Processing of hierarchical syntactic structure in music

Stefan Koelsch[a,1], Martin Rohrmeier[a,b], Renzo Torrecuso[a,c], and Sebastian Jentschke[a]

[a]Cluster: Languages of Emotion, Freie Universität, 14195 Berlin, Germany; [b]MIT Intelligence Initiative, Department of Linguistics and Philosophy, Massachusetts Institute of Technology, Cambridge, MA 02139; and [c]Brain Institute, Federal University of Rio Grande do Norte, 59056-450, Natal, Brazil

Edited* by Dale Purves, Duke-National University of Singapore Graduate Medical School, Singapore, Singapore, and approved July 30, 2013 (received for review January 8, 2013)

Hierarchical structure with nested nonlocal dependencies is a key feature of human language and can be identified theoretically in most pieces of tonal music. However, previous studies have argued against the perception of such structures in music. Here, we show processing of nonlocal dependencies in music. We presented chorales by J. S. Bach and modified versions in which the hierarchical structure was rendered irregular whereas the local structure was kept intact. Brain electric responses differed between regular and irregular hierarchical structures, in both musicians and nonmusicians. This finding indicates that, when listening to music, humans apply cognitive processes that are capable of dealing with long-distance dependencies resulting from hierarchically organized syntactic structures. Our results reveal that a brain mechanism fundamental for syntactic processing is engaged during the perception of music, indicating that processing of hierarchical structure with nested nonlocal dependencies is not just a key component of human language, but a multidomain capacity of human cognition.

syntax | context-free grammar | parsing | electroencephalography | EEG

To process sequential information featuring both local and nonlocal dependencies between elements, nervous systems need to represent information on different time scales, as reflected in different frequencies of oscillatory processes (1, 2) and different types of memory (3, 4). Tonal music has evolved to an extent that composers could make the fullest use of such representations. On the one hand, tonal music involves representations of single events and local relationships on short time scales. On the other hand, many composers designed nested hierarchical syntactic structures spanning longer time scales, potentially up to entire movements of symphonies and sonatas (5, 6). Hierarchical syntactic structure (involving the potential for nested nonlocal dependencies) is a key component of the human language capacity (7–11) and is frequently produced and perceived in everyday life. For example, in the sentence “the boy who helped Peter kissed Mary,” the subject relative clause ”who helped Peter” is nested into the main clause ”the boy kissed Mary,” creating a nonlocal hierarchical dependency between ”the boy” and ”kissed Mary.” Music theorists have described analogous hierarchical structures for music. Schenker (5) was the first to describe musical structures as organized hierarchically, in a way that musical events are elaborated (or prolonged) by other events in a recursive fashion. According to this principle, e.g., a phrase (or set of phrases) can be conceived of as an elaboration of a basic underlying tonic–dominant–tonic progression. Schenker further argued that this principle can be expanded to even larger musical sequences, up to entire musical movements. In addition, Hofstadter (12) was one of the first to argue that a change of key embedded in a superordinate key (such as a tonal modulation away from, and returning to, an initial key) constitutes a prime example of recursion in music. Based on similar ideas, several theorists have developed formal descriptions of the analysis of hierarchical structures in music (13–15). One of these approaches, the Generative Theory of Tonal Music (GTTM) by Lerdahl and Jackendoff (13), has become one of the most influential current theories in music theory and music psychology. Another approach is the Generative Syntax Model (GSM), which provides explicit generative rules modeled in analogy with linguistic syntax (15). However, it has remained unknown whether hierarchical musical structure is perceived by human listeners, or whether hierarchical musical structure is merely a historical convention driven by factors such as notation (where relationships between keys can be surveyed and constructed on paper). The perception of hierarchical structure of music would indicate that this structural property reflects, and is driven by, our capacity to perceive and produce hierarchical, potentially recursive structures (7, 8, 16).

More critically, the theoretical accounts on hierarchical structures in music have been challenged by scholars who argued that the traditional theory of harmony is local and that syntax of tonal music can be captured, e.g., by Markov models (17, 18). Likewise, it has been argued that musical understanding does not centrally involve grasp of large-scale musical dependencies (19). This view assumes that hierarchical accounts are not reflected in the cognitive processing of musical structure and that local models yield the best account of elementary tonal harmony (18).

Empirical evidence on this topic is sparse, but, if anything, then empirical data rather support local accounts, showing that even musically trained listeners are perceptually surprisingly insensitive to drastic manipulations of large-scale musical structure (20), including scrambling the order of the phrases within a single piece (21) or rewriting sections of large tonal pieces so that they end in keys that do not provide tonal closure (22). Notably, all previous studies reporting behavioral or neurophysiological effects of music–syntactic manipulations have tapped into processing of local dependencies, either with frank local violations (such as chords with out-of-key tones or harmonic sequences not ending with an authentic cadence) (23–25), or by manipulating the local transition probability of occurrence of syntactically legal events (26). This was the case even in those studies that used tree models to describe music–syntactic irregularities (27, 28). Thus, behavioral and neurophysiological effects reported in previous studies on music–syntactic processing could have been driven only by the processing of local dependencies (21, 28). Other studies showed recognition of harmonic and melodic reductions, which are predicted by syntactic theories of music like the GTTM or GSM (29, 30) or correlations between hierarchical structure and ratings of tension and relaxation (31), but those studies did not provide evidence for processing of long-distance dependencies (which are also predicted by GTTM and GSM). Thus, although hierarchical musical structures can be described theoretically, there is a striking absence of evidence for the processing of hierarchical syntactic structures involving long-distance dependencies in music.

To investigate this issue, we used two original chorales by J. S. Bach (BWV 302 and 373, Fig. 1, Fig. S1, and Audio File S1), both with a long-distance dependency of the basic form ABA. In addition, we used modified versions of the form A′BA in which

---

Author contributions: S.K. and M.R. designed research; R.T. performed research; S.K. contributed new reagents/analytic tools; S.J. analyzed data; and S.K., M.R., and S.J. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[1]To whom correspondence should be addressed. E-mail: stefan.koelsch@fu-berlin.de.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1300272110/-/DCSupplemental.

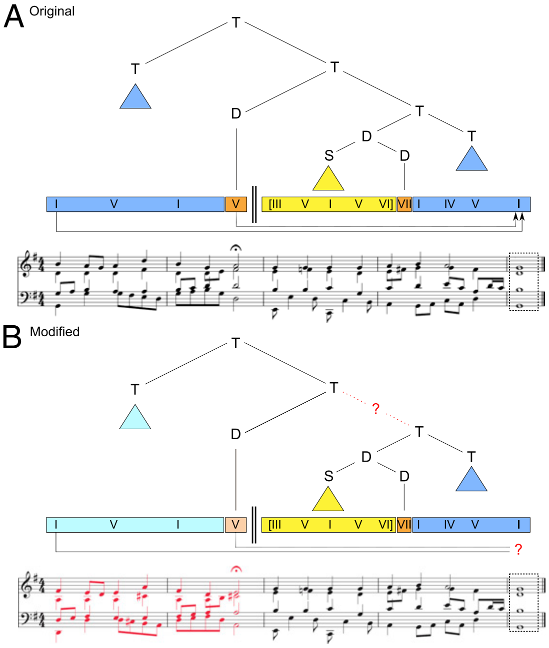


Fig. 1. Illustration of stimuli. (*A*) Original version of J. S. Bach's chorale *Liebster Jesu, wir sind hier* (BWV 373). The first phrase ends on an open dominant (see chord with fermata below orange rectangle), and the second phrase ends on a tonic (dotted rectangle). The tree structure above the scores represents a schematic diagram of the harmonic dependencies (for full tree graphs, see Figs. S2 and S3). The two thick vertical lines (separating the first and the second phrase) visualize that the local dominant (V in orange rectangle) is not immediately followed by a resolving tonic chord but implies its resolution with the final tonic (indicated by the dotted arrow). The same dependency exists between initial and final tonic (indicated by the solid arrow). The tree thus illustrates the nonlocal (long-distance) dependency between the initial and final tonic regions and tonic chords, respectively (also illustrated by the blue rectangles). The chords belonging to a key other than the initial key (yellow rectangle) represent one level of embedding. (*B*) Modified version (the first phrase was transposed downward by the pitch interval of one fourth, red color). The tree structure above the scores illustrates that the second phrase is not compatible with an expected tonic region (indicated by the red dotted line with the red question mark) and that the last chord (a tonic of a local cadence, dotted rectangle) neither prolongs the initial tonic nor closes the open dominant (see solid and dotted lines followed by red question mark). In both *A* and *B*, Roman numerals indicate scale degrees. T, S, and D indicate the main tonal functions (tonic, subdominant, dominant) of the respective part of the sequence (such as functional regions in the GSM). Squared brackets indicate scale degrees relative to the local key (in the original version, the yellow rectangle indicates that the local key of C major is a subdominant region of the initial key G major).

the long-distance dependency between A and A was not fulfilled. Each of the stimuli consisted of two phrases. In the original chorales, the first phrase ended on a half cadence (i.e., on an open dominant) (Fig. 1 and Fig. S1). The second phrase began with a chord other than the tonic (thus not immediately fulfilling the implication of the dominant at the end of the first phrase) and featured a sequence of chords that did not belong to the initial key of the chorale (representing one level of embedding). Then, the second phrase returned to the initial key and ended on an authentic cadence (in analogy with the recursive schema described by Hofstadter) (12). Thus, according to the GTTM and GSM, the final chord of original chorales hierarchically prolonged the first chord of the chorale and closed the established dominant that remained open at the end of the half cadence. Note that the parse trees of the syntactic structures of the two chorales according to the GSM and GTTM (Fig. 1 and Figs. S1–S5) represent recursive hierarchical organization that creates nonlocal dependencies in a way that embedded parts are (recursively) generated by the same set of rules as superordinated parts. As illustrated by the red scores in Fig. 1, we also created modified versions of these chorales by transposing the first phrase either down a forth (BWV 373) (Fig. 1 and Audio File S2), or up a major second (BWV 302) (Fig. S1). By doing so, the second phrase of each modified chorale did not prolong the first chord of the chorale anymore and did not close the open dominant established by the first phrase (see red question marks in Fig. 1 and Fig. S1).

