Early cortical processing of pitch height and the role of adaptation and musicality

Martin Andermann*, Melanie Günther, Roy D. Patterson, André Rupp

*Section of Biomagnetism, Department of Neurology, Heidelberg University Hospital, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany
Department of Physiology, Development and Neuroscience, University of Cambridge, Downing Street, Cambridge, CB2 3EG, United Kingdom

ARTICLE INFO

Keywords:
MEG
Auditory
Pitch
Music
Adaptation
Offset response

ABSTRACT

Pitch is an important perceptual feature; however, it is poorly understood how its cortical correlates are shaped by absolute vs relative fundamental frequency \( f_0 \), and by neural adaptation. In this study, we assessed transient and sustained auditory evoked fields (AEFs) at the onset, progression, and offset of short pitch height sequences, taking into account the listener’s musicality. We show that neuromagnetic activity reflects absolute \( f_0 \) at pitch onset and offset, and relative \( f_0 \) at transitions within pitch sequences; further, sequences with fixed \( f_0 \) lead to larger response suppression than sequences with variable \( f_0 \) contour, and to enhanced offset activity. Musical listeners exhibit stronger \( f_0 \)-related AEFs and larger differences between their responses to fixed vs variable sequences, both within sequences and at pitch offset. The results resemble prominent psychoacoustic phenomena in the perception of pitch contours; moreover, they suggest a strong influence of adaptive mechanisms on cortical pitch processing which, in turn, might be modulated by a listener’s musical expertise.

1. Introduction

Pitch is the perceptual attribute that allows us to order sounds from low to high (ANSI, 1994; low/high refers to the ‘height’ of the sound). Since the note names along the musical scale repeat once per octave, the perception of pitch is commonly represented by a helix with a circular, pitch-chroma dimension and a vertical, pitch-height dimension (see Shepard, 1982, or Warren et al., 2003). This paper reports a magnetoencephalography (MEG) experiment in which pitch chroma is fixed (i.e., we use the same note names at different octaves) and the focus is on the early cortical responses to four note, pitch-height sequences: the onset, sustained, and offset components, and any neural adaptation as the sequence proceeds. The analysis of the results also includes the effect of the listener’s individual musicality.

1.1. Cortical processing of absolute and relative pitch information

The first goal of this study was to assess how early cortical activity is shaped by absolute and relative pitch information in pitch-height sequences, i.e., by the absolute fundamental frequency \( f_0 \) of the current note, and by the interval (distance between notes) and contour (up/down pattern) information that describes the \( f_0 \) relation between successive notes in a sequence.

Neural activity in lateral Heschl’s gyrus (HG) reflects the extraction of pitch information from regular sounds (Patterson et al., 2002; Warren et al., 2003; Penagos et al., 2004; Bendor and Wang, 2006; Seither-Preisler et al., 2006; Schönwiesner and Zatorre, 2008; Andermann et al., 2011, 2014, 2017); more specifically, the latency of the transient response to pitch onset is linked to the absolute \( f_0 \) value (Krumholz et al., 2003; Ritter et al., 2005), and the subsequent sustained field (SF) mirrors the continuous perception of pitch associated with the ongoing sound (Gutschalk et al., 2002, 2004, 2007; Lütkenhöner and Patterson, 2015; Andermann et al., 2020). Beyond the level of single pitch sounds, however, cortical tracking of relative pitch has not received much attention in existing studies. Earlier works have postulated neural ‘frequency shift detectors’ (Demany and Semal, 2018); in line with this notion, Zhang et al. (2003) identified neurons in rat primary auditory cortex that responded preferably to upward or downward sweeps. Cortical processing of relative pitch in humans appears to involve primary auditory areas (Schindler et al., 2013); moreover, Trainor et al. (1999, 2002) demonstrated detection of deviant pitch interval and contour information in the mismatch negativity (MMN; May and Ttilinen, 2010; Paavilainen, 2013; Carbajal and Malmierca, 2018) and P300 components (Linden, 2005; Kujala and Näätänen, 2010) of the evoked potential. Yet, although two recent studies (Quiroga-Martínez et al., 2020a; Regev et al., 2020) report an amplitude increase of the N1 and P2 waves with growing interval size, it is not well understood how relative pitch is encoded in the early auditory evoked field (AEF).
In the current experiment, we studied the early cortical correlates of absolute and relative pitch processing, using short pitch height sequences with carefully balanced contours and intervals. The octave interval was chosen in an effort to vary pitch height independently while keeping pitch chroma constant (Warren et al., 2003). We hypothesized that the neuromagnetic responses to the first note of the pitch sequence would reflect its absolute $f_0$ value in their morphology (Krumbholz et al., 2003; Ritter et al., 2005); moreover, in light of the importance that pitch relations have for speech and music (Janata, 2015), we expected that the neural responses to transitions within sequences would mirror relative $f_0$ information, rather than absolute $f_0$ values.

1.2. Neural adaptation in response to pitch sequences

The second goal of this study was to measure the neural adaptation, i.e., response suppression (Grill-Spector et al., 2006), in the early cortical response to pitch-height sequences when sound probabilities are balanced.

Pitch perception is hierarchically organized, with bottom-up and top-down processes that involve primary and secondary auditory cortex (Patterson et al., 2002; Friston, 2008; Balaguer-Ballester et al., 2009; Kumar et al., 2011), and the processing extracts temporal regularities on multiple time scales (Ulanovsky et al., 2004; Costa-Faíndila et al., 2011). The results of the processing provide the basis for the prediction of how a sound pattern will continue (Winkler et al., 2009). Predictions may refer to different stages of pitch processing, from the note-to-note level (as in the present study) to complex musical structures (Salimpoor et al., 2015); it is argued that neural networks in the brain adapt to the statistical properties of sound sequences to maintain efficient coding (Nelken, 2004; Wark et al., 2007; Yaron et al., 2012; Pérez-González and Malmierca, 2014). For example, an early MEG study by Rupp and Uppenkamp (2005) showed that the amplitude of the N1 wave (Nättäinen and Picton, 1987) decreases much more in response to fixed pitch sequences than random sequences. This finding was later extended to the P2 component (Patterson et al., 2015, 2016), and Globerson et al. (2017) reported an N1 amplitude decrement when presented with a familiar melody as opposed to an unfamiliar melody. The specific influence of predictability on neural adaptation has been demonstrated by Todorovic et al. (2011) and Todorovic and de Lange (2012) who found that response suppression to repetitive sounds increases when stimulus repetition is expected.

Most studies on adaptive auditory processing have focused on the MMN or P300 elicited by rare violations of a repeating standard pattern (Hellbrun and Chait, 2018). Investigations of adaptation in the early transient and sustained AEF are comparably scarce, particularly when they involve experimental paradigms with equally distributed stimulus probabilities. To increase the explanatory power of our experiment, sequences with variable pitch contours ($f_0$ always changes between notes) and fixed pitch contours (no $f_0$ changes between notes) were presented equally often, and each $f_0$ value occurred equally often at each position within a sequence. As a result, the question as to whether the notes of a sequence are fixed or variable must await the presentation of the second note, after which the sequence type is completely determined. Accordingly, we hypothesized that early AEF responses to transitions within a sequence would be larger in variable sequences than in fixed sequences.

1.3. The role of musicality

The third goal of this study was to investigate the role of musicality in the early cortical processing of pitch-height sequences.

Predictions about the progress of a sequence rely on both statistical learning and prior knowledge (Huron, 2006; Pearce, 2018; Morgan et al., 2019). Thus, we might expect that a listener’s musicality might increase predictive precision (Vuust et al., 2018; see, however, Hansen et al., 2016, for the relationship between musical expertise and stylistic entropy). Studies have revealed enhanced MMN and P300 waves in musicians in response to melodic interval and contour deviants (Trainor et al., 1999; Fujikoa et al., 2004), and also to other musical features (Vuust et al., 2012; Tervaniemi et al., 2014; Quiroga-Martínez et al., 2020b); more generally, musical experts perform better in many pitch perception and melody perception tasks (Kishon-Rabin et al., 2001; Michely et al., 2006; Schubert and Stevens, 2006; Bailes, 2010; Strait et al., 2010; Brown et al., 2017). With respect to early neural activity, however, there is little evidence for musically-related differences in melody processing, although numerous experiments have demonstrated that transient waves are larger in musicians (for a review, see Sanju and Kumar, 2016). In particular, it is not clear how musical expertise shapes the early neural correlates of absolute vs relative pitch, and whether it affects their adaptation over the course of a pitch sequence. The current study is designed to determine whether musical experts would show larger transient responses at both pitch onset and subsequent transitions (Sanju and Kumar, 2016), and whether they would exhibit more response suppression.

