Extracting human cortical responses to sound onsets and acoustic feature changes in real music, and their relation to event rate

Niels T. Haumann [a,*], Massimo Lumaca [a], Marina Kliuchko [a,b], Jose L. Santacruz [c], Peter Vuust [a], Elvira Brattico [a,d,*]

[a] Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music, Aarhus/Aalborg, Denmark
[b] BioMag Laboratory, HUS Medical Imaging Center, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
[c] Communications Engineering, ETSIT, University of Málaga, Málaga, Spain
[d] Department of Education, Psychology, Communication, University of Bari Aldo Moro, Italy

A B S T R A C T

Evoked cortical responses (ERs) have mainly been studied in controlled experiments using simplified stimuli. Though, an outstanding question is how the human cortex responds to the complex stimuli encountered in realistic situations. Few electroencephalography (EEG) studies have used Music Information Retrieval (MIR) tools to extract cortical P1/N1/P2 to acoustical changes in real music. However, less than ten events per music piece could be detected leading to ERs due to limitations in automatic detection of sound onsets. Also, the factors influencing a successful extraction of the ERs have not been identified. Finally, previous studies did not localize the sources of the cortical generators. This study is based on an EEG/MEG dataset from 48 healthy normal hearing participants listening to three real music pieces. Acoustic features were computed from the audio signal of the music with the MIR Toolbox. To overcome limits in automatic methods, sound onsets were also manually detected. The chance of obtaining detectable ERs based on ten randomly picked onset points was less than 1:10,000. For the first time, we show that naturalistic P1/N1/P2 ERs can be reliably measured across 100 manually identified sound onsets, substantially improving the signal-to-noise level compared to <10 trials. More ERs were measurable in musical sections with slow event rates (0.2 Hz-2.5 Hz) than with fast event rates (>2.5 Hz). Furthermore, during monophonic sections of the music only P1/P2 were measurable, and during polyphonic sections only N1. Finally, MEG source analysis revealed that naturalistic P2 is located in core areas of the auditory cortex.

1. Introduction

The simplification of real-world phenomena into experimental paradigms is the standard approach in quantitative research. However, this entails a detachment from the more complex reality which raises questions regarding its ecological validity. For this reason, there is now an increasing interest in investigating the human brain under more natural settings as an addition to more controlled experiments (Hasson and Honey, 2012). An early example from vision research on naturalistic designs was the application of inter-subject correlation in functional magnetic resonance neuroimaging (fMRI) while participants watched a popular movie (Hasson et al., 2004). More recently, our team combined computational analysis of the auditory signal, using the Music

Information Retrieval (MIR) Toolbox (Lartillot and Toiviainen, 2007), with the fMRI signal (Alluri et al., 2012; Burunat et al., 2016) to study the cortical processing of single musical features during realistic listening. Other investigations have adopted electroencephalography (EEG) block designs (Bo et al., 2016; Mikutta et al., 2014). These studies showed an association between electrophysiological brain oscillations at narrow frequency bands and the time course of acoustic features. The main drawback of fMRI and EEG block designs is, however, their low temporal resolution, whereby the neurophysiological responses to the temporal details of real music (Fig. 1) are missing or distorted.

Only a few recent studies have adopted the high temporal precision of EEG to examine the relationship between transient evoked cortical responses (ERs) and rapid acoustical changes in real music (Poikonen

et al., 2016a, 2016b; Sturm et al., 2015) and language (Crosse et al., 2016). A popular assumption from studies on the peripheral auditory system is that magnitudes in an acoustical feature, e.g., their sound intensity, are linearly related to ER amplitudes (Holdgraf et al., 2017; Wu et al., 2006). However, auditory detection thresholds (Mäkinen et al., 2004), feedback loops, and inhibitory processes, such as neural adaptation, challenge this assumption by modulating the stimulus-response relationship in the cortex (Holdgraf et al., 2017; Wu et al., 2006). Moreover, in a recent study (Haumann et al., 2018), we were unable to replicate previous findings of ERs to strong acoustical changes. One possible reason is that relatively few ERs have so far been successfully extracted, with less than ten per tested acoustical feature and music piece (see e.g., Poikonen et al., 2016a, 2016b). Despite these challenges, we argue that it is relevant to identify ERs based on acoustical feature changes. We assume that there is a low chance of measuring ERs at random time points not constrained to acoustical feature changes. Also, we suggest that it might be relevant to investigate which additional factors are important for the successful extraction of cortical responses evoked by real music.

Historically, researchers have used sound onsets in controlled stimulus paradigms to extract auditory ERs. However, the detection of sound onsets in real music poses a methodological challenge. The classical approach of MIR algorithms to detect sound onsets in naturalistic music stimuli consists of finding an increase, and setting a threshold for the increase, in two main acoustical features: sound intensity (or RMS energy) and spectral flux (Alías et al., 2016). Spectral flux reflects a combination of a change in sound intensity, pitch, and timbre (e.g., a

different instrument sound). Sound intensity specifically reflects a change in the energy of a sound. The setting of a proper threshold remains, however, a major limitation of these automatic methods. Sound onsets have different spectral spreads and rates of energy increase (Smith and Fraser, 2004). Furthermore, those onsets are particularly difficult to detect from the spectrogram of polyphonic music, which contains a mixture of energies from simultaneous sources of music instruments (Thoshkahna and Ramakrishnan, 2008).

Another relevant factor for extracting cortical ERs is the average number of sound onsets per second, defined as 'event rate' (or 'event density' in the MIR Toolbox). The amplitude of the ER depends negatively on the rate of the stimulus presentation (Naatanen and Picton, 1987). The cortical P1/N1/P2 ERs show higher amplitude at slower event rates between 1 and 5 Hz (Sussman et al., 2008; Wang et al., 2008), and, dependent on the stimulus intensity and pitch, higher N1/P2 amplitudes have been observed for slower event rates in a range between 0.1 Hz and 1 Hz (Naatanen and Picton, 1987). Sound events in naturalistic music stimuli often have event rates higher than 1 Hz; thus, they presumably evoke cortical responses of lower amplitudes. This differs from typical controlled paradigms, where the event rate is slower and optimized to increase the amplitudes of the ERs. To our knowledge, it remains to be empirically validated whether more P1/N1/P2 responses are measurable at slow event rates —compared to fast event rates— in real music stimuli. Therefore, we also investigated whether it might be important to consider the effects of event rate when selecting naturalistic music excerpts.

In summary, here we introduced a novel analysis approach of the

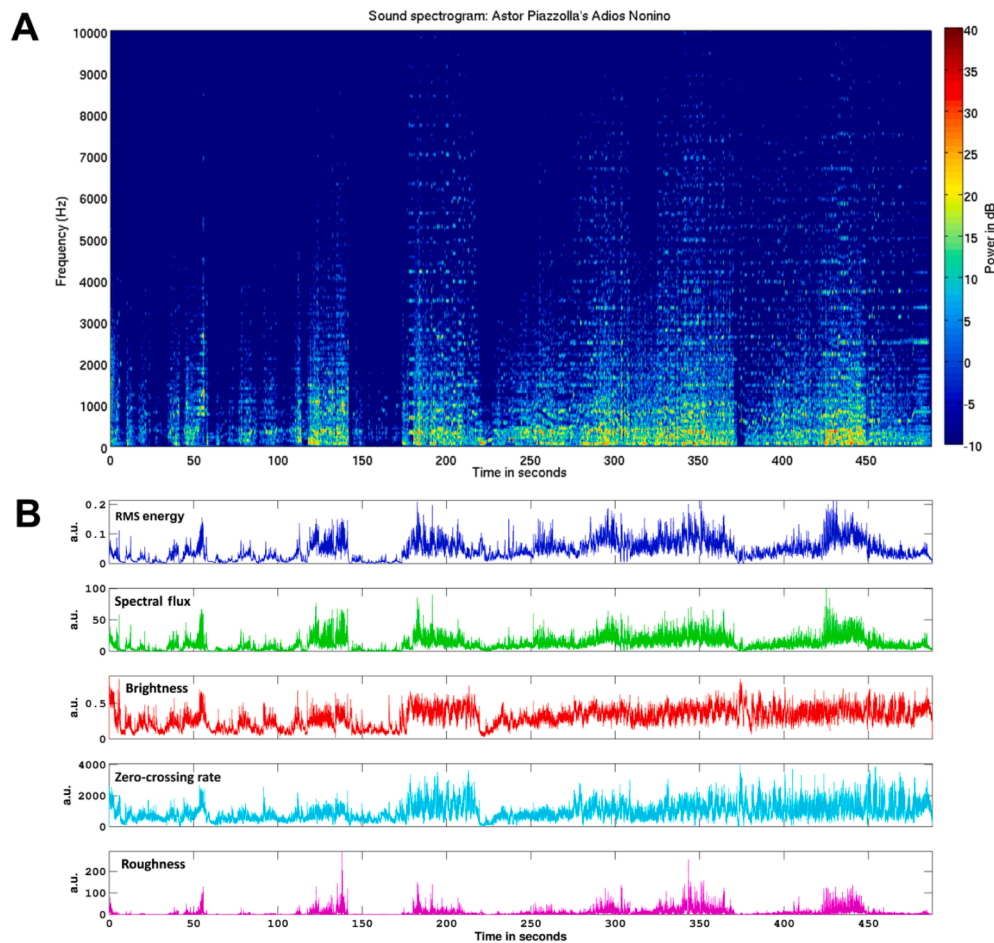

Fig. 1. Sound spectrogram of the music stimuli applied in this study. (A) Sound spectrogram for the naturalistic music piece used in the current study ("Adios Nonino" by Astor Piazzolla). **(B)** Five acoustical features (RMS energy, spectral flux, brightness, zero-crossing rate, roughness) extracted from the same music piece using the MIR toolbox (see details in Experimental Procedures).

naturalistic paradigm by combining sound onset detection and MEG/EEG measurements with the primary goal of extracting cortical ERs during a realistic listening situation. A related goal consisted of testing the reliability of these ERs and in deriving their neural generators. Finally, we aimed to explore the relationship between the ER amplitudes and the rate of the events occurring in realistic music.

2. Results

2.1. Chance level for extracting significant evoked responses at random time points

The chance level for extracting one significant ER to the naturalistic music piece at random time points (without acoustical feature analysis) was lower than 5% both for the EEG channels (maximum $p_{chance,1}$ = 0.026) and the MEG magnetometer channels (maximum $p_{chance,1}$ = 0.039). The chance level for randomly extracting significant ERs in averages across ten trials was estimated to be less than 1 out of 10,000 for the EEG channels ($p_{chance,10} < 0.0001$) and MEG magnetometer channels ($p_{chance,10} < 0.0001$). Moreover, the chance level for extracting significant ERs peaked in EEG channels Fz-Cz and MEG magnetometer left channel 1621 and right channel 1341, which are typical for measuring auditory cortical ERs. This suggests that random exploration of trigger time points is unlikely to lead to the extraction of significant ERs across ten trials by chance.

2.2. Verification of slow and fast event rates for automatic detections

An excerpt from the music piece containing silent breaks was selected as an example of slow event rates. Another excerpt from the music piece without silent breaks was selected as an example of fast event rates. Triggers, time points with salient increases in sound intensity (RMS energy) and spectral flux, assumed to indicate sound onsets that result in cortical ERs, were automatically extracted by an algorithm.

First, it was verified that the event rates differed significantly between the selected excerpts with 'slow' and 'fast' event rates. The event rates for the triggers at the slow event rates were on average 0.30 Hz (SD = 0.48) for sound intensity and 0.20 Hz (SD = 0.42) for spectral flux. By contrast, the event rates for the triggers at the fast event rates were on average 1.10 Hz (SD = 0.88) for sound intensity and 1.38 Hz (SD = 2.39) for spectral flux. The event rate was significantly slower for the triggers at the slow compared to the fast event rates, p=.016.

