NEUROSCIENCE
RESEARCH ARTICLE
S. A. Sauvé et al. / Neuroscience 465 (2021) 187–202

# Mapping Tonal Hierarchy in the Brain

**Sarah A. Sauvé,** [a*] **Alex Cho** [a] **and Benjamin Rich Zendel** [a,b]

[a] *Division of Community Health and Humanities, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland and Labrador A1C 5S7, Canada*

[b] *Aging Research Centre – Newfoundland and Labrador, Memorial University of Newfoundland, Grenfell Campus, Memorial University, Canada*

**Abstract—In Western tonal music, pitches are organized hierarchically based on their perceived fit in a specific tonal context. This hierarchy forms scales that are commonly used in Western tonal music. The hierarchical nature of tonal structure is well established behaviourally; however, the neural underpinnings are largely unknown. In this study, EEG data and goodness-of-fit ratings were collected from 34 participants who listened to an arpeggio followed by a probe tone, where the probe tone could be any chromatic scale degree and the context any of the major keys. Goodness-of-fit ratings corresponded to the classic tonal hierarchy. N1, P2 and the Early Right Anterior Negativity (ERAN) were significantly modulated by scale degree. Furthermore, neural marker amplitudes and latencies were significantly correlated with similar magnitude to both pitch height and goodness-of-fit ratings. This is different from the clearer divide between pitch height correlating with early neural markers (100–200 ms) and tonal hierarchy correlating with late neural markers (200–1000 ms) reported by Sankaran et al. (2020) and Quiroga-Martinez et al. (2019). Finally, individual differences were greater than any main effects detected when pooling participants and brain-behavior correlations vary widely (i.e. *r* = −0.8 to 0.8).** 

**Key words: EEG, music perception, auditory processing, implicit learning, tonal hierarchy.**

## INTRODUCTION

Tonality is a hierarchical structure of pitch in music. The tonal hierarchy reflects the perceptual organization of the relationships between the twelve pitches of the chromatic scale in Western tonal music, and is learned implicitly through passive exposure to music during development (Krumhansl and Keil, 1982; Saffran et al., 1999; Tillmann et al., 2000). Western tonal music is heavily defined by the concept of scales and keys. A scale is a collection of pitches that are a subset of the chromatic scale, the twelve tones that span an octave. The key defines the scale, where the first pitch of that scale is the most structurally stable pitch in the piece of music: it is called the *tonic*. Each pitch of the scale, or *scale degree,* has a particular relationship to the tonic. It will sound more or less perceptually 'close' to the tonic, separate from absolute frequency distance between the pitches. For example, the fifth scale degree, or dominant, is perceived as more closely related to the tonic than the second scale degree, called the supertonic, despite the second scale degree being closer in frequency to the tonic compared to the fifth scale degree. These scale pitches are also referred to as *diatonic tones*. The remaining non-scale pitches, or *chromatic tones*, in the octave from tonic to tonic are considered least related to the tonic. Chromatic tones are also considered out-of-key, while diatonic tones are considered in-key. In this paper, diatonic tones will be referred to by their music theoretical name (i.e. tonic, dominant, etc.) while the notes of the chromatic scale will be referred to by number (i.e. 1st is tonic, 8th is dominant), note name based on a C major scale (i.e. 1st is C, 8th is G) and as *chromatic scale degrees*. The relationships between the 12 pitches of the key are also related to the statistical frequency of these pitches in the Western Music repertoire: diatonic tones most closely related to the tonic occur most frequently while non-scale, chromatic tones occur most infrequently (Huron, 2006).

The tonal hierarchy was empirically defined using a probe tone method (Krumhansl et al., 1987; Krumhansl and Kessler, 1982; Krumhansl and Shepard, 1979). In this method, a tonal *context* is followed by a *probe tone*. Participants rated how well the probe tone fits with the given tonal context. The tonal context, or key, could be established using an arpeggio (the notes of the tonic triad) a chord progression, or a melody. Fig. 1 illustrates the diatonic major scale tonal hierarchy (based on data extracted from Krumhansl and Kessler, 1982), where a high rating corresponds to good fit. Perception of fit is generally

*Corresponding author.
E-mail address: sarah.sauve@mun.ca (S. A. Sauvé).
*Abbreviations:* ERAN, Early Right Anterior Negativity; ERP, event-related potential; FCz, fronto-central electrodes.

https://doi.org/10.1016/j.neuroscience.2021.03.019

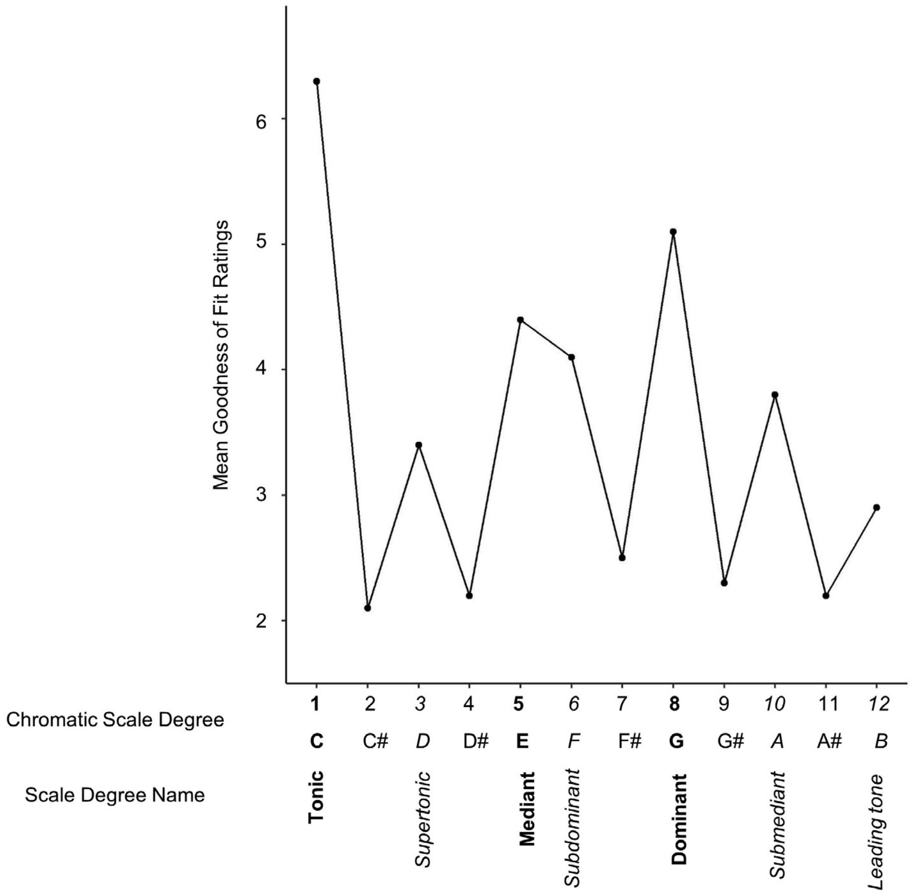


**Fig. 1.** Illustration of the tonal hierarchy for a major scale, with a goodness-of-fit rating (1 = poor fit; 7 = perfect fit) for each chromatic pitch. Chord tones (i.e. tones of the tonic triad) are indicated in bold and diatonic tones in italics.

grouped into three categories: tonic chord tones (C, E, G in Fig. 1; i.e. tones of the tonic triad), diatonic tones (D, F, A, B in Fig. 1) and chromatic tones (C#, D#, F#, G#, A# in Fig. 1). In the first group, we find the tonic, mediant and the dominant scale degrees. In the second, the remaining scale degrees, and in the third, the remaining chromatic pitches.

The relationships of fit between diatonic and chromatic tones give rise to tension and release in music. Pitches further away from the tonic, or those rated with poorer fit, are rarer and therefore less expected by the listener (Huron, 2006). The Early Right Anterior Negativity (ERAN), is an electrophysiological event-related potential (ERP) that is evoked when these musical expectancies are violated (Koelsch et al., 2000, 2001, 2002). The ERAN can be evoked by non-diatonic notes using the probe tone paradigm, when the non-diatonic notes are placed in a melody, or when the tonal context established by a sequence of chords and the probe is an unexpected chord. (Besson and Faïta, 1995; Brattico et al., 2006; Koelsch et al., 2000, 2001, 2002; Lagrois et al., 2018; Zendel et al., 2015; Zendel and Alexander, 2020). Accordingly, the ERAN is evoked by a violation of tonal expectation in a music.

The ERAN is an increase in negativity that occurs around 150–250 ms after the violating stimulus, and is best visualized using a difference wave that is calculated by subtracting the EEG response of a stimulus that violates musical syntax from one that does not. It typically peaks around 200 ms after the syntax-violating stimulus, tends to have a right, fronto-anterior scalp distribution, and can be evoked even when the listener is not attending to the musical stimuli (Koelsch, 2009b; Koelsch et al., 2001; Koelsch and Mulder, 2002; Leino et al., 2007). The ERAN is therefore influenced by the obligatory components of the auditory evoked response that occur during a similar epoch, in this case the N1 and P2 components at that peak at approximately 100 ms and 200 ms post-stimulus respectively (Crowley and Colrain, 2004; Näätänen and Picton, 1987). Accordingly, an ERAN will often manifest as an increase in N1 amplitude, or a decrease in P2 amplitude. The ERAN is, however, unlikely to be related to the same neural populations that give rise to the N1 and P2 because they have different scalp distributions, with the N1/P2 being maximal around fronto-central electrodes, and the ERAN being maximal over fronto-right electrodes (Koelsch, 2009a; Näätänen and Picton, 1987). Moreover, source analyses revealed that the N1/P2 have generators along the superior temporal plane, while the ERAN is generated by sources in inferior frontal regions (Koelsch, 2009a; Näätänen and Picton, 1987).

The goal of this study was to map the amplitude of ERP responses to the tonal hierarchy in order to better understand how tonal structure is processed implicitly during music listening. To date, most studies have focused on the ERAN, and aim to distinguish between diatonic and chromatic tones (Besson and Faïta, 1995; Halpern et al., 2017; Koelsch and Jentschke, 2010; Pearce et al., 2010), or between diatonic chords (i.e., all notes are part of the diatonic scale) and non-diatonic chords (i.e., at least one note in the chord is non-diatonic) (Koelsch et al., 2000; Koelsch and Jentschke, 2010; Koelsch and Mulder, 2002). Here, listeners were exposed to a series of tonal contexts and probe tones covering all twelve chromatic pitches in all twelve major keys while EEG was recorded. Given that the ERAN overlaps the obligatory auditory N1 response, we expect the amplitude of N1 to be related to the goodness of fit rating for that scale degree, with a larger N1 associated with poorer goodness of fit ratings (Omigie et al., 2013; Quiroga-Martinez et al., 2020). Given that the ERAN often extends past the N1 epoch, we also may observe effects on the P2 response, with smaller (i.e., more negative) responses being associated with lower goodness of fit ratings. It is also probable that both the N1 and P2 will be delayed for tones that have low goodness of fit ratings, as the ERAN will overlap these responses, and will make the peaks appear delayed. Finally, with previous evidence that musicians tend to form a sharper tonal hierarchy and therefore stronger expectations (Besson and Faïta, 1995; Krumhansl and Shepard, 1979), we will also consider the effect of musical training and any interaction it may have with the perception of tonal hierarchy, and its neural correlates.

