

PROCESSING OF RHYTHMIC AND MELODIC GESTALTS—AN ERP STUDY

CHRISTIANE NEUHAUS AND THOMAS R. KNÖSCHE
*Max Planck Institute for Human Cognitive and
 Brain Sciences, Leipzig, Germany*

IN TWO EXPERIMENTS WITH EVENT-RELATED potentials (ERPs), we investigated the formation of auditory Gestalts. For this purpose, we used tone sequences of different structure. In the first experiment, we contrasted a rhythmic section to a section with random time values, each embedded in rhythmically irregular context. In the second experiment, melodies were contrasted to randomized sequences. Nonmusicians either had to detect the rhythmic pattern or to memorize short tone excerpts. Random versions in both experiments evoked a significant increase in the amplitude of P1 and P2. Randomized rhythm sections also evoked a late sustained negative potential. The enlarged P1 and P2 for random sequences might reflect stronger integration effort, as the predictability of tone progression was low. Thus, already at the early stage of encoding, sequence processing might be top-down-driven. The late negativity for rhythmically random sections is possibly task-related, reflecting expectancy violation in terms of regularity, since a metrical grid of beats could not be established. The memorizing of tone excerpts did not evoke a late neural correlate. (169)

Received February 16, 2006, accepted September 11, 2006

Key words: encoding, tone predictability, temporal expectancy, auditory Gestalt, event-related potentials

PICTURE PUZZLES, A CUBE IN DIFFERENT perspectives, or the ambiguous “a vase—two faces” figure of E. Rubin, are examples that illustrate that in the visual domain, perception is, first and foremost, holistic. Since 1890, Gestalt psychology has been describing how elements are combined with each other to form a perceptual entity. Ever since, principles such as temporal proximity or similarity are used to explain the coherence between adjacent items, and why objects create the

impression of a perceptual Gestalt (e.g., Köhler, 1929; Koffka, 1935; Wertheimer, 1922). However, it is still not clear if these principles, which are primarily valid for the visual domain, can also be applied to auditory perception and the art of music, which consists of structures that unfold successively over time (e.g., Deutsch, 1994; Reybrouck, 1997).

The first attempt to distinguish between “temporal” and “nontemporal” Gestalt qualities was made by the early Gestaltist von Ehrenfels (1890). Several Gestalt principles seem to be valid for one modality only, either for simultaneous visual Gestalts, or for successive auditory ones. The principle of good continuation, for example, can mainly be assigned to sequential Gestalts of the auditory domain, whereas ambiguity is most obvious in the visual domain (e.g., the Necker cube). Hence, one has to be cautious in applying concepts from one sensory modality to another. However, the most global concepts of perceptual organization, the so-called Ehrenfels-criteria (von Ehrenfels, 1890), “Übersummativität” (the whole is more than the sum of its parts), and “Transponierbarkeit” (a Gestalt can be transposed from one key to another), were originally developed in the music domain, to describe the main attributes of a melody.

Whenever the *process* of melodic Gestalt formation is to be described, the Ehrenfels-criteria are not adequate, since the process of building up is characterized by the successive integration of components. In a more general sense, Treisman and Gelade (1980) proposed that it is focal attention that provides the “glue” to integrate initially separable items into coherent objects. Their hypothesis, which is also known as the “feature-integration theory of attention,” was originally elaborated and proved for the visual domain, but might also be applied to Gestalt formation in music. In detail, this theory suggests that in a first step, physical properties of elements are processed automatically and pre-attentively, whereas in a second step, attention is directed toward the spatiotemporal relationship between adjacent items, to build a coherent percept.

Brain responses in cognitive neuroscience reveal that the integration of successive auditory items already takes place at a very early processing stage, that is, in the

first 200 ms after stimulus-onset. The process of integration is reflected by an early negative component of the event-related potential (ERP), the so-called mismatch negativity (MMN; e.g., Sussman, Winkler, Ritter, Alho, & Näätänen, 1999). In the auditory domain, the MMN is evoked whenever a change within a flow of repetitive stimuli can be identified. If, for example, a stream of constant auditory events (e.g., tones or short words) is given, the brain detects this regularity and forms a so-called *standard*. If this standard trace is violated by a *deviant* sound, an MMN is elicited (for a review, see, e.g., Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001). If stimuli are presented very rapidly, they are even more strongly integrated and form an auditory object that is perceived as a whole (Sussman, Ritter, & Vaughan, 1998).

The processing of sound sequences with a different degree of temporal regularity was also investigated by Griffiths and colleagues (Griffiths, Büchel, Frackowiak, & Patterson, 1998; Griffiths, Uppenkamp, Johnsrude, Josephs, & Patterson, 2001; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002), although their experiments were carried out from the functional-anatomical point of view. PET and fMRI results revealed that auditory sequences, in particular melodies, are processed in a hierarchical manner. Processing on a higher level becomes more asymmetric between hemispheres. That is, in the anterior and posterior parts of the primary auditory cortex (Heschl's gyrus), Griffiths et al. (1998) found bilateral and symmetric activation for pitch extraction, while in the secondary auditory cortex (planum polare and superior temporal gyrus) melodic processing was right-lateralized. However, for regular versus irregular stimulus types, brain-imaging results (Patterson et al., 2002) did not reveal any difference in cortical activation.

In a recent behavioral study, Nakata and Mitani (2005) demonstrated, that even infants between 6 and 8 months of age are able to distinguish between regular and irregular tone sequences. When testing the average looking times for colored flashes, combined either with regular or irregular sound sequences, infants paid more attention to the flashes when the onset-onset distance was constant (173 ms), but not when this distance was variable.

A recent fMRI study by Levitin and Menon (2005) compared brain responses to familiar and unfamiliar pieces of music with such to scrambled versions of the same excerpts. The scrambled versions were generated by cutting the musical excerpts into arbitrary 250 to 350 ms long pieces, which were then reassembled in randomized order. This way, musical structure (i.e., contour,

harmonic functions, metrical accents, phrases, and phrase boundaries) was disrupted, while spectral energy and acoustical features between conditions were the same. Functional resonance imaging with nonmusicians as participants showed high activation in the ventral prefrontal cortex (Brodmann Area 47) for well-structured pieces, but not for those with disrupted syntax. For Brodmann Area 47, Levitin and Menon propose a kind of "structure tracker" that detects underlying syntax, and is deactivated whenever syntax is demolished.

Different from Levitin and Menon, scrambled impressions in the present study were produced by pseudo-random permutation over a set of tones, consisting either of time values (rhythm experiment) or of pitch and time values (melody experiment). This preserved the sequence structure of the stimuli and enabled us to use the ERP components associated with the onset of distinct tones (e.g., P1, N1, P2) as a probe for temporal Gestalt processing. P1, N1, and P2 components are the first high-amplitude voltage deflections that originate in the cortex and that can be observed in the ERP. They are of opposite polarity (P stands for positive; N for negative), and occur roughly between 50 to 100 ms, 100 to 200 ms and 170 to 250 ms, as measured from the onset of a given tone. In the broadest sense, these components reflect the primary processing of acoustic stimulus properties. However, top-down effects, such as attention, can also modify their amplitude and latency. With regard to these ERP components, two hypotheses, and their respective null hypotheses, are formulated and will be tested in our study.

H1: *Early* components with a latency of up to 200 ms reflect the process of Gestalt formation, that is, the buildup of a coherent percept through successive integration of tones, when tone progression is highly predictable. The corresponding null hypothesis is:

H0: *Early* components reflect the bottom-up processing of stimulus properties from the standpoint of psychoacoustics.

