



Title	Which instrument should play here? Decoding predicted musical timbre from EEG signals during omission
Author(s)	Ishida, Kai; Ishida, Tomomi; Nittono, Hiroshi
Citation	Cortex. 2025, 192, p. 64-77
Version Type	VoR
URL	https://hdl.handle.net/11094/102879
rights	This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

<https://ir.library.osaka-u.ac.jp/>

The University of Osaka



Research report

Which instrument should play here? Decoding predicted musical timbre from EEG signals during omission



Kai Ishida^{*}, Tomomi Ishida and Hiroshi Nittono

Graduate School of Human Sciences, The University of Osaka, 1-2 Yamadaoka, Suita, Osaka 565-0871, Japan

ARTICLE INFO

Article history:

Received 15 February 2025

Revised 27 August 2025

Accepted 28 August 2025

Action editor Sascha Frühholz

Published online 8 September 2025

Keywords:

Predictive coding

Omitted stimulus potential

Decoding

Musical timbre

Prediction error

Heartbeat evoked potential

ABSTRACT

The human brain predicts various musical features such as harmony, melody, and rhythm during music perception. A previous electroencephalographic (EEG) study showed that the accuracy of pitch decoding during tone omission was greater when the pitch of the melody was highly predictable than when it was less predictable, reflecting that predictive information of a specific pitch is contained in the EEG signal. However, the specificity of prediction for other musical features has not been fully addressed. The present study investigated whether predicted instruments are decoded from the EEG signal during omission to examine the specificity of prediction in the timbre dimension. Thirty-five participants listened to unfamiliar melodies with simple (high predictability) or complex (low predictability) timbre change rules while watching a silent movie. The EEG was recorded when a tone expected to be played by one of four specific timbres (celesta, electric piano, marimba, organ) was omitted. The results showed that the amplitude of an omitted stimulus potential, oN1, did not differ between high and low predictability conditions. However, the support vector machine was able to decode the type of musical timbre during omission better than random chance in the high predictability condition but not in the low predictability condition. These results suggest that EEG signals contain information about which instrument should be played during omission, but this information is not manifested in traditional event-related potentials. The brain may specifically predict not only the pitch but also other musical dimensions, such as the timbre, of the upcoming tone.

© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. Introduction

The human brain predicts the timing and content of incoming sounds (Denham & Winkler, 2020). In the music domain, the

presence of prediction is widely accepted, and music perception has been studied in terms of predictive coding (Vuust et al., 2022). In predictive coding, humans actively perceive an external input with a prediction, and the prediction error between the prediction and the sensory input is minimized to

^{*} Corresponding author. Graduate School of Human Sciences, The University of Osaka, 1-2 Yamanaka, Suita, Osaka 565-0871, Japan.

E-mail addresses: ishida@hus.osaka-u.ac.jp (K. Ishida), t-ishida@hus.osaka-u.ac.jp (T. Ishida), nittono@hus.osaka-u.ac.jp (H. Nittono).
<https://doi.org/10.1016/j.cortex.2025.08.013>
 0010-9452/© 2025 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC license (<http://creativecommons.org/licenses/by-nc/4.0/>).

explain the sensory input (Friston, 2005, 2010a, 2010b). The prediction is sent as a top-down signal and the prediction error is sent as a bottom-up signal in a hierarchical message passing (Friston & Kiebel, 2009). Following this prediction framework, previous studies have examined various types of predictions, such as chord (Li et al., 2021; Ono et al., 2024), pitch (Ishida & Nittono, 2024c; Quiroga-Martinez et al., 2019), and rhythm (Vuust et al., 2009; Zanto et al., 2006) during music perception (for reviews, see Koelsch et al., 2019; Rohrmeier & Koelsch, 2012; Vuust et al., 2018, 2022; Vuust & Witek, 2014). Although previous studies have demonstrated the presence of prediction, the feature specificity of prediction has not been fully investigated. For the pitch feature, previous studies have shown the specificity of pitch prediction in the simple tone sequence (Chouiter et al., 2015; Demarchi et al., 2019; Hauswald et al., 2024) and musical melody (Ishida et al., 2024). However, it is unclear whether the content prediction of musical timbres (instruments) is also feature specific. Therefore, the present study investigated whether the human brain generates a neural signal specific to the particular musical timbre while listening to musical melodies.

Direct evidence for the presence of prediction is the neural omission response to unexpected omissions (Bendixen et al., 2014; Korka et al., 2020; SanMiguel, Saupe, & Schröger, 2013; van Laarhoven et al., 2017; for review, see Bendixen et al., 2012; Schröger et al., 2015). Neural activation during omission can be a mirror of a top-down predictive signal because there is no external input (SanMiguel, Saupe, & Schröger, 2013; Schröger et al., 2015). Electroencephalographic (EEG) studies have measured the neural omission response through the event-related potential (ERP) recorded during the omission of sensory input in various sensory modalities (visual: Ishida & Nittono, 2024a; Nittono, 2005; Stange et al., 2023; auditory: Dercksen et al., 2020; Horváth et al., 2010; Jongma et al., 2004; Korka et al., 2020; Stekelenburg & Vroomen, 2015; somatosensory: Dercksen et al., 2024; Hernando-Hernández & Hernández-Sánchez, 2017). This ERP is called omitted stimulus potentials (OSPs) or omitted stimulus response (Bullock et al., 1994; Busse and Woldorff, 2003; Hernando-Hernández & Hernández-Sánchez, 2017; Ishida & Nittono, 2024a, 2024b). In the auditory domain, the first OSP is omission N1 (oN1), which appears with a latency of approximately 50–100 msec after omission onset (Ishida & Nittono, 2024a; SanMiguel, Saupe, & Schröger, 2013; SanMiguel, Widmann, et al., 2013; van Laarhoven et al., 2017, 2020). The amplitude of oN1 was larger when the timing and content of the tone were predictable than when they were unpredictable (Dercksen et al., 2020; Ishida et al., 2024; SanMiguel, Saupe, & Schröger, 2013; van Laarhoven et al., 2017). Ishida et al. (2024) discussed that oN1 reflects the precision-weighted prediction error of the unexpected omission, and that its amplitude modulation is driven by the precision of the sensory input due to predictability (certainty).

In the auditory domain, previous studies have shown that neural omission responses reflect predictive information for specific pitch features using decoding methods that predict mental states from neural activity using statistical or machine learning techniques. For example, Demarchi et al. (2019) showed that the accuracy of pitch decoding from magnetoencephalographic responses during tone omission and before

tone presentation in a tone sequence increases with predictability due to sequential regularity, suggesting that pitch expectations generated in auditory prediction become more specific as predictability increases. Hauswald et al. (2024) further demonstrated that the frequency-specific neural pattern during omission was not affected by the prediction of omission positions using the same multivariate pattern analysis decoding. Therefore, the content of tones is specifically predicted, regardless of the position of the omission.

In a musical context, Ishida et al. (2024) reported that oN1 was larger for omission in the familiar melody than in the unfamiliar melody created by shuffling the notes of familiar melodies, reflecting pitch predictability. Ishida et al. further showed that the accuracy of decoding omitted notes from the EEG during omission was higher in the familiar melody context (high predictability) than in the unfamiliar melody context (low predictability) using a support vector machine (SVM), a type of machine learning. They concluded that EEGs in predictable contexts contain more information about musical pitch expectation and that this expectation is more specific than in less predictable contexts. However, in their study, the harmonic and rhythmic structure before the omission was not controlled between the familiar and unfamiliar contexts, except for the preceding one note, due to the random shuffling of notes. Therefore, the amplitude difference in oN1 and decoding accuracy may have been influenced not only by predictions of specific musical pitches but also by overall predictability, such as harmony and rhythm.

The present study investigated whether the omission response reflects the various content predictions in the auditory domain. To this end, two aims were set. The first aim was to examine whether information reflecting predicted instruments is contained in the EEG signal during omission by extending the results of pitch-identity decoding (Ishida et al., 2024) to timbre-identity decoding. Several studies have shown that the representation of distinct timbres is reflected in different neural responses (Caclin et al., 2006; Seol et al., 2011; for a review, Wei et al., 2022). For example, a magnetoencephalography study by Seol et al. (2011) demonstrated a difference in the auditory M50 (P50) and M100 (N1) responses based on timbre characteristics, suggesting that timbre is discriminated during early auditory processing. An ERP study also showed that differences in timbre imagery, manipulated by the spectral centroid, reflect differences in the amplitude of the late positive component (Tužnik et al., 2018). Therefore, separate predictions would be generated for timbre identity.

The second aim was to replicate the oN1 results of Ishida et al. (2024) by manipulating the predictability of the target tone content while controlling other predictive information. Thus, two melodies with the same harmonic and rhythmic structures were composed. The tone at the target positions of these melodies was occasionally omitted, and its timbre was one of four timbres (target timbres). The timbres at positions other than the target were drawn from different set of four timbres (contextual timbres). One melody was used as a high predictability condition and another as a low predictability condition. In the high predictability condition, the contextual timbre was constant and the order of the target timbres was constant. In the low predictability condition, each position in the melody was played in one of four contextual timbres, and

the target timbres at each of the four positions were defined according to which contextual timbres appeared before the target position (see details in the Stimuli section). Thus, the regularity of timbre order was more complex in the low predictability condition than in the high predictability condition. If oN1 reflects the precision-weighted prediction error, the oN1 amplitude would be larger in the high predictability condition than in the low predictability condition (H1). Notably, if the specificity of timbre expectations changes according to the predictability of tone content, the decoding accuracy of the omitted timbre would be greater in the high predictability condition than in the low predictability condition (H2).