## EXPERIMENTAL PROCEDURES

### Participants

Thirty-four participants (15 female) took part in this study and provided written informed consent in accordance with the Interdisciplinary Committee on Ethics in Human Research at Memorial University of Newfoundland. Participants had mean age 24.32 (SD = 6.45) and mean Goldsmiths Musical Sophistication Index (Gold-MSI) training subscale (Müllensiefen et al., 2014) score of 28.11 (SD = 14.05), out of a potential score of 49. All participants reported being healthy and free of any cognitive deficit and received a small cash honorarium for their participation.

### Stimuli

Each trial consisted of a 100 ms burst of white noise followed by eight tones. The first seven tones, the *context*, consisted of an arpeggio followed by a *probe tone* (Fig. 2), which was one of the twelve notes of the chromatic scale. A strong sense of key is based on a context that defines the key, which at minimum requires the tonic, median and dominant tones of the scale. In order to minimize and control the context which establishes a sense of key, the contextual arpeggio in the key of C was: C (tonic; 1st chromatic scale degree), E (mediant; 5th chromatic scale degree), G (dominant; 8th chromatic scale degree), C (+1 octave), G, E, C. Each tone was 500 ms long, the inter-onset-interval (IOI) was 700 ms and the IOI between the last context tone and the probe tone was 1200 ms. Sequences were created in MuseScore and exported to MIDI format. All stimuli can be found on the project's OSF page[‡]. The arpeggio was selected as context instead of a melody, scale or chord progression to limit any sensory confound effect to a few probe tones. A strong confounding effect of the current musical context would predict a behavioral and neural division between chord and non-chord tones, while other contexts would predict a division between in- and out-of-key tones. Using an arpeggio as context results in the tonic probe tone to be an immediate repetition of the last contextual tone, which may result in a repetition suppression bias in response to the tonic. However, this is primarily controlled for in analysis, where all tones are compared to the tonic or in the case of ERAN, where amplitude in response to the tonic is subtracted from all other tones. These analysis measures effectively apply any bias equally to all tones.

‡ https://osf.io/vksne/

### Procedure

The experiment had two parts: an EEG portion and a behavioral portion. In the EEG portion, participants were comfortably seated in a double-walled sound proof booth while set up for EEG recording. Participants first performed a series of eye movements including blinks, lateral and vertical movement for ease of artifact identification in analysis. They then watched a silent subtitled movie (Pettigrew et al., 2004) while stimuli were played via insert headphones (Etymotic E3A) at 75 dB SPL. Trials were presented in four blocks, each block containing all twelve chromatic scale degrees for all twelve major keys, for a total of 576 trials. In the behavioral portion of the experiment, participants rated how well a probe tone fit with its context on a scale of 1–7, where 1 reflects poor fit and 7, excellent fit. In the interest of time, participants rated all chromatic scale degrees for a subset of four keys for a total of 48 trials using over-ear headphones with sounds presented at a comfortable volume. The subset of keys assigned to participants was counterbalanced so that all twelve keys were rated by different participants. Finally, participants filled out the Gold-MSI musical training subscale as a measure of musical training.

### EEG data collection and preprocessing

Neuroelectric brain activity was digitized from 70 electrodes at a sampling rate of 1024 Hz with a highpass filter set at 0.1 Hz using a Biosemi ActiveTwo system (Biosemi Inc., Amsterdam, Netherlands). Sixty-four electrodes were placed on the scalp according to the standard 10–20 layout system. Additionally, six electrodes were placed bilaterally at mastoid, inferior ocular and lateral ocular sites (i.e. M1, M2, IO1, IO2, LO1, LO2).

Data was visualized, preprocessed and averages computed using BESA 6.1. Blinks and eye movements were corrected using the surrogate procedure built into BESA 6.1 (Berg and Scherg, 1994). This procedure involves recording a series of prototypical eye blinks and eye movements before the start of the experiment. A principal component analysis (PCA) of the prototypical artifacts provided a set of spatial components that best explained the eye movements. This linear combination allowed the scalp projections of the artifact components to be subtracted from the experimental ERPs to minimize ocular contamination such as blinks, as well as vertical and lateral eye movements for each individual average with minimal effects on brain activity (Berg and Scherg, 1994). After this correction, trials with greater than 120 μV of activity at electrodes not adjacent to the eyes (i.e., IO1, IO2, LO1, LO2, FP1, FP2, FPz, FP9, and FP10) were considered artifacts and excluded from further analysis. Two participants (one female) were removed due to excessively noisy data, leaving 32 for

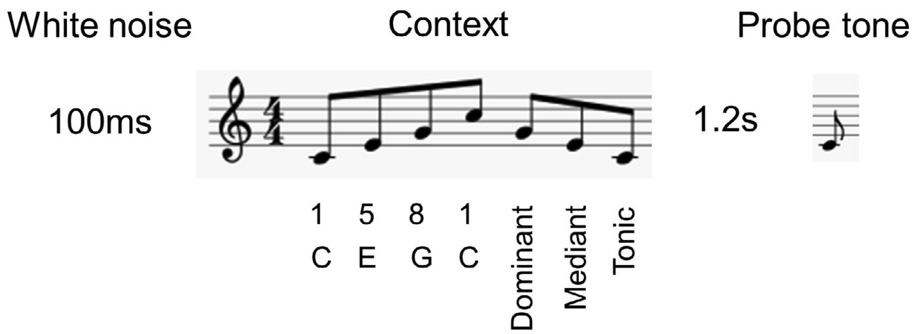


**Fig. 2.** Illustration of a single trial.

analysis. On average 92.6% of trials were accepted across all conditions and these were evenly distributed among conditions, $F$ (1, 370) = 2.80, $p > .01$.

The analysis epoch included 200 milliseconds of prestimulus activity and 1000 milliseconds of poststimulus activity. Continuous EEG was then averaged separately for each condition, into 12 different ERPs, one for each chromatic scale degree, averaged across all 12 keys. Final number of trials across all keys included in each ERP were on average 38.64, 38.58, 38.5, 38.55, 39.14, 39.23, 39.52, 39.47, 39.29, 39.17, 39.08 and 38.88 per participant for chromatic scale degrees 1–12, respectively. Each ERP was corrected using a 200 ms pre-stimulus baseline, band-pass filtered to attenuate frequencies below 0.1 Hz (forward, 6 dB/octave) and above 30 Hz (zero-phase, 12 dB/octave), and referenced to the linked mastoid.

To determine an appropriate electrode montage, a spatial independent component analysis (ICA) was calculated on all the averaged scale degrees across participants to identify the spatial distribution of the independent components that make up the averaged data. The first independent component (IC1) accounted for 32.38% of the spatial variance in the data. This component had a spatial distribution that was maximal over fronto-central electrodes (FCz) and inverted over mastoid electrodes (Fig. 3A). This spatial distribution is likely related to the N1–P2 auditory evoked response (Fig. 3A). The second and third independent components appeared to be remaining eye blink artifacts and accounted for 10.93% and 10.60% of the spatial variance respectively. The fourth independent component (IC4) accounted for 8.25% of the variance in the data. This component had a spatial distribution that was maximal over the fronto-right electrode (AF8) and was inverted over mastoid electrodes. This spatial distribution is likely related to the ERAN auditory evoked response (Fig. 3B). Accordingly, the ERP analysis focused on 2 electrode montages, one based on IC1 (i.e., F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2) and one based on IC2 (i.e., T8, FT8, FC6, F8, F6, F4, F2, Fz, F1, AF8, AF4, AFz, AF3, Fp2, Fpz, Fp1). These 2 electrode montages are highlighted in yellow in Fig. 3.

Peak amplitude and corresponding latency was extracted between 100–160 ms, corresponding to the N1 and 160–240 ms, corresponding to the P2. These data were extracted for each electrode from the montage of fronto-central electrodes identified above (F1, Fz, F2, FC1, FCz, FC2, C1, Cz, C2), selected based on signal distribution and literature precedent (e.g. Alcaini et al., 1994; Bäß et al., 2008; Campbell et al., 2007; Lange, 2011; Paiva et al., 2016). To analyze the ERAN, the mean peak amplitude of the ERAN was calculated on group average data, yielding a latency of 162 ms. To quantify the ERAN, mean amplitudes were extracted between 132–192 ms (mean ERAN latency +/− 30 ms). These data were extracted for each electrode from the montage of right-lateralized fronto-anterior electrodes identified above (T8, FT8, FC6, F8, F6, F4, F2, Fz, F1, AF8, AF4, AFz, AF3, Fp2, Fpz, Fp1), selected based on signal distribution and literature precedent (e.g., Koelsch et al., 2001, 2002; Koelsch and Jentschke, 2008).

## Analysis

First, data from both the behavioral and EEG portion of the study was analyzed in terms of chromatic scale degree. In a second analysis, the data was averaged into three conditions: chord tones (chromatic scale degrees 1, 5, 8; these notes were part of the *context*), diatonic tones (3, 6, 10, 12; these notes were part of the major scale implied by the *context*), and chromatic tones (2, 4, 7, 9, 11; these notes are the notes not in the major scale implied by the *context*). Behavioral and EEG data were analyzed using multiple linear regression modeling implemented in R (3.3.2). All categorical variables were treated as factors, where each level was compared to a base level factor. For the analysis based on chromatic scale degree the *tonic* (1st scale degree) was considered the base level; for the analysis based on scale degree group, *chord tones* was considered the base level; for electrode, *C1* was considered the base level. Models were evaluated using Pearson's correlation between the model's predictions and the data along with the correlation's 95% CIs. Statistical significance of each individual factor level for a given predictor was evaluated using 95% CIs, where an interval not including zero indicates a significant predictor.

In order to evaluate the relationship between the behavioral and the ERP data, Spearman's rank-order correlation between behavioral ratings and neural markers (N1 amplitude, N1 latency, P2 amplitude, P2 latency, ERAN amplitude) was computed for each individual and subsequently analyzed using one-sample $t$-tests to test whether mean correlations were different from zero on average. For all analyses, alpha was set at 0.01 and Bonferroni correction was applied for multiple comparisons.

