"Hearing As": Top-down Processing Affects Early 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

Yeoeun Lim (yeolim@iu.edu),
Cognitive Science Program, Indiana University
1001 E. 10th Street, Bloomington, IN 47405, USA

Megan Kibler (mkibler@iu.edu)
Cognitive Science Program, Indiana University
1001 E. 10th Street, Bloomington, IN 47405, USA

Nazbanou Nozari (bnozari@iu.edu)
Department of Psychological and Brain Sciences & Cognitive Science Program, Indiana University
1101 E. 10th Street, Bloomington, IN 47405, USA

Abstract

Harmonic expectation is an important generator of musical experience often explained through mechanisms of statistical learning. 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 influence remains under-explored. 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 surprising chords when listeners knew that additional context would follow a surprising harmony. We propose that listeners prospectively integrate surprising chords with anticipated future context, rather than responding to them solely through automatic probability assessment. Musical surprisal arises from a dynamic interplay between bottom-up cues and a listener's top-down anticipated syntactic structure.

Keywords: statistical learning; music; EEG; syntax; neuroscience

Background

Music psychologists have identified harmonic expectation (whether fulfilled or subverted) as an important mechanism of musical experience (Huron, 2008; Meyer, 1956). Cognitive scientists mainly explain the mechanisms of expectation in terms of statistical learning: listeners acquire knowledge of statistical dependencies of music through repeated exposure to the regularities of musical styles (Rohrmeier & Pearce, 2018). Further, neuroscientists have identified neural correlates of expectation—often couched in terms of probabilistic musical syntax—using MEG and fMRI to trace 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). Neuroscientists have also leveraged the temporal resolution of EEG to discriminate between components 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. For surprising compared to expected chords, an Early (Right) Anterior Negativity, or E(R)AN (which is sometimes found to be centrally rather than right-distributed), is evoked, and is theorized to index the violation of expectation based on long-term memory of harmonic transition probabilities (Koelsch, Vuust, & Friston, 2019)—in contrast to the Mismatch Negativity (MMN) which is theorized to index local sensory surprisal (Koelsch, 2009).

Such paradigms also evoke later ERP components including the P3a, N5, and P600 (see Goldman et al., 2021), corresponding to different aspects of the cognitive appraisal of surprising chords. The P3a indexes task-irrelevant surprisal for the chord, while the N5 and P600 amplitudes are indexing the effort of integrating the surprising harmony into the context.

Longer priming contexts can increase the E(R)AN amplitude (Leino et al., 2007). Attention increases E(R)AN's amplitude, too (Loui et al., 2005), and a listener's knowledge of an impending surprising chord decreases its latency (Guo & Koelsch, 2016). These latter two findings point to the role of top-down processing in harmonic surprisal.

Different musical styles, for instance, have different expectations because the harmonic transition probabilities differ across those styles (what typically happens in classical music does not necessarily happen in rock music, for instance); behavioral evidence has shown that listeners can change their expectations based on knowledge of what musical genre they are hearing (Vuvan & Hughes, 2019). Along these lines, music theorists have long noted these issues in phenomenological work (Dubiel, 2017; Guck, 2017; Lewin, 1986) which describe how a listener's stance can

