Visually Induced Auditory Expectancy in Music Reading: A Behavioral and Electrophysiological Study

Daniele Schön and Mireille Besson

Abstract

■ The general aim of this experiment was to investigate the processes involved in reading musical notation and to study the relationship between written music and its auditory representation. It was of main interest to determine whether musicians are able to develop expectancies for specific tonal or atonal auditory events based on visual score alone. Can musicians expect an ''atonal'' event or will it always sound odd? Moreover, it was of interest to determine whether the modulations in amplitude of a late positive component (P600) described in previous studies are linked to a general mismatch detection process or to specific musical expectancies. Results showed clearly that musicians are able to expect tonal auditory endings based on visual information and are also able to do so for atonal endings, although to a smaller extent. Strong interactions seem to exist between visual and auditory musical codes and visual information seems to influence auditory processing as early as 100 msec. These results are directly relevant for the question of whether music reading is actually music perception. ■

INTRODUCTION

In the study of reading music, as for reading words, an important question is to understand what types of representations are used. In reading music, the representations may be phonologic (''la''), graphemic (♪), motoric (second finger), and acoustic (sound). Indeed, musicians most often read a score and play it at the same time. For instance, a string player sight-reading a score needs to identify the signs on the score to produce the appropriate motor action, to check whether the played notes are well tuned, and, sometimes, even add a vibrato depending upon the musical importance of each note. This latter point of musical interpretation is of great importance, not only in musical praxis, but also in psychological theory. To interpret the piece musically, the musician needs to assign a ''musical weight'' to each note, thus computing a set of complex operations. These operations comprise the understanding and updating of the metric, rhythmic, melodic, and harmonic structures. Therefore, as noted by Sloboda (1976, 1978, 1984), music reading is not just a visuomotor coding task but rather belongs to the realm of music perception. It has been shown that, in a music-reading task, musicians can better memorize ''good'' (well structured) than ''bad'' musical sequences (Halpern & Bower, 1982). The fact that memory is sensitive to musical structure implies that some musical representations are at work while reading. However, this does not provide precise information about the characteristics of these representations. It may be the case that the representations built from music reading are based on complex relevant visual patterns, on the rules governing music perception in auditory modality, or on an explicit knowledge of musical structure (e.g., the knowledge that in B♭ major, there are 2 flats). Note that these possibilities are not mutually exclusive and it is likely that these different types of representations coexist and interact (Schön & Besson, 2002a; Schön, Semenza, & Denes, 2001). An interesting possibility is that musicians have, to a certain extent, an auditory-like representation of the written music, before they actually play it. Although the visuomotor coding can be sufficient to play the written notes, an anticipated auditory representation of the written music is important for an expressive performance that assigns a different ''weight'' to each single note, depending upon its melodic/harmonic and metric/rhythmic importance. Note also that the sole visuomotor strategy cannot be used with all instruments. For example, string instrument players need to check that the note played is indeed the written note, and this verification process most likely relies on a match/mismatch comparison with their anticipated auditory representation. Indeed, no keys or frets are present on the violin ''fingerboard.'' To test the hypothesis that musicians are able to use an auditory-like representation of written music, we investigated the relationship between reading and listening to music when both are coupled within the same experimental task. We recorded both behavioral (reaction times [RTs] and error rates) and electrophysiological

INCM-CNRS, Marseille, France

data (event-related brain potentials [ERPs]) to examine performance and the processes involved in music reading and listening as they unfold in time. The general hypothesis is that if written music induces musical auditory expectancy (i.e., not just motor expectancy), this may influence the way music is perceived. From studies in other domains, it is well known that a sensory system may, in some cases, degrade, enhance, or modify the perception of another system (Sekuler, Sekuler, & Lau, 1997; McGurk & McDonald, 1976).

Musical notes can be categorized according to the degree to which they are expected. In a given context, some notes are highly stable (Krumhansl, 1990) and plausible, whereas others are unstable and implausible, thus creating musical tension or surprise. Let us take an example. The familiar rhythm and melody snippet ''Shave and a Haircut, Two Bits,'' whose ending is so irresistible to Roger Rabbit, is represented in Figure 1A. If we play the same tune again, raising the last note by half a tone (Figure 1B), the ending will be perceived as highly implausible and surprising. The issue addressed in this article is to determine whether knowing in advance by means of a written score (Figure 1B) that such an odd note is going to be played does influence the expectancy for this note. In other words, is the expectancy of an auditory event in a given musical context influenced by visual information? If yes, at which stage of processing?