The present study further investigated the effect of predictive processing for exteroception on interoceptive processing during omission and tone presentation by recording the heartbeat evoked potential (HEP). HEP is a neurophysiological measure of cardiac interoceptive processing and is extracted by aligning the HEP to the R-peak of the electrocardiogram (ECG) (Coll et al., 2021). The HEP is larger when interoception is attended than when exteroception is attended (Petzschner et al., 2019). One possibility is that internal processing, such as interoceptive and predictive processing, becomes more dominant during omission, whereas external processing, such as bottom-up sensory processing, becomes more dominant during tone presentation. Another possibility is that emotional processing also affects HEP amplitude. The positive HEP amplitude during the high emotional valence and arousal is larger than the low, at around 200–400 msec (Fuseda and Katayama, 2021; Luft & Bhattacharya, 2015). Given that the omission of the expected tone elicits the informative surprise (i.e., the Shannon surprise) because it elicits the prediction error, the omission would cause higher valence and arousal than the tone. Therefore, examining HEP during omission was expected to provide further insight into the predictive nature of the omission response through the lens of interoception. Considering these possibilities, it was expected that HEP amplitude would be larger during the omission than during the tone presentation (H3). Moreover, the difference in HEP amplitude between the high and low predictability conditions was explored separately during tone, omission, and expected omission presentations.

2. Method

2.1. Participants

An effect size of $d_z = .73$ (Ishida et al., 2024) was used to detect the difference in oN1 amplitude. Power analysis using G*power (Faul et al., 2007) resulted in a sample size of $N = 17$ with $\alpha = .05$ and $1 - \beta = .80$. However, considering data exclusion and the small effect size on decoding accuracy, 40 participants were recruited and randomly assigned to four groups with different orders of high and low conditions and combinations of two melodies to counterbalance the order of conditions and the melody-specific effect (two condition orders $\times$ two melodies). Finally, data from 35 participants (16 women and 19 men, 18–27 years old, $M = 21.3$ years) were used for hypothesis testing regarding the OSP analysis. Five

participants were excluded due to excessive noise, which resulted in fewer than 45 artifact-free ERP trials for each stimulus type (180 trials in total). Of these, 33 participants were right-handed and two were ambidextrous, as per the FLANDERS handedness questionnaire (Okubo et al., 2014). Data from 31 participants (16 women and 15 men, 18–27 years old, $M = 21.3$ years) were used for the analysis of HEP. Ten participants were excluded because they had fewer than 100 artifact-free trials in total, with fewer than 50 trials from the high and low predictability conditions, respectively. None of the participants had a history of hearing impairment, neurological or cardiovascular disease. The participants' musical ability was assessed using the Japanese Gold-MSI questionnaire (Sadakata et al., 2023) and is summarized in Supplementary Table S1. They had various histories of formal musical training, with the following distribution: 18 participants reported 0 years of training, 2 had .5 years, 1 had 2 years, 6 had 3–5 years, 2 had 6–9 years, and 6 had over 10 years. The protocol of this study was approved by the Behavioral Research Ethics Committee of Osaka University School of Human Sciences, Japan (HB024-016), and written informed consent was obtained from all participants. Participants received a cash voucher of 4,000 Japanese yen as an honorarium. This study was conducted after another unrelated experiment (approximately 1 h) in which the EEG was recorded while participants listened to a syncopated rhythm while tapping in sync with the rhythm (Ishida & Nittano, 2025). Although performing a different type of musical task beforehand may cause fatigue, the effect is considered minimal because these experiments used completely different musical stimuli and tasks.

2.2. Stimuli

The score of the stimulus is shown in Fig. 1. Two melodies with the same harmonic and rhythmic structures were composed, and four target positions were placed at the same position in each melody. The melodies were in C major and 4/4 time signature, and the total length was 9.6 sec. All tones in the melodies were quarter notes with a duration of 300 msec (BPM = 200). This tone duration and tempo were the same as those used by Ishida et al. (2024), which showed successful decoding of predicted pitch. This enabled us to compare the results of pitch and timbre decoding studies. The tone in the target positions was always A (440 Hz) after the piano tone of G, and its timbre was one of celesta, organ, electric piano, or marimba (target timbres). Using the same pitch consistently as the target tone ensures that decoding was performed solely based on timbre differences. The timbres other than the target positions were those of violin, trumpet, saxophone, and horn (contextual timbres). A melody was used as the high predictability condition. In the high predictability condition, the contextual timbre was constant at the contextual positions within a melody, and the order of the target timbres was constant and uniquely defined by the contextual timbre. For example, if the contextual timbre was violin, the order of the target timbres at the four target positions was celesta–organ–electric piano–marimba (see Fig. 1A).

Another melody was used as the low predictability condition. In the low predictability condition, the timbre at the

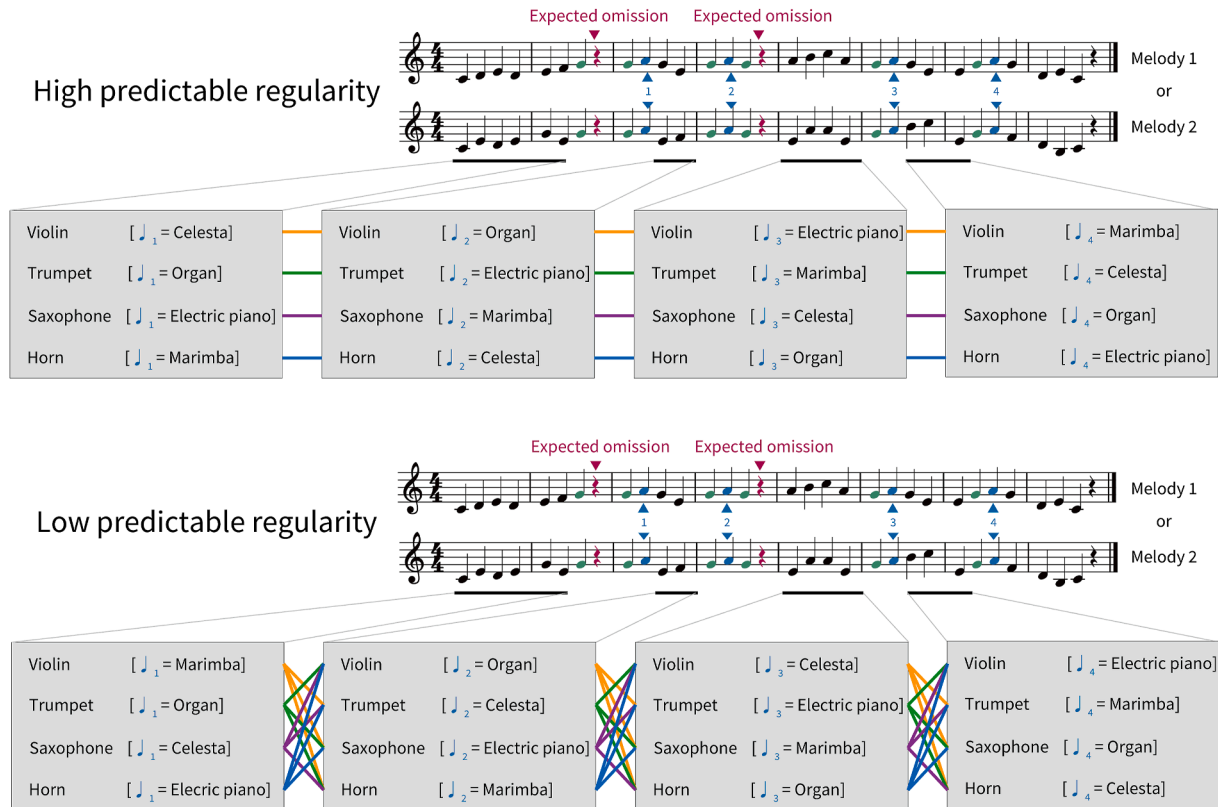


Fig. 1 – Timbre Regularities of High and Low Predictability Conditions

Note. The top panel shows an example of the high predictability condition, and the bottom panel shows an example of the low predictability condition. Blue notes with blue numbers indicate the target positions. In the high predictability condition, the contextual timbre is constant at the contextual positions within a melody. In the low predictability condition, the contextual timbres changed randomly, as indicated by the orange, green, purple, and blue lines. The target tones were omitted with a probability of 50%. The occurrence probability of each target timbre (celesta, organ, electric piano, and marimba) at each target position was the same in the high and low predictability conditions. In both types of melodies, the target timbres were omitted after the piano G, which is colored green. The pink quarter rest is the expected omission where notes were never presented.

target position was determined by which contextual timbre preceded it. However, within a single melody, four contextual timbres were played and their order was randomized, as shown in the bottom panel of Fig. 1. For example, if the order of the contextual timbres was violin–trumpet–horn–saxophone, the order of the target timbres in a trial was marimba–celesta–organ–organ. Therefore, in the low predictability condition, participants had to learn combinations of the four contextual and the four target timbres at each position (i.e., 16 patterns), whereas in the high predictability condition, they only had to learn the four orders of the target timbres. This means that the present study manipulated predictability by changing the stability of timbres, which made it difficult to learn timbre patterns. Because the present study aimed to manipulate predictability based on contextual timbre regularities rather than a positional prediction that a particular timbre would be presented at a particular position, all patterns of each contextual position were presented randomly in each condition so that all four timbres appeared with equal probability at each target position. Note that which of the two melodies was assigned to the

high or low predictability condition was counterbalanced across participants.