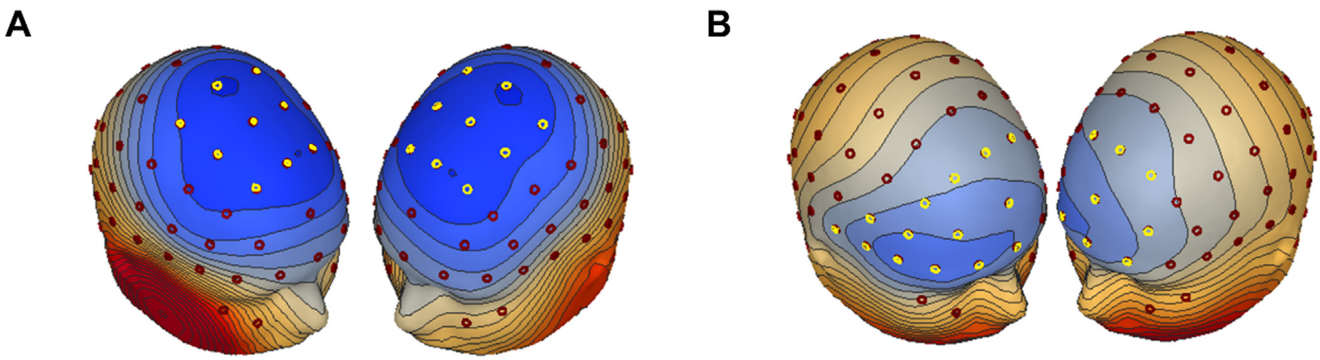


**Fig. 3.** ICA components corresponding to the N1 **(A)** and ERAN **(B)** ERPs. Electrodes highlighted in yellow were selected for data extraction.

## RESULTS

### Behavioral analysis

Probe-tone ratings are presented in dark lines in Fig. 4A. To ease comparisons to previous work using similar paradigms, probe tone ratings from Krumhansl and Kessler (1982) are presented in light gray lines. Although the mean ratings differ slightly, the overall pattern is similar. A multiple linear regression model was fitted with *chromatic scale degree* and *musical training* (as measured by Gold-MSI score) as fixed effect predictors and is summarized in Table 1. Chromatic scale degree was a significant predictor, but musical training was not. Model coefficients (see Table 2) indicate an average rating of 6.58 for the 1st scale degree, or tonic, (see Intercept), very close to the highest possible rating, and estimate the mean difference between each chromatic scale degree and the tonic. For example, the 2nd scale degree is rated 2.69 points lower than the tonic, on average. Confidence intervals all span less than one rating point, which can be considered a fair level of precision for this Likert scale and we can see that confidence intervals for each group of chromatic scale degrees generally overlap. The very small coefficient associated with musical training, a continuous variable, indicates that for every increased point on the Gold-MSI scale, mean ratings are 0.002 points lower, which we do not consider meaningful. Follow-up *t*-tests for all pairs of chromatic scale degrees are mostly significant, with Bonferroni corrected alpha $p < .0007$. Generally, this splits the ratings into above or below approximately 4.5, where chromatic scale degrees (CSDs) that fall in one group are different from CSDs in the other group, but not different from each other. The major exception to this is the tonic, which is rated sig-

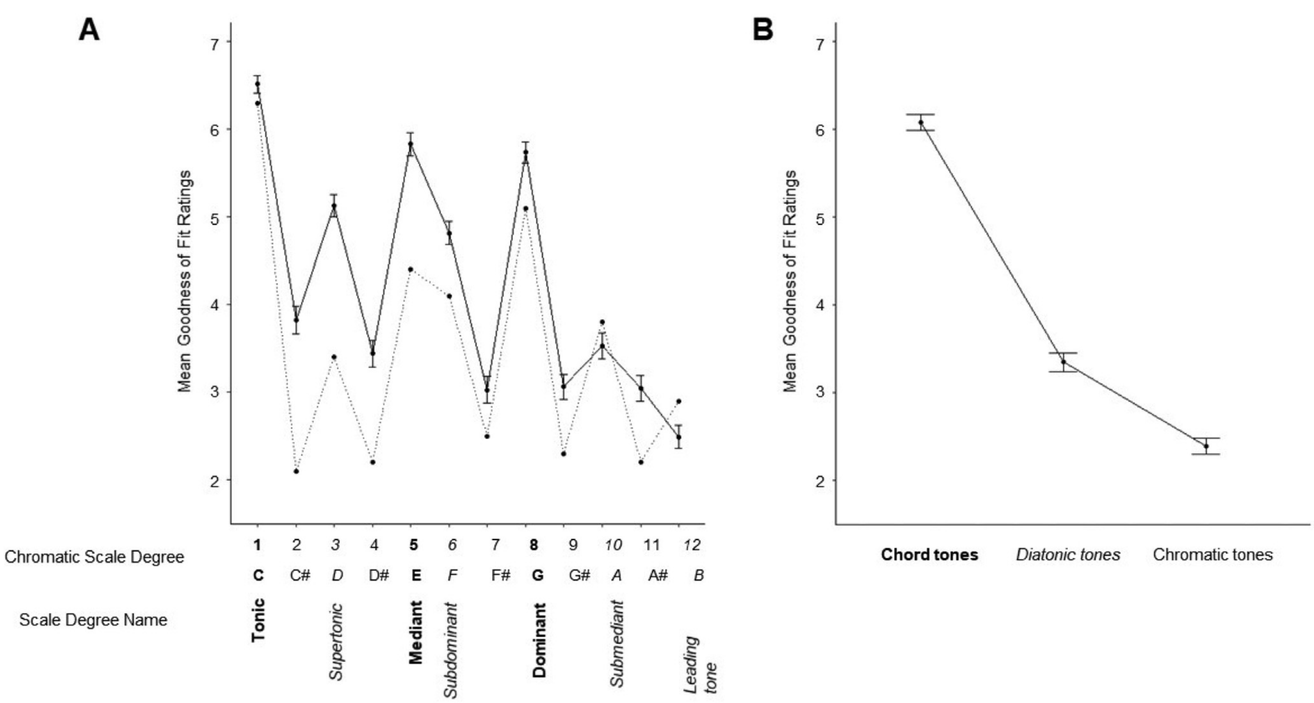


**Fig. 4. (A)** Mean ratings for each chromatic scale degree (solid line) overlaid onto the tonal hierarchy profile from Fig. 1 (dotted line). **(B)** Mean ratings for scale degree groups. Chord tones (i.e. tones of the tonic triad) in **(A)** are highlighted in bold and diatonic tones are italicized. Error bars represent standard error of the mean.

**Table 1.** Summary of the linear model predicting ratings of fit with chromatic scale degree (CSD)

| Predictor | Coefficient (Rating) | 2.5% CI | 98.5% CI | $R^2$ | Predictor | Coefficient (Rating) | 2.5% CI | 98.5% CI | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | 6.58 | 6.27 | 6.89 | 0.00 | Intercept (chord tones) | 6.09 | 5.86 | 6.33 | 0.00 |
| CSD 5 | −0.68 | −1.06 | −0.30 | 0.38 | | | | | |
| CSD 8 | −0.77 | −1.16 | −0.39 | | | | | | |
| | | | | | Diatonic tones | −2.03 | −2.55 | −1.81 | 0.28 |
| CSD 3 | −1.38 | −1.76 | −1.00 | | | | | | |
| CSD 6 | −1.69 | −2.07 | −1.31 | | | | | | |
| CSD 10 | −2.98 | −3.36 | −2.60 | | | | | | |
| CSD 12 | −4.02 | −4.40 | −3.64 | | | | | | |
| CSD 2 | −2.69 | −3.07 | −2.31 | | Chromatic tones | −2.74 | −2.95 | −2.53 | |
| CSD 4 | −3.07 | −3.45 | −2.69 | | | | | | |
| CSD 7 | −3.48 | −3.86 | −3.10 | | | | | | |
| CSD 9 | −3.44 | −3.82 | −3.06 | | | | | | |
| CSD 11 | −3.47 | −3.85 | −3.08 | | | | | | |
| Musical training | −0.002 | −0.008 | 0.002 | 0.00 | | −0.002 | −0.008 | 0.003 | 0.00 |

Note: CSD model: $r$ = 0.62 CIs [0.59, 0.65], $p$ < .01.
Group model: $r$ = 0.53 CIs [0.50, 0.56], $p$ < .01.

nificantly higher than all other CSDs. Chromatic scale degrees not different from each other include CSD 2 compared to CSDs 4 and 10; CSD 3 compared to CSD 6; CSD 4 compared to CSDs 7, 9, 10 and 11; CSD 5 compared to CSD 8; CSD 7 compared to CSDs 9, 10, 11 and 12; CSD 9 compared to CSDs 10, 11 and 12; CSD 10 compared to CSD 11; and CSD 11 compared to CSD 12 (see Fig. 3A).

For the second analysis, a linear mixed effects model was fitted with *scale degree group* and *musical training* as fixed effects, also summarized in Table 1. This model revealed a significant effect of scale degree group, with a mean rating of 6.09 for chord tones, a mean rating 2.03 points lower for diatonic tones, and 2.74 points lower for chromatic tones. Follow-up one-sided $t$-tests for all pairs of scale degree groups are significant, with Bonferroni corrected alpha $p$ < .016: $t$ (950) = 18.98 for chord tones compared to diatonic tones, $t$ (1000) = 28.17 for chord tones compared to chromatic tones and $t$ (1127) = 6.73 for diatonic tones compared to chromatic tones. Fig. 3B illustrates average ratings for each scale degree group.

**Table 2.** Summary of the linear models predicting N1 peak amplitudes

| Predictor | Coefficient (μV) | 2.5% CI | 98.5% CI | $R^2$ | Predictor | Coefficient (μV) | 2.5% CI | 98.5% CI | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | −4.92 | −5.42 | −4.41 | 0.00 | Intercept (chord tones) | −3.8 | −4.25 | −3.52 | 0.00 |
| CSD5 | 0.78 | 0.18 | 1.39 | 0.03 | | | | | |
| CSD8 | 2.31 | 1.70 | 2.91 | | | | | | |
| CSD3 | 2.06 | 1.46 | 2.67 | | Diatonic tones | 0.98 | 0.65 | 1.31 | 0.01 |
| CSD6 | 2.45 | 1.85 | 3.06 | | | | | | |
| CSD10 | 2.04 | 1.43 | 2.64 | | | | | | |
| CSD12 | 1.50 | 0.90 | 2.11 | | | | | | |
| CSD2 | 2.15 | 1.55 | 2.76 | | Chromatic tones | 0.44 | 0.12 | 0.76 | |
| CSD4 | 1.04 | 0.43 | 1.64 | | | | | | |
| CSD7 | 1.73 | 1.13 | 2.34 | | | | | | |
| CSD9 | 1.42 | 0.81 | 2.02 | | | | | | |
| CSD11 | 1.03 | 0.42 | 1.63 | | | | | | |
| Musical Training | 0.03 | 0.02 | 0.04 | 0.01 | Musical training | 0.03 | 0.02 | 0.04 | 0.01 |
| ElecC2 | 0.09 | −0.42 | 0.62 | 0.00 | ElecC2 | 0.09 | −0.43 | 0.62 | 0.00 |
| ElecCz | −0.11 | −0.63 | 0.40 | | ElecCz | −0.11 | −0.64 | 0.41 | |
| ElecF1 | 0.09 | −0.43 | 0.61 | | ElecF1 | 0.09 | −0.44 | 0.62 | |
| ElecF2 | 0.07 | −0.45 | 0.59 | | ElecF2 | 0.07 | −0.45 | 0.60 | |
| ElecFC1 | 0.06 | −0.45 | 0.59 | | ElecFC1 | 0.06 | −0.46 | 0.59 | |
| ElecFC2 | 0.005 | −0.51 | 0.52 | | ElecFC2 | 0.005 | −0.52 | 0.53 | |
| ElecFCz | −0.05 | −0.58 | 0.46 | | ElecFCz | −0.05 | −0.58 | 0.47 | |
| ElecFz | 0.002 | −0.52 | 0.52 | | ElecFz | 0.002 | −0.52 | 0.53 | |

