Let's face the music: A behavioral and electrophysiological exploration of score reading

THOMAS C. GUNTER,[a] BJÖRN-HELMER SCHMIDT,[a] AND MIREILLE BESSON[b]

[a]Max-Planck-Institute of Cognitive Neuroscience, Leipzig, Germany
[b]Centre National de la Recherche Scientifique, Institut de Neuroscience Physiologiques et Cognitives, Marseille, France

Abstract

This experiment was carried out to determine whether reading diatonic violations in a musical score elicits similar endogenous ERP components when hearing such violations in the auditory modality. In the behavioral study, musicians were visually presented with 120 scores of familiar musical pieces, half of which contained a diatonic violation. The score was presented in a measure-by-measure manner. Self-paced reading was significantly delayed for measures containing a violation, indicating that sight reading a violation requires additional effort. In the ERP study, the musical phrases were presented in a "RSVP"-like manner. We predicted that diatonic violations would elicit a late positive component. However, the ERP associated with the measure where a violation was presented showed a negativity instead. The negativity started around 100 ms and lasted for the entire recording period. This long-lasting negativity encompassed at least three distinct effects that were possibly related to violation detection, working memory processing, and a further integration/interpretation process.

Descriptors: Music processing, Music reading, Score reading, Electrophysiology, Reaction time, Electroencephalogram, Event-related brain potential, Early right anterior negativity, Right anterior temporal negativity

From a musical point of view, reading and writing musical scores is a very important skill, as it is the only universal means of communication between composers and musicians. With a score, we can play music that was composed hundreds of years ago. Although the oldest scores date from the 13th century, showing pitches and musical instruments, but with no indication of rhythm, the modern concept of the score was first developed in the 14th century through manuscripts of instrumental music (cf. Sadie, 1995). Similar to written manuscripts for languages, musical scores are an extremely important cultural inheritance, representing the musical memory of our human civilization. From this perspective, it is astonishing that, although extended research has been devoted to language reading, so little cognitive research has been directly aimed at exploring the reading of musical scores. Moreover, it is unfortunate that most of the literature on sight reading[1] and sight singing is anecdotal and almost exclusively focused on output performance (cf. Rogers, 1984; for a more recent example, see Lehmann & Ericson, 1996).

There are, however, a few experiments on eye movements during sight reading (i.e., Rayner, 1998, Rayner & Pollatsek, 1997). Results show that the region around fixation from which information is extracted, the perceptual span, has a width of approximately 1 measure right of the fixation point (Truitt, Clifton, Pollatsek, & Rayner, 1997). The eye-hand span (i.e., how far the eye reads the score before the actual motor programs have been carried out) is, contrary to what musicians believe, relatively small, being between approximately 2 and 4 beats (Rayner & Pollatsek, 1997). Thus, visual processing of skilled music readers is not very far ahead of the hands and the actual position in the score.

Because music reading represents a complex transformation task, it is obvious that many different types of processing abilities must underlie music-reading expertise. Waters, Underwood, and Findlay (1997) showed, for instance, that perceptual pattern recognition was more efficient in experienced musicians in that they were able to perform a score comparison task with fewer and shorter glances between the patterns than less experienced persons. Sight reading might therefore be associated with an ability to rapidly perceive notes or groups of notes.

Although this research provides interesting information regarding the perceptual and output processes during sight reading, it is also important to understand the more central, cognitive, information-processing stages. The use of methods with a very high temporal resolution is therefore necessary. In the present study, we investigate the cognitive aspects underlying music reading by means of electrophysiology.

[1]Note that sight reading and music reading are used interchangeably in this article.

In the last 10 years or so, investigating the auditory perception of music by means of electrophysiology has received increased interest (see, e.g., Peretz & Herbert, 2000). The first event-related brain potential (ERP) experiment using music materials was carried out by Besson and Macar in 1987. Although the authors were not directly interested in music processing itself, but in the language specificity of an ERP component called the N400,[2] it was shown that an out of key note in auditorily presented melodies or scales elicited a P300 and not an N400 component. Similar results were also found by Paller, McCarthy, and Wood (1992), who replicated the Besson and Macar (1987) study using more controlled conditions. Paller et al. showed that deviant endings to melodies elicited only a P300 without an overlapping covert N400. However, in these early studies (see also Verleger, 1990, 1991), the manipulation used for establishing a violation was rather crude. It is therefore not surprising that recent experiments, in which the manipulations are more subtle, clearly show a more differentiated pattern of results.

Besson and Faïta (1995) presented musicians and nonmusicians with both well-known and unfamiliar melodies in which the final note was either correct, a diatonic, a nondiatonic, or a rhythmic violation. As expected, musicians performed better on a postmelody recognition and categorization task than nonmusicians. Moreover, diatonic and nondiatonic violations elicited a late positive component (LPC) with a larger amplitude and earlier peak latency for the nondiatonic than diatonic violations. Finally, musicians showed larger LPCs than nonmusicians. Besson and Faïta suggested that the LPC is sensitive to musical expectancies, and that it is a reliable index of the time course of the mental operations involved in violation detection in the musical domain.

More recently, an important trend can be seen in studies on the electrophysiology of music processing that suggests that there may be interesting parallels between the processing of language and music. Both domains show, for instance, syntactic principles (cf. Lerdahl & Jackendoff, 1983; Sloboda, 1985). Patel, Gibson, Ratner, Besson, and Holcomb (1998) were the first to suggest, on the basis of electrophysiological data, that there are similarities between syntactic processing in language and harmonic processing in music. In this experiment, subjects were presented with auditory sentences and musical phrases that sometimes included syntactic or harmonic violations, respectively. Violations in both domains elicited a slow positive shift with a maximum around 600 ms, which was identified as a P600 (note that the P600 component showed a high similarity with Besson and Faïta's LPC). In language experiments, this component is typically considered as reflecting syntactic processes of structural reanalysis or repair (Osterhout, Holcomb, & Swinney, 1994; Osterhout & Mobley, 1995; see also Gunter & Friederici, 1999; Gunter, Stowe, & Mulder, 1997; Hagoort & Brown, 2000; Osterhout & Hagoort, 1999) or as marker for syntactic integration (cf., Kaan, Harris, Gibson, & Holcomb, 2000). On the basis of its occurrence in music processing, Patel et al. suggested that the P600 probably reflects a more general knowledge-based integration process that is carried out during the analysis of rule-governed sequences. Note that the harmonic violations also elicited an earlier negativity, the so-called N350 or right anterior temporal negativity (RATN). Patel et al. suggested that the N350 might reflect the use of music-specific rules, or music-specific

working memory resources, as this component showed similarities with the so-called left anterior negativity (LAN) elicited in language (cf. Kluender & Kutas, 1993).

