"Hearing As": Prior Knowledge of Syntactic Structure Affects ERP Components for Musical Expectation

Andrew Goldman (angoldm@iu.edu)

Music Theory Department & Cognitive Science Program, Indiana University

200 S. Eagleson Ave., Bloomington, IN 47405 USA

Nazbanou Nozari (bnozari@iu.edu)

Department of Psychological and Brain Sciences & Cognitive Science Program, Indiana University

Yeoeun Lim (yeolim@iu.edu),

Cognitive Science Program, Indiana University

Megan Kibler (mkibler@iu.edu)

Cognitive Science Program, Indiana University

Corresponding author:

Andrew Goldman (angoldm@iu.edu; 200 S. Eagleson Ave., Bloomington, IN 47405 USA)

Abstract

Harmonic expectation is an important mediator of musical experience. EEG research has identified ERP components associated with expectation, including the early (right) anterior negativity (E(R)AN), which is theorized to index harmonic surprisal with reference to long-term memory of the statistical structure of music. However, the role of top-down influences on harmonic predictions remains under-explored. One specific influence concerns how a given harmony can be interpreted in different ways, depending on its syntactic role in a musical context. We present data from a novel paradigm that cues listeners to the syntactic structure of the stimuli (but not whether they contain improbable events). Our main result revealed larger E(R)AN amplitudes for improbable chords when listeners knew that additional context would follow a surprising harmony; P3a and P600 amplitudes were also larger in such cases. Using the theoretical framework of predictive coding, we propose that in such cases, listeners assign higher precision to their predictions, leading to larger prediction errors as indexed by the E(R)AN, P3a, and P600 ERP components, and that prior context alone does not fully explain how unpredictable events are processed. Musical surprisal arises from a dynamic interplay between bottom-up cues and a listener's top-down anticipation within specific syntactic contexts.

Keywords: EEG, Event related potentials, Language: Syntax and Morphology, Perception: Auditory processing

“Hearing As”: Knowledge of Syntactic Structure Affects ERP Components for Musical Expectation

Music psychologists and music theorists have identified harmonic expectation as an important mediator of musical experience (Huron, 2008; Meyer, 1956). Sequences of musical harmonies set up expectations, and composers fulfill or subvert those expectations. In cognitive science, expectation has been investigated in terms of musical syntax, modeling the grammar of music (i.e., rules for ordering harmonies, melodic notes, or other musical elements) with formal principles (Lerdahl & Jackendoff, 1983) and computational models (Temperley, 2004). Formal syntactic models allow music theorists to explain which type of musical events ought to happen in which order, for instance, in considering rules for ordering harmonies, analogous to how one may parse the syntactic structure of a linguistic utterance in terms of parts of speech (Rohrmeier & Pearce, 2018). In addition to such formal models, music syntax has been modeled probabilistically with computational models of statistical structure (Pearce, 2005; White & Quinn, 2018). Through statistical learning, listeners acquire sensitivity to which musical events are most likely to occur, given some context (Vuust et al., 2022).

Indeed, empirical work has demonstrated listeners’ sensitivity to improbable musical continuations, often in harmonic and melodic sequences (Pearce & Rohrmeier, 2018). Motivated in large part to compare neural processes of music and language (Koelsch, Schmidt, & Kansok, 2002; Patel, 2003), neuroscientists have leveraged the temporal resolution of EEG to discriminate between component processes of musical surprisal. Typically, experimental paradigms use a priming sequence to set up the expectation for a particular harmony, and then a target harmony either fulfills or subverts that expectation (Koelsch et al., 2000). For unexpected compared to expected chords, an early anterior negativity—sometimes found to be right-lateralized and labeled the early right anterior negativity (henceforth, E(R)AN)—is evoked, and is theorized to index the violation of expectation based on long-term memory of harmonic transition probabilities (as distinct from the sensory surprisal which evokes an auditory mismatch negativity, or MMN; Koelsch, 2009). Behavioral (Slevc, Rosenberg, & Patel, 2009) and EEG (Koelsch et al., 2005) research has shown that musical syntactic expectation interacts with linguistic syntax, suggesting shared underlying mechanisms. Further, neuroscientists have identified neuro-anatomical correlates of such cognitive processes using MEG and fMRI, tracing expectation violation processes to the inferior part of the pars opercularis (Maess et al., 2001) and the lower and upper parts of BA44 (Koelsch et al., 2002), offering additional corroborating evidence for shared neural networks with linguistic syntax (although, see Chen et al., 2023, who dispute the link).

EEG paradigms of harmonic surprisal have also evoked later ERP components (see Goldman et al., 2021 for a review) including the P3a (Steinbeis, Koelsch, & Sloboda, 2006; Vuust et al., 2009), N5 (Koelsch, 2011), and P600 (Patel et al., 1998; Featherstone et al., 2013), corresponding to later aspects of the cognitive appraisal of surprising chords. The P3a indexes attentional reorienting following the surprising chord, while the N5 and P600 amplitudes index the effort of re-integrating the surprising harmony into the context.

Several experiments have sought to isolate the effects of sensory surprisal that arises from spectral changes in musical stimuli from the cognitive surprisal that arises from a *type* of harmony occurring in the wrong place, for instance by controlling for the number of shared pitches between expected versus unexpected harmonies (Bigand et al., 2003; Koelsch et al., 2007) or by examining the effects of sensory dissonance as distinct from harmonically unexpected chords (Regnault, Bigand, & Besson, 2001). Evidence has generally converged on the E(R)AN distinctly indexing cognitive surprisal, although in a paradigm that used stimuli continuously varying in surprisal (as opposed to the more commonly used categorical differences), no association between the computationally-modeled surprisal and E(R)AN amplitude was found (Goldman et al., 2021).

Various other parameters have also been investigated. Longer priming contexts can increase the E(R)AN amplitude (Leino et al., 2007). The amplitude is larger in adult musicians compared to musical novices (Koelsch, 2002). Work by Loui et al. (2009) showed that listeners can even acquire knowledge of such statistical dependencies—indexed by the E(R)AN component—with an unfamiliar musical system introduced in the experimental session itself, showing it is possible for syntax violations to be explained by statistical learning over short time spans.

Top-down effects and "hearing-as"

Some studies have identified top-down influences on harmonic processing. Loui et al. (2005) showed that the amplitude of the E(R)AN increases when listeners pay attention as compared to engaging in a distractor task and ignoring the auditory stimuli (although, Lee, Jung, & Loui, 2019 subsequently found no effect of attention on E(R)AN amplitudes when examined in a different paradigm). Subsequent studies continued to observe other types of top-down effects. For instance, a listener's veridical knowledge of whether an upcoming stimulus will contain a surprising chord decreases the latency of the E(R)AN but not its amplitude (Guo & Koelsch, 2016), suggesting that veridical expectations do not override schematic ones previously acquired through long-term exposure to syntactic regularities. Vuvan & Hughes (2019) found behavioral evidence—in the form of subjective ratings for how well target events fit a priming sequence—that listeners have different expectations for harmonic continuations when listening to different genres of music; specifically, harmonic expectations for classical music are different from that of rock music.

The attention-mediated effects and the genre-specific effects provide foundational evidence that top-down processing modulates harmonic expectation in some way, although further questions remain about specific top-down processes that modulate harmonic expectation. For instance, genre-based top-down processing may indeed change a listener's overall probabilistic expectations by making them conditioned on style: a given continuation is common in one genre but uncommon in another. In this case, a listener expects different continuations following from the same harmonic event "X" depending on which genre they know they are hearing. That is, if "Y" is another harmonic event, then for the particular continuation $X \rightarrow Y$:

$$p(X \rightarrow Y) \neq p(X \rightarrow Y \mid genre)$$

for at least some $X \rightarrow Y$. Further as Vuvan & Hughes demonstrated, such probabilities may change depending on different genres, that is,

$$p(X \rightarrow Y \mid genre_{rock}) \neq p(X \rightarrow Y \mid genre_{classical})$$

for at least some $X \rightarrow Y$, and listeners are sensitive to such differences (as measured by their subjective ratings of how well harmonies fit). In this case, listeners may have the ability to condition their harmonic predictions on genre, but may still be evaluating particular harmonic transitions in essentially the same manner as if genre did not change the probabilities: by referencing long term memory of the statistics of harmonic transitions. That is, probabilistic models can still explain musical expectations, but become more accurate when accounting for the conditioning effect of genre.