Note: CSD model: $r$ = 0.22, CIs [0.19, 0.25], $p$ < .01
Group model: $r$ = 0.15, CIs [0.12,0.19], $p$ < .01

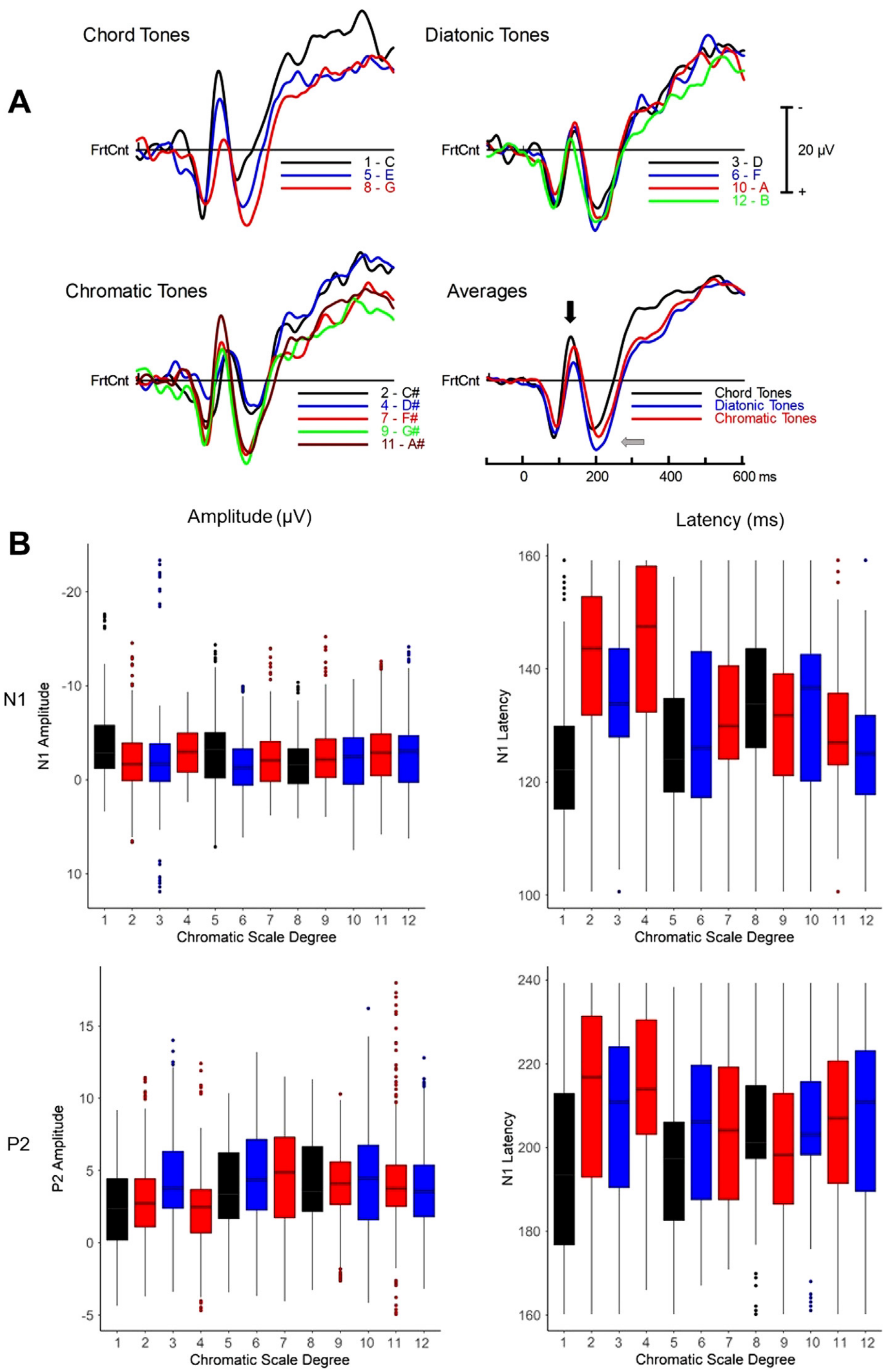


**Fig. 5.** Auditory evoked potential **(A)** averaged across participants and channels for each chromatic scale degree, grouped by chord tones (i.e. tones of the tonic triad), diatonic tones, chromatic tones as well as auditory evoked potential further averaged across groups. N1 and P2 are shown with black and gray arrows respectively in bottom right panel. **(B)** Amplitudes and latencies for the N1 and P2 are plotted for each scale degree. Color legend matches average ERP plots for each scale degree type (**(A)** – bottom right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

## Auditory evoked response analysis

In this section, we present analysis of the N1, P2 and ERAN in order to determine whether these measures are systematically modulated by chromatic scale degree. Fig. 5 presents the auditory evoked potential averaged across participants for each scale degree, illustrating clear P1, N1 and P2 peaks.

### N1

In order to evaluate the potential role of chromatic scale degree, musical background and signal distribution on the N1 amplitude and latency, linear models with *chromatic scale degree*, *musical training* and *electrode* as predictors were fitted. The models are summarized in Tables 2 and 3 for amplitude and latency, respectively. All chromatic scale degrees were significant predictors in both models, where, as the model coefficients indicate, chromatic scale degrees have smaller (less negative) amplitudes and later latencies than the tonic. Musical training was a significant predictor in both the amplitude and the latency models, where for each point scored on the Gold-MSI scale, amplitudes are 0.03 µV less negative and latencies 0.03 ms earlier with higher levels of musical expertise. Once again, though the effects are significant, it takes a difference of 20 points on the Gold-MSI scale to correspond to a difference of less than 1 µV and of 1 ms in neural response, which we do not consider meaningful. Electrode was not a significant predictor, suggesting similar signal strength across the montage.

For the second analysis, linear mixed effects models were fitted for N1 amplitude and latency with *scale degree group*, *musical training* and *electrode* as predictors. The models are summarized in Tables 2 and 3, respectively. For the amplitude model, both the diatonic and chromatic degree groups were significantly different from the chord group, where they have smaller (less negative) amplitudes than the chord group by 0.98 µV and 0.44 µV on average, respectively. For the latency model, both the diatonic and chromatic degree groups were also significantly different than the chord group,

**Table 3.** Summary of the linear models predicting N1 peak latencies

| Predictor | Coefficient (ms) | 2.5% CI | 98.5% CI | $R^2$ | Predictor | Coefficient (ms) | 2.5% CI | 98.5% CI | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | 124.52 | 122.59 | 126.46 | 0.00 | Intercept (chord tones) | 129.29 | 127.82 | 130.76 | 0.00 |
| CSD5 | 2.73 | 0.41 | 5.06 | 0.13 | | | | | |
| CSD8 | 11.56 | 9.23 | 13.89 | | | | | | |
| CSD3 | 12.72 | 10.39 | 15.05 | | Diatonic tones | 2.60 | 1.27 | 3.92 | 0.02 |
| CSD6 | 6.51 | 3.82 | 8.48 | | | | | | |
| CSD10 | 8.76 | 6.43 | 11.09 | | | | | | |
| CSD12 | 1.83 | −0.49 | 4.16 | | | | | | |
| CSD2 | 16.82 | 14.49 | 19.14 | | Chromatic tones | 6.34 | 5.07 | 7.61 | |
| CSD4 | 18.12 | 15.79 | 20.45 | | | | | | |
| CSD7 | 8.36 | 6.03 | 10.69 | | | | | | |
| CSD9 | 7.35 | 5.02 | 9.68 | | | | | | |
| CSD11 | 4.90 | 2.57 | 7.23 | | | | | | |
| Musical training | −0.03 | −0.06 | −0.00 | 0.00 | Musical training | −0.03 | −0.07 | 0.00 | 0.00 |
| ElecC2 | 0.20 | −1.80 | 2.22 | 0.00 | ElecC2 | 0.20 | −1.92 | 2.33 | 0.00 |
| ElecCz | −0.007 | −2.02 | 2.00 | | ElecCz | −0.007 | −2.13 | 2.12 | |
| ElecF1 | 1.78 | −0.23 | 3.80 | | ElecF1 | 1.78 | −0.34 | 3.91 | |
| ElecF2 | 1.96 | −0.05 | 2.97 | | ElecF2 | 1.96 | −0.16 | 4.09 | |
| ElecFC1 | 0.29 | −1.72 | 2.31 | | ElecFC1 | 0.29 | −1.83 | 2.42 | |
| ElecFC2 | 1.58 | −0.43 | 3.60 | | ElecFC2 | 1.58 | −0.54 | 3.71 | |
| ElecFCz | 0.60 | −1.40 | 2.62 | | ElecFCz | 0.60 | −1.52 | 2.73 | |
| ElecFz | 1.53 | −0.48 | 3.54 | | ElecFz | 1.53 | −0.59 | 3.65 | |

Note: CSD model: $r$ = 0.36, CIs [0.33, 0.39], $p$ < .01
Group model: $r$ = 0.18, CIs [0.14, 0.21], $p$ < .01

where latency is later by 2.60 ms and 6.34 ms on average respectively. Musical training was significant in the amplitude but not the latency model, where amplitudes are less negative and latencies earlier for higher levels of musical training, though once again not by a meaningful magnitude. Electrode was not a significant predictor.