This manipulation led to a hierarchical irregularity, while keeping the local structure of the second phrase intact. Several measures guarantee that the hierarchical irregularity does not confound local irregularity. First, despite the transposition of the first phrase of the modified chorales (red scores in Fig. 1 and Fig. S1), the second phrase remained unchanged and did thus not differ acoustically between original and modified chorales (that is, the last nine chords of BWV 373, and the last eight chords of BWV 302, were acoustically identical). Second, it has been shown that local n-gram models of harmony are optimal for a context length of two or three items (32, 33), and that predictions based on such models change only marginally (and to the worse) for longer local context models. Therefore, the local transition probabilities for the final chords were equal in both original and modified versions, and only the long-distance dependency between last and first chord was manipulated (as well as between last chord and open dominant of the half cadence). Consequently, any differences in behavioral or neurophysiological responses to the final chords of the two versions of the Bach chorales can only be due to the processing of the nonlocal, hierarchical structure of the chorale, but not due to local processing. Notably, in contrast to similar experimental designs used in previous research (23), stimuli of the present study contain a center-embedded dependency and end on a locally correct cadence, both of which are required to investigate hierarchical processing without contribution of local processing.

Note that we use the term "hierarchical" here to refer to a syntactic organizational principle of musical sequences by which elements are organized in terms of subordination and dominance relationships (13–15). Such hierarchical structures can be established through the recursive application of rules, analogous to the establishment of hierarchical structures in language (8). In both linguistics and music theory, such hierarchical dependency structures are commonly represented using tree graphs. The term "hierarchical" is sometimes also used in a different sense, namely to indicate that certain pitches, chords, or keys within pieces occur more frequently than others and thus establish a frequency-based ranking of structural importance (34). That is not the sense intended here.

Using electroencephalography (EEG), it has previously been observed that processing of music–syntactical irregularities is reflected electrically in an early right anterior negativity (ERAN) (reflecting music–syntactic processing) (25) and a subsequent late negativity (the so-called N5, reflecting harmonic integration) (28). Whether ERAN and N5 reflect local, hierarchical, or both local and hierarchical processing is not known. In the present study, we tested whether final chords of hierarchically irregular versions (in the absence of any local violation) would evoke ERAN and N5 potentials compared with the hierarchically regular versions. After the EEG session, conclusiveness and emotion ratings of our stimuli were obtained to test the hypothesis that conclusiveness ratings would be higher for original than for modified versions.

Results

Fig. 2*A* shows that, compared with the original versions, final chords of modified versions evoked an early negative brain-electric response that emerged in the N1 range (around 150 ms after chord onset) and was maximal at around 220 ms. This effect had a frontal scalp distribution and a slight (nonsignificant) left-hemispheric weighting. The early effect was followed by a later negativity that emerged at around 500 ms and lasted until about 850 ms after stimulus onset. A global ANOVA (*Materials and*

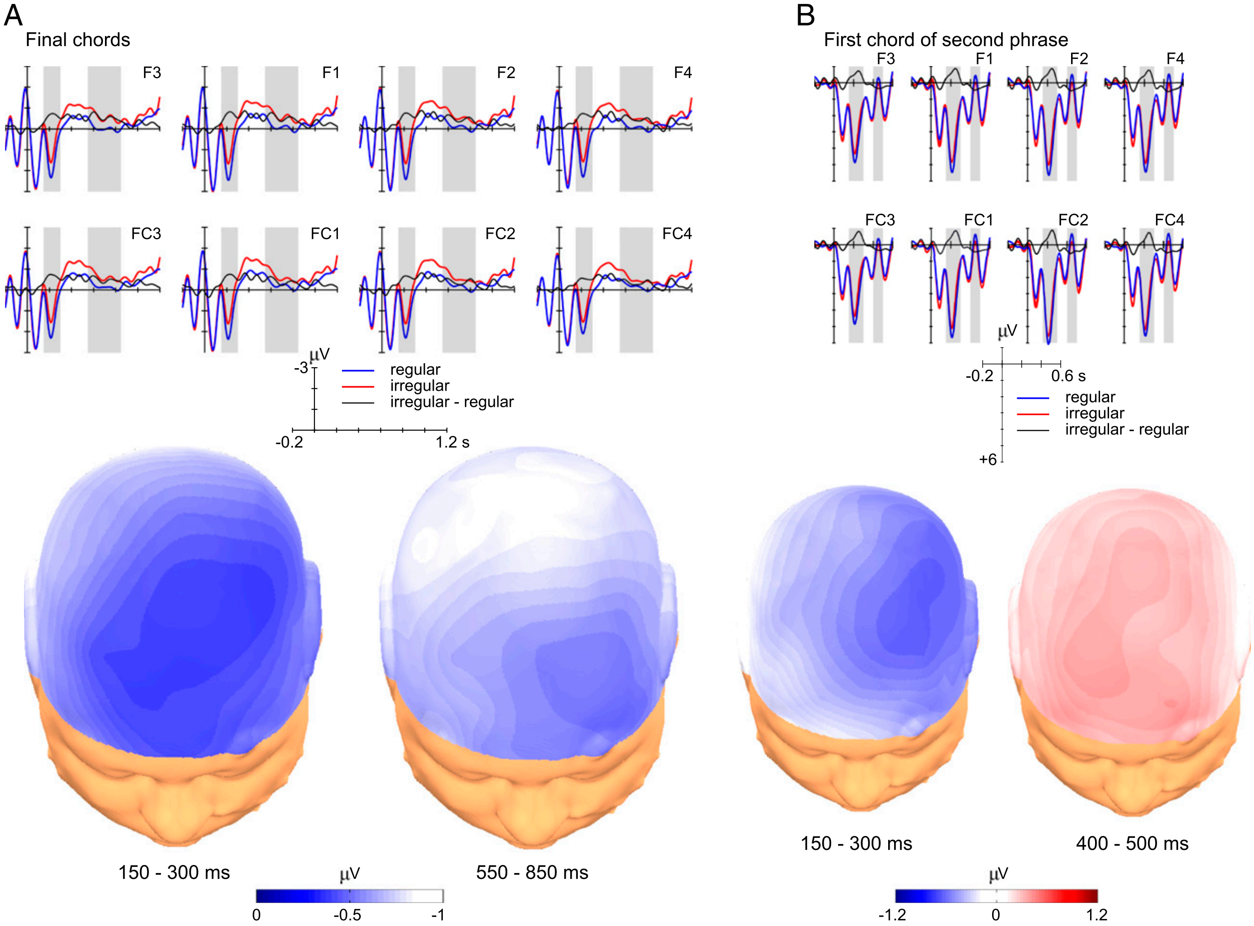


Fig. 2. Brain electric responses to chords. Event-related brain potentials (ERPs) evoked by the final chords are shown in *A*, and ERPs evoked by the first chord of the second phrase are shown in *B*, separately for original (blue waveforms) and modified versions (red waveforms). *Upper* of *A* shows that, compared with ERPs evoked by original versions, modified versions evoked an early negativity that was maximal at around 220 ms, and a later negativity that emerged at around 500 ms, and lasted until around 850 ms (best to be seen in the black difference wave: original subtracted from modified versions). Presentation time of the final chord was 1,200 ms. The lower panel of *A* shows the scalp distribution of the early and late ERP effects elicited by the final chords of modified versions (difference potentials: original subtracted from modified versions). *Upper* of *B* shows that, compared with ERPs evoked by original versions, modified versions evoked an early negativity that was maximal at around 200 ms, and a later positivity between around 400–500 ms (best to be seen in the black difference wave: original subtracted from modified version). Presentation time of chords was 600 ms. *Lower* of *B* shows the scalp distribution of the early and late ERP effects (difference potentials: original subtracted from modified version). Gray-shaded areas indicate time windows used for the statistical analysis reported in the main text. ERPs were recorded from 12 musicians and 12 nonmusicians; none of the ERP effects differed significantly between groups.

Methods) for a time window from 150 to 300 ms (early negativity) indicated an effect of condition [$F(1,22) = 5.39, p = .03$, reflecting that the event-related potentials (ERPs) differed between the original and the modified versions]. The ANOVA also indicated an interaction between condition and anterior–posterior [$F(1,22) = 5.57, p < .03$, reflecting that this effect had an anterior scalp distribution]. A follow-up ANOVA with frontal regions of interest (ROIs) with factors condition, hemisphere, and group indicated an effect of condition ($F(1,22) = 10.25, p = .004$), with no interaction between condition and hemisphere ($p = .59$), nor between condition and group ($p = .93$). Thus, the amplitude of the early negative effect did not differ significantly between musicians and nonmusicians (see also amplitude values provided in Table S1). The global ANOVA computed for a time window from 550 to 850 ms (late negativity) also yielded an effect of condition ($F(1,22) = 6.90, p < .02$), and an interaction between condition, anterior–posterior, and hemisphere ($F(1,22) = 4.64$, $p < .05$). A follow-up ANOVA with frontal ROIs indicated an effect of condition ($F(1,22) = 8.82, p < .01$), with no interaction between condition and hemisphere ($p = .33$), or between condition and group ($p = .98$). Analogous ANOVAs for the intermediate time window (300–550 ms) did not yield an effect of condition (or any interaction between factors), either when computing four ROIs, or when computing two frontal ROIs (Table S1, Final chord). Therefore, irregular terminal chords did not evoke a single tonic effect, but did evoke distinct early and late negative effects.

