eNeuro

Research Article: New Research* | *Cognition and Behavior

Electrophysiological correlates of absolute pitch in a passive auditory oddball paradigm: a direct replication attempt

Marielle Greber[1], Lars Rogenmoser[2], Stefan Elmer[1] and Lutz Jäncke[1,3,4]

[1]*Division Neuropsychology, Department of Psychology, University of Zurich, Zurich, Switzerland*

[2]*Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, Washington, DC, USA*

[3]*University Research Priority Program (URPP), Dynamics of Healthy Aging, University of Zurich, Zurich, Switzerland*

[4]*Department of Special Education, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia*

https://doi.org/10.1523/ENEURO.0333-18.2018

Received: 27 August 2018

Revised: 2 November 2018

Accepted: 22 November 2018

Published: 7 December 2018

Author Contributions: L.J., L.R., S.E., and M.G. designed research; M.G. performed research; M.G. analyzed data; M.G., L.J., L.R., and S.E. wrote the paper.

Funding: Swiss National Science Foundation
320030_163149

Conflict of Interest:

Funding sources: This work was funded by the Swiss National Science Foundation (SNSF), grant number 320030_163149 to LJ.

Correspondence should be addressed to Marielle Greber, marielle.greber@uzh.ch and Lutz Jäncke, lutz.jaencke@uzh.ch

Cite as: eNeuro 2018; 10.1523/ENEURO.0333-18.2018

Title Page

1. Manuscript Title (50 word maximum):

Electrophysiological correlates of absolute pitch in a passive auditory oddball paradigm: a direct replication attempt

2. Abbreviated Title (50 character maximum):

Passive listening in absolute pitch: a replication

3. List all Author Names and Affiliations in order as they would appear in the published article:

Marielle Greber [1]
Lars Rogenmoser [2]
Stefan Elmer [1]
Lutz Jäncke [1, 3, 4]

[1] Division Neuropsychology, Department of Psychology, University of Zurich, Zurich, Switzerland
[2] Laboratory of Integrative Neuroscience and Cognition, Department of Neuroscience, Georgetown University Medical Center, Washington, DC, USA
[3] University Research Priority Program (URPP), Dynamics of Healthy Aging, University of Zurich, Zurich, Switzerland
[4] Department of Special Education, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

4. Author Contributions:

L.J., L.R., S.E., and M.G. designed research; M.G. performed research; M.G. analyzed data; M.G., L.J., L.R., and S.E. wrote the paper.

5. Correspondence should be addressed to either of the following:

Marielle Greber
Binzmühlestrasse 14, Box 25
CH-8050 Zürich
Switzerland
marielle.greber@uzh.ch

Lutz Jäncke
Binzmühlestrasse 14, Box 25
CH-8050 Zürich

Switzerland
lutz.jaencke@uzh.ch

6. Number of Figures: 5
7. Number of Tables: 2
8. Number of Multimedia: 0
9. Number of words for Abstract: 248
10. Number of words for Significance Statement: 120
11. Number of words for Introduction: 745
12. Number of words for Discussion: 2865

13. Acknowledgements:

We wish to thank our research interns Anna Speckert, Chantal Oderbolz, Fabian Demuth, Florence Bernays, Joëlle Albrecht, Kathrin Baur, Laura Keller, Melek Haçan, Nicole Hedinger, Pascal Misala, Petra Meier, Sarah Appenzeller, Tenzin Dotschung, Valerie Hungerbühler, Vanessa Vallesi, and Vivienne Kunz for their invaluable assistance with data collection. Furthermore, we thank Simon Leipold for his important suggestions to improve an earlier draft of this manuscript, Christian Brauchli and Anja Burkhard for their contributions within the larger project on absolute pitch, Silvano Sele for his helpful inputs on the Bayesian analyses, and Isabel Hotz for her assistance with the figures. We would also like to thank Carina Klein and all members of the Auditory Research Group Zurich (ARGZ) for the productive discussions and useful feedback.

14. Conflict of Interest:

The authors declare no conflicts of interest.

15. Funding sources

This work was funded by the Swiss National Science Foundation (SNSF), grant number 320030_163149 to LJ.

Abstract

Humans with absolute pitch (AP) are able to effortlessly name the pitch class of a sound without an external reference. The association of labels with pitches cannot be entirely suppressed even if it interferes with task demands. This suggests a high level of automaticity of pitch labeling in AP. The automatic nature of AP was further investigated in a study by Rogenmoser et al. (2015). Using a passive auditory oddball paradigm in combination with electroencephalography, they observed electrophysiological differences between musicians with and without AP in response to piano tones. Specifically, the AP musicians showed a smaller P3a, an event-related potential (ERP) component presumably reflecting early attentional processes. In contrast, they did not find group differences in the mismatch negativity (MMN), an ERP component associated with auditory memory processes. They concluded that early cognitive processes are facilitated in AP during passive listening and are more important for AP than the preceding sensory processes.

In our direct replication study on a larger sample of musicians with (n = 54, 27 females, 27 males) and without (n = 50, 24 females, 26 males) AP, we successfully replicated the non-significant effects of AP on the MMN. However, we could not replicate the significant effects for the P3a. Additional Bayes factor analyses revealed moderate to strong evidence (Bayes factor > 3) for the null hypothesis for both MMN and P3a. Therefore, the results of this replication study do not support the postulated importance of cognitive facilitation in AP during passive tone listening.

Significance Statement

A better understanding of the neural basis of absolute pitch (AP), the ability to identify a pitch without an external reference, provides valuable insights to the mechanisms of pitch processing in the human brain. Since only a tiny fraction of the population possesses AP, most previous neuroscientific research had small sample sizes. In our direct replication, we used a large sample of musicians (n = 104) with and without AP to confirm an intriguing finding showing that AP musicians process tones more efficiently even when not actively attending them. Using both frequentist and Bayesian analyses, we failed to replicate this effect with an identical experimental setting. This finding highlights the significance of replications and the need for large sample sizes.

Introduction

Replications are an integral part of science. They can help estimate the size of an effect, identify the specific conditions under which it occurs, and — when successful — increase confidence in a scientific claim (Brandt et al., 2014; Nosek et al., 2012). In recent years, the low replicability of published research has become an increasing concern within neuroscience and science in general (Baker, 2016). Possible explanations for the observed low replicability include publication bias, flexibility in data analysis, and low statistical power (Munafò et al., 2017). Due to the resource-intensive data acquisition, many neuroscientific studies use small sample sizes, resulting in low power (Szucs and Ioannidis, 2017). Low power can compromise the conclusions of a study by reducing the probability of detecting a true effect, by increasing the probability that a significant finding does not reflect a true effect, and by overestimating the size of an effect (Button et al., 2013).

Acquiring data from a large sample is even more challenging for studies investigating special populations like individuals with absolute pitch (AP), the rare ability to label the pitch class (chroma) of a sound without an external reference (Levitin and Rogers, 2005; Takeuchi and Hulse, 1993; Zatorre, 2003). AP is often contrasted with relative pitch (RP), the more common ability to identify the musical interval (pitch distance) between two tones (McDermott and Oxenham, 2008). Despite its rarity, AP has received considerable scientific attention, partly because it might help understand different modes of perceptual processing and general aspects of pitch memory (Levitin and Rogers, 2005).

The neural and cognitive mechanisms underlying AP are not yet fully understood, but several studies have demonstrated that the labeling process in AP is at least in part automatic and not suppressible, even if it is disadvantageous for the task at hand (Itoh et al., 2005; Miyazaki and Rakowski, 2002; Schulze et al., 2013). The extent of this automaticity was further investigated by studies recording the electroencephalogram (EEG) during passive listening (Elmer et al., 2013; Matsuda et al., 2013; Rogenmoser et al., 2015; Tervaniemi et al., 1993). Using this approach, one can study the neurophysiological correlates of the automatic labeling process with high temporal resolution while minimizing the influence of top-down processes.

An often-used paradigm is the passive auditory oddball, in which one tone (standard) is presented more frequently than the other tones. The infrequent tones (deviants) are known to reliably elicit two frontal event-related potential (ERP) components: the mismatch negativity (MMN) and the P3a. Both ERP components are usually assessed by subtracting the standard ERP from the deviant ERP. The MMN is a negative deflection on this difference wave that peaks around 100-250 ms after stimulus onset and possibly reflects an automatic memory-based detection of change or rule violation (Garrido et al., 2009; Näätänen et al., 2011; Picton et al., 2000). While the MMN is thought to represent pre-attentive processing, the subsequently occurring positive deflection P3a has been linked to involuntary attention shifts towards unattended stimuli (Escera et al., 1998; Friedman et al., 2001; Kujala et al., 2007; Polich, 2007).

Rogenmoser et al. (2015) were the first to analyze both MMN and P3a in AP, which allowed them to study the influence of the sensory and the early cognitive processes reflected by these ERP components. They recorded EEG from 16 AP musicians and 10 non-AP musicians during a passive auditory oddball paradigm. The analysis of the MMN did not reveal any significant group differences, but AP musicians showed smaller P3a amplitudes than non-AP musicians when the deviations were larger than one semitone. The authors concluded that early cognitive processes are more efficient in AP during passive listening, whereas pre-attentive auditory processing contributes less to AP. This is in accordance with theoretical perspectives describing AP as a mainly cognitive ability (Levitin and Rogers, 2005; Zatorre, 2003).

Within small research fields like AP research, every single study has a high impact on the development of theoretical models. At the same time, the sample sizes are often small, which increases the need for replications. Rogenmoser et al. (2015) showed that AP musicians process tones differently even when not actively attending them. The extent of automaticity implied by this is both interesting and surprising. The aim of the present study was to confirm this finding in an independent and larger sample (n = 104). We attempted a direct replication, using the same stimuli, measures, and statistical analyses as in the original study. In addition, we calculated Bayes factors to quantify the success of the replication.

Material and Methods

Participants

The current study was carried out as part of a broader research project on AP, involving multiple experiments using different imaging modalities (magnetic resonance imaging [MRI] and EEG). Fifty-four self-reported AP possessors and 50 self-reported non-AP possessors between the age of 18 and 44 years were recruited for the current study.

All participants were professional musicians, music students, or highly-trained amateur musicians and received payment for their participation. The research protocol was approved by the local ethics committee in accordance with the Declaration of Helsinki, and all participants provided written informed consent.