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

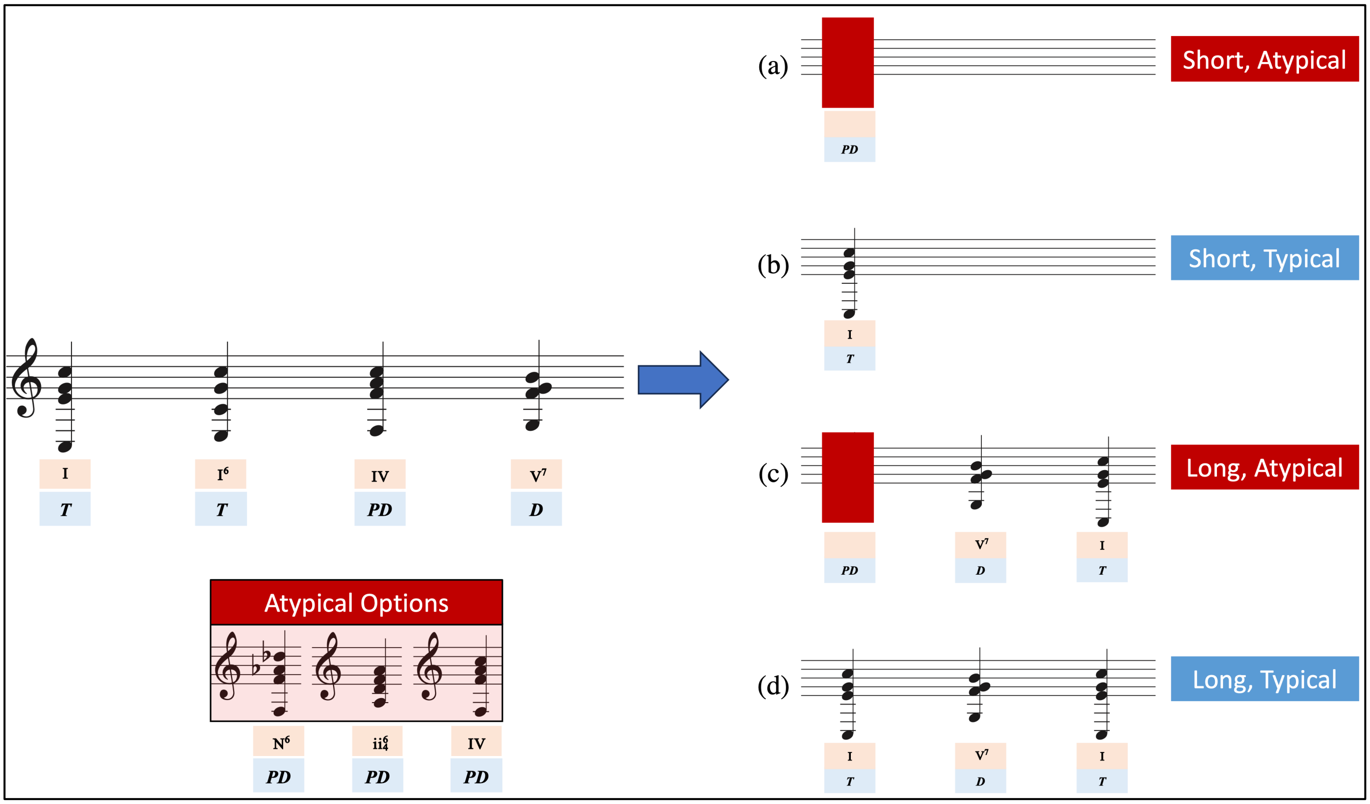

Figure 1: Experimental stimuli. For atypical conditions (a,c), atypical chords can take one of three types. For the long conditions (c,d), atypical chords go on to properly (and predictably) resolve with chords 6 and 7. Roman numerals in boxes below the notated chords indicate specific realizations of chord functions, which are themselves indicated below the numerals.

affect how the same physical stimulus—i.e., a sounding harmony—is perceived.

These observations about top-down processing, however, are global in the sense that listeners can modulate their expectations according to style. However, listeners might be able to modulate their expectations in a more nuanced and local manner. For instance, supposedly surprising chords can be embedded in longer sequences that proceed from the surprising chord in quite predictable ways, rendering that surprising chord less surprising in context. If a listener knows that a surprising event is embedded in a longer context—that is, if they are cued beforehand as to the syntactic structure of the music—then the surprising event may be processed differently.

