The Perception of Melodies: Some Thoughts on Listening Style, Relational Thinking, and Musical Structure

Christiane Neuhaus

Every day, countless numbers of pop songs, ballads, and classical themes are broadcasted via radio stations, sung at schools, or rehearsed by professional ensembles. Given an almost infinite variety of melodies, we may ask which attributes they have in common, and what features we can extract. Composers, however, often decide for the word 'motif' rather than 'melody' to explain their musical ideas, partly due to different sequence length, and partly because both terms point to a different character of tone progression: A 'motif' has the potential for development because of clearly recognizable rhythmic and interval features, enabling us even to build the whole musical architecture. A 'melody', by contrast, is often considered a closed entity (or integrated whole) with elements kept in balance.

In the following chapter we focus on the perception of melodic structure. We first try to define the word 'melody' (Greek: *melōdía*, 'singing, chanting') from the viewpoint of music theory. In my opinion, four structural properties determine the melody's nature, namely 'interval type', 'melodic contour', 'balance between form parts', as well as the 'underlying harmonic framework' (based on the piece's tonality). The first three attributes describe a melody in horizontal direction whereas the fourth refers to its vertical dimension.

Let us quickly go through the details: The word 'interval type' describes the type of tone combination that predominates over several bars. In this regard, most tonal Western melodies either make use of scale segments (or tone steps; in German: 'Skalenmelodik') or of triads in succession (in German: 'Dreiklangsmelodik'), or of a combination of both. The second property, named 'melodic contour' (first mentioned by Abraham and Hornbostel in 1909), refers to the outline of a melody. It describes the curve of melodic movement in rough lines without considering detailed interval progression. For the majority of Western tunes the following contour types are characteristic: continuously ascending or descending, undulating around the pitch axis, or having an arch-like shape (e.g. Adams 1976).

C. Neuhaus (✉)
Systematic musicology, University of Hamburg, Hamburg, Germany
e-mail: chr_neuhaus@t-online.de

From the empirical point of view White (1960) argues convincingly that 'Happy birthday' and other familiar tunes that have been altered with some algebraic procedures can easily be identified when transformations are linear, i.e. when preserving contour while the opposite holds true when transformations are non-linear, i.e. when distorting contour. In detail, subtracting, adding, or multiplying an integer, e.g. i(+)1 or 2i(−)1, keep tunes recognizable but whenever intervals are allowed to change sign, recognition is poor.

Now let us think on 'balance between form parts' which is the third structural property that contributes to the quality of a melody. In this regard lots of tunes are restricted to eight bars which is the common length of the 'musical period'. This type of musical form can further be split up into segments of 4 plus 4, both parts revealing some symmetry due to similar melodic beginnings. (On the empirical findings on musical form perception I will report in detail several pages later).

The fourth structure-determining factor is the assignment of a melody to a latent harmonic framework which is mostly concurrent with the cadence scheme. This assignment allows us to evaluate each tone in its tension towards the harmonic basis, in particular towards inherent vertical anchor points such as the tonic or the dominant, often arousing feelings of tension or relaxation. The involvement of vertical aspects into the horizontal line particularly becomes apparent in the improvisation practice of North Indian music. Here, each tone of a raga becomes established in the listener's mind by constantly playing the drone, i.e. the tonic tone, on the tanpura. This way, distance estimation in vertical direction can be performed easily (see e.g. Danielou 1982).

However, most encyclopedias do not define 'melody' by mere description or analysis of its structural properties but also consider the perceptual side with a particular focus on Gestalt psychology (c.f. van Dyke Bingham 1910; Ziegenrücker 1979). The Gestalt school of psychology (represented by Wertheimer, Krüger, Koffka, and others) has two principles at hand to determine how melodies are perceived, namely the experience of 'unity or completeness' (also known as the 'law of closure'), and the quality of 'structuredness' ('Gegliedertheit') of auditory sequences. The first attribute primarily refers to perceiving as a cognitive process whereas the second is a structural property inherent in the material as such.

Let us briefly explain the points: The impression of completeness arises from cohesive forces regarding the melody's inner structure. Cohesiveness is high when tones are in close distance (i.e. in spatial and temporal proximity) to each other but low when tones are scattered over several octaves. This holistic approach, considering melodies as Gestalt entities, once more becomes obvious through the 'law of transposition', first articulated by the Austrian philosopher Christian von Ehrenfels (1890), the founder of Gestalt theory. According to this perceptual law, melodies remain invariable and can still be recognized when shifting the entire figure to another key.

The second principle that determines how melodies are perceived is based on (pre-)'structuredness' ('Gegliedertheit') as an inherent property of time-based art per se (e.g. West et al. 1991). 'Pre-structuredness' enables the active mind to

subdivide sound patterns into perceivable portions or musical phrases. Interestingly, this perceptual process of segmentation has an equivalent on the neuronal level: For the perception of musical phrase boundaries, a neural correlate has been identified by using event-related potentials as the method of study. This correlate is called Closure Positive Shift (abbreviated CPS), it has a counterpart in the perception of speech. Whenever attention is on the detail, musicians react in the way just described whereas non-musicians respond with an early negativity (cf. Knösche et al. 2005; Neuhaus et al. 2006). This indicates that qualitatively different ERP components (CPS and early negativity) reflect some top-down activation of general but dissimilar phrasing schemata depending on prior musical training.

Besides this issue of (pre-)structuredness which helps to segment and encode the complete musical sequence we should also become aware of another form of experiencing the whole, named 'complex quality' which is a term coined by Krüger in 1926. Complex qualities are indivisible entities of purely sensual character that are evoked by emotions or sensations. This difference in holistic experience—caused by (pre-)structured and perceptually groupable objects on the one hand and indivisible emotions on the other—is indeed the main point that distinguishes the 'Berlin School of Gestaltpsychologie' (represented by Wertheimer, Köhler, and Koffka) from the 'Leipzig School of Ganzheitspsychologie' (represented by Krüger and Wundt).