Besson, Faïta, Peretz, Bonnel, and Requin (1998) explored the interaction between language and music more extensively by using opera scripts in which linguistic (i.e., semantic) and harmonic (i.e., out-of-key) violations were fully crossed. The semantic violations elicited a clear N400 component, a component known to reflect semantic integration in language processing (for a recent review on N400, see Kutas & Federmeier, 2000). The out-of-key tones elicited a clear LPC (cf. P300), and the double violation a (quite early) N400 and a LPC. These data were taken to indicate a strong independence of semantic and harmonic processing and seem to argue against a common cognitive factor underlying semantic processing in language and harmonic processing in music. Note, however, that visual inspection of the data seem to indicate that, at least at the frontal electrodes, no clear signs of independence can be found. The data recorded at these electrodes seem to suggest an interaction between harmony and semantics because both the N400 and the P300 are reduced in the double violation condition.

Koelsch, Gunter, Friederici, and Schroeger (2000) performed four studies in which nonmusicians were presented with chord sequences (i.e., cadences) comprising five chords. Sometimes, the cadence was ended by a so-called Neapolitan sixth chord, a chord that is harmonically correct but highly unexpected. Two ERP components of most interest were elicited: an early right anterior negativity (ERAN), with a maximum amplitude around 150 ms, and an N500, with a frontal distribution. The ERAN was theorized to reflect processing of sound expectancy violations, whereas the N500, because it had characteristics similar to the N400, was hypothesized to reflect contextual integration processes. The ERAN elicited in music and the early left anterior negativity (ELAN) elicited by word category violations in language (cf. Friederici, Pfeifer, & Hahne, 1993) are similar in that they can both be associated with the early detection of a rule-based violation. In recent MEG studies (Friederici, Wang, Herrmann, Maess, & Oertel, 2000; Maess, Koelsch, Gunter, & Friederici, 2001) it was shown that the brain areas associated with both components overlap significantly. Applying dipole localization, Maess et al., for instance, showed that the ERAN (music) and the ELAN (language) are generated in a network of brain structures including the inferior frontal gyrus (Broca's area) and portions of the superior temporal gyrus bilaterally, with a left preponderance for the language-related early negativity and a right preponderance for the music-related negativity. Thus, as in the Patel et al. (1998) study, these results also suggests similarity between music and language processing.

Additional evidence for some overlap in the processing of structural information in language and music is provided by a recent fMRI experiment by Koelsch et al. (2002). It was found that there are large similarities in the network of brain structures activated during the processing of structural violations in music and language. Although of uttermost importance for the general problem of language specificity, these results may not be surprising, when taking into consideration that both music and speech processing imply that a dynamic stream of sounds is transformed into a system of discrete units with a rule-based hierarchical organization.

As mentioned above, another resemblance between music and language is that they both use a written code to store specific

---

[2]The N400 reflects semantic processes (see Kutas & Hillyard, 1980; Kutas & Federmeier, 2000).

information for future purposes. In the linguistic domain, it has been found that, although there are some minor differences, processing speech and written language elicit similar endogenous components like the (E)LAN, N400, and P600 (cf. Gunter, Friederici, & Hahne, 1999; Kutas & Federmeier, 2000). Because all ERP experiments on music processing except one (Schön & Besson, 2002) have been conducted using auditory materials, it is of interest to explore the neurophysiological basis of the reading process in music. If music reading is similar to reading in language, one would expect that the endogenous components elicited during auditory processing should also be observable in music reading.

In the present experiment, the well-known musical phrases used in the Besson and Faïta (1995) study were transcribed by a professional musician. A diatonic violation was introduced at different positions in the middle of the musical phrases but was always located at the beginning of the measure. Although it is clear from the Besson and Faïta study that diatonic violations elicit a smaller P600 compared to nondiatonic violations, it was necessary to use diatonic violations in the present experiment because they cannot be visually identified on the basis of any accidentals. Indeed, in C major, the diatonic tones are: c, d, e, f, g, a, and b whereas, for instance, the nondiatonic sharp tones are c♯, d♯, f♯, g♯, and a♯. These nondiatonic sharp tones do have additional ♯ accidentals in front of the note to indicate that they represent a tone which is played a semi-tone higher.[3] Participants can therefore identify nondiatonic violations on a visual basis alone, without interpreting the score.

Before conducting the ERP experiment we carried out a behavioral study (Experiment 1) to determine whether a measure-by-measure presentation format could be read by musicians. The score was presented in a self-paced manner, such that every time a button was pressed, a new measure appeared in the middle of the screen. Exploring reading processes by means of self-paced reading is a commonly used method in language research (cf. Haberlandt, 1994). At sentence positions where a (parsing) difficulty arises, the self-paced reading time slows down. On the basis of these results, we hypothesized a slowing down of the self-paced music reading times within the measures containing a violation.

In the ERP experiment (Experiment 2), the same musical phrases were presented in a rapid serial visual presentation (RSVP)-like manner (3,000 ms per measure, 200 ms blank screen). If reading music uses processes similar to those used in listening to music, one would expect that sight reading a violation would elicit endogenous components similar to those elicited by hearing such a violation. On the basis of the Besson and Faïta (1995) study, we therefore expected that the diatonic violation would elicit a late positive component in our musician subjects.

EXPERIMENT 1: SELF-PACED READING OF MUSICAL SCORE

Method

Participants

Eight musicians (4 women, 27.6 years, range: 23–32) were paid for their participation. They were students at the Hoch Schule für Musik in Leipzig (i.e., advanced musical school) and had between 17 and 26 years (mean 21 years) of training in classical Western music.

Stimuli and Presentation Sequence

One hundred twenty familiar musical phrases were selected from the materials of the Besson and Faïta (1995) study on the basis of a pretest carried out on 6 musicians not tested in this study. The original phrases were transcribed by a professional musician, who also created a version of every phrase that included a diatonic violation in the middle of the phrase. Variation of the violation position across all experimental items was between 2 and 12 measures after the beginning of the score, making it impossible for participants to anticipate the violation in a very narrow time window. This variation, however, did not affect the overall contour of the musical phrases (i.e., participants in the pretest could recognize the musical phrases on the basis of the previolation measures). Because the visual presentation format of the materials was done on a measure-by-measure basis, the violation was incorporated at the beginning of the measure. This ensured a very stable trigger point. The score was presented in a self-paced manner, such that every time a button was pressed a new measure appeared in the middle of the screen. Half of the phrases contained a violation; the other half were correct. Across the 8 participants, two experimental versions were presented, thus balancing out the harmonic correctness of the musical phrases (i.e., 4 participants per version).