Hillyard, Teder-Sälejärvi, and Münte (1998) provided evidence that, first, acoustic feature encoding takes place during the first 200 ms after stimulus-onset and, second, that it can be modulated by top-down-driven processes. Subjects were given the task of focusing on random sequences of noise bursts, which were delivered from an array of seven loudspeakers. As apparent from the increase of the N1 depending on the attended sound source, amplitude modification occurred in

this early time range and was caused by spatial selective attention. In addition, Pantev et al. (1998), as well as Shahin, Bosnyak, Trainor, and Roberts (2003), provided evidence that amplitudes of the early N1 and P2 are also influenced by the music experience of subjects.¹

With regard to late ERP components, the research hypothesis is:

H1: *Late* components, especially the P3, indicate the processing of random tone order, in contrast to regular tone order. The respective null hypothesis is:

H0: *Late* ERP components do not reflect regularity processing.

The research hypothesis (H1) seems to be more probable than H0, since the P3 is considered to be a reliable marker for stimulus deviation and for incongruence within nonlinguistic context (e.g., Besson & Macar, 1987).

In the present two experiments, regular tone patterns are compared with irregular ones, first in the rhythmic domain, and second in the melodic domain. In the rhythm experiment, tone sequences with an embedded rhythmic pattern were compared with random tone sequences. In the melody experiment, rule-based “melodic Gestalts” were contrasted with random versions that had been generated by the random permutation of tones over all sequences, where both pitch and time values were randomized independently from each other. In this manner, it was ensured that, between conditions, basic properties of tones (pitch, tone duration, etc.) were equally distributed. Any measured difference between conditions must therefore be related to the way tone elements were ordered. In summary, both experiments investigated the processing of tone order versus disorder. Thus, emphasis was on the general aspect of perceiving structure versus randomness, that is, on the processing of auditory Gestalts and non-Gestalts. The study does not intend to investigate the neural processing of *specific* Gestalt principles in the auditory domain, such as good continuation and temporal or spatial proximity. Moreover,

as the task and context were quite different, any comparison between the experiments should be drawn with care. Likewise, this study does not intend to investigate how pitch- and time-based relations *interact* with each other to form a coherent percept (for a description of pitch- and time-based relations, see, e.g., Peretz, & Zatorre, 2005).

Method

Subjects

Participants were undergraduate students from the University of Leipzig (rhythm experiment: 7 males, 7 females, age 21 to 28 years, $M = 24.3$ years; melody experiment: 7 males, 8 females, age 22 to 30 years, $M = 25.0$ years). All of them were nonmusicians with an overall instrumental experience of less than 2 years. At the time of measurement, they did not sing in a choir, nor did they play any instrument. Subjects were in good health with no neurological or hearing disorders. Each subject gave written consent for participation.

Stimuli

RHYTHM EXPERIMENT

The rhythm experiment used 80 sequences, each made up of 60 tones with constant pitch (A_4 , 440 Hz). Forty sequences (condition RAND) consisted of elements with random duration. The other 40 sequences (condition RHY) included one out of four rhythmic patterns embedded in random context. Thus, RHY and RAND differed in the middle section, whereas the preceding and following irregular parts between conditions were exactly the same. In these irregular parts, the average *inter-onset interval* (IOI), that is, the distance from tone-onset to tone-onset, was 0.32 s ($SD = 0.25$); average *tone duration* was 0.28 s ($SD = 0.25$); and the average *inter-stimulus-interval* (ISI), that is, the distance from offset to onset, had a constant value of 0.05 s. For tone duration, ten different values were chosen (minimum = 0.025 s, maximum = 0.8 s, which correspond to values from 1/32 to 1/1 in musical notation). Each tone duration occurred almost equally often and was randomly distributed over all sequences. For comparison of embedded rhythmic and random sections, time parameters were equalized. Therefore, rhythmic sections (tone duration: MIN = 0.05 s, MAX = 0.2 s) were contrasted with a *subset* of random sections, where tone duration was limited to MAX = 0.3 s, instead of the 0.8 s duration used in the irregular parts.

¹Note that to establish the role of components as possible indicators for cognitive processes, their cortical origin is a necessary anatomical precondition. For the early P1 and N1, this was demonstrated in a study by Liégois-Chauvel, Musolino, Badier, Marquis, and Chauvel (1994) by measuring adult patients with intracerebral depth electrodes (stereotaxic method).

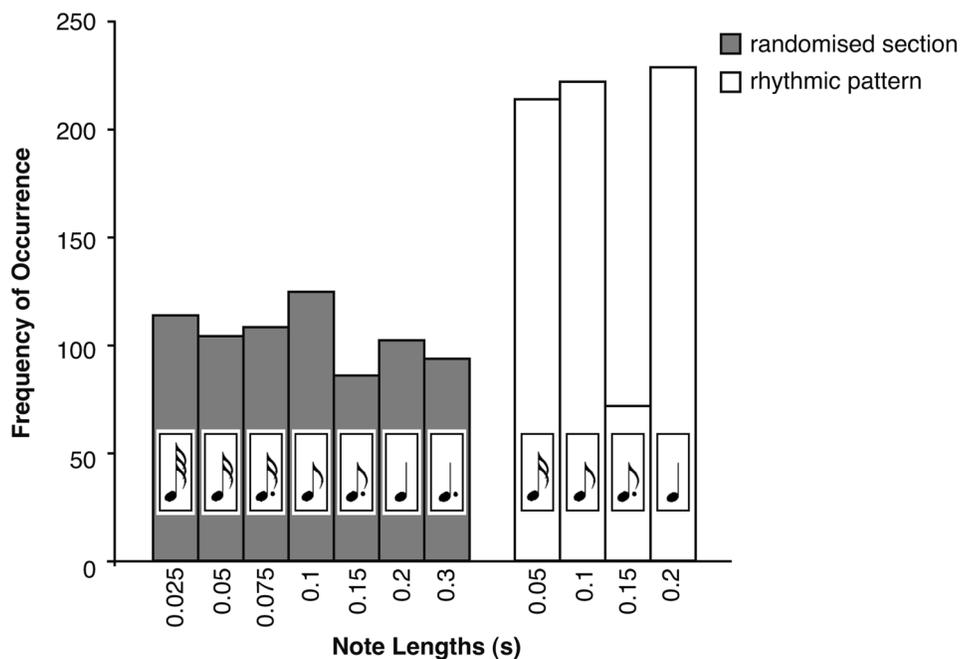


FIG. 1. Distribution of note lengths for rhythm sections (white columns) and random sections (grey columns).



FIG. 2. Rhythmic sections consist of one of the following modules: (a) quarter notes (16 notes), (b) a syncopated rhythm (24 notes), (c) a dotted rhythm (16 notes), and (d) a group of one eighth and two sixteenth notes (24 notes).

For rhythmic sections, average IOI was 0.17 s ($SD = 0.06$), average tone duration was 0.12 s ($SD = 0.06$), and the ISI had a constant value of 0.05 s. Random subset sections had an average IOI of 0.17 s ($SD = 0.08$), an average tone duration of 0.11 s ($SD = 0.08$), and a constant ISI of 0.05 s. Thus, average tone density between conditions was almost the same (see also histograms in Figure 1).

