Inharmonicity enhances brain signals of attentional capture and auditory stream segregation

Krzysztof Basiński[1*], Alexandre Celma-Miralles[2], David R. Quiroga-Martinez[2,3], Peter Vuust[2]

1. Auditory Neuroscience Laboratory, Department of Psychology, Medical University of Gdańsk, Gdańsk, Poland
2. Center for Music in the Brain, Department of Clinical Medicine, Aarhus University & The Royal Academy of Music Aarhus/Aalborg, Universitetsbyen 3, Building 1710, Aarhus C 8000, Denmark
3. Department of Psychology, University of Copenhagen, Copenhagen, Denmark.

* corresponding author

Abstract

Harmonicity is an important feature for auditory perception. However, the neural substrates of processing inharmonic sounds remain unclear. Here, we systematically manipulated the harmonicity of sounds by introducing random jittering to their frequencies. Using electroencephalography, we studied the effect of inharmonicity on markers of auditory prediction errors — mismatch negativity (MMN) and P3a — in a roving oddball paradigm. Inharmonic sounds with a constant jittering pattern generated similar MMN and stronger P3a responses than harmonic sounds. In contrast, MMN responses became undetectable when the jittering pattern changed between consecutive sounds, suggesting that prediction errors are weighted by sequential but not spectral uncertainty. Interestingly, inharmonic sounds generated an object-related negativity, a response associated with the segregation of auditory objects. Our results suggest that inharmonicity induces the segregation of the auditory scene into different streams, captures attention, and gives rise to specific neural processes that are independent from the predictive mechanisms underlying sequential deviance detection.

Introduction

Harmonic sounds consist of acoustic waves containing frequencies that are integer multiples of a common fundamental (F0). Many ecologically salient sounds, such as human voices, animal calls or music, are highly harmonic. Conversely, if the acoustical wave consists of frequencies that are not integer multiples of a common fundamental, it is said to be inharmonic. These sounds are often described as noises, sizzles, pops or rattles[1]. Harmonicity has been suggested to underlie auditory scene analysis[2] and help separating the source of relevant sounds from background noise[3], forming a basic organizational principle of the auditory cortex[4–6]. In language, violations of harmonicity impair the ability to understand concurrent speech and

cause listeners to lose track of target talkers[7]. In music, inharmonic sounds can impair pitch-related abilities such as interval detection[8]. Similarly, inharmonicity disrupts the ability to compare pitch across time delays, which suggests a role in memory encoding[9].

Harmonicity can be understood in the light of predictive processing theories, which posit that perception relies on the brain forming top-down predictions about the incoming stimuli and their causes[10–17]. These predictive processing accounts assume that predictions stem from an internal, generative model that is constantly updated by statistical regularities in the incoming sensory information. An 'error' or 'surprise' response (i.e., *prediction error*) occurs when predictions do not fit the sensory input, which can be used to update the contents of the generative model[12,18]. According to this theory, prediction errors are weighted by uncertainty (or precision) of the context, adjusting the relevance given to sensory inputs, allowing perceptual inference under uncertainty[19–21]. Therefore, if a signal is imprecise or unreliable (as in, for example, seeing in dark conditions or hearing in noisy environments), any prediction errors arising from it would be down-weighted and less likely to influence future predictions. While precision-weighting in the auditory domain has been explored empirically (see[13] for a comprehensive review), the details of this process remain unclear.

Here, we hypothesize that harmonicity might be one of the relevant auditory features (among others, such as duration or intensity) that drive precision-weighting of prediction errors. Harmonic sounds have lower information content (*lower entropy*[22]) than inharmonic sounds, as the spectrum of an ideally harmonic sound will consist only of integer multiples of F0. Thus, many aspects of the sound can be reliably described using only one piece of information: the F0. Conversely, inharmonic sounds have higher information content (*higher entropy*). Consequently, any prediction errors produced by inharmonic sounds should have lower precision and would therefore be less likely to influence future expectations[23]. In the case of inharmonic sounds, precision-weighting might affect prediction errors not only by weakening pitch percepts but, more generally, by down-weighting sensory evidence about the spectral content of sounds.

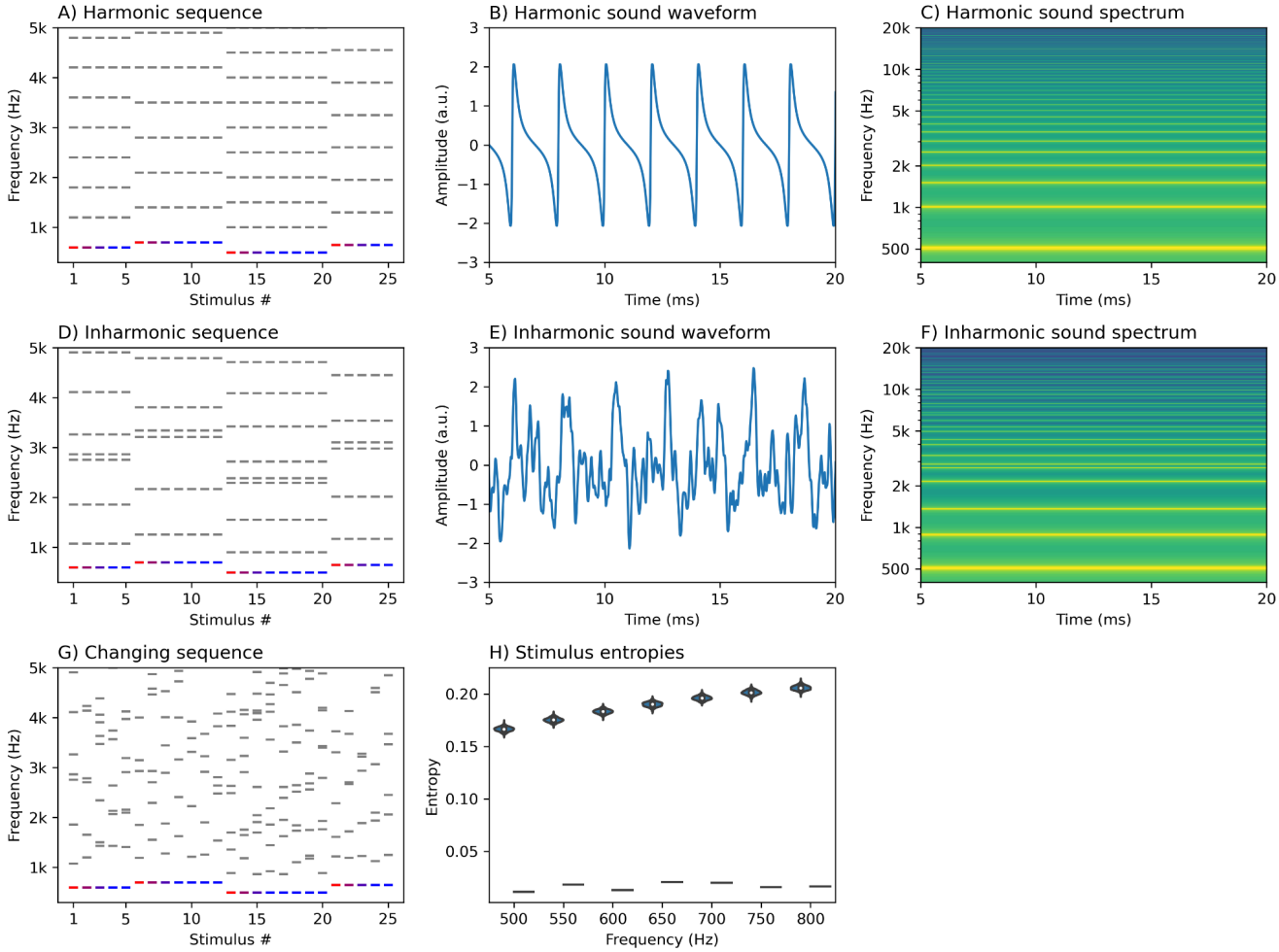
Cortical responses associated with prediction errors can be quantified in event-related potentials (ERPs) using electroencephalography (EEG). Mismatch negativity (MMN) is a widely studied response to a deviation in an otherwise repetitive train of sensory stimuli[24]. It is a well established electrophysiological trace of neural activity associated with precision-weighted prediction errors in the cortical auditory system[12,25–28]. Another neural response relevant to predictive coding is the P3 component, elicited when individuals shift attention (i.e. P3a) or consciously detect deviant stimuli (P3b[29]). In this study, we hypothesize that if harmonic sounds index precision as proposed by predictive processing theories, they should produce larger MMN and P3 responses to pitch deviants in comparison to inharmonic sounds.

Consistent with this idea, we previously showed that MMN responses were more prominent in harmonic piano tones than inharmonic hi-hat cymbal sounds[30] . However, due the naturalistic nature of the stimuli, we could not entirely rule out if other factors, such as the presence of a stable pitch percept, rise and decay of the sound envelope or other spectral differences, played

a role in the modulation of mismatch responses. To address this issue, here we present an EEG experiment using a passive roving oddball paradigm that carefully controlled the harmonic structure of sounds to isolate and measure the effect of harmonicity on auditory prediction errors. We recorded mismatch responses to sounds in three conditions: harmonic, inharmonic and changing. In the *harmonic* condition, sounds were regular harmonic complex tones. In the *inharmonic* condition, we introduced inharmonicity by jittering the frequencies above F0 and applied the same pattern of jitters to all tones in the sequence, thereby increasing the spectral uncertainty of the sounds. Note that the *inharmonic* condition introduces the same kind of uncertainty to the spectrum of each sound. However, MMN responses are also affected by the sequential uncertainty of the sound stream (i.e, what sound follows next). To investigate whether these two types of uncertainty interact, we introduced a *changing* condition in which a different jitter pattern was assigned to each individual sound. This increased sequential uncertainty, and made predictions of subsequent sounds harder. We hypothesize here that the amplitudes of MMN and P3a would be affected by the spectral and sequential uncertainty of the tones. This would be shown as reduced prediction errors in inharmonic sounds, especially when the spectral content is repetitively jittered in the sequence.