None of the participants reported any past or present severe neurological, psychiatric, or audiological disorders. Normal hearing was confirmed by pure-tone audiometry in all participants (MAICO ST 20, MAICO Diagnostic, GmBh, Berlin). The two groups were matched for sex, age, handedness, age of onset of musical training, and cumulative training hours over the lifespan. Handedness was assessed by self-report and validated by the Annett Handedness Questionnaire (Annett, 1970). To control for possible between-group differences in intelligence, the Mehrfachwahl-Wortschatz-Intelligenztest (MWT-B; Lehrl, 2005) was administered. The MWT-B quantifies verbal intelligence and was shown to be a good predictor of global IQ (Lehrl et al., 1995). The musical aptitudes of the participants were assessed based on the total scores in the Advanced Measures of Music Audiation (AMMA; Gordon, 1989). To estimate musical experience in terms of age of onset of musical training and number of training hours, participants filled out an online questionnaire before taking part in the experiment. Demographical information and information on musical experience are given in Table 1.

Pitch-Labeling Test

Pitch-labeling ability was estimated using a web-based behavioral test (adapted from Oechslin et al., 2010), in which participants had to identify the pitch class and pitch height of 108 pure tones. The tones ranged from C3 to B5 (tuning: A4 = 440 Hz), lasted 500 ms, and were each presented three times in a pseudorandomized order with no tones repeated immediately in successive trials. In each trial, 2000 ms of Brownian noise were presented immediately before and after the pure tone. Answers were given by clicking on one label out of a list of all 36 possible labels (C3 to B5). Trials lasted 15000 ms but could be terminated early by clicking on a “next” button. Pitch-labeling ability was determined by the relative frequency of correctly identified tones in terms of pitch chroma and irrespective of octave errors (Deutsch, 2013; Miyazaki, 1989, 1988; Takeuchi and Hulse, 1993).

Stimulus Material and Experimental Procedure

Since the current study was a direct replication, we followed the experimental procedure of the original study as closely as possible. The stimulus material and the code for stimulus presentation were identical to those used in the original study. The auditory stimuli consisted of five piano tones with different fundamental frequencies. Three of the tones were in tune (C4 = 264 Hz, A4 = 440 Hz, A♭4/ G#4 = 416 Hz) and two of the tones were mistuned (1/4-semitone deviation of A♭4/ G#4 = 422 Hz, 1/10-semitone deviation of A4 = 438 Hz). All piano tones were recorded as 16-bit stereo files and had a duration of 200 ms with 5 ms rise and fall time. Their overall amplitude was normalized to ensure equal intensities.

During EEG recording, the auditory stimuli were presented binaurally with HiFi headphones (Sennheiser, HD 25-1, 70 Ω, Ireland) at a sound pressure level of 70 dB. Stimulus presentation was controlled by the Presentation software (Version 18.1, RRID:SCR_002521). The participants were instructed to watch a silent black and white film and to ignore the simultaneously presented auditory stimuli. This passive listening experiment consisted of five blocks, presented in a random order across participants. In each block, one of the five piano tones was presented more frequently (420 times, occurrence probability = 60 %; standard tone) than the other four (70 times each, occurrence probability = 10%; deviant tones). Each piano tone served as standard tone in one block and as deviant tone in all other blocks. As the EEG analyses of the original study, we focused on the blocks with standard tones of 440 Hz (block A) and of 264 Hz (block C). In these blocks, deviation magnitude increased or decreased unambiguously. Therefore, it was possible to test the effect of deviation magnitude on the EEG signal. Table 2 provides an overview of the study design. Presentation of the stimuli was pseudorandomized in each block. To establish a stable memory trace (Näätänen and Winkler, 1999), the first 15 tones were standards. For the remaining trials, deviants were always followed by at least one standard tone, and at least two different deviants were inserted before the same deviant could appear again. The interstimulus interval between the tones was fixed to 550 ms. The entire EEG recording lasted around 45 minutes.

EEG Recording and Preprocessing

EEG data was recorded with a sampling rate of 1000 Hz and an online band-pass filter of 0.1 - 100 Hz using a BrainAmp amplifier (Brainproducts, Munich, Germany). Thirty-two silver/silver-chloride electrodes were placed according to a subset of the 10/10 system, and an electrode on the tip of the nose was used as the reference. Electrode impedance was kept below 10 kΩ by applying an electrically conductive gel.

Preprocessing of the EEG data was conducted with the BrainVision Analyzer software package (Version 2.1, https://www.brainproducts.com/, RRID:SCR_002356). Data were filtered offline with a band-pass filter of 1 - 20 Hz (48 dB/octave) and a notch filter of 50 Hz. Eye movement artifacts (eye blinks and saccades) were corrected using an independent component analysis (ICA; Jung, et al., 2000), and noisy channels were interpolated. Remaining artifacts were removed using an automatic raw data inspection algorithm when a voltage gradient criterion of 50 µV/ms, an amplitude criterion of ±100 µV, or a low activity criterion of 0.5 µV/ 100 ms was exceeded. After preprocessing, the EEG signal was divided into segments of 500 ms (-100 to 400 ms from stimulus onset). These segments were baseline corrected (-100 to 0 ms) and averaged to ERPs. To compute difference waves, the ERPs evoked by the five standard tones were subtracted from the ERPs evoked by the physically identical deviants presented in the two blocks of interest (block A and block C). The grand averages of the difference waves for each deviant over all participants are shown in Figure 1. In Figure 2, the grand averages are presented separately for each group.

We extracted peak values of the resulting difference waves for the MMN and P3a from a pooling of nine frontal and central electrodes (F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4). In the original study, both ERP components elicited maximal amplitudes over these electrodes, and a similar voltage distribution could be observed in the data of the current replication study (see Figure 3; The topographical maps were created using code from the R package *EEGutils* (Craddock, 2018)). Peaks were selected using an automatic peak detection algorithm and verified by visual inspections.

Statistical Analyses

All statistical analyses were carried out in R (version 3.4.3; https://www.r-project.org, RRID:SCR_001905). To compare the groups in terms of demographics and musical experience, we applied Welch's *t*-tests. Effect sizes for *t*-tests are given in Cohen's *d* (Cohen, 1988).

For statistical analyses of the peak amplitudes and latencies, we replicated the null hypothesis statistical testing (NHST) of the original paper (replication analyses) and additionally performed Bayes factor analyses (exploratory Bayesian analyses).

In the replication analyses, a two-way mixed analysis of variance (ANOVA) with two levels of Group (AP and RP) and four levels of Deviation (four deviants) was computed separately for each ERP component and each block of interest using the R package *ez* (version 4.4.0; https://cran.r-project.org/web/packages/ez/index.html). *P*-values and degrees of freedom were adjusted using Greenhouse-Geisser correction when Mauchly's test revealed non-sphericity. For the ANOVAs, generalized eta-squared (η^2_G) is reported as the effect size estimate (Bakeman, 2005). Additionally, we report Cohen's d for the main effect of Group (Cohen, 1988). As in the original study, results with *p*-values less than or equal to 0.05 are termed significant.

Bayes Factors

Using NHST provides direct comparability with the original study. However, because NHST only allows to reject the null hypothesis (H0), but not the alternative (H1), non-significant results cannot differentiate between insensitive data and evidence in favor of H0. To decide whether a replication was successful or not, a quantification of null results is especially useful. Contrary to NHST, Bayes factors allow such conclusions on whether the evidence supports H0, the evidence supports H1, or the evidence is ambiguous (e.g. Dienes, 2014, 2011; Lee and Wagenmakers, 2013; Rouder et al., 2009). Bayes factors express the ratio between the likelihood of the data under one hypothesis (e.g. H0) relative to another hypothesis (e.g. H1). A Bayes factor BF01 of 10 (or the inverse $\frac{1}{BF_{01}}$ = BF10 = 0.1) can be directly interpreted as the data being 10 times more likely to occur under H0 compared to H1. As a consequence, Bayes factors are well suited to interpret non-significant results (Dienes, 2014) and to quantify the success of a replication (Anderson and Maxwell, 2016; Verhagen and Wagenmakers, 2014).

We calculated Bayes factors using the default Cauchy priors (scaling factor r = 0.707) as implemented in the *BayesFactor* package in R (version 0.9.12-4.2; https://cran.r-project.org/web/packages/BayesFactor/index.html) with 100000 iterations. Priors were not based on the effect sizes reported in the original study because small samples often result in inflated effect size estimates (Button et al., 2013; Halsey et al., 2015; Ioannidis, 2008). However, to ensure the robustness of our results, we additionally tested a range of priors (i.e., $r = 0.50$, $r = 1.00$, $r = 1.20$), and the results supported the same main conclusions.

Paralleling the replication analyses, we performed Bayesian ANOVAs (BANOVA; Rouder et al., 2017) on the peak amplitudes and latencies separately for each ERP component in each block. Bayes factors of interaction effects were assessed by comparing the full model (Group + Deviation + Group * Deviation + Subject) to the model without the interaction effect (Group + Deviation + Subject).

To facilitate interpretation, we report BF_{10} when Bayes factors favored the alternative hypothesis and BF_{01} ($\frac{1}{BF_{10}}$) when Bayes factors favored the null hypothesis. Following Jeffreys' (1961; edited by Lee and Wagenmakers, 2013) terminology, a Bayes factor between 1 and 3 is considered anecdotal evidence, between 3 and 10 moderate evidence, between 10 and 30 strong evidence, between 30 and 100 very strong evidence, and above 100 extreme evidence for the respective hypothesis.

Results

Demographics and Behavioral Data

Welch's *t*-tests did not reveal any significant group differences in age ($t_{(100.58)} = 1.39$, $p = .17$, $d = 0.27$), intelligence ($t_{(101.99)} = -1.43$, $p = .15$, $d = 0.28$), age of onset of musical training ($t_{(100.89)} = -1.16$, $p = .25$, $d = 0.23$), and cumulative musical training hours over the lifespan ($t_{(99.49)} = 1.41$, $p = .16$, $d = 0.27$). However, the two groups differed in musical aptitude ($t_{(99.41)}$

= 2.23, p = .028, d = 0.44), and AP musicians performed significantly better in the pitch-labeling test ($t_{(101.75)}$ = 13.77, p < .001, d = 2.70; see Figure 4).