When further elaborating on pre-structuredness, two comments have to be mentioned here, the first made by the Czech musicologist Karbusicky (1979), the second made by Levitin (2009), an American sound producer and neuroscientist. Karbusicky focusses on the difference between the words 'tone structure' and 'tone row'. As mentioned above, 'tone structure' indicates that notes are grown together into an entity whereas 'tone row' describes that notes are loosely connected so that elements can arbitrarily be replaced. Levitin (2009), by contrast, draws our attention to the emotional aspect of this issue, emphasizing that "a randomized or 'scrambled' sequence of notes is not able to elicit the same [emotional] reactions as an ordered one" (p. 10). Levitin focusses in particular on the structure's temporal aspects and their respective neural correlates. By using functional imaging methods he could prove that two brain regions of the inferior frontal cortex—named Brodmann Area 47 and its right-hemispheric homologue—are significantly activated when tone structure is preserved but do not react to arbitrarily scrambled counterparts, having the same spectral energy but lacking any temporal coherence (Levitin and Menon 2005). In addition, Neuhaus and Knösche (2008) could demonstrate by using event-related potentials that the brain responds to time-preserving versus time-permuting tone sequences from a very early stage of sequence processing on. In detail, lack of time order causes a larger increase of the P1 component already around 50 ms (measured from sequence onset) while no such difference in brain activity beyond 250 ms could be observed. In this regard, the brains of musicians and non-musicians react similarly, as both groups of participants have been tested in this study. From these results we conclude that the establishment of a metrical frame, i.e. the intuitive grasp of the underlying beat and the meter's accent structure is essential to integrate new tone items and for processing pitch and time relations (cf. Neuhaus and Knösche 2008) (Fig. 1).
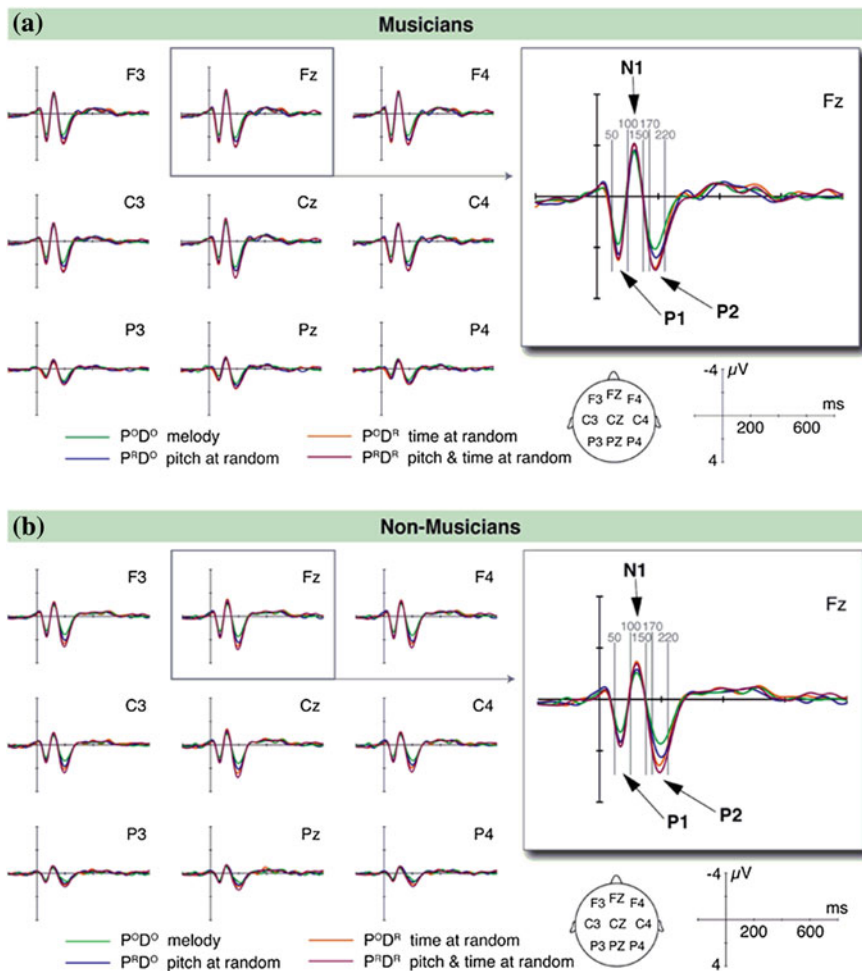
Fig. 1 Grand average event-related potentials of musicians (**a**) and non-musicians (**b**) at selected electrodes. Compressed brain activity over all tones per condition (from tone 6 onward). [*Source* Neuhaus and Knösche (2008)]

Taken non-randomness, i.e. the structured nature of a melody as given, how do we become aware of interval type, melodic contour, and the other properties mentioned above? Which listening mode is the most adequate to grasp these essential musical features? Everyone would agree with me to leave emotional, associative, aesthetic, and distracted styles of listening aside in favor of that one which aims at understanding the logic behind the tones and brings the composer's intention into focus.

The listening style that fits best is called 'structural hearing', a term coined by the Austro-American music theorist Felix Salzer in 1952. According to his younger colleague, the American musicologist Rose Rosengard Subotnik

structural hearing especially stands for modern viewpoints on music and became "the prevalent aesthetic paradigm in Germanic and Anglo-American musical scholarship" around 1950 (cf. dell' Antonio 2004, p. 2). To elaborate on the details, structural hearing "highlights [the] intellectual response to music to the almost total exclusion of human physical presence" (cf. dell' Antonio 2004, p. 8) and should therefore be seen in contrast to some sort of kinesthetic listening style, where the listener may feel the urge to dance, tap his feet, or snap his fingers (cf. Huron 2002; see also Kubik 1973, for the African variant of this motion-inducing listening mode). This 'intellectual response to music', as dell' Antonio puts it, might be restricted to music students, professional instrumentalists, and other people educated in music as it requires that several preconditions such as ear training and self-discipline as well as certain deductive musical abilities like melodic completion and anticipation are fulfilled. Thus, in full consequence, "[structural hearing] gives the listener the sense of composing the piece as it actualizes itself in time." (Subotnik 1988, p. 90). To what extent non-musicians are also capable of doing so, i.e. whether the aforementioned (listening) skills can also be acquired through implicit learning by mere exposure to music remains a question for further empirical research.

In this context we may call to mind that from a sociological point of view music philosopher Adorno (1962) as well as the musicologist and pedagogue Rauhe et al. (1975) made interesting general remarks on music consumption, taking the economic preconditions as well as the different educational standards within society into account. Both worked separately on a specific typology of listening behavior, elaborating carefully on—mostly a priori obtained—categories of reception (in German: 'Rezeptionskategorien').

With regard to structural hearing, Adorno (1962) distinguishes between two types of music listeners. The first is the 'music expert' who is almost exclusively recruited from music professionals, and the second one is the 'good [adequate] listener'. Please note that the 'good listener' should not be confused with another type of recipient, called 'educated listener' (or 'possessor of knowledge'; in German: 'Bildungskonsument'), being familiar with facts and reports about musical pieces and their interprets.

According to Adorno the 'music expert' is able to set past, present, and future moments within a musical piece in relation to each other to comprehend the musical logic behind the piece. Furthermore, while listening to music, the expert is aware of every structural aspect and can report on each musical detail afterwards.

The 'good [adequate] listener', by contrast, possesses these listening skills in a reduced form. Although he can relate musical parts to each other as the 'music expert' does, he is not fully aware of the consequences and implications that specific chords and other pivotal points with regard to musical progression have.

Rauhe et al. (1975) tackle this issue in a slightly different way. From the very start they distinguish between distinct attributes freed from human prototypes. That is, they argue on the basis of listening style rather than on the personnel level. Thus, the distinguishing feature between Rauhe's and Adorno's typologies is a certain flexibility in terms of switching between listening habits.