Other types of top-down influence are possible, though, besides general conditional influences of factors like genre, or effects of attended versus unattended stimuli. Listeners might be able to modulate their expectations in more local, specific ways as well, based on how they *interpret* certain target harmonies. To make this intuition more explicit and precise, consider the phenomenon of "hearing-as" discussed by music theorists, adapted from Wittgenstein's discussion of "seeing-as" (Dubiel, 2017; Guck, 2017; Lewin, 1986). Guck writes that "'[h]earing as' melds perception with thought" (p. 257). Guck further explains the analogy to Wittgenstein's famous duck-rabbit (a perceptually bistable gestalt) in which a viewer can see it as a duck or see it as a rabbit, but not both. Merely seeing the image is different than seeing it *as* a rabbit; by analogy, musical sounds can be heard (their mere sonority) and also *heard-as*, in which a sonority takes on an interpretation. In music, such an interpretation is often syntactic. For instance, a listener can hear the notes C-E-G, which form a major chord (analogous to merely seeing the duck/rabbit image); in context, that sonority could be furthermore *heard as* a *dominant function* chord (implying a particular musical continuation to a *tonic function* chord, usually F-A-C, and analogous to seeing the image as a rabbit). Importantly, the same sonority can often have different interpretations (the sonority C-E-G can take other functions besides a dominant one, such as a *predominant function*), and thus can be heard as different types of sonic objects. For instance, listeners often disagree about whether particular sonorities constitute a particular type of *cadence*, musical events which demarcate phrase and/or formal section endings (Burstein, 2014). Thus, particular harmonies can have different syntactic roles, for instance, as cadential harmonies that conclude a progression, or as harmonies in the middle of a progression that continue on.

This particular account of top-down processing in music perception raises a distinct and specific question in how listeners make harmonic predictions, inherited from music theoretical work. Suppose a listener knows the syntactic position of a target harmony (e.g., whether it is progression-final versus progression-medial), but not whether that harmony will be an expected or an unexpected one. Further, suppose those two conditions are preceded by equivalent priming contexts, so the preceding context does not discriminate between the two conditions, and gives the same information about which harmony will come next. If a listener knows that more context will come *following* the target harmony, does it change how they process the target chord if it turns out to be an unexpected one?

Accordingly, in the present study, we manipulate how listeners interpret the syntactic role of target harmonies, and investigate how differences in interpretation affect the ERP components associated with processing harmonically unexpected events. While it can be difficult to reliably manipulate how a listener interprets a particular harmony, it is not impossible, and we have endeavored to accomplish it here. One way to do so is to tell listeners the number of chords in an upcoming progression. By analogy in language, suppose you are a told a sentence will be six words long. Then, the sentence "she_1 prefers_2 her_3 lattes_4 with_5 **whole**$_6$" may induce some surprise upon hearing the adjective "whole_6" because you may expect a noun (like sugar_6) instead of an adjective. Alternatively, if you know the sentence will be seven words long, the word "whole_6" may be processed differently because it is no longer heard as an agrammatical placement of an adjective, given that a noun is expected to follow, like "milk_7."

In music, too, a target chord following an initial priming sequence can be *heard as* the ending chord, or *heard as* a chord with a subsequent continuation. How unexpected that target harmony is may depend on its syntactic role; if additional harmonies follow the unexpected chord, listeners may adjust their evaluation of the target given knowledge that subsequent context may influence its interpretation. This difference may arise because an unexpected target chord may go on to *resolve* (i.e., proceed in an expected manner).

Typical harmonic syntax (in Western classical music) will go from a tonic type chord (T) to a predominant type chord (PD) to a dominant type chord (D) back to a tonic chord; a common instance is $T_1 \rightarrow T_2 \rightarrow PD_3 \rightarrow D_4 \rightarrow T_5$. By replacing the final T_5 with a PD_5 chord (remembering that PD chords normally *precede* D chords instead of follow them), the syntax is violated: $T_1 \rightarrow T_2 \rightarrow PD_3 \rightarrow D_4 \rightarrow \underline{\mathbf{PD_5}}$. But, by adding additional context *following* this surprising PD_5 such as a subsequent $D_6 \rightarrow T_7$ (which is the typical resolution of a PD chord), that surprising PD_5 is functioning typically again according to its continuation: $T_1 \rightarrow T_2 \rightarrow PD_3 \rightarrow D_4 \rightarrow \underline{\mathbf{PD_5} \rightarrow D_5 \rightarrow T_6}$. If the sequence terminates at the PD_5 chord, that PD chord may be starkly violating syntactic norms, especially if a listener knows that it is the end of the progression; if it continues from it and eventually resolves by the time of the final T_7 chord, the PD_5 chord may still be more unexpected than a T_5 chord, but a listener may adjust their interpretation of it given that additional context will follow. In the 5-chord versus 7-chord progressions, the PD_5 chord has identical preceding context, but can be heard in different ways depending on the progression length, necessitating different syntactic roles for that chord.

In fact, music theorists have different ways to describe this particular compositional technique; one term is an *evaded cadence*. The expected resolution from $D_4 \rightarrow T_5$ is temporarily evaded in positions 4–5 (the T_5 is replaced with a PD_5), but goes on to resolve properly to T_7 at the end when the cadence is eventually reached. This method of building anticipation is well-known to composers and music analysts, allowing the music to subvert an expectation, but eventually provide the final (and often more dramatic) resolution (for a pedagogical treatment of the matter, see Caplin, 2013, pp. 123–165). In this way, indicating the progression length to listeners can be treated as a proxy for providing them with a particular syntactic interpretation of a target harmony. Accordingly, the conditions allow us to compare top-down influences on harmonic expectation processing based on prior knowledge of the syntactic structure of a stimulus.

We note that such a question is distinct from Guo and Koelsch's (2016) contribution noted above, which cued listeners *veridically* to whether or not there would be a surprising musical event in the upcoming stimulus. Here, instead of revealing whether or not a surprising event would be present, we only cue listeners to the syntactic structure of the upcoming stimulus (by telling them the progression length), with an aim to examine how a surprising event is processed when a listener integrates it into different syntactic structures.

If the E(R)AN is theorized to index harmonic surprisal itself, based on a listener's long-term knowledge of harmonic expectations (even if behavioral evidence shows they are style specific), then one might theorize it to be invariant to top-down processing that dynamically adjusts interpretations according to the music's syntactic structure. After all, the context preceding the target chord (chords 1–4) in both the 5-chord and 7-chord conditions is identical. General attentional effects may indeed increase the gain of the E(R)AN, and knowledge of an impending violation may reduce its latency, but it remains to be tested whether the interpretation of syntactic structure (without veridical knowledge of whether or not a violation will occur) can modulate the neural correlates of harmonic expectation.

We thus compare ERP responses to unexpected PD_5 chords (versus T_5 chords) when they occur in *short*, 5-chord-long progressions, versus when they occur in *long*, 7-chord-long progressions. Targets are always the fifth chord in the progression regardless of progression length. Thus, in the short condition, the target chord ends the progression; in the long condition, two additional chords *follow* the target chord (see Figure 1 in Methods). Crucially, for each trial, listeners are cued to the length of the progression—which effectively provides information about the syntactic structure— but not whether it will contain a harmonically surprising event (the fifth chord can either be an expected T_5 chord, or an unexpected PD_5 chord).

Some different possibilities arise. First, it is possible that the ERP differences between atypical and typical conditions are not modulated by length. This could manifest as similar ERP component responses (in difference waves comparing atypical with typical stimuli) regardless of progression length, and could reflect encapsulated processes that evaluate harmonic surprisal solely on the basis of information accumulated from prior context.