Electrophysiological Data: Replication Analyses

The analyses of the MMN amplitudes and latencies showed similar results as in the original study. The original study reported main effects of Deviation for MMN amplitudes and latencies, but only in block A. In the present study, we found a significant main effect of Deviation on MMN amplitudes in both block A ($F_{(2.90,\ 296.15)}$ = 45.60, p < .001, η^2_G = 0.21) and block C ($F_{(2.92,\ 297.71)}$ = 4.28, p = .006, η^2_G = 0.03). However, the generalized eta-squared indicated that the effect in block C was small and comparable to the one obtained in the original study (η^2_G = 0.04). Additionally, as visible in Figure 1 and Figure 5, the amplitudes did not consistently get larger with increasing deviation magnitude in block C. As in the original study, the analysis did not reveal any significant effects of Group (block A: $F_{(1,\ 102)}$ = 0.45, p = .51, η^2_G = 0.002, d = 0.08; block C: $F_{(1,102)}$ = 1.52, p = .22, η^2_G = 0.005, d = 0.14) or significant interactions for MMN amplitudes (block A: $F_{(2.90,\ 296.15)}$ = 0.52, p = .66, η^2_G = 0.003; block C: $F_{(2.92,\ 297.71)}$ = 1.87, p = .14, η^2_G = 0.01).

A similar pattern was found for MMN latencies. There was a significant main effect of Deviation in block A ($F_{(2.52,\ 256.66)}$ = 4.99, p = .004, η^2_G = 0.03) and block C ($F_{(2.86,\ 291.60)}$ = 7.60, p < .001, η^2_G = 0.04), but effect sizes were small. The main effects of Group (block A: $F_{(1,\ 102)}$ = 0.01, p = .94, η^2_G < 0.001, d = 0.008; block C: $F_{(1,102)}$ = 0.42, p = .52, η^2_G = 0.002, d = 0.08) and the interactions (block A: $F_{(2.52,\ 256.66)}$ = 0.78, p = .48, η^2_G = 0.005; block C: $F_{(2.86,\ 291.60)}$ = 0.80, p = .49, η^2_G = 0.004) did not reach significance.

The main result reported in the original study were reduced P3a amplitudes in AP musicians compared to non-AP musicians. P3a latencies were not evaluated in the original study but are reported here for completeness. In line with the original study, the replication analyses showed a significant main effect of Deviation on P3a amplitudes in block A ($F_{(2.63,268.46)}$ = 55.02, p < .001, η^2_G = 0.25), but not in block C ($F_{(2.87,\ 292.91)}$ = 1.39, p = .25, η^2_G = 0.007). However, contrary to the original study, we did not find any significant main effects of Group (block A: $F_{(1,\ 102)}$ = 0.08, p = .78, η^2_G = 0.002, d = 0.03; block C: $F_{(1,102)}$ = 1.19, p = .28, η^2_G = 0.006, d = 0.15) or interaction effects (block A: $F_{(2.63,\ 268.46)}$ = 0.92, p = .42, η^2_G = 0.005; block C: $F_{(2.87,\ 292.91)}$ = 1.14, p = .33, η^2_G = 0.005) for P3a amplitudes (see Figure 5).

The analysis of P3a latencies also revealed a significant main effect of Deviation in block A ($F_{(2.22,\ 226.56)}$ = 5.58, p = .003, η^2_G = 0.04), but no significant main effect of Group ($F_{(1,\ 102)}$ = 0.09, p = .77, η^2_G < 0.001, d = 0.03) and no interaction ($F_{(2.22,\ 226.56)}$ = 0.50, p = .63, η^2_G = 0.003). In block C, there was no significant main effect (Deviation: $F_{(2.87,\ 292.44)}$ = 1.58, p = .20, η^2_G = 0.009; Group: $F_{(1,\ 102)}$ = 0.05, p = .82, η^2_G < 0.001, , d = 0.03) or interaction ($F_{(2.87,\ 292.44)}$ = 0.43, p = .72, η^2_G = 0.002).

Electrophysiological Data: Exploratory Bayesian Analyses

Replication analyses of MMN and P3a amplitudes yielded non-significant results for all group comparisons. To better distinguish between insensitive evidence, evidence for the alternative hypothesis, and evidence for the null hypothesis, we computed Bayes factors.

For MMN amplitudes, the Bayes factors mostly mirrored the results from the replication analyses. In block A, we obtained extreme evidence for an effect of Deviation ($BF_{10} = 7.32 * 10^{21}$), moderate evidence for the absence of an effect of Group (BF_{01} = 5.93) and strong evidence for the absence of an interaction effect (BF_{01} = 21.52). In block C, evidence for an effect of Deviation was less strong than in block A (BF_{10} = 3.25). Further, Bayes factors showed moderate evidence that there was no group difference (BF_{01} = 3.70) and no interaction (BF_{01} = 3.92).

As in the replication analyses, results for the MMN latencies were similar to those obtained for MMN amplitudes. Bayes factors provided evidence for the existence of a difference between deviants in block A (BF_{10} = 9.36) and block C (BF_{10} = 242.91), but not for differences between groups (block A: BF_{01} = 7.17; block C: BF_{01} = 5.10) or for an effect of interaction (block A: BF_{01} = 15.28; block C: BF_{01} = 15.77).

The replication analyses of P3a amplitudes revealed a significant effect of Deviation in block A. All other effects did not reach significance. Bayes factors strongly supported the existence of a difference between deviants in block A ($BF_{10} = 2.06 * 10^{26}$), but not in block C (BF_{01} = 15.86). In terms of group differences, there was moderate evidence for the null hypothesis in both block A (BF_{01} = 7.32) and block C (BF_{01} = 3.14). Bayes factors also strongly favored the null hypothesis regarding the interaction (block A: BF_{01} = 13.40; block C: BF_{01} = 10.40).

For P3a latencies, there was strong evidence for an effect of Deviation in block A (BF_{10} = 26.64). For all other effects, Bayes factors provided support for the null hypothesis in both block A (Group: BF_{01} = 7.29; interaction: BF_{01} = 22.07) and block C (Deviation: BF_{01} = 15.86; Group: BF_{01} = 6.30; interaction: BF_{01} = 10.40).

Electrophysiological Data: Exploratory Subgroup Analyses

The sample of the present study differed from the sample of the original study in three main ways: First, our sample was quite evenly balanced in terms of gender while the original study investigated predominantly female subjects. This might have influenced the results as females have previously been shown to have larger P3a amplitudes than males (visual paradigm: Conroy and Polich, 2007). Second, there was no overlap between the two groups in the pitch-labeling scores in the original study, but there is an overlap in our sample. Third, there was a small but significant difference in musical aptitude (AMMA) between groups in the present study.

Since all these sample differences could account for the differences in the results, we conducted additional subgroup analyses for the P3a amplitude. One subgroup analysis was performed on just the female participants of our study (n_{AP} = 27, $n_{non\text{-}AP}$ = 24). A second subgroup analysis was performed on the third of the participants with the lowest pitch-labeling scores (< 31.79%, n = 35) and the third of the participants with the highest pitch-labeling scores (> 72.83 %, n = 35). This allowed us to check whether the absence of the AP effect on the P3a was due to the more heterogenous groups in the present study. A third subgroup analysis corresponded as closely as possible to the original study in terms of pitch-labeling scores and sample size: only participants with scores < 10 % (n = 9) and > 93 % (n = 15) entered this analysis. Finally, we also performed an analysis of covariance (ANCOVA) with the AMMA score as covariate to test whether the between-group difference in musical aptitude influenced the result.

For the subgroup of females only, analysis of the P3a amplitude revealed an effect of Deviation in block A ($F_{(2.75,\ 134.94)} = 21.83$, $p < .001$, $\eta^2_G = 0.23$, $BF_{10} = 1.13 * 10^{10}$) but no effect of Group ($F_{(1,\ 49)} = 0.20$, $p = .66$, $\eta^2_G = 0.001$, $d = 0.063$, $BF_{01} = 4.95$) or an interaction effect

($F_{(2.75,\ 134.94)} = 0.35$, $p = .77$, $\eta^2_G = 0.004$, $\mathrm{BF}_{01} = 12.72$). No significant effect was found in block C (Group: $F_{(1,\ 49)} = 0.29$, $p = .59$, $\eta^2_G = 0.003$, $d = 0.11$, $\mathrm{BF}_{01} = 3.43$; Deviation: $F_{(2.89,\ 141.73)} = 0.68$, $p = .56$, $\eta^2_G = 0.007$, $\mathrm{BF}_{01} = 17.61$, Interaction: $F_{(2.89,\ 141.73)} = 0.35$, $p = .78$, $\eta^2_G = 0.003$, $\mathrm{BF}_{01} = 12.74$).

Similarly, the analysis with the lowest and highest performing third of participants showed an effect of Deviation in block A ($F_{(2.63,\ 178.59)} = 38.39$, $p < .001$, $\eta^2_G = 0.27$, $\mathrm{BF}_{10} = 9.96 * 10^{17}$) but no effect of Group ($F_{(1,\ 68)} = 0.04$, $p = .83$, $\eta^2_G < 0.001$, $d = 0.09$, $\mathrm{BF}_{01} = 5.18$) or an interaction effect ($F_{(2.63,\ 178.59)} = 0.38$, $p = .74$, $\eta^2_G = 0.003$, $\mathrm{BF}_{01} = 18.79$). Again no significant effects were observed in block C (Group: $F_{(1,\ 68)} = 2.72$, $p = .11$, $\eta^2_G = 0.02$, $d = 0.35$, $\mathrm{BF}_{10} = 1.50$; Deviation: $F_{(2.78,\ 188.84)} = 0.93$, $p = .42$, $\eta^2_G = 0.007$, $\mathrm{BF}_{01} = 18.74$, Interaction: $F_{(2.78,\ 188.84)} = 2.42$, $p = .072$, $\eta^2_G = 0.02$, $\mathrm{BF}_{01} = 2.88$).

Likewise, with even more extreme groups (< 10 % and > 93 % pitch-labeling performance), there was an effect of Deviation in block A ($F_{(2.54,\ 55.91)} = 24.34$, $p < .001$, $\eta^2_G = 0.44$, $\mathrm{BF}_{10} = 5.97 * 10^9$) but no other effect in block A (Group: $F_{(1,\ 22)} = 0.03$, $p = .86$, $\eta^2_G < 0.001$, $d = 0.03$, $\mathrm{BF}_{01} = 3.62$; Interaction: $F_{(2.54,\ 55.91)} = 0.64$, $p = .57$, $\eta^2_G = 0.02$, $\mathrm{BF}_{01} = 4.61$) or block C (Group: $F_{(1,\ 22)} = 2.68$, $p = .12$, $\eta^2_G = 0.06$, $d = 0.55$, $\mathrm{BF}_{01} = 1.03$; Deviation: $F_{(2.67,\ 58.74)} = 1.22$, $p = .31$, $\eta^2_G = 0.02$, $\mathrm{BF}_{01} = 4.61$, Interaction: $F_{(2.67,\ 58.74)} = 0.91$, $p = .43$, $\eta^2_G = 0.02$, $\mathrm{BF}_{01} = 2.94$).

