1 Major classification: Biological Sciences; Minor classification: Psychological & Cognitive

2 Sciences

3

4 **The dynamic emergence of musical pitch structure in human cortex**

5 N. Sankaran[1,2] *, T.A. Carlson[2,4], W.F. Thompson[2,3]

6

7 [1] Auditory Neuroscience Laboratory, School of Medical Sciences, University of Sydney,

8 Sydney, New South Wales 2006, Australia.

9 [2] ARC Centre of Excellence in Cognition and its Disorders, Macquarie University, Sydney,

10 New South Wales 2109, Australia.

11 [3] Department of Psychology, Macquarie University, Sydney, New South Wales 2109,

12 Australia.

13 [4] School of Psychology, University of Sydney, Sydney, New South Wales 2006, Australia.

14 * Corresponding author

15 Current address: Department of Neurological Surgery and Center for Integrative

16 Neuroscience, University of California, San Francisco, 675 Nelson rising Lane, San

17 Francisco, CA 94158, USA.

18 Email: narayan.sankaran@ucsf.edu

19 Number: 415-502-7346

20 ORCID ID: https://orcid.org/0000-0002-6788-3878

21

23

24   **Abstract:**

25   Tonal music the world over is characterized by a hierarchical structuring of pitch, whereby

26   certain tones appear stable and others unstable within their musical context. Despite its

27   prevalence, the cortical mechanisms supporting such a percept remain poorly understood.

28   The current study probed the neural processing dynamics underlying the representation of

29   pitch in Western Tonal Music. Listeners were presented with tones comprising all twelve

30   pitch-classes embedded within a musical context whilst having their

31   magnetoencephalographic (MEG) activity recorded. Using multivariate pattern analysis

32   (MVPA), decoders attempted to classify the identity of tones from their corresponding MEG

33   activity at each peristimulus time sample, providing a dynamic measure of their cortical

34   dissimilarity. Time-evolving dissimilarities between tones were then compared with the

35   predictions of several acoustic and perceptual models. Following tone onset, we observed a

36   temporal evolution in the brain's representation. Dissimilarities between tones initially

37   reflected their fundamental frequency separation, but beyond 200 ms reflected their status

38   within the *tonal hierarchy* of perceived stability. Furthermore, when the dissimilarities

39   corresponding to this latter period were transposed into different keys, cortical relations

40   between keys correlated with the well-known *circle of fifths*. Convergent with fundamental

41   principles of music-theory and perception, current results detail the dynamics with which the

42   complex perceptual structure of Western tonal music emerges in human cortex within the

43   timescale of an individual tone.

44

45

46

47

48

49   **Significance statement:**

50   In music, pitch is organized along a hierarchy of perceived stability. Applying stimulus

51   decoding techniques to the Magnetoencephalographic activity of subjects during music-

52   listening, we examined the structure of this hierarchy in cortex and the dynamics with which

53   it emerges at the timescale of an individual tone. Following its onset, we observed a temporal

54   evolution in the brain's representation of a tone. Activity initially reflected its pitch-value

55   (fundamental frequency) before reflecting its status within the *tonal hierarchy* of perceived

56   stability. 'Transposing' this later period of activity into different musical keys, we found that

57   inter-key distances reflected the well-known *circle of fifths*. Our results provide a link

58   between the complex perceptual structure of tonal music and its dynamic emergence in

59   cortex.

60

61

62

63

64

65

66

67

68

69

70

71

72

73

Introduction

In musical systems throughout the world, pitch is organized hierarchically (1). Depending on the prevailing *key* or *tonality* of a musical passage, certain pitch-classes occur more frequently and occupy positions of melodic, harmonic and rhythmic prominence (2). Perception mirrors this compositional hierarchy, whereby those privileged pitch-classes also have greater *stability* (3, 4). For example, within the Western key of C major, the first scale degree (C) is maximally stable and therefore heads the hierarchy. This is followed by the fifth and third scale degrees (G and E respectively), the other scale tones (D, F, G, A, B), and finally the non-scale or "out-of-key" tones (C#, D#, F#, G#, A#). We refer to this collective structure as the *standard tonal hierarchy* (STH; figure 1A).
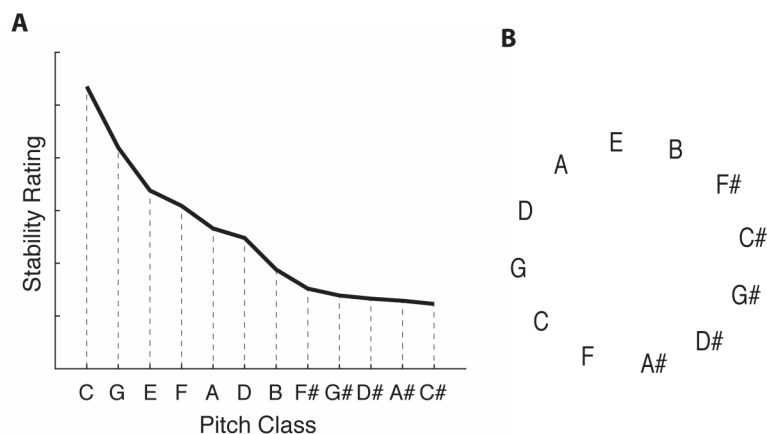
Figure 1. Perceptual descriptions of tonal structure. (A) The standard tonal hierarchy based on listener's ratings of perceived stability reported in Krumhansl & Kessler (1982). (B) The "circle of fifths" conveying the relatedness between the different major musical keys.

Despite functioning as the principle organizing schema of Western Tonal Music, the neural substrates supporting the STH remain unknown. After core auditory areas extract basic

4

94   frequency information from an acoustic signal, a representation of complex pitch is thought

95   to emerge in lateral auditory regions (5-9). How does this isolated sensory representation then

96   acquire the perceived attributes of musical pitch? The surrounding musical context must be

97   integrated, and cortical populations reflecting a prior knowledge of Western tonal structure

98   must be recruited. Both lesion and neuroimaging studies have identified regions implicated in

99   the processing of both melodic (10) and harmonic (11-13) structure, while

100   electrophysiological research has identified cortical response components sensitive to the

101   hierarchical status of evoking tones (14-15). More recently, Sankaran et al. (2018) (16)

102   showed that, independent from acoustics, the tonal *class* of pitches can be decoded from their

103   multivariate patterns of Magnetoencephalographic (MEG) activity, suggesting that the

104   perceptual structure of musical pitch may be directly recoverable from cortical activity.