In addition, event densities for unconstrained triggers (without slow/fast event rate distinction) were on average 1.40 Hz (SD = 2.46) for sound intensity and 1.30 Hz (SD = 2.45) for spectral flux.

2.3. Extraction of P2 responses based on automatic sound onset detections

P2 responses differed significantly from the pre-stimulus baseline at time points when spectral flux and RMS energy increases were identified in slow-rate excerpts (Table 1). However, the acoustical changes at fast event rates and unconstrained stimulus rates were not consistently followed by significant P2 responses (Table 1).

Accordingly, we found that P2 amplitudes were significantly different in the comparison between fast and slow event rates, in the EEG (F(2,74) = 28.55, MSE = 14.24, $p < .001$, $\eta_p^2 = 0.44$), in the MEG magnetometers (F(1.60,62.30) = 45.53, MSE = 9845.38, $p < .001$, $\eta_p^2 = 0.54$) and in the MEG gradiometers (F(1.56,60.87) = 36.47, MSE = 205.13, $p < .001$, $\eta_p^2 = 0.48$). Post-hoc comparisons are shown in Table 2. Overall, these results indicate that the P2 responses are present at the slow event rates but undetectable at the fast event rates (Fig. 2 and Table 1).

The results also showed that the magnetic P2 responses at the slow event rates were significantly stronger in amplitude in the right hemisphere compared to the left hemisphere (magnetometers: F(1,39) = 4.61, MSE = 4366.41, $p = .038$, $\eta_p^2 = 0.11$; gradiometers: F(1,39) = 5.19, MSE = 140.93, $p = .028$, $\eta_p^2 = 0.12$).

2.4. Extraction of P1/N1/P2 responses based on manual sound onset detections

2.4.1. Reliability of automatic vs. manual sound onset detections

Manual sound onset detection was performed by a musicology expert (first author). Generally, the manual sound onset detection method seemed more reliable than the current automatic method. The automatic sound onset detections only partially corresponded with the manual

Table 2
Post-hoc comparisons between P2 responses at slow, unconstrained, and fast event rates. Showing p-values for post-hoc comparisons. Asterisks marks differences at the Bonferroni-corrected significance level at $p = .05/9 = 0.0056$.

Modality	Event rate	Post-hoc comparisons	
		Slow	Unconstrained
EEG	Slow		
	Unconstrained	0.005 *	
	Fast	<0.001 *	<0.001 *
MEG mag.	Slow		
	Unconstrained	<0.001 *	
	Fast	<0.001 *	<0.001 *
MEG grad.	Slow		
	Unconstrained	<0.001 *	
	Fast	<0.001 *	0.701

Table 1
Amplitudes of P2 responses to acoustical changes at slow, unconstrained, and fast event rates. Results of one-sample *t*-tests indicating whether event rates for sound intensity (RMS energy) and spectral flux increases results in amplitudes at the P2 latency significantly differing from the signal measured at the baseline in the EEG and MEG magnetometers. Asterisks marks differences at the Bonferroni-corrected significance level for multiple comparisons at $p = .05 / 18 = 0.0028$.

| Modality	Event rate	Feature	Central				Left				Right			
			df	t	p	mean	df	t	p	mean	df	t	p	mean
EEG	Slow	RMS energy	37	5.34	<0.001*	3.2 μV								
		Spec. flux	37	5.22	<0.001*	2.8 μV								
	Unconstrained	RMS energy	37	1.24	0.224	0.8 μV								
		Spec. flux	37	2.35	0.024	1.2 μV								
	Fast	RMS energy	37	−1.16	0.253	−0.7 μV								
		Spec. flux	37	−3.43	0.002*	−2.6 μV								
MEG mag.	Slow	RMS energy					39	6.17	<0.001*	83 fT	39	7.05	<0.001*	88 fT
		Spec. flux					39	4.53	<0.001*	77 fT	39	8.49	<0.001*	104 fT
	Unconstrained	RMS energy					39	1.01	0.318	9 fT	39	3.15	0.003	34 fT
		Spec. flux					39	2.78	0.008	31 fT	39	2.86	0.007	33 fT
	Fast	RMS energy					39	−3.67	0.001*	−38 fT	39	−0.30	0.768	−4 fT
		Spec. flux					39	−0.46	0.651	−7 fT	39	−1.27	0.211	−22 fT
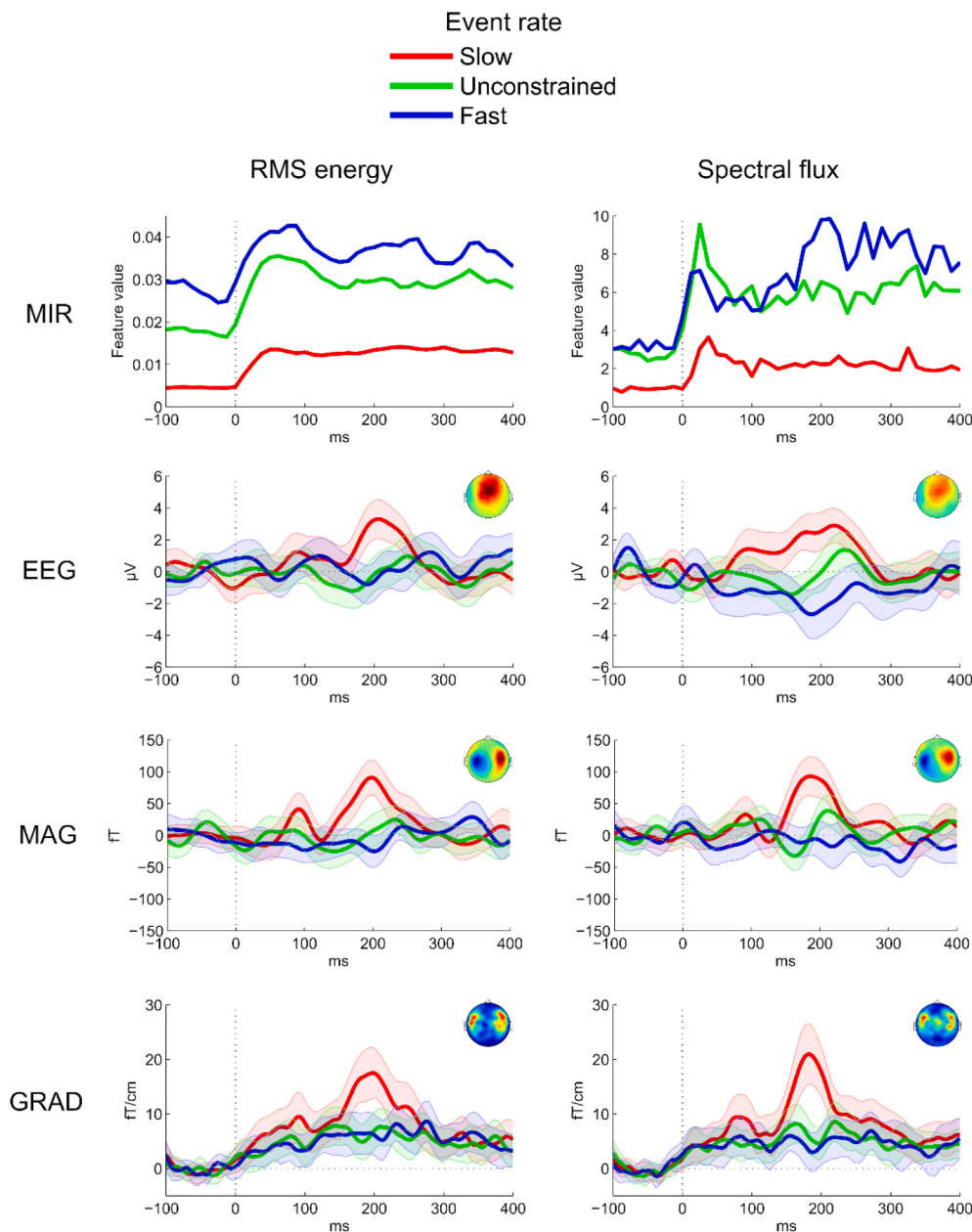
Fig. 2. Automatic detection waveforms showing the average extracted increases in RMS energy (sound intensity) and spectral flux and cortical ERs measured with EEG and MEG. (MAG = MEG magnetometers; GRAD = MEG gradiometers.) The onset of the MIR feature increases is located at time 0 ms (marked by the vertical line). The MEG waveforms show the average ERs across the left and right hemisphere channels. Shaded error bars indicate 95% confidence intervals. P2 topographies for the slow event rate condition are shown with cold to warm colors in the range -/+ 4 μV for the EEG, -/+ 100 fT for the MEG magnetometers, and 0–18 fT/cm for the MEG gradiometers.

sound onset detections, and the automatic detections missed nearly half of the perceivable sound events (Supplementary Material B). This indicates that the automatic method did not reliably detect perceivable sound onsets. Also, the missed sound onsets with the automatic method resulted in underestimated event rates. To our experience, lowering the threshold for detecting sound intensity and spectral flux increases lead to the detection of more unperceivable acoustical changes in addition to perceivable sound onsets, suggesting a relatively low specificity of the current automatic method.

Specifically, the automatically detected time points with MoRI peaks in the sound intensity (RMS energy) feature showed an average minimum deviation from the manual sound onset detection time points by 30 ms (SD = 26 ms). A total of 45.0% of perceivable sound onsets were missing (deviated more than 50 ms from the manual detections) with the automatic sound intensity (RMS)-based detections. Also, the automatically detected time points with MoRI peaks in the spectral flux feature showed an average minimum deviation from the manual sound onset detection time points by 37 ms (SD = 30 ms). A total of 41.6% of

perceivable sound onsets were missing (deviated more than 50 ms from the manual detections) with the automatic spectral flux-based detections.

2.4.2. Stimulus features at slow and fast event rates

The event rate at each manually detected sound onset was calculated as the inverse of the duration in seconds between the previous sound onset and the current sound onset (e.g., 0.5 s = 2 Hz event rate). Slower event rates showed lower sound intensity (RMS) and spectral flux measures prior to the manually detected sound onsets. This means that at slower event rates the sound intensity and spectral change related to the previous sound onset had more time to decay prior to the current sound onset (Fig. 3).

