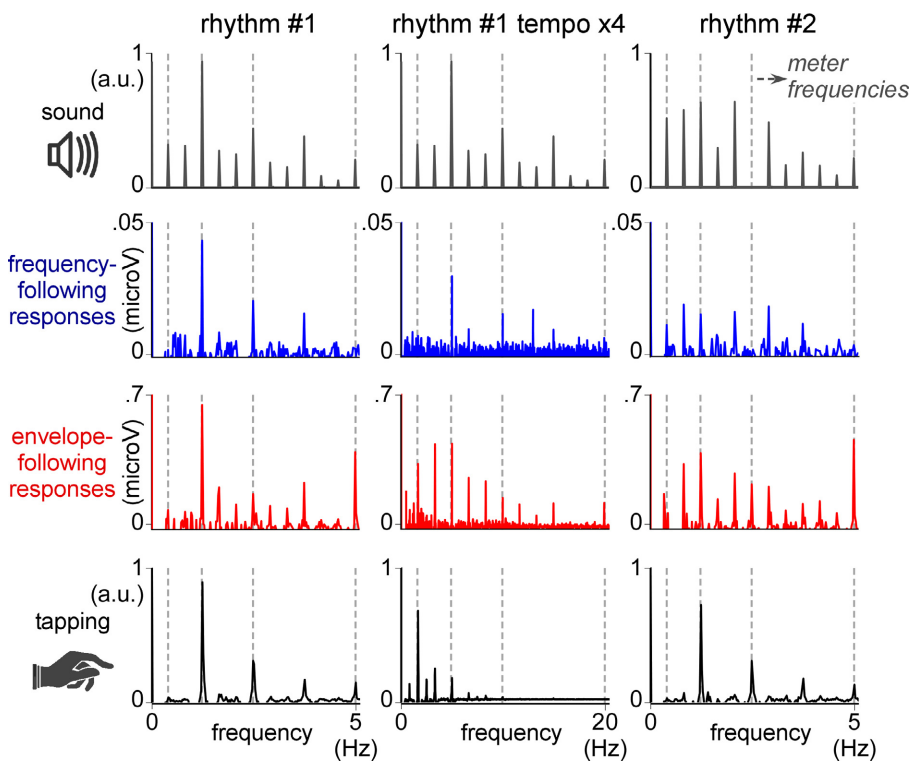

FIG. 2. Spectra of acoustic input, frequency-following and cortical envelope-following responses and tapping signals. Average across channels Fz, FCz and Cz and participants (meter-related frequencies corresponding to the dashed lines). The spectrum is highly similar across these different signals for rhythm #1 (unsyncopated) in which meter-related frequencies are already prominent in the input itself. In contrast, there is a graded increase at meter-related frequencies from the frequency-following responses to the tapping, for rhythm #2 (syncopated) with no prominent meter-related frequencies in the input. Moreover, the responses to the speeded-up rhythm exhibit a low-pass bias in both cortical and tapping signals but not in frequency-following responses, thus revealing a difference in sensitivity to tempo across these different signals. Note that for visualization only the frequency-following responses displayed here were obtained by extracting via Hilbert transform the amplitude of the EEG signal in the frequency range corresponding to the partials and sidebands, instead of direct measures of amplitudes at partials and sideband frequencies (used for statistical tests). [Colour figure can be viewed at wileyonlinelibrary.com].

TABLE 1. Mean and standard deviation, and *t*-tests against zero (after noise subtraction)

Mean ± stdv (μV), *t*-test against zero	Rhythm #1	Rhythm #1 tempo x4	Rhythm #2	White noise	Continuous tone
Envelope frequencies	0.186 ± 0.095	0.192 ± 0.088	0.205 ± 0.125	0.1999 ± 0.106	−0.012 ± 0.047
	t_{19} = 8.729	t_{19} = 9.767	t_{19} = 7.288	t_{19} = 8.400	t_{19} = 1.145
	P < 0.0001	P < 0.0001	P < 0.0001	P < 0.0001	P = 0.266
Partial frequencies	0.001 ± 0.004	0.0009 ± 0.002	0.001 ± 0.002	−0.0006 ± 0.001	0.002 ± 0.004
	t_{19} = 1.422	t_{19} = 1.464	t_{19} = 1.759	t_{19} = 1.876	t_{19} = 2.055
	P = 0.171	P = 0.159	P = 0.094	P = 0.076	P = 0.053
Distortion-product frequencies	0.044 ± 0.023	0.032 ± 0.015	0.038 ± 0.015	0.0003 ± 0.002	0.053 ± 0.020
	t_{19} = 8.578	t_{19} = 9.262	t_{19} = 11.04	t_{19} = 0.670	t_{19} = 11.49
	P < 0.0001	P < 0.0001	P < 0.0001	P = 0.510	P < 0.0001
Sideband frequencies	0.001 ± 0.001	0.0009 ± 0.0007	0.001 ± 0.0009	0.0002 ± 0.0009	0.0001 ± 0.0008
	t_{19} = 4.405	t_{19} = 5.635	t_{19} = 5.647	t_{19} = 1.032	t_{19} = 0.616
	P = 0.003	P < 0.0001	P < 0.0001	P = 0.315	P = 0.544

Significant cortical responses were elicited by all sequences at frequencies of the modulation spectrum, except for the control sequence with no amplitude modulation thus expected not to elicit cortical frequency-tagged responses. In contrast, no significant frequency-following activity was elicited in response to the three partials composing the non-harmonic chord. However, significant distortion-product responses and sidebands were elicited for all rhythmic sequences (except for the control sequence consisting of amplitude modulation of white noise instead of sine tones, which was thus expected not to elicit partials and distortion products). The non-significant results obtained in the control conditions confirm the assumption that, in the absence of significant responses to the input, the noise-subtracted signal tends towards zero.

frequencies are prominent in the input. In contrast, there was a gradual increase of meter-related frequencies from frequency-following responses to tapping, for the syncopated rhythm #2 with no prominent meter-related frequencies in the input. Moreover, this input–output transformation seemed to be sensitive to tempo, as the responses to the speeded-up rhythm exhibit a low-pass bias in both cortical responses and tapping signals but less so in the frequency-following responses.