Results

Paradigm outline

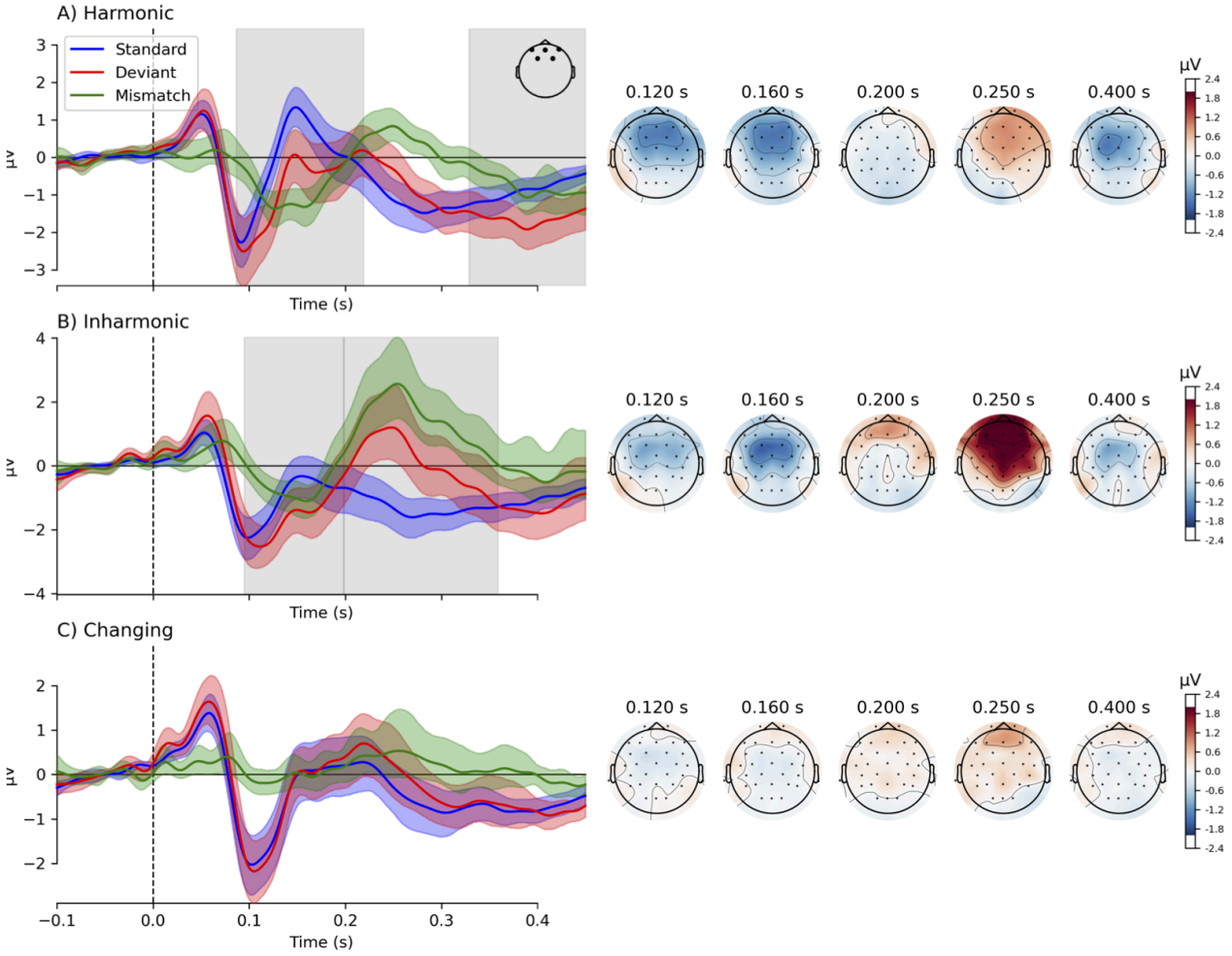


[F**igure 1. Paradigm outline and stimuli characteristics.** Panel A provides a symbolic representation of a roving sequence in the harmonic condition. The fundamental frequency is shown in blue or red, while the upper harmonics are shown in grey. Shades of red represent theoretical deviants, which progressively become standards (blue). Panels B and C present the waveform and spectrum of an example harmonic sound (F0 = 500 Hz). Panels D-F show the sequence representation, waveform and spectrum for the inharmonic condition. Notice that while the distribution of the partials becomes uneven in the inharmonic condition, the pattern of jittering is carried through from one fundamental frequency to the next. Conversely, Panel G shows the sequence representation for the changing condition, where a different jittering pattern is applied to each sound. Panel H shows entropy calculations for harmonic (dashes) and inharmonic (violin plot distributions) sounds for each F0. Note that while we calculated entropies for 1000 different inharmonic sounds present in our sound pool, there was only one harmonic sound for each frequency (thus producing a single entropy value instead of a distribution).]

We used a version of the roving oddball paradigm to generate mismatch responses[31,32]. In a typical roving paradigm, several sounds are presented at a specific frequency, followed by a set of sounds at a different, randomly chosen frequency, which in turn were followed by another set, and so on. The sound immediately after the frequency shift is the deviant sound, but it eventually becomes the standard after a few repetitions. The participant listens to the sounds passively while watching a silent movie. In this study, instead of pure tones we used (in)harmonic complex tones with the roving reflected in changes to the fundamental frequency of the complex. We used the roving paradigm because it ensures that the standards and deviants have the same physical properties (i.e., deviance is induced by the context of the sequence). It also allows for systematic investigation of the effect of the amount of frequency change on the mismatch response, by taking advantage of the random changes to frequencies. Finally, the roving paradigm allows for investigating if mismatch responses are sustained after the first deviant in the sequence by analyzing the second- and third-order deviants (shades of red in Figure 1a, 1d, 1g).

In order to investigate the effect of harmonicity on mismatch responses we presented sounds in three conditions. In the *harmonic* condition all sounds were harmonic complex tones (consisting of F0 and its integer multiples, Figures 1a-c. In the *inharmonic* condition we introduced a random jittering pattern to each frequency above the F0 (Figures 1d-f). Here, the same jittering pattern was applied to each sound within a set (Figure 1d). In the *changing* condition, a different jittering pattern was applied to each sound. Crucially, in each of the conditions the F0 remained unchanged by the jittering. We assumed that keeping the lowest frequency constant would induce the perception of a fundamental (F0) pitch strong enough to form a predictive model based on this percept. Approximate entropy, a measure of the amount of information contained in a time series, was on average lower for harmonic (M = .02, SD = .01) than for inharmonic sounds (M = .19, SD = .01) and this difference was consistent irrespective of the F0 (Figure 1h).

Presence of MMN and P3

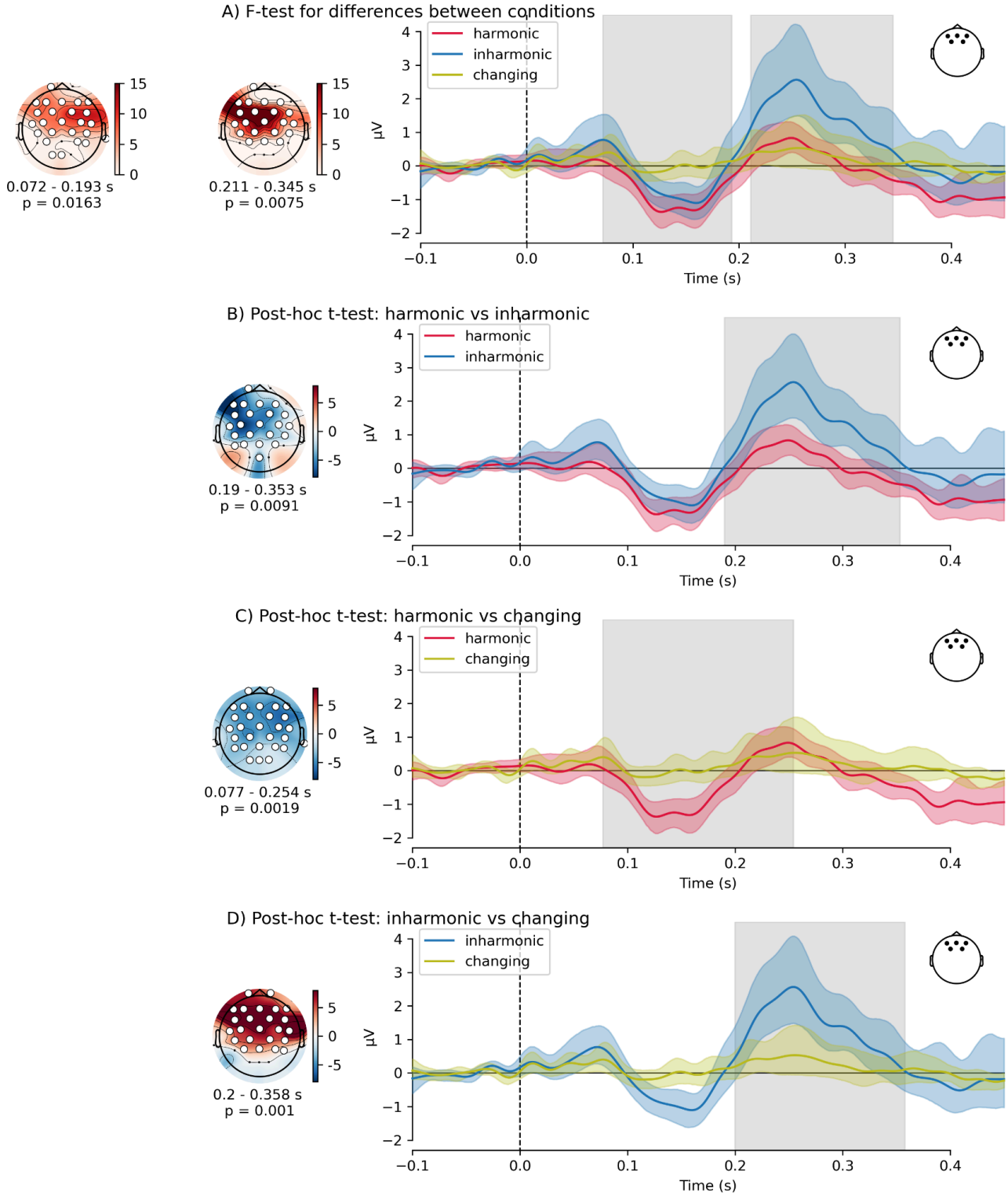


[**Figure 2. Event-related potentials and their topographies in three studied conditions.** Traces on the left show grand-average responses to standards (blue) and deviants (red), and the difference waves (green) for fronto-central channels (F3, Fz, F4, FC1, FC2). Color-shaded areas represent 95% confidence intervals and grey shades represent significant clusters in the mass-univariate analysis. Scalp topographies present the difference wave signal strengths at chosen latencies.]

First we investigated if the three conditions produced significant mismatch responses by testing the differences between standard and deviant responses using a mass-univariate approach. This allows for comparing ERPs without assumptions about spatio-temporal "windows" of activity[33]. We used cluster-based permutations, a non-parametric technique that tests for differences between conditions while controlling for multiple comparisons[34]. In the *harmonic* condition (Figure 2a), we found differences between standards and deviants corresponding to a cluster at 87 - 219 ms (26/30 sensors, p = .0002). Topographically, this mismatch response started as a frontocentral negativity at 120-180 ms typical of an MMN. A cluster at 216 - 286 ms,

reflecting a frontocentral positivity typical of a P3a response, was marginally significant (16/30 sensors, $p = .076$). We also found an additional negative cluster at 329 - 450 ms (21/30 sensors, $p = .002$). In the *inharmonic* condition (Figure 2b), we found differences between conditions corresponding to clusters in the data at 95 - 199 ms (23/30 sensors, $p = .002$) as well as at 198 - 359 ms (23/30 sensors, $p = .0001$). These overlapping clusters formed around typical latencies for MMN and P3a and had typical topographies. However, we found no differences between standards and deviants in the *changing* condition ($p > .05$ for all detected clusters), indicating that this condition did not produce any clear MMN or P3a responses (Figures 2c, Table S1).

Mass univariate analysis of group differences

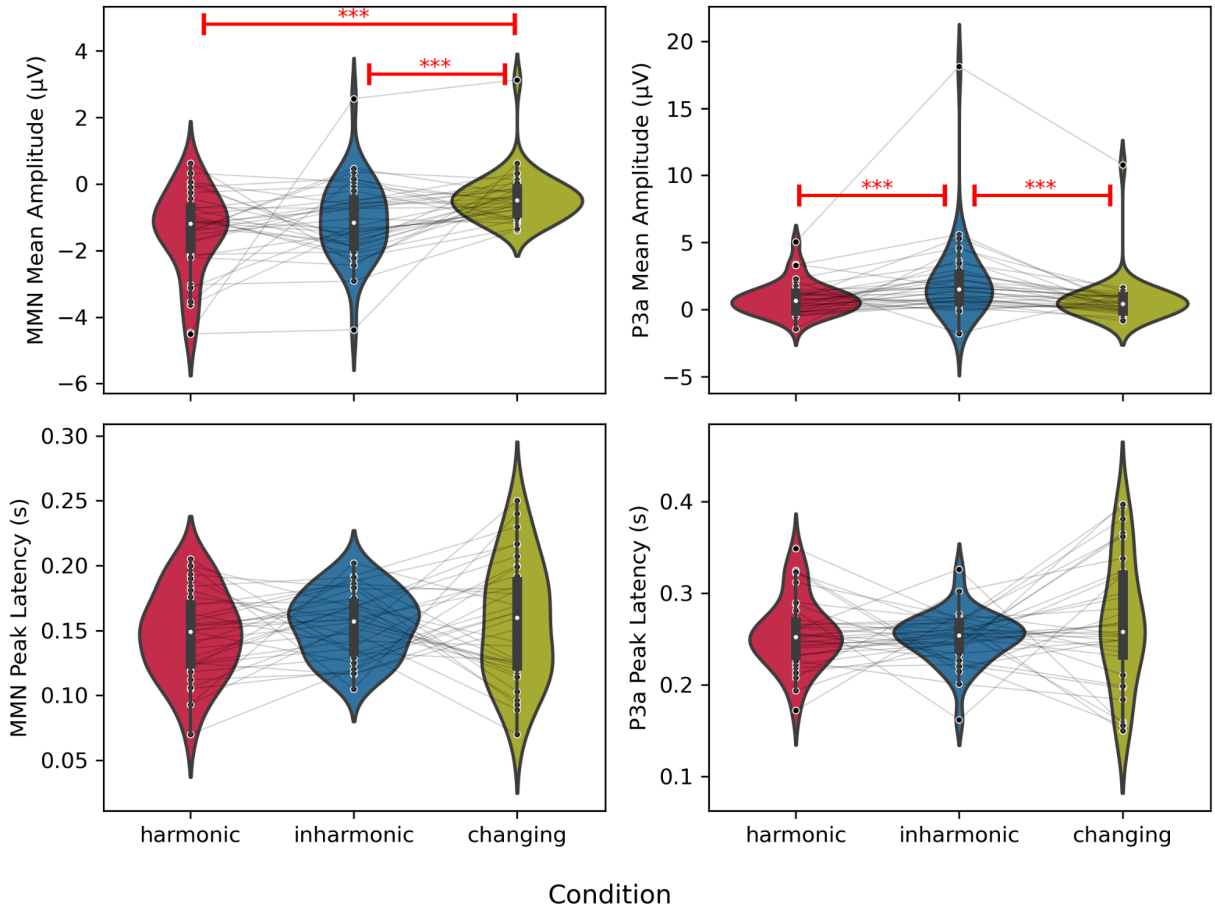


[**Figure 3. Event-related potentials in the harmonic, inharmonic and changing conditions.** Traces show grand average mismatch waves (deviant - standard, corresponding to green