The statistical results verify this finding. The average sound intensity (RMS), $F(5,2112) = 25.14$, MSE = 0.001; $p < .001$, $\eta_p^2 = 0.06$, and spectral flux, $F(1,2112) = 30.51$, MSE = 46.36; $p < .001$, $\eta_p^2 = 0.07$, measures were significantly lower at slower event rates during the 500 ms preceding the manually detected sound onsets. Post hoc comparisons
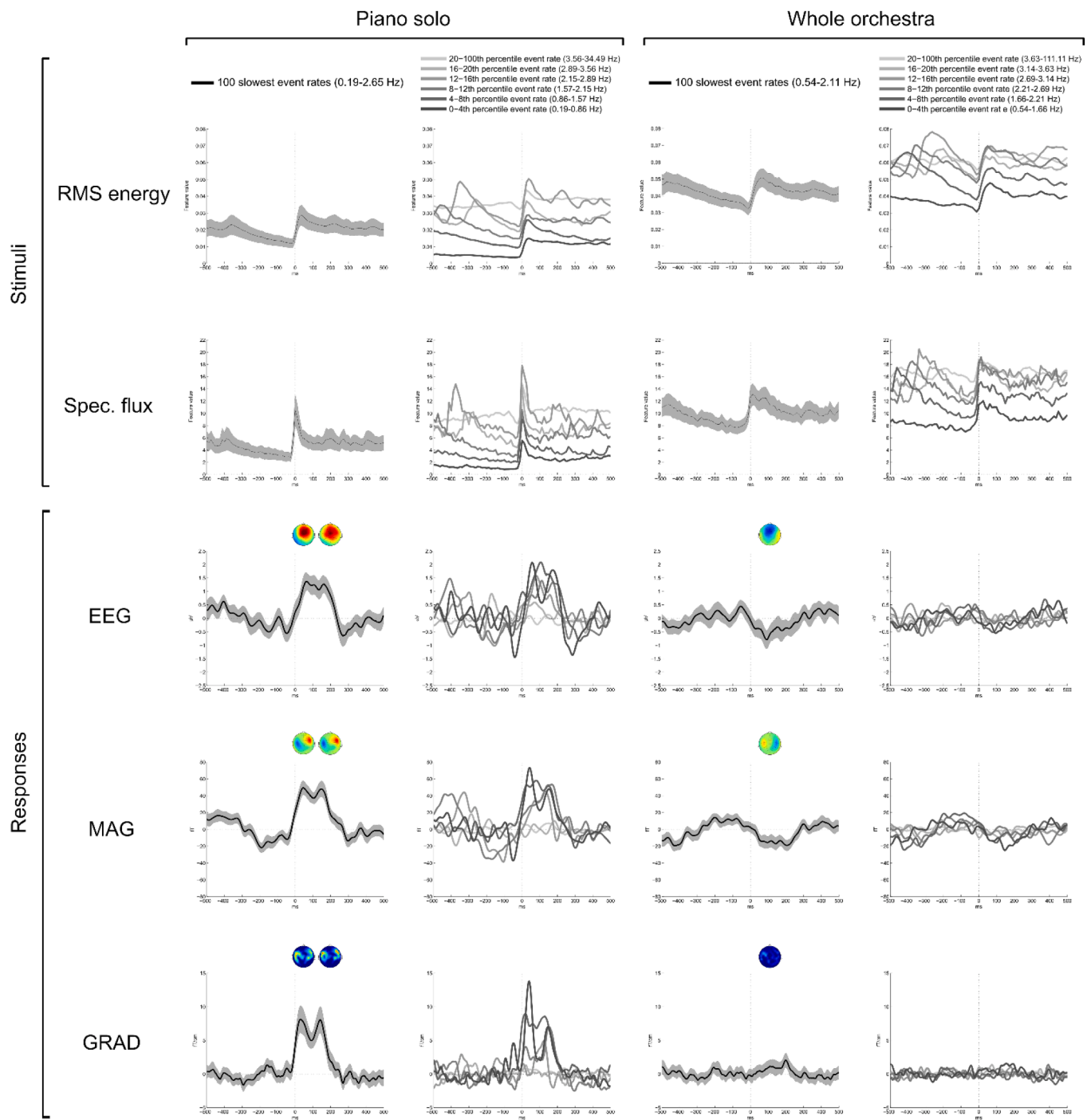
Fig. 3. Average increases in RMS energy (sound intensity), spectral flux, and cortical ERs measured at the manually detected onsets. Analyses of average stimuli and responses for the piano solo (0.117 – 174.182 s) and whole orchestra (174.303 – 484.850 s) across 100 trials with slowest event rates and across trials (~30–60 trials) within percentiles of event rates. (MAG = MEG magnetometers; GRAD = MEG gradiometers.) The manually detected sound onset is located at time 0 ms (marked by the vertical line). The MEG waveforms show the average ERs across the left and right hemisphere channels. Darker colors indicate slower event rates. Shaded error bars indicate 95% confidence intervals. P1/N1/P2 topographies are shown with cold to warm colors in the range -/+ 1 µV for the EEG, -/+ 50 fT for the MEG magnetometers, and 0–6 fT/cm for the MEG gradiometers.

confirmed significantly lower sound intensity (RMS) and spectral flux for the slowest 0-4th percentile event rates compared to all faster event rates ($p < .001−0.01$), for the second slowest 4-8th percentile compared to the faster 12-100th percentile event rates ($p < .001−0.05$), and for the third slowest 8-12th percentile event rates compared to the faster 20-100th percentile event rates ($p < .001-0.05$). Moreover, no significant interaction between event rate and instruments was found on neither the sound intensity (RMS), $F(5,2112) = 1.48$, MSE = 0.001; $p = .192$, $\eta_p^2 =$

0.00, nor the spectral flux, $F(5,2112) = 0.63$, MSE = 46.36; $p = .625$, $\eta_p^2 = 0.00$. This confirms that the stimulus features were lower at slower event rates regardless of the instruments.

Moreover, the solo piano compared to the whole orchestra showed generally lower sound intensity and spectral change measures prior to the manually detected sound onsets (Fig. 3). This means that the sound intensity and spectral flux did not return to as low sound intensity and spectral flux in the whole orchestra as in the piano solo (Fig. 3). The

average sound intensity (RMS), F(1,2112) = 259.38, MSE = 0.001; $p <$ .001, $\eta_p^2 = 0.11$, and spectral flux, F(1,2112) = 162.53, MSE = 46.36; $p < .001$, $\eta_p^2 = 0.07$, measures during the 500 ms preceding the manually detected sound onsets were significantly lower at the piano solo compared to the whole orchestra.

2.4.3. Reliability of manual vs. automatic sound onset detections for extracting P1/N1/P2 responses

Based on a 10 times larger number of successfully extracted trials, the manual sound onset detection method seemed more reliable than the current automatic method for extracting P1/N1/P2 responses.

Across 100 trials average P1 responses diverged significantly from the baseline in the EEG ($p < .001$) and MEG (L and R magnetometers, $p < .001$), N1 responses in the EEG ($p < .001$) and MEG (L and R magnetometers, $p < .001$), and P2 responses in the EEG ($p < .001$) and MEG (L and R magnetometers, $p < .001$) across the study participants when applying the manually detected sound onset time points (Fig. 3). In contrast, based on experience with the automatic method, when allowing to include more than 10 trials, or when lowering the detection threshold for increases in sound intensity (RMS energy) and spectral flux, the average ER amplitudes decreased until no ERs were

measurable.

2.4.4. P1/P2 responses to piano solo and N1 responses to the whole orchestra

The piano solo showed only measurable P1 and P2 responses, whereas the whole orchestra resulted in only measurable N1 responses (Fig. 3). Therefore, statistical comparisons of ER amplitudes at slower and faster event rates were conducted separately for the piano solo and the whole orchestra.

2.4.5. P1/N1/P2 responses measurable at slow event rates

Generally, the manual detections showed higher P1/N1/P2 amplitudes at slower event rates as compared with faster event rates (Fig. 3). During the piano solo, the P1 amplitude was significantly higher at slower event rates in the EEG, F(3.83,157.13) = 9.97, MSE = 4.59; $p < .001$, $\eta_p^2 = 0.20$, MEG magnetometers, F(3.54,162.99) = 30.24, MSE = 3523.11; $p < .001$, $\eta_p^2 = 0.40$, and MEG gradiometers, F(3.50,161.01) = 74.69, MSE = 81.78; $p < .001$, $\eta_p^2 = 0.62$. Post hoc comparisons are shown in Table 3. There was no main effect of the hemisphere on P1 amplitude in neither MEG magnetometers, $p = .439$, nor gradiometers, $p = .135$, and no interaction between hemisphere and event rate in MEG

Table 3

Post-hoc comparisons between P1, N1, and P2 amplitudes tested at different event rates based on manual sound onset detections. Showing *p*-values for post-hoc comparisons.

Component	Modality	Event rate	Post-hoc comparisons				
			0-4th %	4-8th %	8-12th %	12-16th %	16-20th %
P1	EEG	4-8th % (0.86–1.57 Hz)	0.024				
		8-12th % (0.86–2.15 Hz)	0.001	0.179			
		12-16th % (2.15–2.89 Hz)	0.082	0.721	0.141		
		16-20th % (2.89–3.56 Hz)	0.001	0.015	0.211	0.016	
		20-100th % (3.56–34.49)	<0.001	<0.001	<0.001	<0.001	0.149
	MEG mag.	4-8th % (0.86–1.57 Hz)	0.001				
		8-12th % (0.86–2.15 Hz)	<0.001	0.506			
		12-16th % (2.15–2.89 Hz)	<0.001	0.698	0.761		
		16-20th % (2.89–3.56 Hz)	<0.001	<0.001	<0.001	<0.001	
		20-100th % (3.56–34.49)	<0.001	<0.001	<0.001	<0.001	0.140
	MEG grad.	4-8th % (0.86–1.57 Hz)	<0.001				
		8-12th % (0.86–2.15 Hz)	<0.001	0.939			
		12-16th % (2.15–2.89 Hz)	<0.001	0.101	0.122		
		16-20th % (2.89–3.56 Hz)	<0.001	<0.001	0.001	0.024	
		20-100th % (3.56–34.49)	<0.001	<0.001	<0.001	<0.001	<0.001
P2	EEG	4-8th % (0.86–1.57 Hz)	0.244				
		8-12th % (0.86–2.15 Hz)	0.074	0.671			
		12-16th % (2.15–2.89 Hz)	0.046	0.277	0.696		
		16-20th % (2.89–3.56 Hz)	<0.001	0.001	0.010	0.004	
		20-100th % (3.56–34.49)	<0.001	<0.001	0.001	<0.001	0.737
	MEG mag.	4-8th % (0.86–1.57 Hz)	0.100				
		8-12th % (0.86–2.15 Hz)	0.638	0.025			
		12-16th % (2.15–2.89 Hz)	0.500	0.233	0.354		
		16-20th % (2.89–3.56 Hz)	<0.001	<0.001	<0.001	<0.001	
		20-100th % (3.56–34.49)	<0.001	<0.001	<0.001	<0.001	0.777
	MEG grad.	4-8th % (0.86–1.57 Hz)	0.011				
		8-12th % (0.86–2.15 Hz)	0.720	<0.001			
		12-16th % (2.15–2.89 Hz)	0.016	0.774	0.004		
		16-20th % (2.89–3.56 Hz)	<0.001	<0.001	<0.001	<0.001	
		20-100th % (3.56–34.49)	<0.001	<0.001	<0.001	<0.001	<0.001
N1	EEG	4-8th % (1.66–2.21 Hz)	0.432				
		8-12th % (2.21–2.69 Hz)	0.583	0.235			
		12-16th % (2.69–3.14 Hz)	0.260	0.075	0.561		
		16-20th % (3.14–3.63 Hz)	0.019	0.005	0.094	0.301	
		20-100th % (3.63–111.11)	0.028	0.008	0.169	0.579	0.258
	MEG mag.	4-8th % (1.66–2.21 Hz)	0.325				
		8-12th % (2.21–2.69 Hz)	0.004	0.024			
		12-16th % (2.69–3.14 Hz)	0.188	0.422	0.085		
		16-20th % (3.14–3.63 Hz)	0.012	0.034	0.817	0.174	
		20-100th % (3.63–111.11)	<0.001	<0.001	0.067	0.001	0.320
	MEG grad.	4-8th % (1.66–2.21 Hz)	0.591				
		8-12th % (2.21–2.69 Hz)	0.175	0.343			
		12-16th % (2.69–3.14 Hz)	0.154	0.059	0.005		
		16-20th % (3.14–3.63 Hz)	0.113	0.047	0.012	0.829	
		20-100th % (3.63–111.11)	<0.001	<0.001	<0.001	<0.001	<0.001

magnetometers, $p = .164$, and gradiometers, $p = .517$. This suggests that the P1 amplitude was higher at slower event rates as opposed to higher event rates in both hemispheres.