The musical phrase started with the presentation of the bar including clef, key, and time signatures. This information stayed on the screen until the phrase was ended. Two hundred milliseconds after a button was pressed, the first measure was blended in onto the bar including the signs for a maximum of 5,000 ms or until the participant responded. After the next button press, the notes of the first measure were immediately removed and the notes of the next measure were blended onto the bar after 200 ms (cf. Figure 1). This procedure mimics what is called self-paced reading in the domain of language processing (cf. Kieras & Just, 1984). When the end of the score was reached, the participant had to indicate whether or not a violation was present in the score and if she or he knew the musical piece from which the score was derived. Then, the next musical phrase started. Note that only correctly answered items that were familiar[4] to the participant were used in the analysis of the self-paced reading data. Thus, approximately 15% of the data were rejected. The black notes were presented on a light gray background, and had a visual angle between approximately 1 and 8°.

Procedure

A session lasted approximately 1 hr. Participants were instructed to read the score very carefully in a self-paced manner, and to answer the postscore questions as precisely as possible. Before the experiment, a short training session, including five musical phrases, was provided. None of the training materials was used in the following experiment. Participants were comfortably seated in front of a color monitor, at a distance of approximately 110 cm.

---

[3]The flat notes will have an additional ♭ accidental, indicating that they represent a tone that is played a semi-tone lower.

[4]Familiarity was important because the violations were diatonic and are therefore undetectable if the phrase is unfamiliar because they do not constitute a violation in themselves.

Figure 1. Illustration of the presentation format used in both experiments.

Results

As can be seen in Table 1, self-paced reading times were approximately between 1,700 and 3,000 ms. In the measure in which a violation was presented (reading times in the second column, Table 1), a clear and significant difference in reading time was found. The measure including the violation was read around 500 ms slower than the correct one. Interestingly, self-paced reading was somewhat faster in the measures following the violation condition. This effect was, however, nonsignificant except for the fourth measure after the violation. Finally, participants' reading times became overall faster as the musical sequence unfolded.

Discussion

The self-paced reading data indicate that reading a measure that includes a violation costs additional processing effort. Thus, this result is in line with those obtained in language experiments in which a violation in a sentence or text also increased reading times (cf. Haberlandt, 1994). The overall decrease in reading times as the musical sequences unfolded might be related to the position of the violation that was always included in the middle of the score. That is, it was less probable that a violation would be present further downstream in the phrase. Insofar as the detection of a violation is highly relevant for the postphrase question, relief from this detection process will make the self-paced reading easier and therefore faster. Alternatively, one could argue that subjects are "tuning" into the musical phrase, making the score easier to process and self-paced reading faster. Most importantly, however, is the finding that the presentation format (i.e., measure-by-measure presentation) is sensitive to the diatonic violation, and that the musicians were able to read

the scores using such a measure-by-measure presentation.

An RSVP (see Kieras & Just, 1984) experiment was used in the second experiment to explore whether or not in-key violations, as in the Besson and Faïta (1995) study, would elicit a late positive component (P300) when presented visually. This would provide additional information regarding the question of whether written and auditory music processing share similar features.

EXPERIMENT 2: ERPS DURING RSVP READING OF MUSICAL SCORE

Method

Participants
Twenty right-handed musicians were paid for their participation. All participants had not been part of the behavioral study, and had normal or corrected-to-normal vision. They were students at the Hoch Schule für Musik in Leipzig (i.e., advanced musical school) and had between 13 and 25 years (mean 17 years) of training in classical Western music. In the final analysis, only 11 participants reached the criterion of having a music reading performance that was above 60% correct. These 11 participants (8 women) had a mean age of 22.5 years (range 19–31 years) and a mean performance of 74% correct responses.

Stimuli and Presentation Sequence
The same materials were used as in Experiment 1. The presentation format was based on a measure-by-measure presentation using an RSVP procedure (3,000 ms by measure and 200 ms blank screen in between).

Procedure
A session lasted approximately 3 hr. Participants were seated in a dimly lit room, facing a color video screen at a distance of 100 cm. They were instructed to read the score as attentively as possible and to answer two postscore questions (Was there a violation? and Do you know this piece of music?) as accurately as possible. Thus, musicians were first required to indicate the presence or absence of a violation by means of a Yes/No button press. Then, they had to respond a second time to indicate whether or not they knew the piece. The participants were asked to blink after the postscore tasks were completed.

Recordings
The electroencephalogram (EEG) was recorded from 54 Ag–AgCl electrodes (electro-cap) from Af3, AfZ, Af4, F7, F5, F3, Fz, F4, F6, F8, Ft7, Fc5, Fc3, Fcz, Fc4, Fc6, Ft8, Ft7, Fc5, Fc3,

Table 1. *Self-Paced Reading Times (in Milliseconds) of the Measure before the Violation, the Measure Containing the Violation and Four Following Measures*

Measure	Correct phrase	Violated phrase	Violation effect	ANOVA $F(1,7) =$
Violation −1	2,927	3,022	95	0.28, n.s.
Violation	2,551	3,046	495	9.47, $p < .01$
Violation +1	2,467	2,355	−112	3.45, n.s.
Violation +2	2,237	2,014	−223	2.97, n.s.
Violation +3	2,033	1,992	−41	0.2, n.s.
Violation +4	1,873	1,662	−211	8.29, $p < .02$

The first column contains the self-paced reading time of the correct phrase; the second column, incorrect phrase; the third column, the difference between the self-paced reading times of the incorrect and correct phrases; and the fourth column, the results of the ANOVA.

Fcz, Fc4, Fc6, Ft8, T7, C5, C3, Cz, C4, C6, T8, Tp7, Cp5, Cp3, Cpz, Cp4, Cp6, Tp8, P7, P5, P3, Pz, P4, P6, P8, Po7, Po3, Poz, Po4, Po8, O1, Oz, O1, and right mastoid each referred to the left mastoid (nomenclature as proposed by the American Electro-encephalographic Society, 1991). Bipolar horizontal EOG was recorded between electrodes at the outer left and right canthus. Bipolar vertical EOG was recorded between electrodes above and below the subject's right eye. Electrode resistance was kept under 5 KΩ. The signals were recorded continuously with a bandpass between DC and 70 Hz and digitized at a rate of 250 Hz. For presentation purposes only, the wave forms presented in the figures were off-line filtered with a low pass filter of 10 Hz (zero phase shift, finite impulse response type). All analyses were carried out on unfiltered data.