At each target position, notes were omitted with a probability of 50%. All melodies were presented without interstimulus interval. In addition to the physically identical G tone preceding the target tones, the same harmonic and rhythmic structures controlled predictability other than timbre predictability. Thus, the difference in the omission response between the high and low predictability conditions was expected to reflect the predictability of musical timbres based on the complexity of timbre regularity, rather than the late components elicited by the preceding tone before the omission. The presentation of the G tone also equalized the timbre change at the target position between the high and low predictability conditions. In addition, the expected omission, where the tone was never presented, was placed after the piano G tone to examine the difference in ERPs between the unexpected and expected omissions, as this difference was expected to reflect the effect of a tone presentation prediction.

For the melody and trigger output, a stereo audio file was created with the melody sequence in the first channel and the

trigger sound indicating the onset of the tones and omissions in the second channel. These channels were output separately via a stereo-to-monoaural splitter cable, and the melody and trigger sound were completely separated. The melody channel was connected to left and right headphones (MDR-EX650AP; SONY, Tokyo, Japan) at 60 dB SPL. The trigger channel was connected to an auditory signal detector (StimTrak; Brain Products, Gilching, Germany), which immediately (<1 msec) sent a trigger to an EEG amplifier.

2.3. Procedure

Prior to EEG recording, participants were asked to complete the Japanese Gold-MSI (Sadakata et al., 2023) to assess musical ability and the FLANDERS handedness questionnaire (Okubo et al., 2014) to assess handedness. Participants responded to all questionnaire items using a 7-point Likert scale ranging from 1 = “Completely disagree” to 7 = “Completely agree” in the Gold-MSI.

The EEG and ECG recordings consisted of four blocks of the high predictability condition and four blocks of the low predictability condition. The order of the high and low predictability conditions was counterbalanced across participants. In each block, the melody was repeated 32 times, and the tone presentation and omission resulted in a total of 16 trials for each target timbre. Thus, the total number of tone and omission trials in each condition was 256 and 256 trials, respectively (i.e., $64 + 64 + 64 + 64 = 256$ trials). In the high predictability condition, all four contextual instruments were presented within a single block to control for the type of instruments heard and to equalize the sensory processing as much as possible between the high and low predictability conditions. Participants were asked to ignore the melodies while watching a silent movie. Including the online questionnaire session, electrode preparation, and short breaks between blocks, the entire experiment took approximately 2.5 h.

2.4. Data recording

EEG and ECG data were recorded using QuickAmp (Brain Products, Germany) with Ag/AgCl electrodes. Thirty-four scalp electrodes were placed 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 mean of the left and right mastoid electrodes. The sampling rate was 1000 Hz. The online filter was DC–200 Hz. Electrode impedances were kept below 10 k Ω . The ECG was recorded bipolarly using two Ag/AgCl electrodes placed on the left lower rib and the right mastoid sites (as an alternative to the right clavicle electrode).

2.5. Data processing

EEG data were analyzed using EEGLAB (Delorme & Makeig, 2004; version 2024.2.1) and ERPLAB (Lopez-Calderon & Luck, 2014; version 12.00) on MATLAB R2023a (The MathWorks

Inc., Natick, MA). To calculate OSP waveforms, a digital filter composed of a .5-Hz high-pass filter and a 25-Hz low-pass filter was first applied to the data (Ishida et al., 2024; SanMiguel, Widmann, et al., 2013). Then, ocular artifact correction based on independent component analysis was applied. A period of 500 msec (200 msec before and 300 msec after the omissions and tones) was averaged after removing trials with voltages exceeding ± 80 μ V in any channel, as per Ishida et al. (2024). Baseline correction was applied by subtracting the mean amplitude of the 200 msec pre-stimulus period from each point of the waveform. The ERP waveform was calculated by averaging the fronto-central electrodes (Fz, Cz, FC1, FC2) as in Ishida et al. (2024), because the oN1 distributed around that region.

For the statistical evaluation of oN1, the grand mean waveforms of the omissions in the high and low predictability conditions were averaged (averaged grand mean waveforms). The peak of oN1 was then detected in the interval of 60–120 msec of the averaged grand mean waveforms of the omissions in the high and low predictability conditions, and the interval ± 10 msec from the negative peak was defined as the oN1 interval. This extraction method was preliminary determined based on the results of Ishida et al. (2024: oN1 time window was 99–119 msec). Finally, a period of 94–114 msec was defined as the oN1 interval, and the mean ERP amplitudes of tone, omission, and expected omission were calculated from this interval. The number of epochs is summarized in Supplementary Table S2.

To calculate HEP waveforms, R-peaks were automatically detected using HEPLAB (Perakakis, 2019; version 1.0.1), and peaks that were misidentified or not properly detected were manually corrected. R-peaks occurring during tone, omission, and expected omission (i.e., 300 msec duration) were used as triggers for HEP onset. After the detection of R-peaks, a digital filter composed of a .1-Hz high-pass filter and a 30-Hz low-pass filter was applied to the data as in Tanaka et al. (2023), who examined HEP with tone presentation. Independent component analysis was used to remove ocular, muscular, and cardiac field artifacts. An 800 msec period (200 msec before and 600 msec after the R-peak) was averaged after removing trials with voltages exceeding ± 80 μ V in any channel. Baseline correction was applied by subtracting the mean amplitude of the –200 msec to –50 msec period from each point of the waveform. The –50–0 msec were not included in the baseline because this interval was expected to contain an R-wave deflection (Coll et al., 2021). Because the HEP was prominent, the fronto-central electrodes (Fz, Cz, FC1, FC2) were used to calculate the ERP waveform.

For the statistical evaluation of HEP, the grand mean waveforms of the HEP in the tone, omission, and expected omission were averaged. The positive peak was observed around 300–400 msec. The peak was detected in this time window, and the interval ± 20 msec from the peak was used to calculate the averaged ERP amplitude (348–388 msec). Moreover, to examine the difference in the predictability of timbre in HEP amplitude, the difference waveforms were calculated by subtracting the HEP of the low predictability condition from the HEP of the high predictability condition in tone, omission, and expected omission. The grand mean waveforms of all three difference waveforms were averaged. Because the

positive deflection was found around 50–150 msec, the positive peak was detected in this time window and the interval ± 20 msec from the peak was used to calculate the averaged ERP amplitude (77–97 msec). The number of epochs is summarized in [Supplementary Table S2](#).

2.6. Decoding

Similar to [Ishida et al. \(2024\)](#), a participant-based approach to decoding was performed using linear SVM with error-correcting output codes (ECOC). The *templateSVM* function, a MATLAB function, was used to standardize the data, and the linear SVM with ECOC was run using the *fitcecoc* function. The SVM was run separately for each participant at each time point of the OSP waveforms. The features input to the classifier were the scalp electrode potentials from all 34 channels. Before data input, the number of trials was balanced across all classes. That is, for individual-level decoding, the classifier was trained using an equal number of trials by matching the count to the category with the fewest trials within participant. At each time point, a threefold cross-validation was conducted to assess the generalizability of the model. In the threefold cross-validation, all trials of each timbre category were randomly divided into three blocks. Two of the three blocks were used for training, and the remaining block was used for testing the classifier to calculate the decoding accuracy. This process was repeated three times until all three blocks were used as the test block. Then, the averaged decoding accuracy over three test datasets was calculated. For each time point, threefold cross-validation was repeated 10 times (iteration) and the averaged decoding accuracy was calculated. Decoding was performed for the full range of –200–300 msec after the onset of tone, omission, and expected omission, and the full range of –200–600 msec after the onset of the R-peak. The decoding accuracy waveform was compared to the chance level to test whether the accuracy significantly exceeded it using a cluster-based permutation test (see 2.7. Statistical Analysis for details).

2.7. Statistical analysis

To statistically evaluate ERP amplitudes and decoding accuracy, both classical (frequentist) and Bayesian analyses were performed using JASP .19.2 (JASP Team, 2024). The type I error rate (α) was set at .05. Bayes factors were calculated to assess the absence (null hypothesis) or presence (alternative hypothesis) of the difference between conditions. In the Bayesian t-test, the prior distribution for the effect δ was a Cauchy distribution (scale parameter r of .707). A Bayes factor greater than 3 was considered moderate evidence for the alternative and null hypotheses, respectively, as per [Schönbrodt and Wagenmakers \(2018\)](#).