Sequences were presented in four blocks in pseudo-random order. For both types of stimuli, probability was the same. Each block lasted for approximately 10 minutes. Sequences, including a rhythmic section (cond RHY), had an average length of 16.47 s ($SD = 2.41$), and sequences with continuous randomization (cond RAND) had an average length of 19.55 s ($SD = 2.0$). Each sequence was generated in MIDI format. For presentation via soundcard, MIDI files were transformed to Soundblaster™ audio format using “TiMidity” (<http://timidity.sourceforge.net>). All sequences were played in the sound “timpani,” chosen from an open

set of instrument samples (Sound Font no. 47; see <http://freepats.opensrc.org>).² Rhythmic sections (condition RHY) were constructed from random sequences (condition RAND) by replacing either 16 or 24 random tones with a set of rhythmic uniform modules. Rhythmic modules consisted either of a dotted rhythm, of quarter notes, of a syncopation, or of a group with one eighth and two sixteenth notes (see Figure 2). Sequences including the same module type were presented equally often. Each time, rhythmic sections occurred at a different position, that is, in the first, second, or third part of the random

²We are aware of the fact that when listening to sampled timpani-like sounds, the sound envelope is mainly characterized by the attack, and to a lesser extent by the decay. However, due to the fact that soundfonts are controlled through MIDI, we also took account of tone duration, which is a basic parameter in MIDI files. Moreover, we also used tone duration in Figures 3a-d for the visual display of rhythmic and random examples.

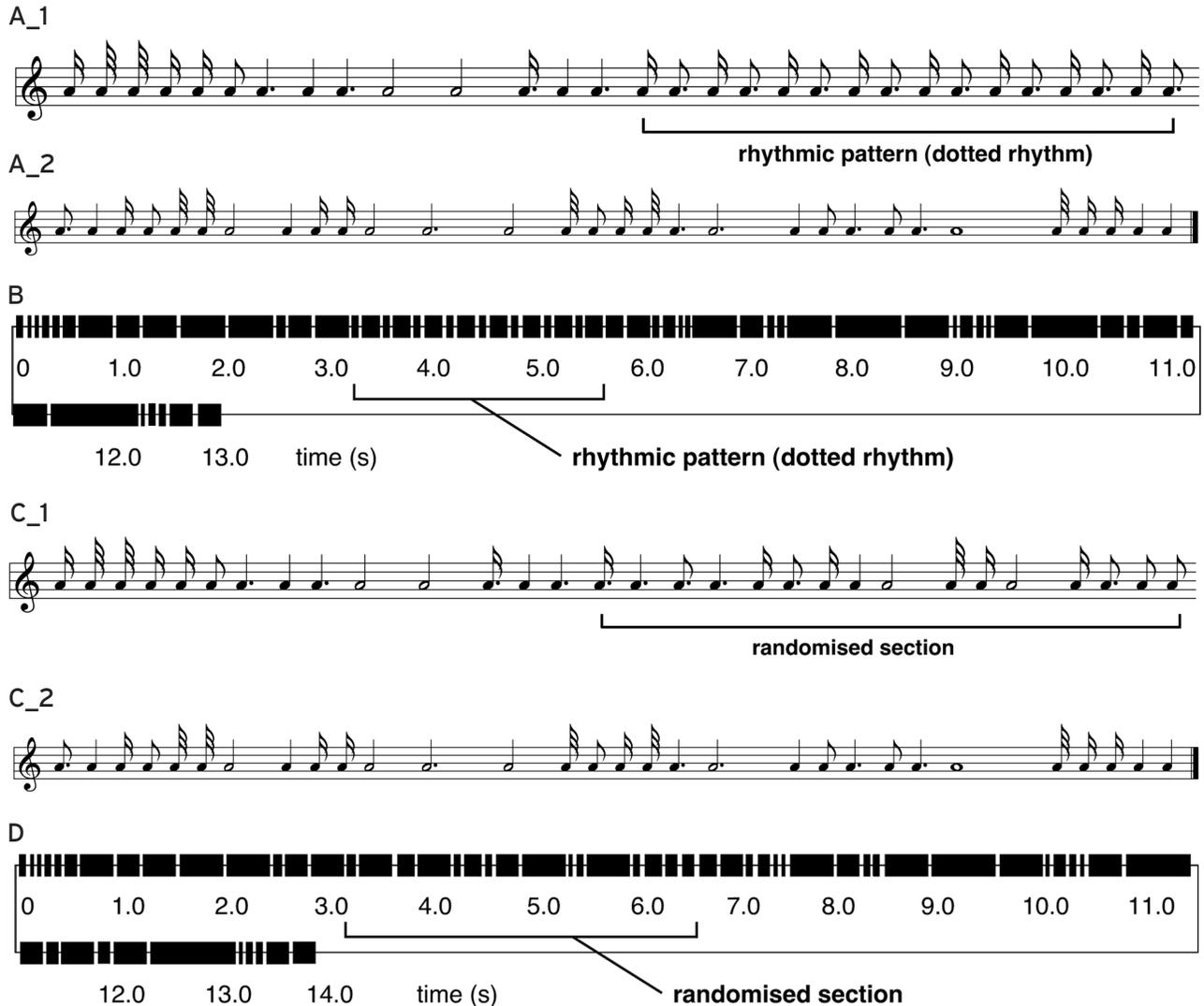


FIG. 3. Examples of tone sequences, including a rhythmic section (RHY), and with continuous randomization of elements (RAND); both examples in musical and experimental notation (a-d). For RHY and RAND, tone number is the same. Squared brackets mark the section that is different between conditions.

context (note examples of rhythm and random sequences in Figure 3).

MELODY EXPERIMENT

The melody experiment compared 80 structured melodies (condition MEL) with 80 randomized sequences (condition RAND). Musical themes were chosen from the classical and baroque period and were instrumental in character. Each theme consisted of eight measures and was built as a musical period; that is, each was made up of two phrases with a small pause, or caesura, in between. Melodic contour was roughly balanced between examples and had one of the following overall shapes: arched (24 times), ascending (18 times), descending (16 times), and oscillating

(22 times). From the viewpoint of Gestalt psychology, each melody was in accordance with basic Gestalt principles. First, successive tones of 0.1 s or less (i.e., an eighth note or shorter) met the Gestalt criteria of “temporal” and “spatial proximity.” Second, phrase and melody endings were consistent with the “principle of closure.”

Random sequences (condition RAND) were generated by independent permutation of pitch and time values (i.e., tone duration plus ISI) over the entire set of stimuli. This resulted in the complete randomization of the elements. Due to permutation of pitch, random sequences and melodies differed in their percentage of steps and leaps, that is, minor and major seconds, minor and major thirds, and leaps larger than a major third. However, tone