With regard to structural hearing, Rauhe et al. (1975) differentiate between a 'structural-analytical' and a 'structural-synthetical' listening mode (in German: 'strukturell-analytische Rezeption' and 'strukturell-synthetisches Hören'), meaning that the attentional focus is either on the structural detail, or that skills are quite similar to those possessed by Adorno's 'musical expert'.

In slight contrast to that, modern American musicologists such as Subotnik, dell' Antonio, or Huron put more emphasis on the real-time listening situation as such, i.e. on versatile listening styles beyond any (a priori drawn) boundaries, taking into account that actual listening behavior may change spontaneously throughout a performance. The American music psychologist Aiello (1994) puts it this way:

> I believe that when we listen, consciously or subconsciously, we *choose* what to focus on. This may be because of outstanding features of the music itself, certain features that are emphasized in the particular performance we are hearing, or because we just choose to focus on this or that element during that particular hearing. A piece of classical music is filled with more information than a listener can process in a single hearing. […] Given the complexity and the richness of the musical stimulus, listening implies choosing which elements to attend to. (p. 276).

More importantly, this principle of selection is confirmed with a series of experiments performed by the Belgian music psychologist Deliège et al. (1996). Deliège and their colleagues provide proof that musicians and non-musicians are able to extract several cues such as 'registral shifts' or 'change of density' from the textural surface of unfamiliar pieces, showing that cue abstraction is an effective means to grasp and encode the most obvious musical aspects in any real-time listening situation.

To extend this approach some sort of 'zoom in'/'zoom out' principle seems best to deal with real-time listening situations, giving the listener the perceptual freedom to either consider the whole or focus on the structural detail. This zooming principle corresponds nicely to Rauhe's structural-analytical and structural-synthetical listening styles. However, to put this zooming principle into practice, several preconditions must be fulfilled, each pointing to higher-order perceptual processes, including the involvement of an active mind. Such preconditions are, first, attentional mechanisms, i.e. some kind of conscious perception of the musical piece, second, strength of mind as well as, third, some form of intentionality, i.e. *directing* attention towards the whole pattern or a specific sound feature.

But which attribute is essential to experience a melody as an organic whole? How shall we use our zoom mechanism when listening to a musical piece in real-time? In my opinion, synthesis is more important than selection, i.e. the mental act of setting parts in relation to each other has priority over, what Deliège calls, 'cue abstraction'. Hugo Riemann, the well-known music theorist of the 19th century, already became aware of this principle in 1877, he called 'relational thinking' (in German: 'Beziehendes Denken'). Relational thinking means to create coherence between constituent elements, intervals, and form parts along the horizontal axis, lasting several seconds up to one minute, i.e. occurring within the temporal limits of working memory. However to describe these processes in modern psychology, the term 'chunking' is much more frequently used.

Theodor Lipps, a late German 19th-century philosopher and psychologist, puts emphasis on the *structural* side of relational thinking, i.e. on properties given in melodies and other musical pieces as such. By using some kind of mathematical formula, he called 'the law of number 2', Lipps managed to specify the relations between constituent tones and intervals as follows:

Whenever tones meet each other that behave as $2^n : 3, 5, 7$ etc., the latter naturally move towards the former; i.e. they have a natural tendency of inner motion to come to rest in the former. The latter "search for" the former as their natural basis, anchor points, or natural centers of gravity.

See also the original wording in German:

Treffen Töne zusammen, die sich zueinander verhalten wie $2^n : 3, 5, 7$ usw., so besteht eine natürliche Tendenz der letzteren zu den ersteren hin; es besteht eine Tendenz der inneren Bewegung, in den ersteren zur Ruhe zu kommen. Jene "suchen" diese als ihre natürliche Basis, als ihren natürlichen Schwerpunkt, als ihr natürliches Gravitationszentrum. (quoted from van Dyke Bingham, 1910, p. 11)

Please note that whenever we argue in favor of *process*, i.e. less in favor of structure, *future-directed* aspects of relational thinking become relevant. They are closely connected to concepts known as 'expectancy' and 'prediction' which either derive from a certain familiarity with the piece, from style-specific knowledge, or, in a statistical sense, from frequent occurrence of some textural features. In this regard, Pearce et al. (2010) distinguish between musical expectations generated on the basis of over learned rules and those built on the basis of associations and co-occurrence of events (with both processes most likely activating different parts of the brain).

From this it follows that processes of 'thinking ahead', be it prediction, anticipation, or 'setting-parts-in-relation-to-each-other', do not only demand the listener's perceptual awareness of the present with a particular focus on the input's acoustical features but also require some sort of re-activation of concepts about style, phrasing, harmonic progression, and other relevant structural issues stored in mind. Accordingly when 'thinking ahead', bottom-up and top-down processes do interact.

How can we 'operationalize' these processes? How can we grasp relational thinking by experiment? We first have to bring to mind that processes such as building expectancies and establishing coherence are two-fold, referring to musical items in horizontal as well as in vertical direction. Regarding the linear dimension, these processes can further relate to small segments or to the large-scale musical form. Let us quickly review the relevant research:

Narmour (1990, 1992), for instance, restricts his elaborations to the detail, i.e. to interval expectation formed within 3-tone segments. To explain how these expectancies for interval progression develop in the listener's mind he proposes a so-called implication-realization model, comprising five principal aspects based on the Gestaltists' perceptual laws, namely proximity, similarity, as well as good continuation. Among these five, the principle of *registral direction* means that in terms of small intervals pitch direction is expected to continue whereas for large intervals, i.e. of a perfect fifth or more, pitch direction is expected to change (see also Krumhansl 1997 for details).

Narmour's model of expectancy formation can simply be tested in a rating experiment, i.e. by asking participants to predict melodic continuation after having interrupted a folksong or some other musical piece at a specific point in time. However, when looking at this testing method a little more closely, Eerola et al. (2002) point out that, since listener's expectations may change in real-time listening situations, it is far better to obtain some continuous data on prediction. Thus, when testing expectation for melodic progression a dynamic approach seems more appropriate, using a coordinate system and a mouse-driven slider to enter the likelihood of forthcoming melodic events on a computer screen.

Krumhansl's probe tone method, in contrast (developed in collaboration with Roger Shepard in 1979) is a valuable means to investigate tone relations in vertical direction. Again, Krumhansl makes use of the rating method, this time by judging the cohesive strength between melodic events and the latent harmonic framework. For this objective, participants have to evaluate on a rating scale how well the twelve probe tones of the chromatic scale fit into the context of the just heard piece, be it tonal, atonal, bi-tonal, or non-Western (e.g. Kessler et al. 1984).