For example, the same target harmony—which follows a priming sequence—may be surprising in one context, but less surprising in another depending on which harmonies *follow* the target harmony. This is because surprising chords often go on to *resolve* (i.e., proceed in an expected manner). Typical harmonic syntax (in Western classical music) will go from a tonic chord (T) to a predominant chord (PD) to a dominant chord (D) back to a tonic chord: T→PD→D→T. By replacing the final T with a PD chord (of which there are several specific exemplars), the syntax is violated: T→PD→D→**PD**. But, by adding additional context *following* this surprising PD such as a subsequent D→T (which is the typical resolution of a PD chord), that surprising PD suddenly is functioning typically again: T→PD→D→**PD**→D→T. Music theorists sometimes call this progression an *evaded cadence*: the expected resolution from D→T is temporarily evaded in positions 3–4 (the T is replaced with a PD), but goes on to resolve properly to T at the end.

If earlier ERP components including the E(R)AN are theorized to index harmonic surprisal itself, based on a listener's long-term knowledge of harmonic expectations, then one might theorize it to be invariant to certain forms of top-down processing that dynamically adjust expectations to local context. Attention 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 other forms of top-down processing can modulate the E(R)AN.

Thus, here we ask whether ERP components known to be sensitive to improbable syntactic events are sensitive to a listener's knowledge of the syntactic structure of a chord progression. If they are, it suggests that listeners can adapt their harmonic expectations dynamically rather than statically and automatically evaluating harmonic predictability.

Methods

Participants

Twenty-six participants took part in the study (14 male, 11 female, 1 non-binary; Age *M* = 22.23 years, *SD* = 6.09 years). Participants all had substantial musical training, as assessed by the self-reported years of formal music lessons (*M* = 12.15, *SD* = 4.94) and the Goldsmiths Musical Sophistication Index (GMSI; Müllensiefen et al., 2014) Musical Training subscale, *M* = 40.27 (87^{th} percentile), *SD* = 4.26, as well as the GMSI Perceptual Abilities subscale, *M* = 53.46 (64^{th} percentile), *SD* = 5.76.

Materials

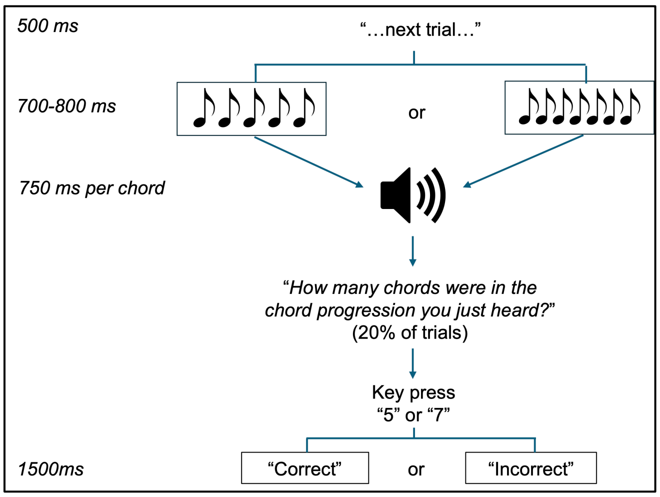

Figure 2: Trial structure.

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). Short progressions were five chords long while long progressions were seven chords long; typical progressions had a syntactically expected chord in position five (a T chord) while atypical progressions had an unexpected chord in position five (a PD chord; see Figure 1). The unexpected chord was always a PD-function chord (using three possible exemplars of PD chords). Because we are interested in harmonic expectation regardless of the absolute pitch level, we varied the key (i.e., pitch level) of the progressions between trials by transposing them; the frequency of the audio was set at 4 different pitch levels. In the full stimulus set, because there were three different types of atypical progressions, we had three times as many typical progressions to make sure there were the same number of typical vs. atypical progressions. There were thus equal numbers of short vs. long and typical vs. atypical progressions in our set, which we repeated 5 times for a total of 240 trials (60 in each of the four conditions).