Second, it is possible that the ERP component amplitudes will differ according to the syntactic interpretation. Any difference between short and long conditions could be interpreted as top-down influence on harmonic expectation because the short and long conditions are acoustically and information-theoretically identical up to the end of the target chord (chord 5). If the response is larger for the short progression compared to the long progression, it suggests that (yet-unheard) future context in the long condition can be taken into account when evaluating the surprisal of a particular musical event, perhaps diminishing its surprisal due to veridical knowledge that it will go on to resolve properly. The short condition has no such opportunity to resolve the unexpected chord, and could thus evoke a greater neural index of surprisal (i.e., greater amplitude ERP components).

Finally, if the ERP responses are larger for the long progression compared to the short progression, it may reflect greater uncertainty in a listener's prediction of the target chord (in predictive coding terms, higher precision), given that longer progressions may sensibly have a PD_5 chord in the target position—even if it is unlikely given the preceding context—because it could predictably *continue*. In this last scenario, the target chord in the long progression is more salient and informative than the target in the short progression because a listener knows that two options are plausible (T_5, and PD_5—given that future context will resolve it properly), and once they hear it, they know which of the two plausible options they are actually hearing. In other words, in the long progression, the fifth chord resolves more uncertainty. By contrast, in the short conditions, an atypical PD_5 chord is simply “wrong” (or, at least, improbable), and means nothing.

It is possible that differences between our conditions could arise from general attentional demand characteristics. Here, it depends on what one means be attention. One sense of the term is more general, contrasting attention with simply ignoring. For instance, in the unattended condition from Loui et al. (2005), participants read a book and were instructed to ignore the auditory stimuli (p. 680). By analogy in our study, participants might ignore the target chord in the short condition, and attend to it in the long conditions because they know there will be subsequent auditory stimulation following the target chord. In a sense, the target chord is more task relevant in the long condition for making sense of the progression as a whole. While we do not believe this is a likely scenario (all conditions are supposed to be attended in our paradigm), we nevertheless assess the possibility by analyzing the N1 ERP component, which has been shown to be sensitive to enhancement when auditory stimuli are attended versus unattended (Hillyard et al., 1973; Lange & Schnuerch, 2014), although Baumann, Meyer, & Jäncke (2008) determined an enhanced N1 to arise from expertise-enhanced differences in perceptual encoding and not attention. In order to consider differences between conditions in sensory processing prior to the harmonic predictions indexed by the E(R)AN, we included an analysis of the auditory N1 ERP component.

However, we believe a more specific sense of the term attention is indeed relevant for our paradigm: attentional gain modulation (Auksztulewicz and Friston, 2015; Feldman & Friston, 2010). This term elides with the predictive coding account of precision: attentional processes can increase the gain of prediction errors. In other words, in the long conditions, the target chord may be afforded higher precision—turning up the attentional gain—because, again, the target chord is more salient and informative than in the short conditions (see Discussion).

Finally, while our paradigm is designed to examine how a listener's knowledge of the syntactic structure of a chord progression affects how they process harmonic surprisal, it also ends up comparing the effect of sequence-final targets versus sequence-medial targets. That is, differences between the short versus long condition might be attributed to the listener's knowledge of the syntactic structure as we intend, or rather due to so-called *wrap-up* effects. Wrap-up effects in linguistics are concerned with cognitive processing at the ends of sentences or clauses in which a listener or reader integrates the syntactic structure of the preceding words, although the term is non-specific and can refer to a variety of underlying processes that may all occur at the end of a sentence or clause (Stowe et al., 2018). This issue has also been addressed in the music cognition literature; Tillmann & Marmel (2013) provided behavioral evidence for perceptual facilitation of expected target harmonies regardless of whether they are sequence-final or sequence-medial. Leino et al. (2007) evoked the E(R)AN component for chords in both medial and final positions for 7-chord sequences (with larger amplitudes when the target chord was in position three or seven versus position five); they offer different possible explanations for the larger amplitude in the final position, including wrap-up effects (in their words, "effects of closure," p. 174) but also the possibility that additional preceding chords may have increased the amplitude due to a more firmly established tonal context ("better established tonality," p. 174). Note that in contrast to Leino et al.'s experiment, our paradigm controls for the amount of preceding context, and has isolated top-down effects based on knowledge of future and yet-unheard stimuli. For our purposes and in our paradigm, wrap-up effects, such as an "effect of closure," are still effects of prior syntactic knowledge. A listener knows whether a target chord is the end, or not, and accordingly may process surprising harmonies differently.

Methods

Participants

Twenty-six participants took part in the study. Of the 26 datasets recorded, five were excluded due to excessive artifacts in the EEG signal (mostly due to sweating), resulting in large numbers of epochs being rejected (range 59–108 epochs rejected out of 240 total, $M = 87.20$, $SD = 19.65$). All further reported data and statistical analyses consider only the 21 included datasets. The number of bad epochs rejected by eye from the included datasets ranged from 0–48 out of 240, $M = 15.43$, $SD = 11.81$.

Hence, 21 participants' data were included in the study (11 male, 10 female; Age $M = 22.76$ years, $SD = 6.65$ years). Participants all had substantial musical training, as assessed by the self-reported years of formal music lessons ($M = 12.19$, $SD = 5.13$) and the Goldsmiths Musical Sophistication Index (GMSI; Müllensiefen et al., 2014) Musical Training subscale, $M = 40.29$ (87th percentile), $SD = 4.54$, as well as the GMSI Perceptual Abilities subscale, $M = 53.67$ (68th percentile), $SD = 5.46$.

We conducted an *a priori* power analysis with G* Power 3 (version 3.1.9.6; Faul et al., 2007). Our main analysis used a 2x2 repeated measures ANOVA design (see Materials), and searched for an interaction between our within-subject factors. Our power analysis thus modeled a repeated measures ANOVA (within factors) with the parameters alpha = .05, power = 0.80, 1 group (i.e., one group of participants with no-between group factors), and 2 measurements (such that the numerator df = 1, as it is for our interaction effect). We set the effect size based on a direct specification of $\eta^2_p = 0.10$ (a small effect). With such parameters, a sample size of 20 is recommended.

Materials

Stimuli

Stimuli consisted of chord progressions presented aurally, organized according to a 2x2 design with the factors length (short vs. long) and typicality (typical vs. atypical; see Figure 1). Short progressions were five chords long while long progressions were seven chords long; typical progressions had a syntactically expected chord in position five ($T_1 \rightarrow T_2 \rightarrow PD_3 \rightarrow D_4 \rightarrow \underline{\mathbf{T_5}}$ for short progressions, and $T_1 \rightarrow T_2 \rightarrow PD_3 \rightarrow D_4 \rightarrow \underline{\mathbf{T_5}} \rightarrow D_6 \rightarrow T_7$ for long progressions), while atypical progressions had an unexpected chord in position five ($T_1 \rightarrow T_2 \rightarrow PD_3 \rightarrow D_4 \rightarrow \underline{\mathbf{PD_5}}$ for short progressions, and $T_1 \rightarrow T_2 \rightarrow PD_3 \rightarrow D_4 \rightarrow \underline{\mathbf{PD_5}} \rightarrow D_6 \rightarrow T_7$ for long progressions). Note that short-typical progressions are identical to the first five chords of long-typical progressions, and short-atypical progressions are identical to the first five chords of long-atypical progressions. Thus, all conditions have equivalent priming context that precede the target chords (chords 1–4).