Moreover, during the piano solo, the P2 amplitude was significantly higher at slower than faster event rates in the EEG, $F(3.78, 154.98) = 8.82$, MSE $= 5.44$; $p < .001$, $\eta_p^2 = 0.18$, MEG magnetometers, $F(3.77, 173.44) = 23.55$, MSE $= 3673.90$; $p < .001$, $\eta_p^2 = 0.34$, and MEG gradiometers, $F(3.72, 171.02) = 70.25$, MSE $= 78.32$; $p < .001$, $\eta_p^2 = 0.61$. Post hoc comparisons are shown in Table 3. Similarly, there was no main effect of the hemisphere on P2 amplitude in neither MEG magnetometers, $p = .342$, nor gradiometers, $p = .443$, and no interaction between hemisphere and event rate in MEG magnetometers, $p = .174$, and gradiometers, $p = .055$, suggesting that the effect of event rate on P2 amplitude affected both hemispheres similarly.

Finally, the N1 responses to the whole orchestra (the polyphonic sections of the music) also showed significantly higher amplitude at slower event rates in the EEG, $F(4.01, 164.47) = 2.79$, MSE $= 2.17$; $p = .028$, $\eta_p^2 = 0.06$, MEG magnetometers, $F(3.54, 162.82) = 5.69$, MSE $= 1902.64$; $p < .001$, $\eta_p^2 = 0.11$, and MEG gradiometers, $F(4.04, 185.89) = 62.88$, MSE $= 15.22$; $p < .001$, $\eta_p^2 = 0.58$. Post hoc comparisons are shown in Table 3. As with the P1 and P2 responses, there was no main effect of the hemisphere on N1 amplitude in neither MEG magnetometers, $p = .321$, nor gradiometers, $p = .251$, and no interaction between hemisphere and event rate in MEG magnetometers, $p = .589$, and gradiometers, $p = .930$, suggesting that the effect of event rate on N1 amplitude affected both hemispheres similarly.

2.5. Localization of the naturalistic P2 responses based on automatic sound onset detections

Results of the source localization analysis suggest that the magnetic P2 responses at the slow event rates originate from the primary auditory cortex. Also, for RMS energy (sound intensity) additional sources were found in the anterior temporal lobe (BA 38) and for spectral flux in secondary auditory areas (BA 22) (Fig. 4).

3. Discussion

In this study, we demonstrated that average P1/N1/P2 responses can be successfully extracted at sound onsets in real musical pieces and with high reliability across 100 sound onsets. This finding substantially improves the signal-to-noise level as well as the efficiency of the M/EEG

recording time, because tenfold more events leading to detectable P1/N1/P2 responses were extracted in comparison to previous studies. We also verified that naturalistic P2 responses are localized to core cortical areas of the auditory cortex. Critically, we found that the chance to extract P1/N1/P2 responses at sound events is low in musical parts with fast event rates (greater than 2.5 Hz), but it can be increased by selecting slower music excerpts (0.2–2.5 Hz). With those excerpts, we observed that P1/P2 responses were predominant in the monophonic/homophonic sections of the music (piano solos), whereas N1 responses were only measurable in polyphonic parts with whole orchestration.

3.1. Naturalistic P1/N1/P2 responses to sound onsets

We observed that the P2 responses were evoked by increases in the acoustical features sound intensity (RMS energy) and spectral flux, which are commonly used to locate the onset times of syllables in speech and the onsets of tones in musical pieces (Alías et al., 2016). Despite the relatively low reliability of detecting sound onsets with the current automatic method, the idea of automatically or semi-automatically detecting sound onsets that evoke P1/N1/P2 seems promising. The focus on detecting sound onsets resulted, for the first time, in average P1/N1/P2 extracted across 100 trials when a manual sound onset method was applied, whereas previous studies have achieved extracting average P1/N1/P2 responses across <10 trials (Haumann et al., 2018; Poikonen et al., 2016a, 2016b). This suggests that with this method we were able to detect more acoustic events generating measurable event-related brain responses than in previous studies.

3.2. Extraction of naturalistic P1/N1/P2 responses at slow event rates

Our findings suggest that ERs can be better extracted at low event rates, which is relevant to consider when selecting music pieces for measuring P1/N1/P2 responses. Past behavioral work on rhythm perception shows an information processing bias towards inter-onset intervals (IOI) within the range of 50–1000 ms (Fraisse, 1978; Handel, 1989; Hirsh, 1959). It is known that the most perceptually salient IOIs in music appear to be within this relatively fast range of event rates between 1 and 20 Hz (Parncutt, 1994), which typically results in decreased ER amplitudes (Campbell et al., 1984; Davis and Zerlin, 1966; Geisler, 1960; Hari et al., 1982; Milner, 1969; Naatanen and Picton, 1987; Nelson and Lassman, 1968; Picton et al., 1977, 1978; Rees et al., 1986; Thompson and Spencer, 1966). However, the neural mechanisms
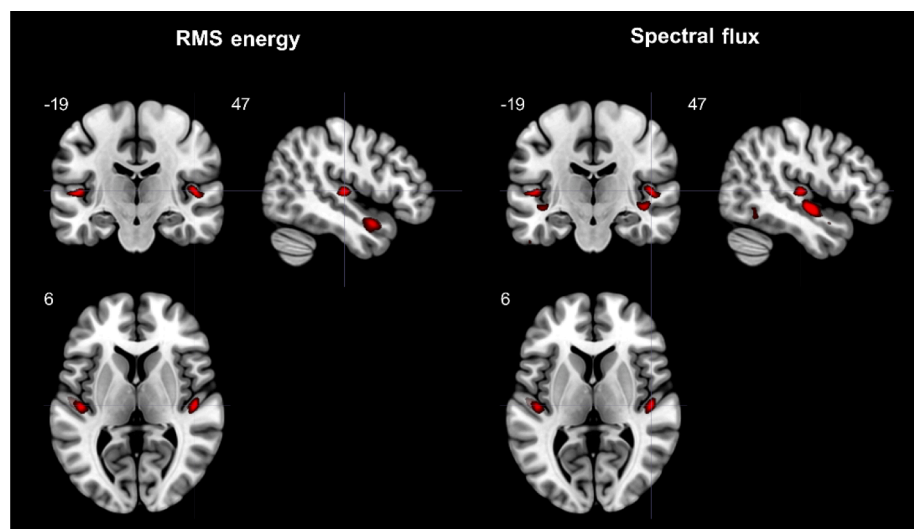

Fig. 4. Source location estimates of the P2 responses. SPM8 standard source imaging estimates of the P2 responses at the slow event rates are plotted with red colored blobs on the MNI standard T1-weighted MRI image. White numbers indicate MNI coordinates in mm of the view centred on the right auditory primary cortex at x = 47, y = −19, z = 6.

underlying the difference in response magnitude at different event rates have not previously been investigated in the context of naturalistic stimuli. Here we observed that naturalistic P1/N1/P2 responses were measurable across 100 trials only when the event rate is in the range of approximately 0.2–2.5 Hz.

The present findings of decreased amplitudes at faster event rates in the context of naturalistic stimuli can be interpreted according to alternative but not mutually exclusive hypotheses: superposition, refractoriness, habituation, or predictive coding (Heilbron and Chait, 2018). The superposition hypothesis focuses on the overlapping cortical responses in the M/EEG waveforms typically observed in fast rate events (Simon et al., 2019; Tan et al., 2015). According to this hypothesis, the sum of the overlapping response waveforms (repeatedly displaced at the rate of the stimulus IOI) will result in cancellation (or summation) between the overlapping positive and negative evoked potentials or fields. This destructive (or constructive) interference might explain the observed lower (or higher) response amplitudes. However, in studies where fast steady-state-responses (SSR, with periodic stimulation greater than 1 Hz) have been simulated by increasing the rate of overlapping cortical ERs (P1/N1/P2), it has been found that the superposition hypothesis is insufficient for characterizing additional amplitude changes across stimulus IOIs (Tan et al., 2015). Apart from stimulus-specific entrainment (e.g., Brenner et al., 2009), also the cortical refractoriness hypothesis becomes relevant to consider. The cortical refractoriness hypothesis assumes that a pool of excited cortical neurons responding to a stimulus become less responsive during a recovery period spanning up to ten seconds (Brattico et al., 2003; Zacharias et al., 2012). This passive adaptation effect is also known as 'neural fatigue', and it is interpreted as a mechanism that increases the processing efficiency of sensory systems (Grill-Spector et al., 2006). In this recovery period, the excitability of the cortical neurons is gradually restored. When the sound features of interest are presented with shorter IOIs, the auditory cortical neurons have a shorter time to fully recover their excitability. In turn, this would result in a lowered cortical excitability and lowered amplitudes of the cortical ERs. Nevertheless, the cortical refractoriness hypothesis has been challenged by recent findings suggesting the involvement of expectations (or 'predictions') (Costa-Faidella et al., 2011; Euler and Ricci, 1958; Pearce et al., 2010; Serkov et al., 1969; Todorovic et al., 2011).

The absence of measurable early cortical responses to repeated stimuli might be explained by the hypothesis of stimulus-specific habituation and dishabituation (Butler, 1968, 1972a, 1972b; Fruhstorfer et al., 1970; Graham, 1973; Gu et al., 2018; Loveless, 1983; Megela and Teyler, 1979; Naatanen and Picton, 1987; Naatanen et al., 1988; Öhman and Lader, 1977; Picton et al., 1978; Thompson and Spencer, 1966; Thompson and Groves, 1973; Woods and Elmasian, 1986) or the more recent predictive coding theory (Bendixen et al., 2009; Brattico et al., 2013; Herrmann et al., 2015; Todorovic et al., 2011; Vuust et al., 2009; Winkler et al., 2009). Early intracranial studies on cats showed that direct stimulation of neurons in the auditory medial geniculate body (MGB) (without stimulation of the ears) evoked typical responses in the primary auditory cortex, but the cortical responses did not decrease in amplitude with shorter IOIs during the MGB stimulation (Euler and Ricci, 1958; Serkov et al., 1969). In another condition, pairs of click sounds were delivered to the ears of the cats through a loudspeaker, and the typical decrease in cortical amplitude for the shorter IOIs reappeared (Euler and Ricci, 1958; Serkov et al., 1969). The authors suggest that response amplitudes for expected stimulus repetitions were reduced by certain inhibitory cortico-thalamic pathways, whereas the direct MGB stimulation might have perturbated these inhibitory pathways (Euler and Ricci, 1958; Serkov et al., 1969). A similar discussion on expectations and stimulus uncertainty has been related to human listeners (Todorovic et al., 2011). In one condition, listeners learned to expect that a click sound would not repeat, and when they heard the click sound repeating with a 500 ms delay the common reduction in N1 amplitude did not appear. In another condition, listeners learned to expect the 500 ms delayed repetition, and the typical pattern of N1 amplitude attenuation was restored. The authors interpret the dependency of the cortical ERs on expectations within the framework of predictive coding. Other recent studies, applying information-theoretic measures on melodic and rhythmic sequences, suggested that early cortical ERs might be affected by the predictability of the auditory stimulus (Lumaca and Baggio, 2016; Lumaca et al., 2018; Pearce et al., 2010). Music often consists of more repetitive and predictable auditory events, e.g., in comparison to spoken language (Huron, 2013), and therefore another possibility is that the fast events in the present study were repetitive and predictable. If the fast event rate stimuli were predictable, a factor of predictability might also explain the absence of measurable P1/N1/P2 responses (e.g., see Neuhaus and Knösche, 2006). In this respect, future studies could investigate whether there might also exist mismatch responses to unpredicted change of acoustical features in naturalistic music pieces (for a case study, see Haumann et al., 2018). Furthermore, it would be relevant to investigate the diminutive effects of neural adaptation on the P1/N1/P2 in clinical populations (Kim, 2015; Zhang et al., 2011) using naturalistic music stimuli.