The ANCOVA with the AMMA score as covariate on the full sample revealed similar results: an effect of Deviation in block A ($\mathrm{F}_{(2.63,\ 268.46)} = 55.02$, $\mathrm{p} < .001$, $\eta^2_G = 0.25$) and no other effects neither in block A (Group: $\mathrm{F}_{(1,\ 102)} = 0.04$, $\mathrm{p} = .85$, $\eta^2_G < 0.001$; Interaction: $\mathrm{F}_{(2.63,\ 268.46)} = 0.92$, $\mathrm{p} = .42$, $\eta^2_G = 0.01$) nor in block C (Group: $\mathrm{F}_{(1,\ 102)} = 1.95$, $\mathrm{p} = .17$, $\eta^2_G = 0.009$; Deviation: $\mathrm{F}_{(2.87,\ 292.91)} = 1.39$, $\mathrm{p} = .25$, $\eta^2_G = 0.007$, Interaction: $\mathrm{F}_{(2.87,\ 292.91)} = 1.14$, $\mathrm{p} = .33$, $\eta^2_G = 0.006$).

We also performed an ANCOVA on the subgroup of participants with comparable sample size and pitch-labeling scores as in the original study. Again, we found an effect of Deviation in block A ($F_{(2.54,\ 55.91)} = 24.34$, $p < .001$, $\eta^2_G = 0.44$) but no other effects in either block A (Group: $F_{(1,\ 22)} = 0.04$, $p = .85$, $\eta^2_G < 0.001$; Interaction: $F_{(2.54,\ 55.91)} = 0.64$, $p = .57$, $\eta^2_G = 0.02$) or block C (Group: $F_{(1,\ 22)} = 3.81$, $p = .064$, $\eta^2_G = 0.08$; Deviation: $F_{(2.67,\ 58.74)} = 1.22$, $p = .31$, $\eta^2_G = 0.03$; Interaction: $F_{(2.67,\ 58.74)} = 0.91$, $p = .43$, $\eta^2_G = 0.02$).

Discussion

In the present study, we attempted to replicate Rogenmoser et al.’s (2015) finding of electrophysiological group differences between AP and non-AP musicians during passive listening. Rogenmoser et al. (2015) investigated the automatic nature of AP by recording EEG during a passive auditory oddball paradigm. By analyzing MMN and P3a, they intended to assess the contribution of both pre-attentive (as reflected by the MMN) and more cognitive processes (as reflected by the P3a) in AP. To compare the tone processing between AP and non-AP musicians under different deviation conditions, they applied a paradigm with multiple tuned and mistuned deviants. In line with previous research (Tervaniemi et al., 1993, Matsuda et al., 2013: condition with tuned tones), they did not find any significant group differences in the MMN. In contrast, Rogenmoser et al. (2015) observed smaller P3a amplitudes in AP musicians. This group difference was only found in conditions in which the deviation magnitude was larger than one semitone (264 Hz deviant in block A and all deviants in block C), suggesting that AP musicians process between-pitch but not within-pitch categories differentially than non-AP musicians. Because the P3a has been associated with an early reallocation of attention (Escera et al., 1998; Friedman et al., 2001; Kujala et al., 2007; Polich, 2007), the smaller amplitudes in AP musicians were interpreted as an indication for more efficient cognitive tone processing in AP. The authors concluded that the “P3a component turned out to be a specific marker for AP”(Rogenmoser et al., 2015).

In the current direct replication study, we found no significant group differences in the MMN, confirming the results of the original study. However, and most critically, there were also no significant group differences in the P3a. Additional Bayes factor analyses revealed that the data is more likely under the null hypothesis, implying that AP and non-AP musicians’ tone processing, as indicated by MMN and P3a peak amplitudes and latencies, does not differ during passive listening. Thus, our results challenge the view of cognitive facilitation in AP during passive listening.

In passive auditory oddball paradigms, the MMN typically occurs in response to a change (deviation) in auditory stimulation within a sequence of repeated stimuli (standard tone). The main generator of the MMN is located in the auditory cortex (for a review, see Näätänen et al., 2007), where the repeated presentation of a stimulus potentially causes the formation of a short-term memory trace (Näätänen and Winkler, 1999). The MMN is generated when a new auditory input differs from the representation in this sensory memory trace. Because this mismatch detection process does not require that the stimuli are attended, it is thought to be automatic (e.g. Paavilainen et al., 2007; Sussman et al., 2003). Accordingly, the MMN is considered an objective measure of auditory discrimination accuracy (Näätänen et al., 2007). Consistent with this view, it has been shown that the amplitude of the MMN increases when discrimination performance improves through training (Atienza et al., 2002; Menning et al., 2000; Näätänen et al., 1993). The MMN amplitude also correlates more generally with behavioral discrimination accuracy (Näätänen et al., 1993; Novak et al., 1990). Similarly, the MMN is also influenced by the deviation magnitude, with larger — and therefore more salient — deviations evoking larger amplitudes and shorter latencies (e.g., Berti et al., 2004; Novitski et al., 2004; Sams et al., 1985).

The original study reported an effect of deviation magnitude for block A but not for block C. The authors provided a possible explanation that in block C, all deviants were clustered around an extreme deviation level, with a distance between eight and nine semitones from the standard tone. Consequently, all deviants were probably equally easy to detect. In accordance with the original study, our results showed larger MMN amplitudes and shorter MMN latencies for larger deviations in block A. In block C, the effect also reached significance, but like in the original study, amplitudes did not unambiguously increase with deviation magnitude (compare Figure 3), suggesting a context effect in this specific block.

More importantly, we also replicated the result of non-significant group differences between the AP and non-AP musicians in MMN measures. The Bayes factor analysis additionally provided support for the null hypothesis. Thus, our data was more likely under the hypothesis that there were no differences in the MMN amplitudes and latencies between the two groups than under the H_1. Our results are not only consistent with the original study but also with other previous research. Using tuned and mistuned pure tones and piano tones, Tervaniemi et al. (1993) did not find group differences between AP and non-AP musicians in MMN amplitudes and latencies. In Matsuda et al.'s (2013) study, MMN amplitudes of AP and non-AP musicians did also not differ for tuned tones, but AP musicians showed larger MMN amplitudes for mistuned tones. However, this effect might have been influenced by the fact that their AP musicians were musically more experienced than the non-AP musicians. Previous research has shown that musical experience can increase MMN amplitudes (Koelsch et al., 1999; Putkinen et al., 2014), specifically in response to mistuned tones (Tervaniemi et al., 2014).

Because the MMN is associated with a passive discrimination process, Tervaniemi et al. (1993) concluded from their results that “pitch naming and discrimination are based on different brain mechanisms”. This coincides with results from behavioral studies showing that pitch-labeling accuracy is not correlated with behavioral pitch-discrimination accuracy (Fujisaki and Kashino, 2002; Sergeant, 1969). Thus, evidence from both behavioral and electrophysiological data suggests that AP does not simply rely on refined pitch discrimination.

In passive auditory oddball paradigms, the MMN is often followed by the P3a, a subcomponent of the P300. Both components have been proposed to play a role in the reallocation of attention to unattended stimuli (Escera et al., 2000; Kujala et al., 2007; Näätänen, 1990), with the processes underlying MMN probably initiating the attention switching and the P3a directly reflecting it. The P3a is affected by the magnitude of deviation in similar ways as the MMN (Berti et al., 2004). As for the MMN, the original study found such a deviation modulation only in block A, probably again due to the more extreme deviation levels in block C. The present study successfully replicated these results. In block A, P3a amplitudes increased and P3a latencies decreased with increasing deviation, and as in the original study, no similar effect was observed in block C. Future studies should more systematically investigate this dependence on specific contexts.

Even though the modulation of the MMN and P3a as a function of deviation magnitude is an interesting aspect of general pitch processing, the main finding of the original study was the reduced P3a amplitudes in AP musicians. This result was compared to findings from the parietal P3b, another subcomponent of the P300, which is elicited in active oddball paradigms and often called P300 in these studies. The P3b has been linked to working memory updating (for a review, see Kok, 2001; Polich, 2007) and has been investigated more thoroughly in AP research than the P3a. The first study to detect differences in ERPs during pitch processing reported the absence of a P3b in individuals with AP (Klein et al., 1984). This was regarded as an indication that individuals with AP did not need to update their auditory working memory during the task because their pitch representations are permanent. Subsequently, some studies replicated the absence or diminution of P3b amplitudes in AP (Crummer et al., 1994; Hantz et al., 1992; Wayman et al., 1992), but others did not (Hantz et al., 1995; Hirose et al., 2002). . This inconsistency was shown to be caused by differential pitch-processing strategies (RP or AP) employed by the participants based on the specific task instructions, the task difficulty, and the individual level of AP (Bischoff Renninger et al., 2003).

Individual differences in listening strategies could explain why we did not replicate the effect of AP on the P3a. However, this seems rather unlikely as the use of top-down strategies was controlled with the help of a distractor task (watching a silent film) in both the original and the replication study. Given how unreliable the effect of AP on ERPs is even in active tasks, we believe it is more plausible that the differences in passive pitch processing are too subtle to be reliably detectable with ERP peak measures. Alternatively, it could also be speculated that the pitch labeling is only initiated when actively attending the auditory stimuli or when performing a labeling-related task (e.g. bimodal Stroop task; Akiva-Kabiri and Henik, 2012). Compelling evidence for an automatic pitch-labeling process comes from behavioral studies, in which the auditory stimuli had to be attended to solve the task. For instance, individuals with AP performed poorer in auditory Stroop tasks when they heard sung tone names and were instructed to repeat the syllable while ignoring the pitch it was sung in (Itoh et al., 2005; Miyazaki, 2004; Schulze et al., 2013). AP also hindered performance in a relative pitch task, in which participants had to compare a visual notation with the auditory presentation of a melody (Miyazaki and Rakowski, 2002). Further evidence for the automaticity of pitch labeling was provided by neuroscientific studies that observed differential electrophysiological or hemodynamic responses in AP musicians during attentive listening (Itoh et al., 2005; Zatorre et al., 1998). Contrary to these studies, in the present study, participants were instructed to focus their attention on a silent film and to ignore the auditory stimuli altogether. AP musicians can label tones fast and effortlessly, but they may not necessarily do so under all circumstances. Apart from the specific task, also other situational factors like stress and fatigue might influence pitch-labeling performance and pitch-labeling automaticity. Additionally, it is also possible that there are considerable interindividual differences in the level of automaticity of AP per se. Future studies will hopefully uncover the role of such influences on this extraordinary ability and its neural underpinnings in more detail.

