Research report in *International Journal of Psychophysiology*

Relationship Between Schematic and Dynamic Expectations of Melodic Patterns in Music Perception

Short title: SCHEMATIC AND DYNAMIC EXPECTATIONS IN MUSIC

Kai Ishida[1], Hiroshi Nittono[1]

[1] Graduate School of Human Sciences, Osaka University, Osaka, Japan

Corresponding author: Kai Ishida, Graduate School of Human Sciences, Osaka University, 1-2 Yamadaoka, Suita, Osaka 565-0871, JAPAN; ishida@hus.osaka-u.ac.jp; ORCID: 0000-0001-6485-0950

Hiroshi Nittono, Graduate School of Human Sciences, Osaka University, 1-2 Yamadaoka, Suita, Osaka 565-0871, JAPAN; nittono@hus.osaka-u.ac.jp; ORCID: 0000-0001-5671-609X

Authorship contribution statement: Kai Ishida: Conceptualization, Methodology, Investigation, Data curation, Formal analysis, Visualization, Writing - original draft, and Funding acquisition, Project administration. **Hiroshi Nittono:** Conceptualization, Methodology, Formal analysis, Writing - review & editing.

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

Funding information: This work was supported by JSPS KAKENHI JP22KJ2199.

Keywords: Schematic expectation, dynamic expectation, tonal expectancy, predictive coding, music prediction

Word count: 4438

Data accessibility:

The sound materials used and datasets analyzed for the present paper are available at https://osf.io/um42f/.

Highlights

- Music-syntactic irregularities violate a schematic expectation in music.
- Melodic-contour deviants violate a dynamic expectation in music.
- Both irregularities and deviants elicited neural prediction error responses.
- The neural response increased multiplicatively when these deviances co-occurred.
- Schematic and dynamic expectations function interactively in music.

Abstract

Prediction is fundamental in music listening. Two types of expectations have been proposed: schematic expectations, which arise from knowledge of tonal regularities (e.g., harmony and key) acquired through long-term plasticity and learning, and dynamic expectations, which arise from short-term regularity representations (e.g., rhythmic patterns and melodic contours) extracted from ongoing musical contexts. Although both expectations are indispensable in music listening, how they interact with each other in music prediction remains unclear. The present study examined the relationship between schematic and dynamic expectations in music processing using event-related potentials (ERPs). At the ending note of the melodies, the schematic expectation was violated by presenting a note with music-syntactic irregular (i.e., out-of-key note), while the dynamic expectation was violated by presenting a contour deviant based on online statistical learning of melodic patterns. Schematic and dynamic expectations were manipulated to predict the same note. ERPs were recorded for the music-syntactic irregularity and the contour deviant, which occurred independently or simultaneously. The results showed that the music-syntactic irregularity elicited an early right anterior negativity (ERAN), reflecting the prediction error in the schematic expectation, while the contour deviant elicited a mismatch negativity (MMN), reflecting the prediction error in the dynamic expectation. Both components occurred within a similar latency range. Moreover, the ERP amplitude was multiplicatively increased when the irregularity and deviance occurred simultaneously. These findings suggest that schematic and dynamic expectations function concurrently in an interactive manner when both expectations predict the same note.

1. Introduction

Music perception has been explained based on the notion of predictive coding (Koelsch et al., 2019), which holds that top-down prediction signals are passed down cortical hierarchies to minimize prediction errors, and these errors are weighted by their expected precision or predictability (Vuust et al., 2022). The predictive process in music entails various expectations (Rohrmeier & Koelsch, 2012). These include schematic expectations that depend on the long-term learning of musical representation and dynamic expectations that are formed immediately and depend on repeated patterns in a musical piece (Vuust et al., 2022). More specifically, schematic expectations are based on the static knowledge of tonality and meter and are exemplified by harmonic expectations (e.g., this type of chord progression will lead to a specific chord). In contrast, dynamic expectations are based on a continuously updated short-term memory of the recent auditory input and are exemplified by expectations of repeated events (e.g., a certain melody will repeat again).

Event-related potential (ERP) studies have investigated schematic expectations of music using early right anterior negativity (ERAN; Koelsch et al., 2000, 2007). The ERAN occurs predominantly in the frontal region, with a latency of 150–200 ms after the onset of music-syntactic irregularities, such as a harmonic irregularity in chord progression (Koelsch et al., 2000, 2007) and an out-of-key note in a tonal melody (Kalda & Minati, 2012; Miranda & Ullmann, 2007). Previous studies have recorded the ERAN even when syntactically regular and irregular notes were presented with equal probability to avoid extracting regularities from the current auditory context (Koelsch et al., 2007), as well as when irregular notes were presented infrequently (Ishida & Nittono, 2022). Thus, the ERAN could represent a prediction error in tonal expectancy generated by the music syntactic and tonal schemata (e.g., tonality and harmony) that exist in a long-term format (Koelsch, 2009).

Dynamic expectations have been investigated using mismatch negativity (MMN), which is elicited by comparing a sound input with the expectations generated by short-term regularity representations extracted online from the current auditory context (Koelsch et al., 2009). Traditionally, the MMN has been recorded using the oddball paradigm, in which an infrequent (deviant) tone has different acoustic features than a frequent (standard) tone (Näätänen et al., 2005; Sussman et al., 2014). The MMN is elicited not only by the irregularity of the frequency of tone occurrence but also by the irregularity of tone transitions based on learning of transitional probabilities (Moldwin et al., 2017; Tsogli et al., 2019). Learning the transitional probability of events from various sensory modalities is called statistical learning (auditory: Batterink et al., 2015; Daikoku et al., 2014, 2015, 2016; visual: Jost et al., 2015; Turk-Browne et al., 2005; multimodal: Conway & Christiansen, 2006; Paraskevopolous et al., 2018). The MMN elicited by statistical learning is referred to as *statistical* MMN, which reflects online learning of probabilistic regularities in auditory sequences (Ishida & Nittono, 2023; Koelsch et al., 2016; Tsogli et al., 2019). Thus, the MMN could be a prediction error based on various types of regularity representations extracted from sensory input (Winkler & Czigler, 2012). However, it remains unclear how schematic and dynamic expectations function interactively in music perception.