The local transition probability between the last chord of the first phrase (see the dominant with the fermata in Fig. 1*A*) and the first chord of the second phrase was lower for modified compared with original versions (*SI Text*). The ERPs of the first chord of the second phrase show that this local effect evoked an early anterior negativity (being maximal at around 200 ms), and a later positivity that was maximal at around 500 ms, and broadly distributed over the scalp (Fig. 2*B*). A global ANOVA for a time window from 200 to 300 ms (early negativity) indicated an effect of condition ($F(1,22) = 5.10, p = .03$) and an interaction between condition and hemisphere ($F(1,22) = 4.84, p < .05$). A follow-up ANOVA with frontal ROIs (with factors condition, hemisphere, and group) indicated an effect of condition ($F(1,22) = 6.28$, $p = .02$), with no interaction between condition and hemisphere ($p = .11$) or between condition and group ($p = .14$; see also amplitude values provided in Table S1, First chord of second phrase). A global ANOVA for a time window from 400 to 500 ms (later positivity) indicated an effect of condition ($F(1,22)=4.91, p<.04$), with no interaction between factors.

To exclude the possibility that these ERP effects (evoked due to the difference in local transition probability between phrases) were simply propagated up to the last chord, or that ERP effects evoked by the last chord were simply a residual of a prolonged effect evoked by the transition between first and second phrase, we also compared brain electric responses to the penultimate chord between original and modified versions. In contrast to ERPs of the final chords, there was no sampling point that showed more negative potentials in reponse to modified, compared with original, versions during the penultimate chord (see Fig. S6*A*; for statistics see Table S1, Penultimate chord). In addition, we sought to exclude the possibility that effects of the last chord were simply a sensory effect or simply an effect of a possible reactivation of the representation of the initial chords or key. Therefore, we also analyzed the tonic chords that were presented in the closing cadence before the final tonics (*Materials and Methods*). Again, modified versions did not show any sampling point at which ERPs were more negative than those of original versions (see Fig. S6*B*; for statistics see Table S1, Prefinal tonic). These findings rule out the possibility that ERP effects elicited by final chords of modified versions (compared with original versions) were due to sensory factors or due to the reactivation of the initial key. Such effects should have been larger on the penultimate chords and prefinal tonics because these chords occurred earlier in time than final tonics and should therefore have evoked even larger effects.

During the EEG session, both musicians and nonmusicians detected 97% of the timbre deviants. The conclusiveness ratings obtained after the EEG session were higher for original than for modified versions in both nonmusicians [original: mean (M) = $7.11, SEM=.35$; modified: $M=6.85, SEM=.39$] and musicians (original: $M=8.0, SEM=.31$; modified: $M=7.7; SEM=.39$). An ANOVA on the conclusiveness ratings with factors version (original, modified) and group (nonmusicians, musicians) indicated a significant effect of version [$F(1,22)=3.09, p<.05$, one-sided according to the directed hypothesis], with no interaction between factors ($p=.87$). Analogous ANOVAs for valence and arousal ratings (also obtained after the EEG session) did not indicate any significant difference between original and modified versions ($p>.40$ in all tests) or any interaction between factors ($p>.25$ in all tests; see Table S2 for details).

Applying the source attribution method (35) (*Materials and Methods*), we also assessed participants' awareness of their knowledge guiding conclusiveness ratings. Of the 288 conclusiveness ratings obtained in total (each of the 24 participants rated six original and six modified stimuli), only one rating (0.3%) was based on knowledge of the piece (provided by a musician). Sixty-five ratings (23%) were based on knowledge of the rule, 51 of which were given by musicians (14 by nonmusicians). Two hundred three ratings (70%) were based on intuition, 111 of which were given by nonmusicians (92 by musicians). Nineteen conclusiveness ratings (7%) were based on guessing, all of which were given by nonmusicians. When considering only conclusiveness ratings that were based on intuition or on guessing, no significant difference was found between the means of ratings for original and modified versions (intuition: $p=.14$; guessing: $p=.2$; both intuition and guessing: $p=.17$). By contrast, conclusiveness ratings based on knowledge of the rule significantly differed between original and modified versions ($p<.05$). A χ^2 test showed that ratings of musicians were overrepresented in the category "knowing the rule" ($p<.0001$).

Discussion

Both electrophysiological and behavioral data show that final chords of stimuli were processed differently, depending on whether or not the final chord closed the hierarchical structure of the harmonic sequence (that is, whether or not the final chord prolonged the first chord; see solid line with arrows in Fig. 1 and Fig. S1). This finding shows that listeners apply cognitive processes that are capable of dealing with long-distance dependencies resulting from hierarchically organized syntactic structures. Our experimental manipulation kept the local structure of the second phrase of sequences identical while manipulating the hierarchical structure by establishing irregular long-distance dependencies between the first and second phrases (see the Introduction and *Materials and Methods*). As will be discussed in more detail below, local models such as Markov models do not plausibly account for this difference. According to the Markov assumption, the probability of the event e_i in a sequence is modeled such that it depends only on the previous $n-1$ elements in the sequence: $p(e_i|e_1^{i-1}) \approx p(e_i|e_{i-(n-1)}^{i-1})$ (33) (in which e_a^b denotes the subsequence $e_a, \ldots, e_b$). Accordingly, nonlocal elements beyond the context length $n-1$ do not affect the prediction of e_i. Therefore, the differences in perception and brain responses observed in our data between regular and irregular sequence endings reflect hierarchical processing involving, e.g., the representation and application of a context-free phrase-structure rule that mandates a nonlocal dependency (such as the tonic prolongation and dominant–tonic implication as described by the GTTM or GSM).

ERPs evoked by hierarchically irregular final chords revealed an early frontal negativity emerging at around 150 ms after stimulus onset, which was maximal at around 220 ms. This observation shows that hierarchically structured harmonic long-distance dependencies are processed as early as about 150–200 ms after the onset of a chord. Notably, this effect was observable even though the attentional focus of participants was not directed on the experimental manipulations (participants watched a silent video and detected the timbre deviants, without being informed about our experimental manipulation). This early negativity is reminiscent of the early right anterior negativity (ERAN) (28) although it was not lateralized to the right (amplitude values were nominally larger over left anterior leads, but this hemispheric weighting was statistically not significant). Previous studies reported similar ERP responses with no hemispheric weighting (36) or even slight (statistically nonsignificant) left-hemispheric weighting (37, 38). More importantly, all previous studies reporting ERAN responses used music–syntactic irregularities that involved both local and nonlocal dependencies (28, 36–38). Therefore, it was not clear whether the ERAN was evoked by local, or hierarchical dependencies, or both. Our data show that an ERAN-like response can be evoked by irregularities that are hierarchical in nature, in the absence of local irregularities.

This early negativity was followed by a later ERP response that is reminiscent of the N5. Both early and late ERP effects were separated by a time interval in which there was no significant difference between original and modified chords. The scalp distribution of the N5 was more anterior than that of the earlier effect, consistent with previous studies (28, 36). The N5 is taken to reflect processes of harmonic integration; in the present study, the N5 evoked by irregular final chords probably reflects the attempt to harmonically integrate a chord that terminates the sequence without closing the hierarchical structure of the sequence.

The ERP responses are consistent with the behavioral results, which also showed significant differences between original and modified sequences. The behavioral ratings (obtained after the EEG session) indicate that participants perceived the original versions as slightly more conclusive than the modified versions. Source attribution ratings suggest that this effect was mainly due to explicit judgment knowledge of some participants (35), rather than due to implicit knowledge. Conclusiveness ratings significantly differed between original and modified versions when participants indicated that their conclusiveness judgment was based on their knowing the rule that differentiated modified from original versions. Conclusiveness ratings did not differ between conditions when participants indicated that they based their rating on intuition or guessing. This finding is in agreement with explicit judgment knowledge found in musical-learning studies (39). Explicit judgment knowledge does not necessarily imply that individuals had explicit structural knowledge (i.e., that they actually knew the rule), but that they were aware of knowledge that guided their responses (35). Analogously, native speakers of a language may detect an ungrammatical sentence with confidence but are often not able to explicitly state the rule.

It is highly unlikely that the behavioral and ERP effects observed in our study were due to local processing (e.g., due to the application of an n-gram model): in Bach chorales, harmonic n-grams obey a Zipf distribution, and even 4- and 5-gram models are extremely sparse (40). That is, very few sequences appear relatively frequently, whereas most remaining sequences appear rarely, or even only once. If the effects observed in the present study were due to local processing, then participants must have processed at least 9-grams (in BWV 302) or even 10-grams (in BWV 373). However, 9- and 10-grams will be unique, even in a very large corpus, and therefore they could have been detected only if participants heard and memorized the chorales before the experiment (which was not the case in our sample). Consequently, our data show that participants applied cognitive processes that are capable of dealing with nonlocal dependencies. This conclusion is substantiated by several observations. First, the local difference between original and modified versions at the beginning of the second phrase (after the fermata; see Fig. 1) evoked an early negative and a later positive ERP effect. These effects can be explained by the different transition probabilities, as well as by possible sensory differences, at this point of the sequences. The ERP effects, however, did not propagate to subsequent events. This was demonstrated by the ERPs of both penultimate chords and prefinal tonics, which did not show any significant ERP effect of modified compared with original versions. Second, although penultimate chords and prefinal tonics did not evoke any significant ERP effects, the hierarchical irregularities at the end of the sequences evoked negative ERP effects. This finding shows that these ERP effects were not due to local or sensory processing or to reactivation of sensory memory traces. None of the negative effects evoked by irregular final chords was observable already before the onset of the final chord. If the ERP effects evoked on the final chords were simply due to such local or sensory factors, then they should have been observed even more strongly on previous chords (which was not the case). Particularly, the observation that the prefinal tonic chords (which were acoustically comparable with the final chords) did not evoke any negative effect renders it highly unlikely that effects evoked by the final chords were simply due to auditory sensory memory processes. Third, the ERP effects evoked by the final chords did not simply reflect a cortical reactivation of a representation of key established by the first chords (41) because such reactivation should already have occurred during the processing of the prefinal tonic, or the penultimate chord.