For rhythm #1 (unsyncopated), there was no significant interaction in the 2 × 12 ANOVA comparing the 12 frequencies elicited by the rhythm across cortical and frequency-following responses ($F_{7.261,6.693}$ = 0.763, P = 0.614). This result thus indicates commensurate cortical and lower-level representations of the rhythm. As for the complementary 2 × 2 ANOVA comparing meter-related and unrelated frequencies, it did not reveal a significant interaction ($F_{0.013,1}$ = 0.673, $p\eta^2$ = 0.034, P = 0.422). This result indicates that activities elicited at meter-related frequencies are larger than those elicited at meter-unrelated frequencies in both frequency-following responses (*post hoc* paired *t*-test: t_{19} = 2.103, r^2 = 0.188, P = 0.049) and cortical responses (t_{19} = 4.459, r^2 = 0.511, P = 0.0003). Moreover, a 2 × 4 repeated-measures ANOVA comparing the cortical and frequency-following responses (factor Response) at the four different meter-related frequencies (factor Frequency) revealed a main effect of Frequency ($F_{2.438,46.314}$ = 10.999, $p\eta^2$ = 0.367, P < 0.0001) characterized by higher amplitude at 1.25 Hz than at the other meter-related frequencies (paired *t*-tests, all P < 0.006), but no main effect of Response and no interaction between the two factors (both P > 0.200). In addition, the input–output distance to the sound was not significantly different between cortical and frequency-following responses elicited at meter-related frequencies (t_{19} = 0.820, r^2 = 0.034, P = 0.422). In other words, both cortical and frequency-following responses presented prominent amplitudes at meter-related frequencies, as found in the input, and these neural representations did not significantly differ (Fig. 3).

This was no longer the case when the same rhythm was played four times faster. In the speeded-up rhythm #1, there was a significant interaction between the factors 'Frequencies' and 'Cortical vs. frequency-following responses' in the 2 × 12 ANOVA ($F_{42.680,5.619}$ = 7.255, P < 0.0001). In the complementary 2 × 2 ANOVA comparing meter-related and unrelated frequencies, the interaction also reached significance ($F_{0.780,1}$ = 4.602, $p\eta^2$ = 0.195, P = 0.045). This interaction was explained by a relatively small contrast between meter-related and unrelated frequencies in the cortical responses (t_{19} = 2.761, r^2 = 0.286, P = 0.012), as compared to the frequency-following responses (t_{19} = 6.217, r^2 = 0.670, P < 0.0001). Moreover, the ANOVA comparing the cortical and frequency-following responses at the four different meter-related frequencies revealed a significant interaction between the factors Response and Frequency ($F_{2.689,51.092}$ = 6.125, $p\eta^2$ = 0.244, P = 0.002). This interaction was attributable to higher amplitude in the cortical responses at the slowest meter-related frequency (1.66 Hz; P = 0.047), and higher amplitude in the frequency-following responses at the faster frequencies (10 and 20 Hz; P = 0.007 and 0.005, respectively), suggesting increased sensitivity to slower tempo for cortical responses compared to lower-level auditory responses. In addition, the input–output distance was significantly larger for cortical responses compared to frequency-following responses elicited at meter-related frequencies (t_{19} = 2.145, r^2 = 0.195, P = 0.045). That is, cortical responses seemed to be overall relatively more constrained by a low-pass function matching the musical tempo range, compared to the lower-level responses (see also section below, on responses at the frequency of individual intervals in the rhythms).

Finally, in rhythm #2 (syncopated), cortical and frequency-following responses convey significantly different representations of the rhythm, as shown by the significant interaction between the factors 'Cortical vs. frequency-following responses' and 'Frequency' in the 2 × 12 ANOVA ($F_{32.529,6.980}$ = 3.701, P = 0.001). The complementary 2 × 2 ANOVA, comparing meter-related and unrelated frequencies, also yielded a significant interaction ($F_{1.072,1}$ = 6.981, $p\eta^2$ = 0.269, P = 0.016). That is, the activity elicited at meter-related frequencies was larger than activity elicited at meter-unrelated frequencies in the cortical responses (t_{19} = 3.189, r^2 = 0.348, P = 0.004), but not in the frequency-following responses (t_{19} = 0.570, r^2 = 0.016, P = 0.575). Moreover, the ANOVA comparing the cortical and frequency-following responses at the four different meter-related frequencies revealed a significant interaction between the factors Response and Frequency ($F_{2.786,52.926}$ = 12.228, $p\eta^2$ = 0.392, P < 0.0001). This interaction was attributable to higher amplitude in the cortical responses at the slowest meter-related frequency (0.416 Hz; P = 0.002), and higher amplitude in the frequency-following responses at the faster frequencies (2.5 and 5 Hz;