1.4. Neuromagnetic activity following pitch offset

The fourth goal of the study was to investigate the cortical activity elicited by the offset of fixed and variable pitch-height sequences.

Generally, sound offset plays an important role in auditory scene analysis (Bregman, 1990); nevertheless, offset responses have not received much attention, despite the fact that they appear in 30 - 70% of neurons in auditory cortex (for a review, see Kopp-Schetzlplug et al., 2018). Onset and offset responses have similar cortical generators (Pantev et al., 1996), and similar sensitivity to signal level, signal-to-noise ratio, and inter-stimulus interval (Takahashi et al., 2004; Yamashiro et al., 2009, 2011; Baltzell and Billings, 2014). More specifically, there is evidence that the timing of offset activity reflects the cessation of temporal regularity in the stimulus (Seither-Preißler et al., 2006; Krishnan et al., 2014), making it a form of pitch-related information in the offset response. The design of the study allows us to determine whether the waveform morphology of transient offset responses reflects the absolute $f_0$ value of the preceding note and/or the nature of the preceding sequence (fixed or variable).

2. Material and methods

2.1. Participants

Twenty-four adult volunteers (14 females; 3 left-handed; mean age: 32.3 ± 10.5 years) participated in the study. None of the subjects reported a history of hearing impairment or neurological or psychiatric disorders; moreover, an audiometric screening confirmed that all subjects had less than 25 dB hearing loss at frequencies below 3 kHz. All subjects gave written informed consent prior to their participation in the study. The experimental procedures were conducted in accordance with the Declaration of Helsinki, and they were approved by the local ethics committee (Medical Faculty, University of Heidelberg, S-441/2016).

The individual musicality of the participants was assessed using the Advanced Measures of Music Audiation (AMMA; Gordon 1989, 1998). In this standardized test, the listeners compare 30 pairs of short melodies, and they are asked to indicate, for each pair, whether the two melodies are identical or whether they differ in melody or rhythm. Neither music training (instrumental or vocal) nor knowledge about musical theory are necessary to complete the test. A maximum of 80 points can be earned from the AMMA, with 40 points in the tonal and rhythm subtests, respectively; high scores are thought to reflect high musical aptitude. The mean overall AMMA score for the participants was 56.3 points (SD: 6.8). For all analyses involving musicality the median overall AMMA score was used to divide the sample into a low-AMMA group (≤ 55 points) and a high-AMMA group (> 55 points). There was no significant correlation between the AMMA score and listener age ($r = –0.23; p = 0.289$).
The participants also completed a brief questionnaire regarding their individual musical experience. There were six ‘true’ non-musicians among the subjects who had never received any music training. The remaining participants were amateur musicians; on average, they had started to play their instrument(s) at the age of 9.6 years (SD: 4.3), and the mean duration of their instrumental practice was 7.0 years (SD: 5.6). The AMMA scores of the listeners correlated strongly with the onset age (r = −0.53; p = 0.014∗) and the duration (r = 0.52; p = 0.016∗) of their music training.

2.2. Stimuli

The auditory stimuli for the MEG experiment were generated online in MATLAB R7.1 (The Mathworks, Inc., Natick, MA) at a sampling rate of 48,000 Hz. Each stimulus sequence consisted of an onset noise, four iterated rippled noise segments (IRN; Yost 1996), and an offset noise. IRN sounds allow for a reliable separation of energy and pitch-related neuromagnetic activity (Krumhholz et al., 2003; Ritter et al., 2005); they are created from a noise signal which is copied and iteratively added to itself with a fixed time delay. The reciprocal of the time delay corresponds to the f0 value of the stimulus. In the current study, we used IRN sounds with eight iterations, and with four different f0 values: 60 Hz, 120 Hz, 240 Hz, and 480 Hz. Each of the six stimulus segments (onset noise, four IRN segments, offset noise) had a duration of 600 ms and was bandpass filtered between 1 and 1920 Hz; further, all segments were ramped on and off with 15 ms Hanning windows, and the overall sound level was set to 70 dB SPL.

The single segments were then assembled into eight different sequences of which four were fixed pitch sequences and four were variable pitch sequences (cf. Fig. 1). The sequences were balanced such that each f0 value occurred equally often in the fixed and variable sequences, respectively, and equally at each position within a sequence; moreover, in the variable sequences, each transition between two different f0 values (−3 octaves, −2 octaves, −1 octaves, +1 octaves, +2 octaves, +3 octaves) occurred equally often. Each of the eight sequences was played 100 times, in pseudorandom order; the inter-stimulus interval between successive sequences varied randomly between 1200 and 1250 ms. The total duration of the stimulation was about 64 min.

Importantly, changes in pitch height are known to contribute to the auditory perception of source size information in natural sounds, together with spectral envelope features (Patterson et al., 2010). However, since the focus of our study was on the processing of pitch height, we did not manipulate the shape of the sound’s spectral envelope, nor did we touch its position along the frequency axis. As a result, the overall timbre of the IRN stimuli was relatively comparable across the four octaves. Moreover, informal listening indicated that the octave shifts in the variable timbre did not appear as changes in the size of the sound source; instead, it was the changes in pitch height that seemed to dominate listener’s perception of our sounds.

2.3. MEG recordings

The neuromagnetic field gradients in response to the auditory stimulation were recorded using a Neuromag-122 whole-head MEG system (Elekta Neuromag Oy, Helsinki, Finland; Ahonen et al., 1993) inside a shielded room (IMEDCO, Hägendorf, Switzerland), with a sampling rate of 1000 Hz and a lowpass filter at 330 Hz. Stimuli were presented via Etymotic Research (ER3) earphones with 90 cm plastic tubes and malleable foam earpieces, using a 24-bit sound card (RME ADI 8DS AD/DA interface), an attenuator (Tucker-Davis Technologies PA-5) and a head-phone buffer (Tucker Davis Technologies HB-7).

Before data acquisition, several head surface points were measured and applied as anatomical landmarks for the subsequent construction of a spherical head model. A Polhemus 3D-Space Isotrak2 system was used to determine the coordinates of two pre-auricular points, the nasion and 100 additional points across the scalp. During the MEG measurement, the participants watched a silent movie of their own choice in order to keep their vigilance stable, and they were instructed to direct their attention to the movie and to ignore the sounds in the earphones.

2.4. Data analysis

Analyses of the individual MEG data were conducted using the BESA 5.2 software (BESA GmbH, Gräfelfing, Germany). First, raw data were visually inspected and noisy channels were removed; then, segments with amplitudes >8000 ft/cm or gradients >800 ft/cm/ms were identified via the automatic artifact rejection tool in BESA and were excluded from further analyses. On average, 86.6% (SD: 11.6) of the segments remained after artifact rejection; these segments were then epoched individually for each participant.

In the next step, the prominent AEF responses at pitch onset (pon, first IRN segment), transitions between IRN segments (pitch change response, pcr, second to fourth IRN segments) and offset (poff, offset of the fourth IRN segment) were assessed by means of spatiotemporal source analysis (Scherg and von Cramon 1985a, b, 1986; Scherg, 1990; Scherg and Berg, 1991), with a spherical head model and a homogeneous volume conductor as implemented in BESA. In this source modeling approach, equivalent current dipoles represent the intracortical sources of the neural activity that is acquired at the scalp level, and their spatial orientation and location is iteratively varied until the source model explains maximum variance in the scalp data. The model output comprises the spatial information (dipole coordinates in Talairach space; Talairach and Tournoux, 1988) and the corresponding neuromagnetic activity across time (source waveform), separately for every current dipole. Source models were generated individually for each participant, based on bandpass filtered data, with one dipole per hemisphere, and separately for every AEF component outlined in Fig. 2 (see below). For each source model, segments of interest were pooled during the generation of the respective model, in an effort to achieve stable dipole fits. For the transient waves (P1, N1, P2), the fitting interval was centered about 30–50 ms around the respective peak; the sustained activity

| \( f_0 \) | fixed | variable |
|---|---|---|
| position # | 1 2 3 4 | position # | 1 2 3 4 |
| 480 Hz | | | |
| 240 Hz | | | |
| 120 Hz | | | |
| 60 Hz | | | |
| 480 Hz | | | |
| 240 Hz | | | |
| 120 Hz | | | |
| 60 Hz | | | |
(SF) was fitted in the interval from 400 to 600 ms within the respective stimulus segment(s). The $N_1_{\text{adapt}}$ and $SF_{\text{adapt}}$ models were based on conjoint dipole fits in the respective pon and pcr segments. Dipole fits were considered stable if they fulfilled all of the following criteria: 1) correct overall waveform polarity, 2) Talairach coordinates in the range of $|x| = 30–60$ and $y = −55–5$ for both dipoles, and 3) robustness of the model solution to small variations in the positioning and duration of the fitting interval. Where necessary, a symmetry constraint was introduced to further stabilize individual fits; no constraints were made regarding dipole orientation. Participants who did not show stable dipole fits in a given source model were excluded from further analyses regarding that model. Table 1 summarizes the relevant information corresponding to the source models in this study; specifically, it reports the bandpass filter settings, the stimulus segments that were pooled for dipole fitting, the number of participants for whom the dipole fit was successful, the number of participants in which a symmetry constraint was applied, and the variance that was explained by that model within the time range of the respective AEF component.