The present study aimed to investigate the relationship between a schematic expectation based on tonal schema and a dynamic expectation based on statistical learning. In the experiment, the participants were repeatedly exposed to two types of melodic contours consisting of single notes while detecting an infrequent timbre change in a sequence of melodies. Two melodies consisting of two phrases ended with a syntactically regular note ascending from the last note of the preceding melody, and an irregular note descending from the last note of the preceding melody. To manipulate the melodic-contour deviance, one melody transitioned to the syntactically regular note with high probability (p = .90, melodic-contour standard) and to the syntactically irregular note with low probability (p = .10, melodic-contour deviant), while another melody transitioned to the syntactically regular note with low probability and to the syntactically irregular note with high probability. To manipulate the music-syntactic irregularity, syntactically regular and irregular notes were presented with equal probability to elicit the ERAN (Koelsch et al., 2000, 2007). A syntactic irregularity could be induced by a single note because the last notes of phrases C and D were leading notes (in C major, B), which induce a strong expectation of the tonic note (in C major, C). The syntactic irregularity and contour deviance were presented independently and simultaneously to examine the relationship between schematic and dynamic expectations.

Several studies have reported that MMN amplitude changes additively (Caclin et al., 2006; Paavilainen et al., 2001; Takegata et al., 1999) or subadditively (Tsogli et al., 2019; Wolf & Schröger, 2001) when two auditory regularity dimensions are deviated simultaneously. The additivity of the MMN suggests that different sets of neurons are activated, and it has been interpreted as evidence for independent processing of two auditory regularity dimensions (Ishida & Nittono, 2022; Takegata et al., 1999). In contrast, the subadditivity of the MMN suggests that the two auditory regularity dimensions share neural resources and are thus processed jointly (Lidji et al., 2009; Tsogli et al., 2019; Wolff & Schröger, 2001). In the present study, the elicitation of ERAN and MMN depended on different regularity representations. However, the prediction process was expected to overlap, because different regularity representations predict the same auditory event (i.e., the pitch of the melody). If this was the case, the ERAN and MMN would occur in the same latency range, and the ERP amplitude would be smaller than the sum of the ERAN and MMN amplitudes when music-syntactic irregularity and melodic-contour deviance occur simultaneously (subadditivity). The schematic expectation and dynamic expectation were expected to interact in the musical context.

2. Method

This study was preregistered before sampling. The preregistration details for each experiment can be found at the following link: https://osf.io/sedt3.

2.1. Participants

A sample size of 36 was predetermined to ensure the detection of a medium effect size (d_z = 0.486) for the interaction of transition probability and physical deviance, which was calculated by $F(1, 20) = 4.95$ (Tsogli et al., 2019) with power $1-\beta = .80$ and error rate $\alpha = .05$, power analysis using G*power (Faul et al., 2007). Taking into account data exclusion, 40 participants were recruited. As a result, data from 37 participants (21 women and 16 men, 18–59 years old, M = 24.1 years) were used for analysis after the exclusion of three participants due to excessive noise and technical errors. Thirty-four participants were right-handed, two were left-handed, and one was ambidextrous (FLANDERS handedness questionnaire; Okubo et al., 2014). None had hearing impairments or a history of neurological disease. The participants had various types of musical experience, with a mean of 5.8 years of extracurricular musical lessons (range 0–16 years). The protocol was approved by the Behavioral Research Ethics Committee of the Osaka University School of Human Sciences, Japan (HB023-007), and written informed consent was obtained from all participants. Participants received a cash voucher of 3,000 Japanese yen as an honorarium.

2.2. Stimuli

Example of the stimuli are shown in Figure 1. Two phrases and an ending note were concatenated to form a melody and played with a piano timbre. Two versions were composed for the first (A, B) and second (C, D) phrases. Each phrase consisted of two eighth notes (each 166.5 ms) and one quarter note (333 ms). Two concatenated phrases consisting of two single phrases were created (i.e., AC and BD) and connected to a final half note (666 ms), which was music-syntactic regular (F) or irregular (E), with high (p = .90, standard) or low (p = .10, deviant) probability. In C major, the syntactically regular note was [note C, variant F], and the syntactically irregular note was [note B♭, variant F]. Therefore, four possible melodies were created, and these were transposed into the 12 major keys and presented in a randomized order to prevent the effect of sensory dissonance due to the out-of-key notes in the syntactic irregular condition. The probabilities of the melodies are shown in Table 1.

Both syntactically regular and irregular notes were separated by a semitone from the preceding note, although the directions differed (i.e., up or down). Four types of melodies were created. To control the effect of the melodic contour, different transitional probabilities were set for Group I and Group II. In Group I, AC transitioned to the syntactically regular note with high probability and to the syntactically irregular note with low probability, whereas BD transitioned to the syntactically irregular note with high probability and to the syntactically regular note with low probability. In

Group II, AC transitioned to the syntactically irregular note with high probability and to the syntactically regular note with low probability, whereas BD transitioned to the syntactically regular note with high probability and to the syntactically irregular note with low probability. All melodies were concatenated without any interstimulus interval, and the transitional probability between the melodies was equal ($p = .50$). Thus, the syntactically regular and irregular notes had an equal probability of occurrence as in previous ERAN studies (Koelsch et al., 2007).

2.3. Procedure

In the EEG recording, all four melodies were randomly presented in the 12 major keys. A timbre change detection task was used as a cover task. The participants were asked to press a button as quickly and accurately as possible when a quarter-note or half-note of the melodies changed to a guitar tone. Cover tasks are commonly used during statistical learning to make participants pay attention to an auditory sequence without focusing on its regularities (Daikoku et al., 2016; Daikoku & Yumoto, 2017; Ishida & Nittono, 2023; Koelsch et al., 2016; Tsogli et al., 2019). They occurred 3–5 times in each melodic sequence (i.e., one block), for a total of 40 changes in the experiment. The timbre change occurred only in the notes of high transitional probability to avoid reducing the number of trials with irregular conditions. Ten blocks were performed with short breaks, and each block lasted approximately six minutes. Within a block, melodies whose phrases were AC and BD were randomly presented 100 times each, and the syntactically regular and irregular notes were presented 100 times each, with the constraint that melodies with low transitional probabilities are not repeated in succession. In Group I, ACE, ACF, BDF, and BDE were presented 900, 100, 900, and 100 times, respectively. In Group II, ACE, ACF, BDF, and BDE were presented 100, 900, 100, and 900 times, respectively.