mismatch traces in Figure 2) for the frontal channel activity (electrodes Fz, F3, F4, FC1, FC2). Panel A compares responses between the three conditions while panels B - D show post-hoc pairwise comparisons. Colored bands around the traces represent 95% confidence intervals around the mean. Grey rectangles indicate statistically significant clusters in the mass univariate analysis. Topographies show F-maps (panel A) or t-maps for post-hoc comparisons (panels B - D) for each significant cluster, with the temporal extent of the clusters indicated below each topography. White markers represent channels that comprise the cluster.]

To investigate the differences between the three conditions we used a cluster-based permutations approach, this time with an F-test. The test revealed differences between the conditions corresponding to clusters at 72 - 193 ms (25/30 sensors, p = .016) and at 211 - 345 ms (21/30 sensors, p = .008), indicating a main effect of condition (Figure 3a). Next, we ran post-hoc pairwise comparisons between conditions using cluster-based t-tests (Figure 3b). To account for multiple comparisons, we assumed a Bonferroni-corrected alpha level of p = .05/3 = .0166. We found significant differences between *harmonic* and *inharmonic* conditions in the P3a latency range (190-353 ms, a cluster comprising 22 sensors, p = .010) but not in the MMN range. Additionally, we found differences between *harmonic* and *changing* conditions for both MMN and P3a latency ranges (77- 254 ms, 24/30 sensors, p = .002). A later cluster (363-450 ms) was not significant after multiple comparisons correction (20/30 sensors, p = .033). Finally, we found differences between *inharmonic* and *changing* conditions in the P3a range (200-358 ms, 24/30 sensors, p = .001). An earlier cluster (98-198 ms) corresponding to the MMN latencies was marginally significant (21/30 sensors, p = .025).

Mean amplitude and peak latency



[**Figure 4. Peak measures differences in harmonic, inharmonic and changing conditions.** Panels show MMN mean amplitude (upper-left), P3a mean amplitude (upper-right), MMN peak latency (lower-left) and P3a peak latency (lower-right). Violin plots show distributions of obtained results. White dots represent the median, narrow rectangles represent quartile 2 and 3 ranges, vertical lines represent minima and maxima. Grey lines connect observations from the same participant. Stars (***) and red bars represent statistically significant post-hoc comparisons (all p-values < .001).]

Next, in order to investigate the effect of harmonicity on the amplitudes and latencies of mismatch responses, we calculated participant-wise mean amplitudes and peak latencies in the latency ranges for both MMN and P3a (Figure 4). For each of these measures we constructed linear mixed models with intercept and condition (*harmonic*, *inharmonic*, *changing*) as fixed effects and a random intercept for each participant (*m1*). We compared these models against null models that contained only the fixed intercept and a random intercept for each participant (*m0*).

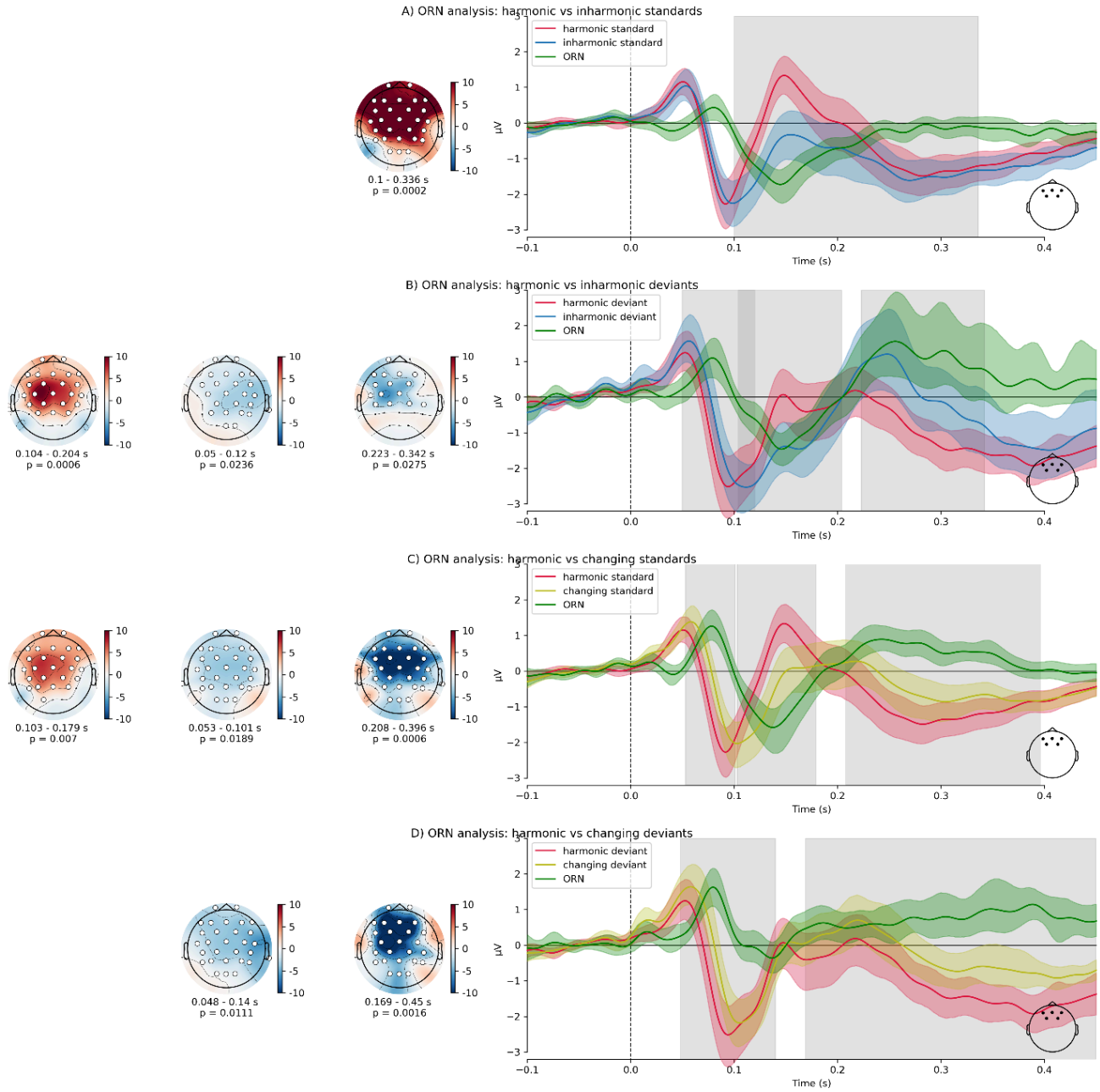
For MMN mean amplitudes, *m1* performed significantly better than *m0* (AIC_{m0} = 336.4, AIC_{m1} = 325.7, $Chi^2(2) = 14.71$, $p = .0006$). Post-hoc comparisons of estimated marginal means revealed

a significant difference between harmonic and changing conditions (contrast estimate = -1.03, SE = .27, t(67.8) = -3.84, *p* = .0008). Similarly, there was a significant difference between inharmonic and changing conditions (contrast estimate = -.70, SE = .27, t(67.8) = -2.57, *p* = .033). The *harmonic - inharmonic* contrast was not statistically significant (contrast estimate = -.34, SE = .27, t(67.8) = -1.24, *p* = .43). Taken together, these results suggest that there are no substantial differences in the MMN amplitude between *harmonic* and *inharmonic* sounds. However the MMN is stronger (more negative) for both *harmonic* and *inharmonic* than for *changing* sounds.

For P3a mean amplitudes, *m1* also performed significantly better than *m0* (AIC_{m0} = 458.8, AIC_{m1} = 444.0, $Chi^2(2)$ = 18.80, *p* < .0001). Post-hoc comparisons revealed significant differences between harmonic and inharmonic conditions (contrast estimate = -1.37, SE = .36, t(68) = -3.85, p = .0007) as well as between inharmonic and changing conditions (contrast estimate = 1.45, SE = .36, t(68) = 4.07, *p* = .0003). No significant differences were found for the harmonic - changing contrast (contrast estimate = .08, SE = .36, t(68) = .23, *p* = .97). These results suggest a significant effect of harmonicity on P3a amplitude, such that the P3a for inharmonic sounds is greater than for both harmonic and changing sounds.

For MMN peak latency measures, *m1* did not perform significantly better than *m0* (AIC_{m0} = -390.4, AIC_{m1} = -388.0, $Chi^2(2)$ = 1.59, *p* = .45). The same was also the case for P3 peak latency (AIC_{m0} = -331.7, AIC_{m1} = -330.2 $Chi^2(2)$ = 2.49, *p* = .29). This indicates that harmonicity does not substantially influence the latency of MMN and P3a mismatch responses.

Object-related negativity



[**Figure 5. Event-related potentials comparison for object-related negativity (ORN) analysis.** Traces show grand average responses for harmonic, inharmonic and changing standards or deviants. Green trace shows the difference wave corresponding to the ORN. Panel A compares harmonic and inharmonic standards, panel B compares harmonic and inharmonic deviants, panel C compares harmonic and changing standards while panel D compares harmonic and changing deviants. Traces represent averaged frontal channel activity (electrodes Fz, F3, F4, FC1, FC2). Colored bands around the traces represent 95% confidence intervals around the mean. Grey rectangles indicate statistically significant clusters in the mass univariate

analysis. Topographies show t-maps for each significant cluster. White markers represent channels that comprise the cluster.]

Contrast	Latency (ms)	Polarity	Sensors	P value
Harmonic vs inharmonic standards	100 - 336	Negative	28/30	.0003
Harmonic vs inharmonic deviants	50 - 120	Positive	23/30	.021
	104 - 204	Negative	23/30	.008
	223 - 342	Positive	15/30	.025
Harmonic vs changing standards	53 - 101	Positive	29/30	.023
	103 - 179	Negative	26/30	.001
	208 - 396	Positive	29/30	.001
Harmonic vs changing deviants	48 - 140	Positive	27/30	.013
	169 - 450	Positive	26/30	.002

[**Table 1.** Cluster-based permutations on harmonic vs. inharmonic and harmonic vs. changing contrasts. Latencies indicate the temporal span of detected clusters while Sensors indicate how many channels contributed to each cluster. Note that only clusters with $p < .05$ are shown here. For a complete list of clusters detected in this analysis, see Table S2.]