The audio was rendered using PsychoPy in a square wave tone with an onramp of 25ms, and an exponential decay of 725ms. The inter-onset interval of the chords was 750ms.

Questionnaire The questionnaire collected data about basic demographic information (age and gender) as well as 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…" (500ms), one of two images indicating the progression length by displaying either 5 or 7 notes (700–800ms, 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.

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 progression, from 50ms before it to 750ms after it. We removed the baseline from each epoch using the average from -50ms to 0ms. Finally, we removed any epochs with significant movement, sweat, and alpha artifacts by eye.

ERP Component Amplitude Measurement Component latencies were identified using standard time ranges in the

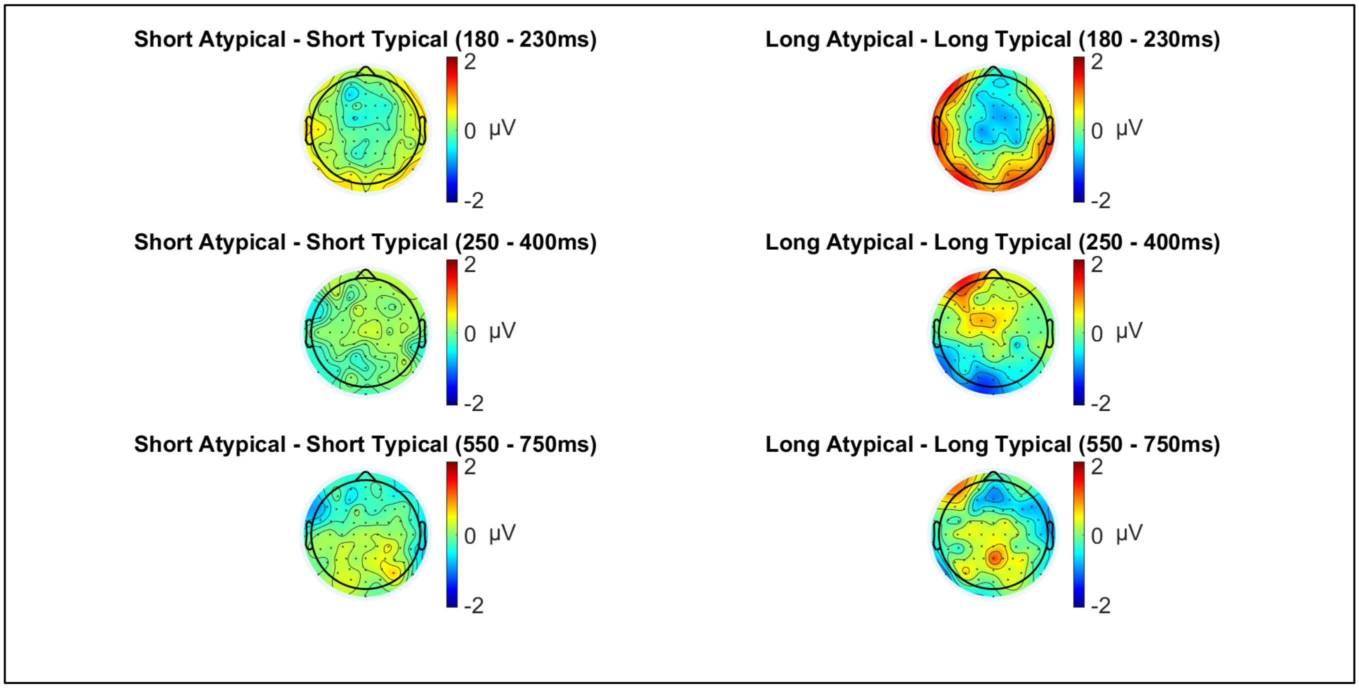

Figure 3: Topographies of amplitudes averaged over latency ranges corresponding to the E(R)AN (top), P3a (middle), and P600 (bottom). ERP components have larger amplitudes for the long conditions.