3.3. Location of naturalistic P2 sources in the auditory cortex

The source localization analysis results showed that the neural generators of the naturalistic P2 responses to sound intensity increases were originating from the primary auditory cortex and the temporal pole (BA 38). This is consistent with previous positron emission tomography (PET) (Belin et al., 1998) and functional magnetic resonance imaging (fMRI) (Jancke et al., 1998) findings of increased activation in the primary auditory cortex and the temporal pole (BA 38) in participants listening to sound intensity changes in non-verbal and verbal stimuli. Moreover, the neural generators of the P2 to the spectral flux increases were estimated to originate from the primary (BA 41, 42) and secondary (BA 22) auditory cortex (BA 22). This is also in line with previous PET (Mirz et al., 1999) and fMRI (Menon et al., 2002) findings of activation in the primary and secondary auditory cortex related to spectral differences in sound.

3.4. Dominance of P1/P2 responses to monophonic/homophonic stimuli and N1 to polyphonic stimuli

One peculiar additional finding was that only P1 and P2 responses were measurable from the piano solo stimuli, whereas the sound onsets in the full orchestra only resulted in measurable N1 responses. Possibly, the different responses might be related to differences in musical texture (monophonic vs polyphonic sounds) between the piano solo and orchestra sounds. Whereas P1/N1/P2 responses to monophonic music stimuli have been thoroughly investigated, only a few studies have addressed the emergence of P1/N1/P2 responses to polyphonic music stimuli with two or more simultaneously sounding instrumental voices appears to be rare (for temporally isolated polyphonic tones, e.g., see Micheyl et al., 2007; Snyder et al., 2006). Though one MMN study includes a finding on standard P1/N1/P2 responses, and interestingly, in line with the present findings, the study shows that monophonic stimuli mainly lead to P1 and P2 responses, whereas responses to polyphonic stimuli were dominated by N1 (Huberth and Fujioka, 2017). The authors mention that the enhanced negativity for polyphonic stimuli is not well known, though, it might be related to the perception of concurrent auditory objects, as in object-related negativity (ORN) responses (i.e., to a mistuned harmonic within a complex tone), or neural processes related to auditory stream segregation (Huberth and Fujioka, 2017). Future studies might shed further light on the behavior of P1/N1/P2 responses to polyphonic stimuli with two or more independently changing auditory streams.

4. Limitations

One limitation of the present and previous studies (Poikonen et al., 2016a, 2016b) was the use of relatively few similar trials for estimating the average cortical ERs to automatically detected rapid changes in an acoustical feature. This means that the EEG or MEG waveforms of the auditory ERs are relatively susceptible to overlap from ongoing auditory ERs to preceding sound events. In particular, the fast event rate condition was more susceptible to overlap of responses to preceding sound events, whereas the slow event rate condition showed a comparably better pre-stimulus baseline interval. Also, with fewer trials there is a higher susceptibility to interferences from intrinsically induced alpha waves during rest or mu waves during movement inhibition. However, we found that the analyzed responses at the P2 latency were not distorted by alpha or mu waves (see Supplementary Material 3). Consistent responses were observable at the P2 latency, the responses differed significantly from the signal measured at the baseline in the slow event rate condition, and the significant responses indicated auditory topographies and sources typical of the auditory P2 (for further details on the statistical distributions of the responses, see Supplementary Material 2). Also, the contrast between triggers at slow versus fast event rates on P2 amplitude explained a large amount of the variance in the P2 amplitude ($\eta_p^2 = 0.44$– 0.54). Moreover, as shown in Supplementary Material 4, ERs in the P2 latency range were more consistently visible in single trials for the slow event rate condition compared to the fast rate condition (despite the relatively low signal-to-noise ratio at the single-trial level). Furthermore, we succeeded in extracting average P1/N1/P2 across 100 trials based on manual detection of sound onsets. The automatic sound onset detection did not seem to be particularly reliable when applied alone, though, future studies might further develop more perceptually valid automatic or semi-automatic sound onset detection methods for extracting ERs to sound onsets, which are faster and less susceptible to bias than exclusively manual sound onset detection. Further exploration of changes in MIR features at time points when perceivable sound onsets are manually indicated might lead to suggestions to develop better sound onset detection algorithms.

Another limitation was that it was impossible to control for confounding variables without intervening with the real music piece, and therefore it is difficult to draw conclusions regarding within-participant stimulus-related effects. By selecting music with different tempi, future studies could test whether an effect of the event rate was directly involved in modulating the P1/N1/P2 response to the naturalistic stimuli. Though, the main goal of the present study was to test the replicability of extracting ERs to naturalistic music excerpts across normal hearing populations. While controlled study designs are the most suitable for explaining within-participant stimulus-related effects, we believe that naturalistic designs are beneficial for group comparisons. The naturalistic paradigms allow testing for group differences in acoustical change detection, e.g., between participants of a normal and clinical population (also see Poikonen et al., 2016b), when groups are exposed to naturalistic stimuli.

5. Conclusion

Our findings support the emerging evidence that cortical ERs can be extracted from naturalistic music events at time points with a change in acoustical features, particularly at slow event rates between approximately 0.2–2.5 Hz. For the first time, we showed evidence that P1/N1/P2 responses across 100 manually detected onset times (i.e., trials) can be reliably extracted at sound onsets and automatically detected P2 responses can be localized to the auditory cortex while participants passively listen to a naturalistic music piece. Moreover, our results suggested that the naturalistic cortical ERs might differ between monophonic/homophonic and polyphonic sections of the music. These findings are highly promising for basic and clinical auditory neuroscience, as they offer novel possibilities to investigate auditory brain function and plastic changes in the normal and impaired hearing under ecologically valid auditory settings.

6. Experimental procedures

6.1. Participants

Forty-eight participants (23 females; mean age 28.3 years, SD = 8.6; 38 musicians and ten non-musicians) with normal hearing and no past neurological or psychiatric disorder were recruited for the study. All but two subjects were right-handed. The dataset is fully anonymized and has already been published elsewhere (e.g., Haumann et al., 2018). All experimental procedures for this study, included in the larger research protocol called "Tunteet", were approved by the Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (approval number: 315/13/03/00/11, obtained on March the 11th, 2012) and were conducted in agreement with the ethical principles of the Declaration of Helsinki. All participants signed an informed consent upon arrival to the laboratory.

6.2. Stimuli and procedures

All participants listened to three real music pieces as stimuli: (1) the first 8 min and 7 s of the tango nuevo piece *Adios Nonino* by Astor Piazzolla (*The Lausanne Concert* album, BMG Music, 1993) (Fig. 1); (2) an edited version of the first part of the modern orchestral ballet music *Rite of Spring* by Igor Stravinsky (7 min and 47 s) (Kirov Orchestra, St. Petersburg, conducted by Valery Gergiev, Decca Music Group Ltd., ℗© 2001 Universal International Music B.V); and (3) the first 7 min and 51 s of the progressive metal/rock piece *Stream of Consciousness* by Dream Theater (*Train of thought* album, Elektra Records, 2003). The presentation order was counterbalanced between subjects.

In a previous study (Haumann et al., 2018), we were unable to extract ERs from the M/EEG dataset applied in the present study, when we did not control for the possible effect of event rate. For the present study, the inclusion criteria were that the music piece must contain silent breaks with durations of at least one second to allow correction for an effect of event rate. The slow event rate criterion was chosen to maximize the chance of extracting high-amplitude non-overlapping ERs. Among the three pieces, the tango nuevo piece *Adios Nonino* was the only one that satisfied this criterion, and it was therefore selected for the current study. The *Rite of Spring* and the *Stream of Consciousness* pieces did not contain silent breaks of at least one second, and the results for these pieces are presented elsewhere (Haumann et al., 2018).

6.3. Acoustic feature extraction with MIR Toolbox

Acoustic features were extracted with the Music Information Retrieval (MIR) Toolbox (version 1.6.1) for Matlab (Lartillot and Toiviainen, 2007). Typically, the time points with sound event onsets in music and speech are estimated based on an increase in sound intensity or spectral flux (Alías et al., 2016) and have been found to evoke P1/N1/P2 responses in previous studies (Poikonen et al., 2016a, 2016b). Sound intensity is in the MIR Toolbox measured in root-mean-square (RMS) values of the audio waveform and partially relates to the perceived loudness. Spectral flux measures the distance between spectrums in adjacent time frames, showing peaks at time points with a change in timbre, pitch, or sound intensity. Since the peak of an evoked early cortical response component (P1/N1/P2) to an acoustical change lasts only a few tens milliseconds (Luck, 2014), a temporal accuracy of approximately 10 ms is necessary for capturing the cortical response. Therefore, the features were extracted in 25 ms time-windows with a 50% overlap, resulting in a sampling rate of 80 Hz (1 / 12.5 ms) for the acoustic features (Poikonen et al., 2016a, 2016b).

6.4. Capture of specific acoustical changes evoking cortical responses

In a preceding EEG study by Poikonen et al. (2016a) an algorithm was tuned by using specific parameter values adapted to each tested music piece and acoustical feature to capture ERs. Parameters were adjusted to identify time points triggering significant ERs across eight trials for a similar version of the music piece (*Adios Nonino* by Astor Piazzolla) investigated in the present study.

One limitation of the previous method was that the sound intensity (RMS energy) should not exceed a specific threshold prior to the trigger onset. This constrains detections to a specific quiet sound intensity level (-10% of the mean = 0.043) in 500 ms preceding the trigger onset (e.g., compare the value 0.043 with the RMS energy values shown in Fig. 5). Real music stimuli can consist of more voices or instrumental sounds (i. e., it can be polyphonic), where a sound onset can occur in one instrument while sounds in other instruments continue, which results in a relatively high sound intensity (RMS energy) preceding a sound onset. Therefore, we also applied an adapted method for detecting acoustical changes regardless of the preceding sound intensity level. Another issue is that previous studies included spectral acoustical features not directly related to sound onsets, whereas we here focused on only sound intensity (RMS) and spectral flux features assumed to relate to sound onsets.

6.5. Automatic capture of acoustical changes related to sound onsets

MIR feature increases assumed to relate to sound onsets that evoke brain responses were automatically identified by following three criteria (further details are shown in Supplementary Material 1):

1. The magnitude of the rapid increase (MoRI) should exceed a perceptual threshold (Poikonen et al., 2016a, 2016b), here set to the 70% percentile of ranked MoRI values.
2. Response overlap from a preceding salient feature increase should be minimized. To achieve this, time points with salient feature increases exceeding a MoRI threshold (criteria 1) in a preceding low increase phase (PLIP) (here defined as the preceding 1 s) are excluded.
3. The number of extracted trigger time points is kept constant for comparable signal-to-noise ratios when averaging across the responses. This is achieved by finding the desired number of trigger time points with maximum distance in time. In previous studies 7–8 trigger time points are applied (Poikonen et al., 2016a, 2016b). Here 10 time points are extracted.

This method allows trigger time points to be identified across a whole music piece regardless of the event rates (Fig. 5). However, the previous success of identifying ERs when detections were constrained to not exceed a specific quiet sound intensity level 500 ms before the trigger onset (Poikonen et al., 2016a) might be due to this constraint resulting in slow event rates, regardless of the strength (MoRI parameter) of the acoustical change. Therefore, we also tested the adapted method on selected excerpts with slow and fast event rates.

6.6. Isolating trigger time points at slow and fast event rates

As mentioned above, the time points with onsets of a sound stimulus in continuous music and speech are generally assumed to be located at the time points with an increase in sound intensity (i.e., indicating a rise in intensity or attack curve) or spectral flux (i.e., indicating a change in timbre, pitch, or intensity) (Alías et al., 2016). This corresponds to the MoRI peaks in RMS energy (sound intensity) and spectral flux.