To examine the difference in omissions with and without prediction, the oN1 amplitude of the omission and expected omission were compared using a paired t-test (one-sided) and its Bayesian version. Then, a paired t-test (two-sided) and its Bayesian version were performed on the oN1 amplitude to compare the difference in the high and low predictability conditions. The same t-tests (two-sided) were performed on the N1 and oN1 amplitude of tone and expected omission. The

Bayes factor was expressed as BF_{-0} for the one-sided test and BF_{10} for the two-sided test.

The decoding results were statistically analyzed using a cluster-based permutation test ([Maris & Oostenveld, 2007](#)). First, the decoding accuracy in classifying the omission and expected omission in the high and low conditions was compared to chance level ($p = .50$), separately, to examine whether the omission with and without prediction was different. Second, the accuracy of decoding using the OSP and ERP of tone in classifying the four timbres was compared to chance level ($p = .25$) in the high and low conditions separately. The chance level was determined by dividing 100% by the number of categories, as with previous studies that used a similar SVM decoding method ([Bae & Luck, 2019](#); [Ishida et al., 2024](#); [Tautvydaitė & Burra, 2024](#)). The decoding accuracy of timbre classification was compared between the high and low conditions for tone and omission decoding separately.

The positive HEP amplitude was compared between the three stimuli (tone, omission, and expected omission), using one-way analysis of variance (ANOVA) and its Bayesian version. Spherical correction was performed to modify the degrees of freedom using Greenhouse-Geisser ϵ . For multiple comparisons, Bonferroni correction was applied to control for type I error rate (α). Furthermore, the decoding accuracy in classifying the HEPs of the three stimuli was compared to chance level ($p = .33$) using a cluster-based permutation test. Finally, the difference in HEP amplitudes (high minus low conditions) in the three stimuli was compared to zero using a one-sample t-test and its Bayesian version.

The electrical sources of the ERP and HEP were estimated using the standardized low-resolution brain electromagnetic tomography analysis (sLORETA) method ([Pascual-Marqui, 2002](#)). sLORETA was applied to the mean voltages of oN1, N1, and positive HEP. Three-dimensional current density magnitudes (sLORETA-xyz values) were compared to zero and within conditions and stimuli using voxel-wise paired t-tests. For the multiple comparison, corrected critical values for significant differences ($p < .05$) were determined using the nonparametric permutation test. Materials and dataset for replicating the analysis are available from <https://osf.io/awx7e/>.

3. Results

3.1. Difference in OSP waveforms between omission and expected omission

[Fig. 2](#) shows the ERP responses to tones and omissions and their source localization by sLORETA with decoding results. Similar ERP waveforms were obtained for the omission and the expected omission. To verify that the ERP amplitude around the oN1 time window was larger in the omission than in the expected omission, the amplitude was compared after collapsing the high and low conditions. The ERP amplitude of the omission ($M = -2.00 \mu V$, $SD = 1.03$) was significantly larger than that of the expected omission ($M = -1.64 \mu V$, $SD = 1.05$), $t(34) = -2.96$, $p = .003$, $d_z = -.50$, $BF_{10} = 14.10$, suggesting that the differential ERP response was due to the oN1 component elicited by the unexpected omission.

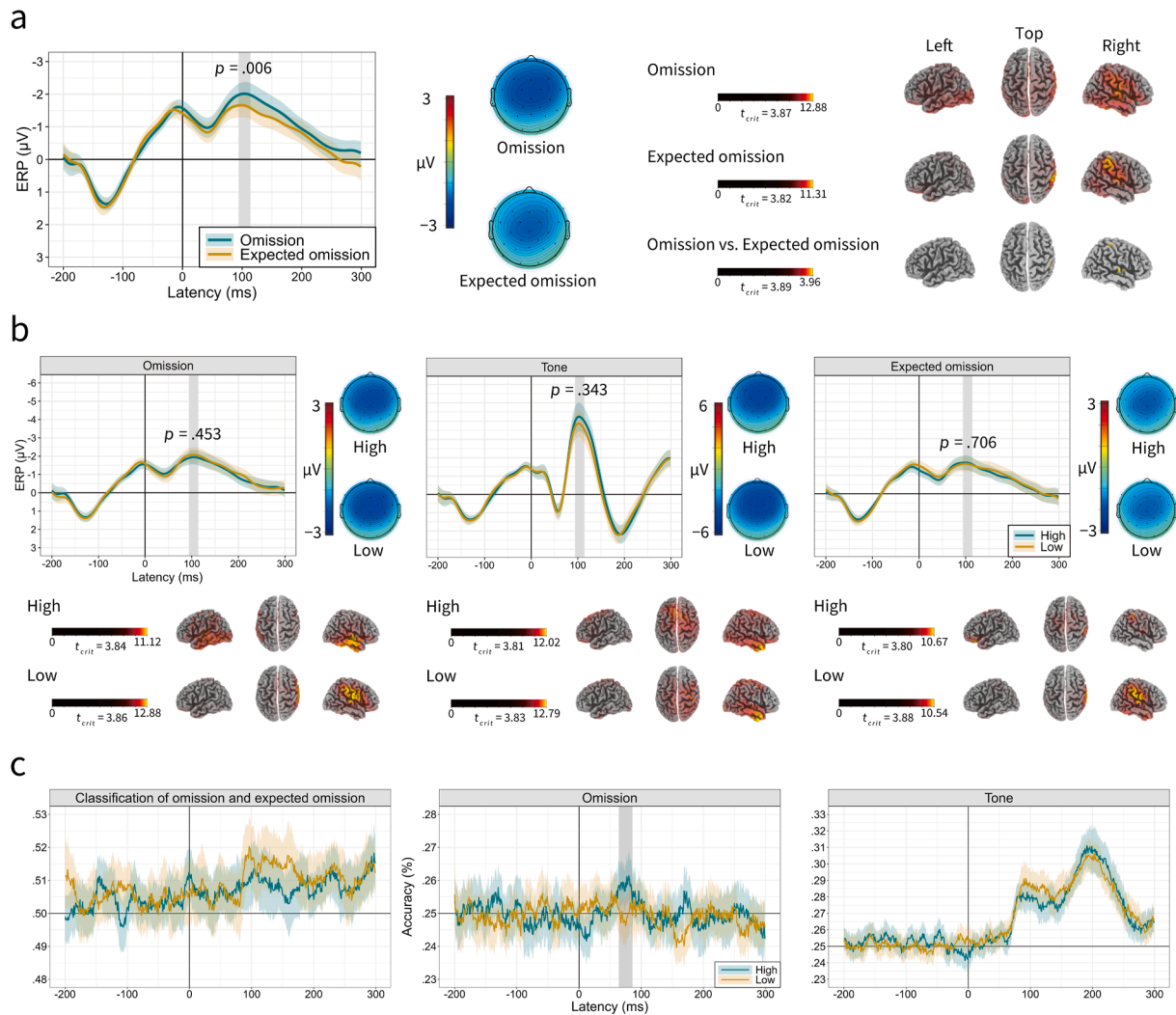


Fig. 2 – ERP Responses and Decoding Results

Note. (a) Grand ERP waveforms of the omission and expected omission when averaging the high and low predictability conditions. The gray area indicates the oN1 interval (94–114 msec) defined as the peak ± 10 msec. The topographic distribution shows the mean ERP amplitude of the oN1 interval. The right panel shows the estimated cortical sources of the omission and the expected omission. (b) The ERP responses to tone, omission, and expected omission. The topographic distribution shows the mean ERP amplitude at 94–114 msec, indicated by the gray area in the waveform panels. The bottom panel shows the estimated cortical sources of each stimulus in the high and low predictability conditions. (c) Decoding accuracy for the omission and expected omission classifications and for the classification of the four timbres using tone and omission. For all panels, the ERP waveforms were calculated by averaging the frontal-central electrodes (Fz, FC1, FC2, and Cz). The light-colored bands on the ERP and decoding accuracy waveforms indicate a 95% confidence interval. For sLORETA results, only voxels with statistically significant activities exceeding the critical t-values (t_{crit}) according to nonparametric t-tests ($p < .05$) are colored.

The right panel of Fig. 2a shows the source localization of the ERP around the oN1 time window (i.e., the mean voltage of the 94–114 msec interval) using sLORETA. The results show that both the omission and the expected omission activated around the superior temporal gyrus (STG) compared to zero. When omission and expected omission were compared, the STG was more activated during the omission than during the expected omission. These results suggest that the oN1 response originates in the auditory cortex and that the

stronger STG activation may reflect the violation of the tone presentation prediction, overlapping with unspecified late ERP components elicited by the preceding tones.

The left panel of Fig. 2c shows the SVM results of the omission and expected omission classification in the high and low predictability conditions. The cluster-based permutation test revealed that the decoding accuracy was significantly above chance level ($p = .50$) in both the high, 77–128 msec cluster $t_{sum} = 123.52$, $p = .010$, and the low, 87–253 msec

cluster, $t_{sum} = 520.20$, $p < .001$, predictability conditions. The difference between omission and expected omission reflects the difference between the presence and absence of an expectation for the tone presentation.

3.2. Comparison of ERP between the high and low predictability conditions

Fig. 2b shows the ERP responses to omission, tone, and expected omission with topographic distribution and its source localization by sLORETA. Table 1 shows the ERP amplitude and the results of the statistical analysis comparing the amplitude between the high and low predictability conditions. The results show that the oN1 amplitude of the omissions and the N1 amplitude of the tone were not significantly different between the high and low predictability conditions.