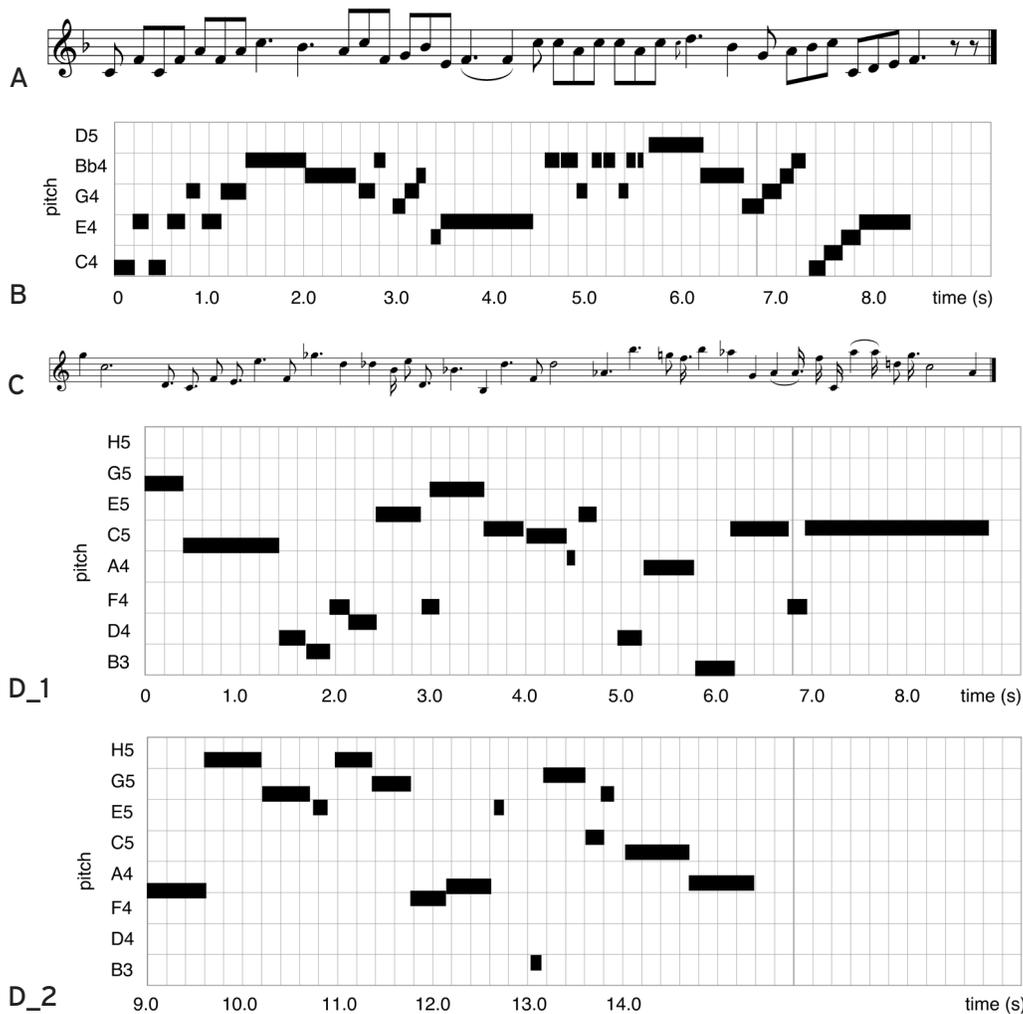


FIG. 4. Examples of a well-balanced melody (MEL), and of a randomized tone sequence (RAND), in musical and experimental notation (a-d). (a) C. Ph. E. Bach, "Air and Variations." For better comparability, notation is without bar lines and time signature. (b) Pitch-time diagram (playing protocol) of "Air and Variations." It shows onset times and tone durations of the recording. (c) Example of pitch and time values (duration combined with ISI), permuted independently over the entire set of examples. (d) Pitch-time diagram shows permutation of performance data, that is, pitch and time values of the playing protocol.

number (2,389 tones) and average pitch height (71.7 in MIDI note number, approximately C_5) between conditions were the same (see Figures 4a-d and 5 for note examples and histograms).

Melodies were played on a programmable keyboard (Yamaha PSR 1000) that was interfaced with a PC. Recorded examples were stored in MIDI format, using the music software Steinberg™ Cubasis VST 4.0. Random sequences were generated off-line by independent permutation of pitch and time values over all sequences. As pure permutation of time values would have caused overlaps in sound (the ISI was partly negative due to legato), tone onsets were partly adjusted within each sequence, so that the ISI was either zero or positive. Due

to this adjustment procedure, the time values between conditions (IOI, tone duration, and ISI) differed slightly from each other. Nonetheless, both conditions were rather similar in sequence length and time parameters. Melodies (cond MEL) had an average length of 10.35 s ($SD = 3.02$), and random sequences (cond RAND) had an average length of 10.3 s ($SD = 3.43$). For melodies, the average IOI was 0.32 s ($SD = 0.21$), the average tone duration was 0.29 s ($SD = 0.25$), and the average inter-stimulus interval was 0.06 s ($SD = 0.12$). Randomized sequences had an average IOI of 0.35 s ($SD = 0.26$), an average tone duration of 0.28 s ($SD = 0.25$), and an average ISI of 0.07 s ($SD = 0.11$). Thus, for both conditions, average tone density was almost the same.

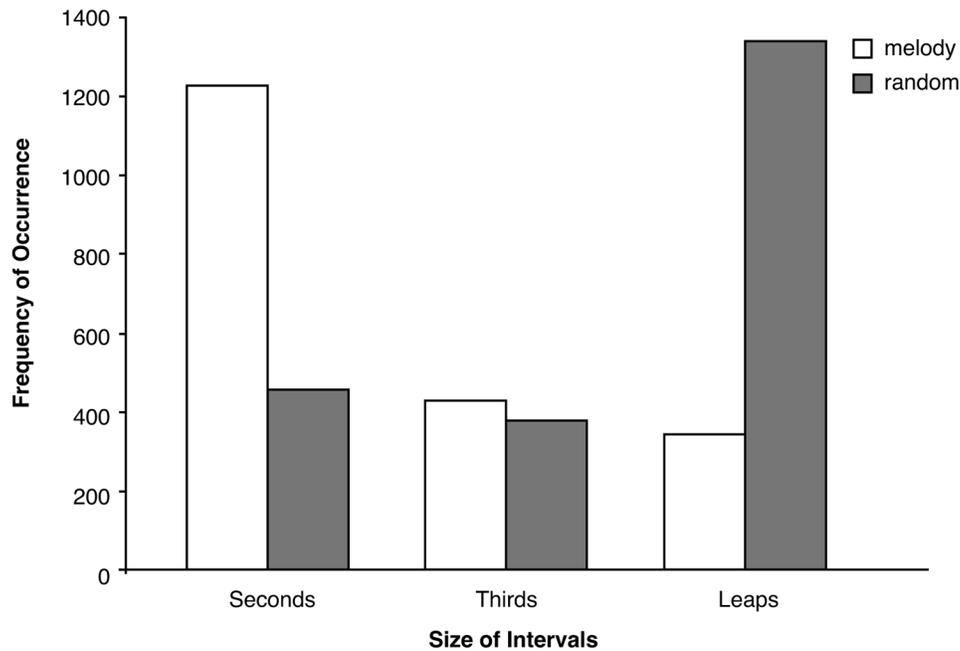


FIG. 5. Distribution of interval sizes for melodies (white columns) and for random examples (grey columns). Frequency of occurrence for seconds (major and minor), thirds (major and minor), and leaps (larger than a major third). Unisons are not included.

For the task, 160 additional tone excerpts, each with a length of 2 s, were prepared. These segments consisted either of the first 2 s or of the last 2 s of the full-length melodies or randomized versions. Melodies, random sequences, and excerpts were presented in four experimental blocks in pseudo-random order. Both sequence types had the same probability. The sound was “acoustic piano” (Sound Font no. 0) taken from the free set of instrumental samples (<http://freepats.opensrc.org>).

Paradigm, Task, and Procedure

For trial presentation, we used the ERTS software (Experimental Run Time System, Version 3.11, BeriSoft 1995). Each trial consisted of three parts. In part 1, a fixation cross was shown for 2 s. In part 2, a tone sequence, and, in the melody experiment, a subsequent additional tone excerpt, was added via loudspeaker. In part 3, subjects had to press the “yes” or “no” button whenever the question “With Pattern?” (rhythm experiment) or “Excerpt Included in Preceding Sequence?” (melody experiment) was displayed on the monitor.

Participants were seated on a comfortable chair in a soundproof, and electrically shielded, chamber. They were instructed to listen attentively to the stim-

uli. For all participants, audio material was the same. In the rhythm experiment, subjects had to detect the rhythmic sections embedded in an irregular context. In the melody experiment, subjects had to compare the tone excerpts with the preceding full-length example. They were asked to decide whether these excerpts, presented after each melody or random example, matched with the initial or the last 2 s sections of the preceding sequence. In each case, subjects had to press either the right or left button of a response-box. Before the measurement, a test run was carried out, and stimulus intensity was balanced individually.

In the recording period, subjects were asked to avoid eye-blinking as much as possible. They had to keep neck and face muscles relaxed, as well as hands, arms, and fingers. After recording, they were asked to fill out a questionnaire about their musical education and their current state of health. Button press responses of each participant were analyzed off-line.