The processing of the hierarchical structure (involving long-distance dependencies) requires working memory (WM) to establish and maintain a representation of the hierarchical structure. Note that original and modified versions had the same length of dependency between first and final chord. Therefore, original and modified versions had identical WM load, and the ERP effects evoked by the final chords of the modified versions cannot simply reflect WM operations only. During the EEG session, participants could have actively held the pitch information of the first chord in their WM and then compared the pitches of the last chord against this memory template. However, this would have required considerable conscious effort on the part of the subjects, and it is unlikely that subjects made such efforts. Participants were instructed to enjoy the silent movie while performing the timbre detection task, and it was easier for participants to merely follow this instruction (notably, none of the participants reported use of such a WM strategy during the debriefing). In addition, no ERP difference was found between musicians and nonmusicians although musicians perform considerably better on such pitch-memory tasks (42, 43).

We assume that previous experiments in which even musically trained listeners were perceptually rather insensitive to drastic manipulations of large-scale musical structure (21, 22) have not found comparable effects for several possible reasons. (*i*) In line with local theories, single exposure to a musical piece may result in only a partial parse of the hierarchical structure whereas multiple listening (as in our study) probably gradually leads to the establishment of representations of more complex dependencies within the musical piece. Such complex dependencies are difficult to learn, and their representation becomes more cognitively demanding the longer the musical dependencies are; this notion is supported by implicit learning research (44) and bevahioral reports on this topic (45). (*ii*) Perhaps EEG is more sensitive (and potentially more direct) than behavioral measures. Previous studies showed recognition of harmonic and melodic reductions, which are predicted by syntactic theories of music like the GTTM or GSM (29, 30). However, those studies did not show processing of long-distance dependencies whereas the present data demonstrate processing of nonlocal, hierarchically organized musical dependencies.

Note that our data show processing of long-distance dependencies that are the result of underlying hierarchically embedded structures. Corroborating syntactic theories of music (13–15), our findings suggest processing of hierarchical structures that operates similarly on different levels of the hierarchy. The structures are predicted by the application of two generative rules (tonic prolongation and dominant–tonic implication) that operate on both local levels (e.g., in a cadence) and nonlocal levels (as in our stimulus material; see arrows in Fig. 1 and Fig. S1). Recursive processing of hierarchical structures in music is consistent with the notion that the linguistic capacities for recursive syntactic processing are shared with music (27) (whether the human brain processes more than one instance of recursively nested center embedding in music needs to be tested in the future). Our findings lend plausibility to the assumption that hierarchical processing is also engaged during the processing of local dependencies, such as when processing a short chord sequence, even though such dependencies can theoretically be processed using local models only. Thus, the ERAN observed in previous studies using chord-sequence paradigms (as well as the ERAN evoked by the first chord of the second phrase of modified versions) (Fig. 2*B*) is probably a conglomerate of potentials due to local processing on the one hand and hierarchical processing on the other.

Our results are important for several reasons. First, they show that music listeners apply cognitive processes that are capable of dealing with nonlocal, hierarchically organized musical dependencies, even without explicit structural knowledge of the underlying syntactic rules. Long-distance dependencies are common in everyday language and can be identified theoretically in most pieces of tonal music. Our data demonstrate that such dependencies have a reality in the mental representation of music listeners, showing that music listeners process long-distance dependencies that are the result of underlying hierarchical and recursive syntactic structure. Second, our data show ERP correlates of syntactic processing involving different time scales (local and nonlocal). Thus, nested processing on different time scales is required to fully grasp the structure of the hierarchically organized sequential information used in our study. This notion challenges approaches in cognitive and brain science that aim at explaining processing of sequential information based on local models only. Third, our results show that a key component of human language, namely processing of hierarchical syntactic structure with nested long-distance dependencies, is engaged during listening to music, and thus is not unique to language. Therefore, our data indicate that representation and processing of information within a temporal hierarchy established by local and nested nonlocal dependencies is a multidomain capacity of human cognition. This finding sheds unique light on the much-debated overlap of music and language as communicative systems (27, 46–53) because our data indicate that both music and language make use of more general resources for the processing of hierarchically organized information than previously believed. Because hierarchical structures of many musical pieces (up to entire movements of a symphony) exceed by far the structural complexity even of the most elaborate sentences, it is tempting to speculate that the human ability to process hierarchical structure in music might be more powerful than linguistic syntax, often considered to be the paragon of human cognitive complexity.

Materials and Methods

Participants. Twelve nonmusicians and 12 musicians without absolute pitch participated in the study (age range 23–39 y, $M = 27.7$; 6 females in each group) (*SI Text*).

Stimuli and Procedure. Original and modified versions were transposed to the twelve major keys, and all stimuli were presented five times in pseudorandomized order with a tempo of 100 beats per minute (*SI Text*). Participants listened to the stimuli through headphones while watching a silent movie without subtitles. The task for the subjects was to monitor the timbre of the musical stimuli and detect infrequently occurring timbre deviants by pressing a response button. Subjects were not informed about the fact that there were original and modified versions of the chorales.

After the EEG session, participants were presented with twelve of the experimental stimuli. After each stimulus, participants rated the ending of each stimulus using nine-point scales with regard to (*i*) its conclusiveness ("How well did the final chord close the entire sequence?"), (*ii*) its valence ("How pleasant/unpleasant did you feel the final chord to be?"), and (*iii*) the degree of physiological arousal evoked by the final chord ("How calming/exciting did you feel the final chord to be?"). Moreover, participants indicated whether their conclusiveness rating was based on (*i*) guessing, (*ii*) their intuition, (*iii*) knowing the rule, or (*iv*) knowing the piece) (*SI Text*).

EEG Recordings and Data Analysis. Continuous EEG data were recorded from 64 electrodes. After filtering and artifact rejection (*SI Text*), data were rereferenced to the algebraical mean of left and right mastoid leads. Grand-average ERPs were computed for the last chord, the first chord of the second phrase (i.e., the chord directly succeeding the chord with the fermata) (Fig. 1 and Fig. S1), the penultimate chords (i.e., the second-to-last chords of the entire sequences), and the prefinal tonics. Prefinal tonics were the tonic chords presented in the closing cadence before the final tonics (for BWV 373, see the *G* depicted in the fourth-to-last leaf in the bottom row of Fig. S2; for BWV 302, see the *D* depicted in the third-to-last leaf in the bottom row of Fig. S4). For the statistical analysis of ERPs, four regions of interest (ROIs) were computed: left anterior, right anterior, left posterior, and right posterior. Global ANOVAs were computed with the within-subject factors condition (original, modified), hemisphere (left, right ROIs), and anterior–posterior distribution (anterior, posterior ROIs), and the between–subjects factor group (musicians, nonmusicians). For additional statistical analyses, see Table S1.

ACKNOWLEDGMENTS. We thank Shuang Guo for assistance in data analysis and W. Tecumseh Fitch as well as Bruno Gingras for valuable discussion.