Cook (1987) as well as Eitan and Granot (2008) take a different approach to relational thinking. They study cohesiveness within *large-scale musical forms* along the horizontal axis. The objective of their studies is to find out whether and to what extent participants become aware of key change and of permuted musical sections while listening to sonata movements and other types of compositional form. A second focus is on the aesthetical impressions created by hybrid and original pieces. Again, insights are gained from rating results. By taking two of Mozart's masterworks (i.e. his piano sonata KV 332 and the earlier KV 280) as the original, and comparing it with a mixed version including some equivalent parts of the respective counterpart, Eitan and Granot (2008) could demonstrate that preferences for hybrid versions in musically trained listeners are strong, and become even stronger when exposed to these pieces several times. Since these rating results do not confirm that priority is given to the original, Eitan and Granot (2008) call the musical logic and the piece's inner unity into question. However, any sweeping generalizations based on these findings should be avoided since judgment results might be restricted to a specific idiolect and style, i.e. to Mozart's way of composing and to the Classical period per se.

In addition, Cook (1987) investigated the effect of tonality on cohesion. In this study, music students rated the degree of completeness of musical pieces of different lengths up to 6 min, ending either on the tonic or a distant key. Since modulations were merely perceptible within a time span of one minute or less, Cook concluded that form-building effects of tonality are weak which advocates for the psychological reality of small-scale over large-scale musical structures.

An even more radical view, called 'concatenationism', is hold by the American philosopher Jerrold Levinson (1997; cited by v. Hippel 2000). Concatenationism in its strictest sense describes a "moment-by-moment listening, faintly tinted by memory and expectation" (p. 135). According to Levinson, 'concatenationism'

denies any conscious influence of large-scale musical form on listening at all, meaning that we simply hear musical sections in succession and remain in the musical present.

To put this issue to the test, I performed a neurocognitive experiment on relational thinking at the Max Planck Institute for Human Cognitive and Brain Sciences Leipzig which I would like to re-report in this context here. To my knowledge it is the first neuroscience approach to musical form perception at all. Brain activity was recorded by using event-related potentials (ERPs) as the measuring method. To control the listening result immediately after presenting a melody, short behavioral feedback was given and registered via button press response. In general, event-related potentials (as well as other neuroimaging methods) have the advantage to map physiological processes in real time, enabling us to watch brain reactions—either (synchronized) extracellular current flows in terms of EEG and ERP or oxygen consumption regarding fMRI—while bundles of nerve cells are active.

The objective of the present study was to find out if the brain responds to the mental act of relational thinking and, if so, whether chunking processes can be made visible by specific component reactions. From the structural point of view I thus tested 'balance between form parts', which is a property considered essential when thinking about the melody's particular characteristics (c.f. p. 1f).

For the clarity of the experiment I decided for the small-scale type of musical form, using the eight-bar 'musical period', also called 'Liedform', in two variants, AABB and ABAB (see Fig. 2).

The figure shows that form parts with equal labelling (**AA**, A**BAB**, AB**AB**) have exactly the same rhythm while intervals are only similar, whereas form parts with different labelling (**AB**) are dissimilar in both, rhythm and interval structure. In terms of sequence succession we assume that relational thinking is affected by the sequential order of form parts, meaning that chunking may either be facilitated or made more difficult when A-parts alternate with B-parts as in ABAB, or when A- and B-parts follow in immediate repetition as in AABB. From these considerations the following hypotheses are deduced:

1. H1: Adjacent form parts of *contrasting* structure (AB) are subsumed to higher-level perceptual units. Melodies of form type ABAB are rated as hierarchical.

   H0: The sequential order of form parts has no effecton building higher-level perceptual units at all.

Fig. 2 Example melodies in *AABB* and *ABAB* forms (modified versions of originals). *First example* L. v. Beethoven. Rondo, WoO 48. *Second example* C. Ph. E. Bach. Thema from Sonata III

2. H1: Structuring small-scale musical forms is an online cognitive process taking place in working memory.
   Specific ERP components serve as indicators.

   H0: Structuring form parts happens in retrospect. Post hoc ratings are necessary to evaluate the entire melody as either 'hierarchical' or 'sequential'.

   From this it follows that melodies of form type ABAB probably elicit a large arc of suspense, increasing coherence, whereas in melodies of form type AABB the arc of suspense is smaller, and cohesive strength is reduced.

   Figure 3 illustrates the task. Participants had to listen to and evaluate the patterns by choosing either a sequential or a hierarchical listening style. Rating results were indicated by press of key buttons. The study was exclusively performed with non-musicians.

1 Methods

1.1 Subjects

Twenty students of different faculties (recruited from the University of Leipzig) participated in the experiment. Age and gender were equally balanced (10 males, 10 females; average age = 25.9 years, $SD = 2.29$).

2 Stimuli and Task

Melodies of form types AABB and ABAB were built on the eight-bar schema of the Classical period split up into $2 + 2 + 2 + 2$- phrase-units (see note examples in Fig. 2). Each form type was realized with 25 different melodies randomly presented in three keys (C major, E major, A flat major), resulting in 75 different

Fig. 3 Illustration of the listening task. Possible segmentations of a melody in real-time. For each given example participants had to decide by individual judgment which listening strategy was best. **a** Patterns are closed sections following in a row. **b** Patterns are related to each other and form higher-order units
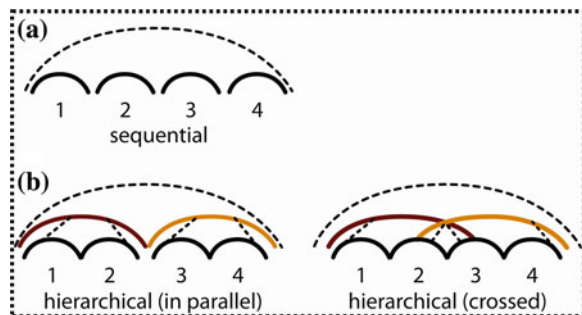
versions of AABB and ABAB, respectively. In terms of melodic contour, an arch-like shaping of all four phrases might intensify the impression of closure, and melodies of that shape might be processed sequentially whereas an upward contour followed by a downward one, each extending over 2 plus 2 bars, strengthens coherence, and listening might be more hierarchical. We therefore took heed that example pieces with arch-like and upward/downward contour were equal in number so that modulating effects caused by contour could be excluded. In addition, pause length, length of the pre-boundary tone, and the underlying harmonic schema were kept constant in each melody to avoid interferences with phrase boundary perception, probably yielding the CPS. In each sequence, the chord progression was "tonic (1st phrase)—dominant (2nd and 3rd phrase)—tonic (4th phrase)". Phrase boundaries had an average pause length of 0.27 s ($SD = 0.12$), and the average duration for each pre-boundary tone was 0.42 s ($SD = 0.25$).

The example pieces were played on a programmable keyboard (Yamaha PSR 1000) in the timbre 'grand piano'. They were stored in MIDI format using the music software Steinberg^{TM} Cubasis VST 4.0. For wave-file presentation via soundcard, MIDI files were transformed to Soundblaster^{TM} audio format.