105   Despite these advances, empirical work is yet to map the neural representational space of

106   musical pitch and explicitly test the predictions of specific perceptual and music-theoretic

107   models. The current study therefore evaluated two major questions: Firstly, do cortical

108   populations encode musical pitch in a manner that precipitates the organization of the STH?

109   Secondly, what are the temporal dynamics with which afferent sensory representations of

110   pitch interface with high-level tonal-schematic ones?

111

112   To probe these questions, we recorded the MEG activity of subjects listening to each pitch-

113   class presented within a tonal musical context. We used Multivariate Pattern Analysis

114   (MVPA) (17) to decode the identity of tones from their corresponding MEG activity. Within

115   this framework, the accuracy with which classifiers can discriminate between the

116   spatiotemporal response patterns elicited by two different tones provides an intuitive measure

117   of their dissimilarity in cortex. As MEG responses were sufficiently time-resolved,

118   classification was applied using a sliding time window, enabling us to track the temporal

119   dynamics of the neural distinctions between tones. Finally, comparing the time-varying MEG

120   dissimilarities with the predictions of relevant acoustic and perceptual models of pitch, we

121   evaluated how the evolving cortical structure between tones relates to stimulus-driven

122   features and the perceptual organization of the STH.

123

124   In addition to examining the brain's representation of pitch within one key, we also measured

125   the relationship between different major keys in cortex. This was motivated by the musical

126   practice of *modulation*, in which a passage shifts from one key to another. In music theory,

127   inter-key distances are described by the *circle of fifths* (figure 1B). In this arrangement, keys

128   separated by intervals of a fifth are closest, and the pattern of relatedness folds back on itself

129   to form a closed circle. Perceptual research has shown that these key-relations emerge when

130   correlating the STH of different keys with one another (4), suggesting that the cognitive basis

131   of tonality resides in the "scaffold" of individual pitch relationships rather than the general

132   accumulation of information across a tonal passage. While prior research has investigated

133   key-relationships using fMRI (18), the relatively poor temporal resolution prohibits an

134   understanding of the neural mechanisms underlying the emergence of tonal structure at the

135   timescale of an individual tone. We therefore derived a neural representation of key-distances

136   using the measured MEG distinctions between tones. Remarkably, the extent to which two

137   keys were related in cortex was predicted by the circle of fifths. Thus, convergent with

138   fundamental principles in both music-theory and perception, current results provide a

139   neuroscientific conceptualization of how complex tonal structure emerges from individual

140   pitch-relationships within music.

141

142   **Results and Discussion**

143   Results are derived from MEG recordings during the presentation of twelve different "probe-

144   tones" that spanned all pitch-classes within an octave range following a C major context (see

145   methods). Discriminant classifiers attempted to decode the MEG activity of two different

146   tones at each time-point in the neural epoch (from -100 ms to 1000 ms relative to onset), and

147   the resulting curve of time-varying accuracy provided a dynamic estimate of the dissimilarity

148   in their neuronal population codes. Applying this classification procedure to every pairwise

149   combination of the twelve different tones, we characterized the dynamic representational

150   structure of musical pitch in cortex.

151

152   **Representational structure of musical pitch in cortex**

153   To examine the dynamics of stimulus-specific information in cortex, we first assessed the

154   average decoding performance when classifying all pairwise combinations of tones (figure

155   2A). As expected, average accuracy was at chance (50%) prior to the onset of tones (t=0) as

156   stimulus-related information was yet to activate cortex. Following onset, neural distinctions

157   between tones first emerged at 100 ms. Distinctions were maximal at 250 ms and remained

158   above chance for the full extent of the neural epoch.

159

160   Next, we studied the dissimilarity between specific tones whose acoustic or perceptual

161   properties generated explicit predictions regarding their representational distance. Firstly, as

162   tones acoustically differed from one another, we reasoned that their distinctions in cortex

163   may be commensurate with their fundamental frequency (f_0) separation, which we term *pitch-

164   height* (PH). Decoding performance was therefore examined for pairwise combinations of

165   tones grouped based on whether their PH difference was small (1–4 semitones), medium (5–7

166   semitones), or large (8–11 semitones). A period from approximately 100 to 250 ms was

167   found in which the above hypothesis held true (figure 2B). For example, cortical distinctions

7

168   between tones that had large PH differences (blue curve) significantly exceeded those

169   between tones that had small PH separation (red curve). Secondly, in addition to acoustic

170   differences, tones differed in their perceived stability given the preceding musical context.

171   We therefore hypothesized that distinctions in their cortical encoding may honor their

172   perceptual differences, embodied by the *Standard Tonal Hierarchy* (STH) of stability. If so,

173   MEG decoding performance would be greatest for tones located at opposite ends of the

174   hierarchy, and poorest for tones that are hierarchically equal. In general, results confirmed

175   this hypothesis (figure 2C). MEG responses to the most stable tone [C] were highly distinct

176   from those of the unstable tones [F#, G#, D#, A#, C#] (green curve), but less discriminable

177   from those of the second and third most stable tones [G and E respectively] (orange curve).

178   Additionally, classifiers performed poorly when attempting to distinguish the neural activity

179   of unstable tones from one another (purple curve). These results suggest that the extent to

180   which the cortical activity elicited by two tones differ corresponds to the difference in their

181   position within the STH. Unlike the earlier neural distinctions based on pitch-height, the

182   correspondence between decoding accuracy and hierarchical distance only emerged

183   approximately 200 ms after onset and persisted throughout the duration of the neural epoch.