The EEG recording was followed by a fitness rating task that took approximately four minutes. This task was conducted as a manipulation check to ensure that the participants had learned the melodic-contour regularities. In the rating task, four types of melodies were presented in all 12 major keys (once for each major key), and participants were asked to rate the fitness of the ending note of the melody using a scale from 1 (*not fit at all)* to 7 (*very fit*). Consecutive trials did not use the same ending note, and the presentation order of the phrase combination was counterbalanced. Participants were informed about the regularity of melodic sequences and the presence of the syntactic deviance at the end of the experiment.

2.4. EEG recording

EEG data were recorded using a QuickAmp (Brain Products, Germany) with Ag/AgCl electrodes. Thirty-four scalp electrodes were used according to the 10–20 system (Fp1/2, F3/4, F7/8, Fz, FC1/2, FC5/6, FT9/10, C3/4, T7/8, Cz, CP1/2, CP5/6, TP9/10, P3/4, P7/8, Pz, O1/2, Oz, PO9/10). Additional electrodes were placed on the left and right mastoids, the left and right outer canthi of the eyes, and above and below the right eye. The data were referenced offline to the algebraic means of the left and right mastoid electrodes. The sampling rate was 1,000 Hz. The online filter was DC–200 Hz. Electrode impedances were kept below 10 kΩ.

2.5. EEG data reduction

EEG data were analyzed using a Brain Vision Analyzer (Brain Products, Germany). First, a digital filter of 0.25 Hz (6 dB/oct) high-pass filter and 25 Hz (48 dB/oct) low-pass filter were applied to the data (Koelsch et al., 2007). Then, ocular artifact correction based on independent component analysis was applied. The detection of ICs associated with artifacts (e.g., ocular, bad connection at a single channel) was performed semiautomatically through visual inspection. A 500 ms period (100 ms before and 400 ms after the ending note) was averaged after removing trials in which voltages exceeded ±80 μV in any channel. Two consecutive trials after the timbre change were removed from the averaging. Baseline correction was applied by subtracting the mean amplitude of the prestimulus 100 ms period from each point of the waveform. For statistical evaluation, five frontal electrodes (F7, F3, Fz, F4, and F8) were clustered. Then, the ERAN was calculated by subtracting ERP waveforms for the note without syntactic irregularity (e.g., mean of ACE and BDE for Group I in Figure 1) from those for the note with syntactic irregularity (e.g., mean of ACF and BDF for Group I in Figure 1). The MMN was calculated by subtracting ERP waveforms for the note without contour deviance (e.g., mean of ACE and BDF for Group I in Figure 1) from those for the note with contour deviance (e.g., mean of ACF and BDE for Group I in Figure 1). Then, the peak of each component was detected in a time window of 100–300 ms on each grand mean difference waveform. The 40-ms period centered around each peak was defined as the time window of the ERAN or MMN amplitude calculation. In addition, to analyze the amplitudes of both components together, a common time window was determined as the 40-ms period centered around the most negative peak between 100 and 300 ms on the average waveform of the ERAN and MMN waveforms. The mean ERP amplitudes of this period were calculated for the four conditions (i.e., standard, syntactic irregularity, contour deviant, double deviant). On average, 816 (698–837), 98 (82–100), 815 (703–837), and 97 (84–100) epochs were used to calculate the standard, contour deviant, syntactic irregularity, and double deviant (i.e., co-occurrence of syntactic irregularity and contour deviant) ERP waveforms, respectively.

2.6. Statistical analysis

Statistical analyses were carried out using JASP 0.17.2 (JASP Team, 2023), which is an open-source statistics program that allows both classical (frequentist) and Bayesian analyses and its output has been verified with other statistical packages. To examine the presence of the ERAN and MMN, a one-sample t-test (two-sided) was conducted on the ERP amplitude of the ERAN and MMN intervals. Then, a Bayesian one-sample t-test was conducted to assess the absence (effect size $\delta = 0$, null hypothesis) or presence of the difference (effect size $\delta > 0$, alternative hypothesis). To examine the difference in the peak latency of the ERAN and MMN, a paired t-test (two-sided) and Bayesian paired t-test were conducted on the latency of the ERAN and MMN. Finally, a two-way analysis of variance (ANOVA) with music-syntactic irregularity (with/without music-syntactic irregularity) and melodic-contour deviance (with/without melodic-contour deviance) was conducted on the ERP amplitudes to examine the interaction of irregularity and deviance factors. This analysis was also conducted using a Bayesian two-way ANOVA to assess the absence or presence of the effects. For frequentist hypothesis testing, the significance levels were set to $\alpha = .05$. For Bayesian hypothesis testing, the Cauchy distribution with a scale parameter r of 0.707 was used as the prior distribution for δ in the t-test. For the Bayesian two-way repeated-measures ANOVA, multivariate Cauchy distribution (fixed effect: scale parameter $r = 0.5$; random effect: scale parameter $r = 1$; covariates: scale parameter $r = .354$) was used as the prior distribution. According to the classification scheme of Schönbrodt and Wagenmakers (2018), a Bayes factor (BF_{01}) greater than 3 was considered moderate evidence for the null hypothesis. The stimulus materials and the data necessary to replicate the findings are available at https://osf.io/um42f/.

3. Results

3.1. Manipulation check

The averaged mean reaction time of the timbre change detection task was 367 ms (SD = 94 ms), and the averaged hit rate was 87.3% (SD = 14.0%). These results suggest that participants focused on the task and attended to the melodic sequence.