Each musical piece was played with an average tempo of 102.36 BPM ($SD = 27.92$), differing slightly between AABB (108.76 BPM, $SD = 32.23$) and ABAB (95.96 BPM, $SD = 21.63$). However, an independent samples t-test yielded a non-significant result (t(48) = 1.65, p > 0.1), giving certainty that tempo should not be considered as a disturbing factor.

Figure 3 shows the given options how to deal with each melody. The drawings were also presented during instruction to illustrate the following task: "Listen carefully and evaluate the parts: Whenever you perceive sections as closed, following in a row—press the left button for 'sequential'. Whenever you perceive sections as related to each other, forming higher-order units—press the right button for 'hierarchical'." Participants were also requested to keep head, neck, arms, hands, and fingers as relaxed as possible and to reduce the amount of eye blinks during ERP recording (cf. general guidelines of ERP measurement, Picton et al. 2001). To get familiar with the task, each recording session started with a test block of ten melodies.

3 Experimental Set-up and Recording

The total duration of each recording session was approximately 55 min. Three blocks with stimuli were presented, each consisting of 50 melodies of types AABB and ABAB in pseudo-random order. Each example piece was part of a trial sequence. It ran as follows: "Fix your eyes onto the monitor (2 s), listen to the added type of melody (approximately 11 s), then evaluate whether the melody is sequential or hierarchical (less than 6 s)." Trial sequences were presented automatically using the software package ERTS (Experimental Run Time System,

Version 3.11, BeriSoft 1995). Each subject was comfortably seated in a dimmed and electrically shielded EEG cabin in front of a monitor. For binaural presentation of stimuli, a loudspeaker was placed at a distance of 1 m.

Brain activity was measured with 59 active Ag/AgCl electrodes (Electro Cap International Inc., Eaton, Ohio) placed according to the 10-10 system onto the head's surface (Oostenveld and Praamstra 2001). Brain activity per electrode was referenced to the left preauricular point (A1), and the sternum was used as the ground electrode. EEG signals were recorded with an infinite time constant and digitised with a sampling rate of 500 Hz. Ocular artefacts were measured with a vertical and a horizontal electrooculogram (EOGV, EOGH). The impedance at each electrode channel was kept below 5 kΩ.

4 Data Analysis

4.1 Pre-processing of Signals

The obtained raw signals were high-pass filtered (cut off frequency 0.50 Hz) and carefully examined for eye blinks, muscle activity, and technical artefacts. Artefact-free trials were merged over melodies and blocks, but averaged separately according to form type (AABB vs. ABAB), phrase onset (2nd, 3rd, and 4th phrase), and electrode channel. The time window for averaging was −200–1,000 ms measured from onset of the respective phrase. ERP traces were baseline-corrected, using a pre-onset interval from −30 to 0 ms. The pre-processing procedures were performed for each subject individually. Figure 5a–c shows the grand average ERP over all subjects at nine representative electrodes (F3, Fz, F4, C3, Cz, C4, P3, Pz, P4). The time range for display is the same as for averaging, namely −200–1,000 ms pertaining to phrase onset.

5 Statistical Analysis

Table 1 shows the button press responses of participants rating melodies as hierarchical or sequential. An additional Chi-square test was performed to prove whether rating results correlated significantly with pattern structure. In contrast to that, Table 2 summarizes the attributes of several 'prototype melodies' in AABB and ABAB that 75 % of participants (n = 15) consistently rated as hierarchical or sequential.

To test the significance of participants' *ERP data*, we computed several ANOVAs (repeated-measures analyses of variance). We started with a four-way analysis so that, if significant, effects of factor and level could be analyzed separately. Repeated measures factors were Time Window (five time ranges: 30–80,

Table 1 Participants' rating scores specified according to Form type

Form type	Rating results	
	"hierarchical"	"sequential"
AABB	1009 (67.45 %)	487 (32.55 %)
ABAB	1109 (74.28 %)	384 (25.72 %)
Across form types	2118 (70.86)	871 (29.14)

Sum of raw scores and percentages

80–140, 140–280, 300–600, 600–800 ms), Onset (2nd, 3rd, 4th phrase), Form (AABB and ABAB), and Channel (36 electrodes). Due to lack of space, results reported here are restricted to only one time window (300–600 ms). Further details were provided with a three-factor ANOVA analysis per time window including the repeated measures factors Onset, Form, and Channel. Dependencies between factors and factor levels were specified with several one-way and two-way post hoc tests. Degrees of freedom were adjusted with Huynh and Feldt's epsilon, and results were considered significant at an α-level of 0.05.

6 Results

6.1 Behavioral Results

Table 1 shows the overall rating scores of participants. The distribution assigns the total number of listening judgments ('hierarchical' vs. 'sequential') to form type ('AABB vs. ABAB'). Altogether, approximately two-thirds of example melodies were judged as 'hierarchical', the remaining one-third was judged as 'sequential'. Further specification by 'Form type' revealed that example pieces in ABAB compared to AABB were slightly more often rated as 'hierarchical' (74.28 vs. 67.25 %). This difference reached statistical significance ($t(38) = 2.13$, $p < 0.05$).

Table 2 Prototype melodies, consistently rated as hierarchical or sequential (14 themes in AABB and 19 themes in ABAB)

Frequency distribution	AABB		ABAB	
	Hierarchical	Sequential	Hierarchical	Sequential
Total number of analyzed melodies	11 [44 %]	3 [12 %]	16 [64 %]	3 [12 %]
Rhythmic contrast between part A and part B	11 [44 %]	1 [4 %]	15 [60 %]	1 [4 %]
Contour				
Archlike	1 [4 %]	2 [8 %]	3 [12 %]	2 [8 %]
Upward–downward	10 [40 %]	1 [4 %]	12 [48 %]	1 [4 %]
Downward–upward			1 [4 %]	

Frequency distributions (raw scores and percentages)

We computed a Chi-square test to evaluate if participants' *overall rating results* differed significantly from chance level. The X^2-value of 16.9 is far beyond the critical value of 10.83 at the 0.001 α-level (X^2 (1, N = 2989) = 16.9, p < 0.001). To quantify correlation strength between 'Form type' and 'Listening style', we additionally computed the adjusted contingency coefficient C^*, yielding scores on a range between 0 and 1. The result ($C^* = 0.86$) shows that the effect size between 'Form type' and 'Listening style' is large, indicating that correlation between both factors is strong.

Table 2 gives information about a subset of example melodies that three-fourths of participants (n = 15) consistently rated as hierarchical or sequential. Two variables are introduced which explain modification *across form types*: 'rhythmical contrast' and 'melodic contour'. The data show that both, sharp rhythmical contrast between A and B and/or upward-downward (instead of arch-like) contours contribute to a strong impression of higher-level junctions—either of 'AB' with 'AB' or of 'AA' with 'BB'.