184
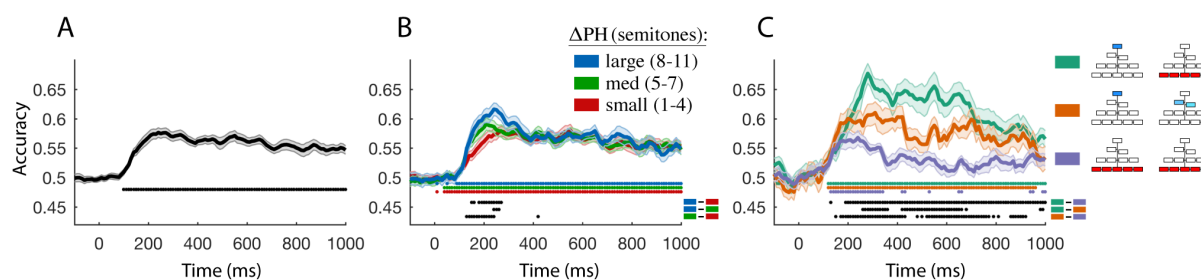
186   **Figure 2. Temporal decoding of tones from evoked MEG responses.** The time axis in all

187   plots are aligned to onset of tones. (A) Average classification accuracy for decoding all

188   pairwise combinations of the twelve tones. (B) Average classification accuracy when

189   decoding tone-pairs grouped based on their pitch-height separation: large (8-11 semitones;

8

190   blue), medium (5-7 semitones; green) and small (1-4 semitones; red). (C) Classification of

191   tone-pairs grouped based on their difference in the hierarchy of perceived stability: large

192   difference (green), medium difference (orange), and little-to-no difference (purple). Colored

193   boxes in the schematic legend specify the hierarchical position of tones being decoded for

194   each curve, with blue and red boxes indicating stable and unstable tones respectively. Results

195   in plots B-C are averaged across all appropriate pairwise combinations of tones. Colored

196   markers underneath curves in B-C indicate timepoints when decoding performance differs

197   significantly from chance levels ($p<0.05$; Wilcoxon sign-rank tests, FDR corrected). Black

198   markers indicate timepoints during which two decoding curves, specified by the bottom-right

199   colored boxes, are significantly different from one another. Shaded regions indicate standard

200   errors across all participants (N=18).

201

202   The results of MVPA suggest that early cortical distinctions between tones reflect their

203   absolute pitch (i.e. f_0) differences, while later distinctions reflect the musical pitch structure

204   of the STH. We next sought to explicitly test this hypothesis within the framework of

205   *representational similarity analysis* (RSA) (19). The set of dissimilarities corresponding to

206   all pairwise combinations of tones were indexed in a time-varying *representational*

207   *dissimilarity matrix* (RDM; figure 3A). For a given subject and timepoint, each cell of the

208   diagonally symmetric RDM indicates the cortical dissimilarity between the tones indexed by

209   the cell's row and column. We found that the RDMs of individual subjects were correlated

210   with one another from 100 ms onwards (figure 3B), verifying that the representational

211   structure was consistent across listeners over the same temporal extent in which average

212   stimulus-distinctions (in figure 3A) were apparent.

213

214   Next, we evaluated the predictive capacity of several models that attempt to explain the

215   observed structure of time-varying cortical RDMs. Each model was coded as a candidate

216   RDM, making its predictions about the expected dissimilarities between tones explicit (figure

217   3C). One candidate RDM was based on the *Standard Tonal Hierarchy (STH)*, where

218   distances between tones corresponded to their difference in perceived stability, as reported by

219   Krumhansl & Kessler (1982) (4). Another candidate RDM coded for differences in *Pitch*

220   *Height (PH)* in order to evaluate whether distinctions between tones were driven by their f_0

221   separation. We additionally tested two purely sensory models; one based on the *Spectral*

222   *Distance (SD)* between tone-pairs, and another based on the differences in their *Spectral*

223   *Overlap (SO)* with the preceding musical context (see methods for details). Each listener's

224   cortical RDM at every time point was compared with the four-different candidate RDMs

225   using a rank-order correlation measure, resulting in four curves tracking neural-model

226   correlation across time (figure 3D). Consistent with the earlier findings from MVPA, the PH

227   and STH models significantly explained cortical RDMs in early (100 – 250 ms) and later

228   (190 ms onwards) regions of neural processing respectively. Crucially, both PH and STH

229   correlations closely tracked the noise ceiling (20), indicating that these models offered

230   optimal predictive power given the noise levels inherent in the MEG data (see methods). The

231   temporal order of model correlations is consistent with dominant conceptions of melodic

232   processing, which posit the extraction of complex pitch before the integration and analysis of

233   broader tonal-harmonic structure (21). Interestingly, from 190 to 250 ms, PH and STH

234   models were both significantly correlated with cortical RDMs, suggesting an intermediary

235   period during which the brain holds a combined representation of both the tone's f_0 and pitch-

236   class.

237

238   To better visualize the above results, neural-model correlations were averaged into three time

239   bins (figure 3E): the first corresponded to a period before stimulus-specific information was

240   present in cortical activity (-100 to 100 ms); the second corresponded to the period during

241   which cortical structure was most strongly correlated with PH differences (100 to 200 ms);

242   and the third corresponded to the remainder of the neural epoch, during which cortical

243   structure reflected the STH (200 to 1000 ms). Time-averaged neural RDMs corresponding to

244   each of the three bins are displayed in the top panels of figure 3F-H. To more intuitively

245   visualize their dissimilarity structure, we applied multidimensional scaling (MDS) to each

246   RDM, obtaining a 2-dimensional solution in each case (Figure 3F-H; bottom panel). The

247   MDS solution in figure 3G clearly demonstrates the organization of pitch from low to high as

248   the space is traversed from upper-left to lower-right respectively. Similarly, the spatial

249   organization of the MDS solution in figure 3H illustrates many key properties of the STH.

250   Traversing the space from right to left reveals the structure of the hierarchy, with the most

251   stable pitch-class (C) situated on the right side, closest to the next most stable classes (G and

252   E) but distant from the cluster of unstable classes (F#, G#, D#, A#, C#) in the lower left

253   corner. Prior behavioral research has underscored the perceptual primacy of this hierarchical

254   arrangement (4). Our findings now provide evidence of its origins in the cortex and reveal the

255   temporal dynamics with which it emerges from the acoustic signal via an intermediate

256   representation of pitch-height.