Figure 1

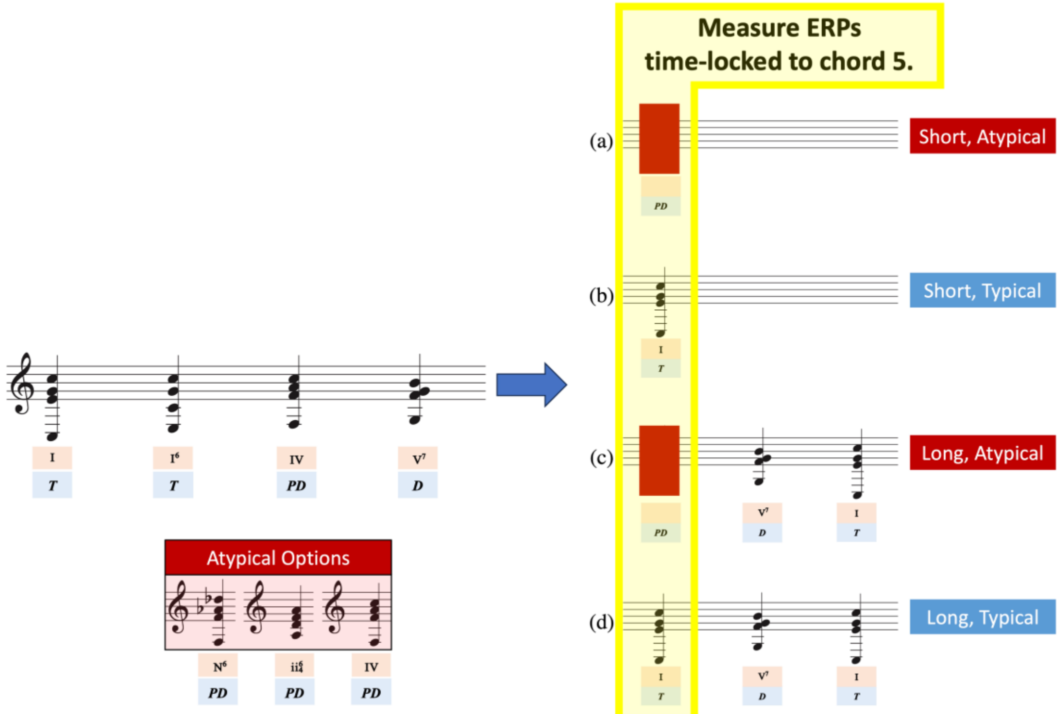

Figure 1: Experimental stimuli. An example is displayed in C major, though the actual pitch level of stimuli varied from trial to trial. For atypical conditions (a,c), atypical chords can be one of three possible PD options. For the long conditions (c,d), atypical chords go on to properly (and predictably) resolve with chords six and seven. Roman numeral music notation in boxes below the notated chords indicate specific realizations of chord functions, which are themselves indicated below the numerals.

The unexpected target chord, for atypical trials, was always a PD-function chord; we used three possible options for PD chords (specifically, using Roman numeral notation from music theory, a IV chord, a ii chord, and an N^6 chord; for example, in the key of C major, these options correspond to an F major chord, a d minor chord, and a Db major chord respectively). Atypical progressions were split equally between the three different PD options. Thus, we included 60 of each type of trial (short-typical, short-atypical, long-typical, long-atypical), totaling 240 stimuli. For each trial type, those 60 trials were further divided into subgroups. For atypical conditions, the 60 trials were divided into three groups of 20 for each of the three PD options (i.e., 20 N^6 chords, 20 IV chords, and 20 ii chords). Each of these groups of 20 trials were further divided into four different pitch levels (varying the musical key), five trials per pitch level. For typical conditions, in which there was only one exemplar for the target T_5 chord, there were also 60 trials; we divided those trials into four groups of 15, one group for each of the four pitch levels.

The audio was rendered using PsychoPy in a square wave tone using a percussive amplitude envelope with an onramp of 25 ms, and an exponential decay of 725 ms. The inter-onset interval of the chords was 750 ms. A square wave tone was used for convenience so that we could easily generate and precisely control the audio parameters within our experimental presentation script.

Questionnaire

The questionnaire collected demographic information (age, gender, and self-reported years of musical training) as well as standardized metrics pertaining to musical training and perceptual abilities using the standardized GMSI.

EEG Apparatus

While listening to the chord progressions, EEG was simultaneously recorded through 64 active scalp electrodes arranged according to the 10-20 system, 2,048 Hz sampling rate, using a BioSemi ActiveTwo System (AD Box version 7.0; 24-bit AD conversion) in a sound attenuated chamber. Participants listened to audio using Etymotic Research 3C insert earphones (10 Ohm).

Procedure

Our study was approved by [our university's] IRB. After consenting to participate, participants were outfitted with the EEG caps and seated in the soundproof booth. Following a short practice session with four trials (one from each of the four conditions), they listened to the 240 chord progressions in a random order (see Figure 2). For each trial, participants saw a fixation cross with the text "…next trial…" (500 ms), one of two images indicating the progression length by displaying either five or seven notes (700–800 ms, uniformly randomly distributed), and then heard the chords one at a time while the image remained on the screen. To test whether participants were paying attention, on 20% of trials (randomly selected), following the audio presentation, participants were prompted to recall how many chords were in the progression they just heard by pressing "5" or "7" on a computer keyboard, and were provided feedback on their response (correct or incorrect).

After completing the EEG portion of the experiment, we removed the EEG caps, and participants completed the questionnaire. Participants were compensated with a $30 gift card.

Figure 2

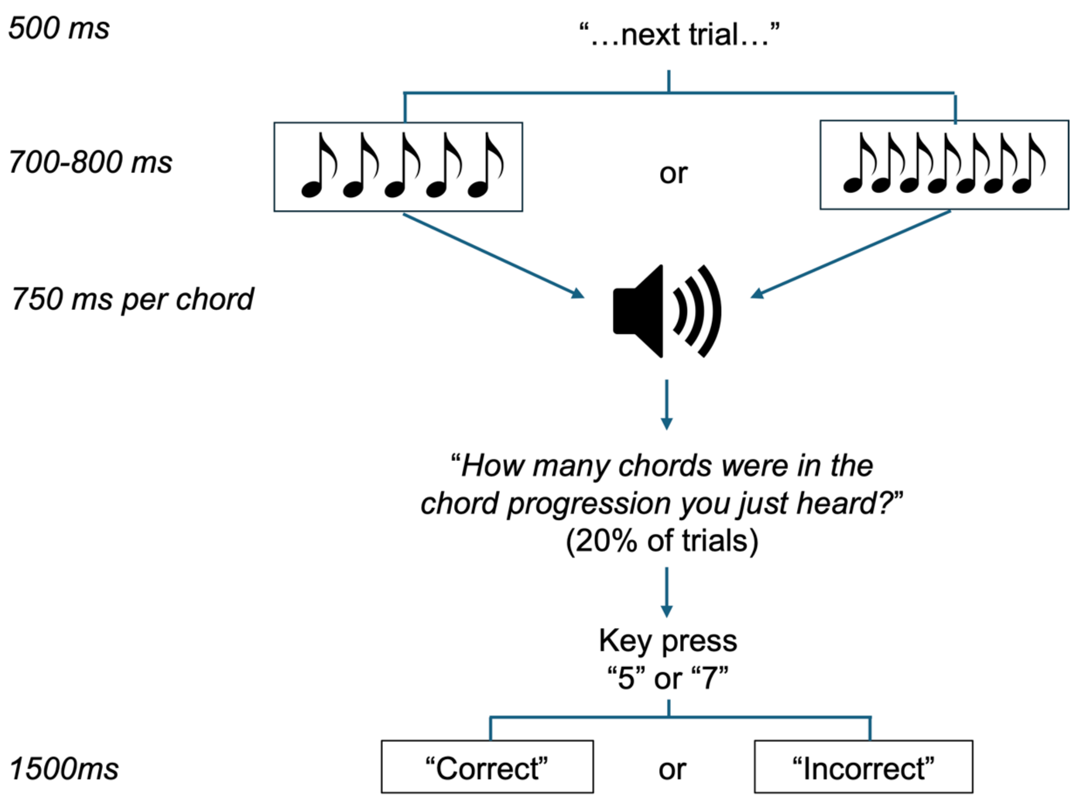

Figure 2: Trial structure.