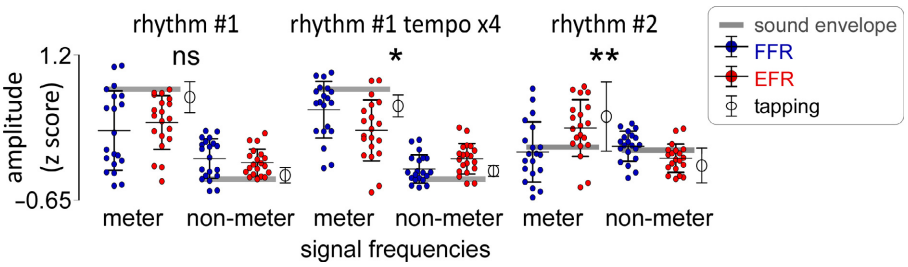

FIG. 3. Statistical comparison of the amplitudes at meter-related and unrelated frequencies. In the unsyncopated rhythm #1, frequency-following responses (FFR, mean and standard deviation, blue points for individual data) and cortical envelope-following responses (EFR, in red) both present prominent responses at meter-related frequencies, as in the input (grey horizontal line), and these responses do not significantly differ (ns, non-significant, for the interaction of the 2 × 2 ANOVA). This result is disrupted when the same rhythm is played four times faster, revealing an increased sensitivity to tempo in cortical responses (*$P < 0.05$). Moreover, for the syncopated rhythm #2 in which meter-related frequencies are not prominent in the input, frequency-following responses also show no prominence at meter-related frequencies, while cortical responses show a relative emphasis at meter-related frequencies (**$P < 0.01$). [Colour figure can be viewed at wileyonlinelibrary.com].

$P = 0.005$ and < 0.0001, respectively). In addition, the input–output distance was significantly greater for cortical responses compared to frequency-following responses elicited at meter-related frequencies ($t_{19} = 2.642$, $r^2 = 0.268$, $P = 0.016$). In other words, frequency-following responses show virtually no prominence at meter-related frequencies, as found in the input, while in contrast cortical responses show a relative emphasis at meter-related frequencies (Fig. 3).

Responses at frequency of individual intervals

As observed in the group-level averaged spectra (Fig. 2), the cortical response elicited at the frequency of the shortest interval in the rhythms (the frequency corresponding to the duration of individual intervals, i.e. corresponding to 12th frequency peak in the modulation spectrum of the inputs) seemed particularly prominent as compared to those of the frequency-following responses. This increased relative amplitude in cortical response compared to frequency-following response was observed in rhythm #1 ($t_{19} = 3.048$, $P = 0.006$) and rhythm #2 (paired *t*-test: $t_{19} = 4.664$, $P = 0.0002$). But this cortical increase was not sensitive to the onset structure of the rhythm (no significant difference between rhythm #1 and rhythm #2, $t_{19} = 1.337$, $P = 0.197$).

However, the cortical response at that frequency was modulated by the tempo. There was a significant reduction in cortical response in the speeded-up rhythm (i.e. when the frequency of individual intervals was of 20 Hz) compared to the same rhythm presented at moderate tempo (i.e. frequency of individual intervals of 5 Hz; $t_{19} = 4.830$, $P < 0.0001$), thus reflecting the low-pass bias shaping cortical responses in favour of activity within the musical tempo range. In the frequency-following responses, in contrast, there was a tendency towards increased amplitude at the speeded-up tempo (i.e. 20 Hz as compared to 5 Hz interval frequency), although it did not reach the threshold for significance ($t_{19} = 1.866$, $P = 0.075$). These results overall indicate that frequency-following responses were relatively less affected by the tempo than cortical responses.

Finally, we found that cortical responses at the frequency of individual intervals were also modulated by other properties of the stimulus such as the spectral content of the tones carrying the amplitude modulation. There was a significant reduction in the cortical responses at that frequency when rhythm #1 was conveyed by white noise ('white noise sequence') compared to the three partials ($t_{19} = 3.313$, $P = 0.003$).

To summarize, the neural activity at the frequency of the individual intervals is increased between frequency-following and cortical responses. This cortical increase is sensitive to properties of the stimulus such as the tempo and the spectral content of the tones carrying the amplitude modulation, but is not sensitive to the onset structure of the rhythm. Hence, neural activity measured at that frequency contributes to the overall cortical increase observed at meter-related frequencies, but it does not fully account for the differences observed between cortical and frequency-following responses to rhythm.

Tapping analysis

Because of technical issues, tapping data were missing for four participants, leaving 16 participants for the tapping analysis. This analysis indicated that the mean intertap interval was 0.797 s for rhythm #1 (median value across participants), 0.799 s for rhythm #1 carried by white noise, 0.798 s for rhythm #2 and 0.600 s for the speeded-up rhythm #1. These behavioural results thus confirm theoretical assumptions and previous behavioural evidence on the preferred frequency for motor entrainment to these rhythms based on grouping by four intervals (three taps per rhythm cycle, i.e. three-beat meter) for rhythms #1 and #2 and by 12 intervals (i.e. one tap per rhythm cycle) for the speeded-up rhythm (Povel & Essens, 1985; Nozaradan *et al.*, 2012b, 2016a). These intertap intervals were also reflected in the tapping sensor spectra (Fig. 2, lower graphs), exhibiting greatest peaks at the corresponding frequencies (1.25 Hz for rhythms #1 and #2 and 1.66 Hz for the speeded-up rhythm). The coefficient of variation of intertap intervals was 5.6% for rhythm #1 (median across participants), 5.1% for rhythm #1 carried by white noise, 6.5% for rhythm #2 and 4.8% for the speeded-up rhythm, with no significant difference between the conditions (one-way ANOVA, $F = 1.581$, $P = 0.207$).