Here, we investigated whether the amplitude of cortical P2 responses is lower in isolated triggers at fast event rates (excerpt without silent breaks of at least one second) compared to triggers isolated at slow event rates (excerpt with silent breaks of at least one second). Also, we included an unconstrained condition without control of the event rate
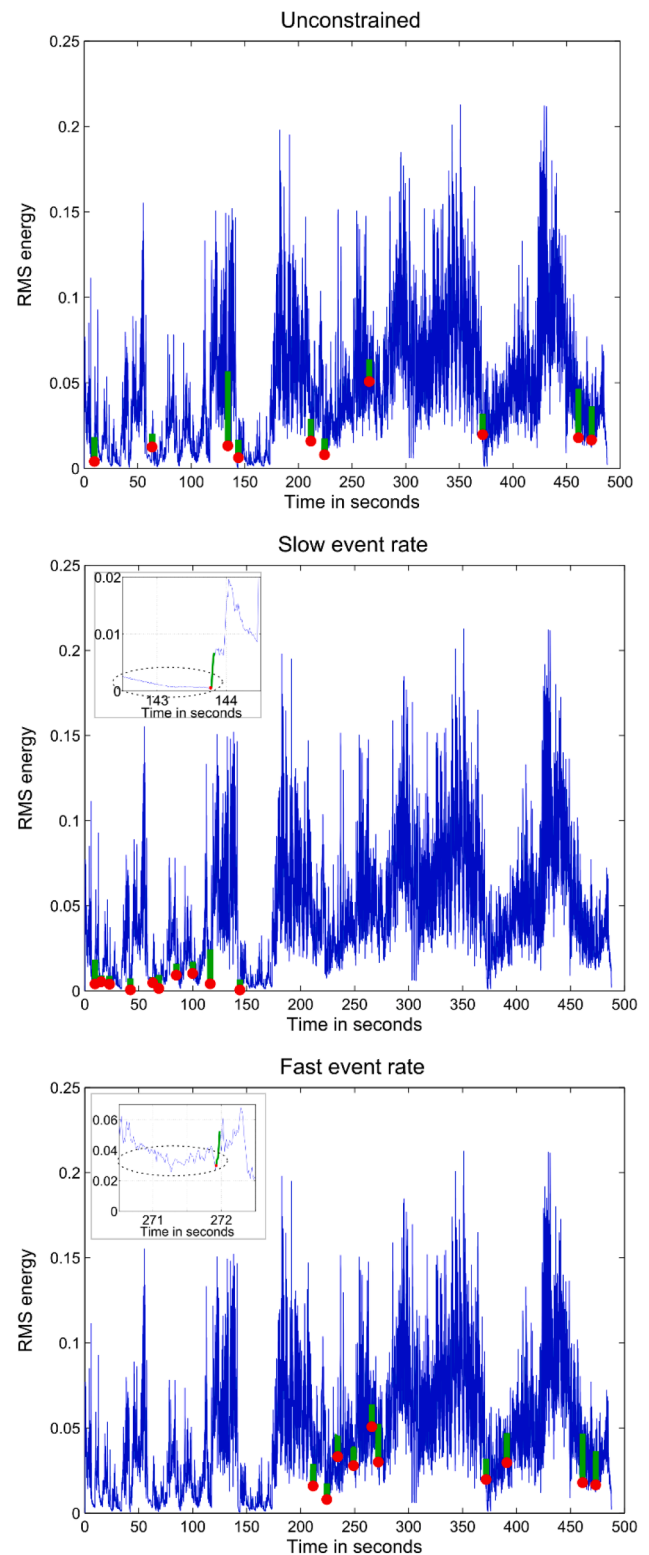

Fig. 5. Examples of trigger time points at unconstrained, slow, and fast event rates based on sound intensity (RMS energy) increases. The acoustic feature curve for the entire musical piece is shown in blue color, while trigger time points (based on MoRI peaks) are marked in red, and the feature increases across the following 50 ms are shown in green. The example shows increases in RMS energy (sound intensity). For more details, see Supplementary Material B.

Table 4

Automatically extracted triggers for the tango nuevo music piece. The number of MoRI peaks, *n* MoRI peaks, corresponds to those detected above the threshold. The assumed best triggers, *n* triggers, satisfy the 70% percentile MoRI threshold and have at least 1 s PLIP and longest distances to previous triggers in seconds. For the assumed best triggers, extracted without further event rate constraints, the median PLIP and distance to previous trigger values are shown with ranges in parenthesis.

Feature	*n* MoRI peaks	*n* triggers	PLIP	Distance to prev.
RMS energy	1773	10	1.502 (1.009–5.604)	14.710 (9.574–94.563)
Spectral Flux	3009	10	2.010 (1.27–5.601)	13.355 (7.932–130.003)

(Table 4, Fig. 5 top).

The MoRI threshold indicates the minimum required magnitude of increase in an acoustical feature for detecting a trigger time point. Since the highest feature increase will be most perceptually salient, the MoRI threshold was defined based on the percentage-wise distribution of the MoRI values sorted from the smallest to the largest acoustical feature increases. Thus, the higher the percentage-wise MoRI threshold, the higher the magnitude of increase applied as the MoRI threshold. Also, since the duration from the previous trigger time point to the next trigger time point depends on the MoRI threshold for detecting the trigger time points, a higher MoRI threshold increases the magnitudes allowed in the duration preceding the onset of a trigger time point. Following this rationale, the trigger time points for the slow event rate condition, with event IOIs of minimum one second, were extracted by applying a lowered MoRI threshold set to the 40% percentile of the ranked MoRI values, which as a result forced the PLIP to contain low feature values (i.e., low sound intensity and spectral flux) (Fig. 5 middle). Thereby, in the slow event rate condition, the acoustic feature increases were preceded by at least one second of low acoustic feature values, resulting in a slow event rate (Fig. 5 middle).

In contrast, the trigger time points for the fast event rate condition were obtained by using the same settings as above with the 70% percentile MoRI threshold, but constraining the trigger time points to the segments in the music piece with the highest acoustic feature values (i.e., highest sound intensity and spectral flux) in the PLIP (Fig. 5 bottom). This means that ongoing events occurred less than one second before each trigger time point, resulting in a fast event rate. Thus, three sets of ten triggers, one set of triggers for the unconstrained and each isolated event rates, were systematically obtained from different parts of the music piece (Fig. 5).

To prove that the event rate differed between the slow and fast event rate conditions we also included an event density measure from the MIR Toolbox (Lartillot and Toiviainen, 2007). The event density measures the average event rate across a moving time window, based on an estimate of the frequency of occurrence of amplitude peaks in the sound. For measuring the event density we applied a time window of 5 s with 10% overlap. Event densities for triggers at slow versus fast event rates were compared with a two-sample *t*-test.

6.7. Manual detection of sound onsets

To verify that the automatic sound onset detections did relate to perceivable sound onsets, we also included manual detection of sound onsets by a musicology expert (first author). All sound onset time points were detected using combined auditory and visual inspection of the sound spectrogram. The detected onsets mainly consisted of onsets of tones, but also a few clapping sound onsets from live audience at the beginning of the recording and a few clearly perceivable mechanical sounds from pressing keys on the accordion without producing tones. In total, 2127 sound events were detected.

6.8. MEG and EEG data acquisition

Simultaneous MEG and EEG data were collected at the Biomag Laboratory of the Helsinki University Central Hospital. The measurements were performed in an electrically and magnetically shielded room (ETS-Lindgren Euroshield, Eura, Finland) with Vectorview™ 306-channel MEG scanner (Elekta Neuromag®, Elekta Oy, Helsinki, Finland) equipped with a compatible EEG system. The MEG system had 102 SQUID sensor elements comprised of 102 axial magnetometers and 102 orthogonal planar gradiometer pairs. A 64-channel EEG electrode cap was used. The reference electrode was placed on the nose tip and the ground electrode was on the right cheek. Blinks, as well as vertical and horizontal eye movements, were measured with four electrodes attached above and below the left eye and close to the external eye corners on both sides. Four head position indicator coils were placed on top of the EEG cap. Their positions were located respectively to the nasion and the preauricular anatomical landmarks by Isotrack 3D digitizer (Polhemus, Colchester, VT, USA). MEG and EEG data were recorded with a sampling rate of 600 Hz. The music stimuli were presented with Presentation software (Neurobehavioral Systems, Ltd.). The sound was delivered through a pair of pneumatic headphones at a sound intensity adjusted to a comfortable level for each individual. There were short recording breaks after each music piece. In some sessions, more stimulation paradigms were presented to subjects, as specified in the Tunteet protocol, and the free-listening condition was kept as the last one.

During the MEG/EEG measurements, subjects were instructed to remain still, listen to the music and keep their eyes open. Moreover, in order to maintain subjects' attention towards the music, they were warned that after each music stimulus, namely during the silent break, they would be asked to provide ratings of familiarity, liking, and sensory pleasure on a scale from 1 to 5.

6.9. MEG and EEG data preprocessing

We applied Elekta Neuromag™ MaxFilter 2.2 Temporal Signal Space Separation (tSSS) (Taulu and Hari, 2009) to minimize the influence of external and nearby noise sources and automatically detect and correct bad MEG channels. This spatial filtering was achieved with the default inside expansion order of 8, outside expansion order of 3, automatic optimization of both inside and outside bases, subspace correlation limit of 0.980, and raw data buffer length of 10 s. The subsequent data processing was performed with FieldTrip version r9093, an open-source toolbox for Matlab (Donders Institute for Brain, Cognition and Behaviour/Max Planck Institute, Nijmegen, the Netherlands) (Oostenveld et al., 2011) and Matlab R2013b (MathWorks, Natick, Massachusetts). On average, 0.17 (range 0–8) EEG channels per participant were bad and corrected by replacing them with interpolation of the neighboring channels. To suppress slow drifts, the M/EEG data were high-pass filtered at 1 Hz, and to minimize muscular artifacts a low-pass filter at 25 Hz was applied (cf. (Poikonen et al., 2016a, 2016b)). The data was decomposed with independent component analysis (ICA) (Makeig et al., 1996), and when one component clearly reflected a vertical eyeblink, one horizontal eye movement, or one ECG artifact, their projections were subtracted from the data. On average, 2.16 (range 1–3) component projections were subtracted from the EEG, average 2.67 (range 1–3) from the MEG magnetometers, and average 2.64 (range 1–3) from the MEG gradiometers. Trials with remaining artifacts were detected and rejected by applying amplitude thresholds of 100 µV for the EEG, 2000 fT for the MEG magnetometers, and 400 fT/cm for the MEG gradiometers. This resulted in a rejection of average 0.92 (range 0–5) trials from the EEG, average 0.09 (range 0–5) trials from the MEG magnetometers, and average 0.05 (range 0–5) trials from the MEG gradiometers. After rejecting trials and removing data of participants with more than half of the trials rejected, from the pool of 48 participants clean EEG data remained for 38 participants with an average of 9.61 trials per condition, while 40 other participants showed clean MEG data with an

average of 9.88 trials per condition. Afterward, the data was baseline corrected by subtracting the average from –100 to 0 ms before the estimated trigger and averaged across the trials.

Since the planar gradiometer sensors measure the difference in the magnetic field across two orthogonal directions, the measures from each couple of longitudinal and latitudinal gradiometer sensors were combined by applying the Pythagorean distance formula, as implemented in FieldTrip, $d = \sqrt{longitudinal^2 + latitudinal^2}$.

6.10. ERP/ERF amplitude analysis

For participant-level analysis, ER amplitudes were calculated from the central EEG electrode Cz (Poikonen et al., 2016a, 2016b), and the hemisphere-specific MEG magnetometer sensors, 0231 (left) and 1341 (right), and MEG combined planar gradiometer sensors 0212 + 13 (left) and 1322 + 23 (right), which showed the largest ER amplitudes above each hemisphere.

For the automatic detections, we focused our analysis on the P2 response, which was the only clearly measurable response. The P2 amplitudes were obtained from the mean amplitude within a 30 ms time window set around the peak latency. The P2 peak latency was identified as the maximum positive peak between 100 and 300 ms in the grand-average waveforms for the slow event rate acoustical changes at the EEG Cz electrode (Sound intensity: 206 ms; Spectral flux: 216 ms), across left (0231) and right (1341) MEG magnetometer peak channels (Sound intensity: 200 ms; Spectral flux: 186 ms), and across left (0212 + 13) and right (1322 + 23) MEG gradiometer peak channels (Sound intensity: 203 ms; Spectral flux: 183 ms).