After dipole fitting, the source models were applied as spatiotemporal filters, and the source waveforms corresponding to each model were derived separately for each condition of interest. In the case of the $SF_{\text{adapt}}$ source model, a principal component analysis (PCA; Berg and Scherg, 1994) was calculated over the last 100 ms of the complete stimulus epoch in each condition to compensate for drift (cf. Gutschalk et al., 2002, 2004, 2007), and the PCA component that explained maximum variance in that condition was included in the model. The source waveforms were then transferred to MATLAB, and each source waveform was adjusted to the baseline calculated as the average of the last 100 ms before the onset of the respective stimulus segment. In MATLAB, the peak latencies and amplitudes of the transient AEF components (P1, N1, P2) were automatically identified within their respective time intervals using the min/max function, separately for each source model, participant, condition/segment, and dipole; the resulting values were double checked through visual inspection. The SF amplitude was determined, in the same manner, as the average activity in the last 200 ms of the respective stimulus segment(s).

Statistical evaluation of the source analytic results was done in SPSS 22.0 (IBM, Corp., Armonk, USA). We conducted repeated measures ANOVAs with main and interaction effects (first and second order interactions), including Mauchly tests for sphericity and adequate Greenhouse-Geisser corrections, separately for the source positions, amplitudes and latencies of the AEF responses outlined in Table 1. Regarding the statistical analyses of the source positions, ANOVAs were calculated for the Talairach coordinates of each source model and included the following factors: hemisphere (HEMI: left vs right dipole), musicality (AMMA: listeners with high vs low AMMA scores), and spatial di-
Table 2
Source waveform ANOVA information. The table presents, separately for each source model, the factors that were included in the repeated measures ANOVAs of the response amplitudes and latencies. HEMI, left vs right dipole; AMMA, listeners with high vs low AMMA scores; PITCH, $f_0$ value = 60, 120, 240, or 480 Hz; SEQ, fixed vs variable sequences; POS, 1st, 2nd, 3rd, or 4th IRN segment; SIZE, $f_0$ shift size of 1, 2, or 3 octaves relative to the previous sound; DIR, upward or downward $f_0$ direction relative to the previous sound.

| model     | HEMI | AMMA | PITCH | SEQ | POS | SIZE | DIR |
|-----------|------|------|-------|-----|-----|------|-----|
| $p_{on}$  | X    | X    | X     |     |     |      |     |
| $N_1p_{on}$ |    | X    | X     |     |     |      |     |
| $P_2p_{on}$ |    | X    | X     |     |     |      |     |
| $S_{Fp}$   |     | X    | X     |     |     |      |     |
| $p_{off}$  |     | X    | X     |     |     |      |     |
| $P_2p_{off}$ |   |     | X    |     |     |      |     |

mension (DIM: Talairach $[x, y, z]$). The factors HEMI and AMMA were also included in all analyses regarding AEF amplitudes and latencies; additionally, the following factors were studied with respect to the AEF responses: $f_0$ value (PITCH: 60, 120, 240, or 480 Hz), sequence type (SEQ: fixed or variable sequence), IRN segment position (POS: 1st, 2nd, 3rd, or 4th position), $f_0$ shift size (SIZE: 1 octave, 2 octaves, or 3 octaves relative to the previous sound), and $f_0$ direction (DIR: upward or downward relative to the previous sound). Table 2 presents an overview of the ANOVA factors that were involved at the different stages of the source waveform evaluation.

The major responses to the auditory stimulation were also studied at the sensor level, based on the temporal MEG gradiometers covering left and right auditory cortex (e.g., Park et al., 2018). After pre-processing of the raw data, the sensor waveforms were exported from BESA, unfiltered and separately for every participant and epoch; then, each MEG channel was linearly detrended in MATLAB, and an RMS curve was calculated for the aggregated activity in all of the selected channels. For the statistical evaluation of the major sensor level responses (repeated measures ANOVAs for peak latencies and amplitudes in the RMS curves), the factor HEMI was excluded. We present summary plots of the sensor waveforms in the supplemental material of this paper, together with the results from the respective statistical analyses.

3. Results

Fig. 2 presents an overview of the unfiltered MEG activity in response to the stimulus sequences (plotted in grey above the source waves); the source waves are based on the $S_{F_{adapt}}$ dipole model and pooled across hemispheres. At the transition from silence to noise, the onset of sound energy elicited a neuromagnetic wave complex with two positive deflections; then, the onset of pitch at the transition to the first IRN segment was followed by a prominent response with two transient components ($N_1p_{on}$, $P_2p_{on}$) and a subsequent $S_{F_{on}}$ that remained stable for the duration of the IRN segments. The second, third, and fourth IRN segments also evoked neuromagnetic responses ($P_1p_{cr}$, $N_1p_{cr}$, $P_2p_{cr}$) that differed strongly between fixed (red) and variable (blue) sequences, and between listeners with high (thin lines) and low (bold lines) AMMA scores. Following the transition from the last IRN segment to the offset noise, a bipolar response occurred ($N_1p_{off}$, $P_2p_{off}$) and the sustained activity returned to near baseline. Finally, the noise-silence transition at the end of the stimulus sequence elicited a transient negative deflection. The following sections present detailed analyses of all pitch-related AEF components outlined above.

3.1. Neuromagnetic responses to $f_0$ pitch onset

We begin with the source waves that arise after the transition from noise to the first IRN segment, i.e., at pitch onset (pon). The left panels of Fig. 3 show close-up views of the source waves from the $N_{1p_{on}}$, $P_{2p_{on}}$, and $S_{F_{on}}$ dipole models, separately for participants with high (thin lines) and low (bold lines) AMMA scores; the right panels of Fig. 3 depict 95% confidence intervals (CI) for the latencies and amplitudes of the respective components. As expected from previous studies (Krumholz et al., 2003; Ritter et al., 2005), the $f_0$ value of the first IRN segment strongly influenced the $N_{1p_{on}}$ morphology: higher $f_0$ values elicited $N_{1p_{on}}$ waves with much shorter latencies ($F(3,63) = 44.25$, $p < 0.001^{+++}$, $\eta^2 = 0.68$) and larger amplitudes ($F(3,63) = 39.50$, $p < 0.001^{+++}$, $\eta^2 = 0.65$). A similar pattern was observed in the $P_{2p_{on}}$ component (latency: $F(3,42) = 20.07$, $p < 0.001^{+++}$, $\eta^2 = 0.59$; amplitude: $F(3,42) = 25.90$, $p < 0.001^{+++}$, $\eta^2 = 0.65$); however, although no direct statistical comparison was made, the CI plots and Bonferroni-corrected posthoc tests (supplemental Table 1) suggest that the relation between $f_0$ and response morphology was clearer in the $N_{1p_{on}}$ than in the $P_{2p_{on}}$ wave. Importantly, there was also a AMMA main effect, with larger $N_{1p_{on}}$ amplitudes in listeners with high AMMA scores ($F(1,21) = 8.34$, $p = 0.009^{++}$, $\eta^2 = 0.28$); moreover, a PITCH*AMMA interaction indicated that the $f_0$-related differences in the $N_{1p_{on}}$ latency were more pronounced in high AMMA listeners ($F(3,63) = 3.83$, $p = 0.035^{*}$, $\eta^2 = 0.15$), but posthoc tests were not significant. There were no main or interaction effects with respect to the $S_{F_{on}}$ magnitude.