These results suggest that the tapping frequency was relatively consistent across participants for the different rhythms. But this was not the case for the phase of tapping relative to interval onsets. As the tapping frequencies in the rhythms played at moderate tempo converged towards three taps per rhythm cycle, the percentage of taps was compared across four possible beat locations (i.e. percentage of taps closest to onsets of intervals #1-5-9 compared to #2-6-10, #3-7-11 or #4-8-12). For rhythm #1, the highest percentage of taps averaged across participants was located on intervals #4-8-12 (33.58%). Surprisingly, when the same rhythm #1 was presented at the same tempo but carried by white noise instead of three tones, the preferred location shifted to intervals #3-7-11 (35.12% of taps). For rhythm #2, the highest percentage of taps was found on intervals #4-8-12, with 27.59% of taps. Finally, for the speeded rhythm, the tapping frequencies converged towards one tap per rhythm cycle, so that the percentage of taps was compared across twelve

possible beat positions, and the highest percentage was found on interval #3 (16.61%). However, in all conditions, there were no significant differences across the possible beat locations (one-way ANOVAs, all P values ≥ 0.231), suggesting that the beat position was generally variable across participants. This overall variability across participants could also be explained in part by shifts in beat position observed in approximately half of the participants over the different rhythms.

Correlation between neural and behavioural measures of entrainment to rhythm

We tested whether the ability to tap to the beat was explained by the relative amplitude at meter-related frequencies obtained in the frequency-following responses vs. the cortical envelope-following responses. This was tested with a multiple linear regression analysis with the coefficient of variation in tapping (averaged across rhythm conditions) as dependent variable and the relative amplitude at meter-related frequencies obtained for cortical and frequency-following responses (averaged across rhythm conditions) as independent variables (note that the independent variables showed very little correlation with each other; Pearson's coefficient $r = 0.189$, $P = 0.484$). The obtained model was not significant ($R^2 = 0.280$, $F_{2,15} = 2.522$, $P = 0.119$). This result may be due to a lack of or an indirect relationship between the neural and tapping data as measured in our study or may be due to the fact that there was insufficient variance in the data (participants were not specifically selected to maximize interindividual differences in tapping performance). However, the partial correlation between tapping and cortical measures reached the threshold for significance ($R = -0.536$, $P = 0.044$), in contrast to the partial correlation between tapping and frequency-following measures ($R = 0.156$, $P = 0.527$). Although these correlation tests do not yield robust statistical evidence, they suggest a particular relationship between tapping and cortical responses, as compared to frequency-following responses. Specifically, participants with lowest variability in their motor entrainment to rhythm tend to be those with larger cortical responses at meter-related frequencies.

Time-domain analysis

As shown in Fig. 4, the time course of frequency-following responses is characterized by a relatively faithful tracking of the envelope of the 'acoustic' pulses. In contrast, the cortical responses display a more complex form of tracking, with slow fluctuations over multiple intervals including 'silent' intervals modulating the responses to individual intervals. Moreover, the time course of responses to the speeded-up rhythm confirms the results of the frequency-domain analysis, by showing that cortical responses are no longer able to track the sound envelope at faster tempo and thus tend to respond to the entire repetition of the rhythm rather than individual intervals. For this reason, further time-domain analyses of the amplitude in response to individual intervals were only performed for rhythms played at moderate tempo (i.e. rhythms #1 and #2), but not for the speeded-up rhythm.

Responses to 'acoustic' vs. 'silent' intervals

The amplitude of frequency-following responses was significantly reduced during 'silent' intervals compared to 'acoustic' intervals, for both rhythms #1 and #2 (significant effect of factor 'acoustic vs. silent intervals': $F_{1,19} = 20.580$, $P < 0.0001$; no effect of factor 'Rhythm' and no interaction between factors, all $P > 0.7$). In contrast, cortical responses showed no significant differences between 'acoustic' and 'silent' intervals ($P = 0.351$; no effect of factor 'Rhythm' and no interaction between factors, all $P > 0.066$).

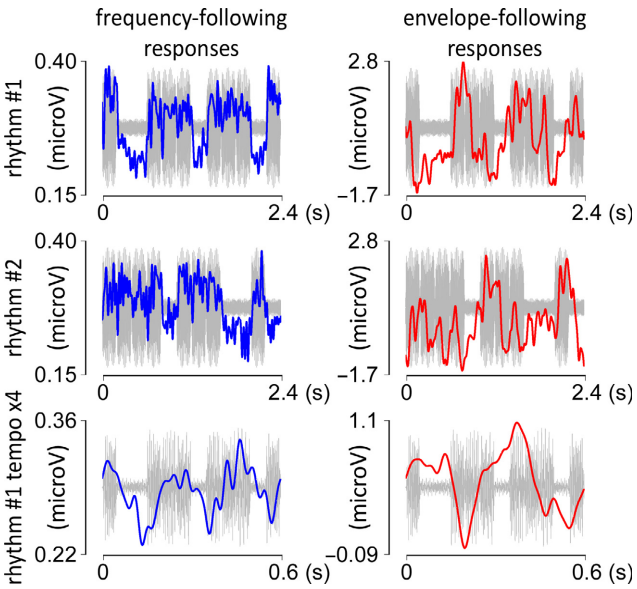

FIG. 4. Time course of frequency-following responses and cortical envelope-following responses (group-level average). The chunks correspond to the time window between each onset of the repeated rhythmic patterns (in grey). The cortical envelope-following responses were obtained by filtering the EEG between 0.1 and 30 Hz, and the frequency-following responses were obtained by extracting the amplitude between 134 and 354 Hz. Frequency-following responses overall exhibit a relatively faithful tracking of the envelope of the 'acoustic' intervals. In contrast, cortical responses display a more complex form of tracking, with processes of grouping and slow fluctuations over multiple intervals. In the speeded-up rhythm, cortical responses show larger grouping across intervals. [Colour figure can be viewed at wileyonlinelibrary.com].