Beyond mismatch responses, the N1-P2 complex patterns in the *inharmonic* and *changing* condition differed noticeably from the *harmonic* condition (Figure 2). A possible candidate for this type of response is the object-related negativity (ORN), an ERP component associated with a separation of concurrently presented sounds into distinct auditory objects[35–37]. Here, we investigated the possibility that inharmonic sounds gave rise to the ORN by contrasting the responses to *harmonic* vs *inharmonic* and *harmonic* vs *changing* sounds, for both standards and deviants. To this end, we performed cluster-based permutations with a *t*-test of differences between conditions (Figure 5, Table 1). We found differences corresponding to negative clusters in the data at latencies consistent with the ORN in *harmonic* vs *inharmonic* contrasts (both for standards and for deviants) and in *harmonic* vs *changing* contrasts (only for standards) (see Table S2 for a complete list of clusters detected in this analysis).

To investigate the relationship between the ORN and the deviance detection processes related to MMN and P3a, we extracted the P2 mean amplitudes of both standards and deviants in each condition, averaged in a window around a global grand-average P2 peak latency (141 ms). We constructed a linear mixed model to predict P2 mean amplitude with *condition*, *deviance* (standard/deviant) and *condition-by-deviance* interaction as fixed effects and participant-wise random intercepts (*m2)*. We compared this model with models that contained only intercept as predictor (*m0*) and intercept and condition (*m1*) as predictors, besides the participant-wise

random intercepts. The model *m2* performed significantly better than both m0 ($Chi^2(2) = 81.28$, $p < .0001$) and m1 ($Chi^2(2) = 41.76$, $p < .0001$); $AIC_{m0} = 707.1$, $AIC_{m1} = 671.6$, $AIC_{m2} = 635.8$. ANOVA on *m2* revealed significant effects of condition ($F(2,170) = 31.38$, $p < .0001$), deviance ($F(1,170) = 37.15$, $p < .0001$) and a significant condition-by-deviance interaction ($F(2,170) = 7.38$, $p = .001$). Post-hoc contrasts revealed that P2 amplitudes were significantly higher for standards than for deviants in the harmonic (contrast estimate = -1.12, SE = .19, t(170) = -5.79, p <.0001) and inharmonic (contrast estimate = -.83, SE = .19, t(170) = -4.262, $p = .0005$) conditions, but not in the changing condition (contrast estimate = -0.01, SE = .19, t(170) = -.51, $p = 0.99$) (Table S5).

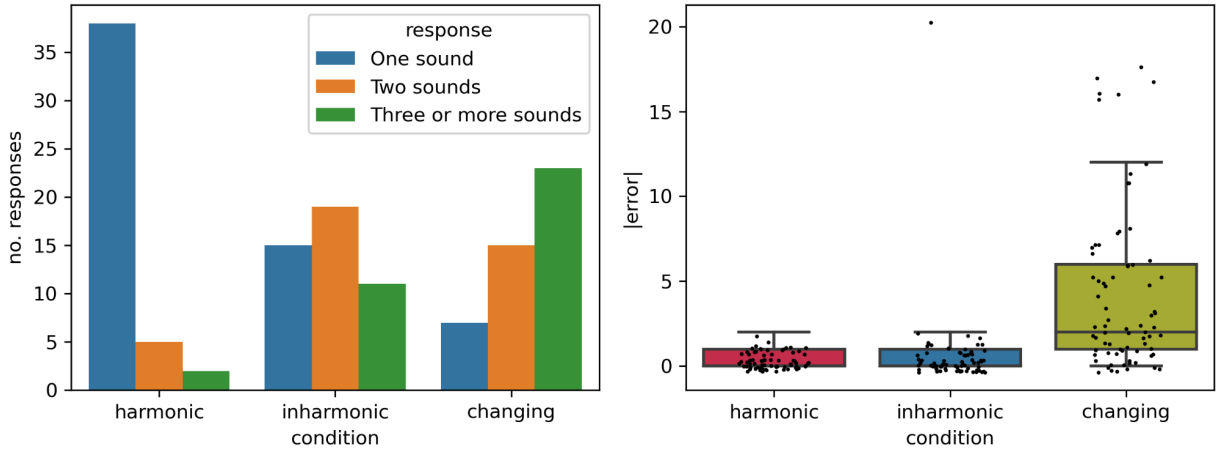
Taken together, these results suggest the presence of a frontocentral negativity in the *harmonic* vs. *inharmonic/changing* difference waves within a latency range of 100 - 200 ms (Figure 5). This pattern was found for all studied comparisons except for the *harmonic vs changing* deviants, likely due to the absence of a mismatch response in the latter (Figure 5D). Note that the inharmonic deviant had a more negative amplitude than the harmonic standard (contrast estimate = -2.06, SE = .19, t(170) = -10.61, p < .0001, see Table 1), suggesting the presence of inharmonicity effects in the absence of deviance effects.

Behavioral experiment

To investigate if the observed responses were related to perceiving multiple auditory objects in both inharmonic conditions, we ran a follow-up behavioral study. The participants were asked to listen to short sequences of either *harmonic*, *inharmonic* or *changing* sounds and judge if they heard one, two or three or more sounds at once. Results revealed that listeners were over sixteen times more likely to judge sounds as consisting of many different objects in the *inharmonic* condition (OR = 16.44, Est. = 2.80, SE = .57, p < .0001) and over sixty two times more likely in the *changing* condition (OR = 62.80, Est. = 4.14, SE = .64, $p < .0001$) in comparison to the *harmonic* condition (Fig 6A).

In the behavioral experiment, we also took the opportunity to examine if the participants were able to consciously perceive the F0 deviants in the *changing* condition. To this end, we exposed the participants to sequences of 20 sounds with roving F0 (following the same rules as in the EEG experiment) and asked them to count the number of deviants. For each sequence, we calculated a counting error metric. We constructed a linear mixed model to predict these errors in this task with intercept and condition as fixed effects and per-participant random intercepts (*m1*). This model performed significantly better than a null model (*m0*) that contained only intercept as predictor and per-participant random effect ($Chi^2(2) = 69.12$, $p < .0001$); $AIC_{m0} = 1203.3$, $AIC_{m1} = 1138.2$. Estimated marginal means for the absolute errors in the *harmonic* (M = 0.31, SE = 0.40) and *inharmonic* (M = 0.62, SE = 0.41) condition were not significantly different (t(206) = -0.630, $p = .80$). However, the absolute error in the *changing* condition (M = 4.25, SE = 0.41) was significantly higher than in both *harmonic* (t(206) = 8.08, p < .0001) and *inharmonic* (t(206) = 7.42, $p < .0001$) conditions. There was no significant difference between *harmonic* and *inharmonic* conditions. These results suggest that while participants were able to generate a

predictive model and consciously perceive deviations from it in both *harmonic* and *inharmonic* conditions, this may not have been the case in the *changing* condition, where a predictive model did not seem to be formed and many sounds were considered deviants.



[**Figure 6. Behavioral experiment results.** Left panel shows the number of responses of each category between the three conditions. The right panel shows mean errors in the three conditions. These errors were calculated as the absolute value of the difference between participant-reported and ground truth number of deviants (lower values = more correct answers). Ground truth values ranged from 1 to 5 (M = 2.8, SD = .72). Dots represent single observations. Horizontal lines represent the median, boxes represent quartile 2 and 3 ranges, whiskers represent minima and maxima.]

Second- and third-order deviants

Next, we took advantage of the roving paradigm to examine if the effects of harmonicity continue for sounds after the first deviant. To this end, we analyzed mismatch responses to second and third sounds after the frequency change (the second- and third-order deviants in each "rove"). We applied the same logic as in the main analysis, first using a mass-univariate approach and then focusing on peak measures. In the mass-univariate comparison of standard and deviant responses, we only found significant differences for the second-order deviant in the inharmonic condition at the latency range of P3a at 215 - 450 ms (19/30 sensors, p = .0042), but nothing at the latency range of MMN. No statistically significant differences were found in the harmonic and changing conditions. Similarly, no significant differences appeared for the third-order deviant in any condition (p > .05 for all clusters). This indicates that the P3a response for inharmonic sounds continued to be detectable in the second-order (but not third-order) deviants, while there was no detectable MMN for these extra deviants. However, when we compared the three conditions using a cluster-based permutations F-test, we found no significant differences for second- as well as third-order deviants (p > .05 for all clusters, Table S3).

We investigated this result further using linear mixed modeling for mean amplitude and peak latency, with participant-wise random intercepts. For the second-order deviant, we found that the model *m1,* which predicted P3a mean amplitude with condition, performed significantly better than the null model *m0*, (AIC_{m0} = 344.6, AIC_{m1} = 340.4, $Chi^2(2)$ = 8.21, p = .016). Post-hoc comparisons revealed significant differences between harmonic and changing (contrast estimate = .67, SE = .26, t(66) = 2.55, p = .035), as well as inharmonic and changing conditions (contrast estimate = .81, SE = .26, t(66) = 3.07, p = .009). No significant differences were found for the harmonic-inharmonic contrast. In the case of third-order deviants, condition did not improve model performance for any of the studied peak measures (p > .05 for all model comparisons). These results suggest that the P3a response gets carried over to the second-order deviant and is stronger for both harmonic and inharmonic sounds than for changing sounds.

Frequency shifts

Finally, we investigated if the effects that have been observed thus far are moderated by the amount of change (deviance) of the F0. In the roving paradigm, the F0 changes randomly in 50 Hz increments from 50 Hz to 300 Hz, both up and down. We extracted the amount of F0 change associated with each first-order deviant and entered it as a fixed effect into the linear mixed models that were analyzed previously. Models containing the main effects of condition and frequency shift (*m2*) performed significantly better than models that included just the condition (*m1*) for MMN mean amplitude (AIC_{m1} = 3470, AIC_{m2} = 3467, $Chi^2(2)$ = 5.15, p = .023) and P3 mean amplitude (AIC_{m1} = 3712, AIC_{m2} = 3703, $Chi^2(2)$ = 11.33, p = .001). However, models containing the interaction of condition and frequency shift (*m3*) did not perform significantly better than *m2* for all studied measures (all p values > .05, see Supplementary Table 4 and Supplementary Figure 2). These results indicate that the effects of harmonicity established in the previous analyses are not moderated by the amount of shift in the fundamental frequency.

Discussion

In this study we showed that harmonicity influences the brain's mismatch responses to both expected and unexpected sounds. Contrary to our hypothesis, inharmonic sounds with a constant jittering pattern (the *inharmonic* condition) generate MMN responses comparable to those of harmonic sounds, and elicit P3a responses that are stronger than in the other two conditions (despite the passive listening nature of the task). In contrast, MMN responses become undetectable when the jittering changes between sounds (the *changing* condition), suggesting that sequential, but not spectral uncertainty induces the precision-weighting effect. Interestingly, the ERPs to both standards and deviants differed between harmonic and inharmonic sounds, suggesting that inharmonicity elicits an ORN response. This result can be further explained by behavioral data, suggesting that for inharmonic sounds listeners are more likely to perceive more than one auditory object at the same time. Overall, our results suggest that inharmonicity does not act as a source of uncertainty as conceived by classic predictive processing theories. Instead, inharmonicity seems to induce the segregation of the auditory

scene into different streams, capturing attention (as reflected in the P3a) and giving rise to specific neural processes that are independent from the predictive mechanisms underlying sequential deviance detection and the MMN.

MMN results

The MMN in the *inharmonic* condition did not differ significantly from the *harmonic* condition. This result is not consistent with the precision-weighting hypothesis, as inharmonic sounds carry more information and should theoretically yield predictions of lower precision[38]. Instead, it suggests that the MMN is sensitive to the uncertainty present in the sequence of consecutive stimuli. This is evidenced by the fact that the MMN was undetectable in the *changing* condition, where spectral uncertainty introduced by inharmonicity was coupled with sequential uncertainty introduced by random jittering of consecutive sounds. In the *inharmonic* condition all partials in the complex tone changed with every deviant, however the relationship between the frequencies remained the same. This would mean that the MMN is sensitive to more global (context-dependent) uncertainty and not to the uncertainty generated by the introduction of inharmonicity.