Recording

Brain electrical activity was registered with small Ag/AgCl scalp electrodes inserted into a special electrode cap (Electro Cap International Inc., Eaton, Ohio). Fifty-seven active channels were used in the rhythm experiment,

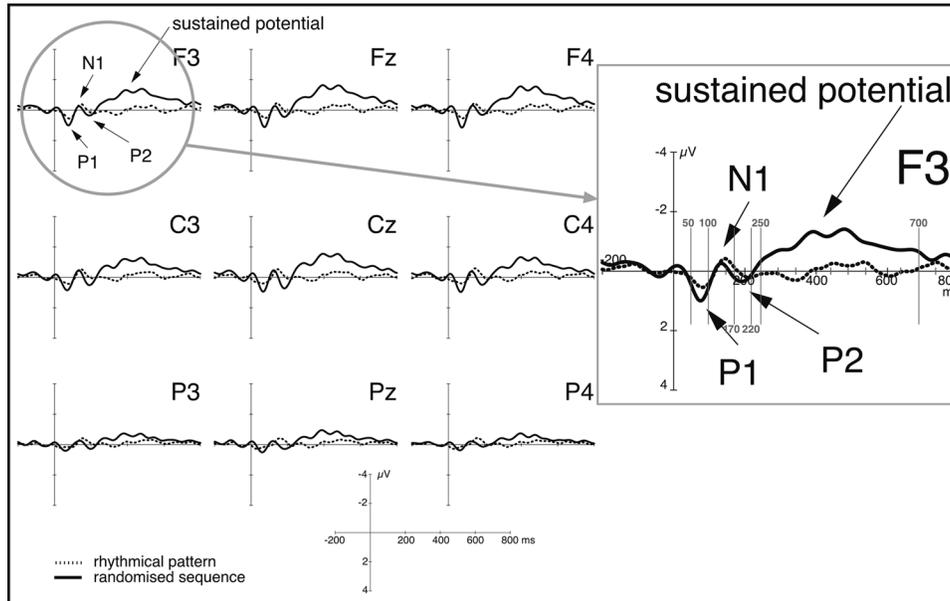


FIG. 6. Grand average ERPs for all subjects. Dotted line: Brain response to the embedded rhythmical section (RHY). Solid line: Brain response to the randomized counterpart (RAND). Recordings are from frontal (top row), central (middle row), and parietal electrode sites (bottom row). For each condition, ERP traces reflect merged activity over all tone onsets and sequences.

and 30 active channels were used in the melody experiment. Electrode placements were in accordance with the 10-10 electrode system (e.g., Oostenfeld & Praamstra, 2001). As an electrically inactive reference point, the left preauricular point (A1) was chosen. The ground electrode was placed on the sternum. For the recording of blink artifacts, vertical and horizontal electrooculograms (EOGH and EOGV) were registered from above and below the right eye, as well as from the outer canthus of both eyes. Impedance was kept below 5 k Ω for each electrode channel.

Data Analysis

PREPROCESSING OF DATA

Electrical signals of the brain were digitized online, with a sampling rate of 500 Hz. The time constant was infinite. To suppress slow potential drifts, the EEG data from each subject were filtered off-line, with a 0.50 Hz high-pass filter. Brain signals were examined for artifacts, in particular, those caused by eye blinks. For subsequent averaging, only those epochs of EEG traces were accepted which did not include amplitude shifts caused by eye blinks larger than 30 μ V (EOGH), and 40 μ V (EOGV). Before averaging, additional trigger points were set at each tone onset. Thus, each average trace for subject, electrode site, and condition consisted of artifact-free trials merged over all sequences and individual

tones. Due to the fact that at the very onset of a tone sequence, N1 and P2 components are larger than for subsequent tones (so-called onset-response), the first tone of each sequence was excluded from further analysis. Each trace was baseline-corrected, using a time window of -100 to 0 ms before tone-onset.³

For analysis and display, grand average potentials were computed over all subjects. With this approach, any individual differences in the onset and shape of the components are averaged out. For purpose of illustration, electrodes F3, Fz, F4; C3, Cz, C4; and P3, Pz, P4 were selected, and the time window was -200 to 800 ms, with respect to trigger-onset (see Figures 6 and 7).

STATISTICAL ANALYSIS

For the repeated-measures analysis of variance (ANOVA), four time windows in the rhythm experiment, and three time windows in the melody experiment, were chosen. Time intervals were of equal size,

³The longer the time window for baseline correction, the more reliable the average value is to free the ERP curve from small potential drifts. In the rhythm experiment, however, this consideration interferes with the ISI of 50 ms, which should guarantee a coherent impression of the tone sequence. However, an additional analysis with a "neutral" time window of 0 to 25 ms post tone-onset (in which brainstem activity is reflected by low-amplitude components) revealed that the selected time window for baseline correction had no substantial effect on the grand average data.

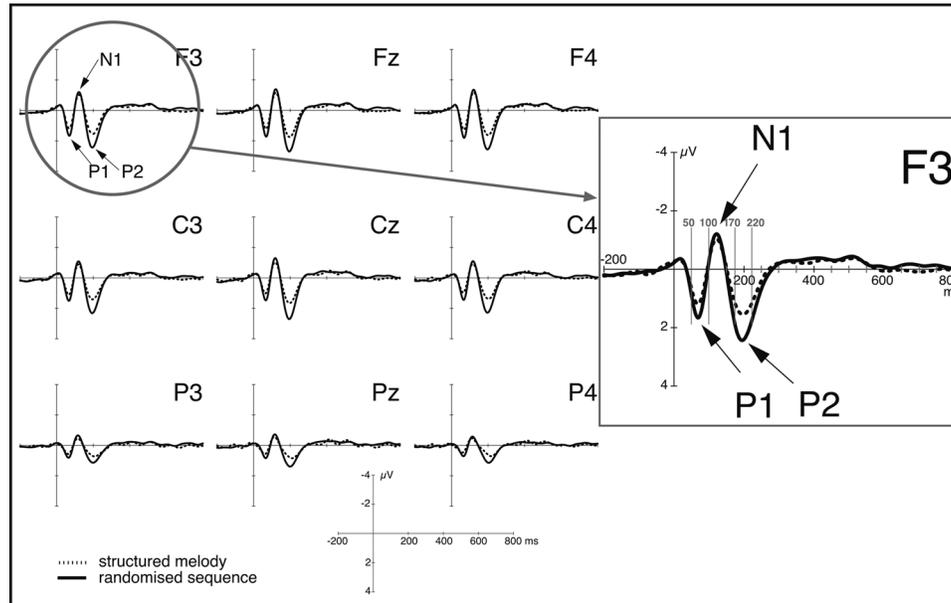


FIG. 7. Grand average ERPs for all subjects. Dotted line: Brain response to structured melodies (MEL). Solid line: Brain response to sequences with randomized pitch and time values (RAND). Recordings are from frontal (top row), central (middle row), and parietal electrode sites (bottom row).

that is, with an equal spacing of 25 ms on either side of component peaks, except for long-lasting waveforms. These time intervals were selected after visual inspection of grand average data and application of a running t test.⁴ Time windows in the rhythm experiment were 50 to 100 ms (P1), 120 to 170 ms (N1), 170 to 220 ms (P2), and 250 to 700 ms (sustained potential). Time windows in the melody experiment were 50 to 100 ms (P1), 100 to 150 ms (N1), and 170 to 220 ms (P2). Data were analyzed from nine electrodes evenly distributed over those areas of the scalp most relevant to auditory processing: F3, Fz, F4, C3, Cz, C4, P3, Pz, and P4. In order to account for topological differences between the ERP patterns, ERP channels were grouped with respect to two topological factors: anterior–posterior (anterior [F3, Fz, F4], central [C3, Cz, C4], and posterior [P3, Pz, P4]) and left–right (left [F3, C3, P3], midline [Fz, Cz, Pz], right [F4, C4, P4]). For each time window, channel, and condition, the dependent variable was the average amplitude (μV) relative to the baseline, for RHY versus RAND (rhythm experiment) and for MEL versus RAND (melody experiment).