Analysis

EEG Data Processing

Data files were processed using EEGLAB version 2023.0 (Delorme & Makeig, 2004). We first downsampled files to 256 Hz. We high-pass filtered the data using an FIR filter (0.50 Hz passband edge, 0.25 Hz cutoff frequency) and then low-pass filtered it using an FIR filter (50 Hz passband edge, 56.25 Hz cutoff frequency). Data with excessive movement or sweat artifacts were removed by eye. We then ran an Independent Components Analysis using the infomax algorithm (first reducing the data to 32 principal components). We removed eye blink independent components from each dataset. We then rereferenced the data to the common average reference, and extracted epochs from the data time locked to the fifth chord in each trial, from 50 ms before it to 750 ms after it. We removed the baseline from each epoch using the average voltage from -50 ms to 0 ms. Finally, we removed any epochs with significant movement, sweat, and other artifacts by eye.

Preprocessed data are available at this project's OSF page: https://osf.io/f4t9d/?view_only=22e672af28ba4ce4b6043b3b05f6449f.

ERP Component Amplitude Measurement

Component latencies were identified using time ranges in previous literature (see Goldman et al., 2021). All time ranges are following the target chord (fifth chord). For the E(R)AN component, we used 180–230 ms; for the P3a component, we used 250–400 ms; for the P600 component, we used 550–750 ms. Note that for long conditions, the sixth chord would sound at 750 ms after the fifth chord and does not overlap with the ERP component time ranges. For anteriorly distributed components (E(R)AN and P3a), we computed the average amplitude over fronto-central electrodes, specifically F1, FC1, F2, FC1, FCz, FC2, C1, Cz, C2. For the

P600, we used posterior-central electrodes CP1, CPz, CP2, P1, Pz, and P2. To statistically model the different amplitudes across conditions, we conducted separate 2-way repeated measures ANOVAs for each ERP component with factors length (short versus long) and typicality (typical versus atypical), and followed up with post-hoc pairwise comparisons using paired *t*-tests.

Before subjecting our data to the ANOVA analysis, we conducted Shapiro-Wilk tests on each condition's data for each component to satisfy the normality assumption. Data for the E(R)AN and P3a analyses satisfied the assumption. However, we discovered that the long-atypical (p = .038) and long-typical ($p < .001$) data for the P600 analysis significantly deviated from normality; upon examining this data, we discovered one participant's data was outlying (z = 3.74 SDs for long-typical and z = 2.96 SD for long-atypical); after removing this participant's data from the P600 ANOVA only, all Shapiro-Wilk tests supported the normality assumption.

Results

All participants got at least 46 out of 48 attention check questions correct (M = 47.52, SD = 0.68), indicating that participants were aware of the progression lengths.

Figure 3 displays the whole scalp topographies for atypical–typical difference ERPs averaged over latencies corresponding to the E(R)AN, P3a, and P600 components. Note the larger amplitudes for the long conditions.

Figure 3

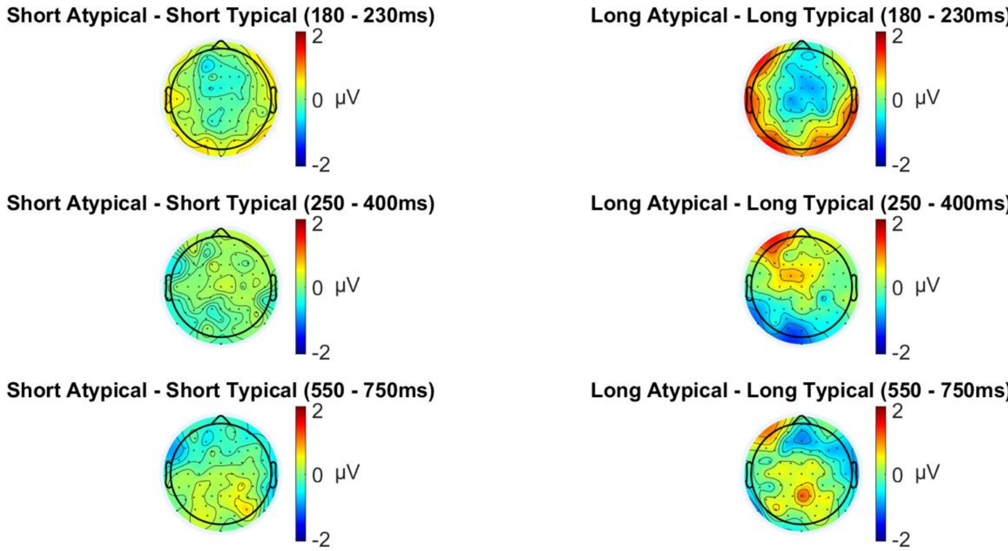

Figure 3: Topographies of amplitudes averaged over latency ranges corresponding to the E(R)AN (top), P3a (middle), and P600 (bottom). ERP components have larger amplitudes for the long conditions.

Figure 4 displays the anterior-central and posterior-central ERP waveforms for all four conditions. Figure 5 displays difference waves (atypical–typical) for the short and long conditions with standard errors.

Figure 4

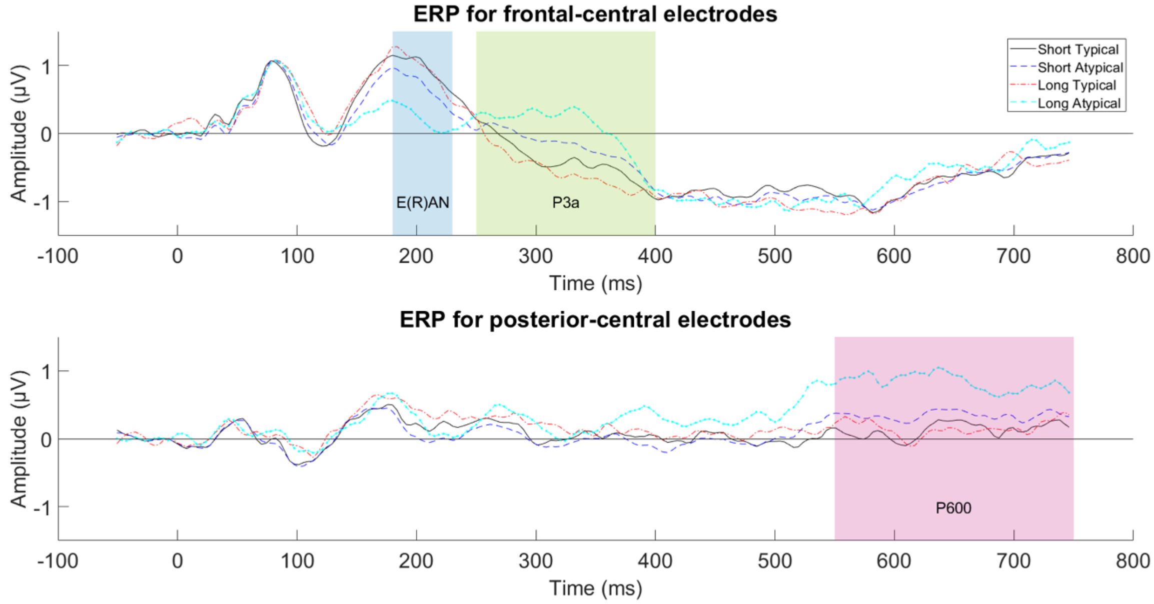

Figure 4: ERPs for the four conditions averaged over different scalp regions. The target chord occurs at time 0. For the long conditions, the following sixth chord would occur at 750 ms.