For the manual detections, P1, N1, and P2 were measurable. The P1/N1/P2 amplitudes were measured within a 30 ms time window around the peak latencies in the grand-average waveform across 100 trials with slowest event rates. The P1 peaked in the EEG at 60 ms, in the MEG magnetometers at 50 ms, and in the MEG gradiometers at 47 ms. The N1 peaked in the EEG at 90 ms, in the MEG magnetometers at 140 ms, and in MEG magnetometers at 110 ms. Finally, the P2 peaked in the EEG at 163 ms, in the MEG magnetometers at 147, and in the MEG gradiometers at 143 ms.

Group-level statistics was performed with SPSS version 20 (IBM, Armonk, New York, USA). First, to consider the reliability of the ERs based on the automatic detections, one-sample t-tests were conducted on the amplitudes of each average ER for each feature and stimulus rate condition, thereby testing whether each average ER differed significantly from the signal measured at the baseline at 0 µV or 0 ft. (One-sample t-tests were not conducted on the MEG gradiometers, due to inflation of the grand-average signal in the positive root-mean-squared values for the combined MEG gradiometers.) Subsequently, ANOVA models were applied to investigate whether the amplitudes showed significant differences across Feature (*RMS energy (Sound intensity)*, *Spectral flux*) and Event rate (*Slow, Unconstrained, Fast*). In addition, the factor Hemisphere (*Left, Right*) was examined for the MEG measurements. Greenhouse-Geisser correction of degrees of freedom was applied in cases where Mauchly's sphericity test showed that the assumption of sphericity was violated.

The correspondence between the automatically and manually detected sound onsets time points was investigated by measuring the minimum difference between each manually detected sound onset time point and the automatically detected time points. For the manual sound onset detections, the event rates were calculated as the inverse of the difference between the onset times of the detected sound onset and the previous sound onset (e.g., 0.5 s = 2 Hz event rate). For more detailed comparisons, event rates were divided into equal sample sizes with slower 0-4th, 4-8th, 8-12th, 12-16th, 16th-20th percentile event rates, and very fast 20-100th percentile event rates. Statistical testing of fixed effects of event rate on sound intensity (RMS energy) and spectral flux averages over 500 ms preceding the sound onsets was conducted with

two-way ANOVAs. Comparisons of P1/N1/P2 amplitudes at different event rates were investigated with repeated measures ANOVAs.

It has recently been argued that due to noise and a large number of time samples in EEG/MEG there is a relatively high chance of observing a significant effect, even when the effect is false and tested conditions are assigned to randomly placed triggers (Luck and Gaspelin, 2017). Therefore, we additionally tested the chance level for extracting significant ERs by placing the triggers randomly. This was investigated by moving the 30 ms amplitude measurement time window across all possible time points from the beginning to the end of the piece and counting the number of significant and non-significant one-sample t-tests at $p < .05$ for each EEG and MEG channel. From this, we derived the chance of extracting one significant ER,

$$p_{chance} = \frac{number\ of\ significant\ tests}{total\ number\ of\ tests}.$$

Also, following the same procedure, we measured the chance of extracting significant average ERs in the 30 ms amplitude measurement time window across ten randomly selected trials over 10,000 one-sample t-tests with significance level at $p < .05$ for each EEG and MEG channel.

6.11. MEG source localization analysis

Distributed source modeling was conducted with SPM8 version r6313 (Wellcome Trust Centre for Neuroimaging, London, United Kingdom) (Friston et al., 2007, 2008). A high-resolution mesh was applied consisting of 20,484 vertices representing the cortex in the canonical T1 MNI image. The forward lead fields were calculated with the single-shell model for MEG (Nolte, 2003) implemented in FieldTrip. Due to low signal-to-noise ratios, it was necessary to estimate the source locations based on the grand-average magnetometer waveforms, instead of the individual single-subject waveforms. The standard GS algorithm in SPM8 was used to estimate the inverse solutions (Friston et al., 2008). The source inversion algorithm is based on multiple sparse priors (MSP) representing patches in the cortex (Friston et al., 2008; Lopez et al., 2014). The resulting localization estimates were rendered on the standard 3D MNI image by applying the MRIcroGL software (Rorden et al., 2007).

Funding

Author contributions

NH designed the study, analyzed the data, and wrote the manuscript. ML contributed to the data analysis and writing of the manuscript. MK collected the data and contributed to the manuscript writing. JS participated in the data analysis and manuscript writing. PV contributed to the writing of the manuscript and provided funding. EB contributed with the study design, data analysis, and writing of the manuscript. All authors have approved the final version of the manuscript and agreed to be accountable for all aspects of the work.

9. Contributions to the field

An emerging research field investigates how brain function relates to the perception of realistic stimuli. Most traditional laboratory experiments apply simplified stimuli, which are less complex than those encountered in real everyday environments. It is, therefore, unknown to which extent previous findings can be generalized to real everyday environments. One general approach is to extract cortical ERs to sound onsets. Another finding observed in classical experiments is the effect of stimulus rate, where the early evoked cortical ERs are diminished in

response to repeated fast events, even though the intensity of the stimulus is not diminished. The effect of event rate on ERs is for example relevant for the study of hearing disorders. However, it is unknown whether the diminishing effect of the event rate is also present in ERs to realistic sound stimuli. In this study, we present an automatic method that extracts acoustic changes in a real music piece, which coincide with 55–58% perceivable sound onsets that trigger neural responses. Also, we test a manual method for detecting sound onsets that trigger neural responses. We validate for the first time that cortical ERs can be extracted across 100 sound onsets, and with MEG source localization analysis we verify that measured brain responses to realistic music stimuli originate from the auditory cortex. Furthermore, we find that cortical ERs were only measurable at slow event rates (0.2–2.5 Hz), and response types varied across different instruments in healthy listeners listening to a real music piece.

CRediT authorship contribution statement

Niels T. Haumann: Conceptualization, Methodology, Validation, Formal analysis, Writing - original draft, Visualization. **Massimo Lumaca:** Conceptualization, Methodology, Writing - review & editing. **Marina Kliuchko:** Investigation, Writing - review & editing. **Jose L. Santacruz:** Conceptualization, Methodology, Writing - original draft. **Peter Vuust:** Writing - review & editing, Funding acquisition. **Elvira Brattico:** Conceptualization, Methodology, Writing - review & editing, Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.brainres.2020.147248.

References

Alías, F., Socoró, J., Sevillano, X., 2016. A Review of physical and perceptual feature extraction techniques for speech, music and environmental sounds. Appl. Sci. 6.

Alluri, V., Toiviainen, P., Jaaskelainen, I.P., Glerean, E., Sams, M., Brattico, E., 2012. Large-scale brain networks emerge from dynamic processing of musical timbre, key and rhythm. Neuroimage. 59, 3677–3689.

Belin, P., McAdams, S., Smith, B., Savel, S., Thivard, L., Samson, S., Samson, Y., 1998. The functional anatomy of sound intensity discrimination. J. Neurosci. 18, 6388–6394.

Bendixen, A., Schroger, E., Winkler, I., 2009. I heard that coming: event-related potential evidence for stimulus-driven prediction in the auditory system. J. Neurosci. 29, 8447–8451.

Bo, H., Li, H., Ma, L., Yu, B., 2016. Time-Course EEG Spectrum Evidence for Music Key Perception and Emotional Effects. Advances in Brain Inspired Cognitive Systems, Vol., ed. Springer International Publishing, Cham, pp. 184-195.

Brattico, E., Tervaniemi, M., Picton, T.W., 2003. Effects of brief discrimination-training on the auditory N1 wave. Neuroreport 14, 2489–2492.

Brattico, E., Bogert, B., Jacobsen, T., 2013. Toward a neural chronometry for the aesthetic experience of music. Front. Psychol. 4.

Brenner, C.A., Krishnan, G.P., Vohs, J.L., Ahn, W.Y., Hetrick, W.P., Morzorati, S.L., O'Donnell, B.F., 2009. Steady state responses: electrophysiological assessment of sensory function in schizophrenia. Schizophr. Bull. 35, 1065–1077.

Burunat, I., Toiviainen, P., Alluri, V., Bogert, B., Ristaniemi, T., Sams, M., Brattico, E., 2016. The reliability of continuous brain responses during naturalistic listening to music. Neuroimage. 124, 224–231.

Butler, R.A., 1968. Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. J. Acoust. Soc. Am. 44, 945–950.

Butler, R.A., 1972a. Frequency specificity of the auditory evoked response to simultaneously and successively presented stimuli. Electroencephalogr. Clin. Neurophysiol. 33, 277–282.

Butler, R.A., 1972b. The auditory evoked response to stimuli producing periodicity pitch. Psychophysiology 9, 233–237.

Campbell, K., Marois, R., Arcand, L., 1984. Ethanol and the event-related evoked-potentials - effects of rate of stimulus-presentation and task-difficulty. Ann. N. Y. Acad. Sci. 425, 551–555.

Costa-Faidella, J., Baldeweg, T., Grimm, S., Escera, C., 2011. Interactions between "What" and "When" in the Auditory System: Temporal Predictability Enhances Repetition Suppression. J. Neurosci. 31, 18590–18597.

Crosse, M.J., Di Liberto, G.M., Bednar, A., Lalor, E.C., 2016. The Multivariate Temporal Response Function (mTRF) Toolbox: A MATLAB Toolbox for Relating Neural Signals to Continuous Stimuli. Front. Hum. Neurosci. 10, 604.

Davis, H., Zerlin, S., 1966. Acoustic Relations of the Human Vertex Potential. J. Acoust. Soc. Am. 39, 109–116.

Euler, C.V., Ricci, G.F., 1958. Cortical evoked responses in auditory area and significance of apical dendrites. J. Neurophysiol. 21, 231–246.

Fraisse, P., 1978. Time and rhythm perception. In: Handbook of Perception. Vol. Vol. 8: Perceptual Coding, E.C.a.F. Carterette, M.P., ed.ˆeds. Academic Press.

Friston, K.J., Ashburner, J.T., Kiebel, S.J., Nichols, T.E., Penny, W.D., 2007. Statistical parametric mapping: The analysis of functional brain images. Vol., Amsterdam, Boston: Elsevier/Academic Press.

Friston, K.J., Harrison, L., Daunizeau, J., Kiebel, S., Phillips, C., Trujillo-Barreto, N., Henson, R., Flandin, G., Mattout, J., 2008. Multiple sparse priors for the M/EEG inverse problem. Neuroimage. 39, 1104–1120.

Fruhstorfer, H., Soveri, P., Jarvilehto, T., 1970. Short-term habituation of the auditory evoked response in man. Electroencephalogr. Clin. Neurophysiol. 28, 153–161.

Geisler, C.D., 1960. Average responses to clicks in man recorded by scalp electrodes. Technical Report. 380, 154–158.

Graham, F.K., 1973. Habituation and dishabituation of responses innervated by the autonomic nervous system. In Habituation. Vol. Vol. 1: Behavioral studies H.V.S.P. M.J. Herz, ed.ˆeds. Academic Press, London, pp. 163-218.

Grill-Spector, K., Henson, R., Martin, A., 2006. Repetition and the brain: neural models of stimulus-specific effects. Trends Cogn. Sci. 10, 14–23.

Gu, F., Wong, L., Chen, F., Huang, W.-T., Wang, L., Hu, A.X., 2018. Lateral inhibition is a neural mechanism underlying mismatch negativity. Neuroscience 385, 38–46.

Handel, S., 1989. Listening: An Introduction to the Perception of Auditory Events. Vol., The MIT Press.