Altogether, two types of ANOVAs were computed. First, we computed an overall repeated-measures

ANOVA over the entire time window (-200 to 800 ms), to test whether the time course of the ERP behaved differently with respect to time windows and experimental conditions. Within-subject factors were time window, the condition (RHY versus RAND, and MEL versus RAND), and left–right and anterior–posterior electrode placements. Second, in order to resolve possible interactions with the time window, we determined main effects and (simple) interactions with condition at each level of time window by computing a three-factor repeated-measures ANOVA, separately for each time window. Within-subject factors were the condition (RHY versus RAND, and MEL versus RAND), as well as left–right (two levels; left, right) and anterior–posterior electrode placements (three levels; anterior, central, posterior). In each analysis, degrees of freedom were corrected with Huynh and Feldt's epsilon. Results were considered significant at $p < .05$. However, marginally significant results, up to $p < .08$, were also included.

Results

Rhythm Experiment

BEHAVIORAL DATA

Subjects solved the detection task with an error rate of 3.04% on average. Errors included misses (rhythmic sections falsely identified as random sections: 3.93%)

⁴A t test without any multiple testing correction, performed at each electrode channel on each sample value (digitized time step).

and false alarms (random sections falsely identified as rhythmic sections: 2.14%). A paired-samples t test (two-tailed) showed that differences between error types (misses and false alarms) were not significant, $t(13) = -1.22, p > .05$.

ERP DATA

Figure 6 shows the grand average ERPs for rhythmic sections in contrast to those with continuous randomization. At each electrode site, curves reflect merged activity over all subjects and tones. Randomized sections revealed higher amplitude values than rhythmic sections for components P1 and P2 but quite similar deflections for the N1.⁵ Randomized versions also elicited a clear sustained negativity in the 250 to 700 ms time window post stimulus-onset. For early components, no effect of lateralization could be found between conditions. In the anterior–posterior direction, P1, P2, and the sustained potential decreased in amplitude.

STATISTICAL RESULTS

For the whole time range of analysis, the overall four-factor repeated-measures ANOVA revealed several significant effects, including the factor time window. In order to further specify these interactions, we computed the condition-related main effects and interactions at each level of time window, with separate three-way ANOVAs for each of these levels. For significant (first-order) interactions, additional post hoc tests were performed. This way, the difference between RHY and RAND was precisely determined for each time window and level of anterior–posterior topography (see Table 1 for details).

Melody Experiment

BEHAVIORAL DATA

When determining whether or not a tone excerpt matched the initial or final part of the preceding sequence, subjects had an overall error rate of 15.16%. Generally, final sections were correctly identified in 86.71% of the cases, whereas for initial sections, the

⁵As it is apparent from visual inspection, as well as from separate statistical analyses for each time window, conditions in both experiments, with regard to N1, do not differ significantly from each other. Thus, statistical data for N1 time windows (120 to 170 ms [rhythm experiment] and 100 to 150 ms [melody experiment]) are not presented in Tables 1 and 2. Likewise, results of the insignificant (simple) interactions Condition \times Left-right are not reported.

TABLE 1. Repeated-measures analysis of variance for condition (RHY vs. RAND)

Four-factor; Overall: -200 to 800 ms
Source df F
Time Window \times Condition 3,39 22.20***
Time Window \times Condition \times Left-right 6,78 4.13**
Time Window \times Condition \times Anterior-posterior 6,78 19.98***
Three-factor; per Time Window: 50 to 100 ms (P1)
Source df F
Condition 1,13 3.74(*)
Condition \times Anterior-posterior 2,26 5.97*
(post hoc: anterior region: Condition 1,13 4.40(**))
170 to 220 ms (P2)
Source df F
Condition 1,13 4.08(*)
Condition \times Anterior-posterior 2,26 3.39(*)
(post hoc: central region: Condition 1,13 5.20*
posterior region: Condition 1,13 5.56*)
250 to 700 ms (sustained potential)
Source df F
Condition 1,13 55.78***
Condition \times Anterior-posterior 2,26 37.24***
(post hoc: anterior region: Condition 1,13 68.56***
central region: Condition 1,13 48.40***
posterior region: Condition 1,13 15.76**)

(*) $p < .08$.

* $p < .05$.

** $p < .01$.

*** $p < .0001$.

percentage of correct answers was 82.98%. Further subdivision, according to sequence type, revealed that error rates for initial and final excerpts were four times higher when the preceding sequence was random. For structured examples, matching was false for 7.58% of initial parts and for 8.0% of final parts, whereas for random examples, the rate of incorrect responses was 26.05% for initial excerpts and 20.88% for final excerpts. Differences between error rates were highly significant for initial segments as parts of melodies, compared to initial segments as parts of random sequences, and the same was true for final segments (paired-samples t test, two-tailed; $t(14) = -11.2, p < .001$, and $t(14) = -7.64, p < .001$, respectively). However, for one sequence category, differences between error rates for initial and final excerpts were not significant: $t(14) = -1.18, p > .05$ for melody excerpts, and $t(14) = 1.83, p > .05$ for random excerpts.

ERP DATA

Figure 7 shows the grand average ERPs for structured melodies and for sequences with random pitch and

time values (tone duration in conjunction with ISI). Curves depict brain activity for both types of sequences averaged over all subjects and tones. Similar to the rhythm experiment, random versions elicited higher amplitudes than melodic versions, with respect to P1 and P2, whereas N1 components were almost congruent between conditions. In the later time window, no difference between conditions could be found. Comparable to ERPs in Figure 6, amplitudes had their maximum at the fronto-central electrodes. With respect to P1 and P2, activation was bilateral.

STATISTICAL RESULTS

Statistical tests were the same as in the rhythm experiment. We obtained a significant main effect of condition, as well as the significant interaction Condition \times Anterior-posterior, indicating general and conditional differences in amplitude and topography.

Interactions including the factor time window are resolved, using the condition-related main effects and interactions at the separate levels of the factor time window. For each level of time window, a three-factor repeated-measures ANOVA was computed. In case of significant first-order interactions (Condition \times Anterior-posterior), additional post hoc tests further specified the amplitude difference between conditions for each time window and level of topography (see Table 2 for details).

Discussion

EARLY COMPONENTS

In both experiments, significant interactions between time window, condition, and brain topography provide support for the differentiability of ERP effects in the analyzed time windows. Accordingly, P1, P2, and the sustained negative potential (in time windows 50 to 100 ms, 170 to 220 ms, and 250 to 700 ms, respectively) seem to be the measurable electrophysiological outcome of different underlying neuronal processes.⁶ Results of Liégeois-Chauvel, Musolino, Badier, Marquis, and Chauvel (1994) are evidence for the cortical origin of the observed components, with the P1 localized in the primary auditory cortex (lateral part of Heschl's gyrus), and the N1 localized in the secondary auditory cortex of

⁶An interaction between time window, brain topography, and experimental condition means that in all time windows, neuronal networks related to the processing of the difference between conditions are not the same. However, this does not implicitly mean that each component represents a single neuronal process independently from generation of the other two components.