The mean fitness ratings (SD) were 6.4 (0.6), 2.8 (1.1), 6.0 (1.1), and 2.5 (1.0) for the standard, syntactic irregularity, contour deviant, and double deviant, respectively. The two-way ANOVA with music-syntactic irregularity and melodic-contour deviance on the fitness ratings revealed the significance of music-syntactic irregularity, $F(1, 36) = 304.50$, $p < .001$, $\eta_p^2 = .894$, $BF_{10} = 1.55\times10^{16}$, and melodic-contour deviance, $F(1, 36) = 9.48$, $p = .004$, $\eta_p^2 = .208$, $BF_{10} = 7.41$. However, the interaction was not significant, $F(1, 36) = 0.21$, $p = .651$, $\eta_p^2 = .006$, $BF_{10} = 0.26$. These results indicate that the participants recognized the music-syntactic irregularity as the musical deviant, and the melodic-contour deviant was not fit based on statistical learning of melodic contours.

Figure 2

3.2. ERP analysis

Figure 2 shows the grand average waveforms and scalp topographies of the ERPs. The music-syntactic irregularity elicited the ERAN with a peak latency of 123 ms, and the ERAN amplitude ($M = -0.72$ µV, $SD = 0.58$) calculated in a period of 103–143 ms was significantly negative, $t(36) = 7.51$, $p < .001$, $d_z = 1.23$, $BF_{10} = 1.78\times10^{6}$. The melodic-contour deviant also elicited the MMN with a peak latency of 155 ms, and the MMN amplitude ($M = -0.66$ μV, $SD = 0.74$) calculated in a period of 135–175 ms was significantly negative, $t(36) = 5.42$, $p < .001$, $d_z = 0.89$, $BF_{10} = 4580.45$. The peak latency of ERAN ($M = 169$ ms, $SD = 63$ ms) and MMN ($M = 184$ ms, $SD = 58$ ms) was not significantly different, $t(36) = 0.99$, $p = .327$, $d_z = 0.16$, $BF_{01} = 3.58$.

Because the ERAN and MMN were elicited in a similar latency range, the interaction of each deviance factor was examined by two-way ANOVA with music-syntactic irregularity and melodic-contour deviance conducted on ERP amplitudes over a period of 134–174 ms. This interval was determined by the peak latency (154 ms) of the grand mean difference waveform of the ERP elicited by the note, with both music-syntactic irregularity and melodic-contour deviance. As shown in the raincloud plots of Figure 2, the deviance-related ERP amplitude was larger for the double deviant than for the single deviants. The means (*SD*) of deviant − standard ERP amplitudes were −0.43 (0.52), −0.38 (0.68), and −1.38 (1.21) for the syntactic irregularity, contour deviant, and double deviant, respectively. All the effects were significant, including the interaction: music-syntactic irregularity, $F(1, 36) = 39.59$, $p < .001$, $\eta_p^2 = .524$, $BF_{10} = 30209.76$, melodic-contour deviance, $F(1, 36) = 29.88$, $p < .001$, $\eta_p^2 = .454$, $BF_{10} = 3854.18$, and interaction, $F(1, 36) = 9.76$, $p = .004$, $\eta_p^2 = .213$, $BF_{10} = 17.56$. Figure 3 shows the deviance-related ERP responses under the presence or absence of the other type of deviance. The post-hoc tests showed that the simple main effect of the music-syntactic irregularity was significant both when the melodic-contour deviance was present (i.e., ERAN for infrequent final notes of the melody), $p < .001$, $dz = 1.02$, and when it was absent (i.e., ERAN for frequent final notes of the melody), $p = .011$, $dz = 0.44$. Also, the simple main effect of the melodic-contour deviance was significant both when the music-syntactic irregularity was present, $p < .001$, $dz = 0.91$ (i.e., MMN for syntactically irregular final notes), and when it was absent (i.e., MMN for syntactically regular final notes), $p = .033$, $dz = 0.37$. These results demonstrated that the effects of music-syntactic irregularity and melodic-contour deviance were significant, irrespective of the presence of another type of deviance. Moreover, the double-deviant ERP amplitude was significantly larger than that of the summed single-deviant ERPs (i.e., syntactic irregularity + contour deviant: $M = -0.81$, $SD = 0.94$), $t(36) = 3.13$, $p = .004$, $dz = 0.51$, $BF_{10} = 10.35$ (see the top panel of Figure 3).

Figure 3

As shown in the difference waveforms of Figure 2, each deviant condition elicited a P1 deflection. The deviant − standard ERP amplitude of this wave looks larger in the double-deviant condition ($M = 0.24$, $SD = 0.70$) than in the single-deviant conditions (Syntactic irregularity: $M = -0.08$, $SD = 0.40$; Contour deviant: $M = -0.08$, $SD = 0.67$). Several studies have reported the effect of statistical learning on the P1 interval (Daikoku et al., 2016, 2017). Therefore, the effect on P1 amplitude was examined exploratorily by a one-way ANOVA with a factor of condition (the syntactic irregular, contour deviant, and double deviant) using a mean amplitude of 53–93 ms (i.e., peak latency 73 ± 20 ms). However, the effect of the condition was not significant, $F(2, 72) = 3.23$, $p = .058$, $\eta_p^2 = .082$, $BF_{10} = 1.72$.

4. Discussion

The present study examined the relationship between schematic and dynamic expectations through statistical learning. The ERAN and the MMN occurred in a similar latency range to reflect the violation of these expectations. Because syntactically irregular notes presented with the same probability as regular notes elicited an ERAN, this study confirmed previous findings that the ERAN is elicited independently of presentation probability (Koelsch et al., 2007). The MMN in response to melodic-contour deviants is considered the statistical MMN. When both the syntactic irregularity and contour deviance co-occurred, the ERP amplitude increased multiplicatively. These results indicate that schematic and dynamic expectations function concurrently and interactively at the early perceptual stage.

The present study observed the interactive effect in MMN and ERAN amplitudes, while the previous study, which recorded the ERAN elicited by a deviance in harmony as a music-syntactic irregularity and the MMN elicited by a deviance in intensity as an acoustic irregularity dimension, demonstrated the additivity of each deviance factor (Ishida & Nittono, 2022). The discrepancy between the additive effect in the previous study and the interactive effect in the current study may be attributed to differences in dynamic expectations. Ishida and Nittono (2022) used an intensity deviant, a separate auditory dimension that was different from the one predicted by schematic expectation. Therefore, the neural prediction errors were summed up, reflecting the parallel detection of music-syntactic irregularity and intensity deviance. In contrast, the patterns of melodic contours were manipulated in the present study so that dynamic expectation predicted the same note as schematic expectation. Therefore, neural prediction errors were multiplied. If each expectation involved a separate prediction system, the co-occurrence of two violations would have caused a deviance response whose magnitude was equal to the sum of the deviance response to each single violation, reflecting the activation of separate neural resources (Caclin et al., 2006; Takegata et al., 1999). However, this was not the case. Thus, the interaction result provides evidence that violations of two expectations are processed jointly for eliciting prediction error signals, although the subadditivity we originally hypothesized was not observed.