Hari, R., Kaila, K., Katila, T., Tuomisto, T., Varpula, T., 1982. Interstimulus-Interval Dependence of the Auditory Vertex Response and Its Magnetic Counterpart - Implications for Their Neural Generation. Electroencephalogr. Clin. Neurophysiol. 54, 561–569.

Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., Malach, R., 2004. Intersubject synchronization of cortical activity during natural vision. Science 303, 1634–1640.

Hasson, U., Honey, C.J., 2012. Future trends in Neuroimaging: Neural processes as expressed within real-life contexts. Neuroimage. 62, 1272–1278.

Haumann, N., Kliuchko, M., Vuust, P., Brattico, E., 2018. Applying acoustical and musicological analysis to detect brain responses to realistic music: a case study. Appl. Sci. 8.

Heilbron, M., Chait, M., 2018. Great expectations: is there evidence for predictive coding in auditory cortex? Neuroscience 389, 54–73.

Herrmann, B., Henry, M.J., Fromboluti, E.K., McAuley, J.D., Obleser, J., 2015. Statistical context shapes stimulus-specific adaptation in human auditory cortex. J. Neurophysiol. 113, 2582–2591.

Hirsh, I.J., 1959. Auditory perception of temporal order. J. Acoust. Soc. Am. 31, 759–767.

Holdgraf, C.R., Rieger, J.W., Micheli, C., Martin, S., Knight, R.T., Theunissen, F.E., 2017. Encoding and decoding models in cognitive electrophysiology. Front. Syst. Neurosci. 11, 61.

Huberth, M., Fujioka, T., 2017. Neural representation of a melodic motif: Effects of polyphonic contexts. Brain Cogn. 111, 144–155.

Huron, D., 2013. A psychological approach to musical form: the habituation-fluency theory of repetition. Curr. Musicol. 96.

Jancke, L., Shah, N.J., Posse, S., Grosse-Ryuken, M., Muller-Gartner, H.W., 1998. Intensity coding of auditory stimuli: an fMRI study. Neuropsychologia. 36, 875–883.

Kim, J.R., 2015. Acoustic change complex: clinical implications. J. Audiol. Otol. 19, 120–124.

Lartillot, O., Toiviainen, P., 2007. A Matlab Toolbox for Musical Feature Extraction From Audio. International Conference on Digital Audio Effects.

Lopez, J.D., Litvak, V., Espinosa, J.J., Friston, K.J., Barnes, G.R., 2014. Algorithmic procedures for Bayesian MEG/EEG source reconstruction in SPM. Neuroimage. 84, 476–487.

Loveless, N.E., 1983. The orienting response and evoked potentials in man. In Orienting and habituation: Perspectives in human research. Vol., D. Siddle, ed.ˆeds. John Wiley & Son, New York, pp. 71-108.

Luck, S.J., 2014. An introduction to the event-related potential technique. Vol. MIT press.

Luck, S.J., Gaspelin, N., 2017. How to get statistically significant effects in any ERP experiment (and why you shouldn't). Psychophysiology 54, 146–157.

Lumaca, M., Baggio, G., 2016. Brain potentials predict learning, transmission and modification of an artificial symbolic system. Soc. Cogn. Affect. Neurosci. 11, 1970–1979.

Lumaca, M., Trusbak Haumann, N., Brattico, E., Grube, M., Vuust, P., 2018. Weighting of neural prediction error by rhythmic complexity: A predictive coding account using mismatch negativity. Eur. J. Neurosci.

Makeig, S., Bell, A.J., Jung, T.P., Sejnowski, T.J., 1996. Independent component analysis of electroencephalographic data. Adv. Neural. Inf. Process Syst. 8 (8), 145–151.

Mäkinen, V., May, P., Tiitinen, H., 2004. Transient brain responses predict the temporal dynamics of sound detection in humans. Neuroimage. 21, 701–706.

Megela, A.L., Teyler, T.J., 1979. Habituation and the human evoked-potential. J. Comp. Physiol. Psychol. 93, 1154–1170.

Menon, V., Levitin, D.J., Smith, B.K., Lembke, A., Krasnow, B.D., Glazer, D., Glover, G.H., McAdams, S., 2002. Neural correlates of timbre change in harmonic sounds. Neuroimage. 17, 1742–1754.

Micheyl, C., Carlyon, R.P., Gutschalk, A., Melcher, J.R., Oxenham, A.J., Rauschecker, J. P., Tian, B., Courtenay Wilson, E., 2007. The role of auditory cortex in the formation of auditory streams. Hear. Res. 229, 116–131.

Mikutta, C.A., Maissen, G., Altorfer, A., Strik, W., Koenig, T., 2014. Professional Musicians Listen Differently to Music. Neuroscience 268, 102–111.

Milner, B.A., 1969. Evaluation of auditory function by computer techniques. Int. Audiol. 8, 361–370.

Mirz, F., Ovesen, T., Ishizu, K., Johannsen, P., Madsen, S., Gjedde, A., Pedersen, C.B., 1999. Stimulus-dependent central processing of auditory stimuli - A PET study. Scand. Audiol. 28, 161–169.

Naatanen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound – A review and an analysis of the component structure. Psychophysiology 24, 375–425.

Naatanen, R., Sams, M., Alho, K., Paavilainen, P., Reinikainen, K., Sokolov, E.N., 1988. Frequency and Location Specificity of the Human Vertex N1-Wave. Electroencephalogr. Clin. Neurophysiol. 69, 523–531.

Nelson, D.A., Lassman, F.M., 1968. Effects of intersignal interval on the human auditory evoked response. J. Acoust. Soc. Am. 44, 1529–1532.

Neuhaus, C., Knösche, T.R., 2006. Processing of Rhythmic and Melodic Gestaltsâ An ERP Study. Music Perception. 24, 209–222.

Nolte, G., 2003. The magnetic lead field theorem in the quasi-static approximation and its use for magnetoencephalography forward calculation in realistic volume conductors. Phys. Med. Biol. 48, 3637–3652.

Öhman, A., Lader, M., 1977. Short-term changes of the human auditory evoked potentials during repetitive stimulation. . In Auditory evoked potentials. Vol., J.E. Desmedt, ed.ˆeds. S. Karger, Basel, Munchen, Paris, London, New York, Sydney, pp. 93-118.

Oostenveld, R., Fries, P., Maris, E., Schoffelen, J.M., 2011. FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. Computational Intelligence and Neuroscience.

Parncutt, R., 1994. A perceptual model of pulse salience and metrical accent in musical rhythms. Music Perception. 11, 409–464.

Pearce, M.T., Ruiz, M.H., Kapasi, S., Wiggins, G.A., Bhattacharya, J., 2010. Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. Neuroimage. 50, 302–313.

Picton, T.W., Woods, D.L., Baribeaubraun, J., Healey, T.M.G., 1977. Evoked-Potential Audiometry. J. Otolaryngol. 6, 90–119.

Picton, T.W., Woods, D.L., Proulx, G.B., 1978. Human Auditory Sustained Potentials. 2. Stimulus Relationships. Electroencephalogr Clin Neurophysiol. 45, 198–210.

Poikonen, H., Alluri, V., Brattico, E., Lartillot, O., Tervaniemi, M., Huotilainen, M., 2016a. Event-Related Brain Responses While Listening to Entire Pieces of Music. Neuroscience 312, 58–73.

Poikonen, H., Toiviainen, P., Tervaniemi, M., 2016b. Early auditory processing in musicians and dancers during a contemporary dance piece. Sci. Rep. 6.

Rees, A., Green, G.G.R., Kay, R.H., 1986. Steady-State Evoked-Responses to Sinusoidally Amplitude-Modulated Sounds Recorded in Man. Hear. Res. 23, 123–133.

Rorden, C., Karnath, H.O., Bonilha, L., 2007. Improving lesion-symptom mapping. J. Cognit. Neurosci. 19, 1081–1088.

Serkov, F.N., Leovova, E.F., Shelest, I.I., 1969. Evoked potentials of the auditory cortex on paired stimuli. Neurophysiology. 1, 42–49.

Simon, J., Balla, V., Winkler, I., 2019. Temporal boundary of auditory event formation: An electrophysiological marker. Int. J. Psychophysiol. 140, 53–61.

Smith, L.S., Fraser, D.S., 2004. Robust sound onset detection using leaky integrate-and-fire neurons with depressing synapses. IEEE Trans. Neural Networks 15, 1125–1134.

Snyder, J.S., Alain, C., Picton, T.W., 2006. Effects of attention on neuroelectric correlates of auditory stream segregation. J. Cognit. Neurosci. 18, 1–13.

Sturm, I., Dahne, S., Blankertz, B., Curio, G., 2015. Multi-Variate EEG Analysis as a Novel Tool to Examine Brain Responses to Naturalistic Music Stimuli. PLoS ONE 10, e0141281.

Sussman, E., Steinschneider, M., Gumenyuk, V., Grushko, J., Lawson, K., 2008. The maturation of human evoked brain potentials to sounds presented at different stimulus rates. Hear. Res. 236, 61–79.

Tan, X.D., Yu, X.F., Lin, L., Wang, T., 2015. Simulation on the Comparison of Steady-State Responses Synthesized by Transient Templates Based on Superposition Hypothesis. Comput. Math. Methods Med.

Taulu, S., Hari, R., 2009. Removal of Magnetoencephalographic Artifacts With Temporal Signal-Space Separation: Demonstration With Single-Trial Auditory-Evoked Responses. Hum. Brain Mapp. 30, 1524–1534.

Thompson, R.F., Spencer, W.A., 1966. Habituation: a model phenomenon for the study of neuronal substrates of behavior. Psychol. Rev. 73, 16–43.

Thompson, R.F., Groves, P.M.T., T. J., Roemer, R.A., 1973. A dual-process theory of habituation: Theory and behavior. In Habituation. Vol. Vol. 1. Behavioral studies, H. V.S.P.M.J. Herz, ed.ˆeds. Academic Press, London, pp. 239-271.

Thoshkahna, B., Ramakrishnan, K., 2008. A psychoacoustics based sound onset detection algorithm for polyphonic audio. In 2008 9th International Conference on Signal Processing. Vol., ed.ˆeds. IEEE, pp. 1424-1427.

Todorovic, A., van Ede, F., Maris, E., de Lange, F.P., 2011. Prior Expectation Mediates Neural Adaptation to Repeated Sounds in the Auditory Cortex: An MEG Study. J. Neurosci. 31, 9118–9123.

Vuust, P., Ostergaard, L., Pallesen, K.J., Bailey, C., Roepstorff, A., 2009. Predictive coding of music - Brain responses to rhythmic incongruity. Cortex. 45, 80–92.

Wang, A.L., Mouraux, A., Liang, M., Iannetti, G.D., 2008. The Enhancement of the N1 Wave Elicited by Sensory Stimuli Presented at Very Short Inter-Stimulus Intervals Is a General Feature across Sensory Systems. PLoS ONE 3.

Winkler, I., Denham, S.L., Nelken, I., 2009. Modeling the auditory scene: predictive regularity representations and perceptual objects. Trends in Cognitive Sciences. 13, 532–540.

Woods, D.L., Elmasian, R., 1986. The Habituation of Event-Related Potentials to Speech Sounds and Tones. Electroencephalogr. Clin. Neurophysiol. 65, 447–459.

Wu, M.C., David, S.V., Gallant, J.L., 2006. Complete functional characterization of sensory neurons by system identification. Annu. Rev. Neurosci. 29, 477–505.

Zacharias, N., Konig, R., Heil, P., 2012. Stimulation-history effects on the M100 revealed by its differential dependence on the stimulus onset interval. Psychophysiology 49, 909–919.

Zhang, F.W., Hammer, T., Banks, H.L., Benson, C., Xiang, J., Fu, Q.J., 2011. Mismatch negativity and adaptation measures of the late auditory evoked potential in cochlear implant users. Hear. Res. 275, 17–29.