TABLE 2. Repeated-measures analysis of variance for condition (MEL vs. RAND)

Four-factor; Overall: -200 to 800 ms	
Source	df F
Condition	1,14 21.75**
Condition \times Anterior-posterior	2,28 26.39***
Time Window \times Condition	2,28 25.97***
Time Window \times Condition \times Left-right	4,56 2.44(*)
Time Window \times Condition \times Anterior-posterior	4,56 6.35**
Three-factor; per Time Window: 50 to 100 ms (P1)	
Source	df F
Condition	1,14 15.61**
Condition \times Anterior-posterior	2,28 8.60**
(post hoc: anterior region: Condition	1,14 15.32**
central region: Condition	1,14 7.92*)
170 to 220 ms (P2)	
Source	df F
Condition	1,14 71.38***
Condition \times Anterior-posterior	2,28 19.55**
(post hoc: anterior region: Condition	1,14 38.68***
central region: Condition	1,14 33.84***
posterior region: Condition	1,14 5.68*)

(*) $p < .08$.

* $p < .05$.

** $p < .01$.

*** $p < .0001$.

the brain (mainly planum temporale). Conditions RHY versus RAND and MEL versus RAND already show amplitude differences for the P1. Accordingly, processing of regular and irregular patterns seems to take place already in primary auditory regions. Thus, our P1 and P2 results are in contrast to Griffiths et al. (2001), and Patterson et al. (2002). For sequence processing, they observed brain activation only in secondary auditory areas but could not provide evidence for further differentiation between regular and irregular versions. Note that in our study, bottom-up effects (e.g., refractory) can be excluded, since tone duration, average pitch, and distribution of inter-onset distances between tones (IOIs) were (nearly) the same between conditions.

With respect to amplitudes, P1 and P2 components reveal much higher values in the melody experiment than in the rhythm experiment. This result can be explained with the different inter-onset intervals between experiments (on average 0.171 ms for the rhythm experiment, and 0.320 ms for the melody experiment), suggesting that in the rhythm experiment, the recovery period for activated neurons is too short to develop components with full amplitude size. Moreover, because rhythmic and random sequences were presented with constant pitch (A₄, 440 Hz), the

same tonotopic area in the primary auditory cortex (Heschl's gyrus) was activated repetitively, which made the recovery period for pyramidal cells of the same neural tissue even shorter (for refractory effects in sequences with identical tones, see, e.g., Budd, Barry, Gordon, Rennie, & Michie, 1998; Kuriki, Kanda, & Hirata, 2006). Whereas early components *between* both experiments differ from each other with regard to physiological refractoriness (due to different IOIs), this is not the case *within* the same experimental setting. Thus, mere bottom-up effects cannot explain the observed conditional P1 and P2 differences in the same paradigm. Furthermore, both conditions did not differ substantially from each other with respect to average pitch height and tone duration. We therefore conclude that P1 and P2 differences, with regard to conditions, might reflect early cognitive processes relating to the regularity and/or the spatial and temporal cohesion between adjacent tones, as basic attributes of structure.

To summarize, P1 and P2 differences in both experiments might indicate that, depending on the degree of regularity in auditory sequences, the mental integration effort for current tone elements is different. As long as irregular pitch and/or time order has to be integrated to build a perceptual unit, mental effort is strong, and higher deflections for early components can be found. This means that whenever in (irregular) tone progression predictability—as a general means for certainty and descriptiveness of items—is low,⁷

⁷However, we should also think about the fact that listeners form strong expectations for contour reversal when tones follow a melodic leap, but not when tones follow a melodic step (see, e.g., von Hippel & Huron, 2000). If, in the current melody experiment, this is the case, expectations might differ with regard of context (random versus structure). However, due to the fact that in the random condition, tone progression was completely irregular, we assume that it was difficult for the listener to form expectations with regard to interval type, and the subsequent change of melodic contour. In order to disentangle tone predictability, as a meta-effect, from interval size as the local aspect, we computed grand average potentials for melodic reversals when the preceding interval was an upward third (major and minor) either in structured context (condition MEL) or in random context (condition RAND; not displayed for lack of space). Visual inspection of the grand average, in particular of early components (P1, N1, and P2), provided evidence that, although interval size was kept constant between sequence types, amplitudes of the P1, N1, and P2 were extensively larger for random versions as opposed to structured versions. These results confirm our opinion that amplitude differences between conditions cannot be attributed to interval size alone. Rather, it might be tone predictability, as a global factor, that has an effect on the encoding and integration of random contents versus structured contents, and interval size, as the local aspect, is probably incorporated. Further research should clarify the relationship between tone predictability, interval size, and listener's expectancy in detail.

higher shifts in amplitude can be observed, indicating stronger integration effort. Thus, already in this early stage of stimulus encoding (time ranges up to 200 ms), auditory processing might be influenced by top-down processes, such as predictability and selective attention. This interpretation seems to correspond well with the feature-integration theory of attention, proposed by Treisman and Gelade (1980). According to their theory, focused attention is a necessary top-down strategy, so that separate items can be integrated correctly, and coherent percepts are built. However, this interpretation might be challenged by recent results from the visual domain. With regard to perceptual integration of facial and body expression, (with either congruent or incongruent emotional meaning), Meeren, van Heijnsbergen, and de Gelder (2005) found amplitude differences in the occipital P1 (approximately 115 ms after stimulus-onset). This result suggests that higher-order processing might already take place at a very early encoding stage regardless of whether attention might play a role. Similarly, P1 differences in the present study might indeed be a correlate for top-down-driven integration processes, whereas the almost identical N1 might indicate that selective attention for both structured and random versions is constant. This also means that aspects of the task, and task difficulty, which require attention as a necessary precondition, cannot fully explain the observed component differences in the time range up to 170 ms.

LATE COMPONENTS

Randomized compared to rhythmic sections evoked a clear sustained negativity in the 250 to 700 ms time window. Amplitudes were largest at fronto-central electrode sites (Figure 6). Components in the later time window are commonly regarded as markers for higher cognitive and often task-relevant processes—for example, updating of the context in working memory (see, e.g., Comechero & Polich, 1999; Squires, Donchin, Herning, & McCarthy, 1977). In this study, the late sustained negativity might indicate a mismatch between stimulus structure, (i.e., completely random time events) and the mind's endeavor to detect regularity. Obviously, this matching procedure requires a high amount of effort.

This explanation is corroborated by the ideas of Large (2001; Large & Palmer, 2002), who brings the aspects of metrical structure and temporal expectancy into play. The idea behind is the role of meter as an active listening strategy, rather than a

static framework. According to Large (2001, 2002), this approach enables temporal expectation based on a grid of internal oscillations. With regard to the random sections in our rhythmic experiment, this grid of tiny internal oscillations is missing, so that the listener's expectation of underlying oscillations and rhythmic regularity is not fulfilled. Following the ideas of Large, violation of expectancy in the time domain, due to absence of a metrical grid, might be the reason for the stronger perceptual integration effort during the rhythmical random sections, and for this, the late sustained negativity is a neural correlate.

In contrast to results of the rhythm experiment, the processing of random sequences versus melodic Gestalts did not reveal any component difference in late time windows. Although button press responses show that the matching between tone excerpts and initial or final parts of the preceding full-length examples are four times more demanding for random sequences than for melodic sequences, memorizing, as opposed to pattern detection, did not elicit a task-related sustained negativity, or any other late component.⁸ With ERP, activation of neural cell assemblies underlying the process of memorization could therefore not have been made visible. Moreover, temporal expectancy, which seems to be the explanatory variable in the rhythm experiment, might be less pronounced in the melody experiment, where a combined effect of *various* pitch- and time-based relations over an entire tone sequence was examined. In summary, there seems to be no ERP effect that corresponds to the fMRI results of Levitin and Menon (2003, 2005), who found higher inferior prefrontal activation for structured musical pieces, but not for scrambled musical syntax. However, the ERP and fMRI methods feature quite different sensitivity profiles, in particular, with respect to the temporal and spatial extent of brain activity. ERP can detect very brief processes but requires the concurrent activation of a large mass of parallel pyramidal cells. FMRI, by contrast, can detect much smaller active brain areas. It does not rely on asymmetric cell morphology and the parallel alignment of cells but

tends to miss brief activations that do not consume much energy. Therefore, it is quite plausible that both methods observe partially different processes related to musical structure. While Levitin and Menon (2005) might indeed have found some kind of "structure tracker" located in Brodmann Area 47, our results with ERP reflect the cognitive effort to detect and integrate structure, when there is no regularity. Thereby, we conclude that structural processing starts very early (< 100 ms).