Modulation by the number of immediately preceding 'acoustic' intervals

The cortical 'acoustic'- and 'silent'-evoked responses were significantly modulated by the number of immediately preceding 'acoustic' intervals (Fig. 5; 'silent'-evoked cortical responses: Kruskal–Wallis ANOVA $H_5 = 12.430$, $P = 0.010$; 'acoustic'-evoked cortical responses: $H_5 = 30.280$, $P < 0.0001$). For cortical 'silent'-evoked responses, the significant difference across instances preceded by different number of 'acoustic' intervals was characterized by a significant positive linear trend (slope: 0.11, $P = 0.045$, linear fit: $R^2 = 0.819$). That is, the greater the number of immediately preceding 'acoustic' intervals, the larger the amplitude of cortical 'silent'-evoked responses. In contrast, for cortical 'acoustic'-evoked responses, there was no significant linear trend ($P = 0.463$), indicating that 'acoustic'-evoked responses did not linearly follow a neural adaptation model. Instead, *post hoc* Mann–Whitney U-tests revealed significant reduction in amplitude only for the instance preceded by one 'acoustic' interval (all $P < 0.038$).

In contrast with cortical responses, frequency-following responses were not modulated by the number of immediately preceding 'acoustic' intervals (frequency-following responses to 'silent' intervals: $H_5 = 0.350$, $P = 0.986$; frequency-following responses to 'acoustic' intervals: $H_5 = 0.347$, $P = 0.986$). These results suggest that these auditory responses were relatively stable over time or that these measures were not sensitive enough to capture an effect.

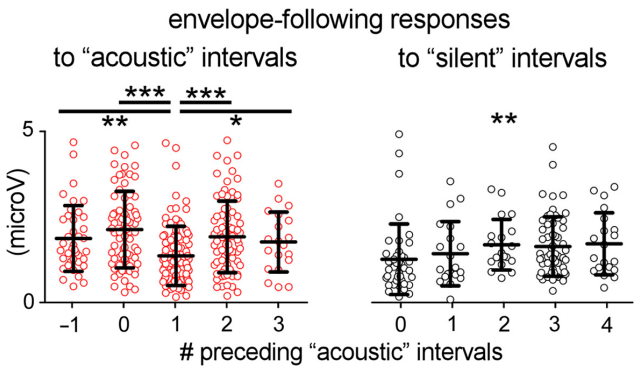

FIG. 5. Cortical responses to individual 'acoustic' and 'silent' intervals. Mean and standard deviation; the individual points represent the cortical responses of each participant at each individual instances where an 'acoustic' or 'silent' interval was immediately preceded by a given number of 'acoustic' intervals, merged across rhythms #1 and #2. The 'acoustic'-evoked responses to the second of a cluster of 'acoustic' intervals were decreased compared to the responses to the first, penultimate or last 'acoustic' interval of a cluster. In contrast, 'silent'-evoked responses increased with the number of immediately preceding 'acoustic' intervals (*$P < 0.05$, **$P < 0.01$, ***$P < 0.001$). [Colour figure can be viewed at wileyonlinelibrary.com].

On-beat vs. off-beat responses

The on-beat vs. off-beat difference was calculated for a three-beat meter (i.e. in accordance with the mean intertap intervals obtained in the tapping data) across the four different possible beat locations. When calculated with the sound input of rhythm #1 (by taking 'acoustic' intervals = 1 and 'silent' intervals = 0), this yielded a quadratic function with values [0.44, 0, 0, 0.44] across four possible beat positions. A similar function across beat positions was thus expected if the neural responses were mainly driven by the acoustic input. This was tested using polynomial contrast analysis. A similar quadratic contrast was found for frequency-following responses ($F = 10.278$, $P = 0.005$) but not for cortical responses to rhythm #1 ($P > 0.642$), which instead followed a cubic function favouring the second and fourth beat positions ($F = 10.508$, $P = 0.004$). For rhythm #2, the on-beat vs. off-beat difference calculated with the sound input yielded a function close to a linear or cubic function in favour of the third beat position ([−0.44, 0, 0.44, 0] across the four possible beat positions). A linear contrast was found at threshold for frequency-following responses ($F = 4.277$, $P = 0.053$) and was also found for cortical responses ($F = 5.085$, $P = 0.036$), though cortical responses were best explained by a cubic function favouring the second and fourth beat positions ($F = 12.187$, $P = 0.002$). Together, these results suggest that frequency-following and cortical responses did not match each other or the stimulus with respect to on-beat vs. off-beat amplitude differences. Finally, the same analysis conducted on the tapping data yielded no significant effect for the polynomial contrast (rhythms #1 and 2: all $P > 0.424$), confirming the overall variability in target beat positions across participants in the tapping data.

Discussion

How the human brain performs the transformation from rhythmic acoustic inputs into perceptual metric units is a key issue in understanding the neural mechanisms enabling rhythmic entrainment. To address this question, we brought together physical measures of rhythmic acoustic inputs, physiological measures of neural outputs at different stages of auditory processing and behavioural measures of entrainment to the beat.