This result relates to a larger issue of how precision-weighting is related to MMN, an ERP that is thought to reflect prediction error responses[25] (Garrido et al., 2009). In general, precision-modulated evoked responses to unexpected stimuli could theoretically exhibit both larger and smaller amplitudes. Smaller ERP amplitude could result from a smaller predicted difference between the standard and a deviant (smaller prediction error). However, it could also result from a larger predicted difference that is down-weighted by precision[13,39]. Recent simulation work has shown that these two cases could in practice be disassociated, because any change to precision-weighting would necessarily be accompanied by a change in the latency of the ERP peaks[40]. We have not found any evidence of latency effects on any of the studied components.

The results of this study show that inharmonic sounds with *changing* jitter patterns produce weaker mismatch responses than harmonic sounds. This is consistent with the precision-weighting hypothesis, as the *changing* condition has higher sequential uncertainty, making the inference about pitch of each consecutive sound more uncertain (however see *P3a, ORN and multiple auditory objects* section for an alternative explanation). This is also in line with previous research suggesting a role of harmonicity in precision-weighting for auditory features such as timbre, intensity and location[30]. However, the stimuli used in that study were sounds of musical instruments (piano and hi-hat) that differed in other acoustical properties apart from harmonicity (e.g., there were substantial differences in the spectra and in the amplitude envelopes). Furthermore, that study could not investigate pitch differences per se, due to the pitch percept being imperceivable for the hi-hat sound. Here we provide a systematic manipulation of harmonicity using synthetic sounds, ensuring that all other acoustic parameters apart from harmonicity of the stimuli remain the same between the conditions. The robustness

of these results is further reinforced by the lack of any interaction between frequency shift and condition, which indicates that the observed effects are not frequency-dependent.

P3a, ORN and multiple auditory objects

Surprisingly, we found that P3a amplitude was higher for inharmonic than for harmonic sounds. Following the conventional interpretation of the P3a[29], this result may indicate that inharmonic sounds capture attention more than harmonic sounds. A possible explanation might be that the auditory system treats inharmonic sounds as multiple, different auditory objects, while it collapses harmonic sounds into a single prediction about the F0. The F0 change associated with the deviant in the *inharmonic* condition would then require an update to not one but many different predictive models, resulting in overall stronger prediction error responses. In contrast, in the case of the changing condition where sounds constantly shift,, the auditory system attempts (and fails) to track changes in too many partials at once, because the only stationary frequency component is the F0. This logic could explain the P3a results in the present study, as well as the absence of both P3a and MMN components in the *changing* condition.

An unexpected result came from direct comparisons between responses to harmonic and inharmonic sounds (both standard or deviant). The introduction of inharmonicity produced a response pattern characterized by a more negative P2 peak. We interpret this as an ORN, an ERP related to the separation of concurrently presented sounds into distinct auditory objects [35–37]. The ORN is thought to reflect the auditory system performing concurrent sound segregation. Conventionally, it is elicited in paradigms with complex tones that include one mistuned harmonic or a harmonic with an asynchronous onset [36,41]. Other ORN-eliciting cues include dichotic pitch[42], simulated echo[43], differences or discontinuities in location[44], and onset asynchrony[45]. Importantly, the ORN is not a mismatch response, i.e. it does not arise in response to a violation of any global rule established with an oddball paradigm. This fact enabled us to evaluate the ORN using standard vs. standard and deviant vs. deviant comparisons. Importantly, the ORN is thought to arise from different neural sources than the MMN[46] however we did not see topographic differences in our data (see Appendix and Supplementary Figure 3). We found that it was elicited by both inharmonic standards and inharmonic deviants. One explanation of this result is that the auditory system interprets the inharmonic sound as multiple different sound sources and performs auditory object segregation. This is independent of pitch deviance detection indexed by the MMN. This conclusion is also reinforced by the presence of ORN in the harmonic standard vs. changing standard comparison. Finally, the P3a results discussed above also support this conclusion.

Limitations and future directions

When synthesizing sounds for this experiment, we used every frequency in the harmonic series up to the Nyquist limit at 24 kHz. This approach was motivated by the need to ensure that the harmonicity manipulation is applied equally throughout the frequency spectrum and that the F0 remains clearly audible throughout the sequence. We rejected frequency-jittering patterns that

could produce beating artifacts arising from two harmonics being too close to each other in frequency. Thus, our stimuli were broadband, spanning a large portion of the human hearing range. However, in psychophysical studies on pitch perception the stimuli are often high-pass or band-pass filtered specifically to reduce the amplitudes of the low frequency harmonics[8,47]. This procedure makes the pitch perception system focus on the so-called 'temporal fine structure' (the high-frequency content) of the sound and does not rely on the resolvability of the lower harmonics[48]. This issue can be addressed in future by experimentally varying the amount of low frequency information available.

A related issue is the fact that the lack of a significant MMN in the changing condition can be (at least in part) explained by the amount of jittering applied to the harmonics. We used jittering rates sampled from a uniform distribution in the range between -0.5 and 0.5 which corresponds to up to 50% change in the frequency. This change is quite salient perceptually and leads to a situation where the frequency deviants in the *changing* condition were not discernible from the continuously changing standards, as evidenced by the behavioral results. This value was chosen because it was used in a series of behavioral studies that established effects of harmonicity on pitch-related and memory tasks[8,9,49]. However, it is possible that with lower jitter values the MMN and P3a components would be apparent in the *changing* condition as well. Thus, the relationship between the strength of mismatch responses and jitter rate should be thoroughly addressed in future research. Relatedly, more research is needed to establish the behavioral thresholds of pitch detection as a function of jitter rate. The behavioral task used in this study was

It is important to note that any precision-weighting mechanisms in the brain can only make use of information that is already encoded by the peripheral sensory system. In this sense, a more valid metric of precision would be the entropy associated with the output from the cochlea. While any cochlear transformation would necessarily retain many aspects of the harmonic structure of sound, we acknowledge that the correlation between the entropy of the acoustical wave and the entropy of its neural representation is not perfect. Future research should address this issue, perhaps by using computational models of the cochlea and the auditory nerve.

In sum, this work suggests that inharmonicity does not serve as an index of precision-weighting for low level, short timescale auditory predictions, as evidenced by the MMN results. Instead, it encourages the auditory system to perform object segregation, as evidenced by the ORN results. The tracking of multiple objects (or multiple predictions) engages attention and leads to larger P3a responses to unexpected changes as long as multiple objects are trackable within a sequence (as in the *inharmonic* condition). This work offers a new perspective on the neural mechanisms underlying the auditory processing of low level acoustic features, as well as expanding previously established behavioral effects of inharmonicity on pitch perception and memory.

Methods

Participants

We recruited 37 participants for the EEG experiment. One participant completed the study but was removed from analysis due to audio equipment failure during the procedure. Another participant was removed from analysis because of high levels of noise present in the data (over 60% of epochs excluded in the *autoreject* procedure, see below). The final sample size was 35 participants (median age 27 years, 18 female). We included participants with no reported neurological or psychiatric illness, age between 18 and 45 years old, normal hearing, normal sight or corrected normal sight (e.g., contact lenses) and no use of medication that affects the central nervous system (e.g. opioids, pain medications). All participants received monetary compensation for their participation and signed a consent form. The study was approved by the department Institutional Review Board at Aarhus University (reference number DNC-IRB-2022-009) and followed the Declaration of Helsinki.

The number of participants was calculated with power analysis before the experiment. We used linear mixed models on a simulated dataset with the hypothetical differences between harmonic and inharmonic conditions (see *Stimuli and procedure*) set at 1 μV. This choice was informed by the results of a previous study where the difference in peak amplitude between harmonic piano and inharmonic hi-hat tones was around 1μV[50]. Here we aimed to be able to detect a similar difference between harmonic and inharmonic conditions, with an assumption that *changing* would not differ significantly from the *inharmonic* condition (see *Procedure* for the descriptions of the conditions). We ran this simulation 10 000 times for each N value in the range between 25 and 40. The analysis showed that in order to achieve a statistical power of 0.8, at least N = 33 participants were required.

For the behavioral study, 15 participants were recruited (median age 20 years, 13 female). Inclusion criteria were identical to the ones used in the EEG experiment. In this case, the participants did not receive monetary compensation.

Stimuli and procedure

The stimuli consisted of complex tones made by adding sine waves of varying fundamental and harmonic frequencies up to the Nyquist limit (24 kHz). Harmonics were added with equal amplitudes, in sine phase. Each tone had a duration of 70 ms and was amplitude modulated with 5 ms onset and offset ramps. To synthesize inharmonic stimuli, a procedure was adapted from McPherson and McDermott (2018). Each frequency in the harmonic series above F0 was jittered by a random value to introduce inharmonicity. The jitter value was calculated by multiplying the original frequency by a number drawn from a uniform distribution U(-0.5, 0.5). In order to avoid beating artifacts, rejection sampling was used in cases where two frequencies would be spaced closer than 30 Hz apart. In the inharmonic condition, the same jitter pattern was applied to all sounds in the stimulus block. In the inharmonic-changing condition, each stimulus train had a new, randomly generated jitter pattern (Fig. 1). A sound pool of one harmonic and one thousand inharmonic sounds per each F0 was generated offline before the

study as 16-bit .wav files with a sample rate of 48 kHz. All sounds were loudness-normalized to -12 LUFS using *pyloudnorm*.

Tones were presented using an oddball roving paradigm with an inter-stimulus interval of 600ms. Within each stimulus train, all standard tones had the same F0 and were followed by a train of stimuli with different F0 (chosen randomly from a 500 Hz - 800 Hz range in 50Hz intervals). In this paradigm, the first tone of a new stimulus train serves as the deviant stimulus and a potential source of mismatch negativity in context of prior stimuli. After a few repetitions, this deviant tone is established as the new standard until the next stimulus train is presented. The number of stimuli in a given train varied pseudo-randomly from 3 to 11, with 3 to 7 repetitions being four times more probable than 8 to 11. The entire procedure was divided into six blocks (two for each harmonicity condition), each 6 minutes long. Each block consisted of 600 sounds, with 98 deviants on average. The participants were asked not to move during the auditory stimulation, however there were short breaks between the blocks to provide rest and relaxation. All stimuli were administered passively and participants were watching a silent movie throughout the procedure. The stimuli were randomized and played back using PsychoPy (version 2022.1.3). Headphones (Beyerdynamic DT 770 PRO) were used for binaural sound presentation. The entire experiment lasted 80-90 minutes on average, including participant preparation and debriefing.

The behavioral experiment consisted of two parts. In the first part, participants listened to a sequence of 8 sounds at a stationary F0 and were asked if they perceived ‘one’, ‘two’ or ‘three or more’ sounds ‘at the same time, at any given moment’. Depending on the condition, the sequences could be either *harmonic*, *inharmonic* or *changing*, following the principles from the EEG experiment and using the exact same sound pool. Three sequences were presented for each condition for a total of nine trials per participant.

In the second part of the behavioral task, the participants listened to a sequence of 20 sounds with roving F0s following the rules of the EEG experiment. The task was to silently count the number of deviants (‘please count, how many times the sounds change their pitch’) and report the number after each sequence. This was preceded by a demo trial that included a harmonic sequence and informed the participant about the number of deviants in that sequence. The participants could repeat the demo trial multiple times until they become familiar with the task. Overall, five sequences were presented for each condition (*harmonic*, *inharmonic* or *changing*) for a total of 15 trials per participant. The entire experiment lasted about 5-7 minutes on average and was performed on a laptop running PsychoPy in a quiet room. The sounds were presented binaurally using Beyerdynamic DT 770 PRO headphones.

Experimental setup and EEG analysis

Scalp EEG potentials were recorded with a 32-channel active system (BrainProducts GMBH) at a sampling rate of 1000 Hz. All EEG signal processing was performed in Python v.3.11 using *MNE* v.1.5[51,52], *Numpy* v.1.24[53] and *Pandas* v.2.1[54]. *Matplotlib* v.3.7 was used for plotting[55].