257
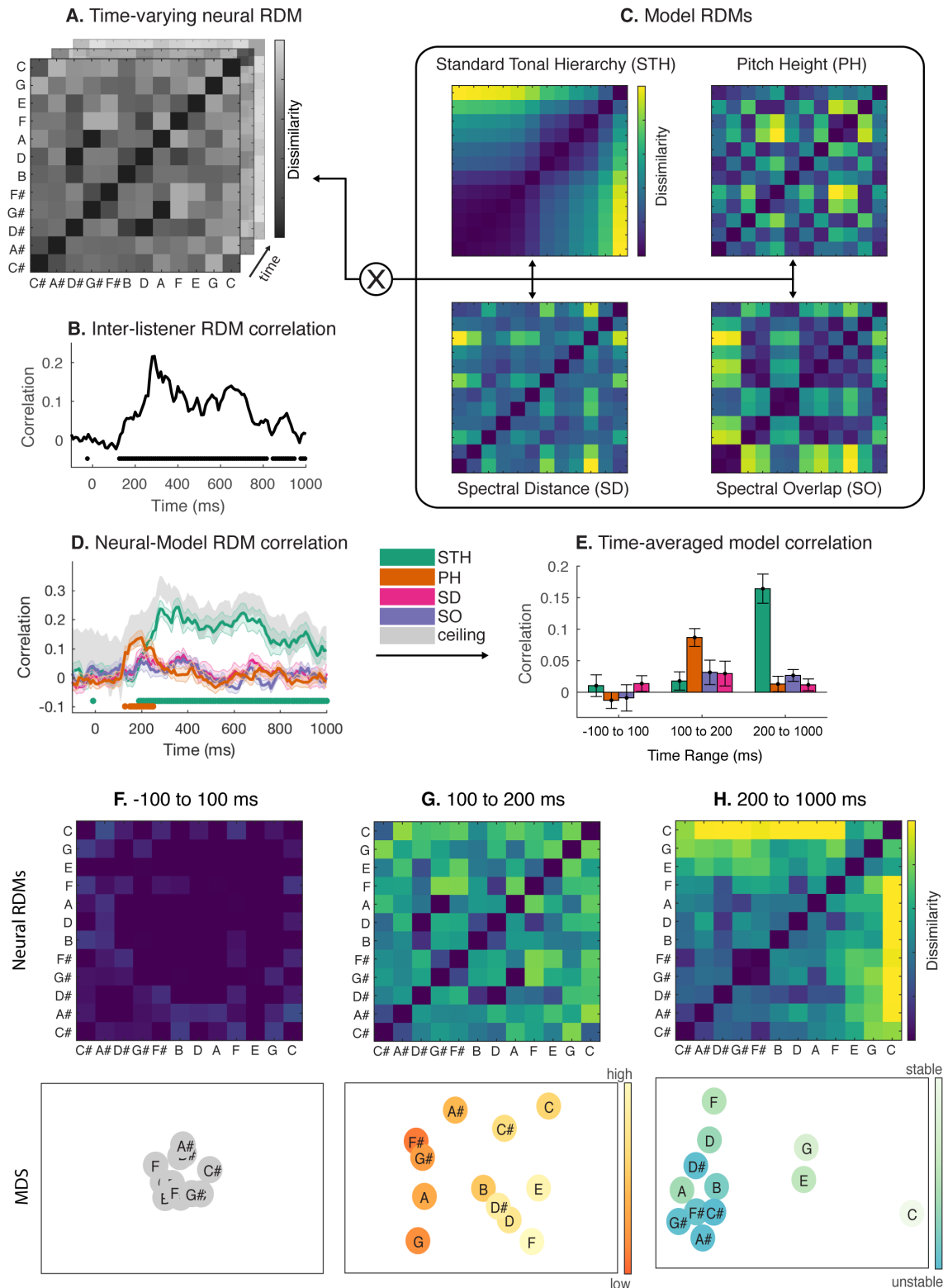
Figure 3. Representational similarity analysis of pitch-class. (A) Neural representational similarity matrix (RDM) indexing measured cortical dissimilarities between pairs of pitch-classes at each time point in the neural epoch. (B) The mean rank-order correlation between the RDMs of individual listeners (N=18). Significant time points are indicated underneath

263   the curve (p<0.05; randomization test, FDR corrected). (C) Four different candidate RDMs

264   based on models that attempt to explain neural dissimilarities. (D) Rank-order correlations

265   between each model and neural RDMs at each time point. Shaded regions indicate standard

266   errors across listeners. Significant time points are indicated by colored markers beneath

267   curves (p<0.05; Wilcoxon sign-rank tests, FDR corrected). (E) For visualization purposes,

268   neural-model RDM correlations were averaged across three different peristimulus time bins.

269   (F-H) Average neural RDMs (top) and multidimensional scaling solutions (bottom) for the

270   three regions in E. Colormaps indicate pitch-height (low to high) or perceived stability

271   (unstable to stable) in G and H respectively.

272

273   **Representation of major musical keys in cortex**

274   In tonal music, the perceptual structure that exists between individual pitch-classes is thought

275   to generate the second-order percept of a harmonic center or "key" (22). Logically therefore,

276   we reasoned that two keys should be related in the brain to the extent that they impose a

277   similar neural structure amongst the constituent tones. Adapting the procedure of prior

278   behavioral research (4) to the neural domain, we next used the MEG-based distinctions

279   between tones to derive an empirical measure of inter-key distances, comparing the resulting

280   structure with the *circle of fifths* (figure 1B).