7 Electrophysiological Results

We expect that chunking tendencies were strongest at the immediate onset of A- and B-parts. For that reason, three trigger points were set per example piece, one at each phrase onset. Note that at each phrase onset the average tone length was 0.24 s ($SD = 0.17$) while mode values, i.e. tone lengths occurring most frequently, were 0.07, 0.08, and 0.18 s over examples. We therefore proceed on the assumption that each phrase onset displays the average brain response to at least two or three tones rather than to the onset tone alone.

The first idea was that the brain reflects the interrelation between 'Form type' (AABB vs. ABAB) and 'Listening style' (hierarchical vs. sequential). Figure 4 shows the grand average ERP at the onset of the second phrase which is the crucial phrase of the eight-measure theme following the initial phrase.

From 300 ms onward, we observe a splitting of traces according to 'Form type' (AABB vs. ABAB) but no further division for subjective listening style

Fig. 4 Grand average ERPs at phrase onset 2. Division of traces by form type and listening style. Cortical activity at electrode *Fz*
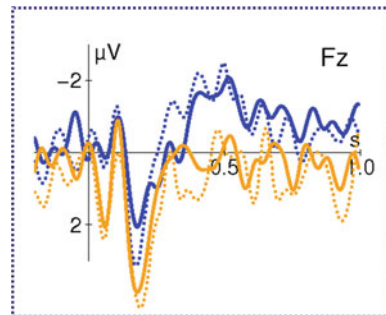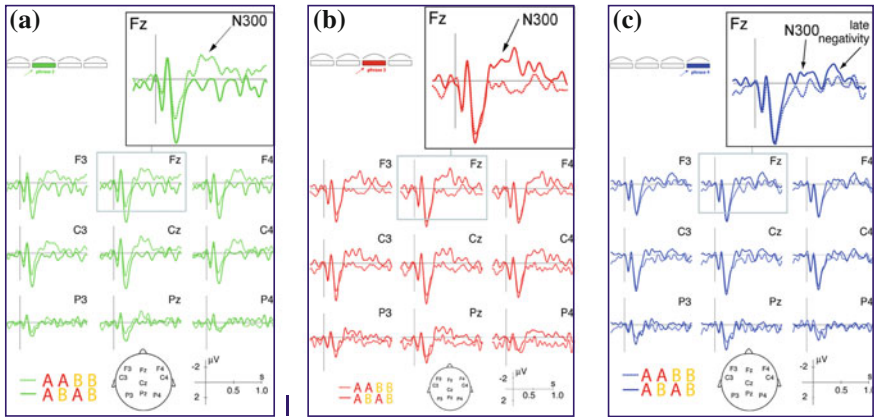
Fig. 5 a–c Grand average ERPs at phrase onsets 2, 3, and 4. Curve splittings according to 'Form type' (*AABB* vs. *ABAB*)

(hierarchical vs. sequential). This observation is validated with statistics. The ANOVA analysis yields a main effect of Form type ($F(1,19) = 7.53$, $p < 0.01$, 36 channels) whereas no interaction between 'Form type' and 'Listening style' could be found. The ANOVA results for curve splitting therefore suggest that the brain is merely sensitive to pattern structure (but not to the respective listening style).

Figure 5a–c displays the event-related potentials *as a mere function of form type*. The most interesting result is the broad negative shift between 300 and 600 ms measured from phrase onset. Since the amplitude maximum is fronto-central, we call this negative shift 'anterior negativity' (anterior N300). As you can see from Fig. 5a–c polarity changes between the onsets of the second and the third (and fourth) phrase. At phrase onset 2, the anterior negativity is up for patterns **AA** (immediate repetition of A) compared to AB, while at phrase onset 3 it is up for **ABA** (non-immediate repetition of A) compared to AAB. At phrase onset 4 we observe an anterior N300 for A**BAB** compared to AABB, although amplitude is reduced. Due to this pattern reversal, a main effect of 'Form type' is only marginally significant, whereas the interaction 'Form type x Onset' is highly significant. The anterior N300 also decreases from anterior to posterior. This is validated by a highly significant interaction 'Onset x Form type x Channel', mostly pronounced at phrase onset 2. (Due to lack of space ANOVA results are not displayed here).

8 Discussion

What are the study's main results in particular with regard to relational thinking? Let us briefly call them to mind:

1. [*behavioral*]. Sequences with form parts in alternating order (ABAB) are more often judged as 'hierarchical' than sequences with form parts in immediate succession (AABB) (cf. Table 1).
2. [*behavioral*]. Judgment results are more pronounced when rhythmical contrast between A- and B-parts is sharp and when melodic contours are upward-downward (cf. Table 2).
3. [*electrophysiological*]. The brain seems to react to pattern similarity at the onsets of adjacent (A|A) and of non-adjacent (e.g. AB|A) form parts. It does *not* respond to contrasting pattern structure (A-B).
4. [*behavioral and electrophysiological*]. Decisions on how form parts are perceived are not reflected in the ERP as a 'real-time protocol of brain activity'. Instead, the impression of the entire melody seems necessary.

The *behavioral results* suggest that participants are able to distinguish between example pieces of strong and weak cohesion. As already mentioned several sentences before, cohesiveness is strong when the following textural properties are given: (1) Form patterns alternate with each other as in ABAB, (2) melodic contours are upward-downward, and (3) rhythmic contrast between A- and B-parts is sharp. Among these points causing the impression of hierarchy, 'sequential order' obviously is the essential one while features regarding rhythm and contour are merely supplementary. Thus, the choice of the adequate listening style and the ease of evaluation seem largely to depend on the extent to which these properties are developed.

Note that due to the given task the primary focus was on listening strategy and on stimulus evaluation, i.e. on the *pre-requisites* for relational thinking. That is, chunking, defined as a cognitive process of 'subsuming form parts to higher-order units' probably requires a still larger amount of mental activity as needed by the current task.

Please note further that this experiment was exclusively performed with non-musicians, meaning that the process of 'setting-form-parts-in-relation-to-each-other' obviously belongs to a number of implicitly acquired musical skills. In this regard, Bigand and his colleagues were able to show that untrained listeners process tension and relaxation as well as several structural aspects above chance level and often with same results as professional musicians (Bigand 2003; Bigand and Poulin-Charronnat 2006). In this context Tillmann and Bigand (2004) write the following:

> Nonmusician listeners tacitly understand the context dependency of events' musical functions and, more generally, the complex relations between tones, chords, and keys. (p. 212).

Now let us try to explain the electrophysiological data. First, form parts with equal labelling (AA, ABAB, ABAB) are of exactly the same rhythm while intervals are only similar whereas form parts with different labelling (AB) differ in both, interval size and tone duration. We therefore suggest that the brain detects pattern similarity *on the basis of identical rhythmic structure*. In this context, the

anterior N300 may serve as a neural correlate, indicating the amount of mental effort which is necessary to identify the structural properties at phrase onsets. For these cognitive processes working memory resources are needed that are mainly found in the brain's frontal part. Whenever working memory is active, we observe amplitude maxima at anterior parts of the brain (cf. Patel 2003; and Fig. 5a–c).