Musicians were asked to judge whether the last note of a 5-note auditory musical sequence matched or mismatched the information provided on a score presented on a computer screen (Figure 2). There were two types of visual stimuli (all comprising 5 notes): In one type, the last written note was stable, whereas in the other type, the last note was unstable with respect to the previous tonal context.[1] The sequence of auditory events always matched the sequence of written notes, except for the final notes that could either match or mismatch with the last visual stable or unstable notes. In the visual–auditory mismatching conditions, two types of mismatches were introduced: Although the played note was always different from the one written on the score (i.e., mismatch), it could be either plausible (2 and 5) or implausible (3 and 6) with respect to the previous tonal context. To summarize, the experimental design comprised 6 experimental conditions with two types of visual sequences (ending with stable or unstable notes relative to the previous context), each being associated to three types of auditory sequences (ending with notes that match the visual sequence, that mismatch but are still musically plausible, and that mismatch and are musically implausible).

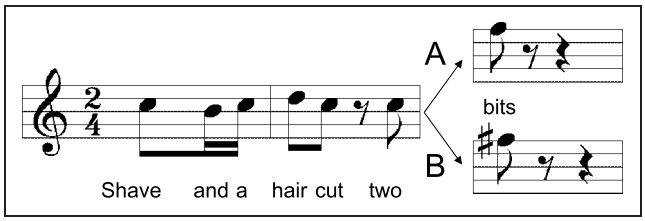

Figure 1. Example of expected/stable (A) and unexpected/unstable endings (B). A is the tonic, the most expected note in this musical context, and B is the minor second, completely out of context.

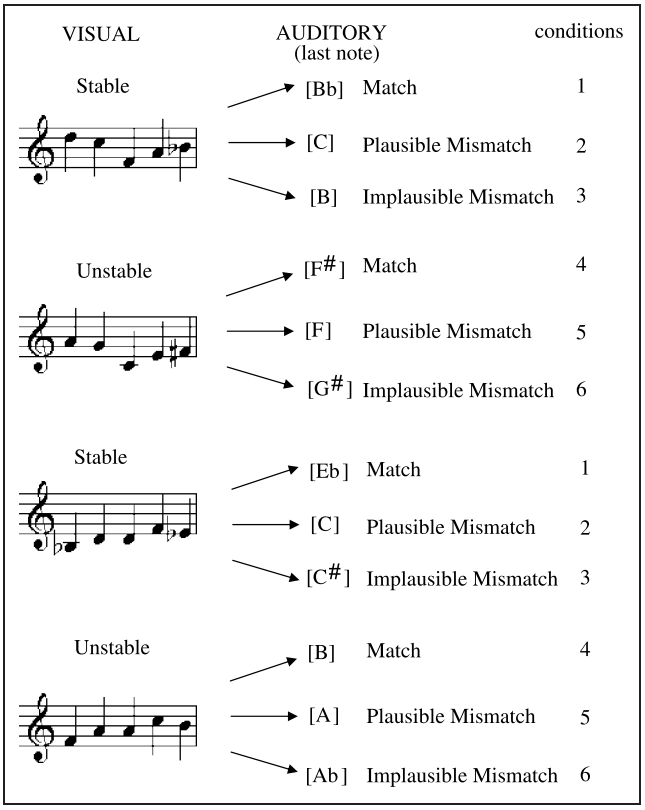

Figure 2. Illustration of the experimental conditions.

The first aim was to determine whether musicians would be able to anticipate unstable auditory events based on visual information? If so, the two matching conditions (1 and 4) should be processed similarly, that is, independently of whether the final auditory event matching the score is stable (henceforth, stable match, Condition 1) or unstable (henceforth, unstable match, Condition 4). By contrast, if musicians are unable to anticipate unstable events, comparison of the stable and unstable matching conditions should show differences in behavioral data (longer RTs and more errors) and in the electrophysiological components (larger amplitude and longer latency, see below) described in the literature as correlated with the violation of musical expectancy (e.g., Schön & Besson, 2002b; Koelsch, Gunter, Friederici, & Schroger, 2000; Besson & Macar, 1987). The second aim was to further track the functional significance of the late positive component (LPC, or P600), peaking between 300 and 600 msec, that, in a