3.2. Response adaptation within the pitch sequence

Fig. 2 reveals pronounced differences between the neuromagnetic responses to IRN transitions in fixed and variable stimulus sequences (pcr): fixed sequences elicited only very small $P_{1p_{cr}}$, $N_{1p_{cr}}$, and $P_{2p_{cr}}$ components. This pattern was also visible when studying the data at the level of temporal MEG gradiometers (upper panel of Fig. 4). The sensor data indicate that the $N_{1p_{cr}}$ activity was much larger in variable sequences than in fixed sequences; moreover, the $P_{1p_{cr}}$ and $P_{2p_{cr}}$, deflections were more pronounced in variable sequences than in fixed sequences. Because response peaks could not be clearly identified for every listener in the fixed sequences, the related statistical evaluation (amplitude comparison in the response to fixed vs variable sequences) was done with a non-parametric bootstrapping technique (Efron and Tibshirani, 1993). This procedure confirmed that the differences between fixed and variable sequences were highly significant ($P_{1p_{cr}}$: $p < 0.001^{+++}$, $|d| = 1.62$;
compared factors, scores; develops analyses on consequence, N1 (bootstrap lines) and The M. (3,66) < 0.001 (latency: M. • Andermann, pcr fixed, the latency = 63.15, p < 0.001**, η² = 0.74) and position within the sequence (latency: F(3,66) = 14.16, p < 0.001**, η² = 0.39; amplitude: F(3,66) = 68.00, p < 0.001***, η² = 0.76), as well as interactions between both factors (latency: F(3,66) = 8.36, p = 0.001**, η² = 0.28; amplitude: F(3,66) = 53.36, p < 0.001***, η² = 0.71). This pattern indicates that compared with the N1pon at pitch onset, the N1_per responses to the second, third and fourth IRN segments were strongly attenuated in the fixed, but not in the variable sequences (supplemental Table 2). There were also differences with respect to the musicality of the participants: Generally, high AMMA listeners showed larger N1adapt responses than low AMMA listeners (F(1,22) = 7.94, p = 0.010*, η² = 0.27); moreover, significant interactions of AMMA with sequence type (F(1,22) = 12.53, p = 0.002**, η² = 0.36) and within-sequence position (F(1,22) = 12.53, p = 0.002**, η² = 0.36) as well as a second-order interaction between all three factors (F(3,66) = 9.11, p = 0.001**, η² = 0.29) revealed that the N1_per amplitude decrement in fixed relative to variable sequences was more pronounced in high AMMA listeners than in low AMMA listeners; this is also supported by the pattern of posthoc tests reported in Table 2 of the supplement. Together, these results demonstrate that a strong response suppression of the N1 response is present during fixed, but not variable sequences, and that this effect is larger in listeners with high musical aptitude.

Fig. 2 shows that the SF develops differently in listeners with high and low AMMA scores over the course of fixed vs variable sequences.
There was a second-order interaction between sequence type, position, and AMMA (F(3,45) = 3.47, p = 0.034*, n^2 = 0.19); posthoc tests indicated that only high AMMA listeners had significantly larger SFs in response to variable as compared to fixed sequences (supplemental Table 2). The interaction between sequence type and position was also significant across both groups (F(3,45) = 5.51, p = 0.005**, n^2 = 0.27), but posthoc tests shortly missed significance, perhaps due to high SF magnitude variability in the low AMMA listeners (lower panel of Fig. 4).

3.3. Neural correlates of absolute and relative pitch within the sequence

The left panels of Fig. 5 (absolute pcr) show close-up views of the source waves from the P1_pcr, N1_pcr, and P2_pcr dipole models, together with 95% CIs of the respective components, and separately for participants with high (thin lines) and low (bold lines) AMMA scores. Consistent with the results presented in the previous section, the N1_pcr peak was earlier (F(1,21) = 13.05, p = 0.002**, n^2 = 0.38) and larger (F(1,21) = 57.59, p < 0.001****, n^2 = 0.73) in variable than in fixed sequences, and it was larger in high AMMA listeners than in low AMMA listeners (F(1,21) = 6.41, p = 0.019*, n^2 = 0.23). Furthermore, the N1_pcr amplitude difference between fixed and variable sequences was larger in high AMMA listeners (F(1,21) = 10.87, p = 0.003**, n^2 = 0.34). There was also a pitch main effect in the N1_pcr amplitude (F(3,63) = 5.25, p = 0.012*, n^2 = 0.20), and these differences were more pronounced in variable sequences than in fixed sequences (F(3,63) = 22.15, p < 0.001****, n^2 = 0.51); importantly, however, there was no obvious relation between the N1_pcr morphology and the absolute f0 values of the respective IRN segments, in contrast to the N1_pron at pitch onset. With regard to N1_pcr latency, we also found a second-order interaction between AMMA, sequence type and pitch (F(3,63) = 3.76, p = 0.038*, n^2 = 0.15); posthoc tests revealed that the N1_pcr wave occurred later in fixed than in variable sequences, but only in high AMMA listeners (middle panel of Fig. 4 and supplemental Table 3). Further, there was an interaction of AMMA with hemisphere (F(1,21) = 6.14, p = 0.022*, n^2 = 0.23), but the corresponding posthoc tests were not significant.

Within the variable sequences, transitions between source sequences also evoked P1_pcr, and P2_pcr components (lower left panels of Fig. 5), and these deflections exhibited significant pitch main effects in their latencies and amplitudes (P1_pcr latency: F(3,51) = 20.50, p < 0.001****, n^2 = 0.55; P1_pcr amplitude: F(3,51) = 6.15, p = 0.004**, n^2 = 0.27; P2_pcr latency: F(3,57) = 15.21, p < 0.001****, n^2 = 0.45; P2_pcr amplitude: F(3,57) = 14.48, p < 0.001****, n^2 = 0.43). However, similar to the N1_pcr wave, there was no clear relation between the morphology of the P1_pcr and P2_pcr components and the absolute f0 values of the respective IRN segments. No other main or interaction effects occurred regarding the amplitudes and latencies of the P1_pcr, and P2_pcr waves, with the exception of a main effect of hemisphere indicating that the P2_pcr emerged with a slightly longer latency in the left compared to the right hemisphere (F(1,19) = 5.35, p = 0.032*, n^2 = 0.22).

The results described so far suggest that the link between the P1, N1, P2 components and the absolute f0 values of the IRN segments is more direct at pitch onset (pon) than at the transitions within variable pitch sequences (pcr). We therefore analyzed, in a second step, how the f0 relations (i.e., the size and direction of pitch shifts) between successive IRN segments were reflected in the transient AEF responses to f0 transitions in the variable sequences (i.e., IRN segments #2–4). The right panels of Fig. 5 (relative pcr) show close-up views of the source waves and 95% CIs from the P1_pcr, N1_pcr, and P2_pcr dipole models in response to transitions with different step sizes (one, two or three octaves) and directions (upward or downward shifts). The data show that the P1_pcr, N1_pcr, and P2_pcr source waves were influenced by the size and direction of the f0 shifts between successive IRN segments (supplemental Table 4). Compared with downward shifts, upward f0 transitions elicited earlier (F(1,17) = 27.39, p < 0.001****, n^2 = 0.62) and smaller (F(1,17) = 12.50, p = 0.003**, n^2 = 0.42) P1_pcr components, but larger N1_pcr amplitudes (F(1,20) = 6.82, p = 0.017*, n^2 = 0.25), and the P2_pcr latency was prolonged in response to upward transitions (F(1,19) = 39.28, p < 0.001****, n^2 = 0.67). Importantly, larger f0 shifts were associated with larger N1_pcr (F(2,40) = 36.46, p < 0.001****, n^2 = 0.65) and P2_pcr (F(2,38) = 87.86, p < 0.001****, n^2 = 0.82) deflections. The data also reveal an interesting interplay between the size and direction of f0 transitions and the temporal dynamics of the N1_pcr and P2_pcr waves: The shift-size related differences in the N1_pcr latency (F(2,40) = 28.30, p < 0.001****, n^2 = 0.59) were more pronounced in upward than in downward f0 shifts (F(2,40) = 26.41, p < 0.001****, n^2 = 0.57), and the P2_pcr latency difference between upward and downward shifts (F(2,38) = 12.61, p < 0.001****, n^2 = 0.40) was larger when the size of the f0 transition was small (F(2,38) = 4.92, p = 0.021*, n^2 = 0.21). There was also an interaction effect between shift direction and hemisphere with respect to the N1_pcr latency (F(1,20) = 4.67, p = 0.043*, n^2 = 0.19), but corresponding posthoc tests were not significant (supplemental Table 4). Together, the neuromagnetic responses to transitions within the variable sequences appear to reflect f0 relations, rather than absolute f0 values.