Leaving considerations about rhythmical sameness aside, we can also explain the data from the viewpoint of expecting and predicting musical structure. In a recent study on pitch expectations in church hymns Pearce and colleagues (2010) managed to demonstrate that high-probability tones (perceived as expected) compared to low-probability tones (perceived as unexpected) elicit a fronto-central negativity in the time range from 300 to 600 ms similar to the anterior N300 of the present study but slightly different in shape. However, on closer examination, we have to keep in mind that the anticipated events of both studies are of different character and size, using highly probable single tones on the one hand and highly probable form parts (stretching over 2 bars) on the other. Even so, the similarity between the anterior brain components of both studies is striking.

With regard to 'relational thinking'—what conclusions should we draw from this study?

In my opinion the key aspect is that this 'setting-parts-in-relation-to-each-other' is not part of a superior process named 'structural listening' occurring in the same time range, but rather a separate process taking place several seconds later. This means that, initially, the anterior N300 indicates the processing of structural features whenever the brain becomes aware of rhythmical sameness in real-time listening situations, whereas relational thinking, or, more exactly, the choice of the adequate listening style combined with subsequent rating decisions, needs a complete perceptual unit to evaluate the coherence of form parts. Thus, so far, no specific neural correlate for pure active, higher-order mental acts such as chunking or Gestalt perception has been found, showing that relational thinking seems measurable only in retrospect by using traditional behavioral methods. The present ERP study therefore makes clear that a new paradigm seems more appropriate in order to investigate the idealized form of 'expert structural listening' referring to real-time listening skills such as reconstructing and anticipating the score.

Having elaborated on Riemann's 'relational thinking' and 'balance between form parts' from a structural point of view, we should close this chapter with some general remarks on structural listening and on grasping the musical idea, this time by also including the composer's point of view. The first point we should think about is that spatio-temporal concepts depend on the type of cognitive process, i.e. there are conceptual differences between imagining and creating on the one hand and simply listening to music on the other.

Let us first consider the composer's point of view: When describing the creative act, they often report on aspects of simultaneity and spatiality, i.e. on the overall-structure of a musical piece. Mozart, for instance, obviously sees the complete musical architecture in his mind's eye (c.f. a letter written by J. F. Rochlitz 1813) whereas, Schönberg uses the term "timeless entity" to describe this overall framework (quoted from Cook 1990, p. 226). Thus, a main problem when

composing music is to transfer spatial concepts into time-based sequences, i.e. to unfold basic ideas successively step by step. Listening, by contrast, proceeds in the opposite direction in that the overall idea has to be grasped (and compressed) from the sequential 'spread out', i.e. from tones presented in succession. Thus, in terms of transferring sound information, processes of composing/imagining and listening are diametrically opposite (see also Mersmann 1952).

In sharp contrast to that, Riemann completely omits this spatio-temporal aspect. He skips also every facet regarding structure and the sound material as such. Instead, he puts emphasis on the immediate mental exchange between the composer's and the listener's minds for rapid transfer of thought and direct flow of ideas. See the following quotation (Riemann 1914/15):

> The Alpha-Omega in music is not the sound, i.e. real music, or the audible tones as such, but rather the idea of tone relation. Before putting them into notation, these relations live in the imagination of the creative artist and then in the listener's imagination .…In other words, the essential approach to music, i.e. the key to its inner nature is neither provided by acoustics nor by tone physiology and tone psychology but rather by imagination of tone. (p. 15f)

See here the original wording in German:

> daß nämlich gar nicht die wirklich erklingende Musik, sondern vielmehr die in der Tonphantasie des schaffenden Künstlers vor der Aufzeichnung in Noten lebende und wieder in der Tonphantasie des Hörers neu erstehende Vorstellung der Tonverhältnisse das Alpha und Omega der Tonkunst ist. ….. Mit anderen Worten: den Schlüssel zum innersten Wesen der Musik kann nicht die Akustik, auch nicht die Tonphysiologie und Tonpsychologie, sondern nur eine "Lehre von den Tonvorstellungen" geben. (Riemann 1914/15, p. 15f)

However, several constraints are imposed on Riemann's idea of communicating from mind-to-mind, mainly caused by attitudes, experiences, expectations, wishes, and situational needs on the side of the listener (cf. Rauhe et al. 1975). In other words, encoding and decoding or the sending and receiving of the overall musical idea are largely dissimilar processes, even more when they are split between two people, since they are modified by some mental acts and attributes in-between. In this regard, Jackendoff distinguishes between a 'compositional grammar' and a 'listening grammar', although admitting in the spirit of Riemann that "the best music arises from an alliance of a compositional grammar with the listening grammar" (1988, p. 255; quoted from Cook 1994, p. 87). Even so, it makes more sense to assume that one person is real and the other fictional, meaning that a composer constructs his 'ideal listener' to enhance creativity and shape the (creative) work like poets or journalists do when having their 'ideal reader' in mind (cf. dell' Antonio 2004).

Altogether we may conclude that the perception of musical structure as well as related issues such as structural listening and relational thinking need an interdisciplinary approach, i.e. the move from music theory to music psychology to Gestalt psychology and back, focusing on both, the sound object, i.e. the tone material, and the respective perceptual process.

Let me close with a quotation from Miller and Gazzaniga (1984) on the interrelationship between structure and process in its broadest sense:

> There seems to be general agreement that the objects of study are cognitive *structures* and cognitive *processes*, although this distinction is drawn somewhat differently in different fields. In American psychology, for example, a long tradition of functional psychology has made it easier to think in terms of processes—perceiving, attending, learning, thinking, speaking—than in terms of structures. Yet, something has to be processed. … Linguists, on the other hand, generally find it easier to think in terms of structures—morphological structures, sentence structures, lexical structures—and to leave implicit the cognitive processes whereby such structures are created or transformed. Computer scientists seem to have been most successful in awarding equal dignity to representational structures and transformational processes. … In principle, however, it is agreed that both aspects must be considered together, but that once the process is understood the structure is easier to describe. Thus, cognition is seen as having an active and a passive component: an active component that processes and a passive component that is processed. (p. 8f)

References

Abraham, O., & v. Hornbostel, E. M. (1909). Vorschläge für die Transkription exotischer Melodien. In C. Kaden & E. Stockmann (Eds., 1986), Tonart und Ethos - Aufsätze zur Musikethnologie und Musikpsychologie (pp. 112–150). Leipzig: Reclam.

Adams, C. R. (1976). Melodic contour typology. *Ethnomusicology, 20*(2), 179–215.