Source estimation of sLORETA showed significant activation mainly in the right temporal cortex (i.e., middle temporal gyrus: MTG, STG) and right frontal cortex (inferior frontal gyrus: IFG) in omission, tone, and expected omission when the activation of each predictability condition of the stimuli was compared to zero. However, the difference between the high and low predictability conditions for tone, omission, and expected omission was not statistically significant. Activation in the MTG and posterior STG was significantly different between the oN1 for omission and the N1 for tone.

3.3. SVM decoding from ERP response

The timbres were decoded using the OSP waveforms in the high and low predictability conditions separately. Decoding accuracy was significantly above chance level ($p = .25$) only in the high predictability condition, at around 64–86 msec, $t_{sum} = 64.51$, $p = .023$, but not in the low predictability condition, the largest cluster being detected was around 154–171 msec, $t_{sum} = 48.26$, $p = .082$. However, decoding accuracy was not significantly different between the two conditions, despite the largest cluster being detected around 72–82 msec, $t_{sum} = 64.51$, $p = .392$. The mean accuracy in this cluster was $M = 25.91\%$ ($SD = 1.61$) for the high predictability condition and $M = 24.88\%$ ($SD = 1.67$) for the low predictability condition, suggesting that not all participants' data showed successful decoding, even in the high predictability condition. These results suggest a qualitative difference in the prediction of timbre content between the high and low predictability conditions. Specifically, decodable content prediction was present in the ERP of the high predictability condition, but not in the low predictability condition.

Table 1 – Comparison of the mean ERP amplitude between high and low conditions.

N = 35	High M (SD)	Low M (SD)	t df = 34	p	dz	BF ₁₀
oN1 (omission)	−1.93 (1.13)	−2.07 (1.19)	.76	.453	.13	.24
N1 (tone)	−4.15 (2.22)	−4.34 (2.24)	.96	.343	.16	.28
ERP (expected omission)	−1.67 (1.14)	−1.61 (1.13)	−.38	.706	−.06	.19

Note. Results of paired t-test and Bayesian paired t-test are shown.

The same timbre decoding was performed using the ERP waveforms elicited by the tone in the high and low predictability conditions separately. Decoding accuracy was significantly above chance level ($p = .25$) in both the high, 68–300 msec cluster, $t_{sum} = 1513.53$, $p < .001$, and low, 60–300 msec cluster, $t_{sum} = 1656.88$, $p < .001$, predictability conditions. Decoding accuracy was not significantly different between the two conditions (high: $M = 27.42\%$, $SD = 2.35$; low: $M = 28.63\%$, $SD = 2.74$), the largest cluster was around 118–120 msec, $t_{sum} = -6.80$, $p = .960$.

3.4. HEP difference in stimuli and timbre predictability

Fig. 3 shows the HEP and accuracy waveforms from decoding the HEPs of the three stimuli. The positive HEP waveforms (348–388 msec interval) of the three stimuli (i.e., tone, omission, and expected omission), calculated by averaging the high and low predictability conditions, are shown in the left panel of Fig. 3a. A one-way ANOVA revealed a significant main effect of stimulus type (tone, omission, and expected omission), $F(2, 60) = 18.74$, $p < .001$, $\eta_p^2 = .384$, Greenhouse-Geisser $\epsilon = .931$, $BF_{incl} = 37207.66$. The post-hoc t-tests revealed that the positive HEP was significantly larger during omission and expected omission than during tone, corrected $ps < .001$. The positive HEP was not statistically significant between omission and expected omission, corrected $p = .535$. SVM decoding was performed to classify the HEPs of the three stimuli. The peak of decoding accuracy was found around the positive HEP, and the cluster-based permutation test revealed that the 167–485 msec cluster was significantly above chance level, $t_{sum} = 1396.33$, $p < .001$. These results indicate that the HEPs of the three stimuli were dissociable.

The sLORETA results of the source estimation of the positive HEP are shown in Fig. 3b. The source was detected around the insula in both the omission and expected omission when compared to zero, but not in the tone. When the activation was compared between omission and tone, and between expected omission and tone, the activation of the insula was significantly stronger for omissions than for tone. However, the significant difference between omission and expected omission was not observed.

Fig. 3c shows the comparison of the HEP response between the two conditions during each stimulus type. To examine the difference in the HEP amplitude between the high and low predictability conditions, the difference HEP amplitudes (high minus low) at the 77–97 msec interval was compared to zero using one-sample t-test. No significant differences were found for tone, $M = -.16 \mu V$ ($SD = 1.17$), $t(30) = -.76$, $p = .452$, $dz = -.14$, $BF_{10} = .25$, omission, $M = .56 \mu V$ ($SD = 1.73$), $t(30) = 1.82$, $p = .079$, $dz = .33$, $BF_{10} = .82$, or expected omission, $M = -.04 \mu V$ ($SD = 1.41$), $t(30) = -.15$, $p = .879$, $dz = -.03$, $BF_{10} = .19$. Predictability had no effect on HEP amplitude.

4. Discussion

The present study examined whether decoding accuracy changes according to timbre predictability and whether oN1 amplitude differs between contexts with different timbre predictability. The oN1 amplitude did not differ significantly



Fig. 3 – HEP Responses to Tones and Omissions with its Source Localization Estimated by sLORETA

Note. (a) Grand HEP waveforms of tone, omission, and expected omission when averaging the high and low predictability conditions. The gray area indicates the positive HEP interval (348–388 msec) defined as the peak ± 20 msec. The topographic distribution shows the mean ERP amplitude of the positive HEP interval. The right panel shows the waveform of the decoding accuracy of the HEPs during the three stimuli. (b) The estimated cortical sources of tone, omission, and expected omission with the comparison results of the stimuli. Only voxels that showed statistically significant activities exceeding the critical t-values (t_{crit}) according to nonparametric t-tests ($p < .05$) are colored. (c) The grand difference HEP waveforms to tone, omission, and expected omission. The topographic distribution shows the mean ERP amplitude at 77–97 msec, indicated by the gray area in the waveform panels. The right panel shows the topographic distribution of HEP in the gray area. For all panels, the HEP waveforms were calculated by averaging the frontal-central electrodes (Fz, FC1, FC2, and Cz). The light-colored bands on the ERP and accuracy waveforms indicate a 95% confidence interval.

between the high and low predictability conditions (not supporting H1). Although decoding accuracy was not quantitatively different between the high and low predictability conditions, decoding accuracy was above chance level only in the high predictability condition, reflecting a qualitative difference (partially supporting H2). These results suggest that the EEG during omission contains feature-specific predictive information only in the high predictability condition. The source of oN1 and N1 responses was estimated around the STG and MTG, suggesting the computation of prediction error in the auditory cortex. The HEP amplitude was significantly greater during tone omission (expected and unexpected omission) than during tone (supporting H3). The results of OSP and HEP are discussed in terms of predictive processing.

The ERP responses to the omission and the expected omission were similar. This is unexpected, since the OSPs should not be observed at the expected omission position because the participants knew that no tone would be presented. One possibility is that the ERP responses to both the omission and the expected omission contained a late ERP component elicited by the preceding tone. The preceding tone was physically the same for both omissions, and this may have led to similar late ERP components contaminating the ERPs observed during both. Nevertheless, the ERP amplitude around the oN1 time window and activation of the STG were significantly greater in the omission than in the expected omission. Prete et al. (2022) interpreted the negative component, calculated by subtracting the ERP of the expected

omission from the unexpected omission, as a reflection of prediction error. Similarly, in the present study, the enhanced ERP response around the oN1 time window for the omission compared to the expected omission is attributable to the prediction error elicited by the omission. Moreover, the successful decoding of the omission and the expected omission supports the difference between two omissions. Therefore, the difference in the ERP amplitude around the oN1 time window indicates the presence of prediction-related activity (i.e., OSP).

The oN1 amplitude was not different between the high and low predictability conditions. The present study strictly manipulated content predictability while controlling for the other possible predictabilities, in contrast to [Ishida et al. \(2024\)](#), where the modulation of oN1 amplitude may have been influenced by the overall predictability, such as the pitch alignment, harmony, and rhythm of the melodies, rather than solely by the content predictability of the target tone. Moreover, because the present study used an unfamiliar melody in both conditions, participants had to learn pitch and timbre patterns simultaneously to form predictions in an inattentive state. It may also be challenging for participants to learn timbre regularity with the current procedure, where all four contextual instruments were presented within a single block, even in the high predictability condition. Thus, one possible explanation for the current results is that subtle differences in predictability arise due to the complexity of the regularity and difficulties in learning, which may have prevented a clear distinction between the high and low predictability conditions.