Even though this study could not demonstrate a cognitive facilitation in AP during passive listening, we believe our results do not challenge existing cognitive theories of AP, like the two-component model (Levitin, 1994). The two-component model focuses on the use of long-term pitch memory representations and their association with labels in AP. This mechanism in turn poses less demands on working memory in some tasks than using relative pitch (e.g. Itoh et al., 2005; Klein et al., 1984; Schulze et al., 2009). In contrast to these mnemonic processes, the P3a in passive auditory oddball paradigms is mostly associated with attentional processes, which are not explicitly postulated as part of AP by the two-component model. Further research should be undertaken to determine the influence of attention on pitch processing in AP.

We attempted a direct replication of the original study, still there are some mentionable differences between the original and the replication study that might have influenced the results. While questionnaires on musical experience and the pitch-labeling test were assessed with paper-pencil in the original study, we used online questionnaires and an online pitch-labeling test in the present study. Because our participants underwent an extensive test protocol in the context of the larger AP project spanning several days during which they participated in various (f)MRI and EEG experiments, we tried to keep the travel burden for them as low as possible by providing the opportunity to work on several tests at home. For our statistical analyses, we used the software R instead of SPSS , and we performed Welch's *t*-tests instead of Student's *t*-test because they are more robust for groups with unequal sample sizes (Delacre et al., 2017; Ruxton, 2006). For ANOVAs, we reported generalized eta-squared instead of partial eta-squared as recommended by Bakeman (2005). Like in the original study, groups were defined based on self-report. Contrary to the original study, in our replication study, the non-AP musicians performed above chance in the pitch-labeling test. Accordingly, it could be argued that the groups were less homogenous than in the original study and that this is the reason for the unsuccessful replication. However, because trials in the pitch-labeling test lasted 15 s instead of 5 s participants probably had enough time to employ RP strategies in our test. It can be expected that highly-trained musicians perform above chance levels when given the opportunity to use RP strategies. For the same reason, it is possible that the pitch-labeling performance of AP musicians was also overestimated. The longer maximal trial duration was due to the online implementation of the pitch-labeling test. In a pilot study, we tested a version with the original trial duration of 5 s, which turned out to be very demanding and difficult to solve even for AP musicians because of the multiple-choice format with 36 answer options. We would recommend future studies to measure reaction times in pitch-labeling tests to be able to better disentangle the effortless and fast AP strategy from the slower RP strategy, or to apply a pitch-labeling test that impedes the usage of RP strategies (e.g. as suggested in Wengenroth et al., 2014). Yet, it still remains unclear which is the best way to objectively identify AP ability and if it is even possible to do so, a question that has been asked frequently and was also discussed in an early influential review on AP (Takeuchi and Hulse, 1993). The authors addressed several methods to quantify AP, ranging from producing tones to different variants of pitch-labeling tests. Up to date, the pitch-labeling tests applied in AP research differ considerably in procedure (e.g. trial duration, answer registration, sine tones/instrumental tones), the number of used tones, and the presentation technique (e.g. online vs. lab). Most importantly, no specific cut-off has been established to distinguish AP from non-AP possessors. Thus, in the present study, the pitch-labeling test only served as a validation tool. For group assignment, we relied on self-report since only the participants themselves can judge whether they possess the ability to employ AP strategies. In addition, as demonstrated in the exploratory subgroup analyses, the conclusions of the results remained the same even when just considering participants with the lowest and highest pitch-labeling scores, suggesting that this sample difference between studies did not cause the absence of the AP effect. Similarly, conclusions about the P3a amplitude did not change when just looking at the female participants. Thus, even though the original study was less balanced in terms of gender than the present study, the absence of an effect of AP on the P3a amplitude in the present study does not seem to be caused by gender distribution differences between studies. Also, according to current scientific understanding gender differences in neuroscientific cognitive studies are most often due to small sample sizes and should only be interpreted when the influence of hormonal levels was controlled for (Jäncke, 2018). It should also be mentioned that in the present study, the AP and non-AP musicians showed a statistically significant – albeit small in absolute terms (< 3 points out of 80 possible points) – difference in musical aptitude (AMMA). However, scores are comparable to those reported in the original study, and additional covariance analyses with the AMMA score as covariate showed the same results as the replication analyses.

Finally, it is important to note that a single replication study can never conclusively confirm or disconfirm previous findings. Nevertheless, our results cast reasonable doubt that there is cognitive facilitation in AP during passive tone processing as indicated by the P3a. The more so since our sample was four times the size of the original study, and Bayes factors analyses provided evidence that the proposed effect does not exist. Although it is possible that additional factors we did not control for moderated the effect, we reduced such moderators to a minimum by doing a direct replication. Thus, if an effect of AP on the P3a really exists, its true effect size is probably much smaller than reported in the original study as it is not reliably detectable in a large sample, and its generalizability might be limited.

Considering the large effect size obtained in the original study, the results of the current study demonstrate that only through replications a better estimate of the true effect can be obtained. We believe replications are desirable in science in general and particularly in research fields that are prone to false-positive results and to overestimations of effect sizes due to small samples. Neuroscientific studies often use small samples because of the high financial costs and time-consuming data acquisition and analysis. Collaborative efforts between multiple research groups are suggested as a means to recruit larger sample sizes.

In summary, our direct replication of Rogenmoser et al. (2015) successfully replicated the non-significant results for group differences in the MMN. In contrast, we did not replicate the finding of smaller P3a amplitudes in AP musicians. Taken together, our study does not support electrophysiological differences between AP and non-AP musicians during passive listening. It is conceivable that the different pitch-processing modes of AP and RP can only be reliably distinguished either with more sensitive measures or in more attention-engaging tasks. In more general terms, the results of the present study underline both the importance of replications and of larger sample sizes in neuroscientific research.

References

Akiva-Kabiri L, Henik A (2012) A unique asymmetrical stroop effect in absolute pitch possessors. Exp Psychol 59:272–278.

Anderson SF, Maxwell SE (2016) There’s more than one way to conduct a replication study: Beyond statistical significance. Psychol Methods 21:1–12.

Annett M (1970) A classification of hand preference by association analysis. Br J Psychol 303–321.

Atienza M, Cantero JL, Dominguez-Marin E (2002) The time course of neural changes underlying auditory perceptual learning. Learn Mem 9:138–50.

Bakeman R (2005) Recommended effect size statistics for repeated measures designs. Behav Res Methods 37:379–384.

Baker M (2016) Is there a reproducibility crisis? Nature 533:452–454.

Berti S, Roeber U, Schröger E (2004) Bottom-up influences on working memory: Behavioral and electrophysiological distraction varies with distractor strength. Exp Psychol 51:249–257.

Bischoff Renninger L, Granot RI, Donchin E (2003) Absolute pitch and the P300 component of the event-related potential: An exploration of variables that may account for individual differences. Music Percept 20:357–382.

Brandt MJ, IJzerman H, Dijksterhuis A, Farach FJ, Geller J, Giner-Sorolla R, Grange JA, Perugini M, Spies JR, van ’t Veer A (2014) The Replication Recipe: What makes for a convincing replication? J Exp Soc Psychol 50:217–224.

Button KS, Ioannidis JPA, Mokrysz C, Nosek BA, Flint J, Robinson ESJ, Munafò MR (2013) Power failure: Why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14:365–76.

Cohen J (1988) Statistical power analysis for the behavioral science, 2nd ed. Hillsdale, NJ: Erlbaum.

Conroy MA, Polich J (2007) Normative variation of P3a and P3b from a large sample: Gender, topography, and response time. J Psychophysiol 21:22–32.

Craddock M (2018) craddm/eegUtils: eegUtils.

Crummer GC, Walton JP, Wayman JW, Hantz EC, Frisina RD (1994) Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch. J Acoust Soc Am 95:2720–2727.

Delacre M, Lakens D, Leys C (2017) Why psychologists should by default use Welch’s t-test instead of Student’s t-test. Int Rev Soc Psychol 30:92.

Deutsch D (2013) Absolute Pitch In: The Psychology of Music , pp141–182. San Diego: Elsevier.

Dienes Z (2014) Using Bayes to get the most out of non-significant results. Front Psychol 5:1–17.

Dienes Z (2011) Bayesian versus orthodox statistics: Which side are you on? Perspect Psychol Sci 6:274–290.

Elmer S, Sollberger S, Meyer M, Jäncke L, Jancke L (2013) An empirical reevaluation of absolute pitch: Behavioral and electrophysiological measurements. J Cogn Neurosci 25:1736–1753.

Escera C, Alho K, Schröger E, Winkler I (2000) Involuntary attention and distractability as evaluated with event-related potentials. Audiol Neurootol 5:151–166.

Escera C, Alho K, Winkler I, Näätänen R (1998) Neural mechanisms of involuntary attention to acoustic novelty and change. J Cogn Neurosci 10:590–604.

Friedman D, Cycowicz YM, Gaeta H (2001) The novelty P3: An event-related brain potential (ERP) sign of the brain’s evaluation of novelty. Neurosci Biobehav Rev 25:355–373.

Fujisaki W, Kashino M (2002) The basic hearing abilities of absolute pitch possessors. Acoust Sci Technol 23:77–83.

Garrido MI, Kilner JM, Stephan KE, Friston KJ (2009) The mismatch negativity: A review of underlying mechanisms. Clin Neurophysiol 120:453–463.

Gordon EE (1989) Manual for the advanced measures of music audiation. Chigaco: GIA Publication.

Halsey LG, Curran-Everett D, Vowler SL, Drummond GB (2015) The fickle P value generates irreproducible results. Nat Methods 12:179–185.

Hantz EC, Crummer GC, Wayman JW, Walton JP, Frisina RD (1992) Effects of musical training and absolute pitch on the neural processing of melodic intervals: A P3 event-related potential study. Music Percept 10:25–42.

Hantz EC, Kreilick KG, Braveman AL, Swartz KP (1995) Effects of musical training and absolute pitch on a pitch memory task: An event-related potential study. Psychomusicology A J Res Music Cogn 14:53–76.