Horizontal and vertical eye movements (EOG) as well as heart rate (ECG) were recorded with additional electrodes. Raw EEG data were high-pass filtered at 0.2 Hz, divided into epochs (from -100 ms to 450 ms) and entered into the *autoreject* algorithm v.0.4.2[56] for automatic bad channel and bad epoch selection. This method uses cross-validation and a robust evaluation metric to estimate optimal peak-to-peak thresholds for each EEG sensor. This thresholding is applied epoch-wise and the signals are either interpolated or (in case of many bad sensors) rejected. Afterwards, an independent components analysis (ICA) was performed in order to remove eye movement and heart-related artifacts. ICA components were marked for removal using an automated procedure based on data from EOG and ECG channels (the functions *find_bads_eog()* and *find_bads_ecg()* from the *MNE* package). However, these automatic choices were inspected visually. A maximum of four ICA components were removed from any participant. The ICA solution was applied to data before artifact rejection and *autoreject* was performed once more, as suggested by Jas et al.[56]. Overall, 6.9% of epochs were rejected. Finally, the epoched data were low-pass filtered at 30 Hz, re-referenced to the mastoids and baseline-corrected with a baseline of 100 ms pre-stimulus.

We treated the sound as a deviant if it was the first, second or third sound after the change of F0 in the roving paradigm. All non-deviant sounds were treated as standard apart from the first five sounds at the start of each experimental block. ERPs were estimated by averaging the responses to standard and deviant sounds across participants and across conditions. Difference waves were calculated by subtracting the standard from the deviant ERP, for each deviant type (first, second or third), each participant and condition. For ORN analysis, differences between harmonic and inharmonic or changing ERPs were calculated, separately for standards and (first-order) deviants. In the mass-univariate approach, entire ERPs were subject to cluster-based permutations analyses. In peak-based analyses, we extracted participant-wise latencies for MMN and P3 peaks in the fronto-central EEG channels. For MMN, we extracted the latency of the most negative peak between 70 ms and 250 ms. For P3a, we found the latency of the most positive peak between 150 and 400 ms. The mean amplitude of each component was calculated as the average signal amplitude in the 50 ms window centered at the peak.

Cluster-based permutations were performed using MNE's built-in functions. Thresholds were chosen automatically based on the number of valid observations and assuming a *p*-value of .05. For each test, 1024 permutations were performed. Linear mixed models were fitted in R v.4.4.1 [57] using *lme4* v.1.1 [58]. Model comparison was performed with a likelihood ratio Chi^2 test. Post-hoc comparisons were performed by comparing the estimated marginal means calculated with *emmeans* v.1.10[59]. *P*-values were corrected for multiple comparisons using the Tukey HSD method. Approximate entropies were calculated with *AntroPy* v.0.1.6[60]. Entropy was estimated and averaged for all inharmonic sounds within the sound pool.

The behavioural experiment was performed on a laptop computer running PsychoPy. The sounds were presented binaurally using Beyerdynamic DT 770 PRO headphones. Participants gave their responses by pressing the keyboard. Count error metrics were calculated by taking the absolute value of the difference between participant-reported and ground truth number of

deviants. A cumulative link mixed model regression with participant-wise random intercepts was used to estimate odds ratios in the first task. A linear mixed model was fit in the second task.

Data availability

The raw data from this experiment is freely available at https://doi.org/10.5281/zenodo.13939897

Code availability

Code that was used to perform this experiment and analyze the results is freely available at https://doi.org/10.5281/zenodo.15236581 .

Author contributions

KB: conceptualization, methodology, software, data analysis, investigation, data curation, funding acquisition, writing - original draft, review & editing, visualization. ACM: conceptualization, methodology, investigation, writing - review & editing. DRQM: conceptualization, writing - review & editing. PV: funding acquisition, supervision, writing - review & editing.

Competing interests

The authors declare no competing interests.

Acknowledgements

We would like to thank Thies Lübeck, Pelle de Deckere and Marc Barcelos for help in data collection and data preprocessing. We would also like to thank Alejandro Blenkmann and Andrew Oxenham for inspiring discussions that lead to some of the ideas presented in this paper. Krzysztof Basiński is supported by grants from the National Agency for Academic Exchange, Poland (Bekker, no. BPN/BEK/2021/1/00358/U/00001) and National Science Centre, Poland (Sonata, no. 2022/47/D/HS6/03323). Alexandre Celma-Miralles and Peter Vuust are funded by the Danish National Research Foundation (DNRF117) supporting the Center for Music in the Brain. David R. Quiroga-Martinez is supported by the Carlsberg Foundation (CF23-1491).

References

1. Oxenham, A. J. How We Hear: The Perception and Neural Coding of Sound. *Annu. Rev.*

Psychol. **69**, 27–50 (2018).

2. de Cheveigné, A. Harmonic Cancellation—A Fundamental of Auditory Scene Analysis. *Trends Hear.* **25**, 23312165211041422 (2021).
3. McDermott, J. H. The cocktail party problem. *Curr. Biol.* **19**, 1024–1027 (2009).
4. Feng, L. & Wang, X. Harmonic template neurons in primate auditory cortex underlying complex sound processing. *Proc. Natl. Acad. Sci.* **114**, E840–E848 (2017).
5. Wang, X. The harmonic organization of auditory cortex. *Front. Syst. Neurosci.* **7**, (2013).
6. Wang, X. Cortical Coding of Auditory Features. *Annu. Rev. Neurosci.* **41**, 527–552 (2018).
7. Popham, S., Boebinger, D., Ellis, D. P. W., Kawahara, H. & McDermott, J. H. Inharmonic speech reveals the role of harmonicity in the cocktail party problem. *Nat. Commun.* **9**, 2122 (2018).
8. McPherson, M. J. & McDermott, J. H. Diversity in pitch perception revealed by task dependence. *Nat. Hum. Behav.* **2**, 52–66 (2018).
9. McPherson, M. J. & McDermott, J. H. Time-dependent discrimination advantages for harmonic sounds suggest efficient coding for memory. *Proc. Natl. Acad. Sci.* **117**, 32169–32180 (2020).
10. Clark, A. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* **36**, 181–204 (2013).
11. Clark, A. *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. (Oxford University Press, 2015).
12. Friston, K. J. A theory of cortical responses. *Philos. Trans. R. Soc. B Biol. Sci.* **360**, 815–836 (2005).
13. Heilbron, M. & Chait, M. Great Expectations: Is there Evidence for Predictive Coding in Auditory Cortex? *Neuroscience* **389**, 54–73 (2018).
14. Hohwy, J. *The Predictive Mind*. (OUP Oxford, 2013).
15. Rao, R. P. N. A sensory–motor theory of the neocortex. *Nat. Neurosci.* **27**, 1221–1235

(2024).

16. Rao, R. P. N. & Ballard, D. H. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* **2**, 79–87 (1999).
17. Vuust, P., Heggli, O. A., Friston, K. J. & Kringelbach, M. L. Music in the brain. *Nat. Rev. Neurosci.* (2022) doi:10.1038/s41583-022-00578-5.
18. Friston, K. J. Does predictive coding have a future? *Nat. Neurosci.* **21**, 1019–1021 (2018).
19. de Lange, F. P., Heilbron, M. & Kok, P. How Do Expectations Shape Perception? *Trends Cogn. Sci.* **22**, 764–779 (2018).
20. Kanai, R., Komura, Y., Shipp, S. & Friston, K. Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. B Biol. Sci.* **370**, 20140169 (2015).
21. Yon, D. & Frith, C. D. Precision and the Bayesian brain. *Curr. Biol.* **31**, R1026–R1032 (2021).
22. Shannon, C. E. A Mathematical Theory of Communication. 55 (1948).
23. Basiński, K., Quiroga-Martinez, D. R. & Vuust, P. Temporal hierarchies in the predictive processing of melody − From pure tones to songs. *Neurosci. Biobehav. Rev.* **145**, 105007 (2023).
24. Näätänen, R., Paavilainen, P., Rinne, T. & Alho, K. The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clin. Neurophysiol.* **118**, 2544–2590 (2007).
25. Garrido, M. I., Kilner, J. M., Stephan, K. E. & Friston, K. J. The mismatch negativity: A review of underlying mechanisms. *Clin. Neurophysiol.* 11 (2009).
26. Koelsch, S., Vuust, P. & Friston, K. Predictive Processes and the Peculiar Case of Music. *Trends Cogn. Sci.* **23**, 63–77 (2019).
27. Lumaca, M., Trusbak Haumann, N., Brattico, E., Grube, M. & Vuust, P. Weighting of neural prediction error by rhythmic complexity: A predictive coding account using mismatch negativity. *Eur. J. Neurosci.* **49**, 1597–1609 (2019).

28. Winkler, I. & Czigler, I. Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *Int. J. Psychophysiol.* **83**, 132–143 (2012).
29. Polich, J. & Criado, J. R. Neuropsychology and neuropharmacology of P3a and P3b. *Int. J. Psychophysiol.* **60**, 172–185 (2006).
30. Quiroga-Martinez, D. R. *et al.* Enhanced mismatch negativity in harmonic compared to inharmonic sounds. *Eur. J. Neurosci.* **n/a**, (2022).
31. Cowan, N., Winkler, I., Teder, W. & Näätänen, R. Memory Prerequisites of Mismatch Negativity in the Auditory Event-Related Potential (ERP). *J. Exp. Psychol. Learn. Mem. Cogn.* **19**, 909–21 (1993).
32. Garrido, M. I. *et al.* The functional anatomy of the MMN: A DCM study of the roving paradigm. *NeuroImage* **42**, 936–944 (2008).
33. Groppe, D. M., Urbach, T. P. & Kutas, M. Mass univariate analysis of event‐related brain potentials/fields I: A critical tutorial review. *Psychophysiology* **48**, 1711–1725 (2011).
34. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* **164**, 177–190 (2007).
35. Alain, C., Arnott, S. R. & Picton, T. W. Bottom–up and top–down influences on auditory scene analysis: Evidence from event-related brain potentials. *J. Exp. Psychol. Hum. Percept. Perform.* **27**, 1072 (2001).
36. Alain, C., Schuler, B. M. & McDonald, K. L. Neural activity associated with distinguishing concurrent auditory objects. *J. Acoust. Soc. Am.* **111**, 990–995 (2002).
37. Bendixen, A., Jones, S. J., Klump, G. & Winkler, I. Probability dependence and functional separation of the object-related and mismatch negativity event-related potential components. *NeuroImage* **50**, 285–290 (2010).
38. Feldman, H. & Friston, K. J. Attention, Uncertainty, and Free-Energy. *Front. Hum. Neurosci.* **4**, (2010).

39. Bowman, H., Filetti, M., Wyble, B. & Olivers, C. Attention is more than prediction precision. *Behav. Brain Sci.* **36**, 206–208 (2013).

40. Bowman, H., Collins, D. J., Nayak, A. K. & Cruse, D. Is predictive coding falsifiable? *Neurosci. Biobehav. Rev.* **154**, 105404 (2023).

41. Weise, A., Schröger, E. & Bendixen, A. The processing of concurrent sounds based on inharmonicity and asynchronous onsets: An object-related negativity (ORN) study. *Brain Res.* **1439**, 73–81 (2012).

42. Hautus, M. J. & Johnson, B. W. Object-related brain potentials associated with the perceptual segregation of a dichotically embedded pitch. *J. Acoust. Soc. Am.* **117**, 275–280 (2005).

43. Sanders, L. D., Joh, A. S., Keen, R. E. & Freyman, R. L. One sound or two? Object-related negativity indexes echo perception. *Percept. Psychophys.* **70**, 1558–1570 (2008).

44. McDonald, K. L. & Alain, C. Contribution of harmonicity and location to auditory object formation in free field: Evidence from event-related brain potentials. *J. Acoust. Soc. Am.* **118**, 1593–1604 (2005).

45. Lipp, R., Kitterick, P., Summerfield, Q., Bailey, P. J. & Paul-Jordanov, I. Concurrent sound segregation based on inharmonicity and onset asynchrony. *Neuropsychologia* **48**, 1417–1425 (2010).