Cortical responses showed increased amplitudes at meter-related frequencies, regardless of the prominence of meter-related frequencies in the modulation spectrum of the rhythmic inputs. Cortical activity was also found to correlate with behavioural measures of motor entrainment to the rhythms and to be disrupted at fast tempo. In contrast, frequency-following responses were less affected by the tempo and did not show increased amplitudes at meter-related frequencies for the more challenging syncopated rhythm with no prominent meter-related frequencies in the input. These findings indicate critical transformations between cortical and lower-level auditory representations of rhythm in humans, possibly linked to the emergence of a perceptual representation of meter across different stages of auditory processing.

Slow and fast time scale responses to rhythm

Frequency-following responses were captured in response to three non-harmonic partial tones above 150 Hz. This neural activity was elicited at the frequencies corresponding to the distortion products across the three tones and at sideband frequencies corresponding to the rhythmic amplitude modulation of these tones. Fast time scale responses > 150 Hz have long been assumed to originate mainly from the inferior colliculus, located in the human brainstem (see, e.g. Skoe & Kraus, 2010). However, recent work has cast doubt on this assumption, providing evidence that these responses could instead constitute a mixture of activity not only from auditory brainstem nuclei but also from cortical sources, in particular from the earliest auditory cortical stages (Coffey *et al.*, 2016, 2017). Moreover, brainstem nuclei are known to transmit auditory information to cortical areas through ascending fibres, but these subcortical nuclei also receive corticofugal inputs modulating their activity in turn (see, e.g. Winer, 2006; Winer & Lee, 2007; Bajo & King, 2013). Hence, we cannot reject the hypothesis that in the current study cortical processes contributed to the frequency-following responses recorded here with surface EEG, either through corticofugal transmission or as an actual portion of neural activity elicited at fast time scale. Irrespective of the answer to this question, our findings indicate that the representation of rhythm conveyed by these fast time scale responses is not redundant with those conveyed by cortical responses captured at much slower time scale (< 20 Hz). Rather, slower time scale responses show a relatively greater dissociation from the acoustic input, with stronger emphasis of meter-related frequencies regardless of the prominence of these frequencies in the input, compared to frequency-following responses, which seem to be relatively more directly driven by the stimulus.

This finding fits well with the view that slow time scale responses reflect activity from cortical sources, especially from primary and secondary auditory areas, which are themselves functionally connected to higher-level associative and/or motor areas. This cortical activity, due to its overall relatively slow lingering dynamics, may enable temporal integration and higher-level encoding of the temporal properties of sounds, compared to frequency-following responses. Cortical activity would thus be able to convey greater level of abstraction from the acoustic stimulus compared to frequency-following responses, which reflect activity from the earliest auditory cortical stages and subcortical auditory nuclei. This view is also consistent with recent data recorded directly in the human auditory cortex using intracranial depth electrodes in Heschl's gyrus and the planum temporale of epileptic patients (Nozaradan *et al.*, 2016c). In this work, a selective enhancement of the neural activity at meter-related frequencies was observed in the primary auditory cortex, but was even more pronounced in the secondary auditory cortex

(planum temporale), especially for a syncopated rhythm with no prominent meter-related frequencies in the input.

From selective neural enhancement to rhythmic motor entrainment

Our results show a graded selective enhancement of meter-related frequencies across different stages of auditory processing. The exact mechanisms leading to this input–output transformation of the rhythms remain unclear. For example, this process could find its roots in low-level auditory properties that are shared across most mammalian species regardless of their abilities to synchronize movements to rhythmic inputs (Merchant & Honing, 2014). Along this line, midbrain adaptation has been proposed as a low-level process which, by encoding the temporal properties of sounds, would create points of neural emphasis that may influence the perceptual emergence of meter (Rajendran *et al.*, 2017). Such a process of neural emphasis could also provide a neurobiological foundation to previous models that explained the perceptual emergence of meter based on a set of rules by which points of perceptual emphasis or perceptual accents, are generated according to the onset structure and tempo of the rhythmic input (Povel & Essens, 1985).

Relatedly, the time-domain analyses performed in the current study revealed modulations of neural responses by the number of immediately preceding sounds. Specifically, the response to a sound was significantly reduced when directly preceded by another sound, compared to the responses to sounds located at the first, penultimate or last position in a cluster of sounds. However, these modulations were not found in the frequency-following responses. Most importantly, these modulations are not readily explained by a process of adaptation, as found in the midbrain of non-human models. This discrepancy between human and non-human animal data could be due to differences in the measurements used in these studies (firing rate and local field potential measures in response to acoustic events in two anesthetized gerbils vs. surface EEG recording during active listening in humans). But it could also be related to the differences across animal species in the general ability to entrain to rhythmic inputs (see, e.g. the gradual audiomotor evolution hypothesis; Merchant & Honing, 2014). On the one hand, some species such as gerbils would present rudiments of meter processing in the form of neural emphasis at instances coinciding with the meter while listening to rhythmic patterns, and this neural emphasis would be produced by low-level properties of subcortical auditory neurons widely shared across animals (Rajendran *et al.*, 2017). On the other hand, other species such as humans would be equipped with functional connections between these low-level sensory processes and higher-level associative and/or motor areas, thus allowing rhythmic entrainment to emerge with substantial flexibility and invariance with respect to the rhythmic input.

This dissociation with respect to the input is reflected clearly in our behavioural measures of motor entrainment to the rhythms, showing consistent frequency but variable phase across participants. Indeed, while the intertap intervals generally converged towards the theoretically predicted beat period, there was an overall absence of a consistent set of beat locations across participants, as observed by comparing the percentage of taps performed on-beat vs. off-beat. Also, the high variability in phase of tapping across participants was partly produced by shifts in beat positions observed over the course of trials in some participants. This high variability in phase but not in period across participants suggests that the selection of the phase of the perceived metrical structure, that is the selection of the anchor points relative to the rhythmic input, may rely on mechanisms that are distinct from the selection of the frequencies of the perceived metrical structure (Repp & Saltzman, 2002).