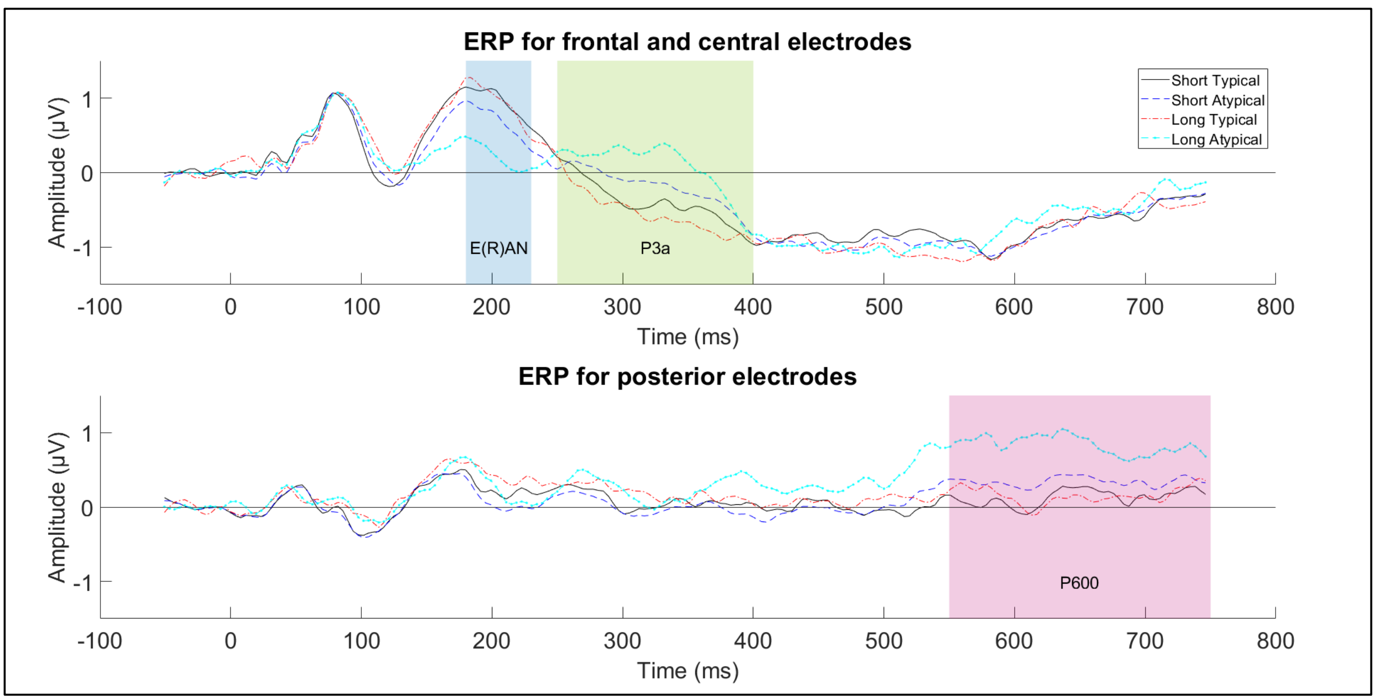

Figure 4: ERPs for the four conditions averaged over different scalp regions.

literature. For the E(R)AN component, we used 180–230ms; for the P3a, we used 250–400 ms; for the P600, we used 550–750 ms. 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.

Results

Of the 26 datasets recorded, five were not included due to excessive sweat artifacts in the signal. All further reported statistics consider only included datasets.