The interaction effect can be explained in several ways. When the same note was predicted by both expectations, the deviance-related ERP amplitude increased multiplicatively when violations occurred simultaneously compared to when each violation occurred independently. Although this finding is uncommon in research on the double deviant MMN, the multiplicative effect may be explained if the prediction from one type of expectation had an antagonistic effect on the prediction from the other type of expectation. When one expectation was violated while the other was not (i.e., single deviant), the size of the prediction error may have been reduced, reflecting the antagonism between the violation and non-violation of expectations. In the present study, the ERAN and MMN were more pronounced when another deviance was present than when another deviance was absent (see Figure 3). This finding may indicate that the prediction error of expectation was inhibited when the other type of expectation was met. In contrast, when the two expectations were violated together (i.e., double deviant), the neural prediction error was fully emitted because there was no contradiction. Therefore, in the present study, the ERP response to the double deviant showed multiplicativity rather than subadditivity. Alternatively, it is also possible that the combination of two deviants made the event surpassingly salient, and it produced a multiplicatively large ERP response.

The predictive coding framework assumes that the prediction error is used to update the generative model for precise inference of the environment (Friston, 2010; Friston & Kiebel, 2009). If the elicitation of the ERAN/MMN simply reflects deviance detection, the multiplicative effect elicited by the double deviant note cannot be explained, because deviance was detectable by both of the two expectations alone. The multiplicativity of the ERAN/MMN in the present study suggests the separate updating of auditory generative models for prediction. Schematic expectations seem to serve as a higher-level prediction compared to dynamic expectations, because the latter require bottom-up regularity extraction. In the present study, therefore, the prediction errors may have ascended the auditory cortical hierarchies to update predictions at both lower (dynamic expectations) and higher (schematic expectations) levels, leading to more accurate predictions at each level. This is consistent with previous findings that neural activations reflecting prediction errors were observed differently between local and global deviants, corresponding to separate formations of lower- and higher-level predictions (Hofmann-Shen et al., 2020; Jiang et al., 2022; Recasens et al., 2014). However, further evidence is required to draw conclusions about the hierarchical structure in the prediction mechanisms of the present study, as predictability can be inferred post hoc, as criticized in the review by Denham and Winkler (2020).

The modulation of the effect of melodic-contour deviance may also be explained by the effect of prior knowledge on statistical learning. Previous studies have empirically demonstrated that prior knowledge affects statistical learning (Elazar et al., 2022; Rogers et al., 2021; Siegelman et al., 2018; Stärk et al., 2022). For example, Elazar et al. (2022) demonstrated that auditory statistical learning of artificial words with high co-occurrence syllables in a native language resulted in higher recognition and discrimination performance than that of words with low co-occurrence syllables. This finding suggests that prior knowledge modulates auditory statistical learning. In the present study, because the contour deviant could be syntactically regular (i.e., the single deviant of the contour deviant), the statistical learning could be influenced by the predictability of the schematic expectation, leading to antagonism between the schematic and dynamic expectations. As in the language domain, the effect of prior musical knowledge on statistical learning in music should be considered in future research.

In summary, the present study demonstrated that schematic and dynamic expectations functioned interactively when both expectations predicted the same note. The interactive effect suggests that violations of two expectations are processed jointly for eliciting prediction error signals. The multiplicative interaction may be due to the antagonism between schematic and dynamic expectations, or due to a higher salience of the double deviant. Moreover, the current findings support the presence of hierarches in predictive coding, in which prediction errors ascend to different levels in cortical hierarchies to update predictions in both schematic (higher level) and dynamic (lower level) expectations, although further research is required to confirm this.

References

Batterink, L. J., Reber, P. J., Neville, H. J., & Paller, K. A. (2015). Implicit and explicit contributions to statistical learning. *Journal of Memory and Language*, *83*, 62–78. https://doi.org/10.1016/j.jml.2015.04.004

Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Giard, M. H., & McAdams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, *18*, 1959–1972. https://doi.org/10.1162/jocn.2006.18.12.1959

Conway, C. M., & Christiansen, M. H. (2006). Statistical learning within and between modalities: Pitting abstract against stimulus-specific representations. *Psychological Science*, *17*(10), 905–912. https://doi.org/10.1111/j.1467-9280.2006.01801.x

Daikoku, T., Yatomi, Y., & Yumoto, M. (2014). Implicit and explicit statistical learning of tone sequences across spectral shifts. *Neuropsychologia*, *63*, 194–204. https://doi.org/10.1016/j.neuropsychologia.2014.08.028

Daikoku, T., Yatomi, Y., & Yumoto, M. (2015). Statistical learning of music- and language-like sequences and tolerance for spectral shifts. *Neurobiology of Learning and Memory*, *118*, 8–19. https://doi.org/10.1016/j.nlm.2014.11.001

Daikoku, T., Yatomi, Y., & Yumoto, M. (2016). Brain and Cognition Pitch-class distribution modulates the statistical learning of atonal chord sequences. *Brain and Cognition*, *108*, 1–10. https://doi.org/10.1016/j.bandc.2016.06.008

Daikoku, T., & Yumoto, M. (2017). Single, but not dual, attention facilitates statistical learning of two concurrent auditory sequences. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-10476-x

Denham, S. L., & Winkler, I. (2020). Predictive coding in auditory perception:

challenges and unresolved questions. *European Journal of Neuroscience*, *51*(5), 1151–1160. https://doi.org/10.1111/ejn.13802

Elazar, A., Alhama, R. G., Bogaerts, L., Siegelman, N., Baus, C., & Frost, R. (2022). When the “tabula” is anything but “rasa:” What determines performance in the auditory statistical learning task? *Cognitive Science*, *46*(2). https://doi.org/10.1111/cogs.13102