Hirose H, Kubota M, Kimura I, Ohsawa M, Yumoto M, Sakakihara Y (2002) People with absolute pitch process tones with producing P300. Neurosci Lett 330:247–250.

Ioannidis JPA (2008) Why most discovered true associations are inflated. Epidemiology 19:640–648.

Itoh K, Suwazono S, Arao H, Miyazaki K, Nakada T (2005) Electrophysiological correlates of absolute pitch and relative pitch. Cereb Cortex 15:760–769.

Jäncke L (2018) Sex/gender differences in cognition, neurophysiology, and neuroanatomy. F1000Research 7:805.

Jeffreys H (1961) Theory of probability, 3rd ed. New York, NY: Oxford University Press.

Klein M, Coles MGH, Donchin E (1984) People with absolute pitch process tones without producing a P300. Science (80-) 223:1306–1309.

Koelsch S, Schröger E, Tervaniemi M (1999) Superior pre-attentive auditory processing in musicians. Neuroreport 10:1309–1313.

Kok A (2001) On the utility of P300 amplitude as a measure of processing capacity. Psychophysiology 38:557–577.

Kujala T, Tervaniemi M, Schröger E (2007) The mismatch negativity in cognitive and clinical neuroscience: theoretical and methodological considerations. Biol Psychol 74:1–19.

Lee MD, Wagenmakers E-J (2013) Bayesian cognitive modeling: A practical course. Cambridge: Cambridge University Press.

Lehrl S (2005) Mehrfachwahl-Wortschatz-Intelligenztest MWT-B. Balingen: Spitta.

Lehrl S, Triebig G, Fischer B (1995) Multiple-Choice Vocabulary-Test MWT as a valid and short test to estimate premorbid intelligence. Acta Neurol Scand 91:335–345.

Levitin DJ (1994) Absolute memory for musical pitch - evidence from the production of learned melodies. Percept Psychophys 56:414–423.

Levitin DJ, Rogers SE (2005) Absolute pitch: perception, coding, and controversies. Trends Cogn Sci 9:26–33.

Matsuda A, Hara K, Watanabe S, Matsuura M, Ohta K, Matsushima E (2013) Pre-attentive auditory processing of non-scale pitch in absolute pitch possessors. Neurosci Lett 548:155–158.

McDermott JH, Oxenham AJ (2008) Music perception, pitch, and the auditory system. Curr Opin Neurobiol 18:452–463.

Menning H, Roberts LE, Pantev C (2000) Plastic changes in the auditory cortex induced by intensive frequency discrimination training. Neuroreport 11:817–822.

Miyazaki K (2004) The auditory Stroop interference and the irrelevant speech/pitch effect: Absolute-pitch listeners can’t suppress pitch labeling. 18th Int Congr Acoust 3619–3622.

Miyazaki K (1989) Absolute pitch identification: effects of timbre and pitch region. Music Percept An Interdiscip J 7:1–14.

Miyazaki K (1988) Musical pitch identification by absolute pitch possessors. Percept Psychophys 44:501–512.

Miyazaki K, Rakowski A (2002) Recognition of notated melodies by possessors and nonpossessors of absolute pitch. Percept Psychophys 64:1337–1345.

Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie Du Sert N, Simonsohn U, Wagenmakers EJ, Ware JJ, Ioannidis JPA (2017) A manifesto for reproducible science. Nat Hum Behav 1:1–9.

Näätänen R (1990) The role of attention in auditory information processing as revealed by event-related potentials and other brain measures of cognitive function. Behav Brain Sci 13:201–288.

Näätänen R, Kujala T, Winkler I (2011) Auditory processing that leads to conscious perception: A unique window to central auditory processing opened by the mismatch negativity and related responses. Psychophysiology 48:4–22.

Näätänen R, Paavilainen P, Rinne T, Alho K (2007) The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin Neurophysiol 118:2544–2590.

Näätänen R, Paavilainen P, Tiitinen H, Jiang D, Alho K (1993) Attention and mismatch negativity. Psychophysiology.

Näätänen R, Winkler I (1999) The concept of auditory stimulus representation in cognitive neuroscience. Psychol Bull 125:826–859.

Nosek BA, Spies JR, Motyl M (2012) Scientific utopia: II. Restructuring incentives and practices to promote truth over publishability. Perspect Psychol Sci 7:615–631.

Novak GP, Ritter W, Vaughan HG, Wiznitzer ML (1990) Differentiation of negative event-related potentials in an auditory discrimination task. Electroencephalogr Clin Neurophysiol 75:255–275.

Novitski N, Tervaniemi M, Huotilainen M, Näätänen R (2004) Frequency discrimination at different frequency levels as indexed by electrophysiological and behavioral measures. Cogn Brain Res 20:26–36.

Oechslin MS, Meyer M, Jäncke L (2010) Absolute pitch-functional evidence of speech-relevant auditory acuity. Cereb Cortex 20:447–455.

Paavilainen P, Arajärvi P, Takegata R (2007) Preattentive detection of nonsalient contingencies between auditory features. Neuroreport 18:159–163.

Picton TW et al. (2000) Guidelines for using human event-related potentials to study cognition : Recording standards and publication criteria. Psychophysiology 37:127–152.

Polich J (2007) Updating P300: An integrative theory of P3a and P3b. Clin Neurophysiol 118:2128–2148.

Putkinen V, Tervaniemi M, Saarikivi K, de Vent N, Huotilainen M (2014) Investigating the effects of musical training on functional brain development with a novel Melodic MMN paradigm. Neurobiol Learn Mem 110:8–15.

Rogenmoser L, Elmer S, Jäncke L (2015) Absolute Pitch: Evidence for early cognitive facilitation during passive listening as revealed by reduced P3a amplitudes. J Cogn Neurosci 27:623–637.

Rouder JN, Morey RD, Verhagen J, Swagman AR (2017) Bayesian analysis of factorial designs. Psychol Methods 22:304–321.

Rouder JN, Speckman PL, Sun D, Morey RD, Iverson G (2009) Bayesian t tests for accepting and rejecting the null hypothesis. Psychon Bull Rev 16:225–237.

Ruxton GD (2006) The unequal variance t-test is an underused alternative to Student’s t-test and the Mann-Whitney U test. Behav Ecol 17:688–690.

Sams M, Paavilainen P, Alho K, Näätänen R (1985) Auditory frequency discrimination and event-related potentials. Electroencephalogr Clin Neurophysiol 62:437–448.

Schulze K, Gaab N, Schlaug G (2009) Perceiving pitch absolutely: Comparing absolute and relative pitch possessors in a pitch memory task. Bmc Neurosci 10.

Schulze K, Mueller K, Koelsch S (2013) Auditory stroop and absolute pitch: An fMRI study. Hum Brain Mapp 34:1579–1590.

Sergeant D (1969) Experimental investigation of absolute pitch. J Res Mus Ed 17:135–143.

Sussman E, Winkler I, Wang W (2003) MMN and attention: competition for deviance detection. Psychophysiology 40:430–435.

Szucs D, Ioannidis JPA (2017) Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLoS Biol 15:1–18.

Takeuchi H, Hulse SH (1993) Absolute pitch. Psychol Bull 113:345–361.

Tervaniemi M, Alho K, Paavilainen P, Sams M, Naatanen R (1993) Absolute pitch and event-related brain potentials. Music Percept 10:305–316.

Tervaniemi M, Huotilainen M, Brattico E (2014) Melodic multi-feature paradigm reveals auditory profiles in music-sound encoding. Front Hum Neurosci 8.

Verhagen J, Wagenmakers E-J (2014) Bayesians tests to quantify the success or failure of a replication attempt. J Exp Psychol Gen 143:1457–1475.

Wayman JW, Frisina RD, Walton JP, Hantz EC, Crummer GC (1992) Effects of musical training and absolute pitch ability on event-related activity in response to sine tones. J Acoust Soc Am 91:3527–3531.

Wengenroth M, Blatow M, Heinecke A, Reinhardt J, Stippich C, Hofmann E, Schneider P (2014) Increased volume and function of right auditory cortex as a marker for absolute pitch. Cereb Cortex 24:1127–1137.

Zatorre RJ (2003) Absolute pitch: A model for understanding the influence of genes and development on neural and cognitive function. Nat Neurosci 6:692–695.

Zatorre RJ, Perry DW, Beckett CA, Westbury CF, Evans AC (1998) Functional anatomy of musical processing in listeners with absolute pitch and relative pitch. Proc Natl Acad Sci U S A 95:3172–3177.

Legends

Figure 1. Grand averages of the difference waves (deviant ERP minus standard ERP). ERPs from the fronto-central pooling of electrodes were averaged over all participants for each deviation condition. The lines represent the means, the shaded areas indicate 95 % within-subject confidence intervals. Darker colors illustrate larger deviation magnitudes. In

block A (standard tone 440 Hz), amplitudes increase with larger deviation magnitude. In block C (standard tone 264 Hz), no such clear relationship can be observed.

Figure 2. Grand averages of the difference waves (deviant minus standard) for absolute pitch (AP, in red) and non-absolute pitch (non-AP, in blue) musicians. Deviation magnitude increases from top to bottom. The lines represent the group means, the shaded areas represent the 95 % between-subject confidence interval.

Figure 3. Voltage distributions over the scalp for the Mismatch Negativity (MMN) and P3a for each group and each deviant in block A (standard tone 440 Hz) and block C (standard tone 264 Hz). Topographies are shown at the timepoint of the peak according to the Grand Average of the specific deviation condition and group. Deviation magnitude increases from left to right. Both MMN and P3a are maximally expressed at fronto-central electrode sites. AP = absolute pitch, non-AP = non-absolute pitch.

Figure 4. Performance in the pitch-labeling test for absolute pitch (AP) and relative pitch (RP) musicians. Octave errors were treated as correct answers, resulting in a chance level of 8.33 % (dashed line). AP musicians are depicted in red, non-AP musicians in blue. AP musicians performed significantly better than non-AP musicians ($t_{(101.75)} = 13.77$, $p < .001$, $d = 2.70$).

Figure 5. MMN and P3a amplitudes of musicians with absolute pitch (AP) and without absolute pitch (non-AP) for all deviation conditions in block A (standard tone 440 Hz) and block C (standard tone 264 Hz). Deviants are ordered from left to right according to increasing deviation magnitude. Amplitudes of AP musicians are shown in red, amplitudes of non-AP musicians are shown in blue.