3.4. Neuromagnetic activity following pitch offset

Fig. 2 shows that two transient waves (N1_poff and P2_poff) occur in the AEF after pitch offset (poff), i.e., following the transition from the final IRN segment to the offset noise. We were interested in whether the morphology of these deflections would reflect the (absolute) f0 value of the final IRN segment and/or the type of the preceding sequence (fixed or variable). Fig. 6 presents the source waves and CI plots from the N1_poff and P2_poff dipole models, separately for epochs that were preceded by fixed (red lines) and variable (blue lines) sequences. The amplitudes of both the N1_poff (F(3,57) = 11.00, p < 0.001****, n^2 = 0.37) and the P2_poff (F(3,66) = 62.39, p < 0.001****, n^2 = 0.74) waves varied strongly with the absolute f0 value of the preceding IRN segment (supplemental Table 5): Higher f0 values were associated with larger amplitudes, and they elicited P2_poff deflections with shorter latencies (F(3,66) = 38.84, p < 0.001****, n^2 = 0.64); in addition, the amplitude of the P2_poff wave was larger (F(1,22) = 9.13, p = 0.006*, n^2 = 0.29) in high AMMA listeners (thin lines) than in low AMMA listeners (bold lines). This pattern closely resembles the results that were obtained for the N1_pwave at pitch onset. There were also differences regarding the type of the preceding stimulus sequence: The offset of fixed sequences led to larger N1_poff (F(1,19) = 33.17, p < 0.001****, n^2 = 0.64) and P2_poff (F(1,22) = 6.52, p = 0.018*, n^2 = 0.23) amplitudes, and to prolonged P2_poff latencies (F(1,22) = 19.86, p < 0.001****, n^2 = 0.47); moreover, the f0-related N1_poff amplitude (F(3,57) = 9.31, p < 0.001****, n^2 = 0.33) and P2_poff (F(3,66) = 14.97, p < 0.001****, n^2 = 0.41) amplitude differences were more pronounced following fixed sequences than after variable sequences. An interaction between AMMA and sequence type also indicated that low AMMA listeners had somewhat larger P2_poff latency differences in response to fixed versus variable sequences (F(1,22) = 7.53, p = 0.012*, n^2 = 0.26). Moreover, the P2_poff amplitude showed larger f0-related differences in the right than in the left hemisphere (F(3,66) = 3.15, p = 0.048*, n^2 = 0.13), and there was a second-order interaction between AMMA, pitch and hemisphere in the N1_poff amplitude (F(3,57) = 4.25, p = 0.034*, n^2 = 0.18), but posthoc tests were not significant (supplemental Table 5). In summary, neuromagnetic activity at pitch offset reflects both the absolute f0 value of the preceding segment and the history of a preceding fixed or variable stimulus sequence.

3.5. Source locations

The average locations of the dipole sources in the pon (N1_pon, P2_pon, SF_pon), pcr (P1_pcr, N1_pcr, P2_pcr), and poff (N1_poff, P2_poff) segments are shown in Fig. 7, projected onto axial auditory cortex maps (Leonard et al., 1998); the coordinates of the sources along the dimensions of the Talairach space (Talairach and Tournoux, 1988) are reported in supplemental Table 6. The transient AEFSs originated from
**Fig. 5.** Neuromagnetic activity in response to the second, third, and fourth IRN segments (pcr). The left panels (absolute pcr) show close-up views of the source waves and 95% CIs from the P1	extsubscript{pcr}, N1	extsubscript{pcr}, and P2	extsubscript{pcr} dipole models in the fixed (red lines) and variable (blue lines) stimulus sequences, separately for listeners with high (thin lines) and low (bold lines) AMMA scores. The right panels (relative pcr) shows the source waves and 95% CIs, separately for transitions with different step sizes (one, two or three octaves) and directions (upward or downward shifts). CIs are bias-corrected and accelerated as suggested by Efron and Tibshirani (1993).

anterolateral Heschl’s gyrus, a sub-region of auditory cortex that is involved in the processing of pitch information (Patterson et al., 2002; Krumbholz et al., 2003; Gutschalk et al., 2004; Ritter et al., 2005; Schönwiesner and Zatorre, 2008; Andermann et al., 2014, 2017, 2020); the sources of the sustained activity (SF	extsubscript{pon}) were located slightly medial and anterior. An interaction effect of AMMA with spatial dimension ($F(2,34) = 4.50, p = 0.020^*$, $\eta^2 = 0.21$) indicated that the P1	extsubscript{pcr} generators were located more anterior in high AMMA listeners than in low AMMA listeners. Further, there were interactions between spatial dimension and hemisphere in the N1	extsubscript{pon} ($F(2,42) = 3.62, p = 0.044^*$,
following, we will discuss how these results add to our understanding of cortical pitch processing, and the influence of neural adaptation and musical expertise thereon.

4.1. Absolute and relative $f_0$ information in cortical pitch processing

In our study, the absolute $f_0$ value of the first IRN segment strongly affected the N1$_\text{poff}$ morphology at pitch onset, in line with previous findings (Krumholz et al., 2003; Ritter et al., 2005). Importantly, however, the neural responses to transitions within variable sequences reflect $f_0$ relations, and this result nicely resembles a number of perceptual phenomena. When a melody is transposed to another key, humans can identify it without problems or prior experience; their ability to recognize a pitch sequence is not based on absolute values but relies on interval and contour information (Plantinga and Trainor, 2005; McDermott and Oxenham, 2008; Janata, 2015). Our observation that larger $f_0$ shifts evoke larger N1$_\text{poff}$ and P2$_\text{poff}$ waves is consistent with recent studies (Quiroga-Martínez et al., 2020a; Regev et al., 2020), and it parallels the common practice of composers to use wide intervals in order to increase intensity and attract attention (e. g., in the first movement of Mozart’s ‘Haffner’ symphony, KV 385, or in the well-known ‘Star Wars’ opening fanfare by John Williams; see also Vos and Troost, 1989); similarly, the fact that upward transitions elicit larger N1$_\text{poff}$ amplitudes and longer P2$_\text{poff}$ latencies is compatible with the psychoacoustic finding that the auditory system prefers upward over downward shifts (Schouten, 1985; Neuhoff and McBeath, 1996; Kishon-Rabin et al., 2004). There is some discrepancy between our results and the study of Bidelman (2015) who found no direction effects in the cortical pitch response; however, Bidelman’s experiment used $f_0$ glides that might have been too slow (200 ms) to affect the activity in the N1 time window, whereas our paradigm employed discrete $f_0$ steps that are more prevalent in musical melodies and are clearly reflected in the AEF morphology. In summary, our results provide evidence that early cortical AEFs reflect both absolute (onset) and relative (transitions) $f_0$ information, in a way that parallels important aspects of pitch perception.

4.2. Adaptation in cortical pitch processing

In line with previous findings (Rupp and Uppenkamp, 2005; Patterson et al., 2015, 2016), our data show that the early cortical AEFs (P1$_\text{poff}$, N1$_\text{poff}$ and P2$_\text{poff}$) decrease strongly over the course of fixed sequences, but not in variable sequences. Regarding the fixed sequences, these results may be explained by ‘classic’ repetition suppression (Grill-Spector et al., 2006). The situation is, however, more complex with respect to the variable sequences: here, no magnitude decrement was visible in the transient AEFs, not even in response to the third and fourth note that were perfectly predictable once the second note was played (at least at the level of the overall sequence type, i. e., fixed or variable). This pattern might perhaps be understood in terms of stimulus-specific adaptation (SSA), a basic aspect of deviance processing (Nelken, 2014; Pérez-González and Malmierca, 2014). SSA is a neural correlate of regularity-based change detection that is sensitive to fundamental sound features and operates on multiple time scales (Ulansky et al., 2003, 2004; Winkler et al., 2009; Grimm et al., 2016); the larger transients in variable sequences would then reflect temporal regularity changes during the sequence, no matter whether these are expected or not. On the other hand, neural adaptation might not necessarily mean attenuation; instead, it has been suggested that activity is up-regulated when stimuli are highly predictable, to focus perception on stable environmental features (Feldman and Friston, 2010). Predictable sound sequences would then be expected to evoke larger responses than random sequences (Heilbron and Chait, 2018), which contradicts our current results; however, other recent findings are well in line with this notion (Barascud et al., 2016; Southwell et al., 2017; Southwell and Chait, 2018).