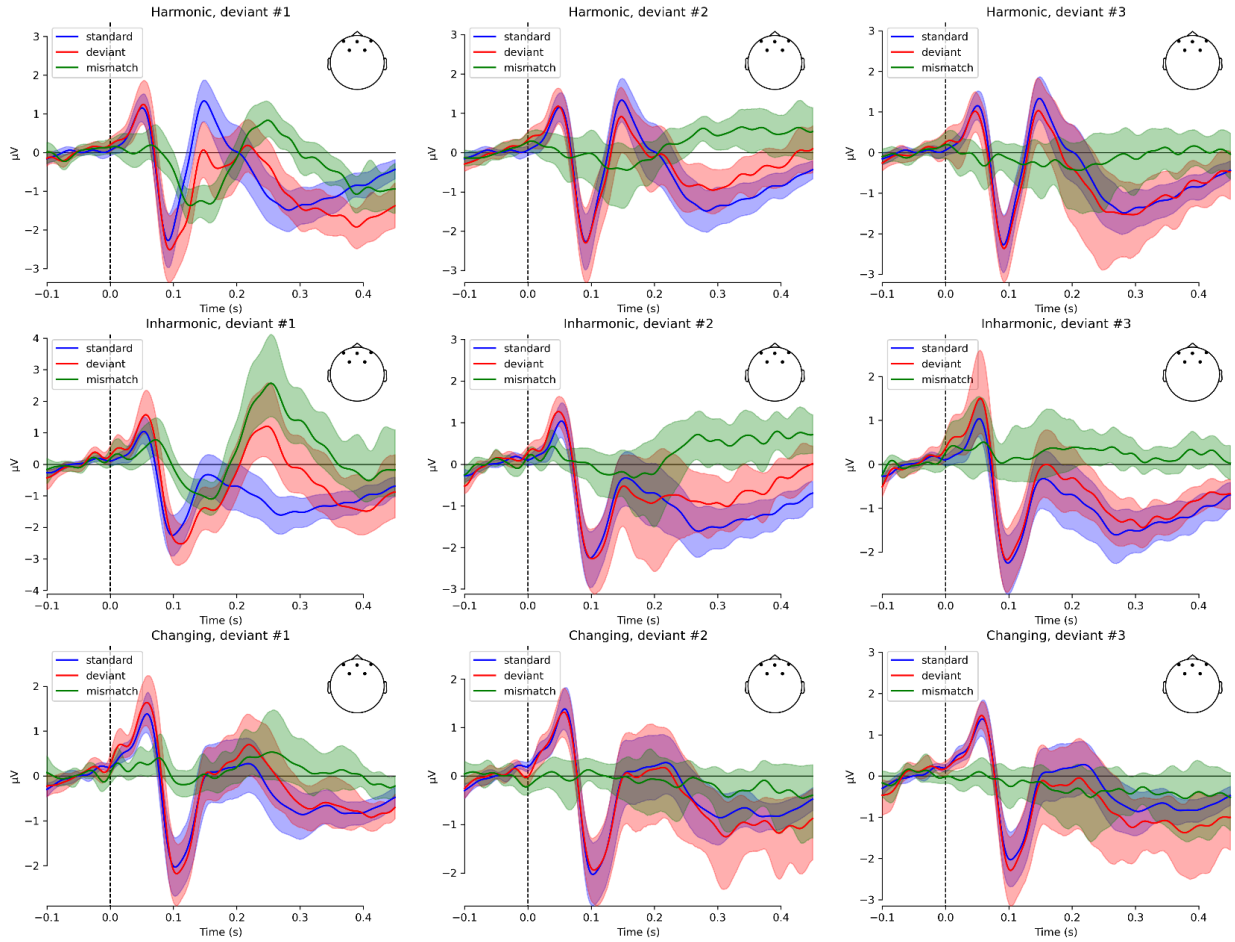
46. Alain, C. & Izenberg, A. Effects of Attentional Load on Auditory Scene Analysis. *J. Cogn. Neurosci.* **15**, 1063–1073 (2003).

47. Guest, D. R. & Oxenham, A. J. Human discrimination and modeling of high-frequency complex tones shed light on the neural codes for pitch. *PLOS Comput. Biol.* **18**, e1009889 (2022).

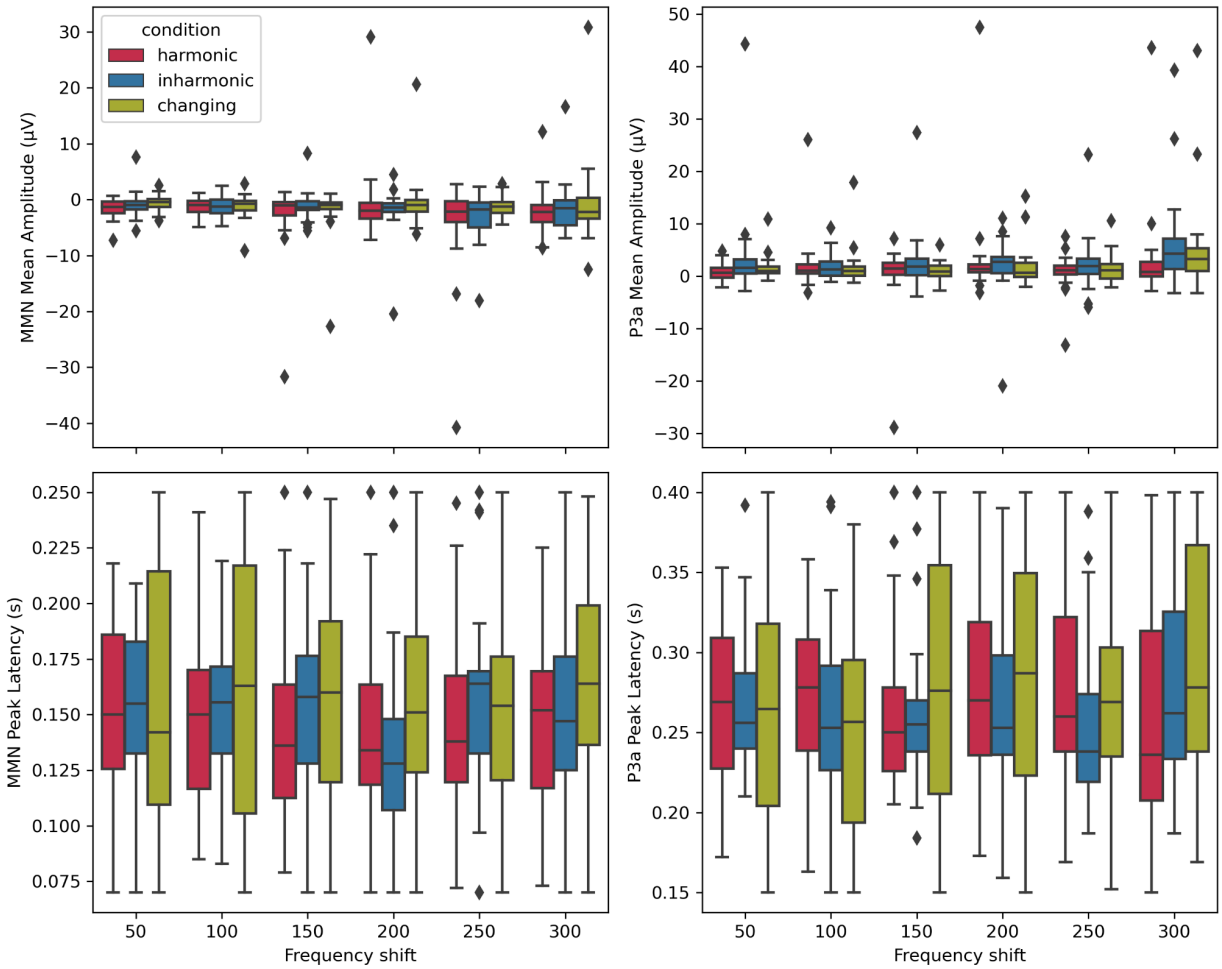
48. Shackleton, T. M. & Carlyon, R. P. The role of resolved and unresolved harmonics in pitch perception and frequency modulation discrimination. *J. Acoust. Soc. Am.* **95**, 3529–3540 (1994).

49. McPherson, M. J., Grace, R. C. & McDermott, J. H. Harmonicity aids hearing in noise. *Atten. Percept. Psychophys.* **84**, 1016–1042 (2022).
50. Quiroga-Martinez, D. R. *et al.* Enhanced mismatch negativity in harmonic compared with inharmonic sounds. *Eur. J. Neurosci.* **56**, 4583–4599 (2022).
51. Gramfort, A. MEG and EEG data analysis with MNE-Python. *Front. Neurosci.* **7**, (2013).
52. Larson, E. *et al.* MNE-Python. Zenodo https://doi.org/10.5281/zenodo.14519545 (2024).
53. Harris, C. R. *et al.* Array programming with NumPy. *Nature* **585**, 357–362 (2020).
54. McKinney, W. Data Structures for Statistical Computing in Python. in *Proceedings of the 9th Python in Science Conference* 56–61 (Austin, Texas, 2010). doi:10.25080/Majora-92bf1922-00a.
55. Hunter, J. D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
56. Jas, M., Engemann, D. A., Bekhti, Y., Raimondo, F. & Gramfort, A. Autoreject: Automated artifact rejection for MEG and EEG data. *NeuroImage* **159**, 417–429 (2017).
57. R Core Team. R: The R Project for Statistical Computing. https://www.r-project.org/ (2021).
58. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
59. Lenth, R. V. *et al.* emmeans: Estimated Marginal Means, aka Least-Squares Means. (2021).
60. Vallat, R. AntroPy: entropy and complexity of (EEG) time-series in Python. (2021).

Appendix



Supplementary Figure 1. Event-related potentials in three studied conditions for first-, second- and third-order deviants. Traces show grand-average responses to standards (blue), deviants (red) and the difference waves (green) for fronto-central channels (F3, Fz, F4, FC1, FC2). Color-shaded areas represent 95% confidence intervals.



Supplementary Figure 2. Boxplots showing distributions of mean MMN (left panels) and P3a (right panels) mean amplitudes (top panels) and peak latencies (bottom panels) in the three conditions as a function of frequency shift.

Supplementary Table 1. Cluster-based analysis results for contrasts between standards and deviants. Table shows cluster p-values, start and end times (in seconds) and number of sensors in each cluster.

Condition	Cluster ID	P-value	T start (s)	T end (s)	No. sensors
harmonic	0	0.8989	-0.081	-0.062	6
	1	0.9597	-0.054	-0.038	4
	2	0.9997	0.03	0.039	1
	3	0.0001	0.087	0.219	26
	4	0.998	0.223	0.239	1
	5	0.9961	0.28	0.301	1
	6	0.0025	0.329	0.45	21
	7	0.9999	-0.037	-0.031	1
	8	0.9998	-0.005	0.003	1
	9	0.9055	0.087	0.134	1
	10	0.0763	0.216	0.286	16
	11	0.9989	0.313	0.326	1
inharmonic	0	0.9998	-0.086	-0.081	1
	1	0.9997	-0.054	-0.047	1
	2	0.0025	0.095	0.199	23
	3	0.0618	0.356	0.45	10
	4	0.4093	-0.035	-0.017	15
	5	0.999	-0.003	0.01	2
	6	0.5749	0.002	0.022	10
	7	0.1492	0.039	0.09	14
	8	0.0001	0.198	0.359	23
	9	0.9984	0.338	0.352	1
changing	0	1	-0.078	-0.073	1
	1	1	-0.071	-0.068	1
	2	0.9996	-0.05	-0.037	1
	3	1	-0.042	-0.039	1
	4	0.9848	0.034	0.062	1
	5	0.997	0.087	0.104	1
	6	0.7332	0.1	0.138	6
	7	1	0.162	0.165	1
	8	0.9824	0.371	0.39	3
	9	0.9887	0.402	0.419	3
	10	0.8501	0.432	0.45	6
	11	0.9996	-0.065	-0.054	2
	12	0.9999	-0.023	-0.013	1
	13	0.9924	0.008	0.022	3

Condition	Cluster ID	P-value	T start (s)	T end (s)	No. sensors
	14	1	0.043	0.044	1
	15	1	0.054	0.058	1
	16	1	0.064	0.072	1
	17	0.997	0.087	0.104	1
	18	0.7025	0.205	0.245	6
	19	0.904	0.252	0.29	2

Supplementary Table 2. Cluster-based permutations analysis for object-related negativity (ORN). Table shows cluster p-values, start and end times (in seconds) and number of sensors in each cluster.