All participants got at least 46 out of 48 attention check questions correct ($M = 47.52$, $SD = 0.68$). The number of bad

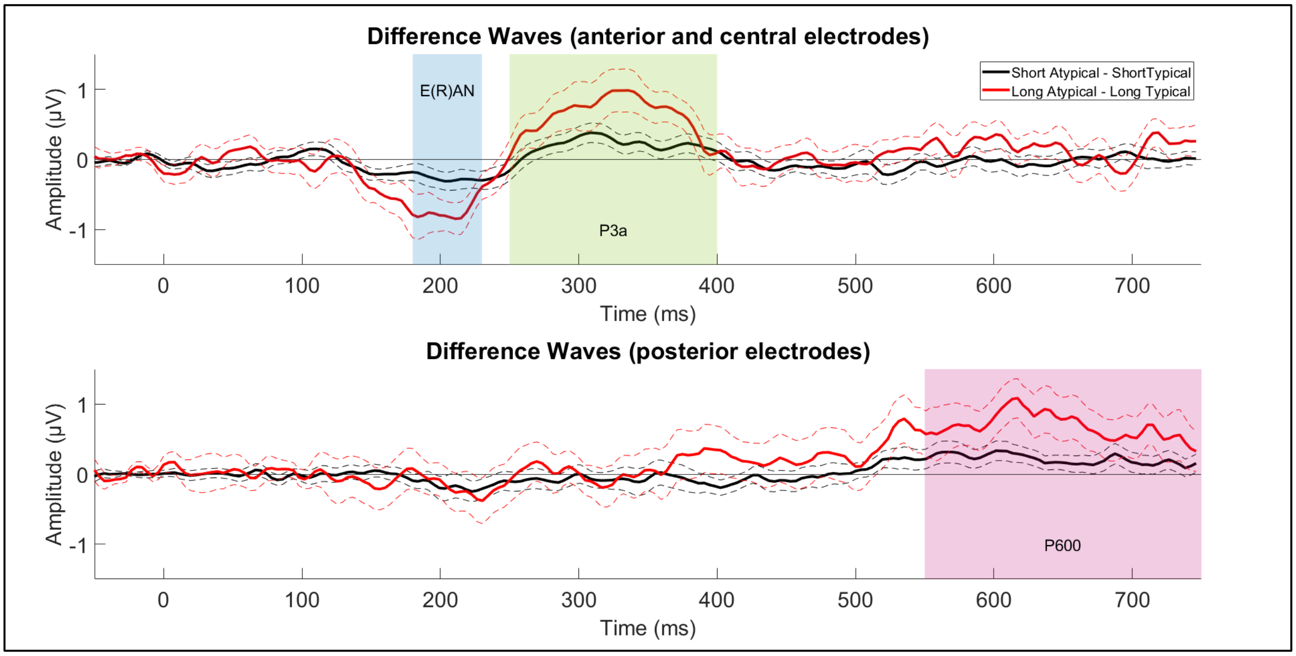

Figure 5: Difference waves (atypical minus typical) for the short and long conditions.

epochs rejected from the datasets ranged from 0–48 (out of 240 total epochs), $M = 15.43$, $SD = 11.81$.

Figure 3 displays the topographies for differences between atypical and typical conditions at latencies corresponding to the E(R)AN, P3a, and P600 components. Note the larger amplitudes for the long conditions.

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

The 2-way repeated measures ANOVA for the E(R)AN component found a significant main effect of Typicality, $F(1,20) = 13.61$, $p = .001$, $\eta^2_p = .40$, but 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 for the long progressions was much stronger.

The 2-way repeated measures ANOVA for the P3a component found a significant main effect of Typicality, $F(1,20) = 9.87$, $p = .005$, $\eta^2_p = .33$, but 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, $F(1,20) = 9.42$, $p = .006$, $\eta^2_p = .32$. 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 ($p = .081$), but the comparison between long atypical and long typical progressions was significant, $t(20) = 2.62$, $p = .017$, Cohen's $d = 0.44$. 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.

Discussion

This study investigated whether listeners are able to dynamically modulate their expectations based on knowledge of the length of a musical chord progression: in situations in which a surprising harmony occurs, we asked if listeners can adapt their expectations based on what will follow that chord.