In the current study, time-domain measures of on-beat vs. off-beat amplitude differences were used to test whether the selective neural enhancement of meter-related frequencies observed in the frequency domain was reflected in a consistent increase at specific equally spaced intervals across participants. In other words, this method allowed the consistency of beat positions across participants to be examined, in contrast with the frequency-domain analysis, which is blind to this information. Interestingly, the time-domain analysis yielded an overall absence of consensus beat positions across participants and across the cortical and frequency-following responses. This result could be due to high interindividual variability in the anchor points of the perceived metrical structure, as found in the tapping data. But this result could also be attributable to the fact that the time-domain approach required averaging the signals across a large number of cycles and binning the averaged signals into 200-ms temporal windows corresponding to individual intervals, which may not fully capture the neural dynamics elicited by these rhythmic inputs. These considerations thus highlight the relevance of frequency-domain approaches for characterizing processes of selective neural enhancement at specific frequencies, to complement time-domain approaches.

Taken together, the results of our study suggest that the selection of a subset of frequencies serving as perceived metrical structure could arise at the earliest auditory cortical stage, perhaps even at subcortical stage when these frequencies are relatively prominent in the acoustic input. However, this process of temporal selection is critically shaped at the cortical level, especially for complex rhythms in which the meter does not systematically match the onset structures. This critical transformation may possibly arise through functional connections between the auditory cortex and other, movement-related, brain structures. Future research capturing this input–output transformation of rhythm across different cortical regions (e.g. via multiple depth-electrode recordings) may provide a more detailed picture of these processes as embodied throughout the nervous system, and how they differ across animal species.

Acknowledgements

This work was supported by the Australian Research Council (DE160101064 to S.N. and FT140101162 to P.E.K.), by the Fonds the recherche du Quebec-Sante (FRQS). We wish to thank Mihaela Felezeu, Laura Caron-Derochers and Gabrielle Laufer for their help with EEG recording.

Conflict of interest

The authors declare no competing financial interest.

Author contributions

SN, AL and MS designed the study. SN, TL analysed the data. SN, MS, PK, TL and AL wrote the manuscript.

Data accessibility

The data are available upon request.

References

Bajo, V.M. & King, A.J. (2013) Cortical modulation of auditory processing in the midbrain. *Front. Neural. Circuit.*, **6**, 114.

Bidelman, G.M. & Krishnan, A. (2011) Brainstem correlates of behavioral and compositional preferences of musical harmony. *Neuroreport*, **22**, 212–216.

Burger, B., London, J., Thompson, M.R. & Toiviainen, P. (2017) Synchronization to metrical levels in music depends on low-frequency spectral components and tempo. *Psychol. Res*. https://doi.org/10.1007/s00426-017-0894-2 [Epub ahead of print.]

Celma-Miralles, A., de Menezes, R.F. & Toro, J.M. (2016) Look at the beat, feel the meter: top-down effects of meter induction on auditory and visual modalities. *Front. Hum. Neurosci.*, **10**, 108.

Chandrasekaran, B. & Kraus, N. (2010) The scalp-recorded brainstem response to speech: neural origins and plasticity. *Psychophysiology*, **47**, 236–246.

Chemin, B., Mouraux, A. & Nozaradan, S. (2014) Body movement selectively shapes the neural representation of musical rhythm. *Psychol. Sci.*, **25**, 2147–2159.

Cirelli, L.K., Spinelli, C., Nozaradan, S. & Trainor, L.J. (2016) Measuring neural entrainment to beat and meter in infants: effects of music background. *Front. Neurosci.*, **10**, 229.

Coffey, E.B., Herholz, S.C., Chepesiuk, A.M., Baillet, S. & Zatorre, R.J. (2016) Cortical contributions to the auditory frequency-following response revealed by MEG. *Nat. Commun.*, **7**, 11070.

Coffey, E.B.J., Musacchia, G. & Zatorre, R.J. (2017) Cortical correlates of the auditory frequency-following and onset responses: EEG and fMRI evidence. *J. Neurosci.*, **37**, 830–838.

Draganova, R., Ross, B., Borgmann, C. & Pantev, C. (2002) Auditory cortical response patterns to multiple rhythms of AM sound. *Ear Hearing*, **23**, 254–265.

Grahn, J.A. (2012) Neural mechanisms of rhythm perception: current findings and future perspectives. *Top. Cogn. Sci.*, **4**, 585–606.

Hove, M.J., Marie, C., Bruce, I.C. & Trainor, L.J. (2014) Superior time perception for lower musical pitch explains why bass-ranged instruments lay down musical rhythms. *Proc. Natl. Acad. Sci. USA*, **111**, 10383–10388.

Lehmann, A., Arias, D.J. & Schönwiesner, M. (2016) Tracing the neural basis of auditory entrainment. *Neuroscience*, **337**, 306–314.

London, J. (2004) *Hearing in Time: Psychological Aspects of Musical Meter*. Oxford University Press, London, UK.

McAuley, J.D. (2010) Tempo and rhythm. In Jones, M.R. *et al.* (Eds), *Music Perception, Springer Handbook of Auditory Research*, 36. Springer, New York, USA, pp. 165–199.

Merchant, H. & Honing, H. (2014) Are non-human primates capable of rhythmic entrainment? Evidence for the gradual audiomotor evolution hypothesis. *Front. Neurosci.*, **7**, 274.