1. Singer W (1995) Development and plasticity of cortical processing architectures. *Science* 270(5237):758–764.
2. Giraud AL, Poeppel D (2012) Cortical oscillations and speech processing: Emerging computational principles and operations. *Nat Neurosci* 15(4):511–517.
3. Näätänen R, et al. (1997) Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature* 385(6615):432–434.
4. Baddeley A (2003) Working memory: Looking back and looking forward. *Nat Rev Neurosci* 4(10):829–839.
5. Schenker H (1956) *Neue Musikalische Theorien und Phantasien: Der Freie Satz* (Universal Edition, Vienna), 2nd Ed.
6. Salzer F (1962) *Structural Hearing: Tonal Coherence in Music* (Dover, Mineola, NY), Vol 1.
7. Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: What is it, who has it, and how did it evolve? *Science* 298(5598):1569–1579.
8. Chomsky N (1995) *The Minimalist Program*. Current Studies in Linguistics, ed Keyser SJ (MIT Press, Cambridge, MA), Vol 28.
9. Fitch WT, Hauser MD (2004) Computational constraints on syntactic processing in a nonhuman primate. *Science* 303(5656):377–380.
10. Friederici AD, Bahlmann J, Heim S, Schubotz RI, Anwander A (2006) The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proc Natl Acad Sci USA* 103(7):2458–2463.
11. Nevins A, Pesetsky D, Rodrigues C (2009) Pirahã exceptionality: A reassessment. *Language* 85:355–404.
12. Hofstadter DR (1979) *Gödel, Escher, Bach* (Basic Books, New York).
13. Lerdahl F, Jackendoff R (1983) *A Generative Theory of Tonal Music* (MIT Press, Cambridge, MA).
14. Steedman MJ (1984) A generative grammar for jazz chord sequences. *Music Percept* 2(1):52–77.
15. Rohrmeier M (2011) Towards a generative syntax of tonal harmony. *J Math Music* 5:35–53.
16. Jackendoff R, Lerdahl F (2006) The capacity for music: What is it, and what's special about it? *Cognition* 100(1):33–72.
17. Huron DB (2006) *Sweet Anticipation: Music and the Psychology of Expectation* (MIT Press, Cambridge, MA).
18. Tymoczko D (2011) *A Geometry of Music: Harmony and Counterpoint in the Extended Common Practice* (Oxford Univ Press, New York).
19. Levinson J (1997) *Music in the Moment* (Cornell Univ Press, Ithaca, NY).
20. Tillmann B, Bigand E (2004) The relative importance of local and global structures in music perception. *J Aesthet Art Crit* 62:211–222.
21. Tillmann B, Bigand E, Madurell F (1998) Local versus global processing of harmonic cadences in the solution of musical puzzels. *Psychol Res* 61:157–174.
22. Cook N (1987) The perception of large-scale tonal closure. *Music Percept* 5:197–205.
23. Bigand E, Madurell F, Tillmann B, Pineau M (1999) Effect of global structure and temporal organization on chord processing. *J Exp Psychol Hum Percept Perform* 25: 184–197.
24. Patel AD, Gibson E, Ratner J, Besson M, Holcomb PJ (1998) Processing syntactic relations in language and music: An event-related potential study. *J Cogn Neurosci* 10(6): 717–733.
25. Maess B, Koelsch S, Gunter TC, Friederici AD (2001) Musical syntax is processed in Broca's area: An MEG study. *Nat Neurosci* 4(5):540–545.
26. 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.
27. Patel AD (2003) Language, music, syntax and the brain. *Nat Neurosci* 6(7):674–681.
28. Koelsch S (2012) *Brain and Music* (Wiley, New York).
29. Serafine ML, Glassman N, Overbeeke C (1989) The cognitive reality of hierarchic structure in music. *Music Percept* 6(4):397–430.
30. Dibben N (1994) The cognitive reality of hierarchic structure in tonal and atonal music. *Music Percept* 12:1–25.
31. Lerdahl F, Krumhansl CL (2007) Modeling tonal tension. *Music Percept* 24:329–366.
32. Rohrmeier M, Graepel T (2012) Comparing feature-based models of harmony. Proceedings of the 9th International Symposium on Computer Music Modelling and Retrieval. pp 357–370. Available at http://cmmr2012.eecs.qmul.ac.uk/sites/cmmr2012.eecs.qmul.ac.uk/files/pdf/papers/cmmr2012_submission_95.pdf. Accessed August 19, 2013.
33. Pearce M, Wiggins G (2004) Improved methods for statistical modelling of monophonic music. *J New Music Res* 33:367–385.
34. Krumhansl CL, Cuddy LL (2010) A theory of tonal hierarchies in music. *Music Percept* 36:51–87.
35. Dienes Z, Scott R (2005) Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychol Res* 69(5-6):338–351.
36. Loui P, Grent-'t-Jong T, Torpey D, Woldorff M (2005) Effects of attention on the neural processing of harmonic syntax in Western music. *Brain Res Cogn Brain Res* 25(3):678–687.
37. 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.
38. Steinbeis N, Koelsch S, Sloboda JA (2006) The role of harmonic expectancy violations in musical emotions: Evidence from subjective, physiological, and neural responses. *J Cogn Neurosci* 18(8):1380–1393.
39. Rohrmeier M, Rebuschat P (2012) Implicit learning and acquisition of music. *Topics Cogn Sci* 4(4):525–553.
40. Rohrmeier M, Cross I (2008) Statistical properties of tonal harmony in Bach's chorales. *Proceedings of the 10th International Conference on Music Perception and Cognition*, pp 619–627. Available at http://www.mus.cam.ac.uk/~ic108/PDF/MP081391.PDF. Accessed August 19, 2013.
41. Janata P, et al. (2002) The cortical topography of tonal structures underlying Western music. *Science* 298(5601):2167–2170.
42. Deutsch D (1999) *The Psychology of Music* (Academic, New York).
43. Schulze K, Zysset S, Mueller K, Friederici AD, Koelsch S (2011) Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum Brain Mapp* 32(5):771–783.
44. Kuhn G, Dienes Z (2005) Implicit learning of nonlocal musical rules: Implicitly learning more than chunks. *J Exp Psychol Learn Mem Cogn* 31(6):1417–1432.
45. Woolhouse M, Cross I, Horton T (2006) The perception of nonadjacent harmonic relations. *International Conference on Music Perception and Cognition*, Vol 9. Available at http://www.mus.cam.ac.uk/~ic108/PDF/204.pdf. Accessed August 19, 2013.
46. Longuet-Higgins HC (1979) The perception of music. *Proc R Soc Lond B Biol Sci* 205(1160):307–322.
47. Bernstein L (1981) *The Unanswered Question: Six Talks at Harvard* (Harvard Univ Press, Cambridge, MA).
48. Peretz I, Coltheart M (2003) Modularity of music processing. *Nat Neurosci* 6(7):688–691.
49. Fitch WT (2006) The biology and evolution of music: A comparative perspective. *Cognition* 100(1):173–215.
50. Ross D, Choi J, Purves D (2007) Musical intervals in speech. *Proc Natl Acad Sci USA* 104(23):9852–9857.
51. Cross I (2009) The evolutionary nature of musical meaning. *Music Sci* 13:179–200.
52. Zatorre RJ, Baum SR (2012) Musical melody and speech intonation: Singing a different tune. *PLoS Biol* 10(7):e1001372.
53. Han Se, Sundararajan J, Bowling DL, Lake J, Purves D (2011) Co-variation of tonality in the music and speech of different cultures. *PLoS ONE* 6(5):e20160.

Supporting Information

Koelsch et al. 10.1073/pnas.1300272110

S1 Text

Participants. Twelve musicians and 12 nonmusicians participated in the study (age-range 23–39 y, $M = 27.7$, 6 females in each group). Musicians were recruited from the Universität der Künste Berlin and had at least 10 y of formal musical training. Exclusion criteria included past or present neurological or psychiatric disorders. Only musicians without absolute pitch were admitted to the study, and nonmusicians were admitted only if they had not received any formal musical training outside of normal school education. All participants were right-handed and had normal hearing (according to self-report). Written informed consent was obtained from all subjects; the study was conducted according to the Declaration of Helsinki and approved by the ethics committee of the Psychology Department of the Freie Universität.

Stimuli. We used the first two phrases of two chorales by J. S. Bach (BWV 373 and BWV 302, both in major keys) (Fig. 1, Fig. S1, and Audio File S1), henceforth referred to as original versions. In both chorales, these two phrases consisted of five bars, the first phrase ending with a half cadence (i.e., on the dominant), the second beginning with a chord other than the tonic (thus not immediately fulfilling the implication of the dominant at the end of the first phrase) and ending on the initial tonic by means of an authentic cadence. Therefore, according to the GTTM and GSM, the final chord of each original version hierarchically prolonged the first chord of the chorale (and closed the established dominant that remained open at the end of the first phrase). From the original versions, modified versions were created. As illustrated by the red scores in Fig. 1*B* and Fig. S1*B*, the modified versions were created such that the first phrase was transposed down a fourth (BWV 373) (Audio File S2) or up a major second (BWV 302). Thus, the final chord of the second phrase of each modified version did not prolong the first chord of the chorale anymore and, furthermore, did not close the dominant established by the first phrase. Importantly, the second phrase of original and modified version was identical (compare Fig. 1*A* and Fig. 1*B*, as well as Fig. S1*A* and Fig. S1*B*). Therefore, the local probabilities for the transition between penultimate and final chords were equal in both original and modified versions, and the manipulation of our stimulus material led only to an irregular long-distance dependency.

Transition Probabilities Between Last Chord of First Phrase and First Chord of Second Phrase. The probabilities for the local transition between the last chord of the first phrase (see the dominant with the fermata in Fig. 1) and the subsequent chord, as estimated from a corpus analysis of Bach chorales (1), was 0.07 for each of the two original versions (dominant–submediant progression), 0.03 for the modified version of BWV 373 (dominant–supertonic), and 0.001 for the modified version of BWV 302 (dominant–minor dominant). Thus, although the transition between first and second phrase was plausible in both original and modified versions, the local transition probabilities were lower in the modified versions compared with the original versions (these probabilities did not, however, necessarily correspond to the actual expectancies of our participants).

Stimulus Processing. Using musical instrument digital interface (MIDI) format, the two original versions and the two modified versions were transposed to the twelve major keys, and exported as wav files with a piano sound and a tempo of 100 beats per minute (600 ms per quarter note) using Sibelius 6.2 software (Avid Tech. Inc.). To guarantee that the second phrases of both original and modified versions were acoustically identical, the second phrase of the wav file of each original version was copied and pasted as the second phrase of a corresponding modified version using Audacity 2.0 (audacity.sourceforge.net). This procedure resulted in 48 different experimental stimuli in total (2 chorales × 2 versions × 12 keys).

In addition to this stimulus set, each stimulus was also modified such that in one bar of the chorale one voice was not played with a piano sound, but with a bassoon sound. The bars with these timbre deviants were distributed equally among the bars across chorales, voices, and keys. These stimuli were used for a timbre detection task, and not included in the analysis of the event-related potentials (ERPs).

Experimental Procedure. During the electroencephalographic (EEG) recording session, participants listened to the stimuli presented with 60 dB sound pressure level (SPL) through headphones while watching a silent movie without subtitles (*March of the Penguins*, Warner, ASIN B000BI5KV0). The task for the subjects was to monitor the timbre of the musical stimuli, detect the timbre deviants, and indicate the detection of the timbre deviants by pressing a response button. Subjects were not informed about the original and the modified versions. Each of the 48 stimuli without timbre deviants was presented five times, randomly intermixed with 25 sequences containing a timbre deviant, amounting to 265 stimuli in total, and a duration of an experimental session of about 53 min. Stimuli were presented in pseudorandom order such that (*i*) each stimulus was presented in a key that differed from the key of the second phrase of the previous sequence, (*ii*) each chorale (BWV 302 or BWV 373) was maximally presented three times in a row (independently of whether it was an original or a modified version), and (*iii*) there were maximally three original or three modified versions presented in a row.