Figure 5

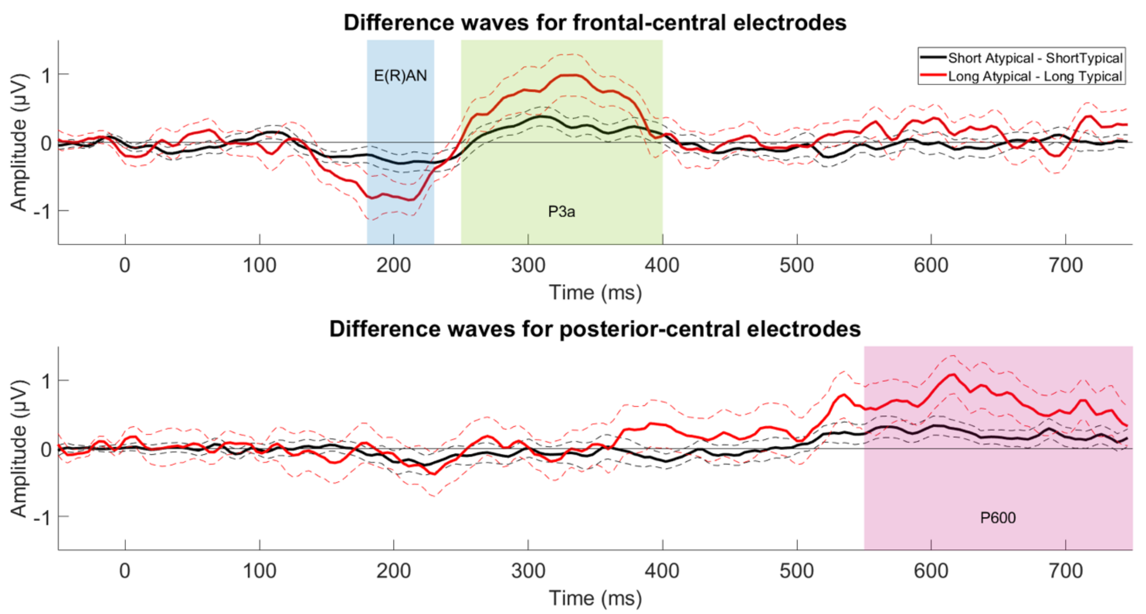

Figure 5: Difference waves (atypical minus typical) for the short and long conditions. Dotted lines represent one standard error above and below the mean. The target chord occurs at time 0. For the long conditions, the following sixth chord would occur at 750 ms.

The 2-way repeated measures ANOVA for the E(R)AN component found a significant main effect of typicality, $F(1,20) = 13.61$, $p = .001$, $\eta^2_p = .40$; component amplitudes were significantly more negative for atypical compared to typical trials. There was no significant main effect of length. There was a significant interaction between length and typicality, $F(1,20) = 6.80$, $p = .017$, $\eta^2_p = .25$. Post-hoc *t*-tests showed a significant difference between short-atypical and short-typical conditions, $t(20) = -2.29$, $p = .033$, Cohen's $d = -0.24$, and a difference between long-atypical and long-typical conditions, $t(20) = -3.69$, $p = .002$, Cohen's $d = -0.92$. Both short and long progressions show a significant E(R)AN for atypical vs. typical stimuli, but the effect size for the long progressions (-0.92) was much larger than for the short progressions (-0.24).

The 2-way repeated measures ANOVA for the P3a component found a significant main effect of typicality, $F(1,20) = 9.87$, $p = .005$, $\eta^2_p = .33$; component amplitudes were significantly more positive for atypical compared to typical trials. There was no significant main effect of length. The interaction between length and typicality was marginally significant, $F(1,20) = 4.01$, $p = .059$. Post-hoc *t*-tests between short-atypical and short-typical progressions did not have a significant difference for the P3a component ($p = .078$), but the comparison between long-atypical and long-typical progressions was significant, $t(20) = 2.95$, $p = .008$, Cohen's $d = 0.71$. Both short and long progressions showed a P3a for atypical vs. typical stimuli, but it was only significant for the long progressions, with a medium-large effect size.

The 2-way repeated measures ANOVA for the P600 component found a significant main effect of typicality, $F(1,19) = 8.57$, $p = .009$, $\eta^2_p = .31$; component amplitudes were significantly more positive for atypical compared to typical trials. There was no main effect of length and no interaction. Post-hoc *t*-tests between short-atypical and short-typical progressions did not have a significant difference for the P600 component (p = .061), but the comparison between long-atypical and long-typical progressions was significant, $t(19) = 2.42$, $p = .026$, Cohen's $d = 0.62$. Short and long progressions considered together showed a P600 for atypical vs. typical stimuli; post-hoc tests showed it was only significant for the long progressions, with a medium effect size.

Finally, all Spearman correlation tests between participants' GMSI musical training subscale scores and ERP component amplitudes (as calculated by difference waves between typical and atypical conditions) were not significant ($p > .05$), nor were the Spearman correlation tests between the GMSI perceptual abilities subscale scores and ERP component amplitudes (all $p > .05$).

N1 Component Analysis

In order to rule out an interpretation of our main findings based on a global attentional difference between the short and long conditions, we also compared N1 component amplitudes across the four conditions with the same ANOVA model structure as for the other components. We measured the N1 component amplitude at electrode Cz, with a latency range of 80–120 ms, following Lange & Schnuerch (2014). A two-way repeated measures ANOVA did not return a significant main effect for length (p = .329), typicality (p = .439), or interaction between length and typicality (p = .914). N1 component amplitude did not differ systematically by condition.

Discussion

This study investigated whether listeners' prior knowledge of the syntactic structure of a musical chord progression (but not its specific content) would modulate how they process unexpected harmonic events within that progression, as indexed by ERP components. In accordance with a broad collection of previous studies, we found that atypical (surprising) harmonies evoked the ERP components normally associated with such unexpected stimuli, including the E(R)AN, P3a, and P600. Crucially, though, we found that these difference wave component amplitudes had greater magnitude for the long chord progressions with harmonies succeeding the target than those without, indicating that a listener's knowledge of the syntactic structure of the chord progressions indeed modulates these ERP components. Again, note that nothing differed in the physical stimuli between short and long conditions until after the target chord (in the long conditions, additional chords followed). The only difference between the short and long conditions is whether more chords would come *after* the fifth chord, and the listener's knowledge of whether or not there would be that subsequent context. Thus, differences in the ERP component could only be due to top-down processing; in this case, that processing is related to a listener's knowledge of the syntactic structure of the stimuli.

Before offering an explanation for the direction of these effects, that is, that the ERP components apparent in the difference waves for longer progressions had *larger* amplitudes than for the shorter progressions, we argue against two possible interpretations of the data. First, the larger amplitudes of the difference waves for long compared to short conditions may have been a result of different general attentional demands; in the long condition, participants must "listen on" from the fifth chord, perhaps ignoring it in the short condition in which the fifth chord is the final stimulus in the sequence. After all, attention has sometimes been shown to increase the amplitude of the E(R)AN component (Loui et al., 2005), so perhaps differences between the short and long conditions were merely due to attentional differences. However, purely attentional effects have mixed results in the literature: Lee et al., 2019 did not find such an effect. Further, if some global attention effect were driving the differences between conditions, we would also have expected to see earlier differences in the auditory ERP, specifically, in the auditory N1 component, as attention (Hillyard, et al., 1973), including temporally-oriented attention (Lange & Schnuerch, 2014), has been shown to increase auditory N1 amplitudes. In our case, there was no difference in the N1 amplitude, providing evidence that the observed differences were not due to general attentional demand characteristics of the short versus long conditions. However, a more precise notion of attention may indeed explain our findings, namely, attentional gain modulation as discussed below.

Secondly, perhaps the short and long conditions differ with regard to wrap-up effects. We intended to compare conditions in which participants adopted different syntactic interpretations of a target chord, but our specific implementation ended up comparing sequence-final with sequence-medial target chords. Perhaps the sequence-final targets in our short-atypical condition were processed differently than the sequence-medial targets in our long-atypical condition by virtue of being at the end of the progression. There are two reasons we do not favor this interpretation of the data. First, in previous linguistic ERP wrap-up effects studies that compared P600 amplitudes for sequence-final vs. sequence-medial syntactic errors, sequence-final targets produced either similar (Osterhout, 1997) or *larger* amplitudes (Kuperberg et al., 2003, p. 123) than sequence-medial targets, and in music studies of the E(R)AN, sequence-final syntactic anomalies produced larger amplitudes than sequence-medial ones (Koelsch et al., 2002; Leino et al., 2007). The received interpretation of such differences is that some additional sequence-final cognitive integration process (a wrap-up effect) occurs at sequence boundaries. But in our findings, the sequence-final targets, i.e., those in the short-atypical condition, produced *smaller* amplitudes than in the sequence-medial targets (i.e., targets in the long conditions). Secondly, our goal was to investigate whether knowledge of syntactic structure would modulate the processing of unexpected events. It is true that “wrap-up” may be part of such differences in processing, but even in that case, participants would only know to wrap-up because we cued them to the length of the progression—again, a cue to syntactic structure. In this sense, whether or not the difference in surprisal processing is due to the general influence of syntactic processes or the somewhat more specific wrap-up syntactic processes, the findings still support our more general hypothesis that knowledge of the syntactic role of a target chord affects how listeners process unexpected events. Still, we acknowledge that it could be helpful to know whether such syntactically driven effects might *also* occur with stimuli designed to compare targets at two sentence-medial positions.