Data Analysis

Average ERPs, starting 200 ms before and lasting 2,000 ms after the presentation of a measure were computed for each electrode position for both the correct and violated measures. Averages were computed for the measure at the violation and at the first and second measures after the violation. Only trials in which both postscore questions were correctly answered were included in the averages.

Approximately 5% of the trials were excluded from the averages due to ocular- and/or amplifier-saturation artifacts (EOG rejection $\pm 50 \mu V$). Averages were aligned to a 200-ms prestimulus baseline. To describe the onset and length of the ERP effects in reasonable details within the first 1,000 ms, statistical analyses were carried out using 20 latency windows of 50 ms duration each. The interval name refers to the onset of the specific latency window (e.g., interval "50" represents the mean amplitude between 50 and 100 ms, etc.). In the time range between 1,000 and 2,000 ms, statistical analyses were performed using 10 latency windows of 200 ms duration each.

Violation was the within-subjects variable in the repeated measures ANOVAs that were used to quantify the effects statistically. Whenever significant effects in a particular latency range are reported (e.g., 100–350 ms), all smaller windows in this latency range showed significant effects, $p < .05$. To minimize the danger of false positives due to the large number of comparisons, effects are only presented when two or more successive latency windows showed significant effects. To analyze scalp distribution of the N1 (170–190 ms), ERAN (100–150 ms), RATN (200–350 ms), and of components in the middle (500–1,000 ms) and late time ranges (1,000–2,000 ms), representative time windows were computed for four regions of interest (ROIs) at left anterior (LA: FT7, FC5, FC3, F7, F5, F3, AF7, AF3), right anterior (RA: FT8, FC6, FC4, F8, F6, F4, AF8, AF4), left posterior (LP: TP7, CP5, CP3, P7, P5, P3, PO7, PO3), and right posterior (RP: TP8, CP6, CP4, P8, P6, P4, PO8, PO4) electrodes. In these scalp distibution analyses, violation (2), anterior/posterior (2) and hemisphere (2) were used as within-subjects variables. The Geisser–Greenhouse correction (Vasey & Thayer, 1987; Geisser & Greenhouse, 1959) was always applied when evaluating effects with more than one degree of freedom in the numerator.

Results

ERP Effects in an Early Time Range (up to 1,000 ms)

As can be seen in Figure 2A–C, reading a measure gave rise to several ERP components: A clear N1 component was found at
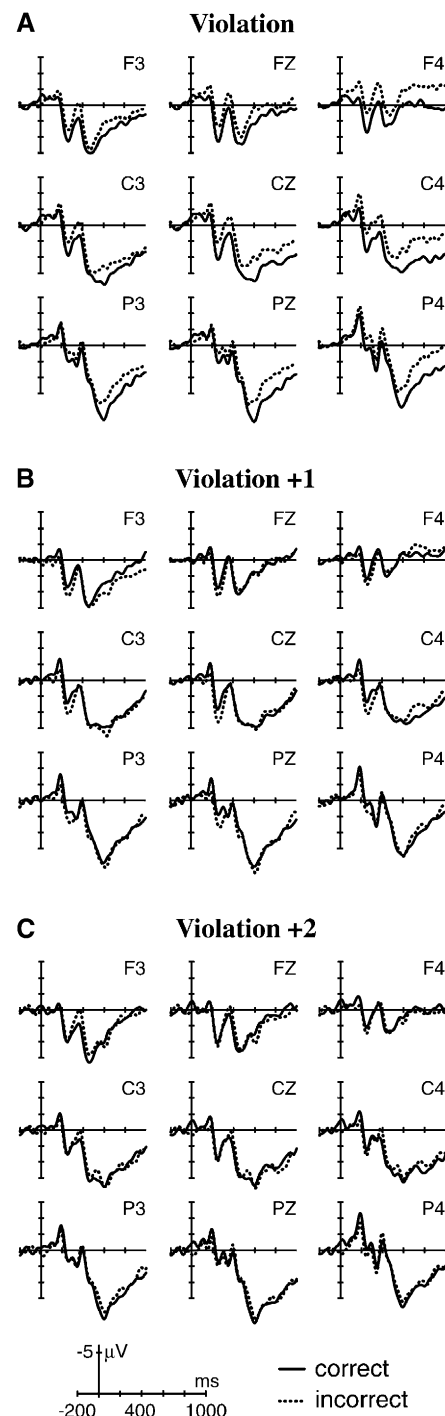

Figure 2. ERPs recorded in the measure containing the correct and incorrect (i.e., nondiatonic violation) notes (A) and in one or two measures after the violation was presented (B and C, respectively).

the occipital electrodes with maximum amplitude at the right side. The N1 was followed by a P2, after which an N350 (or RATN; see Patel et al., 1998) was elicited with a maximum at right frontal electrodes. Around 600 ms, these components were followed by positivity at parieto-occipital electrodes and by an increased negativity at frontal electrodes (e.g., F4). For the purpose of the present experiment, it is important to know how these components were modulated by the reading of a musical violation.
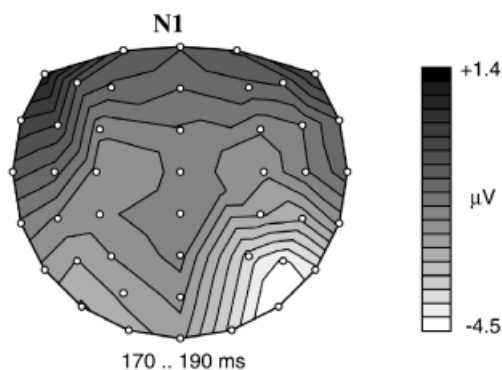
Figure 3. Scalp distribution of the occipital N1 component, which had a latency of approximately 180 ms.