After the EEG session, participants were presented with a questionnaire to assess whether they could differentiate between the two versions of the chorales, and (if so) on which kind of knowledge such differentiation was based. Applying the source attribution method (2), it was addressed whether participants consciously knew that their answer was correct, whether they were guessing, or whether they were following their intuition. Moreover, to assess potential emotional effects of our experimental manipulation, we also used standard dimensional emotion measures of valence (pleasant/unpleasant) and physiological arousal (calm/excited) (3). Twelve of the stimuli used in the EEG session (6 originals and 6 modified versions from each chorale, each stimulus in a different key) were presented to participants. Using nine-point scales, participants rated the ending of each stimulus with regard to (*i*) its conclusiveness (”How well did the final chord close the entire sequence?”), (*ii*) its valence (”How pleasant/unpleasant did you feel the final chord to be?”), and (*iii*) the degree of physiological arousal evoked by the final chord (”How calming/exciting did you feel the final chord to be?”). Scales ranged from 1 (low conclusiveness, low valence, and low arousal) to 9 (high conclusiveness, high valence, and high arousal). Finally, for each stimulus participants indicated whether their conclusiveness rating was based on (*i*) guessing, (*ii*) their intuition, (*iii*) knowing the rule, or (*iv*) knowing the piece.

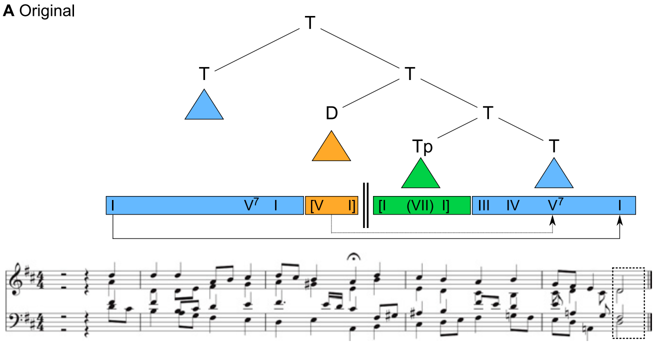
EEG Recordings and Data Analysis. Continuous EEG data were recorded from 64 electrodes (extended 10–20 system), referenced to M1. Four electrodes were used for recording the electrooculogram (EOG): two electrodes were placed above and below the left eye to record the vertical EOG, and two electrodes were positioned at the outer canthus of each eye to record the horizontal EOG. Impedance was kept below 5 kΩ, and sampling rate was 500 Hz (low and high cut off was direct current and 1,000 Hz, respectively).

Data were analyzed offline using the EEGLAB Toolbox (4). To remove slow waves (such as electrode saturation or drifts), raw data were filtered with a 0.25-Hz high-pass filter with finite impulse response (FIR) and a filter order of 13,750 points. Then, data were filtered with a 49- to 51-Hz band-stop filter (FIR, 2,750 points) to eliminate line noise. An Independent Component Analysis (ICA) was carried out, and components representing artifacts (eye blinks, eye movements, and muscular activity) were removed. Afterward, data were filtered with a 25-Hz low-pass filter (FIR, 550 points) to remove remaining high-frequency noise (such as muscle activity that was not removed using the ICA). Subsequently, data were epoched. To remove further possible artifacts, sampling points were rejected whenever the SD of a 200-ms or 800-ms gliding window exceeded 25 μV at any EEG electrode. Then, data were rereferenced to the algebraical mean of left and right mastoid leads. Finally, using a baseline from −200 to 0 ms, nonrejected epochs were averaged for the last chord from −200 to 1,200 ms relative to the onset of final chords, or from −200 to 600 ms relative to the onset of (*i*) the first chord of the second phrase (i.e., the chord directly succeeding the chord with the fermata) (Fig. 1 and Fig. S1), (*ii*) the onset of the penultimate chords (i.e., the second-to-last chords of the entire sequences), and (*iii*) the prefinal tonics. Prefinal tonics were the tonic chords presented in the closing cadence before the final tonics (for BWV 373, see the G depicted in the fourth-to-last leaf in the bottom row of Fig. S2; for BWV 302, see the D depicted in the third-to-last leaf in the bottom row of Fig. S4).

For the statistical analysis of ERPs, four regions of interest (ROIs) were computed: left anterior (AF3, F1, F3, F5, C1, C3, C5), right anterior (AF4, F2, F4, F6, C2, C4, C6), left posterior (CP1, CP3, CP5, P1, P3, P5, PO3), and right posterior (CP2, CP4, CP6, P2, P4, P6, PO4). Then, global ANOVAs were conducted with the within-subject factors condition (original, modified), hemisphere (left, right ROIs), and anterior–posterior distribution (anterior, posterior ROIs), and the between-subjects factor group (musicians, nonmusicians). The time window for statistical analysis of the ERAN was 150–300 ms, and for the N5, 550–850 ms. Additional statistical analyses are provided in Table S1.

1. Rohrmeier M, Cross I (2008) Statistical properties of tonal harmony in Bach's chorales. *Proceedings of the 10th International Conference on Music Perception and Cognition*. Available at http://www.mus.cam.ac.uk/~ic108/PDF/MP081391.PDF. Accessed August 18, 2013.
2. Dienes Z, Scott R (2005) Measuring unconscious knowledge: Distinguishing structural knowledge and judgment knowledge. *Psychol Res* 69(5-6):338–351.
3. Bradley MM, Lang PJ (1994) Measuring emotion: The self-assessment manikin and the semantic differential. *J Behav Ther Exp Psychiatry* 25(1):49–59.
4. Delorme A, Makeig S (2004) EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods* 134(1):9–21.

A Original



B Modified

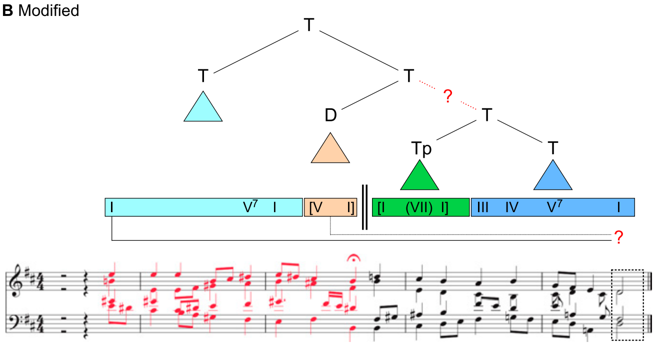


Fig. S1. Score and syntactic structure of J. S. Bach's chorale *Ein feste Burg ist unser Gott* (BWV 302). (*A*) The score of the original version. The tree structure above the score represents a schematic diagram of the harmonic dependencies (for full tree graphs according to the GSM and GTTM, see Figs. S4 and S5). Roman numerals indicate scale degrees. Triangles denote an abbreviation of dependency structures that are not depicted in detail. T, S, and D indicate the main tonal functions (tonic, subdominant, dominant) of the respective part of the sequence (such as functional regions in the GSM). Squared brackets indicate scale degrees relative to the local key (here, the local key of B minor is a submediant region of the initial key D major; see green rectangle). The two thick parallel vertical lines (separating the first and the second phrase) visualize the fact that the local dominant is not immediately followed by a resolving tonic chord, but implies its resolution with the final tonic (indicated by the dotted arrow). The same dependency exists between intial and final tonic (indicated by the solid arrow). The tree thus illustrates the nonlocal (long-distance) dependency between the initial and final tonic regions and tonic chords, respectively (also illustrated by the blue rectangles). The chords belonging to a key other than the initial key (green rectangle) represent one level of embedding. (*B*) The score of the modified version, in which the first phrase (red color) was shifted a major second upwards. Note that the second phrase (beginning after the fermata) was acoustically identical to the second phrase of the original version. The tree structure above the scores illustrates that the second phrase is not compatible with an expected tonic region (indicated by the red dotted line with the red question mark), and that the last chord (a tonic of a local cadence, dotted rectangle) neither prolongs the initial tonic nor closes the open dominant (see solid and dotted lines followed by red question mark).

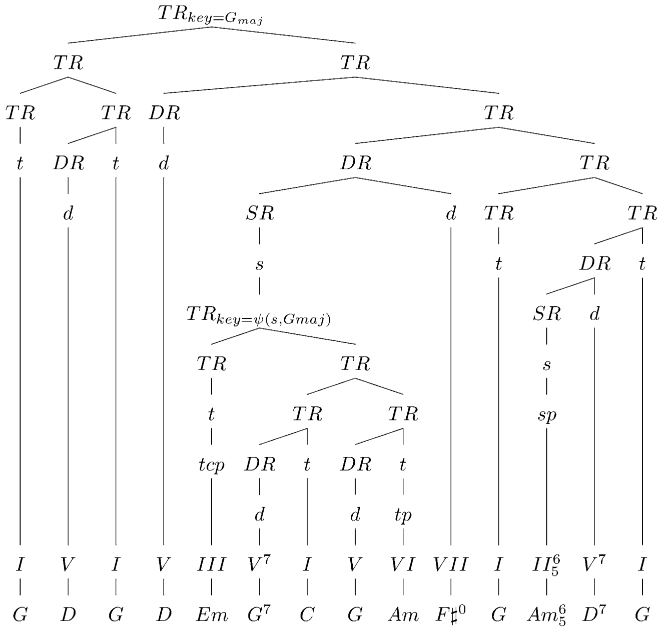


Fig. S2. Analysis of the first two phrases of the chorale *Liebster Jesu, wir sind hier* (harmonized by J. S. Bach, BWV 373) according to the Generative Syntax Model (GSM) (1, 2). For scores and abbreviated analysis, see Fig. 1. The phrase-structure level (top) is indicated by the uppercase symbols (*TR*,*DR*,*SR*), the functional level is indicated by the lowercase letters (*t*,*s*,*d*,*tp*,*sp*,*dp*,*tcp*), the scale-degree level by Roman numeral notation, and the surface level by the chord symbols. *DR*, dominant region; *SR*, subdominant region; *TR*, tonic region.