281

282   In order to only capture the cortical processing of tonal-schema in our analysis, neural RDMs

283   were first averaged across 250–1000 ms; a time during which only the STH model

284   significantly predicted cortical RDMs. Next, we used MDS to geometrically express the

285   dissimilarity structure between the twelve tones as points in representational space. An

286   eleven-dimensional MDS solution was found for each subject's neural RDM, noting that n

287   objects will perfectly fit into n-1 dimensions (23). To transpose the representational structure

288   between tones (measured in the key of C major) into different keys, the configuration of

289   tones was shifted in MDS-space by the appropriate number of steps entailed by a given

290   transposition. For example, to transpose the MDS structure from C major to G major, the

291   point representing the tone 'C' was shifted to that occupied by 'G', the point occupied by

292   'C#' was shifted to that of 'G#', and the process was repeated for all twelve pitch-classes.

293   The overall dissimilarity between two keys was computed as the mean Euclidean distance

294   across all twelve tone-translations. Application of this procedure to all pairwise combinations

295   of the twelve major keys resulted in a cortical *inter-key* RDM in which the rows and columns

296   correspond to different keys and cells code the corresponding distance between two keys.

297   The average inter-key RDM across subjects is displayed in figure 4A, alongside a candidate

298   RDM based on the circle of fifths (figure 4B). Rows and columns are ordered such that

299   adjacent cells progress in intervals of a fifth. We found that the two structures were

300   significantly correlated (Kendall's Tau$_A$ = 0.26; p = 0.002) and shared several essential

301   properties. For example, keys separated by fifths were most proximate (e.g. C major and G

302   major), while those separated by 6-semitones were most distant (e.g. C major and F# major).

303   The generative nature of tonal music has been established by decades of perceptual research -

304   showing that the perceived structure between keys emerges directly from the constituent

305   structure that exists between tones (24, 25, 26, 22). Current results establish the

306   neurophysiological basis of this generative property, deriving the same musical key-relations

307   directly from the MEG response structure to individual tones.

308

A. Neural inter-key RDM
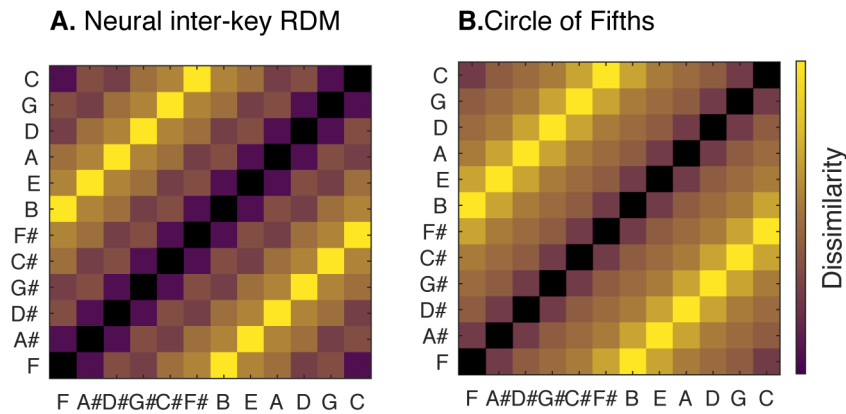
B. Circle of Fifths

Figure 4. Inter-key relationships. (A) Neural RDM indicating the pairwise distances

between the twelve different major keys. (B) A candidate RDM based on the "circle of

fifths".

We have dynamically characterized distinctions in the spatiotemporal patterns of cortical

activity encoding the different classes of Western musical pitch. Our results suggest that, as a

tone is received by the auditory system, an evolution exists in the underlying information

contained in its cortical population codes. Initially, they represent a localized and intrinsic

attribute of the tone. Eventually however, they contain information reflecting its integration

with the surrounding context and an acquired knowledge of the pitch-structure of Western

tonal music. In elucidating the representational dynamics underlying musical pitch

perception, we shed light on the neural underpinnings of domain-general perceptual

processes in which incoming sensory signals interact with internal structural knowledge. It

remains the goal of future work to further detail the neural computations involved in

integrating these two sources of information to arrive at an ultimate percept.

Materials & Methods

Participants.

328   Eighteen subjects with a minimum of 5 years of formal music training (mean = 11.9 years)

329   were recruited through the Sydney Conservatorium of Music and Macquarie University to

330   partake in the study. All subjects reported having no known hearing loss or brain

331   abnormalities and did not possess absolute pitch. The study was approved by the Human

332   Research Ethics Committee at Macquarie University (REF 5201300804) and all methods

333   were carried out in accordance with the stated guidelines. Informed consent was obtained

334   prior to testing, after all experimental details and potential risks were explained.

335

336   **Apparatus.**

337   Data were collected with a whole-head MEG system (Model PQ1160R-N2; KIT, Kanazawa,

338   Japan) consisting of 160 coaxial first-order gradiometers with a 50 mm baseline (27, 28).

339   Prior to recording, each participant's head shape was measured with a pen digitizer

340   (Polhemus Fastrack, Colchester, VT, USA) and the positions of five marker coils on the

341   surface of the scalp were registered. During recording, MEG data was bandpass filtered

342   online from 0.1 – 200 Hz using first-order RC filters and digitized at 1000 Hz. Participants

343   were in a supine position within a magnetically shielded room containing the MEG sensors.

344   During experimental trials, they were instructed to direct their gaze at a fixation cross. Both

345   the fixation cross and experimental instructions were projected by an InFocus IN5108 LCD

346   back projection system (InFocus, Portland, Oregon, USA) to a screen located above the

347   participant at a viewing distance of 113 cm. Sound stimuli were delivered via Etymonic ER-

348   30 insert headphones at a sampling frequency of 44.1 kHz.

349

350   **Stimuli & Design.**

351   Stimuli were piano tones recorded at 44.1kHz and sampled using Max/MSP (Cycling '74,

352   San Francisco, CA) to construct tones that were 500ms in duration with an additional 150ms

353   decay. Prior to testing, all tones were passed through a time varying loudness model (29) to

354   normalize for differences in perceived loudness. For each tone, the maximum short-term-

355   loudness (STL_{max}) was computed and normalized to the mean value of all four tones.