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127–138. https://doi.org/10.1038/nrn2787

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Hofmann-Shen, C., Vogel, B. O., Kaffes, M., Rudolph, A., Brown, E. C., Tas, C., Brüne, M., & Neuhaus, A. H. (2020). Mapping adaptation, deviance detection, and prediction error in auditory processing. *NeuroImage*, *207*, 116432. https://doi.org/10.1016/j.neuroimage.2019.116432

Ishida, K., & Nittono, H. (2022). Relationship between early neural responses to syntactic and acoustic irregularities in music. *European Journal of Neuroscience*, *56*(12), 6201–6214. https://doi.org/10.1111/ejn.15856

Ishida, K., & Nittono, H. (2023). Statistical learning of chord-transition regularities in a novel equitempered scale: An MMN Study. *Neuroscience Letters, 815*, 137478.

https://doi.org/10.1016/j.neulet.2023.137478

Jiang, Y., Komatsu, M., Chen, Y., Xie, R., Zhang, K., Xia, Y., ... & Wang, L. (2022). Constructing the hierarchy of predictive auditory sequences in the marmoset brain. *eLife*, *11*, e74653. https://doi.org/10.7554/ELIFE.74653

Jost, E., Conway, C. M., Purdy, J. D., Walk, A. M., & Hendricks, M. A. (2015). Exploring the neurodevelopment of visual statistical learning using event-related brain potentials. *Brain Research*, *1597*, 95–107. https://doi.org/10.1016/j.brainres.2014.10.017

Kalda, T., & Minati, L. (2012). Detecting scale violations in absence of mismatch requires music-syntactic analysis: A further look at the early right anterior negativity (ERAN). *Brain Topography, 25*(3), 285–292. https://doi.org/10.1007/s10548-011-0208-8

Koelsch, S. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology, 46*(1), 179–190. https://doi.org/10.1111/j.1469-8986.2008.00752.x

Koelsch, S., Busch, T., Jentschke, S., & Rohrmeier, M. (2016). Under the hood of statistical learning: A statistical MMN reflects the magnitude of transitional probabilities in auditory sequences. *Scientific Reports, 6*, 19741. https://doi.org/10.1038/srep19741

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: “Nonmusicians” are musical. *Journal of Cognitive Neuroscience, 12*(3), 520–541. https://doi.org/10.1162/089892900562183

Koelsch, S., Jentschke, S., Sammler, D., & Mietchen, D. (2007). Untangling syntactic and sensory processing: An ERP study of music perception. *Psychophysiology,* *44*(3), 476–490. https://doi.org/10.1111/j.1469-8986.2007.00517.x

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive Processes and the Peculiar Case of Music. *Trends in Cognitive Sciences, 23*(1), 63–77. https://doi.org/10.1016/j.tics.2018.10.006

Lidji, P., Jolicœur, P., Moreau, P., Kolinsky, R., & Peretz, I. (2009). Integrated preattentive processing of vowel and pitch: A mismatch negativity study. *Annals of the New York Academy of Sciences*, *1169*, 481–484. https://doi.org/10.1111/j.1749-6632.2009.04770.x

Miranda, R. A., & Ullman, M. T. (2007). Double dissociation between rules and memory in music: An event-related potential study. *NeuroImage, 38*(2), 331–345. https://doi.org/10.1016/j.neuroimage.2007.07.034

Moldwin, T., Schwartz, O., & Sussman, E. S. (2017). Statistical learning of melodic patterns influences the brain's response to wrong notes. *Journal of Cognitive Neuroscience, 29*(12), 2114–2122. https://doi.org/10.1162/jocn

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology, 42*(1), 25–32. https://doi.org/10.1111/j.1469-8986.2005.00256.x

Okubo, M., Suzuki, H., & Nicholls, M. E. R. (2014). A Japanese version of the FLANDERS handedness questionnaire. *Shinrigaku Kenkyu*, *85*(5), 474–481. https://doi.org/10.4992/jjpsy.85.13235

Paavilainen, P., Valppu, S., & Näätänen, R. (2001). The additivity of the auditory feature analysis in the human brain as indexed by the mismatch negativity: 1+1≈2 but 1+1+1<3. *Neuroscience Letters*, *301*, 179–182. https://doi.org/10.1016/S0304-3940(01)01635-4

Paraskevopoulos, E., Chalas, N., Kartsidis, P., Wollbrink, A., & Bamidis, P. (2018). Statistical learning of multisensory regularities is enhanced in musicians: An MEG study. *NeuroImage*, *175*, 150–160. https://doi.org/10.1016/j.neuroimage.2018.04.002

Recasens, M., Grimm, S., Wollbrink, A., Pantev, C., & Escera, C. (2014). Encoding of nested levels of acoustic regularity in hierarchically organized areas of the human auditory cortex. *Human Brain Mapping*, *35*(11), 5701–5716. https://doi.org/10.1002/hbm.22582

Rogers, L. L., Park, S. H., & Vickery, T. J. (2021). Visual statistical learning is modulated by arbitrary and natural categories. *Psychonomic Bulletin and Review*, *28*(4), 1281–1288. https://doi.org/10.3758/s13423-021-01917-w

Rohrmeier, M. A., & Koelsch, S. (2012). Predictive information processing in music cognition. A critical review. *International Journal of Psychophysiology*, *83*(2), 164–175. https://doi.org/10.1016/j.ijpsycho.2011.12.010

Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin and Review, 25*(1), 128–142. https://doi.org/10.3758/s13423-017-1230-y

Siegelman, N., Bogaerts, L., Elazar, A., Arciuli, J., & Frost, R. (2018). Linguistic entrenchment: Prior knowledge impacts statistical learning performance. *Cognition*, *177*, 198–213. https://doi.org/10.1016/j.cognition.2018.04.011

Stärk, K., Kidd, E., & Frost, R. L. A. (2022). The effect of children's prior knowledge and language abilities on their statistical learning. *Applied Psycholinguistics*, *43*(5), 1045–1071. https://doi.org/10.1017/S0142716422000273

Sussman, E. S., Chen, S., Sussman-Fort, J., & Dinces, E. (2014). The five myths of