N1. As can be seen in Figure 3, the N1 component elicited during music reading was lateralized to the right. This was confirmed by the ROI analysis in the N1 time window in which only the two posterior ROIs were used (i.e., LP and RP). This analysis showed a main effect of hemisphere, $F(1,10) = 17.7$, $p < .002$, and an interaction of Hemisphere $\times$ Violation, $F(1,10) = 7.63$, $p < .02$. This interaction was due to a right frontally distributed violation related negativity (see below) that also extended to more posterior electrodes. The main effect, however, shows that the N1 had a right posterior preponderance.

ERAN/RATN. As can be seen in Figure 2A, the reading of the violations gave rise to a prolonged negative shift over the right hemisphere. This negative shift was no longer present in the next two measures (see Figure 2B,C). The 50-ms interval analysis showed that this slow negative shift comprises three different effects. Two early fronto-central effects between 100 and 350 ms, $F(1,10) = 5.29–15.26$, $p < .05$, and a more broadly distributed right-sided negativity between 500 and 1,000 ms (see next paragraph). The early fronto-central effects started between 100 and 150 ms at right frontal sites and shifted toward right central sites between 200 and 350 ms (see Figure 4A,B). The difference between the early (100–150 ms) and later (200–350 ms) part of the negative shift (i.e., ERAN vs. RATN) was confirmed by the ROI analyses. The main analyses, which included the additional within-subjects variable component (early vs. late part) showed a highly significant three-way interaction of Component with Anterior/Posterior and Hemisphere, $F(1,10) = 32.21$, $p < .0002$, which remained significant after McCarthy and Wood's (1985) correction procedure, $F(1,10) = 9.73$, $p < .01$. This interaction suggests that the scalp distribution of the early and later part of the negativity is

different. However, it is critical to know whether the distributional difference also holds for the right anterior electrodes and whether the interactions with violation are significant. An additional analysis was therefore carried out using two new ROIs that describe the scalp distribution within both time intervals as precisely as possible (right frontal: F6, F8, FC6, and FC8; middle centro-frontal: F4, FC4, C4, and C6). This analysis showed significant main effects of component, $F(1,19) = 19.63$, $p < .001$, and violation, $F(1,10) = 7.98$, $p < .02$, and a marginally significant three-way interaction of Component with ROI and Violation, $F(1,10) = 3.43$, $p < .09$. This interaction became significant after the McCarthy and Wood (1985) correction procedure was applied, $F(1,10) = 6.48$, $p < .029$. Although less optimal from a statistical point of view, but backed up by the data of the correction procedure, we feel confident enough to suggest that the early and late time windows of the negativity between 100 and 350 ms have different scalp distributions, in that the violation effect shifts from a right anterior to a more middle central position.

Negativity after 500 ms. After 500 ms, the negativity became much broader in its scalp distribution and more developed over the more posterior electrodes. As can clearly be seen in Figure 4C, the negativity had a right preponderance (cf. 50-ms analyses: AF4, F4, F6, FC4, FC6, C4, C6, CP4, CP6, P6, PO4, PO8, O2; $F(1,10) = 4.98–40.35$, $p < .05$). This observation was confirmed by the ROI analysis in the 500–1,000 ms interval. Significant main effects of hemisphere, $F(1,10) = 23.75$, $p < .0007$, and violation, $F(1,10) = 5.04$, $p < .05$, were found. The two-way interaction of Hemisphere $\times$ Violation was significant, $F(1,10) = 18.62$, $p < .0015$, even after the McCarthy and Wood (1985) procedure was applied, $F(1,10) = 7.68$, $p < .02$. Thus, the ROI analysis clearly shows that the middle latency negativity had a right hemisphere preponderance. It is important to note that the ERPs elicited in the measures after the violation was presented (Figure 2B,C) did not show any significant violation-related difference in the time range up to 1,000 ms.

ERP Effects in the Late Time Range (1,000–2,000 ms)
Analyses in this late time window were computed using 10 intervals of 200 ms each, covering 1,000 to 2,000 ms (note that the presentation time of one measure was 3,000 ms). The negativity that started to develop after 500 ms (see above) is a long-lasting effect (see Figure 5A). Significant differences between correct notes and violations were only found for electrodes on the right side in all 200-ms time intervals (1,000–2,000 ms: $F(1,10) = 3.91–19.25$, $p < .05$). Midline electrodes showed such differences 400 ms later (1,600–2,000 ms: $F(1,10) = 3.36–10.92$, $p < .05$). To test the scalp distribution in
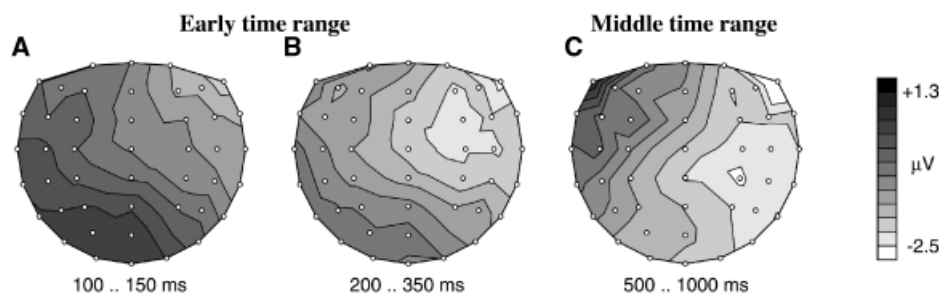

Figure 4. Scalp distribution of the differences between correct and violated musical phrases in the measure where the violation was presented. Data between a time range of (A) 100 to 150 ms, (B) 200 to 350 ms, and (C) 500 to 1,000 ms.
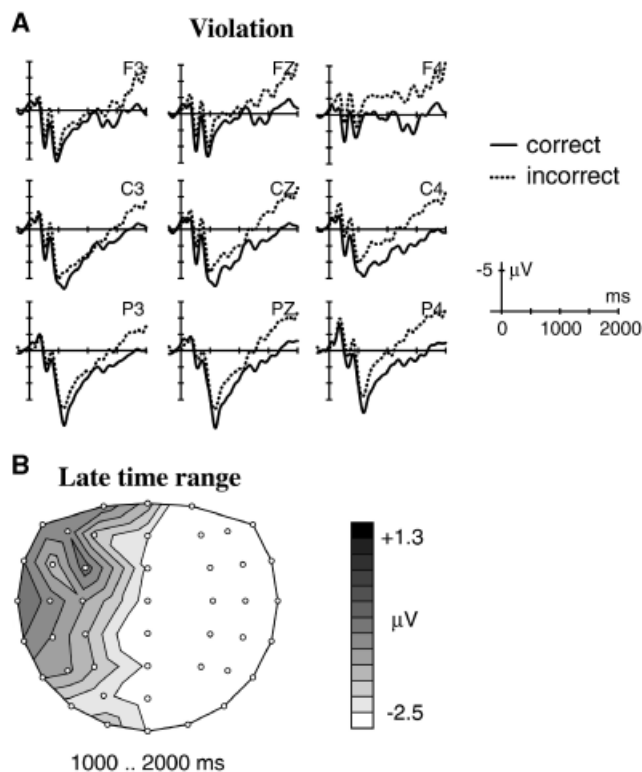
A Violation