1. Rohrmeier M (2011) Towards a generative syntax of tonal harmony. *J Math Music* 5:35–53.
2. Rohrmeier M (2007) A generative grammar approach to diatonic harmonic structure. *Proceedings of the 4th Sound and Music Computing Conference*. Available at http://www.smc-conference.org/smc07/SMC07%20Proceedings/SMC07%20Paper%2015.pdf. Accessed August 19, 2013.

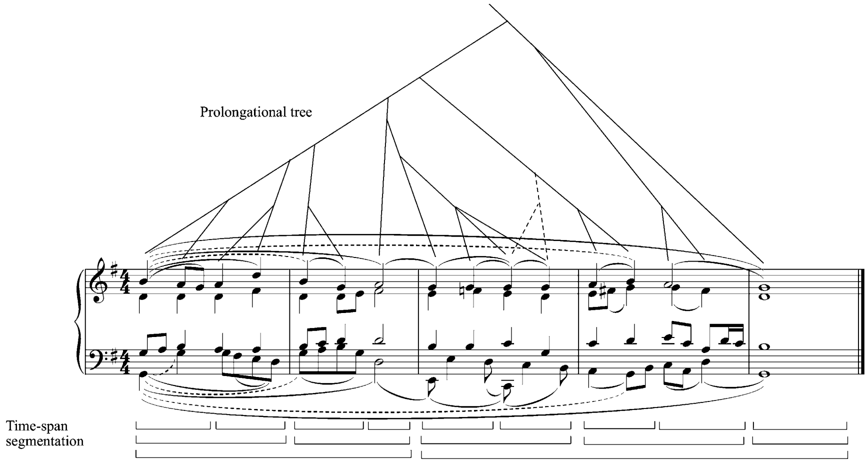


Fig. S3. Basic analysis of the first two phrases of the chorale *Liebster Jesu, wir sind hier* (harmonized by J. S. Bach, BWV 373) according to the Generative Theory of Tonal Music (GTTM) and Tonal Pitch Space theory (TPS) (1). The diagram represents a prolongational analysis (to which the syntactic analysis of the GSM is analogous). The dashed lines indicate double derivations of a pivot chord that can be analyzed as being dependent of two different subtrees. Courtesy of Fred Lerdahl.

1. Lerdahl F (2001) *Tonal Pitch Space* (Oxford Univ Press, New York).

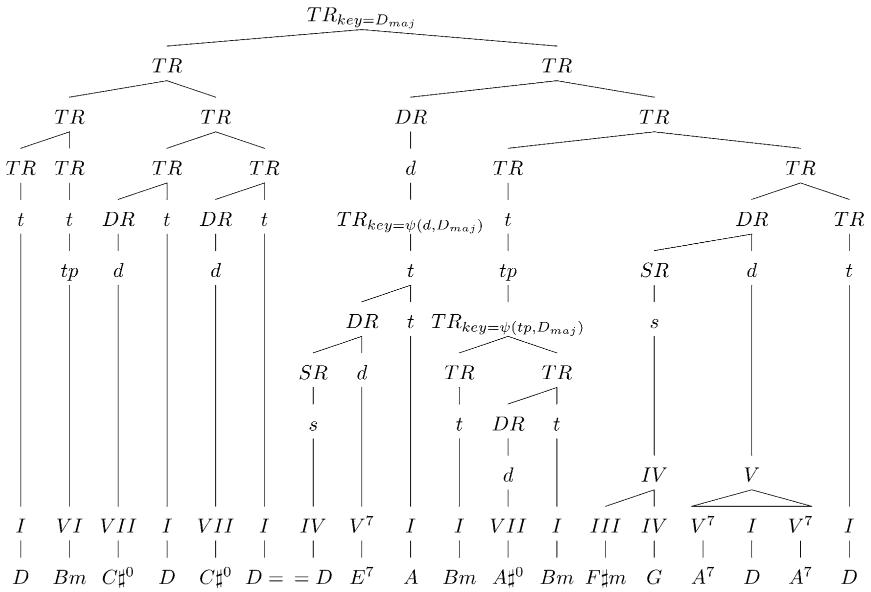


Fig. S4. Analysis of the first two phrases of the chorale *Ein feste Burg ist unser Gott* (harmonized by J. S. Bach, BWV 302) according to the Generative Syntax Model (GSM) (1, 2). For scores and abbreviated analysis, see Fig. S1). The = sign in the bottom row indicates that both instances of the D major chord refer to the identical surface pivot chord. The triangle visualizes a local dominant prolongation by a passing chord (I). The phrase-structure level (top) is indicated by the uppercase symbols (*TR*,*DR*,*SR*), the functional level is indicated by the lowercase letters (*t*,*s*,*d*,*tp*,*sp*,*dp*,*tcp*), the scale-degree level by Roman numeral notation, and the surface level by the chord symbols. *DR*, dominant region; *SR*, subdominant region; *TR*, tonic region.

1. Rohrmeier M (2011) Towards a generative syntax of tonal harmony. *J Math Music* 5:35–53.
2. Rohrmeier M (2007) A generative grammar approach to diatonic harmonic structure. *Proceedings of the 4th Sound and Music Computing Conference*, pp 97–100. Available at http://www.smc-conference.org/smc07/SMC07%20Proceedings/SMC07%20Paper%2015.pdf. Accessed August 19, 2013.

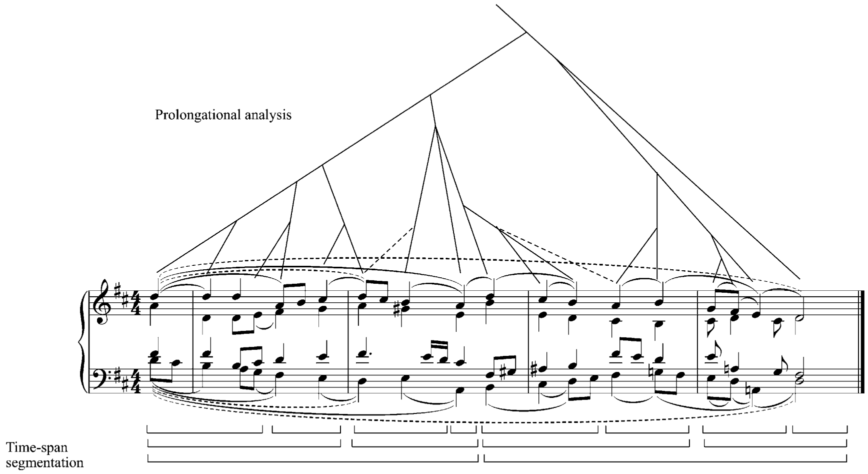


Fig. S5. Basic analysis of the first two phrases of the chorale *Ein feste Burg ist unser Gott* (harmonized by J. S. Bach, BWV 302) according to the Generative Theory of Tonal Music (GTTM) and Tonal Pitch Space theory (TPS) (1). The diagram represents a prolongational analysis (to which the syntactic analysis of the GSM is analogous). The dashed lines indicate double derivations of a pivot chord that can be analyzed as being dependent of two different subtrees. Courtesy of Fred Lerdahl.

1. Lerdahl F (2001) *Tonal Pitch Space* (Oxford Univ Press, New York).

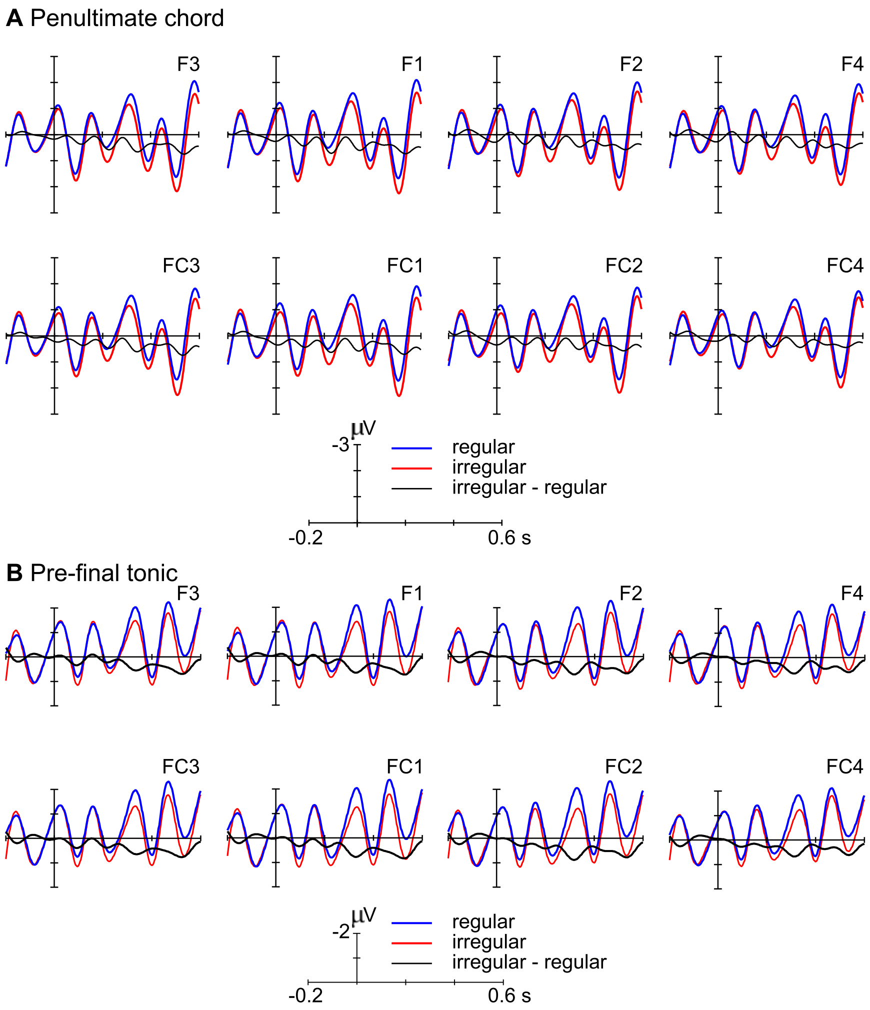


Fig. S6. Brain electric responses evoked by penultimate chords, i.e., the chords preceding the final tonic (*A*), and prefinal tonics, i.e., the tonic chord preceding the final tonic chord in the cadence ending the second phrase (*B*). The blue line indicates ERPs evoked by original versions, and the red line ERPs evoked by modified versions; the black line shows the difference wave (original subtracted from modified version). Note that eighth notes were presented in both BWV 302 and BWV 373; therefore, the ERPs show two P1, N1, and P2 waves (and each P2 is followed by another negative potential). ERPs of modified versions did not evoke any negative effect (compared with original versions; best to be seen in the difference wave), in contrast to the ERPs evoked by the final chords of modified versions.