4. Discussion

Four major findings arise from our study: 1) the early cortical AEFs reflect absolute $f_0$ information at pitch onset and relative $f_0$ information at transitions within pitch height sequences; 2) fixed pitch sequences lead to larger within-sequence response adaptation (i. e., decrease in subsequent AEF amplitudes) than variable sequences; 3) musical listeners show stronger $f_0$-related AEFs and larger response suppression to fixed vs variable sequences; 4) cortical offset responses reflect both absolute $f_0$ information and the history of a preceding fixed or variable sequence, in a way that is modulated by the listener’s musicality. In the
At this point, we should emphasize a specific feature of our study that assessed early transient (P1, N1, P2) and sustained (SF) AEsFs, in an experimental setup with equally distributed stimulus probabilities; in contrast, previous studies on SSA and predictive processing mainly focused on MMN or P300 responses elicited in oddball paradigms. Direct comparison with earlier findings is therefore challenging, and although neural adaptation clearly is at work in the current data, it appears difficult to specify the proportion by which the above-described mechanisms contribute to the responses. Recent studies (Quiroga-Martinez et al., 2020a; Hofmann-Shen et al., 2020; Regev et al., 2020) argue that the N1/P2 response decrement reflects early sensory adaptation of auditory neuron populations to low-level acoustic features (e.g., interval size), rather than higher order probabilistic predictions, and that it is tempting to interpret the results of our study along those lines; however, the respective experiments did not use sounds with a noisy spectrum like IRN, and they report activity in primary auditory cortex regions (Quiroga-Martinez et al., 2020a; Hofmann-Shen et al., 2020), whereas the responses in our study seem to arise from somewhat more anterolateral portions of HG (cf. the middle panel of Fig. 7). A promising way to tackle the issue in future studies could be to combine our paradigm with model simulations of neural adaptation (Hansen and Pearce, 2014; Kudela et al., 2018; Regev et al., 2020); further, adding a behavioral task might help to quantify effects of top-down expectations and attention. Interestingly, there were almost no lateralization effects in our data, which is in line with earlier studies on cortical melody processing (Fujikoa et al., 2004; Uppenkamp and Rupp, 2005; Schindler et al., 2013; Patterson et al., 2015; Allen et al., 2017); carefully speaking, this might also argue in favor of some earlier process involving symmetric activity in both hemispheres.

4.3. The role of musicality

Our data reveal some interesting effects that relate to the participants’ musicality. Generally, the major responses to pitch onset (N1pon) and transitions (N1pct) were enhanced in listeners who achieved high scores in a musicality test (the AMMA). This pattern is largely consistent with earlier findings on differences between musicians and nonmusicians in auditory cortex structure and function (Shahn et al., 2003; Bermudez et al., 2009; Itoh et al., 2012; Sanju and Kumar, 2016). In contrast, no AMMA-associated interactions occurred with respect to the neural correlates of f0 transitions, i.e., the size and direction of f0 transitions. This seems surprising in light of Cuddy’s and Cohen’s (1976) observation that musicians perform better than nonmusicians in perceptual tasks on relative pitch processing; however, such superior performance is also found in other psychoacoustic measures (Asaridou and McQueen, 2013; Coffey et al., 2017) and might point to some general advantage in these listeners.

Importantly, we also observed AMMA-related differences with respect to neural adaptation in pitch sequences: high AMMA listeners showed a much larger N1pct amplitude decrement than low AMMA listeners in response to fixed sequences, and they had larger SFs in variable than in fixed stimulus conditions. Stimulus statistics and prior knowledge both shape musical expectations (Huron, 2006; de Lange et al., 2018; Pearce, 2018; Morgan et al., 2019), and predictive precision has been found to modulate neural activity while listening to music (Quiroga-Martinez et al., 2019). In this vein, enhanced predictive precision in musical experts (Hansen and Pearce, 2014; Vuust et al., 2015) might well underlie their stronger N1pct response suppression during fixed sequences in our experiment. Regarding sustained cortical processing, our source level data indicate that the SF reflects the combined impact of pitch sequence features and the subject’s musicality. This is in line with previous MEG studies that revealed effects of stimulus importance and history on the SF (Gutschalk et al., 2007; Herdman, 2011; Okamoto et al., 2011); further, the SF magnitude is larger when the sound has a semantic meaning for the listener (Fan et al., 2017). Based on the idea that sequences with varying pitch have a higher relevance for musical experts, a similar mechanism might also apply to our results; moreover, this would explain the discrepancy between our current findings and the study of Patterson et al. (2016) who found no SF difference in fixed vs melodic pitch conditions, but did not consider musicality effects.

4.4. Cortical activity at pitch offset

Cortical offset responses have been rarely studied in auditory neuroscience. Our data show that absolute f0 information is present in the N1pon and P2pon waves at pitch offset, i.e., after the last IRN segment of the sequence: here, higher f0 values were associated with larger response magnitudes and (in the P2pon) shorter latencies. Importantly, the P2pon was enhanced in musical listeners; moreover, the fact that fixed sequences elicited larger N1pon and P2pon waves resembles earlier observations that the magnitude of the offset response increases with the duration of the preceding sound (Takahashi et al., 2004; Ross et al., 2009). Generally, the pattern of the offset AEsFs is very similar to the N1pon dynamics that mirror temporal regularity onset, and it seems plausible that the AEsFs following pitch offset reflect the cessation of regularity.
(Seither-Preisler et al., 2006; Krishnan et al., 2014). Regarding the low variability in the N1\textsubscript{poff} latency, Seither-Preisler et al. (2006) have argued that it might be easier for the brain to detect regularity offset than onset; however, our results suggest that it is, in fact, the P2\textsubscript{poff} component that might be viewed as an offset counterpart to the N1\textsubscript{pasp} wave at pitch onset. Future studies should test this assumption by assessing whether the N1\textsubscript{poff} and P2\textsubscript{poff} responses reflect the salience of the preceding pitch and how their waveform morphologies relate to computer simulations of auditory processing (Patterson et al., 1995; Tabas et al., 2019).

4.5. Strengths and limitations of the study

The current experiment incorporates some unique features that, to the best of our knowledge, have not been combined before when assessing cortical pitch processing. It is the first study that comprehensively investigates transient and sustained neural activity at the onset, progression, and offset of pitch sequences, using spatiotemporal source analysis and a carefully balanced stimulus paradigm with equally distributed probabilities that allows to distinguish absolute from relative pitch information. Moreover, when assessing the role of musical expertise, we did not compare extreme groups (i.e., professional musicians vs. nonmusicians); rather, we studied a broad variety of amateur musicians, and we applied a standardized musicality test that was based on melody comparisons and validated through self-reports of our listeners.

However, there are also some methodological issues that should be considered when drawing conclusions from our experiment. First, it is important to note that no natural sounds were used in the current study; instead, we employed IRN stimuli (Yost, 1996) that have been utilized in numerous experiments on pitch and melody processing (Patterson et al., 2002; Krumbholz et al., 2003; Ritter et al., 2005; Griffiths et al., 2010). IRN sounds allow for energy-balanced transitions between stimuli, but their noisy timbre probably precludes timbre-specific AEF enhancement in musical listeners (Pantev et al., 2001). Second, the f\textsubscript{0} shifts in our paradigm were based on octaves, in an effort to study pitch height separately from pitch chroma (Warren et al., 2003). The octave is clearly above perceptual threshold and it is of particular importance in various musical cultures (McDermott and Oxenham, 2008; Hoeschele et al., 2012). Yet, the effects of f\textsubscript{0} relations on the transient AEFs in our data seem robust and in good agreement with known realities, so one would expect them to be largely preserved even in smaller intervals.

A third aspect relates to the way in which the MEG data were analyzed in our experiment. Within our comprehensive approach, we wished to develop source models for all transient and sustained AEF components of interest, individually for each participant; this, however, was not successful in every listener, component and condition. Statistical evaluation was therefore performed based on varying sample sizes and separately for each model. Despite this fact, all source models were able to explain large portions of variance in the data (Table 1), and the statistical results comprise substantial effect sizes. Therefore, our findings provide a robust foundation for future studies that will continue to elucidate the neural basis of pitch processing.