## P2

In order to evaluate the potential effect of chromatic scale degree, musical training and signal distribution on P2 amplitude and latency, linear models were fitted with *chromatic scale degree*, *musical training* and *electrode* as predictors. The models are summarized in Tables 4

**Table 4.** Summary of the linear models predicting P2 peak amplitudes

| Predictor | Coefficient (μV) | 2.5% CI | 98.5% CI | $R^2$ | Predictor | Coefficient (μV) | 2.5% CI | 98.5% CI | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | 1.30 | 0.86 | 1.74 | 0.00 | Intercept (chord tones) | 2.39 | 2.07 | 2.71 | 0.00 |
| CSD5 | 1.37 | 0.84 | 1.89 | 0.05 | | | | | |
| CSD8 | 1.90 | 1.37 | 2.42 | | | | | | |
| CSD3 | 1.81 | 1.29 | 2.34 | | Diatonic tones | 0.69 | 0.40 | 0.98 | 0.008 |
| CSD6 | 2.16 | 1.63 | 2.68 | | | | | | |
| CSD10 | 1.82 | 1.29 | 2.34 | | | | | | |
| CSD12 | 1.35 | 0.82 | 1.87 | | | | | | |
| CSD2 | 0.31 | −0.21 | 0.83 | | Chromatic tones | 0.05 | −0.21 | 0.33 | |
| CSD4 | −0.03 | −0.56 | 0.48 | | | | | | |
| CSD7 | 2.04 | 1.52 | 2.57 | | | | | | |
| CSD9 | 1.72 | 1.19 | 2.24 | | | | | | |
| CSD11 | 1.70 | 1.18 | 2.23 | | | | | | |
| Musical training | 0.03 | 0.03 | 0.04 | 0.02 | Musical training | 0.03 | 0.03 | 0.04 | 0.02 |
| ElecC2 | 0.04 | −0.40 | 0.50 | 0.00 | ElecC2 | 0.04 | −0.41 | 0.51 | 0.00 |
| ElecCz | 0.09 | −0.36 | 0.55 | | ElecCz | 0.09 | −0.37 | 0.56 | |
| ElecF1 | −0.11 | −0.57 | 0.33 | | ElecF1 | −0.11 | −0.58 | 0.34 | |
| ElecF2 | −0.02 | −0.48 | 0.43 | | ElecF2 | −0.02 | −0.49 | 0.44 | |
| ElecFC1 | 0.10 | −0.35 | 0.56 | | ElecFC1 | 0.10 | −0.36 | 0.57 | |
| ElecFC2 | 0.21 | −0.23 | 0.67 | | ElecFC2 | 0.21 | −0.24 | 0.68 | |
| ElecFCz | 0.28 | −0.17 | 0.73 | | ElecFCz | 0.28 | −0.18 | 0.74 | |
| ElecFz | −0.07 | −0.52 | 0.38 | | ElecFz | −0.07 | −0.53 | 0.39 | |

Note: CSD model: $r$ = 0.28, CIs [0.25, 0.31], $p$ < .01
Group model: $r$ = 0.19, CIs [0.16,0.22], $p$ < .01

**Table 5.** Summary of the linear models predicting P2 peak latencies

| Predictor | Coefficient (ms) | 2.5% CI | 98.5% CI | $R^2$ | Predictor | Coefficient (ms) | 2.5% CI | 98.5% CI | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | 195.63 | 192.91 | 198.34 | 0.00 | Intercept (chord tones) | 198.31 | 196.32 | 200.29 | 0.00 |
| CSD5 | −1.05 | −4.32 | 2.21 | 0.06 | | | | | |
| CSD8 | 9.10 | 5.82 | 12.37 | | | | | | |
| CSD3 | 11.02 | 7.75 | 14.29 | | Diatonic tones | 7.72 | 5.92 | 9.52 | 0.03 |
| CSD6 | 8.74 | 5.47 | 12.01 | | | | | | |
| CSD10 | 10.19 | 6.92 | 13.47 | | | | | | |
| CSD12 | 11.65 | 8.38 | 14.92 | | | | | | |
| CSD2 | 15.75 | 12.47 | 19.02 | | Chromatic tones | 9.73 | 8.01 | 11.45 | |
| CSD4 | 19.10 | 15.83 | 22.37 | | | | | | |
| CSD7 | 9.51 | 6.24 | 12.78 | | | | | | |
| CSD9 | 7.00 | 3.73 | 10.28 | | | | | | |
| CSD11 | 10.70 | 7.43 | 13.97 | | | | | | |
| Musical training | −0.01 | −0.06 | 0.03 | 0.00 | Musical training | −0.01 | −0.06 | 0.03 | 0.00 |
| ElecC2 | 0.88 | −1.95 | 3.71 | 0.00 | ElecC2 | 0.88 | −1.99 | 3.76 | 0.00 |
| ElecCz | −0.28 | −3.11 | 2.55 | | ElecCz | −0.28 | −3.16 | 2.59 | |
| ElecF1 | 0.37 | −2.46 | 3.20 | | ElecF1 | 0.37 | −2.50 | 3.25 | |
| ElecF2 | 1.77 | −1.05 | 4.61 | | ElecF2 | 1.77 | −1.10 | 4.65 | |
| ElecFC1 | 1.48 | −1.34 | 4.32 | | ElecFC1 | 1.48 | −1.39 | 4.36 | |
| ElecFC2 | 2.09 | −0.73 | 4.93 | | ElecFC2 | 2.09 | −0.78 | 4.97 | |
| ElecFCz | 1.23 | −1.60 | 4.06 | | ElecFCz | 1.23 | −1.64 | 4.11 | |
| ElecFz | 1.00 | −1.83 | 3.83 | | ElecFz | 1.00 | −1.87 | 3.88 | |

Note: CSD model: $r$ = 0.26, CIs [0.23, 0.29], $p < .01$
Group model: $r$ = 0.19, CIs [0.16, 0.22], $p < .01$

and 5 for amplitude and latency, respectively. In the amplitude model, all but two chromatic scale degrees were significant and in the latency model, all but one, where chromatic scale degrees have larger amplitudes and later latencies than the tonic. In the case of amplitude, CSD2 and CSD4, both chromatic tones, were not different from the tonic, while for latency, CSD5, a chord tone, was not different from the tonic. Musical training was a significant predictor for amplitude but not for latency, though once again the effect of musical training on amplitude is not meaningful, equivalent to a change of 0.03 μV for each point of the scale. Electrode was not a significant predictor.

For the second analysis, linear mixed effects models were fitted for P2 amplitude and latency with *scale degree group*, *musical training* and *electrode* as predictors. The models are summarized in Tables 4 and 5, respectively. For the amplitude model, only the diatonic degree group was significantly different from the chord group, where it has smaller amplitude than the chord group by 0.69 μV on average. For the latency model, both the diatonic and chromatic degree groups were significantly different than the chord group, where latency is later by 7.72 ms and 9.73 ms on average respectively. Musical training was significant for the amplitude model but not latency, where amplitudes are less negative and latencies earlier for higher levels of musical training, though once again not by a meaningful magnitude. Electrode was not a significant predictor.

## ERAN

The ERAN was computed by subtracting the mean amplitude in response to the tonic from the mean amplitude of each chromatic scale degree between 132 and 192 ms. Fig. 6 illustrates the difference waves for each chromatic scale degree. A linear model with *chromatic scale degree*, *musical training* and *electrode* as predictors was fitted and is summarized in Table 6. All but three chromatic scale degrees, CSD2, CSD4 and CSD11, all chromatic tones, were significantly different from the tonic. The ERANs were small and varied in latency between the different scale degrees. Therefore, by using a 60 ms wide window for the mean amplitude analysis, we captured positive-going activity that occurred before and after the ERAN peak. Small, negative-going peaks are present for all scale-degrees compared to the tonic, as can be seen in Fig. 6A. Electrode was not a significant predictor.

The second analysis revealed a significant effect of *scale degree group*, where chord tones here exclude the tonic, as there is no difference wave for it. As the coefficients suggest, chord tones and diatonic tones were similar, with the mean amplitude of their difference waves approximately 1 μV smaller (i.e., less negative) than the intercept, here at −1.07 μV; chromatic tones had mean amplitudes 0.41 μV larger (i.e., more negative). Once again, musical training was a significant but not meaningful predictor for both these models. Electrode was not a significant predictor. Both these models have larger confidence intervals than the N1 models but are similar to the P2 models; however, correlation between model predictions and data is poorer than for the amplitude and latency of N1 and P2 models, as is $R^2$.

## Behavioral and neural relationship analysis

In this section, we present analysis exploring how the behavioral ratings and neural markers related to each other. To do this, Spearman's rank-order correlation coefficient was calculated between mean behavioral

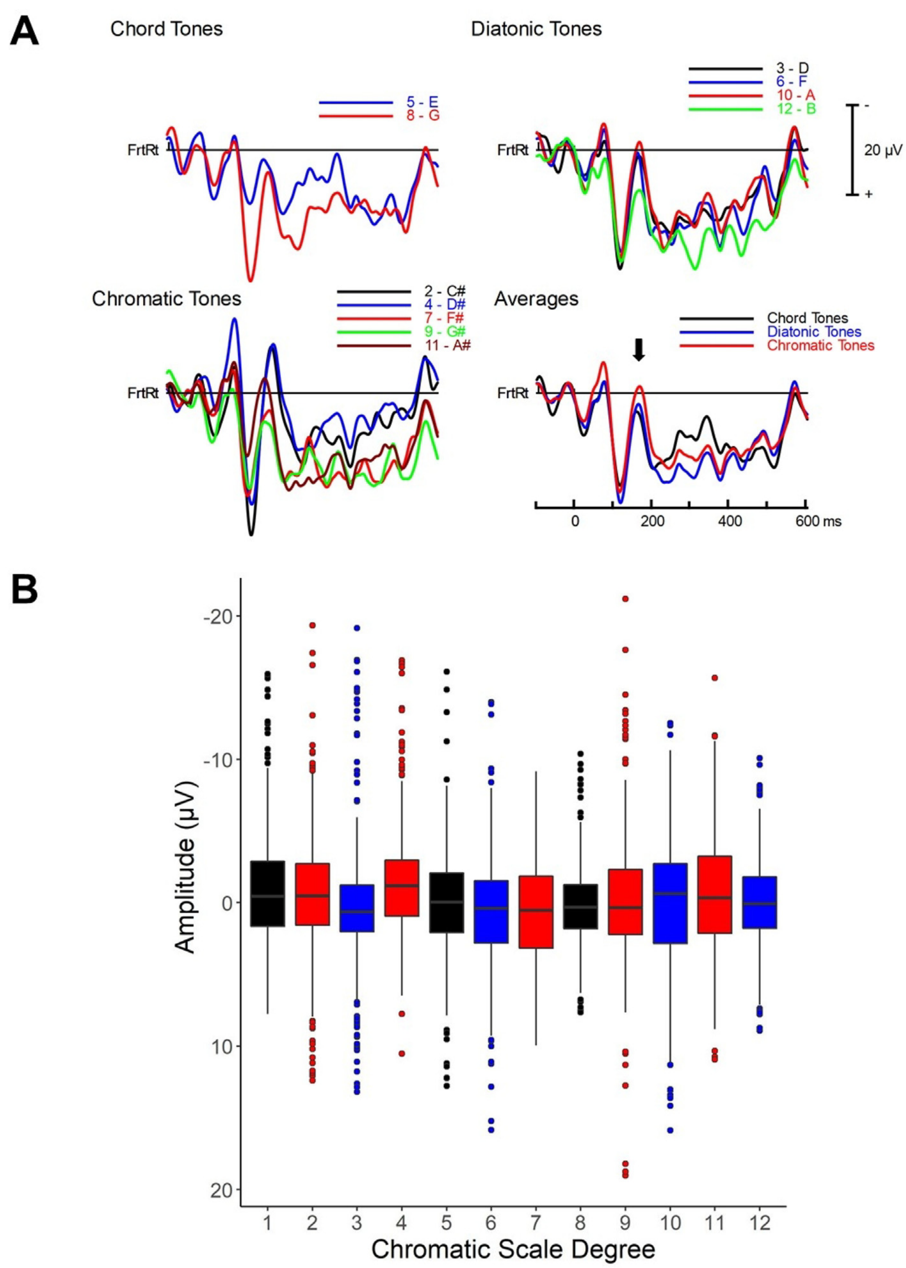


**Fig. 6.** **(A)** Difference waves averaged across each participant for each scale degree grouped by type in comparison to the tonic: chord tones (i.e. tones of the tonic triad), diatonic tones and chromatic tones along with difference waves averaged across each scale degree group. ERAN is shown with a black arrow in bottom right panel. **(B)** Amplitude of the ERAN for each scale degree. Color legend matches average ERP plots for each scale degree type (**(A)** – bottom right). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

ratings for each chromatic scale degree and mean neural marker (N1 amplitude and latency, P2 amplitude and latency and ERAN mean amplitude) for each chromatic scale degree, for each participant. This process ranks the chromatic scale degrees by value (i.e. mean rating or mean neural marker measure) and ascertains how similar the rankings are between the behavioral and neural measures.