MMN: Redefining how to use MMN in basic and clinical research. *Brain Topography*, *27*(4), 553–564. https://doi.org/10.1007/s10548-013-0326-6

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters*, *266*, 109–112. https://doi.org/10.1016/S0304-3940(99)00267-0

Tsogli, V., Jentschke, S., Daikoku, T., & Koelsch, S. (2019). When the statistical MMN meets the physical MMN. *Scientific Reports, 9*(1), 5563. https://doi.org/10.1038/s41598-019-42066-4

Turk-browne, N. B., Junge, J. A., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–564. https://doi.org/10.1037/0096-3445.134.4.552

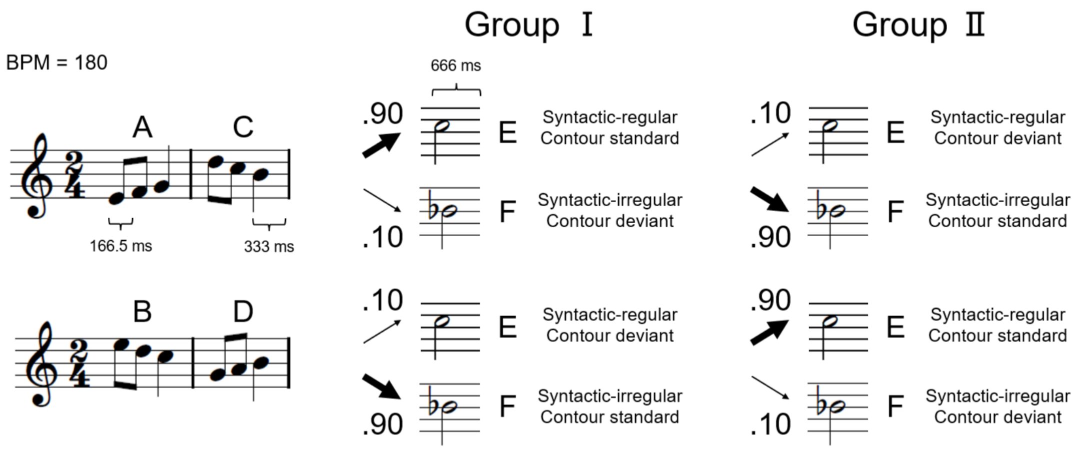
Vuust, P., Heggli, O. A., Friston, K. J., & Kringelbach, M. L. (2022). Music in the brain. *Nature Reviews Neuroscience, 23*(5), 287–305. https://doi.org/10.1038/s41583-022-00578-5

Winkler, I., & Czigler, I. (2012). Evidence from auditory and visual event-related potential (ERP) studies of deviance detection (MMN and vMMN) linking predictive coding theories and perceptual object representations. *International Journal of Psychophysiology*, *83*(2), 132–143. https://doi.org/10.1016/j.ijpsycho.2011.10.001

Wolff, C., & Schröger, E. (2001). Human pre-attentive auditory change-detection with single, double, and triple deviations as revealed by mismatch negativity additivity. *Neuroscience Letters*, *311*, 37–40. https://doi.org/10.1016/S0304-3940(01)02135-8

Figure 1.

Combinations of phrases and ending notes



Note. Two phrases and ending notes were concatenated to create four types of melodies: with or without music-syntactic irregularity and with or without melodic-contour deviance. Standard: syntactically regular and frequent final notes of the melody. Contour deviant: syntactically regular and infrequent final notes of the melody. Syntactic deviant: syntactically irregular and frequent final notes of the melody. Double deviant: syntactically irregular and infrequent final notes of the melody.

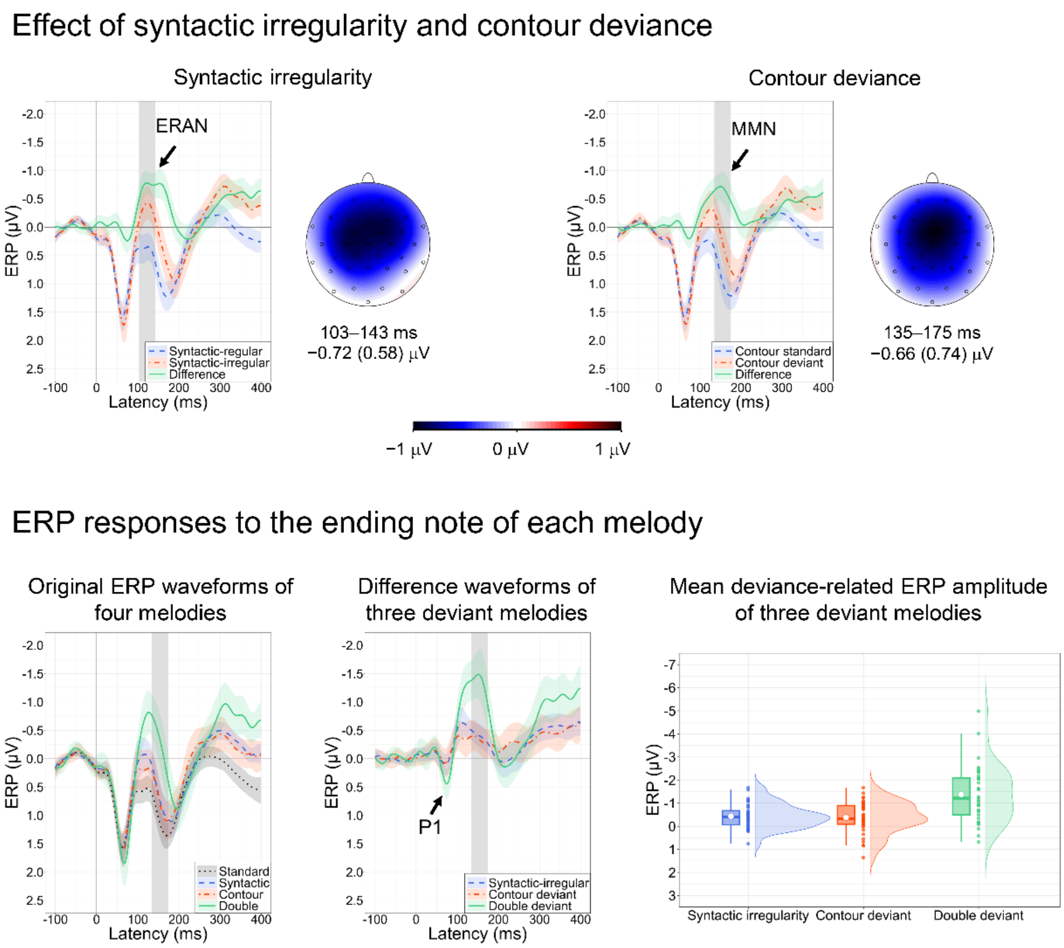
Table 1.