5. Data and code availability statement

All raw data, trigger codes and dipole model outputs (source coordinates and source waves) are available in OSF (https://osf.io/c79xe/), together with an audio example of the stimulation.

Declarations of Competing Interest

None.

Credit authorship contribution statement

Martin Andermann: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing - original draft, Writing - review & editing.
Melanie Günther: Investigation, Methodology, Formal analysis, Writing - review & editing. Roy D. Patterson: Conceptualization, Writing - review & editing. André Rupp: Conceptualization, Formal analysis, Project administration, Resources, Supervision, Validation, Writing - review & editing.

Acknowledgments

The authors would like to express their gratitude to the Dietmar Hopp Foundation for kindly supporting their research (Grant 2301 1239; Recipient: A.R.); moreover, they would like to thank three anonymous reviewers for their helpful and encouraging comments on earlier versions of this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2020.117501.

References

Ahonen, A.I., Hämäläinen, M.S., Kajola, M.C., Knuutila, J.E.T., Laine, P.P., Lounasmaa, O.V., Parkkonen, L.T., Simola, J.T., Tesche, C.D., 1993. 122-channel SQUID instrument for investigating the magnetic signals from the human brain. Phys. Scr. T49, 198-205. doi:10.1088/0031-8949/1993/T49A/033.
Allen, E.J., Burton, P.C., Olfman, C.A., Oxenham, A.J., 2017. Representations of pitch and timbre variation in human auditory cortex. J. Neurosci. 37, 1284-1293. doi:10.1523/jneurosci.2236-16.2016.
Andermann, M., van Duyker, R., Patterson, R.D., Rupp, A., 2011. Neurametric representation of musical register information in human auditory cortex. Neuroimage 57, 1499-1506. doi:10.1016/j.neuroimage.2011.05.049.
Andermann, M., Patterson, R.D., Geldhausner, M., Sieroka, N., Rupp, A., 2014. Difftufts pitch: neurametric representation and auditory modeling. J. Neurophysiol. 112, 2616-2627. doi:10.1152/jn.00898.2013.
Andermann, M., Patterson, R.D., Vogt, C., Winterstetter, L., Rupp, A., 2017. Neurametric correlates of voice pitch, vowel type, and speaker size in auditory cortex. Neuroimage 158, 79-89. doi:10.1016/j.neuroimage.2017.06.065.
Andermann, M., Patterson, R.D., Rupp, A., 2020. Transient and sustained processing of musical consonance in auditory cortex and the effect of musicality. J. Neurophysiol. 123, 1320–1331. doi:10.1152/jn.00876.2018.
ANSI, 1994. American National Standard Acoustical Terminology (Acoustical Society of America).
Asaridou, S.S., McQueen, J.M., 2013. Speech and music shape the listening brain: evidence for shared domain-general mechanisms. Front. Psychol. 4, 321. doi:10.3389/fpsyg.2013.00321.
Bailes, F., 2010. Dynamic melody recognition: distinctiveness and the role of musical expertise. Mem. Cognit. 38, 641-650. doi:10.3758/MC.38.5.641.
Balguer, J.C., Bleiler, E., Clark, N.V., Goath, M., Krumbholz, K., Denham, S.L., 2009. Understanding pitch perception as a hierarchical process with top-down modulation. PLoS Comput. Biol. 5, e1000301. doi:10.1371/journal.pcbi.1000301.
Bahnell, L.S., Billings, C.J., 2014. Sensitivity of offset and onset cortical auditory evoked potentials to signals in noise. Clin. Neurophysiol. 125, 370-380. doi:10.1016/j.clinph.2013.08.003.2.
Barascud, N., Pearce, M.T., Griffiths, T.D., Friston, K.J., Chait, M., 2016. Brain responses in humans reveal ideal observer-like sensitivity to complex acoustic patterns. Proc. Natl. Acad. Sci. USA 113, 8616-8625. doi:10.1073/pnas.1508523113.
Bendor, D., Wang, X., 2006. Cortical representations of pitch in monkeys and humans. Curr. Opin. Neurobiol. 16, 391-399. doi:10.1016/j.conb.2006.07.001.
Berg, F., Scherg, M., 1994. A multiple source approach to the correction of eye artifacts. Electroencephalogr. Clin. Neurophysiol. 90, 229-241. doi:10.1016/0013-4694(94)90094-9.
Bermudez, P., Lerch, J.P., Evans, A.C., Zatorre, R.J., 2009. Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. Cereb. Cortex 19, 1583-1596. doi:10.1093/cercor/bhn196.
Bidelman, G.M., 2015. Sensitivity of the cortical pitch onset response to height, time-variance, and directionality of dynamic pitch. Neurosci. Lett. 603, 89–93. doi:10.1016/j.neulet.2015.07.018.
Bregman, A.S., 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. MIT Press.
Brown, C.J., Jeon, E.-K., Dritschill, V., Mussoi, B., Deshpande, S.B., Gfeller, K., Abbas, P.J., 2017. Effects of long-term musical training on cortical auditory evoked potentials. Ear Hear. 38, e74-e84. doi:10.1097/AUD.0000000000000375.
Carbajal, G.V., Malmierca, M.S., 2018. The neuronal basis of predictive coding along the auditory pathway: from the subcortical roots to cortical deviance detection. Trends Hear. 22, 1–33. doi:10.1177/2331325618784922.
Coffey, E.B., J., Chepuesik, A.M.P., Herholz, S.C., Baillet, S., Zatorre, R.J., 2017. Neural correlates of early sound encoding and their relationship to speech-in-noise perception. Front. Neurosci. 11, 479. doi:10.3389/fnins.2017.00479.
Kujala, T., Nätänäinen, R. 2010. The adaptive brain: a neurophysiological perspective. Proc. Neurobiol. 91, 55–67. doi:10.1016/j.neurobiol.2010.01.006.

Kumar, S., Sedlery, W., Nourski, K.V., Kawahsi, H., Oya, H., Patterson, R.D., Howard, M.A., Friston, K.J., Griffiths, T.D. 2010. Predictive coding and pitch processing in the auditory cortex. J. Cogn. Neurosci. 23, 3084–3094. doi:10.1162/jocn.2010.00021.

Leonard, C.M., Puranik, C., Kuldau, J.M., Lombardino, L.J. 1998. Normal variation in the frequency and location of human auditory cortex landmarks. Heschl’s gyrus: where is it? Cereb. Cortex 8, 397–401. doi:10.1093/cercor/8.4.397.

Linden, D.E.J. 2005. The P300: where in the brain is it produced and what does it tell us? Neuroscientist 11, 563–576. doi:10.1016/j.neuroimage.2004.05.024.

Lüütkenhoven, B., Patterson, R.D. 2015. Disruption of the auditory response to a regular versus an irregular pitch train by EcoG. Exp. Brain Res. 233, 1875–1892. doi:10.1007/s00221-015-4260-6.

May, P.J.C., Tiitinen, H. 2010. Mismatch negativity (MMN), the deviance-elicited auditory deflection, explained. Exp. Brain Res. 47, 66–122. doi:10.1111/j.1469-8686.2009.01856.x.

McDermott, J.H., Oxenham, A.J. 2008. Music perception, pitch, and the auditory system. Curr. Opin. Neurobiol. 18, 452–463. doi:10.1016/j.conb.2008.09.005.

Micheyl, C., Delhommeau, K., Perrot, X., Oxenham, A.J. 2006. Influence of musical and psychoacoustical training on pitch discrimination. Hear. Res. 219, 36–47. doi:10.1016/j.heares.2005.06.004.

Morgan, E., Fogel, A., Nair, A., Patel, A.D. 2019. Statistical learning and gestalt-like principles predict melodic expectations. Cognition 189, 23–34. doi:10.1016/j.cognition.2018.12.015.

Nätänäinen, 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. doi:10.1111/j.1469-8986.1987.tb02171.x.

Nelken, I., 2004. Processing of complex stimuli and natural scenes in the auditory cortex. Curr. Opin. Neurobiol. 14, 474–480. doi:10.1016/j.conb.2004.06.005.

Nelken, I., 2014. Stimulus-specific adaptation and deviance detection in the auditory system: experiments and models. Biol. Cybern. 108, 655–663. doi:10.1007/s00422-014-0643-9.