Mean rank-order correlations across participants were 0.01 (SD = 0.25) for N1 amplitude, −0.07 (SD = 0.26) for N1 latency, −0.07 (SD = 0.26) for P2 amplitude, −0.18 (SD = 0.28) for P2 latency and 0.01 (SD = 0.25) for ERAN mean amplitude. Only mean rank-order correlations for P2 latency were significantly different from zero, $t(30) = -3.53$, with Bonferroni correction $p < .002$. Fig. 7 illustrates the relationship between each of the N1 and P2 wave amplitude or latency and ERAN amplitude with mean ratings by plotting both measures for comparison.

However, some participants have negative correlations between ratings and neural markers while some have positive correlations between the two (see Fig. 8A), resulting in mean correlations near 0. Taking the absolute value to better measure the magnitude of the effect, mean rank-order correlations were 0.21 (SD = 0.15) for N1 amplitude, 0.25 (SD = 0.17) for N1 latency, 0.19 (SD = 0.15) for P2 amplitude, 0.25 (SD = 0.20) for P2 latency and 0.19 (SD = 0.15) for ERAN mean amplitude, all significantly different from a correlation of zero, $p < .002$. To find out whether the relationship between brain and behavior is consistent within a participant, Fig. 8B plots the same information as Fig. 8A but puts the participant on the ordinal axis. Based on our hypothesis, we would expect a positive correlation between brain and behavior for N1 and ERAN amplitudes, where a higher rating corresponds to a less negative wave and a negative correlation for P2 amplitude and latency and N1 latency, where a higher rating corresponds to a smaller amplitude and earlier latency. Only one participant, participant 118 (see Fig. 8B, red box), matches these predictions exactly. Otherwise, participants tend to have mostly negative or mostly positive brain-behavior correlations, meaning that some correlations match our hypothesis while others do not and suggesting a high rate of individual differences in the neural encoding of tonal information.

To investigate any potential effects of pitch height, Spearman rank-order correlations were also computed between neural measures and scale degree number. The means of the absolute value rank-order correlations

**Table 6.** Summary of the linear models predicting ERAN mean amplitudes

| Predictor | Coefficient (μV) | 2.5% CI | 98.5% CI | $R^2$ | Predictor | Coefficient (μV) | 2.5% CI | 98.5% CI | $R^2$ |
|---|---|---|---|---|---|---|---|---|---|
| Intercept | −1.07 | −1.71 | −0.42 | 0.00 | Intercept | −1.07 | −1.71 | −0.42 | 0.00 |
| CSD5 | 0.89 | 0.33 | 1.46 | 0.01 | Chord tones | 0.99 | 0.50 | 1.48 | 0.007 |
| CSD8 | 1.08 | 0.51 | 1.64 | | | | | | |
| CSD3 | 1.31 | 0.75 | 1.87 | | Diatonic tones | 1.16 | 0.71 | 1.61 | |
| CSD6 | 1.48 | 0.92 | 2.04 | | | | | | |
| CSD10 | 0.89 | 0.33 | 1.45 | | | | | | |
| CSD12 | 0.96 | 0.40 | 1.53 | | | | | | |
| CSD2 | 0.33 | −0.23 | 0.89 | | Chromatic tones | 0.41 | −0.01 | 0.85 | |
| CSD4 | −0.46 | −1.02 | 0.10 | | | | | | |
| CSD7 | 1.23 | 0.67 | 1.80 | | | | | | |
| CSD9 | 1.70 | 0.14 | 1.27 | | | | | | |
| CSD11 | 0.27 | −0.29 | 0.83 | | | | | | |
| Musical training | 0.038 | 0.030 | 0.046 | 0.01 | Musical training | 0.01 | 0.008 | 0.02 | 0.01 |
| ElecAF3 | −0.05 | −0.70 | 0.59 | 0.00 | ElecAF3 | −0.05 | −0.71 | 0.59 | |
| ElecAF8 | 0.12 | −0.52 | 0.77 | | ElecAF8 | 0.12 | −0.52 | 0.77 | |
| ElecAFz | −0.19 | −0.84 | 0.45 | | ElecAFz | −0.19 | −0.84 | 0.45 | |
| ElecF1 | −0.008 | −0.65 | 0.64 | | ElecF1 | −0.008 | −0.66 | 0.64 | |
| ElecF2 | −0.007 | −0.65 | 0.64 | | ElecF2 | −0.007 | −0.65 | 0.64 | |
| ElecF4 | −0.15 | −0.80 | 0.49 | | ElecF4 | −0.15 | −0.80 | 0.49 | |
| ElecF6 | −0.04 | −0.69 | 0.60 | | ElecF6 | −0.04 | −0.69 | 0.60 | |
| ElecF8 | −0.34 | −0.99 | 0.30 | | ElecF8 | −0.34 | −0.99 | 0.30 | |
| ElecFC6 | −0.05 | −0.70 | 0.59 | | ElecFC6 | −0.05 | −0.70 | 0.59 | |
| ElecFp1 | 0.12 | −0.52 | 0.77 | | ElecFp1 | 0.12 | −0.53 | 0.77 | |
| ElecFp2 | 0.004 | −0.64 | 0.65 | | ElecFp2 | 0.004 | −0.64 | 0.65 | |
| ElecFpz | 0.05 | −0.59 | 0.70 | | ElecFpz | 0.05 | −0.59 | 0.70 | |
| ElecFT8 | −0.19 | −0.84 | 0.45 | | ElecFT8 | −0.19 | −0.84 | 0.46 | |
| ElecFz | −0.16 | −0.81 | 0.48 | | ElecFz | −0.16 | −0.82 | 0.48 | |
| ElecT8 | −0.29 | −0.94 | 0.35 | | ElecT8 | −0.29 | −0.94 | 0.36 | |

Note: CSD model: $r$ = 0.17, CIs [0.14, 0.19], $p < .01$
Group model: $r$ = 0.14, CIs [0.12, 0.17], $p < .01$

were 0.24 (SD = 0.20) for N1 amplitude, 0.22 (SD = 0.15) for N1 latency, 0.31 (SD = 0.23) for P2 amplitude, 0.22 (SD = 0.16) for P2 latency and 0.26 (SD = 0.17) for ERAN mean amplitude, all significantly different from zero, $p < .002$. Once again, individual differences are large and no participant matches the expected predictions, namely a negative correlation between brain and pitch height for N1 and ERAN amplitudes, where scale degrees furthest from the tonic result in the most negative amplitudes, and a positive correlation for P2 amplitude and latency and N1 latency, where a scale degree further from the tonic would correspond to a larger amplitude and later latency.

## DISCUSSION

The current study explored the relationship between tonal hierarchy and the auditory evoked response. The main finding from the study was that ERPs evoked during a passive probe-tone task were well correlated to both pitch height and behavioral goodness-of-fit ratings in a probe tone paradigm, with large individual differences in the magnitude and direction of the correlation. In order to ensure our data was consistent with previous behavioral probe tone studies, we examined how our participants rated each probe tone and found that our participants performed the task similarly to participants in previous studies. The next sections will explore these findings in more detail.

### Behavioral ratings

Behavioral ratings (see Fig. 4) matched the expected pattern very closely, where the tonic was rated with the best fit, with chord tones, diatonic scale degrees and chromatic tones rated with increasingly worse fit. One noteworthy exception to the well-established hierarchy pattern was the response to the 12th chromatic scale degree, or the leading tone. In Fig. 1, this tone was rated as having better fit than the tone immediately preceding it, the 11th chromatic scale degree; our participants rated the leading tone as having worse fit than the flat leading tone. This may be due to the high level of musicality in our participant sample, where many perceived the flat leading tone as a consonant continuation to a tonic triad. A tonic triad with a flat seventh extension is more common than a natural seventh extension and is considered a consonant, dominant-function chord.

### Neural response

The N1, the P2 and the ERAN were all impacted by the scale degree of the probe tone. We evaluated how each marker was modulated by chromatic scale degree, or chromatic scale degree grouped by type (i.e., chord, diatonic, chromatic) as well as the relationship between neural markers and behavioral ratings.

First, our results demonstrate that neural markers were significantly modulated by chromatic scale degree,

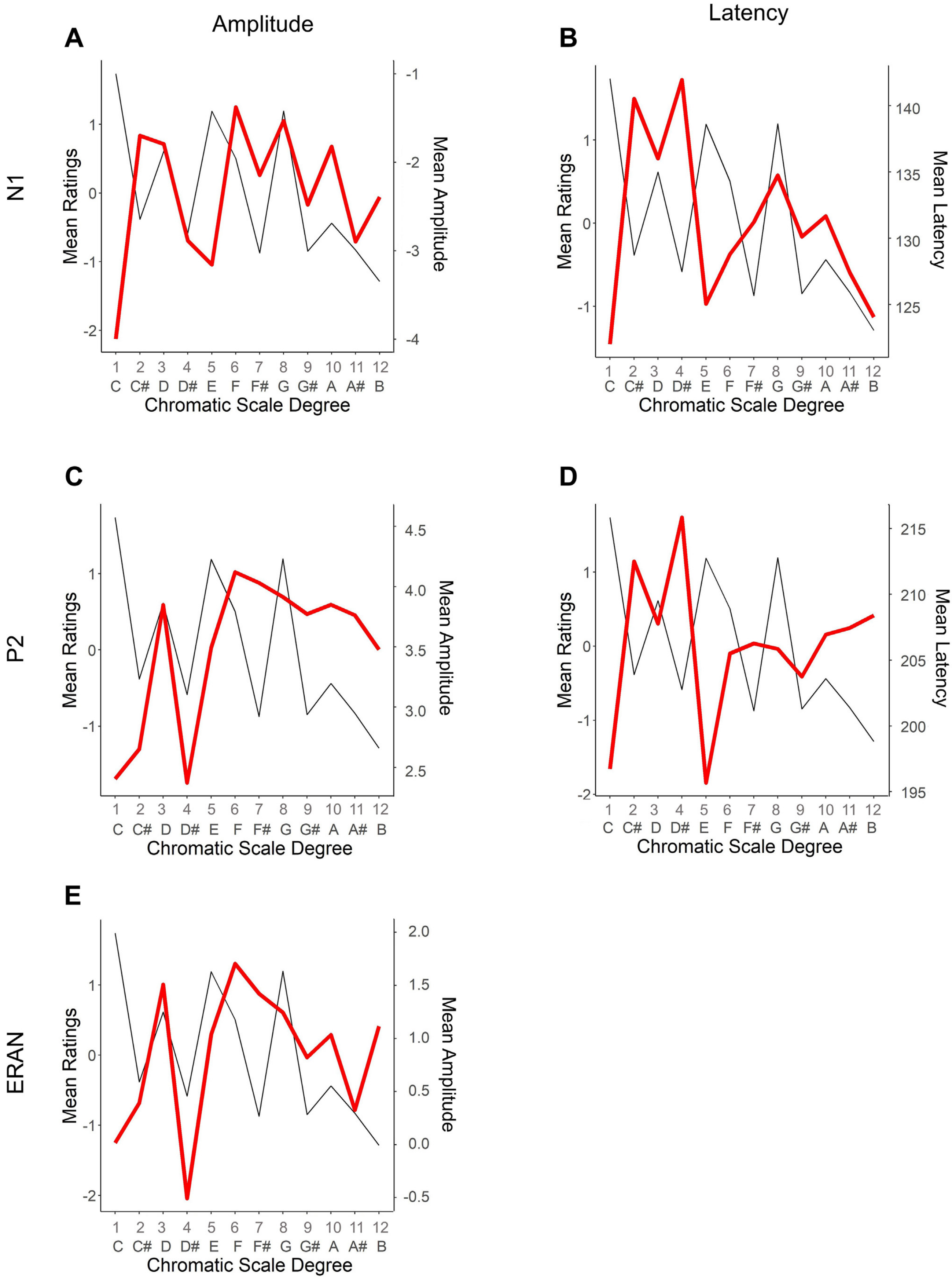


**Fig. 7.** Black lines illustrate mean ratings for each chromatic scale degree while red lines show mean amplitude and latency for each chromatic scale degree for the N1, P2 and ERAN respectively.