We found that atypical (surprising) harmonies evoked the ERP components normally associated with such stimuli including the E(R)AN, P3a, and P600. Crucially, though, we found that, for the E(R)AN and P3a, the components had larger amplitude for the long chord progressions than the short ones, indicating that a listener's knowledge of the syntactic structure of the chord progressions indeed influences these ERP components. Note that nothing differs in the physical stimuli before the target chord; the short and long progressions have equivalent context (chords 1–4). The only difference between the short and long conditions is what comes *after* the fifth chord (and, the listeners knowledge of whether or not there will be subsequent context), which can only affect the ERPs through top-down processing, that is, the listener's psychological state of expectation by being cued to the syntactic structure of the stimuli.

The most general way to state the meaning of these findings is that the E(R)AN, often thought to be an index of harmonic surprisal processing, in fact depends on a listener's knowledge of local musical context. Listeners can change their stance, and hear the surprising harmony as more or less surprising depending on expectations of what will follow the chord. This provides evidence that listeners can directly modulate their harmonic expectations. Also, the E(R)AN still appears to index harmonic surprisal, but its neural generators evidently depend on a listener's local syntactic expectations as well. That is, listeners can apply neural resources to make sense of surprising harmonies (even for this early component) as opposed to automatically detecting them for further processing in a purely bottom-up process.

This top-down modulation based on listeners' expectations appears to involve multiple processing mechanisms. One key mechanism appears to be listeners' proactive adjustment of processing strategies based on contextual information (here, progression length). Rather than simply responding to surprising harmonies after they occur, listeners seem to prepare differently for these harmonies when they know more context will follow. This pattern of anticipatory processing can be interpreted within Friston's (2002) predictive coding framework. 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). In our study, participants anticipating longer progressions may have generated stronger predictions about harmonic continuity, leading to larger prediction error signals when encountering unexpected chords. Second, harmonic expectancy appears to be dynamically regulated through an interaction between long-term memory (i.e., containing statistically learned harmonic probabilities) and working memory (i.e., integrating current contextual information). When listeners know additional context will follow, they may maintain the surprising harmony in working memory differently, anticipating its potential resolution or function in the broader progression. This differs fundamentally from traditional views of the E(R)AN as an automatic detector of harmonic violations based solely on stored statistical regularities. Instead, our findings suggest that even at this early processing stage, listeners actively engage with the musical material, using their knowledge of upcoming context to shape how they process current harmonic information.

There are several limitations that warrant further exploration. Our study focused on Western tonal harmony with musically trained participants, so future research should examine whether these effects generalize to other musical systems or to listeners with less formal training. Additionally, manipulating the certainty of progression length or varying the position of the surprising harmony could further clarify how top-down expectations and bottom-up processes interact. Despite these constraints, 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.

References

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including inde- pendent component analysis. *Journal of Neuroscience Methods*, *134*(1), 9–21.

Dubiel, J. (2017). Music analysis and kinds of hearing-as. *Music Theory and Analysis (MTA)*, *4*(2), 233–242.

Friston, K. (2002). Beyond phrenology: What can neuroimaging tell us about distributed circuitry?. *Annual Review of Neuroscience*, *25*(1), 221–250.

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.

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.

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

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

Huron, D. (2008). *Sweet anticipation: Music and the psychology of expectation*. MIT Press.

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

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.

Koelsch, S., Vuust, P., & Friston, K. (2019). Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences*, *23*(1), 63–77.

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.

Lewin, D. (1986). Music theory, phenomenology, and modes of perception. *Music Perception*, *3*(4), 327–392.

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.

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.
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.
Rohrmeier, M. & Pearce, M. (2018). Musical Syntax I: Theoretical Perspectives. In R. Bader (Ed.), *Springer handbook of systematic musicology*. Springer.
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.
Vuvan, D. T., & Hughes, B. (2019). Musical style affects the strength of harmonic expectancy. *Music & Science*, *2*, 1–9.