The degree of difference in predictability may be an important factor in oN1 modulation. Previous studies that observed amplitude differences in oN1 ([Dercksen et al., 2020](#); [SanMiguel, Saupe, & Schröger, 2013](#)) compared content predictability between identical, where the same tone was always presented, and random, where 48 tones were presented randomly and the tone content was unpredictable. In contrast, the present study compared high and low predictability conditions, where predictions were able to form in both conditions, although the predictability was different. If prediction is present in both conditions and the difference in predictability is small (i.e., there is no clear distinction between predictable and unpredictable conditions), the difference in oN1 amplitude may not be observed. Supporting this possibility, in the visual domain, [Kimura and Takeda \(2018\)](#) showed that the amplitude of omission P3 was significantly larger in the one-stimulus condition, in which bars with identical slopes were always presented, than in the two-, four-, and eight-stimulus conditions, in which bars with two, four, or eight different slopes were presented, respectively, but did not differ between the two-, four-, and eight-stimulus conditions. Therefore, a clear difference in predictability would be required to observe the difference in OSP amplitude.

The neural source of oN1 was estimated to be around the temporal lobe, similar to the N1 response of tone. The activation of the STG and MTG during omission is consistent with previous studies that reported the source of oN1 in the STG and MTG ([Ishida & Nittono, 2024a](#); [SanMiguel Saupe et al., 2013](#); [Stekelenburg & Vroomen, 2015](#)). The source similarity of oN1 and N1 suggests that oN1 is a modality-specific

omission response, as reported in [Ishida and Nittono \(2024a\)](#). However, in contrast to tone, which dominantly activated the anterior temporal gyrus, omission dominantly activated the middle and posterior areas of the temporal gyri, such as MTG and STG. These results are similar to those of [Cho et al. \(2023\)](#), who reported the dominance of high-frequency EEG power in the posterior and middle STG for syllable omission using an electrocorticogram. The activation of the posterior STG for omission may be due to higher salience, considering the results of [Downar et al. \(2002\)](#), who reported greater activation of the temporo-parietal junction for novel than for familiar stimuli in auditory, visual, and tactile modalities as discussed in [Cho et al. \(2023\)](#). Therefore, oN1 and N1 do not fully share neural pathways. However, due to the low spatial resolution of EEG, these sLORETA results should be replicated in an MEG or fMRI study to verify the current interpretation of the source localization.

Timbre decoding accuracy exceeded chance level only in the high predictability condition, although accuracy was not quantitatively different between the high and low predictability conditions. This result suggests that the prediction of musical tone was feature specific not only in the pitch domain but also in the timbre domain when the timbre was highly predictable. In the high predictability condition, the cluster latency (64–86 msec), which was significantly above chance level, was similar to [Ishida et al. \(2024\)](#), who found a significant difference in decoding accuracy between the high predictability (familiar) and low predictability (unfamiliar) conditions of around 58–83 msec. The qualitative difference in decoding accuracy between the two conditions suggests that the predictive information was contained only in the high predictability condition. The multivariate decoding method could be a powerful tool to detect the subtle difference in predictability that was not observed in the rough amplitude indices. Similar to previous studies that used omission responses to decode predicted tone features ([Demarchi et al., 2019](#); [Hauswald et al., 2024](#); [Ishida et al., 2024](#)), the present study successfully decoded feature-specific predictions during music perception and extended the pitch domain to the timbre domain. The timbres were selected to ensure participants could clearly identify the instruments. Even though it is difficult to determine whether the current decoding results reflect predictions related to specific timbre features like the spectral centroid due to the instrument selection, the current results still support that the decoding was due to the broad difference of the musical instruments.

It should be noted that the decoding accuracy in the current high predictability condition (peak accuracy was 26.14%) was low compared to the familiar melody condition (peak accuracy was 31.08%) in [Ishida et al. \(2024\)](#). This may be due to the reduction of predictive information in the present study compared to [Ishida et al. \(2024\)](#) because the possible predictive information that would be used to infer the target tone was strictly controlled to manipulate content predictability only. Alternatively, a slower tempo would be required to sufficiently encode timbre regularity compared to pitch, because the processing of the timbre dimension relies on more complex spectrotemporal information ([Town & Bizley, 2013](#); [Wei et al., 2022](#)). It is known that the theoretical chance level, which is calculated by dividing 100% by the number of classes, can be

exceeded by chance, especially when the sample size is small (Combrisson & Jerbi, 2015). Although the current decoding accuracy only slightly exceeded the theoretical chance level, the latency of the significant cluster was similar to the cluster that significantly exceeded the chance level in a previous decoding study of EEG signals during omission (Ishida et al., 2024). This replicability suggests that the current decoding result is not entirely due to a false positive.

The peak latency of decoding accuracy differed between the omission and the tone. The decoding accuracy waveform shows the two peaks around the peaks of N1 and P2, which are the auditory evoked potentials (Eggermont & Ponton, 2002; Picton et al., 1974). However, the peak accuracy of the high predictability condition was earlier than the N1 and oN1 peaks when decoding was performed using omission. Similarly, Ishida et al. (2024) reported peak accuracy of decoding before the N1 and oN1 peaks. Previous studies (Bendixen et al., 2009; Ishida et al., 2024; SanMiguel Widmann et al., 2013) have discussed that the sensory template for the expected sound is generated until it is interrupted by omission detection in the early latency range (approximately 50–100 msec after omission). The scalp-recorded ERP is thought to largely reflect the superficial pyramidal cell (Jackson & Bolger, 2014), whose activation reflects the forward transmission of prediction errors (Friston & Kiebel, 2009), although top-down prediction signals may partially contribute to the scalp-recorded ERP (Schröger et al., 2015). For this reason, it is difficult to dissociate prediction signals from prediction error signals. The oN1 response is a prediction error even if it mirrors a prediction (SanMiguel, Saupe, & Schröger, 2013; Schröger et al., 2015). Therefore, in the case of an omission, the prediction signal can be reflected in the EEG because no neural activations caused by sensory input contaminate until the omission deviant is detected and the prediction signal is superimposed by the prediction error signal as recorded in oN1. This may lead to successful decoding before the oN1 peak. The current results further support that the early part of the OSP contains the sensory template for the expected tone content.

The positive HEP was larger during omission compared to tone. The neural generators of the positive HEP of the omission and the expected omission were estimated in the right insula. Moreover, activation of the insula was significantly stronger during the omission and expected omission presentations compared to the tone presentation. HEP amplitude is larger when participants attend to the interoceptive signal than when they attend to the exteroceptive signal (Petzschner et al., 2019), even during tone omission (Banellis and Cruse, 2020). Thus, the most parsimonious explanation is that the enhancement of HEP amplitude during omissions was due to switching between interoceptive and exteroceptive processing and focused attention on interoception. However, due to the low spatial resolution of sLORETA, further research using imaging techniques is necessary to definitively conclude these interpretations, especially when estimating the source to the insula, which is located within the lateral sulcus. Another possibility is that the positive HEP was due to increased arousal to the unexpected omission. The positive HEP

becomes more pronounced around 200–400 msec when stimuli with high emotional valence and arousal occur (Fuseda and Katayama, 2021; Luft & Bhattacharya, 2015). In the present study, the expected omission, in addition to the omission, may have caused some surprise. This is because the expected omission was inserted in the middle of a melody where quarter notes were constantly presented, specifically in a non-attending situation. Therefore, the informatic surprise (i.e., the Shannon surprise) of the omission may have increased arousal, and positive HEP was more pronounced during the omission than during the tone presentation.

Decoding of the three HEPs during tone, omission, and expected omission exceeded chance level at approximately 167–485 msec. This suggests that HEP differs not only between the tone and the omissions but also between the omission and the expected omission, although this difference was subtle and did not differ significantly in the HEP waveform. The difference in HEP between omissions may reflect the difference in prediction. HEP amplitude is modulated by the prediction of external stimuli, such as tones synchronized to the heartbeat (Banellis and Cruse, 2020) and expected facial expressions (Gentsch et al., 2019). Moreover, Ono et al. (2024) reported that the current source density (CSD) of HEP was more positive for the harmonic chord than the non-harmonic chord in the frontal region around 300–500 msec. In their study, the CSD-transformed HEP amplitude was positively correlated with subjective ratings of uncertainty, suggesting that HEP reflects the predictability of the chord. The omission and the expected omission in the current study differed only in content prediction, and the latency at which decoding accuracy exceeded chance level was similar to the HEP latency that was statistically significant in Ono et al. (2024). Therefore, the HEP difference between omissions may also reflect the difference between the presence and absence of content prediction of the external tone. However, the amplitude difference was not observed between the high and low predictability conditions for all three stimuli. The present study exploratorily examined HEP during omission and timbre predictability. In future research, with a sufficient number of trials and sample size, it will be possible to test whether the predictability of exteroceptive signals influences interoceptive signal processing.

In summary, the present study showed that the decoding accuracy of the timbres was significantly above chance only in the high predictability condition but not in the low predictability condition, reflecting that the EEG during omission contained predictive information about the specific timbre. However, the oN1 amplitude did not differ between the high and low predictability conditions. These results suggest that the EEG during omission may contain predictive information about the specific timbre that is not reflected in the traditional ERP amplitude. The present study provides further evidence that the brain's prediction is not only feature specific in the pitch domain but also in the timbre domain. Moreover, the enhancement of the positive HEP response during omission and expected omission compared to tone may reflect the difference in interoceptive or internal predictive processing.

The effect of the predictability of the external signal on the interoceptive processing reflected in the HEP is a topic for future research.