where amplitudes were less negative (N1, ERAN) or more positive (P2), and latencies were delayed for almost all chromatic scale degrees when compared to the tonic. This is the expected direction for the N1 and P2, but not for the ERAN, where we would expect more negative peaks for the most violating chromatic scale degrees. Furthermore, while latencies were progressively later from chord tones to diatonic tones to chromatic tones as expected, the N1 and P2 amplitudes in response to chord tones and chromatic tones were similar, with diatonic tones being the most different from the two other groups. This is contrary to what is expected as chromatic tones were perceptually most different from both chord tones and diatonic tones. While the contingent negative variation (CNV; Walter et al., 1964) is typically generated by fixed duration inter-onset intervals, as in the stimuli context in this study and may overlap with ERPs, the CNV is typically associated with the expectancy, anticipation and task preparation found in active tasks (Kononowicz and Penney, 2016; Mento, 2017; Tarantino et al., 2010; Van Rijn et al., 2011; Wilkinson and Ashby, 1974). This suggests that possible interference generated by a CNV in this study was unlikely. The presence of chord tones in the context and probe tone portions of the stimuli represent a potential low-level sensory confound; however, given that the probe tone paradigm includes all twelve chromatic tones, it is impossible to create a context that implies the correct key without these tones. Despite this limitation, studies using chords have dissociated sensory processing by presenting listeners with a probe chord that violates musical syntax, but is made up of tones that were all presented in the priming context, and compared that to a probe chord that was syntactically correct, but contained notes that were not presented in the priming context (Koelsch et al., 2007; Koelsch and Sammler, 2008). These studies found that the ERAN was evoked for the syntactic violation, providing support that the ERAN is evoked by music syntactic violations, and not low-level sensory information (Koelsch et al., 2007; Koelsch and Sammler, 2008). This suggests that it is unlikely that the ERAN responses observed in the current study were the result of acoustic repetition; however, future research should use a modified version of the probe tone paradigm to better control for sensory repetitions.

It is also surprising that both the N1 and P2 amplitudes and latencies in response to the dominant (G, SD8) were more different from the tonic than the mediant (E, SD5) and that the P2 response to the chromatic tones F#, G# and A# (SD7, SD9 and SD11) was much larger than the response to chromatic tones C# and D# (SD2 and SD4), where they would be expected to be similar. Both of these particular patterns suggest a relationship between pitch height and brain response. Such a relationship is supported by a recently published article that explored a similar question to the present study, but required participants to make a judgement about the probe tone while brain activity was recorded. Quiroga-Martinez et al. (2020) collected MEG data from participants listening to melodies composed in the classical Western idiom and investigated the N1m, MMNm and P3m response in relation to levels of information content in the melodies. Information content is the negative log probability of a note, given its context, in this case modeled by the probabilistic computational model IDyOM (Pearce, 2018). Therefore, a note that is highly probable will have low information content, and is perceived as expected, and a note that is highly improbable will have high information content and is perceived as unexpected. They found that the early responses (N1m and MMNm) were more strongly related to pitch distance, while the late response (P3m) was related to information content. Similarly, Sankaran et al. (2018), Sankaran et al. (2020) found that early responses (approx. 100–200 ms) to probe tones with a chordal context collected using MEG were more closely related to pitch distance, while late responses (+200 ms) reflected the structure tonal hierarchy.

We found a significant correlation between both brain response and scale degree and brain response and behavioral probe tone ratings for all neural measures, ranging from 0.22 to 0.31 and 0.19 to 0.25 respectively. With overlapping margins of error, these correlations were not significantly different from one another, suggesting that pitch height and the perceived tonal hierarchy structure explain early (N1, P2) neural responses similarly. This is different to Quiroga-Martinez et al. (2020) and Sankaran et al. (2018), Sankaran et al. (2020), who report a more distinct separation between the influence of pitch height and tonal hierarchy (or information content) on neural responses. It is possible that the differences the orientation of brain activity that EEG (this study) and MEG (Quiroga-Martinez et al., 2020; Sankaran et al., 2018, 2020) detect, and the difference between a passive (this study; Quiroga-Martinez et al., 2020) and an active (Sankaran et al., 2018, 2020) task may lead to these differences in results. More specifically, Quiroga-Martinez et al. (2020) employ what we would call a passive listening task, where participants watch a silent movie and ignore the stimuli. Sankaran et al. (2018), Sankaran et al. (2020) use two types of active tasks: one where participants attend to the tonal information, which we would call explicit attention, and another where participants attend to the stimuli but not the tonal information (in this case timbre), which would call implicit attention. Future work should explicitly compare active and passive listening tasks in order to better characterize the role of attention in processing of the tonal hierarchy. It is also possible that recording EEG during the behavioral rating task in our study, resulting in an active task, may have increased the correlation between brain response and behavioral ratings. However, the three studies discussed here were not yet published at the time the study reported in this paper was designed and the goal was to begin with the simplest design possible. With research into this topic still very new, future work will be needed to clarify the influences of pitch height, perceived tonality and/or information content, and attention on the neural correlates of the tonal hierarchy. It would also be interesting to investigate how musical training interacts with attention, as we have found it to have little influence in the current passive listening context. It is also important

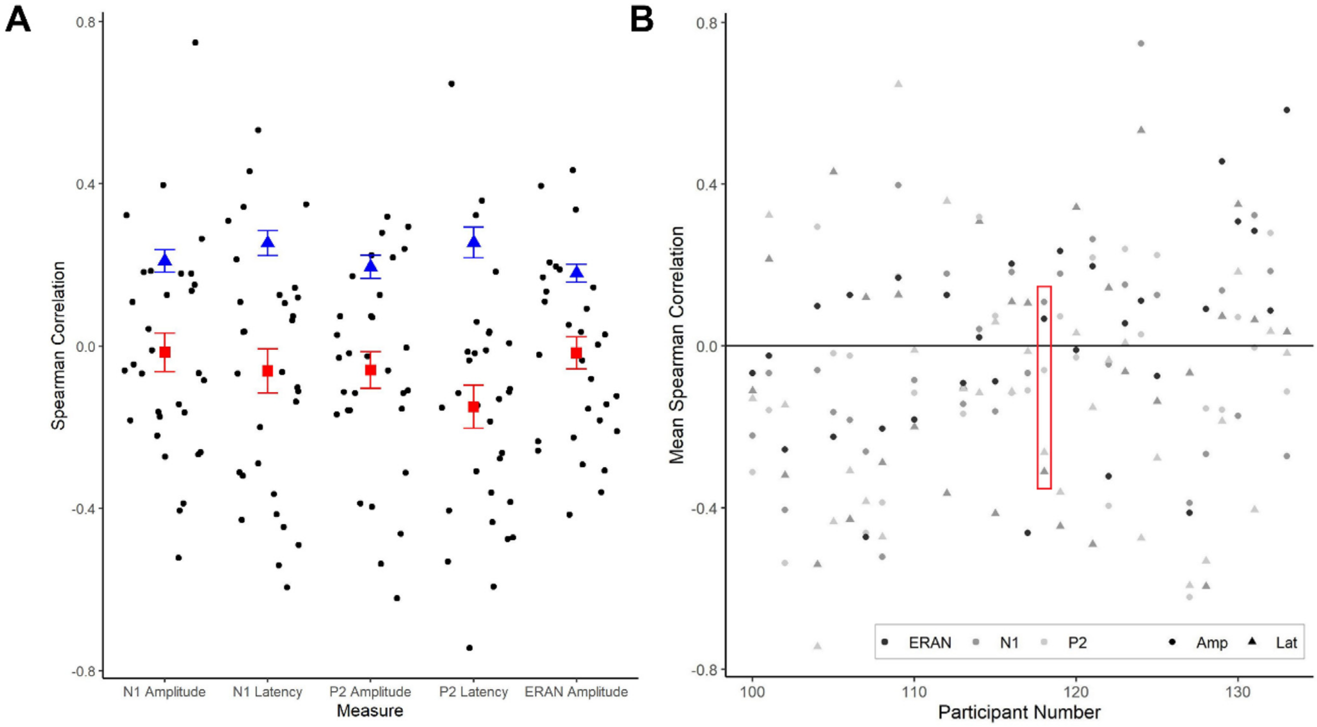


**Fig. 8.** **(A)** Spearman rank correlation between mean ratings and each neural measure; each point represents a participant, red boxes represent mean correlation for each neural measure; blue triangles represent mean magnitude of correlations for each neural measure. Error bars are standard error of the mean. **(B)** Mean Spearman rank correlation for all negative and positive wave neural measures for each participant. Red box highlights participant 118, whose brain-behavior correlations match predicted correlation directions for each neural measure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

to note that the conclusions discussed here are based on absolute value correlations, which demonstrate the strength of the relationship between brain response and behavioral response for each individual, but does not suggest the nature of that relationship. The degree of high individual differences in correlation direction and the fact that the correlations almost never match the expected direction suggest that different mechanisms may be at work and that there is still a lot we do not understand about the encoding and perception of the tonal hierarchy.