356   Differences in STL_{max} between all probe-tones did not exceed 3 phones.

357   Subjects were presented with a series of trials while having their MEG activity recorded.

358   Each trial consisted of a tonal context followed by a single tone (hereafter referred to as the

359   *probe-tone*). The tonal context consisted of four major chords written in four-part harmony

360   outlining an I-IV-V-I harmonic progression in the key of C major. The context and probe-

361   tone were separated by a silent period equivalent to one beat (650ms, 92 bpm). This temporal

362   separation was introduced in order to prevent the sensory processing of the context from

363   contaminating evoked responses to probe-tones whilst maintaining metric regularity. On a

364   given trial, the subsequent probe-tone was one of twelve notes spanning the chromatic range

365   between F#3 (185 Hz) and F4 (349 Hz). This range was chosen to minimize the average

366   pitch-distance between the probe-tone and its preceding context. On each trial, presentation

367   of probe-tones was randomized but constrained to avoid repeated presentation across adjacent

368   trials. To ensure participants were attending to stimuli (30), participants judged whether the

369   probe-tone on each trial was 'in-key' or 'out-of-key', registering their response after the

370   occurrence of the probe-tone by pressing one of two buttons. Participants were instructed to

371   use their left and right thumbs to register the two respective responses. The mapping of in-

372   key/out-of-key to left/right button was interchanged every two blocks to control for the

373   effect, if any, of motor activity on the measured neural responses. No trial-by-trial feedback

374   was provided during the MEG recording. On average, subjects responded correctly on 78%

375   of the trials (SD = 16.3%). All trials, including those with incorrect responses, were included

376   in the subsequent neural analysis (31). Once the response was registered, inter-trial-intervals

377   were randomly roved between 0.5 - 1 sec. Before testing, subjects completed a training

378   session consisting of 20 trials with an identical behavioral task to that of the main

379   experiment. Feedback was provided after each training trial and the experimenter ensured

380   that subjects could perform the task (using a threshold of $\geq 75\%$ correct) before proceeding to

381   the MEG recording session. Each participant's MEG data were collected in a single hour-long

382   session. The total experiment comprised 672 trials, yielding 56 neural observations of each of

383   the 12 probe-tones. Testing was divided into 8 blocks, each comprised of 84 trials and

384   separated by one-minute breaks.

385

386   **Analysis.**

387   *MEG pre-processing.*

388   Pre-processing of MEG data was performed in MATLAB. Data corresponding to each

389   participant was first epoched from 100 ms before to 1000 ms after onset of probe-tones

390   before being down-sampled to 100 Hz with a low-pass Chebyshev Type 1 filter. Down-

391   sampling improved the overall SNR while still retaining a suitable level of temporal

392   resolution to examine the time course of neural pitch-processing. Next, spatial Principal

393   Components Analysis (PCA) was applied to the dataset of each participant using the MEG

394   sensor channels as input features. We retained components that cumulatively explained 99%

395   of the variance. On average, PCA reduced the dimensionality of datasets from 160 sensor

396   channels to 28 principle components (SD = 5.4). PCA has been found to be an efficient pre-

397   processing step for optimizing data for MEG decoding analyses (32). In a single step, PCA

398   reduces the dimensionality of the data, and obviates the need for additional artefact rejection

399   or de-noising procedures, as classifiers can "learn" to suppress nuisance variables isolated by

400   PCA, e.g. eye-blinks and environmental noise.

401

402   *Multivariate pattern classification of MEG activity.*

403   To measure the neural dissimilarity between tones, we applied Multivariate pattern analysis

404   (MVPA) (17), whereby a binary classifier learns features of the evoked MEG activity that

405   best distinguishes two different tones. MVPA was applied to each subject's pre-processed

406   dataset using MATLAB. Prior to classification, we averaged the MEG responses of 2 trials

407   within the same pitch category in order to boost the overall SNR of classification (32). We

408   used a naïve Bayes implementation of linear discriminate analysis (LDA) (33) to perform

409   classification for each pairwise combination of tones. Generalization of the classifier was

410   evaluated using k-fold cross validation with a 9:1 training to test ratio. Specifically, MEG

411   data corresponding to the two classes being classified were randomly assigned to 10 bins of

412   equal size, with a balanced number of observations from each class in every bin. Next, nine

413   of the bins were pooled together and used to train the classifier, and the trials in the

414   remaining bin were used to test the classifier. This procedure was repeated 10 times such that

415   each bin was utilized for testing once. The reported accuracy is the average across all 10

416   cross-validation runs. A sliding classification time-window was used on the MEG time-

417   series, resulting in a curve of classifier accuracy across time that tracks the dynamic

418   emergence of stimulus-related information in the cortex. The classifier window was 50ms

419   long and adjacent classification runs traversed the neural epoch in 10ms steps. Importantly,

420   the neural response at each adjacent time point within the 50 ms window mapped onto a new

421   dimension in the classification feature space. In this fashion, classifiers not only

422   discriminated between responses based on their spatial activation patterns at each time, but

423   also their fine-grained temporal response structure across multiple time-points. Classifier

424   performance at each time point was evaluated in terms of balanced accuracy (32), whereby

425   accuracy is evaluated individually for each class and then averaged. Significance at the group

426   level (N = 18) at each time sample was evaluated using two-sided Wilcoxon sign-rank tests

427   (p<0.05). Multiple comparisons were corrected by controlling the false discovery rate (FDR)

428   (34, 35) with $\alpha = 0.05$.

429

430   ***Representational similarity analysis of tones.***

431   Application of MVPA as described above to every pairwise combination of the twelve tones

432   resulted in a 12x12 diagonally symmetric Representational Dissimilarity Matrix (RDM) for

433   every subject and time sample. To check for consistency across subjects, the mean inter-

434   subject RDM correlation (figure 3B) was calculated at each time sample by averaging the

435   rank-order correlation (Kendall's Tau_A) (20) of all pairwise combinations of individual

436   subjects' RDMs (N=18). Significance was assessed by way of randomization testing. Briefly,

437   the columns of subjects' RDMs were randomly permuted before being correlated, and this

438   procedure was repeated 1000 times, resulting in a correlation noise floor. Significance was

439   based on the true mean correlation rising above the 95% distribution of the noise floor (FDR

440   corrected). Next, MEG RDMs for each subject and at each time sample were compared with

441   four candidate RDMs coded according to the predictions of several perceptual and sensory