Table 1. Demographics and musical experience. Continuous measures are given as mean (standard deviations in parentheses). MWT-B, Mehrfachwahl-Wortschatz-Intelligenztest; AMMA, Advanced Measures of Music Audiation. [a] Raw scores, [b] Units are given in 1 x 104

Table 2. Study design. Deviant tones are listed from left to right according to increasing deviation magnitude.

Tables

Table 1.

Demographics and Musical Experience

	Absolute Pitch Musicians (*n* = 54)		**Non-Absolute Pitch** Musicians (*n* = 50)	
Sex				
Female	27		24	
Male	27		26	
Age (years)	26.67	(5.49)	25.30	(4.51)
Handedness				
Right-handed	47		45	
Left-handed	4		4	
Both-handed	3		1	
Intelligence (MWT-B) [a]	27.69	(5.10)	29.06	(4.68)
Age of Onset of Musical Training (years)	5.93	(2.39)	6.48	(2.46)
Lifetime Cumulative Training (hours) [b]	1.66	(1.22)	1.36	(0.96)
Musical Aptitude (AMMA) [a]	66.11	(6.31)	63.22	(6.86)
Pitch-labeling Test (%)	76.41	(19.55)	24.31	(19.01)

Continuous measures are given as mean (standard deviations in parentheses). MWT-B, Mehrfachwahl-Wortschatz-Intelligenztest; AMMA, Advanced Measures of Music Audiation.

[a] Raw scores

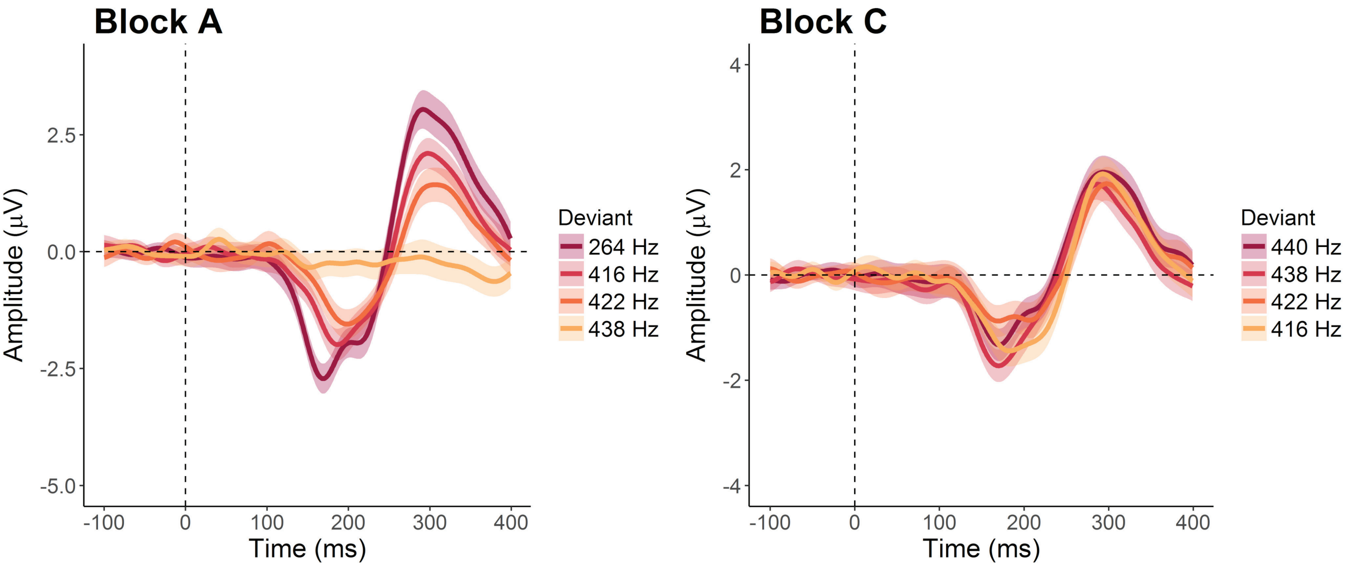
[b] Units are given in 1×10^4

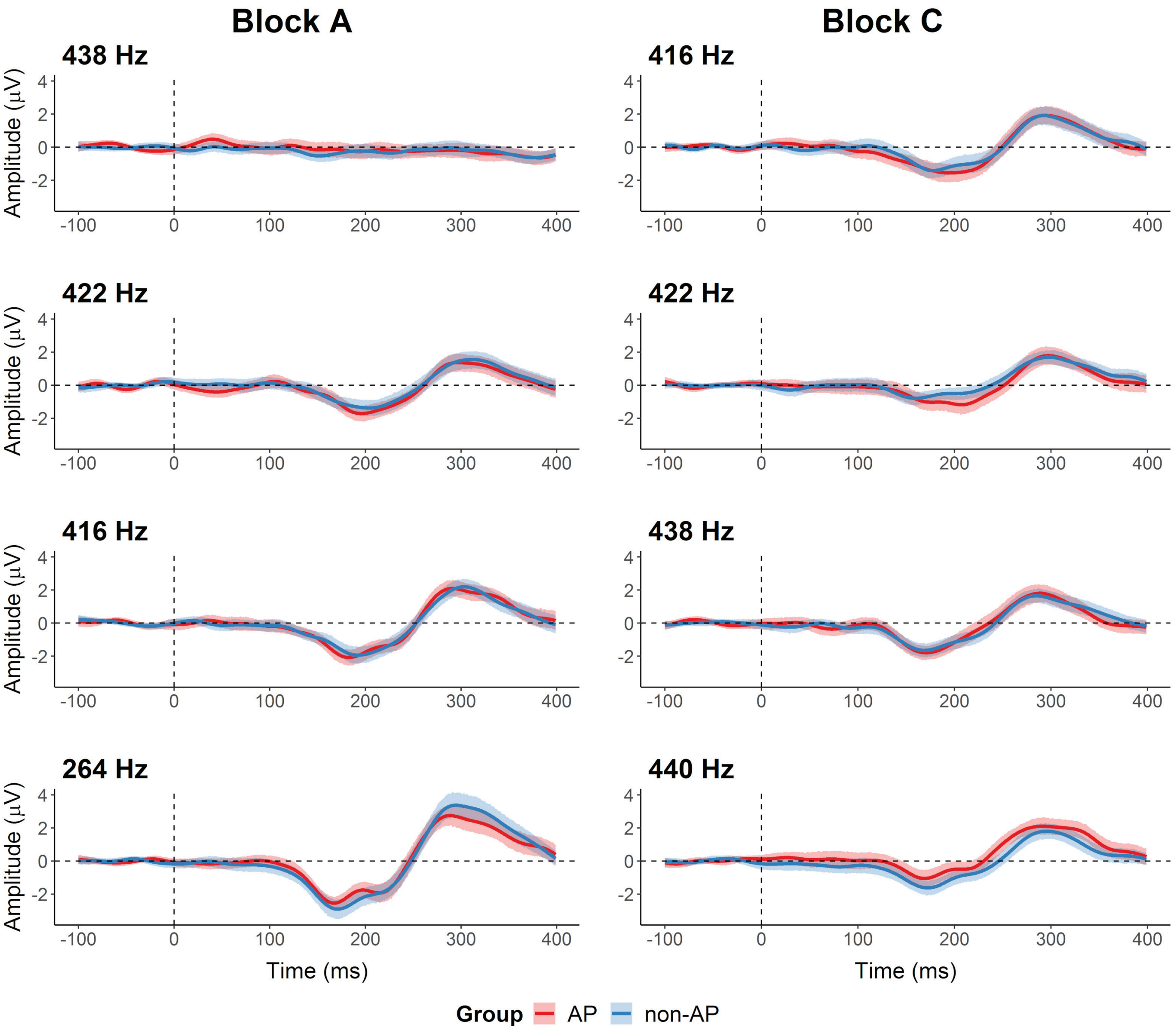
Table 2.