B Late time range

1000 .. 2000 ms

Figure 5. A: Same data as Figure 1a, but plotted over a longer time frame. B: Scalp distribution of the long-lasting negative difference between correct and violated musical phrases between 1,000 and 2,000 ms.

the late time range, a ROI analysis was carried out between 1,000 and 2,000 ms. A main effect of violation, $F(1,10) = 8.18$, $p < .017$, and a Hemisphere $\times$ Violation interaction, $F(1,10) = 22.23$, $p < .0008$, was found, even after the McCarthy and Wood (1985) correction was applied, $F(1,10) = 11.9$, $p < .006$. No main effects or interactions with the factor anterior/posterior were found. This analysis, therefore, indicated that the effect was clearly lateralized to the right (see Figure 5B). Finally, as found in the analyses over the earlier time scale, no violation-related differences were found in the 1,000–2,000-ms range for measures presented after the violations.

In summary, the analyses showed that in-key violations elicited a large, long-lasting negativity compared to correct notes presented in the same measure. Interestingly, three effects can be distinguished on the basis of their scalp topography: an early right fronto-anterior negative effect (between 100 and 150 ms) followed by a right fronto-central negative effect (between 200 and 350 ms) and a late negative effect that lasted until the end of the recording period and showed a broadly distributed right-sided lateralization.

Discussion

In this experiment, musicians were presented with the score of familiar melodies. The score was presented measure by measure, and, in half of the trials, it contained a diatonic violation somewhere in the middle of the score. Results of the behavioral experiment suggested processing difficulty for the measure containing the violations, as self-paced reading slowed down by about 500 ms compared to correct measures. In the ERP

experiment, the score was presented in a RSVP format in which the measures were presented for 3,000 ms each. The ERPs to the violations showed an N1 component followed by a long-lasting negative shift that encompassed three different effects that will be discussed in turn.

N1

Generally speaking, the occipital N1 component is assumed to reflect modality-specific processing in the visual pathway that is sensitive to the location of visual attention in the visual field (Mangun & Hillyard, 1990). One could therefore suggest that the rightward distribution of the N1 reflects directed attentional resources to the left visual field. This would fit the left-to-right reading pattern needed in (music) reading. Note also, that such N1 distribution fits the fMRI-data of Nakada, Fujii, Suzuki, and Kwee (1998) and Schön, Anton, Roth, and Besson (2002) very well. In the Nakada et al. experiment, a comparison between reading Japanese, English, and music (in Japanese piano players) was made. It was found that only music reading shows an activation of the cortex flanking the right transverse occipital sulcus. Interestingly, using a quite different task, Schön et al. also found a very localized focus of activation in the right occipito-temporal junction when contrasting music reading with reading words or numbers in Arabic notation. These fMRI findings suggests that the N1 lateralization is probably not a "reading"-specific effect, but is instead sight reading specific.

Early Negativity

In the time range between 100 and 350 ms, an early negativity was found that had its maximum amplitude within the first 100 ms, at right frontal electrodes. Between 200 and 350 ms, it shifted more centrally leading to an anterior temporal scalp distribution (see Figure 4). Such topographical differences seem to indicate that this negativity encompasses two distinct effects, an early and a late effect.

Between 100 and 150 ms, a right fronto-anterior negativity was found that shows similarities with the so-called ERAN component found by Koelsch et al. (2000) in an experiment where chord sequences were presented auditorily. The ERAN is theorized to reflect the early detection of sound expectancy violations. It is very interesting to find such an early right anterior negativity also with visually presented materials, because it suggests that early violation detection mechanisms are independent of stimulus modality. This mirrors recent findings in language processing where visually and auditorily presented word category violations both elicit an early left anterior negativity (ELAN; cf. Gunter et al., 1999). The ELAN is thought to reflect an early (syntactic) violation detection mechanism on the basis of an abstract linguistic feature (i.e., the category of a word). Thus, it may be that the ERAN found in music and the ELAN found in language both reflect a functionally similar process, namely an early violation detection process on the basis of a salient feature in the domain of concern. In this respect, it is interesting to note that Maess et al. (2000) suggested that the ERAN reflects music "syntactic" processing in a harmonic context. Note, however, that the present experiment does not contain any syntactic violations. Therefore, one should remain careful with respect to the functional interpretation of the present early negativity, also because Koelsch et al. as well as Maess et al. used chords sequences whereas the present experiments used simple but expected melodic lines (i.e., single notes). Interestingly, visual inspection

of the data from Besson and Faïta (1995) also seems to show an early negativity for auditorially presented diatonic violations processed by musicians. Whether the scalp distribution of this component is similar to the present one is very difficult to determine on the basis of visual inspection of their Figure 4, because a very limited number of electrodes is shown.

Immediately after the ERAN, in the time range up to approximately 400 ms, the negativity shifted in distribution to show an anterior temporal predominance that is reminiscent of the RATN, described by Patel et al. (1998). This component has been suggested to correlate either with the application of musical rules or with working memory (WM) processes. Because a musical rule has not been violated here, but only the expectancy for a certain note in a familiar/known musical piece, it is more likely that the RATN-like response in the present experiment is associated with WM processes. In the present violation detection paradigm, it is necessary for the subject to compare the presented notes with the long-term memory representation of the musical piece to perform the task. Such retrieval and comparison processes probably take place in working memory. It is highly questionable that the comparison is made at the (visual) note level. Probably, one has to assume that the visual symbols are transferred into the auditory domain before interpretation and violation detection can take place. This may be comparable with grapheme to phoneme conversion or "inner speech," as described by Perfetti (1998). It is interesting to note that such a subvocal singing or humming strategy may be used during auditory imagination of familiar melodies and consequently activates the supplementary motor area (SMA), that would reflect the associated motor planning (Halpern & Zatorre, 1999). How such a strategy is actually used during music reading is not an issue that can be solved by this study. The representation of the long-term memory trace of a familiar melody, however, may consist of both auditory and motor information (note that during the perception of well-known piano pieces, piano players show motor activation of the contra-lateral hand area; cf. Haueisen & Knosche, 2001). Thus, in addition to retrieval and auditory comparison processes, visual-to-auditory transfer processes may also take place in WM. Which of these processes are reflected by the RATN remains unclear at the moment and will need further experimentation.