Deviance	Condition	Cluster ID	P-value	T start (s)	T end (s)	Sensors
standard	harmonic-inharmonic	0	1	-0.1	-0.096	1
		1	1	-0.1	-0.094	1
		2	0.9993	-0.066	-0.059	2
		3	0.9803	0.004	0.021	2
		4	0.9944	0.033	0.051	1
		5	0.934	0.036	0.059	3
		6	0.0002	0.1	0.336	27
		7	0.3997	0.369	0.398	15
		8	0.4096	0.414	0.45	13
		9	0.9999	-0.019	-0.012	1
		10	0.9991	-0.019	-0.006	1
		11	0.9997	-0.017	-0.007	1
		12	0.2283	0.059	0.099	15
standard	harmonic-changing	0	1	-0.097	-0.095	1
		1	0.9357	-0.071	-0.056	5
		2	0.9995	-0.07	-0.059	1
		3	0.9999	-0.025	-0.02	2
		4	0.9979	0.053	0.068	1
		5	0.007	0.103	0.179	24
		6	0.9349	0.242	0.276	2
		7	0.7852	0.413	0.45	4
		8	0.9995	-0.07	-0.059	1
		9	0.9999	-0.043	-0.036	1
		10	0.3674	-0.024	0.023	14
		11	0.9165	0.023	0.034	8
		12	0.0189	0.053	0.101	28
		13	0.0006	0.208	0.396	25
standard	contrast	0	0.9994	0.173	0.183	1
		1	0.9886	0.251	0.273	1
		2	0.9997	0.429	0.438	1
		3	1	-0.1	-0.099	1
		4	1	-0.1	-0.098	1
		5	0.195	-0.002	0.04	19
		6	0.0556	0.051	0.099	24
		7	0.0001	0.149	0.45	27

Deviance	Condition	Cluster ID	P-value	T start (s)	T end (s)	Sensors
deviant	harmonic-inharmonic	0	0.9533	-0.1	-0.082	3
		1	0.9814	-0.1	-0.088	4
		2	0.9705	0.077	0.103	1
		3	0.0006	0.104	0.204	23
		4	1	-0.08	-0.076	1
		5	0.5343	-0.042	-0.011	15
		6	0.0236	0.05	0.12	24
		7	0.0275	0.223	0.342	14
		8	0.9939	0.265	0.291	1
		9	0.9991	0.296	0.314	1
		10	0.9996	0.341	0.356	1
		11	1	0.381	0.391	1
		12	0.9993	0.381	0.396	2
deviant	harmonic-changing	0	0.9998	-0.055	-0.044	1
		1	0.7981	0.08	0.14	1
		2	0.8054	0.123	0.147	5
		3	1	0.135	0.141	1
		4	0.9999	-0.015	-0.007	1
		5	0.9665	0.011	0.024	3
		6	0.0111	0.048	0.14	28
		7	0.9364	0.161	0.181	3
		8	0.0016	0.169	0.45	23
		9	1	0.44	0.444	1
deviant	contrast	0	1	-0.073	-0.068	1
		1	1	-0.03	-0.028	1
		2	1	0.042	0.044	1
		3	0.9614	0.046	0.058	5
		4	0.9978	0.099	0.117	1
		5	0.9873	0.121	0.148	1
		6	0.2476	0.237	0.319	16
		7	0.9958	0.331	0.352	1
		8	0.997	0.371	0.389	1
		9	1	-0.06	-0.054	2
		10	1	-0.053	-0.044	1
		11	1	0.022	0.029	1
		12	0.0008	0.055	0.224	25
		13	0.1494	0.346	0.45	8
		14	0.9945	0.37	0.389	2

Supplementary Table 3. Cluster-based F-test for first-, second- and third-order deviants. Table shows cluster p-values, start and end times (in seconds) and number of sensors in each cluster.

Deviant	Cluster ID	P-value	T start (s)	T end (s)	No. sensors
First-order	0	1	-0.05	-0.047	1
	1	1	-0.035	-0.026	1
	2	1	-0.019	-0.015	1
	3	0.9994	-0.014	-0.007	2
	4	1	0.04	0.045	1
	5	1	0.048	0.056	1
	6	1	0.06	0.069	1
	7	0.0163	0.072	0.193	18
	8	0.9849	0.085	0.107	1
	9	1	0.193	0.204	1
	10	0.9935	0.201	0.215	2
	11	0.0075	0.211	0.345	19
	12	0.9989	0.264	0.278	1
	13	1	0.337	0.346	1
	14	0.1561	0.37	0.45	7
Second-order	0	0.9712	-0.1	-0.087	3
	1	0.9999	-0.092	-0.086	1
	2	0.5027	-0.022	0.01	9
	3	1	-0.009	-0.006	1
	4	0.9996	0.028	0.037	1
	5	0.9996	0.028	0.037	1
	6	0.0643	0.243	0.403	14
	7	0.9996	0.379	0.391	1
	8	0.5895	0.417	0.45	7
Third-order	0	0.9899	-0.062	-0.053	2
	1	1	-0.053	-0.048	1
	2	1	-0.047	-0.045	1
	3	0.9997	-0.01	-0.002	1
	4	0.9865	0.032	0.049	1
	5	1	0.071	0.075	1
	6	0.9993	0.102	0.112	1
	7	0.5878	0.114	0.148	4
	8	1	0.2	0.206	1
	9	0.9398	0.359	0.397	1
	10	0.9994	0.38	0.389	1

Supplementary Table 4. Model comparisons for the frequency shift analysis. The models include condition-only (m1), condition and frequency shift without interaction (m2), condition and frequency shift with interaction (m3). Model comparison is performed with a likelihood ratio chi squared test. Akaike Information Criteria (AIC) are presented for each model.

DV	Model	AIC	Statistic	Df	P-value
MMN mean amplitude	m1	3470.80			
	m2	3467.65	5.151	1	0.023
	m3	3470.09	1.556	2	0.459
MMN peak latency	m1	-2114.62			
	m2	-2112.99	0.369	1	0.543
	m3	-2109.27	0.276	2	0.871
P3 mean amplitude	m1	3712.23			
	m2	3702.90	11.333	1	0.001
	m3	3706.07	0.832	2	0.660
P3 peak latency	m1	-1705.49			
	m2	-1705.68	2.199	1	0.138
	m3	-1705.19	3.510	2	0.173

Supplementary Table 5. Post-hoc comparisons for P2 analysis of variance (model m2). T-tests with Tukey HSD correction for multiple comparisons.

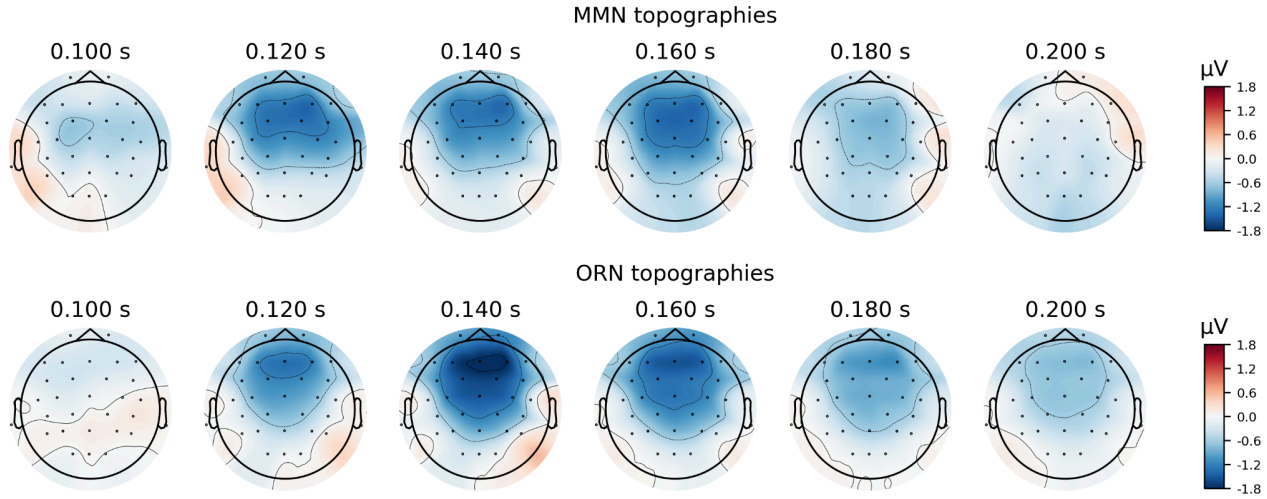
Contrast	Estimate	Std. error	df	Statistic	P-value
harmonic deviant - inharmonic deviant	0.938	0.194	170	4.828	<.0001
harmonic deviant - changing deviant	-0.030	0.194	170	-0.154	1.0000
harmonic deviant - harmonic standard	-1.124	0.194	170	-5.787	<.0001
harmonic deviant - inharmonic standard	0.110	0.194	170	0.566	0.9931
harmonic deviant - changing standard	-0.128	0.194	170	-0.661	0.9859
inharmonic deviant - changing deviant	-0.967	0.194	170	-4.982	<.0001
inharmonic deviant - harmonic standard	-2.061	0.194	170	-10.615	<.0001
inharmonic deviant - inharmonic standard	-0.828	0.194	170	-4.262	0.0005
inharmonic deviant - changing standard	-1.066	0.194	170	-5.489	<.0001
changing deviant - harmonic standard	-1.094	0.194	170	-5.634	<.0001
changing deviant - inharmonic standard	0.140	0.194	170	0.720	0.9793
changing deviant - changing standard	-0.098	0.194	170	-0.507	0.9959
harmonic standard - inharmonic standard	1.234	0.194	170	6.353	<.0001
harmonic standard - changing standard	0.995	0.194	170	5.127	<.0001
inharmonic standard - changing standard	-0.238	0.194	170	-1.227	0.8233

P3 mean amplitude analysis - a possible outlier?

Visual inspection of the P3 mean amplitudes (Fig. 4) revealed a participant with P3 amplitudes higher than others. To verify the possibility that this outlier was driving the effect of harmonicity on P3 amplitudes, we conducted the same analysis with data from this participant excluded. The results showed that *m1* performs better than *m0*, (AIC_{m0} = 458.8, AIC_{m1} = 443.9, $Chi^2(2)$ = 18.80, $p < .0001$). Post-hoc comparisons revealed significant differences between harmonic and inharmonic conditions (contrast estimate = -1.36, SE = .36, t(68) = -3.85, p = .0008) as well as between inharmonic and changing conditions (contrast estimate = 1.45, SE = .36, t(68) = 4.07, *p* = .0004). No significant differences were found for the harmonic - changing contrast (contrast estimate = .08, SE = .36, t(68) = .23, *p* = .97). Taken together, these results suggest that the effect of harmonicity on P3 mean amplitude was not driven by an outlier observation.

Differences in topographies between ORN and MMN

To investigate if the topographies of ORN and MMN differed, we used a spatio-temporal cluster 1-sample t-test to compare MMN and ORN difference waves. We did this analysis for the time window of 100 - 200 ms where both components are negative. We contrasted MMNs in *harmonic* and *changing* conditions to ORNs for *inharmonic standards, inharmonic deviants and changing standards*. Since no MMN was detected for the *changing* condition, nor no ORN was detected for *changing deviants*, we did not include these contrasts in the analysis. Overall, six pairwise comparisons were made and none revealed significant clusters (all p > .05). These results do not support the notion that ORN and MMN have different topographical distributions. However, because we did not perform source analysis, this result does not preclude different neural sources of the components.



Supplementary Figure 3. Topographies of MMN and ORN responses. Upper panels show topographies of the *harmonic* MMN condition and lower panels show topographies of the *inharmonic standards* ORN.