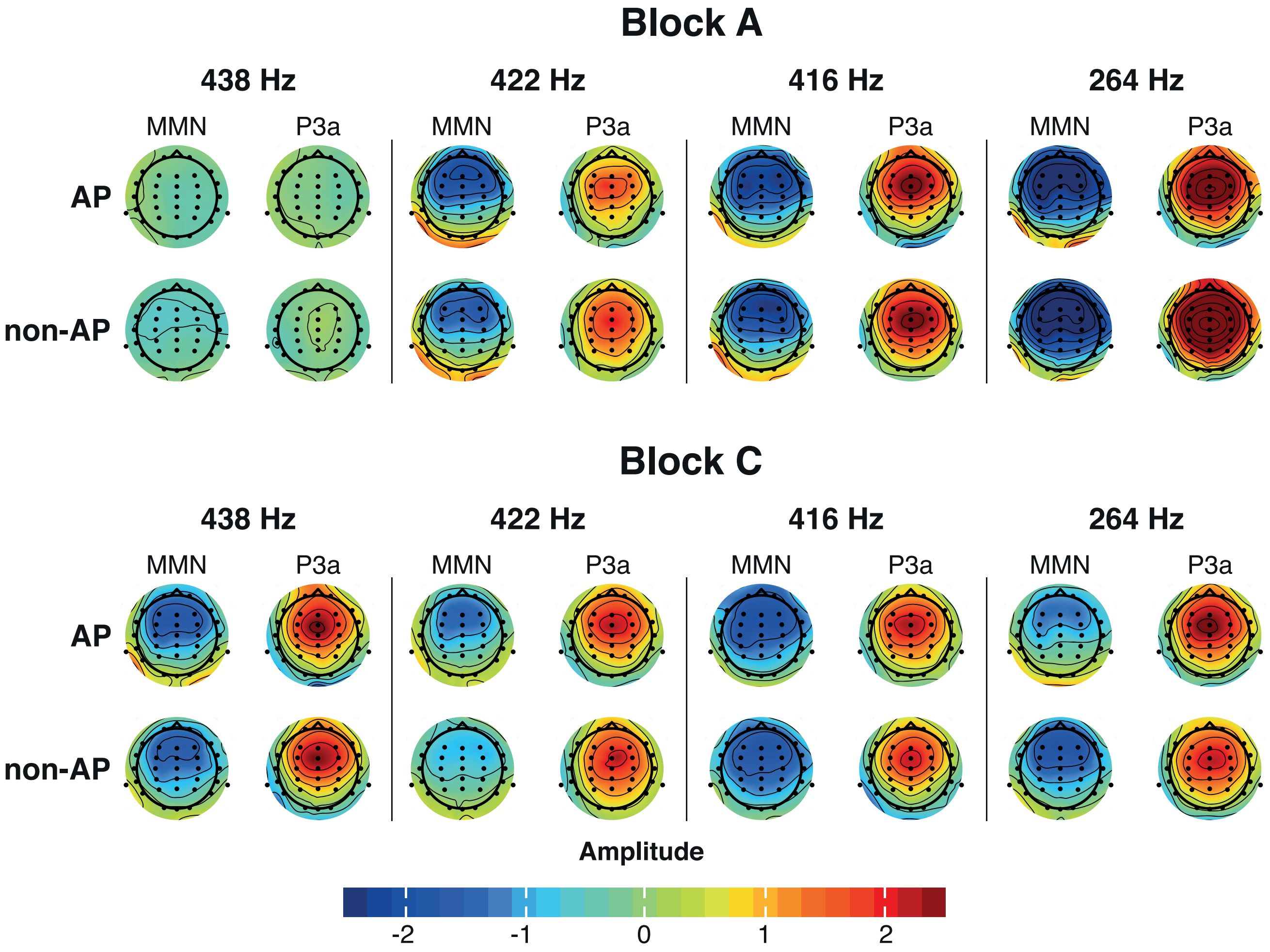
Study Design

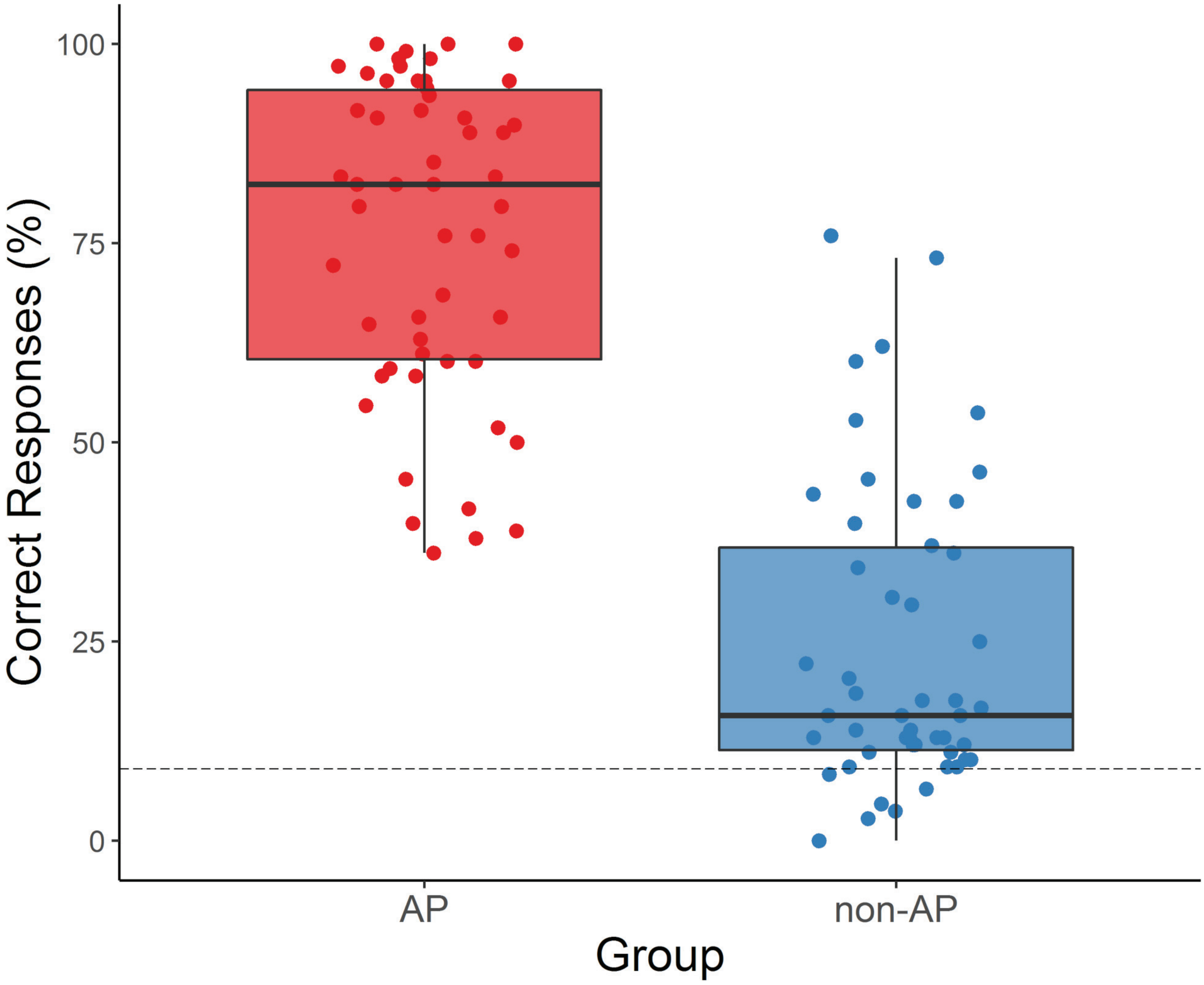
	Standard Tone	*Deviant Tones*			
Block A	440 Hz	438 Hz	422 Hz	416 Hz	264 Hz
Block C	264 Hz	416 Hz	422 Hz	438 Hz	440 Hz

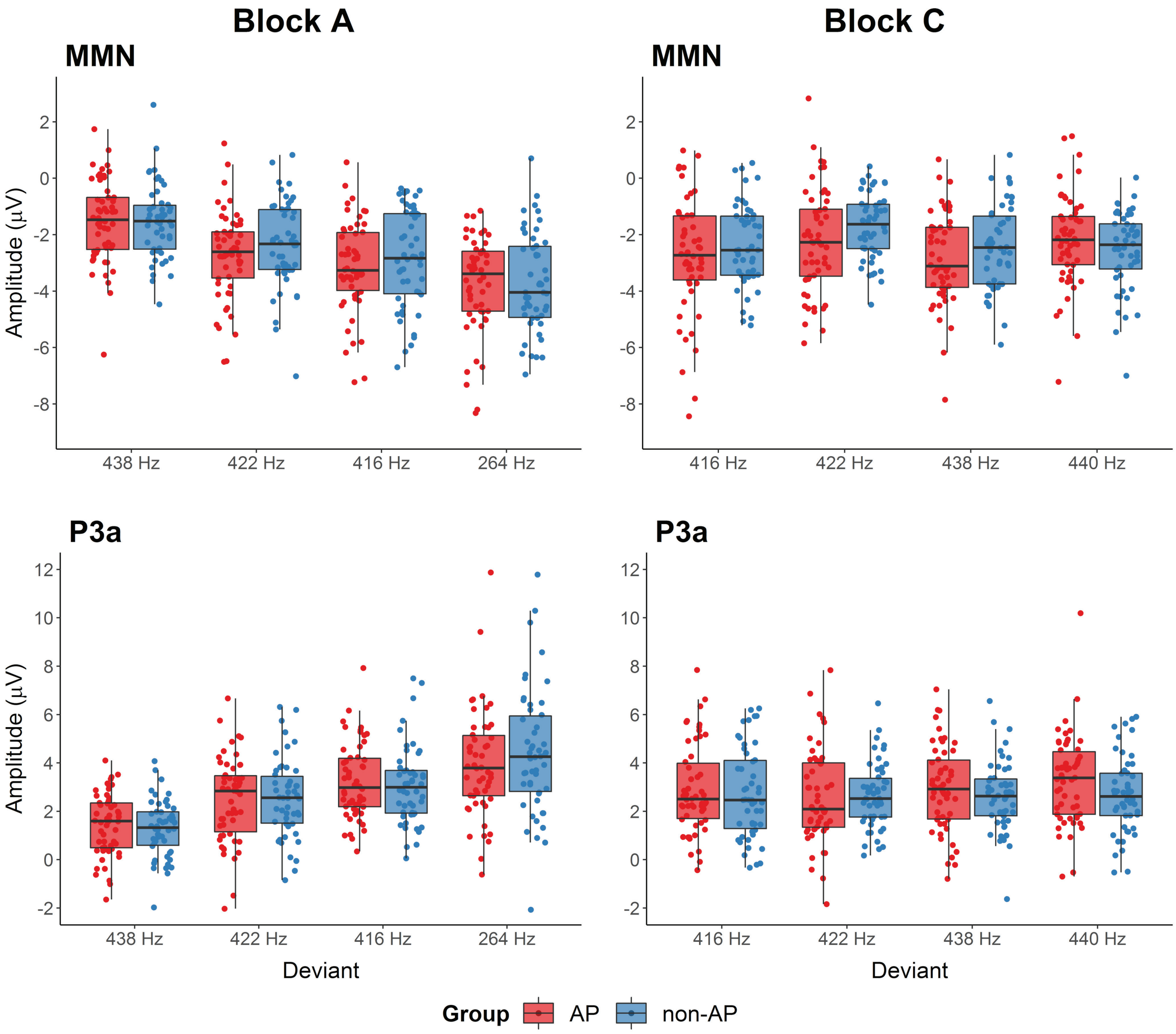
Deviant tones are listed from left to right according to increasing deviation magnitude.


Block A
2.5
0.0
-2.5
-5.0
Amplitude (µV)
-100
0
100
200
300
400
Time (ms)
Deviant
264 Hz
416 Hz
422 Hz
438 Hz
Block C
4
2
0
-2
-4
Amplitude (µV)
-100
0
100
200
300
400
Time (ms)
Deviant
440 Hz
438 Hz
422 Hz
416 Hz


Block A
Block C
438 Hz
416 Hz
422 Hz
422 Hz
416 Hz
438 Hz
264 Hz
440 Hz
Amplitude (µV)
4
2
0
-2
-100
0
100
200
300
400
Time (ms)
Group
AP
non-AP


Block A
438 Hz
422 Hz
416 Hz
264 Hz
MMN
P3a
AP
non-AP
Block C
Amplitude
-2
-1
0
1
2


Correct Responses (%)
100
75
50
25
0
AP
non-AP
Group


Block A
Block C
MMN
MMN
Amplitude (μV)
2
0
-2
-4
-6
-8
438 Hz
422 Hz
416 Hz
264 Hz
416 Hz
422 Hz
438 Hz
440 Hz
P3a
P3a
Amplitude (μV)
12
10
8
6
4
2
0
-2
438 Hz
422 Hz
416 Hz
264 Hz
416 Hz
422 Hz
438 Hz
440 Hz
Deviant
Deviant
Group
AP
non-AP