Conclusion

The processing of irregular sequences, in contrast to rhythmically regular patterns or rule-based melodic Gestalts, reveals the top-down influence of cognitive strategies at a very early stage of stimulus encoding. As in both experiments physical properties of tones were quite similar between conditions (i.e., pitch height, average tone duration, and IOI), mere bottom-up processing cannot account for the observed component differences. Instead, due to random permutation of pitch and/or time values, increased P1 and P2 amplitudes might indicate a stronger cognitive effort to integrate the tone elements of the current input, with low predictability. In the sense of feature-integration, proposed by Treisman and Gelade (1980), this means that integration effort is stronger when spatiotemporal conjunctions, that is, rule-based inner structures, are missing.

When a sequence of irregular time values is scanned for rhythmic regularity, violation of temporal expectancy based on a (missing) grid of tiny internal oscillations is reflected by a broad sustained negative potential. By contrast, irregular sequences, as opposed to *melodic* Gestalts, did not elicit any ERP correlate in later time windows when a memory task is given.

Author Note

We would like to thank Sofia Dahl and two anonymous reviewers for valuable comments on previous versions of the manuscript. We thank Ina Koch and Heike Böthel for careful EEG measurements.

Address correspondence to: Christiane Neuhaus, Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstraße 1a, 04103 Leipzig, Germany. E-MAIL neuhaus@cbs.mpg.de

⁸In the melody experiment, ERP data were analyzed for each full-length example, but not for the following tone excerpt. Thus, potential curves *do* reflect the period of memorizing but exclude the period of excerpt matching, which occurred in the time interval when button press responses were requested.

References

- BESSON, M., & MACAR, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, 24, 14–25.
- BUDD, T. W., BARRY, R. J., GORDON, E., RENNIE, C., & MICHIE, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: Habituation vs. refractoriness. *International Journal of Psychophysiology*, 31, 51–68.
- COMECHERO, M. D., & POLICH, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, 110, 14–30.
- DEUTSCH, D. (1994). Die Wahrnehmung audativer Muster. In N. Birbaumer, D. Frey, & J. Kuhl (Series Eds.) & W. Prinz, & B. Bridgeman (Vol. Eds.), *Enzyklopädie der Psychologie: Vol. 1. Wahrnehmung* (pp. 339–389). Göttingen: Hogrefe.
- GRIFFITHS, T. D., BÜCHEL, C., FRACKOWIAK, R. S. J., & PATTERSON, R. D. (1998). Analysis of temporal structure in sound by the human brain. *Nature Neuroscience*, 1, 422–427.
- GRIFFITHS, T. D., UPPENKAMP, S., JOHNSTRUDE, I., JOSEPHS, O., & PATTERSON, R. D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nature Neuroscience*, 4, 633–637.
- HILLYARD, S. A., TEDER-SÄLEJÄRVI, W. A., & MÜNTE, T. F. (1998). Temporal dynamics of early perceptual processing. *Current Opinion in Neurobiology*, 8, 202–210.
- KOFFKA, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace, & World.
- KÖHLER, W. (1929). *Gestalt psychology*. New York: Liveright.
- KURIKI, S., KANDA, S., & HIRATA, Y. (2006). Effects of musical experience on different components of MEG responses elicited by sequential piano-tones and chords. *The Journal of Neuroscience*, 26, 4046–4053.
- LARGE, E. W. (2001). Periodicity, pattern formation, and metric structure. *Journal of New Music Research*, 30, 173–185.
- LARGE, E. W., & PALMER, C. (2002). Perceiving temporal regularity in music. *Cognitive Science*, 26, 1–37.
- LEVITIN, D. J., & MENON, V. (2003). Musical structure in processed in “language” areas of the brain: A possible role for Brodmann Area 47 in temporal coherence. *NeuroImage*, 20, 2142–2152.
- LEVITIN, D. J., & MENON, V. (2005). The neural locus of temporal structure and expectancies in music: Evidence from functional neuroimaging at 3 Tesla. *Music Perception*, 22, 563–575.
- LIÉGEAIS-CHAUVÉL, C., MUSOLINO, A., BADIÉ, J. M., MARQUIS, P., & CHAUVÉL, P. (1994). Evoked potentials recorded from the auditory cortex in man: Evaluation and topography of the middle latency components. *Electroencephalography and Clinical Neurophysiology*, 92, 204–214.
- MEEREN, H. K. M., VAN HEIJNSBERGEN, C. C. R. J., & DE GELDER, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences*, 102, 16518–16523.
- NÄÄTÄNEN, R., TERVANIEMI, M., SUSSMAN, E., PAAVILAINEN, P., & WINKLER, I. (2001). “Primitive intelligence” in the auditory cortex. *Trends in Neurosciences*, 24, 283–288.
- NAKATA, T., & MITANI, C. (2005). Influences of temporal fluctuation on infant attention. *Music Perception*, 22, 401–409.
- OOSTENVELD, R., & PRAAMSTRA, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clinical Neurophysiology*, 112, 713–719.
- PANTEV, C., OOSTENVELD, R., ENGELIEN, A., ROSS, B., ROBERTS, L. E., & HOKE, M. (1998). Increased auditory cortical representation in musicians. *Nature*, 392, 811–814.
- PATTERSON, R. D., UPPENKAMP, S., JOHNSTRUDE, I. S., & GRIFFITHS, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, 36, 767–776.
- PERETZ, I., & ZATORRE, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, 56, 89–114.
- REYBROUCK, M. (1997). Gestalt concepts and music: Limitations and possibilities. In M. Leman (Ed.), *Music, gestalt, and computing: Studies in cognitive and systematic musicology* (pp. 57–69). Berlin: Springer.
- SHAHIN, A., BOSNYAK, D. J., TRAINOR, L. J., & ROBERTS, L. E. (2003). Enhancement of neuroplastic P2 and N1c auditory evoked potentials in musicians. *Journal of Neuroscience*, 23, 5545–5552.
- SQUIRES, K. C., DONCHIN, E., HERNING, R. I., & MCCARTHY, G. (1977). On the influence of task relevance and stimulus probability on event-related potential components. *Electroencephalography and Clinical Neurophysiology*, 41, 1–14.
- SUSSMAN, E., RITTER, W., & VAUGHAN, H. G. (1998). Predictability of stimulus deviance and the mismatch negativity. *NeuroReport*, 9, 4167–4170.
- SUSSMAN, E., WINKLER, I., RITTER, W., ALHO, K., & NÄÄTÄNEN, R. (1999). Temporal integration of auditory stimulus deviance as reflected by the mismatch negativity. *Neuroscience Letters*, 264, 161–164.
- TREISMAN, A. M., & GELADE, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97–136.
- VON EHRENFELS, C. (1890). Über “Gestaltqualitäten.” *Vierteljahrsschrift für wissenschaftliche Philosophie*, 14, 249–292.
- VON HIPPEL, P., & HURON, D. (2000). Why do skips precede reversals? The effects of tessitura on melodic structure. *Music Perception*, 18, 59–85.
- WERTHEIMER, M. (1922). Untersuchungen zur Lehre von der Gestalt. I. Prinzipielle Bemerkungen. *Psychologische Forschung*, 1, 47–58.