CRediT authorship contribution statement

Kai Ishida: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Tomomi Ishida:** Writing – review & editing, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hiroshi Nittono:** Writing – review & editing, Validation, Supervision, Software, Resources, Methodology, Investigation, Formal analysis, Conceptualization.

Data accessibility

The sound materials used and datasets analyzed for the present paper are available at <https://osf.io/awx7e/>.

Declaration of competing interest

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

Acknowledgement

This work was supported by JSPS KAKENHI JP22KJ2199.

Scientific transparency statement

DATA: All raw and processed data supporting this research are publicly available: <https://osf.io/awx7e/files>.

CODE: All analysis code supporting this research is publicly available: <https://osf.io/awx7e/files>.

MATERIALS: All study materials supporting this research are publicly available: <https://osf.io/awx7e/files>.

DESIGN: This article reports, for all studies, how the author(s) determined all sample sizes, all data exclusions, all data inclusion and exclusion criteria, and whether inclusion and exclusion criteria were established prior to data analysis.

PRE-REGISTRATION: No part of the study procedures was pre-registered in a time-stamped, institutional registry prior to the research being conducted. No part of the analysis plans was pre-registered in a time-stamped, institutional registry prior to the research being conducted.

For full details, see the Scientific Transparency Report in the supplementary data to the online version of this article.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.cortex.2025.08.013>.

REFERENCES

- Bae, G. Y., & Luck, S. J. (2019). Decoding motion direction using the topography of sustained ERPs and alpha oscillations. *Neuroimage*, 184, 242–255. <https://doi.org/10.1016/j.neuroimage.2018.09.029>
- Banellis, L., & Cruse, D. (2020). Skipping a beat: Heartbeat-evoked potentials reflect predictions during interoceptive-exteroceptive integration. *Cerebral Cortex Communications*, 1(1). <https://doi.org/10.1093/texcom/tgaa060>
- Bendixen, A., SanMiguel, I., & Schröger, E. (2012). Early electrophysiological indicators for predictive processing in audition: A review. *International Journal of Psychophysiology*, 83(2), 120–131. <https://doi.org/10.1016/j.ijpsycho.2011.08.003>
- Bendixen, A., Scharinger, M., Strauß, A., & Obleser, J. (2014). Prediction in the service of comprehension: Modulated early brain responses to omitted speech segments. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior*, 53(1), 9–26. <https://doi.org/10.1016/j.cortex.2014.01.001>
- Bendixen, A., Schröger, E., & Winkler, I. (2009). I heard that coming: Event-related potential evidence for stimulus-driven prediction in the auditory system. *Journal of Neuroscience*, 29(26), 8447–8451. <https://doi.org/10.1523/JNEUROSCI.1493-09.2009>
- Bullock, T. H., Karamiirsel, S., Achimowicz, J. Z., Mcclune, M. C., & Başar-Eroglu, C. (1994). Dynamic properties of human visual evoked and omitted stimulus potentials. *Electroencephalography and Clinical Neurophysiology*, 91(1), 42–53. [https://doi.org/10.1016/0013-4694\(94\)90017-5](https://doi.org/10.1016/0013-4694(94)90017-5)
- Busse, L., & Woldorff, M. G. (2003). The ERP omitted stimulus response to “no-stim” events and its implications for fast-rate event-related fMRI designs. *Neuroimage*, 18(4), 856–864. [https://doi.org/10.1016/S1053-8119\(03\)00012-0](https://doi.org/10.1016/S1053-8119(03)00012-0)
- Caclin, A., Brattico, E., Tervaniemi, M., Näätänen, R., Morlet, D., Ne Giard, M.-H., & Mcadams, S. (2006). Separate neural processing of timbre dimensions in auditory sensory memory. *Journal of Cognitive Neuroscience*, 18(12), 1959–1972. <https://doi.org/10.1162/jocn.2006.18.12.1959>
- Cho, H., Fonken, Y. M., Adamek, M., Jimenez, R., Lin, J. J., Schalk, G., Knight, R. T., & Brunner, P. (2023). Unexpected sound omissions are signaled in human posterior superior temporal gyrus: An intracranial study. *Cerebral Cortex*, 33(14), 8837–8848. <https://doi.org/10.1093/cercor/bhad155>
- Chouiter, L., Tzovara, A., Dieguez, S., Annoni, J. M., Magezi, D., de Lucia, M., & Spierer, L. (2015). Experience-based auditory predictions modulate brain activity to silence as do real sounds. *Journal of Cognitive Neuroscience*, 27(10), 1968–1980. https://doi.org/10.1162/jocn_a_00835
- Coll, M. P., Hobson, H., Bird, G., & Murphy, J. (2021). Systematic review and meta-analysis of the relationship between the heartbeat-evoked potential and interoception. *Neuroscience and Biobehavioral Reviews*, 122, 190–200. <https://doi.org/10.1016/j.neubiorev.2020.12.012>
- Combrisson, E., & Jerbi, K. (2015). Exceeding chance level by chance: The caveat of theoretical chance levels in brain signal classification and statistical assessment of decoding accuracy. *Journal of Neuroscience Methods*, 250, 126–136. <https://doi.org/10.1016/j.jneumeth.2015.01.010>
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- Demarchi, G., Sanchez, G., & Weisz, N. (2019). Automatic and feature-specific prediction-related neural activity in the human auditory system. *Nature Communications*, 10(1). <https://doi.org/10.1038/s41467-019-11440-1>

- Dercksen, T. T., Widmann, A., Noesselt, T., & Wetzels, N. (2024). Somatosensory omissions reveal action-related predictive processing. *Human Brain Mapping*, 45(4), Article e26550. <https://doi.org/10.1002/hbm.26550>
- Dercksen, T. T., Widmann, A., Schröger, E., & Wetzels, N. (2020). Omission related brain responses reflect specific and unspecific action-effect couplings. *Neuroimage*, 215, Article 116840. <https://doi.org/10.1016/j.neuroimage.2020.116840>
- Downar, J., Crawley, A. P., Mikulis, D. J., & Davis, K. D. (2002). A cortical network sensitive to stimulus salience in a neutral behavioral context across multiple sensory modalities. *Journal of Neurophysiology*, 87(1), 615–620. <https://doi.org/10.1152/jn.00636.2001>
- Eggermont, J. J., & Ponton, C. W. (2002). The neurophysiology of auditory perception: From single units to evoked potentials. *Audiology and Neurotology*, 7, 71–99. <https://doi.org/10.1159/000057656>
- Faul, F., E. E. L. A. G., & B. 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. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- Friston, K. (2010a). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <https://doi.org/10.1038/nrn2787>
- Friston, K. (2010b). 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>
- Fuseda, K., & Katayama, J. (2021). A new technique to measure the level of interest using heartbeat-evoked brain potential. *Journal of Psychophysiology*, 35(1), 15–22. <https://doi.org/10.1027/0269-8803/a000257>
- Gentsch, A., Sel, A., Marshall, A. C., & Schütz-Bosbach, S. (2019). Affective interoceptive inference: Evidence from heart-beat evoked brain potentials. *Human Brain Mapping*, 40(1), 20–33. <https://doi.org/10.1002/hbm.24352>
- Hauswald, A., Benz, K. R., Hartmann, T., Demarchi, G., & Weisz, N. (2024). Carrier-frequency specific omission-related neural activity in ordered sound sequences is independent of omission-predictability. *European Journal of Neuroscience*, 60(1), 3812–3820. <https://doi.org/10.1111/ejn.16381>
- Hernández, O. H., & Hernández-Sánchez, K. M. (2017). Omitted stimulus potential depends on the sensory modality. *Acta Neurobiologiae Experimentalis*, 77(4), 297–304. <https://doi.org/10.21307/ane-2017-062>
- Horváth, J., Müller, D., Weise, A., & Schröger, E. (2010). Omission mismatch negativity builds up late. *Neuroreport*, 21(7), 537–541. <https://doi.org/10.1097/WNR.0b013e3283398094>
- Ishida, K., Ishida, T., & Nittono, H. (2024). Decoding predicted musical notes from omitted stimulus potentials. *Scientific Reports*, 14(1), Article 11164. <https://doi.org/10.1038/s41598-024-61989-1>
- Ishida, T., & Nittono, H. (2024a). Effects of sensory modality and task relevance on omitted stimulus potentials. *Experimental Brain Research*, 242(1), 47–57. <https://doi.org/10.1007/s00221-023-06726-2>
- Ishida, T., & Nittono, H. (2024b). Visual omitted stimulus potentials are not retinotopic. *Neuroscience Letters*, 830, Article 137777. <https://doi.org/10.1016/j.neulet.2024.137777>
- Ishida, K., & Nittono, H. (2024c). Relationship between schematic and dynamic expectations of melodic patterns in music perception. *International Journal of Psychophysiology*, 196, 112292. <https://doi.org/10.1016/j.ijpsycho.2023.112292>
- Ishida, K., & Nittono, H. (2025). Active inference in music perception: Motor engagement to syncopation modulates rhythmic prediction error. *Psychophysiology*, 62(7). <https://doi.org/10.1111/psyp.70113>
- Jackson, A. F., & Bolger, D. J. (2014). The neurophysiological bases of EEG and EEG measurement: A review for the rest of us. *Psychophysiology*, 51(11), 1061–1071. <https://doi.org/10.1111/psyp.12283>
- Jongsma, M. L. A., Quiroga, R. Q., & Rijn, C. M. van (2004). Rhythmic training decreases latency-jitter of omission evoked potentials (OEPs) in humans. *Neuroscience Letters*, 355(3), 189–192. <https://doi.org/10.1016/j.neulet.2003.10.070>
- Kimura, M., & Takeda, Y. (2018). Omission P3 after voluntary action indexes the formation of action-driven prediction. *International Journal of Psychophysiology*, 124, 54–61. <https://doi.org/10.1016/j.ijpsycho.2017.12.006>
- 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>
- Korka, B., Schröger, E., & Widmann, A. (2020). What exactly is missing here? The sensory processing of unpredictable omissions is modulated by the specificity of expected action-effects. *European Journal of Neuroscience*, 52(12), 4667–4683. <https://doi.org/10.1111/ejn.14899>
- Li, C. W., Guo, F. Y., & Tsai, C. G. (2021). Predictive processing, cognitive control, and tonality stability of music: An fMRI study of chromatic harmony. *Brain and Cognition*, 151, Article 105751. <https://doi.org/10.1016/j.bandc.2021.105751>
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213. <https://doi.org/10.3389/fnhum.2014.00213>
- Luft, A. D. B., & Bhattacharya, J. (2015). Aroused with heart: Modulation of heartbeat evoked potential by arousal induction and its oscillatory correlates. *Scientific Reports*, 5(1), Article 15717. <https://doi.org/10.1038/srep15717>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Nittono, H. (2005). Missing-stimulus potentials associated with a disruption of human-computer interaction. *Psychologia*, 48(2), 93–101. <https://doi.org/10.2117/psysoc.2005.93>
- Okubo, M., Suzuki, H., & Nicholls, M. E. R. (2014). A Japanese version of the FLANDERS handedness questionnaire. *Shinrigaku Kenkyu: the Japanese Journal of Psychology*, 85, 474–481. <https://doi.org/10.4992/jjpsy.85.13235>
- Ono, K., Mizuuchi, R., Yamamoto, K., Sasaoka, T., & Ymawaki, S. (2024). Exploring the neural underpinnings of chord prediction uncertainty: An electroencephalography (EEG) study. *Scientific Reports*, 14(1), 4586. <https://doi.org/10.1038/s41598-024-55366-1>
- Pascual-Marqui, R. D. (2002). Standardized low-resolution brain electromagnetic tomography (sLORETA): Technical details. *Methods Find Exp Clin Pharmacol*, 24(Suppl D), 5–12.
- Perakakis, P. (2019). HEPLAB: A Matlab graphical interface for the preprocessing of the heartbeat-evoked potential. Zenodo. <https://doi.org/10.5281/zenodo.2649943>. Version v1.0.0.
- Petzschner, F. H., Weber, L. A., Wellstein, K. v., Paolini, G., Do, C. T., & Stephan, K. E. (2019). Focus of attention modulates the heartbeat evoked potential. *Neuroimage*, 186, 595–606. <https://doi.org/10.1016/j.neuroimage.2018.11.037>
- Picton, T. W., Hillyard, S. A., Krausz, H., & Galambos, R. (1974). Human auditory evoked potentials. I: Evaluation of components. *Electroencephalography and Clinical Neurophysiology*, 36, 179–190. [https://doi.org/10.1016/0013-4694\(74\)90155-2](https://doi.org/10.1016/0013-4694(74)90155-2)