Table S1. Observer-independent analysis of ERPs

	All subjects		Musicians		Nonmusicians	
Time window	Frontal L	Frontal R	Frontal L	Frontal R	Frontal L	Frontal R
Final chord						
0–100	0.08 (0.20)	0.23 (0.18)	0.02 (0.23)	0.09 (0.23)	0.14 (0.35)	0.36 (0.29)
100–200	0.45 (0.21)	0.40 (0.21)	0.21 (0.25)	0.16 (0.29)	0.69 (0.33)	0.65 (0.30)
200–300	**0.73 (0.19)**	**0.67 (0.16)**	**0.74 (0.29)**	**0.60 (0.27)**	**0.73 (0.25)**	**0.74 (0.19)**
150–300	**0.69 (0.18)**	**0.64 (0.16)**	**0.63 (0.26)**	**0.53 (0.26)**	**0.75 (0.26)**	**0.75 (0.21)**
300–400	0.49 (0.29)	0.52 (0.22)	0.60 (0.44)	0.61 (0.38)	0.38 (0.38)	0.43 (0.25)
400–500	0.46 (0.32)	0.45 (0.29)	0.56 (0.49)	0.48 (0.46)	0.36 (0.44)	0.43 (0.36)
500–600	0.40 (0.28)	0.37 (0.28)	0.57 (0.38)	0.49 (0.42)	0.23 (0.41)	0.25 (0.37)
600–700	0.43 (0.15)	0.26 (0.17)	0.35 (0.23)	0.20 (0.30)	0.50 (0.19)	0.31 (0.18)
700–800	**0.40 (0.15)**	**0.32 (0.15)**	**0.33 (0.18)**	**0.21 (0.18)**	**0.47 (0.24)**	**0.43 (0.25)**
550–850	**0.43 (0.13)**	**0.32 (0.14)**	**0.41 (0.18)**	**0.27 (0.24)**	**0.45 (0.18)**	**0.37 (0.16)**
800–900	0.27 (0.14)	0.22 (0.15)	0.41 (0.21)	0.34 (0.25)	0.13 (0.20)	0.11 (0.17)
900–1000	0.18 (0.15)	0.07 (0.20)	0.16 (0.20)	0.08 (0.28)	0.20 (0.23)	0.06 (0.31)
1,000–1,100	0.24 (0.14)	0.30 (0.16)	−0.10 (0.15)	−0.06 (0.16)	0.58 (0.19)	0.66 (0.23)
1,100–1,200	0.13 (0.22)	0.09 (0.21)	−0.02 (0.23)	−0.05 (0.27)	0.28 (0.37)	0.23 (0.31)
First chord of second phrase						
0–100	0.05 (0.19)	0.15 (0.17)	−0.25 (0.18)	−0.07 (0.16)	0.34 (0.32)	0.38 (0.30)
100–200	−0.06 (0.18)	0.06 (0.19)	−0.30 (0.17)	−0.16 (0.20)	0.18 (0.32)	0.29 (0.32)
200–300	**-0.61 (0.23)**	**−0.43 (0.21)**	**−1.00 (0.25)**	**−0.67 (0.18)**	**-0.22 (0.35)**	**−0.19 (0.38)**
150–300	**-0.53 (0.21)**	**−0.37 (0.20)**	**−0.92 (0.23)**	**−0.63 (0.18)**	**-0.14 (0.33)**	**−0.10 (0.35)**
300–400	0.11 (0.22)	0.22 (0.20)	−0.22 (0.23)	0.04 (0.21)	0.44 (0.35)	0.40 (0.33)
400–500	**0.40 (0.24)**	**0.43 (0.17)**	**0.07 (0.35)**	**0.29 (0.27)**	**0.74 (0.31)**	**0.57 (0.20)**
500–600	0.22 (0.27)	0.25 (0.22)	0.04 (0.37)	0.09 (0.27)	0.41 (0.39)	0.41 (0.36)
Penultimate chord						
0–100	−0.07 (0.10)	−0.05 (0.13)	−0.15 (0.15)	−0.06 (0.20)	0.00 (0.15)	−0.04 (0.16)
100–200	−0.14 (0.19)	−0.14 (0.17)	0.11 (0.20)	0.07 (0.22)	−0.38 (0.31)	−0.35 (0.26)
200–300	−0.26 (0.35)	−0.23 (0.28)	0.14 (0.30)	0.16 (0.26)	−0.66 (0.63)	−0.62 (0.48)
300–400	−0.39 (0.26)	−0.32 (0.21)	−0.47 (0.30)	−0.32 (0.23)	−0.31 (0.43)	−0.33 (0.36)
400–500	−0.26 (0.25)	−0.27 (0.23)	0.02 (0.31)	0.04 (0.34)	−0.54 (0.40)	−0.58 (0.31)
500–600	−0.54 (0.27)	−0.39 (0.24)	−0.42 (0.45)	−0.33 (0.43)	−0.66 (0.30)	−0.45 (0.23)
Prefinal tonic						
0–100	0.11 (0.17)	0.08 (0.17)	−0.25 (0.24)	−0.29 (0.22)	0.47 (0.19)	0.45 (0.20)
100–200	0.07 (0.15)	0.16 (0.18)	−0.05 (0.20)	−0.10 (0.23)	0.19 (0.21)	0.42 (0.26)
200–300	0.24 (0.22)	0.30 (0.25)	−0.12 (0.28)	−0.18 (0.33)	0.60 (0.31)	0.78 (0.33)
150–300	0.17 (0.19)	0.26 (0.22)	−0.13 (0.23)	−0.16 (0.28)	0.47 (0.27)	0.68 (0.30)
300–400	0.42 (0.27)	0.35 (0.28)	0.25 (0.41)	0.04 (0.38)	0.58 (0.37)	0.66 (0.41)
400–500	0.39 (0.33)	0.31 (0.31)	−0.11 (0.34)	−0.14 (0.34)	0.89 (0.53)	0.76 (0.49)
500–600	0.40 (0.30)	0.35 (0.27)	−0.30 (0.41)	−0.27 (0.35)	1.09 (0.35)	0.97 (0.33)

Mean amplitude values (with SD in parentheses) of differences between conditions (difference-potentials: original subtracted from modified versions). Potentials are provided separately for left frontal and right frontal regions of interest (ROIs), and separately for all subjects, musicians, and nonmusicians. The time windows (outermost left column) span in 100-ms steps the entire duration of the final chord (Final chord), the first chord of the second phrase, i.e., the chord directly following the chord with the fermata in Figs. 1 and S1 (First chord of second phrase), the penultimate chord (Penultimate chord), and the prefinal tonic, i.e., the tonic chord preceding the final tonic chord in the cadence ending the second phrase (Prefinal tonic). In addition, time windows reported in the main text are included. Bold font indicates that amplitude differences between original and modified versions were statistically significant at frontal ROIs ($p < .05$) as indicated by an effect of condition in an ANOVA with factors condition (original, modified), hemisphere, and group. None of these ANOVAs with frontal ROIs yielded any interaction between factors. In addition, for all time windows indicated in bold font, ANOVAs with four ROIs (left anterior, right anterior, left posterior, and right posterior) indicated interactions between condition and anterior-posterior, or between condition, anterior-posterior, and hemisphere ($p < .05$ in each test). None of such interactions was yielded for any other time window.

Table S2. Behavioral data (means, with SEM in parentheses)

Rating	Nonmusicians	Musicians
Conclusiveness		
Original	7.11 (0.35)	8.0 (0.31)
Modified	6.85 (0.39)	7.7 (0.39)
Valence		
Original	6.35 (0.37)	7.15 (0.28)
Modified	6.41 (0.35)	6.75 (0.44)
Arousal		
Original	3.1 (0.37)	3.25 (0.41)
Modified	3.08 (0.36)	3.19 (0.37)

Scales for ratings of conclusiveness, valence, and arousal ranged from 1 (very low) to 9 (very high).

Audio File S1. Original (hierarchically regular) version of J. S. Bach's chorale "Liebster Jesu, wir sind hier." For scores and detailed information, see legend of Fig. 1*A*.

Audio File S1

Audio File S2. Modified (hierarchically irregular) version of J. S. Bach's chorale "Liebster Jesu, wir sind hier." For scores and detailed information, see legend of Fig. 1*B*.

Audio File S2