Instead of general attention and wrap-up effects, we offer two interpretations of our findings. The first interpretation is general. On the basis of our findings, we argue that the E(R)AN, and subsequent P3a and P600 ERP components—broadly theorized to be an index of harmonic surprisal processing—also depends on a listener's knowledge of the specific syntactic structure of local musical context. Our data show that listeners can change how they process such harmonies depending on expectations of what will follow the chord. The neural signals point to listeners directly modulating their harmonic expectations based on their specific knowledge of the syntactic structure of a musical passage, and that harmonic surprisal processing is not an automatic and encapsulated process purely dependent on accumulating evidence from context prior to a target chord.

Secondly, and more specifically, our findings can be interpreted within the predictive-coding framework (Friston, 2002), and specifically following Koelsch et al. (2019). According to this framework, the brain constantly generates expectations about upcoming sensory events, and violations of these expectations lead to prediction error signals (Friston, & Kiebel, 2009; Koelsch et al., 2019; Vuust et al., 2009). Given some priming sequence, listeners can make *first-order* *predictions* of what will happen next; also, listeners can make *second-order predictions* by evaluating how certain they ought to be about their own prediction, a value termed *precision*. Essentially, higher precision means affording greater weight to prediction errors when they are processed further. Koelsch et al. (2019) interpret the E(R)AN as a signal indexing this second-order prediction error. They reference a study by Vuvan, Zendel, & Peretz (2018) in which listeners were asked to judge whether target stimuli fit a tonal context or not; by providing them with random (deceptive) feedback about their response accuracy, Vuvan et al. were able to increase listeners' E(R)AN amplitudes, which returned to normal levels after the feedback became veridical in later blocks. Vuvan et al. interpret this increase as an attentional effect, which Koelsch et al. (2019) later recast as a second-order prediction effect: the deceived listeners increased their precision on the unexpected events.

Similarly, we argue that target chords in our long condition are afforded a greater precision: the prediction error is more salient, more informative, and increases the attentional gain. Our study also found increased amplitudes in a high precision condition (our long-atypical condition), but here, the precision was not artificially manipulated with deceptive feedback, but rather by cueing listeners to the syntactic structure of the stimuli.

Following this theoretical framework, in our short conditions, chords 1–4 encourage listeners to make a strong prediction about chord five, but *with low precision*: they most likely predict a T chord, and given that it is the end of the sequence, they are confident about their predictions. In this case, if the final chord is not what was predicted (short-atypical condition), the low precision means the listener will disregard their prediction error. It is not informative; after all, the stimuli in the experiment are known to be unpredictable, and contain the syntactic violation 50% of the time. It is schematically still unpredictable given long-term exposure to the regularities of harmonic syntax heard outside the laboratory, so some E(R)AN is still evoked. However, in the long conditions, listeners may still predict that chord five will be a T chord, but with *higher precision* because now that chord becomes informative, helping the listener understand and integrate what comes next. One continuation is when the typical resolution (T_5) is reinforced with a subsequent $D_6 \rightarrow T_7$ (in musical terms, a short codetta that reinforces the cadence), and the option in which an unexpected event (PD_5) goes on to properly resolve (in musical terms, an evaded cadence that ultimately resolves later). Effectively, they still predict T for chord 5, but know that another chord is also plausible (a PD chord) given that there will be subsequent context. Hence, we argue that the target chords in the long-atypical condition are more informative, and generate a larger E(R)AN.

This interpretation explains the differences in the P3a and P600 amplitudes as well. Following the detection of the improbable stimulus, the P3a reflects an attentional reorientation and information integration process (Goldman et al., 2021; Vuust et al., 2009). Then, the P600 component is indexing a re-evaluation process given the contextual incongruity of the atypical target chords (Featherstone et al., 2013) which is more relevant for the long condition in which an atypical fifth chord urges the listener to reinterpret that chord as "belonging" to the succeeding chords (hearing the final $PD_5 \rightarrow D_6 \rightarrow T_7$ part of the progression) as opposed to being the resolution of the preceding $PD_3 \rightarrow D_4$.

We note that future studies of musical expectation could systematically manipulate precision using a variety of music-theoretically motivated comparisons. Which kinds of music-contextual features may necessitate greater attentional gain is a rich and colorful question; the link we provide here between canonical observations of syntax-related ERP components and syntactic manipulations of musical expectations suggest other possible manipulations of listeners' certainty in their predictions.

One limitation of our study was that we worked only with musically trained participants. Previous work has found more robust responses to improbable musical events for musicians compared to non-musicians (Kim, Kim, & Chung, 2011; Koelsch et al., 2002). Given the novelty of our experimental question, we wished to maximize the chances of detecting a difference between our conditions if indeed one would be present; thus, we worked with musicians. On the other hand, it remains unclear the degree to which expertise interacts with our findings. Perhaps knowledge of syntactic structure only affects musicians, given their explicit training in music theory or more extensive experience with the syntax of our Western tonal harmony-based stimuli. At least for our sample, our measures of musical ability did not correlate with component amplitudes, although again, our participants were in the upper percentiles of the scale. Nevertheless, our findings still refine the understanding of ERP components associated with harmonic surprisal, and explain top-down effects of neural processing in music listening for at least some listeners.

In conclusion, our results advance the understanding of musical harmony processing, showing that even early neural responses to harmonic violations are shaped by complex interactions among bottom-up statistical learning, top-down contextual knowledge, and active integration processes. This highlights the sophisticated nature of musical expectation and underscores its essential role in shaping our experience of music.

Data availability statement

Our project has an OSF page: https://osf.io/f4t9d/?view_only=22e672af28ba4ce4b6043b3b05f6449f. Upon acceptance, we would provide a non-anonymous link to the page, make the page public, and add raw data files as well.

Author contributions

To be shared if the manuscript is accepted.

Acknowledgments

To be shared if the manuscript is accepted.

Funding information

To be shared if the manuscript is accepted.

References

Auksztulewicz, R., Friston, K. (2015) Attentional enhancement of auditory mismatch responses: A DCM/MEG Study, *Cerebral Cortex*, *25*(11), 4273–4283, https://doi.org/10.1093/cercor/bhu323

Baumann, S., Meyer, M., & Jäncke, L. (2008). Enhancement of auditory-evoked potentials in musicians reflects an influence of expertise but not selective attention. *Journal of Cognitive Neuroscience*, *20*(12), 2238–2249. https://doi.org/10.1162/jocn.2008.20157

Bigand, E., Poulin, B., Tillmann, B., Madurell, F., & D'Adamo, D. A. (2003). Sensory versus cognitive components in harmonic priming. *Journal of Experimental Psychology: Human Perception and Performance*, *29*(1), 159–171. https://doi.org/10.1037/0096-1523.29.1.159

Burstein, L. P. (2014). The half cadence and other such slippery events. *Music Theory Spectrum*, *36*(2), 203–227. https://doi.org/10.1093/mts/mtu010

Caplin, W. E. (2013). *Analyzing classical form: An approach for the classroom*. Oxford University Press.