Adorno, T. W. (1962). Typen musikalischen Verhaltens. In Einleitung in die Musiksoziologie—zwölf theoretische Vorlesungen (pp. 13–31). Frankfurt a.M.: Suhrkamp.

Aiello, R. (1994). Can listening to music be experimentally studied? In R. Aiello & J. Sloboda (Eds.), *Musical perceptions* (pp. 273–282). Oxford: Oxford University Press.

Bigand, E. (2003). More about the musical expertise of musically untrained listeners. In G. Avanzini et al. (Eds.), The neurosciences and music II: From perception to performance. Annals of the New York Academy of Sciences 999 (pp. 304—312). New York: The New York Academy of Sciences.

Bigand, E., & Poulin-Charronat, B. (2006). Are we "experienced listeners"? A review of the musical capacities that do not depend on formal musical training. *Cognition, 100*, 100–130.

Cook, N. (1987). The perception of large-scale tonal closure. *Music Perception, 5*(2), 197–206.

Cook, N. (1990). *Music, imagination, and culture*. Oxford: Oxford University Press.

Cook, N. (1994). Perception: A perspective from music theory. In R. Aiello & J. Sloboda (Eds.), *Musical perceptions* (pp. 64–95). Oxford: Oxford University Press.

Danielou, A. (1982). Einführung in die indische Musik (2nd ed.; Series: Taschenbücher zur Musikwissenschaft 36). Wilhelmshaven: Heinrichshofen.

Deliège, I., et al. (1996). Musical schemata in real-time listening to a piece of music. *Music perception, 14*(2), 117–160.

Dell'Antonio, A. (2004). Beyond structural listening: Postmodern modes of hearing. Berkeley: University of California Press (Ed. 2002).

Eerola, T., et al. (2002). Real-time predictions of melodies: Continuous predictability judgments and dynamic models. In Stevens, C., et al. (Eds.) Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney (pp. 473–476).

Eitan, Z., & Granot, R. Y. (2008). Growing oranges on Mozart's apple tree: "Inner form" and aesthetic judgment. *Music Perception, 25*(5), 397–417.

Huron, D. (2002). Listening styles and listening strategies. Talk presentation and handout. Society for Music Theory Conference, Columbus, Ohio, 1st Nov 2002.

Karbusicky, V. (1979). Systematische Musikwissenschaft. (Series: Uni-Taschenbücher 911). München: Wilhelm Fink.

Kessler, E. J., et al. (1984). Tonal schemata in the perception of music in Bali and the West. *Music Perception, 2*, 131–165.

Knösche, T. R., et al. (2005). Perception of phrase structure in music. *Human Brain Mapping, 24*(4), 259–273.

Krüger, F. (1926). *Komplexqualitäten, Gestalten und Gefühle*. München: C.H. Beck.

Krumhansl, C. (1997). Effects of perceptual organization and musical form on melodic expectancies. In M. Leman (Ed.), *Music, gestalt, and computing: Studies in cognitive and systematic musicology* (pp. 294–321)., Lectures notes in computer science1317 Berlin: Springer.

Kubik, G. (1973). Verstehen in afrikanischen Musikkulturen. In P. Faltin & H.-P. Reinecke (Eds.), *Musik und Verstehen: Aufsätze zur semiotischen Theorie, Ästhetik und Soziologie der musikalischen Rezeption* (pp. 171–188). Köln: Hans Gerig.

Levitin, D. J. (2009). The neural correlates of temporal structure in music. *Music and Medicine, 1*(1), 9–13.

Levitin, D. J., & Menon, V. (2005). The neural locus of temporal structure and expectancies in music: Evidence from functional imaging at 3 Tesla. *Music Perception, 22*(3), 563–575.

Mersmann, H. (1952). Musikhören. Frankfurt a. M.: Hans F. Menck.

Miller, G. A., & Gazzaniga, M. S. (1984). The cognitive sciences. In M. S. Gazzaniga (Ed.), *Handbook of cognitive neurosciences* (pp. 3–11). New York: Springer.

Narmour, E. (1990). *The analysis and cognition of basic melodic structures: The implication-realization model*. Chicago: University of Chicago Press.

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication–realization model*. Chicago: University of Chicago Press.

Neuhaus, C., et al. (2006). Effects of musical expertise and boundary markers on phrase perception in music. *Journal of Cognitive Neuroscience, 18*(3), 472–493.

Neuhaus, C., & Knösche, T. R. (2008). Processing of pitch and time sequences in music. *Neuroscience Letters, 441*(1), 11–15.

Oostenveld, R., & Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology, 112*, 713–719.

Patel, A. D. (2003). Language, music, syntax, and the brain. *Nature Neuroscience, 6*, 674–681.

Pearce, M. T., et al. (2010). Unsupervised statistical learning underpins computational, behavioural, and neural manifestations of musical expectation. *NeuroImage, 50*(1), 302–313.

Picton, T. W., et al. (2001). Guidelines for using human event-related potentials to study cognition: Recording standards and publication criteria. *Psychophysiology, 37*, 127–152.

Rauhe, H., et al. (1975). *Hören und Verstehen: Theorie und Praxis handlungsorientierten Musikunterrichts*. München: Kösel.

Riemann, H. (1877). *Musikalische Syntaxis: Grundriß einer harmonischen Satzbildungslehre*. Leipzig: Breitkopf and Härtel.

Riemann, H. (1914/15). Ideen zu einer "Lehre von den Tonvorstellungen". In B. Dopheide (Ed., 1975), Musikhören (pp. 14–47). Darmstadt: Wissenschaftliche Buchgesellschaft.

Subotnik, R. R. (1988). Toward a deconstruction of structural listening: A critique of Schönberg, Adorno, and Stravinsky. In L. B. Meyer et al. (Eds.) Explorations in music, the arts, and ideas: Essays in honor of Leonard B. Meyer (pp. 87–122). Stuyvesant, New York: Pendragon Press.

Tillmann, B., & Bigand, E. (2004). The relative importance of local and global structures in music perception. *Journal of Aesthetics and Art Criticism, 62*(2), 211–222.

Van Dyke Bingham, W. (1910). Studies in melody. *Psychological Review, 50*, 1–88.

Von Ehrenfels, C. (1890). Ueber 'Gestaltqualitäten'. *Vierteljahrsschrift für wissenschaftliche Philosophie, 14*, 249–292.

Von Hippel, P. (2000). Review of the book Music in the moment by Jerrold Levinson [1997. Ithaca, NY: Cornell University Press]. Music Analysis 19 (1), 134–141.

West, R., et al. (1991). Musical structure and knowledge representation. In P. Howell et al. (Eds.) Representing musical structure. (Cognitive science series 5, pp. 1–30). London: Academic Press.

White, B. (1960). Recognition of distorted melodies. *American Journal of Psychology, 73*, 100–107.

Ziegenrücker, W. (1979). *Allgemeine Musiklehre*. Mainz, München: Goldmann Schott.