Distribution of syntactic irregularity and transition probabilities in the ending note of melodies

	Musical syntax	
Melodic contour	Regular (.50)	Irregular (.50)
Standard (.90)	.45	.45
Deviant (.10)	.05	.05

Figure 2.

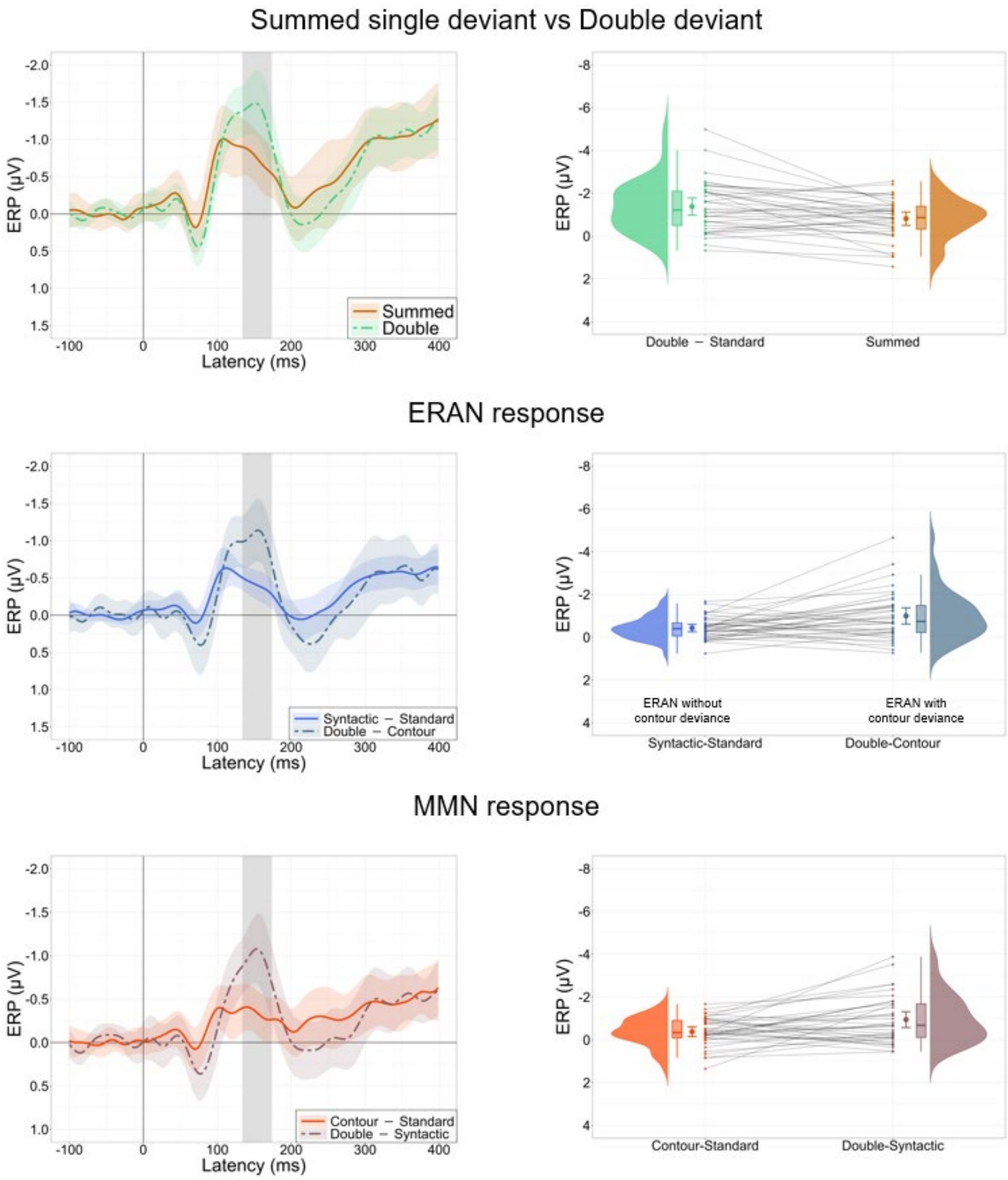
Deviance-related responses and interactive effect of irregularity and deviance factors



Note. The top panel shows the effects of the music-syntactic irregularity and the contour deviance. The distribution of the mean ERP amplitude is shown in each topographic map; below that is the mean ERP amplitude (standard deviation) for that window. The bottom panel shows the ERP responses to the ending note of each melody. The two figures on the left show the grand average ERP waveforms (means of the five frontal electrodes: F7, F3, Fz, F4, and F8) with 95% confidence intervals. In the right raincloud plots, the white dots indicate the grand mean ERP amplitudes across the participants.

Figure 3.

Deviance-related ERP responses under the presence or absence of the other type of deviance



Note. The top panel shows the comparison of the summed single-deviant ERP amplitude (syntactic irregularity + contour deviant) and the double-deviant ERP amplitude. The middle panel shows the ERAN responses calculated by syntactic irregularity − standard

(ERAN without contour deviance) and double deviant − contour deviant (ERAN with contour deviance). The bottom panel shows the MMN responses calculated by contour deviance − standard (MMN without syntactic irregularity) and double deviant − syntactic irregularity (MMN with syntactic irregularity). The ERP waveforms (means of the five frontal electrodes: F7, F3, Fz, F4, and F8) are shown with 95% confidence intervals. The raincloud plots show the mean ERP amplitudes (134–174 ms). The large dots indicate the grand mean ERP amplitudes across the participants with 95% confidence intervals.