Middle Latency Negativity

Comparing the present data with that of the Besson and Faïta (1995) study, a very pronounced difference is seen in the time range between 400 and 1,000 ms. In the Besson and Faïta results, a clear late positive component can be seen for diatonic violations. Although the present results also show a clear P3b component around 600 ms, it is not larger, but smaller, for the violations than for the correct notes. Thus, instead of a positive-going difference, a negative-going difference is found (see Figure 4C). This middle latency negativity was widely distributed across the right hemisphere. Around the peak of the P3b, however, it was clearly centrally distributed (see Figure 6).

Taking the central and widely distributed scalp distribution of the violation effect into account, one might want to suggest that the "negativity" in the P3b time range is a modulation of this component. There are many studies reporting a decrease in P3b amplitude when task load is increased during target detection tasks (for a recent review, see Kok, 2001). For instance, in memory search tasks, it is found that more search leads to a smaller P3b. One could think of this effect as a P3b with
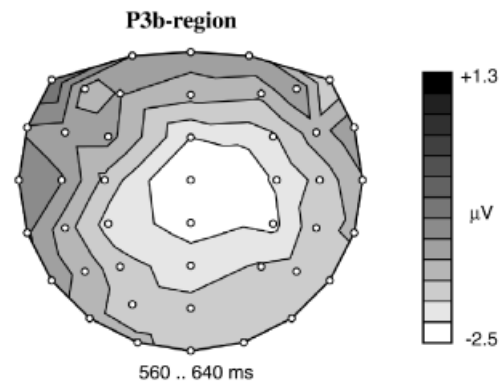

Figure 6. Scalp distribution of the difference between correct and violated musical phrases around the peak of the P3b (560–640 ms).

overlapping processing negativity (cf. Wijers, Otten, Feenstra, Mulder, & Mulder, 1989). Interestingly, Kok suggested that P3b reflects processing underlying recognition memory. Such a conception of P3b would fit well into the situation of sight reading a known musical phrase in which a violation needs to be detected.

A large difference between the results of the present experiment and that of Besson and Faïta (1995) is the position of the violation. In Besson and Faïta, the violations were always presented at the end of the musical phrase, whereas in the present experiment, they always were somewhere in the middle of the phrase. It could therefore be that the accumulation of the musical context and, consequently, the expectancy for a certain note is different between the two experiments. However, a recent experiment performed by Schmidt, Gunter, and Kotz (2002) shows that LPC-like components can be elicited by auditory in-key violations when played in the middle of a familiar musical phrase.[5] Thus, the position of the violation per se cannot account for the large differences between the present and the Besson and Faïta study. It may be that the effects reported here are specific reflections of mechanisms used during music reading. Results of a recent experiment point in the same direction. Schön and Besson (2002) explored the reading of notes using a priming-like paradigm. First, participants were presented with a bar including clef, key, and time signatures. After a SOA of 1,000 ms, a measure containing a note was presented. Participants were asked to judge whether the note was the tonic and/or had the correct duration. The violations elicited a long-lasting negativity. Thus, even very shallow musical contexts are able to elicit music-reading-related negativities. Moreover, the Schön and Besson results also indicate that a negativity is elicited when participants are involved in reading notes presented at the end of a trial.

Long Lasting Negativity

It is clear that the integration of notes into a musical context lasts beyond the point when the violation was presented. The violation was always presented at the beginning of a measure (to avoid excessive eye movements) in the middle of the score. Thus, after the detection of the violation, repair processes are required, as well as processes that integrate the rest of the measure into the

---

[5]Note that in this experiment, the positivity was followed by a long-lasting negativity, as in the present experiment, which had a length of approximately one measure after the occurrence of the violation.

musical context. Insofar as the detection of violations was associated with very early effects, the information contained in the whole score was probably not taken into account during this detection, because this would have taken more than the 100–150-ms time range within which the early negativity developed. It is, therefore, not unreasonable to propose that the long-lasting right-sided negativity present until the end of the recording period is related to such integration processes. This is also in line with suggestions that the right hemisphere is dominantly involved during music processing (see Zatorre, Belin, & Penhune, 2002, see also Nakada et al., 1998 and Maess et al., 2001).

General Conclusions

The present study was undertaken to explore whether sight reading of diatonic violations in a musical phrase elicits electrophysiological reflections similar to those elicited by hearing such violations. If this had been the case, a clear parallel to language processing could have been established, because reading and hearing violations in language elicit similar endogenous ERP components. Compared to data found in the literature on auditory music processing, the present data on music reading showed some similarities as well as important differences. The detection of the violation, as reflected in an early right anterior negativity, is comparable to what is found in auditory music processing. Thus, it seems that violation detection in music relies on modality-independent and abstract processing mechanisms. Following the early violation detection processes, WM-memory related (RATN) and recognition memory (P3b) processes may take place.

In a later time frame (after approximately 500 ms), the major differences between sight reading and hearing music become apparent in the ERPs. In contrast to the presently found violation-related negativities, the experiment of Besson and Faïta (1995) showed a larger LPC component for the same type violations presented in the auditory modality. Although there were differences between the two studies in the relative position where the violation was presented, the very different results obtained in the two studies suggest that reading music does not elicit the same pattern of endogeneous ERP components as does hearing the same music. Thus, in contrast to similarities found in reading and speech processing, reading and listening to music have less commonalities. Whether such differences between language and music are due to differences in symbolic representation or to differences in underlying cognitive processing needs to be further explored.

REFERENCES

American Electroencephalographic Society. (1991). Guidelines for standard electrode position nomenclature. *Journal of Clinical Neurophysiology*, *8*, 200–202.

Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparisons of musicians with nonmusicians. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 1278–1296.

Besson, M., Faïta, F., Peretz, I., Bonnel, A. M., & Requin, J. (1998). Singing in the brain: Independence of lyrics and tunes. *Psychological Science*, *9*, 494–498.

Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, *24*, 14–25.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, *1*, 183–192.

Friederici, A. D., Wang, Y., Herrmann, C. S., Maess, B., & Oertel, U. (2000). Localization of early syntactic processes during word processing in the human brain: An fMRI investigation of different word types. *Human Brain Mapping*, *11*, 1–11.