Chen, X., Affourtit, J., Ryskin, R., Regev, T. I., Norman-Haignere, S., Jouravlev, O., Malik-Moraleda, S., Kean, H., Varley, R., & Fedorenko, E. (2023). The human language system, including its inferior frontal component in "Broca's area," does not support music perception. *Cerebral Cortex*, *33*(12), 7904–7929. https://doi.org/10.1093/cercor/bhad087

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

Dubiel, J. (2017). Music analysis and kinds of hearing-as. *Music Theory and Analysis (MTA)*, *4*(2), 233–242. https://doi.org/10.11116/MTA.4.2.4

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

Featherstone, C. R., Morrison, C. M., Waterman, M. G., & MacGregor, L. J. (2013). Semantics, syntax or neither? A case for resolution in the interpretation of N500 and P600 responses to harmonic incongruities. *PLoS One*, *8*(11), e76600. https://doi.org/10.1371/journal.pone.0076600

Feldman, H., & Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Frontiers in human neuroscience*, *4*, 215.

Friston, K. (2002). Beyond phrenology: What can neuroimaging tell us about distributed circuitry?. *Annual Review of Neuroscience*, *25*(1), 221–250. https://doi.org/10.1146/annurev.neuro.25.112701.142846

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

Goldman, A., Harrison, P. M., Jackson, T., & Pearce, M. T. (2021). Reassessing syntax-related ERP components using popular music chord sequences: A model-based approach. *Music Perception*, *39*(2), 118–144. https://doi.org/10.1525/MP.2021.39.2.118

Guck, M. A. (2017). Perceptions, Impressions: When Is Hearing "Hearing As"? *Music Theory and Analysis (MTA)*, *4*(2), 243–254. https://doi.org/10.11116/MTA.4.2.5

Guo, S., & Koelsch, S. (2016). Effects of veridical expectations on syntax processing in music: Event-related potential evidence. *Scientific Reports*, *6*(1), 19064. https://doi.org/10.1038/srep19064

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177–180. https://doi.org/10.1126/science.182.4108.177

Huron, D. (2008). *Sweet anticipation: Music and the psychology of expectation*. MIT Press. https://doi.org/10.7551/mitpress/6575.001.0001

Kim, S. G., Kim, J. S., & Chung, C. K. (2011). The effect of conditional probability of chord progression on brain response: An MEG study. *PloS one*, *6*(2), e17337. https://doi.org/10.1371/journal.pone.0017337

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., Gunter, T. C., Cramon, D. Y. V., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: a cortical "language-network" serves the processing of music. *Neuroimage*, *17*(2), 956–966. https://doi.org/10.1006/nimg.2002.1154

Koelsch, S., Schmidt, B. H., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology*, *39*(5), 657–663. https://doi.org/10.1017/S0048577202010508

Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, *17*(10), 1565–1577. https://doi.org/10.1162/089892905774597290

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. (2009). Music-syntactic processing and auditory memory: Similarities and differences between ERAN and MMN. *Psychophysiology*, *46*(1), 179–190.

Koelsch, S. (2011). Towards a neural basis of processing musical semantics. *Physics of Life Review*, 8, 89–105. https://doi.org/10.1016/j.plrev.2011.04.004

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

Kuperberg, G. R., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research*, *17*(1), 117–129. https://doi.org/10.1016/S0926-6410(03)00086-7

Lange, K., & Schnuerch, R. (2014). Challenging perceptual tasks require more attention: The influence of task difficulty on the N1 effect of temporal orienting. *Brain and Cognition*, *84*(1), 153–163. https://doi.org/10.1016/j.bandc.2013.12.001

Lee, D. J., Jung, H., & Loui, P. (2019). Attention Modulates Electrophysiological Responses to Simultaneous Music and Language Syntax Processing. *Brain Sciences*, *9*(11), 305. https://doi.org/10.3390/brainsci9110305

Leino, S., Brattico, E., Tervaniemi, M., & Vuust, P. (2007). Representation of harmony rules in the human brain: Further evidence from event-related potentials. *Brain Research*, 1142, 169–177. https://doi.org/10.1016/j.brainres.2007.01.049

Lerdahl, F., & Jackendoff, R. (1983). *A Generative Theory of Tonal Music*. MIT Press. https://doi.org/10.7551/mitpress/12513.001.0001

Lewin, D. (1986). Music theory, phenomenology, and modes of perception. *Music Perception*, *3*(4), 327–392. https://doi.org/10.2307/40285344

Loui, P., Grent-'t-Jong, T., Torpey, D., & Woldorff, M. (2005). Effects of attention on the neural processing of harmonic syntax in Western music. *Cognitive Brain Research*, *25*, 678–687. https://doi.org/10.1016/j.cogbrainres.2005.08.019

Loui, P., Wu, E. H., Wessel, D. D., & Knight, R. T. (2009). A generalized mechanism for perception of pitch patterns. *Journal of Neuroscience*, *29*(2), 454–459. https://doi.org/10.1523/jneurosci.4503-08.2009

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, *4*(5), 540–545. https://doi.org/10.1038/87502

Meyer, L. (1956). *Emotion and meaning in music.* The University of Chicago Press.

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLOS ONE*, *9*(2), e89642. https://doi.org/10.1371/journal.pone.0101091

Osterhout, L. (1997). On the brain response to syntactic anomalies: Manipulations of word position and word class reveal individual differences. *Brain and Language*, *59*(3), 494–522. https://doi.org/10.1006/brln.1997.1793

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. https://doi.org/10.1038/nn1082

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, *10*(6), 717–733. https://doi.org/10.1162/089892998563121

Pearce, M., & Rohrmeier, M. (2018). Musical Syntax II: Empirical Perspectives. In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 487–505). Springer. https://doi.org/10.1007/978-3-662-55004-5_26

Pearce, M. T. (2005). *The construction and evaluation of statistical models of melodic structure in music perception and composition* [Doctoral dissertation, City University London].

Regnault, P., Bigand, E., & Besson, M. (2001). Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: Evidence from auditory event-related brain potentials. *Journal of Cognitive Neuroscience*, *13*(2), 241–255. https://doi.org/10.1162/089892901564298

Rohrmeier, M., & Pearce, M. (2018). Musical Syntax I: Theoretical Perspectives. In R. Bader (Ed.), *Springer handbook of systematic musicology* (pp. 473–486). Springer. https://doi.org/10.1007/978-3-662-55004-5_25

Slevc, R. R., Rosenberg, J. C., & Patel, A. D. (2009). Making psycholinguistics musical: Self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychonomic Bulletin & Review*, *16*(2), 374–381. https://doi.org/10.3758/16.2.374

Steinbeis, N., Koelsch, S., & Sloboda, J. A. (2006). The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *Journal of Cognitive Neuroscience*, *18*(8), 1380–1393. https://doi.org/10.1162/jocn.2006.18.8.1380

Stowe, L. A., Kaan, E., Sabourin, L., & Taylor, R. C. (2018). The sentence wrap-up dogma. *Cognition*, *176*, 232–247. https://doi.org/10.1016/j.cognition.2018.03.011

Temperley, D. (2004). *The Cognition of Basic Musical Structures*. MIT Press.

Tillmann, B., & Marmel, F. (2013). Musical expectations within chord sequences: Facilitation due to tonal stability without closure effects. *Psychomusicology: Music, Mind, and Brain*, *21*(1), 1–5. https://doi.org/10.1037/a0030454

Vuust, P., Heggli, O.A., Friston, K.J., & Krigelbach, 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*, *45*(1), 80–92. https://doi.org/10.1016/j.cortex.2008.05.014

Vuvan, D. T., Zendel, B. R., & Peretz, I. (2018). Random feedback makes listeners tone-deaf. *Scientific Reports*, *8*(1), 7283. https://doi.org/10.1038/s41598-018-25518-1

Vuvan, D. T., & Hughes, B. (2019). Musical style affects the strength of harmonic expectancy. *Music & Science*, *2*, 1–9. https://doi.org/10.1177/2059204318816066

White, C. W., & Quinn, I. (2018). Chord context and harmonic function in tonal music. *Music Theory Spectrum*, *40*(2), 314–335O. https://doi.org/10.1093/mts/mty021