## Musicianship and individual differences

Based on the musical training sub-scale of the Goldsmiths Musical Sophistication Index (Müllensiefen et al., 2014), we had a participant group with relatively high musical sophistication, with a mean score 28.11 out of 49, or over 50%. Musical training was not a consistent nor always significant predictor of neural markers in relation to chromatic scale degree or groups. Where it was a significant predictor, the coefficient was very small, ranging from less than 0.01 to 0.05 micro-volts or milliseconds, which is meaningless in a practical sense when other coefficients are associated with variability as large as 2.91 μv and 18.65 ms. Musical training also had no significant effect on behavioral ratings, demonstrating the robust nature of the tonal hierarchy regardless of formal musical knowledge. Though not reported, models including maximally fitted random effects had $R^2$ values $> 0.95$, where near-zero variance was attributed to the chromatic scale degree or group. This demonstrates that each participant's response was more different than any averaged effect of chromatic scale degree or tone group on the investigated neural responses. The high degree of variance between participants was also evident in the wide range (approx. $-0.8$ to 0.8) and patterns of brain-behavior correlations across all neural measures.

The goal of this study was to identify the neural correlate(s) of the perceived tonal hierarchy. Given the highly robust nature of this hierarchy and the link between syntactical violation and the ERAN, finding a relationship between the tonal hierarchy and early ERP components should have been fairly straightforward. The fact that it was not is in itself an interesting finding. While we find that scale degree does modulate the N1, P2 and ERAN, the pattern of neural marker amplitudes and latencies does not have the clarity of the behavioral tonal hierarchy. On the other hand, we found that both pitch height and behavioral probe tone ratings correlate similarly with these neural markers, though the correlation is small-medium. Previous studies investigating the neural underpinnings of the tonal hierarchy have found that pitch height modulated early neural markers while the tonal hierarchy or information content modulated later neural markers (Quiroga-Martinez et al., 2020; Sankaran et al., 2018, 2020); however, these effects are also small to medium. For example, in the case of Quiroga-Martinez et al. (2020), IDyOM's interval model explained only 15% of the variance in N1m amplitude, while Sankaran et al.'s (2020) neural-model correlations between pitch height and brain data and tonal hierarchy and brain data hovered around 10–20%, respectively. These low values imply a high degree of individual differences, as we explicitly showed in our data. The high degree of individual variance combined with influence from both pitch height and tonal hierarchy could be explained by the use of multiple listening and decoding strategies unique to individuals. This, combined with individual differences in brain structure, may explain why it is difficult to observe a clear pattern with a much closer relationship to the tonal hierarchy from either EEG or MEG. Much future work is needed to better understand this surprisingly, and therefore excitingly challenging question. It would also be interesting to test the neural underpinnings of the minor tonal hierarchy, which is less stable than the major tonal hierarchy, and investigate the development of a musical hierarchy using a novel musical system by collecting goodness-of-fit ratings and neural data at regular time points in the learning of the novel system.

## CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

**Sarah A. Sauvé:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Visualization, Writing - original draft, Writing - review & editing. **Alex Cho:** Investigation, Writing - review & editing. **Benjamin Rich Zendel:** Conceptualization, Funding acquisition, Resources, Supervision, Writing - review & editing.

## ACKNOWLEDGEMENTS

The authors acknowledge that this work was undertaken on the unceded, unsurrendered ancestral Lands of the Mi'kmaq and Beothuk. We would also like to acknowledge the Inuit of Nunatsiavut and NunatuKavut and the Innu of Nitassinan, and their ancestors, as the original peoples of Labrador.

## FUNDING

This work was supported by Benjamin Rich Zendel's Canada Research Chair.

## REFERENCES

Alcaini M, Giard M-H, Echallier J-F, Pernier J (1994) Selective auditory attention effects in tonotopically organized cortical areas: A topographic ERP study. Hum Brain Mapp 2(3):159–169.

Berg P, Scherg M (1994) A multiple source approach to the correction of eye artifacts. Electroencephalogr Clin Neurophysiol 90 (3):229–241.

Besson M, Faïta F (1995) An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. J Exp Psychol Hum Percept Perform 21(6):1278.

Brattico E, Tervaniemi M, Näätänen R, Peretz I (2006) Musical scale properties are automatically processed in the human auditory cortex. Brain Res 1117(1):162–174. https://doi.org/10.1016/j.brainres.2006.08.023.

Campbell T, Winkler I, Kujala T (2007) N1 and the mismatch negativity are spatiotemporally distinct ERP components: Disruption of immediate memory by auditory distraction can be related to N1. Psychophysiology 44(4):530–540.

Crowley KE, Colrain IM (2004) A review of the evidence for P2 being an independent component process: Age, sleep and modality. Clin Neurophysiol 115(4):732–744.

Halpern AR, Zioga I, Shankleman M, Lindsen J, Pearce MT, Bhattarcharya J (2017) That note sounds wrong! Age-related effects in processing of musical expectation. Brain Cogn 113:1–9.

Huron, D. (2006). Sweet anticipation: Music and the psychology of expectation (2006-08683-000). The MIT Press.

Koelsch S, Gunter T, Friederici AD, Schröger E (2000) Brain indices of music processing: "nonmusicians" are musical. J Cognit Neurosci 12(3):520–541.

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

Koelsch, Stefan. (2009b). Neural substrates of processing syntax and semantics in music. In Music that works (pp. 143–153). Springer.

Koelsch Stefan, Gunter TC, Schröger E, Tervaniemi M, Sammler D, Friederici AD (2001) Differentiating ERAN and MMN: an ERP study. NeuroReport Neurosci Res 12(7):1385–1389.

Koelsch S, Jentschke S (2008) Short-term effects of processing musical syntax: an ERP study. Brain Res 1212:55–62.

Koelsch S, Jentschke S (2010) Differences in electric brain responses to melodies and chords. J Cognit Neurosci 22 (10):2251–2262.

Koelsch Stefan, Jentschke S, Sammler D, Mietchen D (2007) Untangling syntactic and sensory processing: An ERP study of music perception. Psychophysiology 44(3):476–490.

Koelsch S, Mulder J (2002) Electric brain responses to inappropriate harmonies during listening to expressive music. Clin Neurophysiol 113(6):862–869.

Koelsch S, Sammler D (2008) Cognitive components of regularity processing in the auditory domain. PLoS ONE 3(7).

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.

Kononowicz TW, Penney TB (2016) The contingent negative variation (CNV): Timing isn't everything. Curr Opin Behav Sci 8:231–237. https://doi.org/10.1016/j.cobeha.2016.02.022.

Krumhansl CL, Keil FC (1982) Acquisition of the hierarchy of tonal functions in music. Mem Cogn 10(3):243–251.

Krumhansl CL, Kessler EJ (1982) Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. Psychol Rev 89(4):334.

Krumhansl CL, Sandell GJ, Sergeant DC (1987) The perception of tone hierarchies and mirror forms in twelve-tone serial music. Music Percept 5(1):31–78. https://doi.org/10.2307/40285385.

Krumhansl CL, Shepard RN (1979) Quantification of the hierarchy of tonal functions within a diatonic context. J Exp Psychol Hum Percept Perform 5(4):579.

Lagrois M-É, Peretz I, Zendel BR (2018) Neurophysiological and behavioral differences between older and younger adults when processing violations of tonal structure in music. Front Neurosci 12:54.

Lange K (2011) The reduced N1 to self-generated tones: An effect of temporal predictability? Psychophysiology 48(8):1088–1095. https://doi.org/10.1111/j.1469-8986.2010.01174.x.

Leino S, Brattico E, Tervaniemi M, Vuust P (2007) Representation of harmony rules in the human brain: further evidence from event-related potentials. Brain Res 1142:169–177.

Mento G (2017) The role of the P3 and CNV components in voluntary and automatic temporal orienting: A high spatial-resolution ERP study. Neuropsychologia 107:31–40. https://doi.org/10.1016/j.neuropsychologia.2017.10.037.

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.0089642.

Näätänen R, Picton T (1987) The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. Psychophysiology 24(4):375–425.

Omigie D, Pearce MT, Williamson VJ, Stewart L (2013) Electrophysiological correlates of melodic processing in congenital amusia. Neuropsychologia 51(9):1749–1762. https://doi.org/10.1016/j.neuropsychologia.2013.05.010.

Paiva TO, Almeida PR, Ferreira-Santos F, Vieira JB, Silveira C, Chaves PL, Barbosa F, Marques-Teixeira J (2016) Similar sound intensity dependence of the N1 and P2 components of the auditory ERP: Averaged and single trial evidence. Clin Neurophysiol 127 (1):499–508.

Pearce MT (2018) Statistical learning and probabilistic prediction in music cognition: Mechanisms of stylistic enculturation. Ann N Y Acad Sci 1423(1):378–395. https://doi.org/10.1111/nyas.13654.

Pearce MT, Ruiz MH, Kapasi S, Wiggins GA, Bhattacharya J (2010) Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. NeuroImage 50(1):302–313. https://doi.org/10.1016/j.neuroimage.2009.12.019.

Pettigrew CM, Murdoch BE, Ponton CW, Kei J, Chenery HJ, Alku P (2004) Subtitled videos and mismatch negativity (MMN) investigations of spoken word processing. J Am Acad Audiol 15 (7):469–485.

Quiroga-Martinez DR, Hansen NC, Højlund A, Pearce M, Brattico E, Vuust P (2020) Decomposing neural responses to melodic surprise in musicians and non-musicians: Evidence for a hierarchy of predictions in the auditory system. NeuroImage 215:116816. https://doi.org/10.1016/j.neuroimage.2020.116816.

Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cognition 70(1):27–52. https://doi.org/10.1016/S0010-0277(98)00075-4.

Sankaran N, Carlson TA, Thompson WF (2020) The rapid emergence of musical pitch structure in human cortex. J Neurosci.

Sankaran N, Thompson WF, Carlile S, Carlson TA (2018) Decoding the dynamic representation of musical pitch from human brain activity. Sci Rep 8(1):839.

Tarantino V, Ehlis A-C, Baehne C, Boreatti-Huemmer A, Jacob C, Bisiacchi P, Fallgatter AJ (2010) The time course of temporal discrimination: An ERP study. Clin Neurophysiol 121(1), 43–52. 10.1016/j.clinph.2009.09.014.

Tillmann B, Bharucha JJ, Bigand E (2000) Implicit learning of tonality: A self-organizing approach. Psychol Rev 107(4):885.

Van Rijn H, Kononowicz TW, Meck WH, Ng KK, Penney TB (2011) Contingent negative variation and its relation to time estimation: A theoretical evaluation. Front Integr Neurosci 5. https://doi.org/10.3389/fnint.2011.00091.

Walter W, Cooper R, Aldridge VJ, McCallum WC, Winter AL (1964) Contingent negative variation: An electric sign of sensori-motor association and expectancy in the human brain. Nature 203:380–384.

Wilkinson RT, Ashby SM (1974) Selective attention, contingent negative variation and the evoked potential. Biol Psychol 1 (3):167–179. https://doi.org/10.1016/0301-0511(74)90009-X.

Zendel BR, Alexander EJ (2020) Autodidacticism and Music: Do self-taught musicians exhibit the same auditory processing advantages as formally trained musicians? Front Neurosci.

Zendel BR, Tremblay C-D, Belleville S, Peretz I (2015) The impact of musicianship on the cortical mechanisms related to separating speech from background noise. J Cognit Neurosci. Available from: https://www.mitpressjournals.org/doi/full/10.1162/jocn_a_00758.

*(Received 29 January 2021, Accepted 16 March 2021)*
*(Available online 25 March 2021)*