Geisser, S., & Greenhouse, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

Gunter, Th. C., & Friederici, A. D. (1999). Concerning the automaticity of syntactic processing. *Psychophysiology*, *36*, 126–137.

Gunter, Th. C., Friederici, A. D., & Hahne, A. (1999). Brain responses during sentence reading: Visual input affects central processes. *NeuroReport*, *10*, 3175–3178.

Gunter, Th. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology*, *34*, 660–676.

Haberlandt, K. (1994). Methods in reading research. In Ann Gernsbacher Morton (Ed.), *Handbook of psycholingiustics* (pp. 1–31). San Diego: Academic Press.

Hagoort, P., & Brown, C. M. (2000). ERPs effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia*, *38*, 1513–1549.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, *9*, 697–704.

Haueisen, J., & Knosche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience*, *13*, 786–792.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The P600 as an index of syntactic integration difficulty. *Language and Cognitive Processes*, *15*, 159–201.

Kieras, D. E., & Just, M. A. (1984). *New methods in reading comprehension*. Hillsdale, NJ: Lawrence Erlbaum.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbounded dependencies. *Journal of Cognitive Neuroscience*, *5*, 196–214.

Koelsch, S., Gunter, T., Friederici, A. D., & Schroeger, E. (2000). Brain indices of music processing: Nonmusicians are musical. *Journal of Cognitive Neuroscience*, *12*, 520–541.

Koelsch, S., Gunter, T. C., von Cramon, D. Y., Zysset, S., Lohmann, G., & Friederici, A. D. (2002). Bach speaks: A cortical 'language network' serves the processing of music. *NeuroImage*, *17*, 956–966.

Kok, A. (2001). On the utility of P3 amplitude as a measure of processing capacity. *Psychophysiology*, *38*, 557–577.

Kutas, M., & Federmeier, K. D. (2000). Electrophysiology reveals semantic memory use in language comprehension. *Trends in Cognitive Sciences*, *4*, 463–470.

Kutas, M., & Hillyard, S. A. (1980). Event-related brain potentials to semantically inappropriate and surprisingly large words. *Biological Psychology*, *11*, 99–116.

Lehmann, A. C., & Ericsson, K. A. (1996). Performance without preparation: Structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology*, *15*, 1–29.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT press.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: An MEG study. *Nature Neuroscience*, *4*, 540–545.

Mangun, G. R., & Hillyard, S. A. (1990). Allocation of visual attention to spatial location: Event-related brain potentials and detection performance. *Perception and Psychophysics*, *47*, 532–550.

McCarthy, G., & Wood, C. C. (1985). Scalp distribution of event-related potentials: An ambiguity associated with analysis of variance models. *Electroencephalography and Clinical Neurophysiology*, *69*, 218–233.

Nakada, T., Fujii, Y., Suzuki, K., & Kwee, I. L. (1998). 'Musical brain' revealed by high-field (3 Tesla) functional MRI. *NeuroReport*, *9*, 3853–3856.

Osterhout, L., & Hagoort, P. (1999). A superficial resemblance doesn't necessarily mean that you're part of a family: Counterarguments to Coulson, King & Kutas (1998) in the P600/SPS-P300 debate. *Language and Cognitive Processes*, *1*, 1–14.

Osterhout, L., Holcomb, P. J., & Swinney, D. A. (1994). Brain potentials elicited by garden-path sentences: Evidence of the application of verb information during parsing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 786–803.

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language*, *34*, 739–773.

Paller, K. A., McCarthy, G., & Wood, C. C. (1992). Event-related potentials elicited by deviant endings to melodies. *Psychophysiology*, *29*, 202–206.

Patel, A., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. (1998). Processing syntactic relations in language and music: An event related potential study. *Journal of Cognitive Neuroscience*, *10*, 717–733.

Peretz, I., & Herbert, S. (2000). Toward a biological account of music. *Brain and Language*, *42*, 131–134.

Perfetti, C. A. (1998). Comprehending written language: A blueprint of the reader. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 167–208). Oxford, UK: Oxford University Press.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*, 372–422.

Rayner, K., & Pollatsek, A. (1997). Eye movements, the eye-hand span, and the perceptual span during sight-reading of music. *Current Directions in Psychological Science*, *6*, 49–53.

Rogers, M. R. (1984). *Teaching approaches in music theory*. Carbondale, IL: Southern Illinois University Press.

Sadie, S. (1995). *The New Grove Dictionary of Music and Musicians*. Middleton, WI: Music Library Association.

Schmidt, B. H., Gunter, T. C., & Kotz, S. A. (2002). Expectancy in familiar melodies: An ERP investigation. In C. Stevens, D. Burnman, G. McPherson, E. Schubert, & J. Renwick (Eds.), *Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney 2002* (pp. 716–719). Adelaide, Australia: Causal Productions.

Schön, D., Anton, J. L., Roth, M., & Besson, M. (2002). An fMRI study of music sight-reading. *NeuroReport*, *13*, 2285–2289.

Schön, D., & Besson, M. (2002). Processing pitch and duration in music reading: A RT-ERP study. *Neuropsychologia*, *40*, 868–878.

Sloboda, J. (1985). *The musical mind: The cognitive psychology of music*. Oxford, UK: Clarendon Press.

Truitt, F. E., Clifton, C., Pollatsek, A., & Rayner, K. (1997). The perceptual span and eye-hand span in sight reading music. *Visual Cognition*, *4*, 143–161.

Vasey, M. W., & Thayer, J. F. (1987). The continuing problem of false positives in repeated measures ANOVA in psychophysiology: A multivariate solution. *Psychophysiology*, *4*, 479–486.

Verleger, R. (1990). P3-evoking wrong notes: Unexpected, awaited, or arousing? *International Journal of Neuroscience*, *55*, 171–179.

Verleger, R. (1991). Slow potentials in a melody recognition task. *Archives of Psychology*, *142*, 225–244.

Waters, A. J., Underwood, G., & Findlay, J. M. (1997). Studying expertise in music reading: Use of a pattern-matching paradigm. *Perception and Psychophysics*, *59*, 477–488.

Wijers, A. A., Otten, L. J., Feenstra, S., Mulder, G., & Mulder, L. J. M. (1989). Brain potentials during selective attention, memory search and mental rotation. *Psychophysiology*, *26*, 452–467.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*, 37–46.