Mouraux, A., Iannetti, G.D., Colon, E., Nozaradan, S., Legrain, V. & Plaghki, L. (2011) Nociceptive steady-state evoked potentials elicited by rapid periodic thermal stimulation of cutaneous nociceptors. *J. Neurosci.*, **31**, 6079–6087.

van Noorden, L. & Moelants, D. (1999) Resonance in the perception of musical pulse. *J. New Music Res.*, **28**, 43–66.

Nozaradan, S. (2014) Exploring how musical rhythm entrains brain activity with electroencephalogram frequency-tagging. *Philos. T. Roy. Soc. B*, **369**, 20130393.

Nozaradan, S., Peretz, I., Missal, M. & Mouraux, A. (2011) Tagging the neuronal entrainment to beat and meter. *J. Neurosci.*, **31**, 10234–10240.

Nozaradan, S., Peretz, I. & Mouraux, A. (2012b) Selective neuronal entrainment to the beat and meter embedded in a musical rhythm. *J. Neurosci.*, **32**, 17572–17581.

Nozaradan, S., Zerouali, Y., Peretz, I. & Mouraux, A. (2015) Capturing with EEG the neural entrainment and coupling underlying sensorimotor synchronization to the beat. *Cereb. Cortex*, **25**, 736–747.

Nozaradan, S., Mouraux, A., Jonas, J., Colnat-Coulbois, S., Rossion, B. & Maillard, L. (2016a) Intracerebral evidence of rhythm transform in the human auditory cortex. *Brain Struct. Funct.*, **222**, 2389–2404.

Nozaradan, S., Peretz, I. & Keller, P.E. (2016b) Individual differences in rhythmic cortical entrainment correlate with predictive behavior in sensorimotor synchronization. *Sci. Rep.*, **6**, 20612.

Nozaradan, S., Schönwiesner, M., Caron-Desrochers, L. & Lehmann, A. (2016c) Enhanced brainstem and cortical encoding of sound during synchronized movement. *Neuroimage*, **142**, 231–240.

Nozaradan, S., Mouraux, A. & Cousineau, M. (2017a) Frequency-tagging to track the neural processing of contrast in fast continuous sound sequences. *J. Neurophysiol.*, **118**, 243–253.

Nozaradan, S., Schwartze, M., Obermeier, C. & Kotz, S.A. (2017b) Specific contributions of basal ganglia and cerebellum to the neural tracking of rhythm. *Cortex*, **95**, 156–168.

Nozaradan, S., Keller, P.E., Rossion, B. & Mouraux, A. (2017c) EEG frequency-tagging and input-output comparison in rhythm perception. *Brain Topogr*. https://doi.org/10.1007/s10548-017-0605-8 [Epub ahead of print.]

Patel, A.D. & Iversen, J.R. (2014) The evolutionary neuroscience of musical beat perception: the Action Simulation for Auditory Prediction (ASAP) hypothesis. *Front. Psychol.*, **8**, 56.

Povel, D.J. & Essens, P.J. (1985) Perception of temporal patterns. *Music Percept.*, **2**, 411–441.

Rajendran, V.G., Harper, N.S., Garcia-Lazaro, J.A., Lesica, N.A. & Schnupp, J.W.H. (2017) Midbrain adaptation may set the stage for the perception of musical beat. *Proc. Biol. Sci.*, **284**, 20171455.

Repp, B.H. & Saltzman, E.L. (2002) Influences of grouping and metrical structure on the kinematics of rhythmic finger tapping. Unpublished manuscript. Available http://www.haskins.yale.edu/haskins/STAFF/repp.html.

Repp, B.H. & Su, Y.H. (2013) Sensorimotor synchronization: a review of recent research (2006–2012). *Psychon. B. Rev.*, **20**, 403–452.

Retter, T.L. & Rossion, B. (2016) Uncovering the neural magnitude and spatio-temporal dynamics of natural image categorization in a fast visual stream. *Neuropsychologia*, **91**, 9–28.

Ross, J.M., Iversen, J.R. & Balasubramaniam, R. (2016) Motor simulation theories of musical beat perception. *Neurocase*, **22**, 558–565.

Skoe, E. & Kraus, N. (2010) Auditory brain stem response to complex sounds: a tutorial. *Ear Hearing*, **31**, 302–324.

Stupacher, J., Wood, G. & Witte, M. (2017) Neural entrainment to polyrhythms: a comparison of musicians and non-musicians. *Front. Neurosci.*, **11**, 208.

Tal, I., Large, E.W., Rabinovitch, E., Wei, Y., Schroeder, C.E., Poeppel, D. & Zion Golumbic, E. (2017) Neural entrainment to the beat: the "Missing-Pulse" phenomenon. *J. Neurosci.*, **37**, 6331–6341.

Tierney, A. & Kraus, N. (2013) The ability to move to a beat is linked to the consistency of neural responses to sound. *J. Neurosci.*, **33**, 1481–1488.

Tierney, A. & Kraus, N. (2016) Getting back on the beat: links between auditory-motor integration and precise auditory processing at fast time scales. *Eur. J. Neurosci.*, **43**, 782–791.

Winer, J.A. (2006) Decoding the auditory corticofugal systems. *Hearing Res.*, **212**, 1–8.

Winer, J.A. & Lee, C.C. (2007) The distributed auditory cortex. *Hearing Res.*, **229**, 3–13.

Wojtczak, M., Mehta, A.H. & Oxenham, A.J. (2017) Rhythm judgments reveal a frequency asymmetry in the perception and neural coding of sound synchrony. *Proc. Natl. Acad. Sci. USA*, **114**, 1201–1206.