442   models of pitch. Candidate RDMs were as follows: [1] An RDM based on the *Standard*

443   *Tonal Hierarhcy (STH)* was constructed in which each cell coded the difference in perceived

444   stability between the two corresponding tones using the major-profile ratings reported in

445   Krumhansl & Kessler (1982). [2] In order to test the hypothesis that MEG dissimilarities

446   reflected the difference in each tone's fundamental frequency (f_0), we constructed a *Pitch-*

447   *Height (PH)* RDM, in which each cell corresponded to the semitone difference in f_0 for the

448   two tones in question. [3] To assess whether neural dissimilarities between tones reflected

449   their sensory differences, a *Spectral Distance (SD)* RDM was constructed in which each cell

450   corresponded to the Euclidean distance between the 128-channel stimulus spectrograms of

451   two tones. Spectrograms for each tone were extracted by passing the raw audio through a

452   biologically inspired model of the auditory periphery (36). The model consisted of three main

453   stages: (i) a cochlear filter bank comprised of 128 asymmetric filters equally distributed in

454   log-frequency, (ii) a hair cell stage consisting of a low-pass filter and nonlinear compression

455   function, and finally (iii) a lateral inhibitory network modelled as a first-order derivative

456   along the tonotopic axis followed by a half-wave rectifier. (4) Lastly, although the tonal

457   context and probe-tone were separated by 650ms (see experiment design), models of auditory

458   short-term memory involve time-constants of up to 4 seconds (37, 38). Thus, it was possible

459   that neural dissimilarities between tones were driven by the sensory memory of the context.

460   To test this possibility, we constructed a *Spectral Overlap (SO)* RDM. First, the context

461   stimulus waveform was passed through the auditory peripheral model described above in

462   order to obtain a context spectrogram. Next, the Euclidean distance between the context and

463   each probe-tone was calculated from their respective spectrograms. Each cell in the SO RDM

464   was then coded as the difference in spectral distance between context and probe-tone for the

465   two tones in question. Additional perceptual models were considered - for example the "basic

466   space" of the *Tonal Pitch Space Theory* (25). However, the candidate RDMs arising from

467   such models shared an identical rank-order structure to that of the STH and were therefore

468   precluded from the analysis. Using the framework of Representational Similarity Analysis

469   (RSA) (19), we studied the brain's emerging representation by comparing each candidate

470   RDM with the empirical time-varying MEG RDM (see statistical analysis below).

471   Correlations between neural and model RDMs were assessed by computing a rank-order

472   correlation measure (Kendall's Tau_A). We used the 'noise ceiling' as a benchmark for testing

473   model performance (20). The noise ceiling uses inter-subject variance in RDMs to estimate

474   the magnitude of the expected correlation between a "true" model RDM and the empirical

475   RDM given measurement noise. To visualize the structure of RDMs, Multidimensional

476   Scaling (MDS) was applied using Kruskal's normalized stress 1 criterion.

477

478   *Representational similarity analysis of musical keys.*

479   Inter-key distances in the cortex were derived by adapting the analytical approach established

480   in Krumhansl & Kessler (1982) (4) to the neural domain. First, MEG RDMs from 250 – 1000

481   ms were averaged to obtain a single time-averaged neural RDM for each subject. In order to

482   geometrically express the RDM distances between tones as points in representational space,

483   we applied nonmetric MDS to the time-averaged RDMs of each subject. Because *n* objects

484   will always fit into *n-1* dimensions (23), MDS solutions were obtained in eleven dimensions.

485   Accordingly, all solutions had stress equal to zero, indicating that the MDS decomposition

486   perfectly preserved distance information in the RDMs. To transpose the representational

487   structure between the twelve tones into different keys, the twelve points corresponding to

488   each pitch-class were shifted in MDS-space by the appropriate number of steps implicated by

489   the transposition. For example, to transpose the MDS structure from C major to G major

490   (seven semitone steps), the point representing the tone 'C' was shifted to that occupied by

491   'G', the point occupied by 'C#' was shifted to that of 'G#', and the process was repeated for

492   all twelve pitch-classes. The distance between two keys was then defined as the mean

493   Euclidean distance between the original and new positions of all twelve tones. In this fashion,

494   distances were computed between all twelve major keys, resulting in a neural inter-key RDM

495   for each subject, in which rows and columns correspond to different musical keys and each

496   cell codes the corresponding distance between two keys (figure 4A). Finally, neural inter-key

497   RDMs of each subject were rank-order correlated (using Kendall's Tau$_A$) with a candidate

498   inter-key RDM based on the circle of fifths (figure 4B).

499

500   **Acknowledgements.**

505

References.

507   1.   Lerdahl F (1992) Cognitive constraints on compositional systems. *Contemporary*

508   *Music Review*, *6*(2), 97-121.

509

510   2.   Vos P-G, Troost J-M (1989) Ascending and descending melodic intervals: Statistical

511   findings and their perceptual relevance. *Music Perception: An Interdisciplinary*

512   *Journal*, *6*(4), 383-396.

513

514   3.   Krumhansl C-L, Shepard R-N (1979) Quantification of the hierarchy of tonal

515   functions within a diatonic context. *Journal of experimental psychology: Human*

516   *Perception and Performance*, *5*(4), 579.

517

518   4.   Krumhansl C-L, Kessler E-J (1982) Tracing the dynamic changes in perceived tonal

519   organization in a spatial representation of musical keys. *Psychological review*, *89*(4),

520   334.

521

522   5.   Hall D-A, Johnsrude I-S, Haggard M-P, Palmer A-R, Akeroyd M-A, Summerfield A-

523   Q (2002) Spectral and temporal processing in human auditory cortex. *Cerebral*

524   *Cortex*, *12*(2), 140-149.

525

526   6. Wessinger C-M, VanMeter J, Tian B, VanLare J, Pekar J, Rauschecker J-P (2001)

527   Hierarchical organization of the human auditory cortex revealed by functional

528   magnetic resonance imaging. *Journal of cognitive neuroscience*, *13*(1), 1-7.