- Prete, D. A., Heikoop, D., McGillivray, J. E., Reilly, J. P., & Trainor, L. J. (2022). The sound of silence: Predictive error responses to unexpected sound omission in adults. *European Journal of Neuroscience*, 55(8), 1972–1985. <https://doi.org/10.1111/ejn.15660>
- Quiroga-Martinez, D. R., Hansen, N. C., Højlund, A., Pearce, M. T., Brattico, E., & Vuust, P. (2019). Reduced prediction error responses in high-as compared to low-uncertainty musical contexts. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior*, 120, 181–200. <https://doi.org/10.1016/j.cortex.2019.06.010>
- 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>
- Sadakata, M., Yamaguchi, Y., Ohsawa, C., Matsubara, M., Terasawa, H., von Schnehen, A., Müllensiefen, D., & Sekiyama, K. (2023). The Japanese translation of the Gold-MSI: Adaptation and validation of the self-report questionnaire of musical sophistication. *Musicae Scientiae*, 27(3), 798–810. <https://doi.org/10.1177/10298649221110089>
- SanMiguel, I., Saupe, K., & Schröger, E. (2013a). I know what is missing here: Electrophysiological prediction error signals elicited by omissions of predicted “what” but not “when.”. *Frontiers in Human Neuroscience*, 7, 407. <https://doi.org/10.3389/fnhum.2013.00407>
- SanMiguel, I., Widmann, A., Bendixen, A., Trujillo-Barreto, N., & Schröger, E. (2013b). Hearing silences: Human auditory processing relies on preactivation of sound-specific brain activity patterns. *Journal of Neuroscience*, 33(20), 8633–8639. <https://doi.org/10.1523/JNEUROSCI.5821-12.2013>
- Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin & Review*, 25(1), 128–142. <https://doi.org/10.3758/s13423-017-1230-y>
- Schröger, E., Marzecová, A., & Sanmiguel, I. (2015). Attention and prediction in human audition: A lesson from cognitive psychophysiology. *European Journal of Neuroscience*, 41(5), 641–664. <https://doi.org/10.1111/ejn.12816>
- Seol, J., Oh, M. A., Kim, J. S., Jin, S. H., Kim, S. il, & Chung, C. K. (2011). Discrimination of timbre in early auditory responses of the human brain. *Plos One*, 6(9), Article e24959. <https://doi.org/10.1371/journal.pone.0024959>
- Stange, L., Ossandón, J. P., & Röder, B. (2023). Crossmodal visual predictions elicit spatially specific early visual cortex activity but later than real visual stimuli. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 378(1886), Article 20220339. <https://doi.org/10.1098/rstb.2022.0339>
- Stekelenburg, J. J., & Vroomen, J. (2015). Predictive coding of visual-auditory and motor-auditory events: An electrophysiological study. *Brain Research*, 1626, 88–96. <https://doi.org/10.1016/j.brainres.2015.01.036>
- Tanaka, Y., Ito, Y., Terasawa, Y., & Umeda, S. (2023). Modulation of heartbeat-evoked potential and cardiac cycle effect by auditory stimuli. *Biological Psychology*, 182, Article 108637. <https://doi.org/10.1016/j.biopsycho.2023.108637>
- Tautvydaitė, D., & Burra, N. (2024). The timing of gaze direction perception: ERP decoding and task modulation. *Neuroimage*, 295, Article 120659. <https://doi.org/10.1016/j.neuroimage.2024.120659>
- Town, S. M., & Bizley, J. K. (2013). Neural and behavioral investigations into timbre perception. *Frontiers in Systems Neuroscience*, 7, 88. <https://doi.org/10.3389/fnsys.2013.00088>
- Tužnik, P., Augustynowicz, P., & Francuz, P. (2018). Electrophysiological correlates of timbre imagery and perception. *International Journal of Psychophysiology*, 129, 9–17. <https://doi.org/10.1016/j.ijpsycho.2018.05.004>
- van Laarhoven, T., Stekelenburg, J. J., Eussen, M. L. J. M., & Vroomen, J. (2020). Atypical visual-auditory predictive coding in autism spectrum disorder: Electrophysiological evidence from stimulus omissions. *Autism: the International Journal of Research and Practice*, 24(7), 1849–1859. <https://doi.org/10.1177/1362361320926061>
- van Laarhoven, T., Stekelenburg, J. J., & Vroomen, J. (2017). Temporal and identity prediction in visual-auditory events: Electrophysiological evidence from stimulus omissions. *Brain Research*, 1661, 79–87. <https://doi.org/10.1016/j.brainres.2017.02.014>
- Vuust, P., Dietz, M. J., Witek, M., & Kringelbach, M. L. (2018). Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Sciences*, 1423(1), 19–29. <https://doi.org/10.1111/nyas.13622>
- 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>
- Vuust, P., Ostergaard, L., Pallesen, K. J., Bailey, C., & Roepstorff, A. (2009). Predictive coding of music - Brain responses to rhythmic incongruity. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior*, 45(1), 80–92. <https://doi.org/10.1016/j.cortex.2008.05.014>
- Vuust, P., & Witek, M. A. G. (2014). Rhythmic complexity and predictive coding: A novel approach to modeling rhythm and meter perception in music. *Frontiers in Psychology*, 5, 1111. <https://doi.org/10.3389/fpsyg.2014.01111>
- Wei, Y., Gan, L., & Huang, X. (2022). A review of research on the neurocognition for timbre perception. *Frontiers in Psychology*, 13, Article 869475. <https://doi.org/10.3389/fpsyg.2022.869475>
- Zanto, T. P., Snyder, J. S., & Large, E. W. (2006). Neural correlates of rhythmic expectancy. *Advances in Cognitive Psychology*, 2(2), 221–231. <https://doi.org/10.2478/v10053-008-0057-5>