Neuhoff, J.G., McBeath, M.K. 1996. The Doppler illusion: the influence of dynamic intensity change on perceived pitch. J. Exp. Psychol. Hum. 22, 970–985. doi:10.1037/0096-1523.22.4.970.

Okamoto, H., Stracke, H., Bermudez, P., Pantev, C., 2011. Sound processing hierarchy within human auditory cortex. J. Cogn. Neurosci. 23, 1855–1863. doi:10.1162/jocn.2011.201251.

Pavilides, P., 2013. The mismatch negativity (MMN) component of the auditory event-related potential to violations of abstract regularities: a review. Int. J. Psychophysiol. 88, 109–123. doi:10.1016/j.ijpsycho.2013.03.015.

Patterson, R.D., Allerhand, M.N. 2005. Neural mechanisms for speech and pitch processing in the anterior superior temporal gyrus. J. Neurophysiol. 93, 378–395. doi:10.1152/jn.00568.2004.

Park, J.M., Chung, C.K., Kim, J.S., Lee, K.M., Seo, J., S.W., YI, 2018. Musical expectations enhance auditory cortical processing in musicians: a magnetoencephalography study. Neuroscience 369, 325–335. doi:10.1016/j.neuroscience.2017.11.038.

Patterson, R.D., Allerhand, M.N. 2005. Time domain modeling of peripheral auditory processing: a modular architecture and a software platform. J. Acoust. Soc. Am. 98, 1890–1894. doi:10.1121/1.141456.

Patterson, R.D., Upperkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of tone sequences and pitch movement in audition in auditory cortex. Neuron 36, 767–776. doi:10.1016/S0896-6273(02)00670-X.

Patterson, R.D., Gaudrain, E., Walters, T.C. 2010. The perception of family and register in musical tonal. In: Jones, M.R., Fay, R.R., Popper, A.N. (Eds.), Springer Handbook of Auditory Research, Vol. 36. Music Perception. Springer Science + Business Media, pp. 13–50.

Patterson, R.D., Andermann, M., Bussel, M.J. 1995. Time domain modeling of peripheral auditory processing: a modular architecture and a software platform. J. Acoust. Soc. Am. 98, 1890–1894. doi:10.1121/1.141456.

Patterson, R.D., Upperkamp, S., Johnsrude, I.S., Griffiths, T.D., 2002. The processing of tone sequences and pitch movement in audition in auditory cortex. Neuron 36, 767–776. doi:10.1016/S0896-6273(02)00670-X.

Patterson, R.D., Gaudrain, E., Walters, T.C. 2010. The perception of family and register in musical tonal. In: Jones, M.R., Fay, R.R., Popper, A.N. (Eds.), Springer Handbook of Auditory Research, Vol. 36. Music Perception. Springer Science + Business Media, pp. 13–50.

Patterson, R.D., Andermann, M., Bussel, M.J. 1995. Time domain modeling of peripheral auditory processing: a modular architecture and a software platform. J. Acoust. Soc. Am. 98, 1890–1894. doi:10.1121/1.141456.
Quiroga-Martínez, D.R., Hansen, N.C., Hojlund, A., Pearce, M.T., Brattico, E., Vuust, P., 2020b. Musical prediction error responses similarly reduced by predictive uncertainty in musicians and non-musicians. Eur. J. Neurosci. 51, 2250–2269. doi:10.1111/jen.14667.

Reger, T.L., Markusfeld, G., Deouell, L.Y., Nelken, I., 2020. Context sensitivity across multiple time and frequency scales. bioRxiv (preprint) doi:10.1101/2020.06.08.141044.

Ritter, S., Donch, H.G., Specht, H.-J., Rupp, A., 2005. Neuromagnetic responses reflect the temporal pitch change of regular interval sounds. NeuroImage 27, 533–543. doi:10.1016/j.neuroimage.2005.05.003.

Ross, B., Snyder, J.S., Aalto, M., McDonald, K.L., Dyson, B.J., Schneider, H., Alain, C., 2009. Neural encoding of sound duration persists in older adults. NeuroImage 47, 678–687. doi:10.1016/j.neuroimage.2009.04.051.

Rupp, A., Uppenkamp, S., 2005. Neuromagnetic representation of short melodies in the auditory cortex. In: Fastl, H., Frahm, M. (Eds.), Proceedings of the Fortschritte der Akustik-DAGA 2005. DEGA e.V., Berlin, pp. 473–474.

Salimpoor, V.N., Zald, D.H., Zatorre, R.J., Dagher, A., McIntosh, A.R., 2015. Predictions and the brain: how musical sounds become rewarding. Trends Cognit. Sci. 19, 86–91. doi:10.1016/j.tics.2014.12.001.

Sanju, H.K., Kumar, P., 2016. Enhanced auditory evoked potentials in musicians: a review of recent findings. J. Otol. 11, 63–72. doi:10.1016/j.joto.2016.04.002.

Scherg, M., von Cramon, D., 1985a. Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalogr. Clin. Neurophysiol. 62, 32–44. doi:10.1016/0168-5597(85)90033-4.

Scherg, M., von Cramon, D., 1985b. A new interpretation of the generators of the BAEP waves I–V: results of a spatio-temporal dipole model. Electroencephalogr. Clin. Neurophysiol. 62, 290–299. doi:10.1016/0168-5597(85)90006-1.

Scherg, M., von Cramon, D., 1986. Evoked dipole source potentials of the human auditory cortex. Electroencephalogr. Clin. Neurophysiol. 65, 344–360. doi:10.1016/0168-5597(86)90014-6.

Scherg, M., 1990. Fundamentals of dipole source potential analysis. Adv. Audiol. 6, 40–69.

Scherg, M., Berg, P., 1991. Use of prior knowledge in brain electromagnetic source analysis. Brain Topogr. 4, 143–150. doi:10.1007/bf01127711.

Schindler, A., Herder, M., Bartels, A., 2013. Coding of melodic gestalt in human auditory cortex. Cereb. Cortex 23, 2987–2993. doi:10.1093/cercor/bhs289.

Schönwiesner, M., Zatorre, R.J., 2008. Depth electrode recordings show double dissociation between pitch processing in lateral Heschl’s gyrus and sound onset processing in medial Heschl’s gyrus. Exp. Brain Res. 187, 97–105. doi:10.1007/s00221-008-1286-z.

Schouen, M.E.H., 1985. Identification and discrimination of sweep tones. Percept. Psychophys. 37, 369–376. doi:10.3758/bf03211361.

Schuh, E., Stevens, C., 2006. The effect of implied harmony, contour and musical expertise on judgments of similarity of familiar melodies. J. New. Music Res. 35, 161–174. doi:10.1080/09298220600835500.

Seith-Pfeiderer, A., Patterson, R.D., Krumholtz, K., Seither, S., Lütkenhöner, B., 2006. From noise to pitch: transient and sustained responses of the auditory evoked field. Hear. Res. 218, 50–63. doi:10.1016/j.heares.2006.04.005.

Shein, A., Bosnyak, D.J., Trainor, L.J., Roberts, L.E., 2003. Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. J. Neurosci. 23, 5545–5555. doi:10.1523/jneurosci.23-13-2003.5545.2003.

Shepard, R.N., 1982. Geometrical approximations to the structure of musical pitch. Psychol. Rev. 89, 305–333 https://psychnet.apa.org/about/id/10033-295x.89.4.305.

Southwell, R., Baumann, A., Gal, C., Barascud, N., Friston, K., Chait, M., 2017. Is predictability salient? A study of attentional capture by auditory patterns. Philos. Trans. R. Soc. Lond. B Biol. Sci. 372, 20160105–20160111. doi:10.1098/rstb.2016.0105.

Southwell, R., Chait, M., 2018. Enhanced deviant responses in patterned relative to random sound sequences. Cortex 109, 92–103. doi:10.1016/j.cortex.2018.08.032.2.

Strait, D.L., Kraus, N., Parbery-Clark, A., Ashley, R., 2010. Musical experience shapes top-down auditory mechanisms: evidence from masking and auditory attention performance. Hear Res. 261, 22–29. doi:10.1016/j.heares.2009.12.021.

Tabas, A., Andermann, M., Schubert, V., Ehrle, H., Balaguer-Ballester, E., Rupp, A., 2019. Modeling and MEG evidence of early consonance processing in auditory cortex. PLoS Comput. Biol. 15, e1006820. doi:10.1371/journal.pcbi.1006820.