529

530   7. Zatorre R-J, Evans A-C, Meyer E (1994) Neural mechanisms underlying melodic

531   perception and memory for pitch. *Journal of Neuroscience*, *14*(4), 1908-1919.

532

533   8. Griffiths T-D, Büchel C, Frackowiak R-S, Patterson R-D (1998) Analysis of temporal

534   structure in sound by the human brain. *Nature neuroscience*, *1*(5), 422.

535

536   9. Patterson R-D, Uppenkamp S, Johnsrude I-S, Griffiths T-D (2002) The processing of

537   temporal pitch and melody information in auditory cortex. *Neuron*, *36*(4), 767-776.

538

539   10. Lee Y-S, Janata P, Frost C, Hanke M, Granger R (2011) Investigation of melodic

540   contour processing in the brain using multivariate pattern-based

541   fMRI. *Neuroimage*, *57*(1), 293-300.

542

543   11. Klein M-E, Zatorre R-J (2011) A role for the right superior temporal sulcus in

544   categorical perception of musical chords. *Neuropsychologia*, *49*(5), 878-887.

545

546   12. Foo F, King-Stephens D, Weber P, Laxer K, Parvizi J, Knight R-T (2016) Differential

547   processing of consonance and dissonance within the human superior temporal

548   gyrus. *Frontiers in human neuroscience*, *10*, 154.

549

550 13. Fedorenko E, McDermott J-H, Norman-Haignere S, Kanwisher N (2012) Sensitivity

551   to musical structure in the human brain. *Journal of Neurophysiology*, *108*(12), 3289-

552   3300.

553

554 14. Krohn K-I, Brattico E, Välimäki V, Tervaniemi M (2007) Neural representations of

555   the hierarchical scale pitch structure. *Music Perception*, *24*(3), 281-296.

556

557 15. Brattico E, Tervaniemi M, Näätänen R, Peretz I (2006) Musical scale properties are

558   automatically processed in the human auditory cortex. *Brain research*, *1117*(1), 162-

559   174.

560

561 16. Sankaran N, Thompson W-F, Carlile S, Carlson T-A (2018) Decoding the dynamic

562   representation of musical pitch from human brain activity. *Scientific reports*, *8*(1),

563   839.

564

565 17. Haxby J-V, Connolly A-C, Guntupalli J-S (2014) Decoding neural representational

566   spaces using multivariate pattern analysis. *Annual review of neuroscience*, *37*, 435-

567   456.

568

569 18. Janata P, Birk J-L, Van Horn J-D, Leman M, Tillmann B, Bharucha J-J (2002) The

570   cortical topography of tonal structures underlying Western music. *science*, *298*(5601),

571   2167-2170.

572

19. Kriegeskorte N, Mur M, Bandettini P-A (2008) Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, *2*, 4.

20. Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A toolbox for representational similarity analysis. *PLoS computational biology*, *10*(4), e1003553.

21. Koelsch S (2011) Toward a neural basis of music perception–a review and updated model. *Frontiers in psychology*, *2*, 110.

22. Krumhansl C-L (2001) Cognitive foundations of musical pitch. Oxford University Press.

23. Kruskal J-B, Wish M (1978) Multidimensional scaling. Sage University Paper series on quantitative applications in the social sciences (11). Beverly Hills, CA: Sage.

24. Lerdahl F (2004) *Tonal pitch space*. Oxford University Press.

25. Lerdahl F (1988) Tonal pitch space. *Music Perception: An Interdisciplinary Journal*, *5*(3), 315-349.

26. Chew E (2000) *Towards a mathematical model of tonality* (Doctoral dissertation, Massachusetts Institute of Technology).

599   27. Kado H, Higuchi M, Shimogawara M, Haruta Y, Adachi Y, Kawai J, Ogata H,

600   Uehara G (1999) Magnetoencephalogram systems developed at KIT. *IEEE*

601   *transactions on applied superconductivity*, *9*(2), 4057-4062.

602

603   28. Uehara G, Adachi Y, Kawai J, Shimogawara M, Higuchi M, Haruta Y, Ogata H,

604   Kado H (2003) Multi-channel SQUID systems for biomagnetic measurement. *IEICE*

605   *transactions on electronics*, *86*(1), 43-54.

606

607   29. Glasberg B-R, Moore B-C (2002) A model of loudness applicable to time-varying

608   sounds. *Journal of the Audio Engineering Society*, *50*(5), 331-342.

609

610   30. Loui P, Grent T, Torpey D, Woldorff M (2005) Effects of attention on the neural

611   processing of harmonic syntax in Western music. *Cognitive Brain Research*, *25*(3),

612   678-687.

613

614   31. VanRullen R (2011) Four common conceptual fallacies in mapping the time course of

615   recognition. *Frontiers in psychology*, *2*, 365.

616

617   32. Grootswagers T, Wardle S-G, Carlson T-A (2017) Decoding dynamic brain patterns

618   from evoked responses: A tutorial on multivariate pattern analysis applied to time

619   series neuroimaging data. *Journal of cognitive neuroscience*, *29*(4), 677-697.

620

621   33. Duda R-O, Hart P-E, Stork D-G (2012) *Pattern classification*. John Wiley & Sons.

622

623   34. Benjamini Y, Yekutieli D (2001) The control of the false discovery rate in multiple

624   testing under dependency. *Annals of statistics*, 1165-1188.

625

626   35. Nichols T-E (2012) Multiple testing corrections, nonparametric methods, and random

627   field theory. *Neuroimage*, *62*(2), 811-815.

628

629   36. Chi T, Ru P, Shamma S-A (2005) Multiresolution spectrotemporal analysis of

630   complex sounds. *The Journal of the Acoustical Society of America*, *118*(2), 887-906.

631

632   37. Leman M (2000) An auditory model of the role of short-term memory in probe-tone

633   ratings. *Music Perception: An Interdisciplinary Journal*, *17*(4), 481-509.

634

635   38. Huron D, Parncutt R (1993) An improved model of tonality perception incorporating

636   pitch salience and echoic memory. *Psychomusicology: A Journal of Research in

637   Music Cognition*, *12*(2), 154.

638
639

640