melody, are associated to incongruous/unexpected endings (e.g., Besson & Macar, 1987). The amplitude and the latency of these late positivities have been shown to be sensitive to musical expertise, to the familiarity of the musical piece, and to the degree of incongruity (Besson & Faïta, 1995). However, in previous studies (e.g., Regnault, Bigand, & Besson, 2001; Patel, Gibson, Ratner, Besson, & Holcomb, 1998; Besson & Faïta, 1995; Besson & Macar, 1987), the congruous stimulus was always a match, whereas the incongruous stimulus was always a mismatch, because the notes/chords were not appropriate in the musical context. Moreover, insofar as the task, in most cases, was to decide whether the note/chord was expected or not within the previous context, the mismatching aspect of the incongruous notes/chords was clearly task-relevant. Therefore, the extent to which the occurrence of the LPCs reflects a general mismatch process or a more specific musical (tonal) mismatch process remains an open question. In the present experiment, the use of two types of auditory mismatches (plausible and implausible) allowed us to address this question. If the amplitude of the LPC is mainly sensitive to a general mismatching process, auditory mismatching events should be processed similarly, regardless of whether they are plausible or implausible within the previous tonal context. By contrast, if this component is sensitive to more specific musical expectancy processes, governed by the rules of musical ''syntax,'' one should find differences between the two types of mismatching conditions, with the implausible mismatches showing larger LPCs than the plausible mismatches.

The third aim was to study the relationship between the expectancy created by the visual score (stable vs. unstable endings) and the tonal dimension of the auditory stimuli in the 2 mismatching conditions (plausible vs. implausible). Indeed, musicians may perceive the same event (e.g., plausible or implausible mismatch) differently when they expect a stable ending from when they expect an unstable ending, based on visual information.

RESULTS

Behavioral Results

Error Rate

A 2-way repeated measures ANOVA was carried out, including visual expectancy (stable/unstable) and auditory match/mismatch dimensions as factors (Figure 3). The main effect of visual expectancy was significant, $F(1,13) = 14.2$, $p = .002$, with more errors in the unstable-ending conditions (9.3% vs. 5.6%). Moreover, the visual expectancy by auditory match/mismatch interaction was also significant, $F(2,26) = 9.1$, $p = .001$. Post hoc comparisons showed that only the plausible mismatch in the visual unstable-ending condition was significantly different from all the other conditions (always $p < .008$) that did not differ from one another (always $p > .47$).

Reaction Times

Again, a 2-way repeated measures ANOVA was carried out, including the same factors as above (Figure 3). The main effect of visual expectancy was significant, $F(1,13) = 4.7$, $p = .049$, with slower RTs in the unstable-ending conditions (682 vs. 639 msec). The auditory match/mismatch main effect was also significant, $F(2,26) = 12.8$, $p = .0001$, with slower RTs to mismatching than matching auditory events. Moreover, the visual expectancy by auditory match/mismatch interaction was also significant, $F(2,26) = 3.7$, $p = .04$. Two further analyses were conducted: t tests revealed that RTs in the matching conditions were faster with stable than unstable visual endings (573 vs. 632 msec, $p = .04$). By contrast, results of a 2-way repeated measures ANOVA, including visual expectancy (stable/unstable) and the 2 mismatching conditions only (plausible/implausible mismatch) as factors, showed no main effect of visual expectancy ($p > .05$). However, plausible mismatches were associated with longer RTs than implausible mismatches (709 vs. 679 msec), $F(1,13) = 17.1$, $p = .001$, especially when the visual endings were unstable [visual expectancy by mismatch type interaction: $F(1,13) = 4.6$, $p = .05$].

Discussion

Overall, accuracy and RTs showed a consistent pattern of results (i.e., no speed–accuracy trade-off). First, the matching conditions were performed faster and with fewer errors overall than the mismatching conditions. Second, although the 2 matching conditions did not differ in accuracy as a function of the stability of the musical sequence, RTs were shorter to stable than to unstable endings. Finally, the differences in accuracy and RTs between plausible and implausible mismatches were only significant when the visual endings were unstable.

The finding that RTs were shorter and accuracy higher for matching than mismatching auditory events is in line with many results in the literature. In the present experiment, a target note with a specific pitch was expected based on visual information. Thus, although participants could anticipate matching targets, and therefore be ready to respond, they could not anticipate mismatching targets, thus explaining why RTs were slower in this latter case. Most importantly, this was true for both visual endings, showing that, upon visual information, musicians were able not only to anticipate stable, but also unstable endings. Such a tentative conclusion needs, however, to be further considered in light of the results in the matching conditions (1 vs. 4). Indeed, RTs were shorter for stable than unstable matching endings. Two possible interpretations may account for

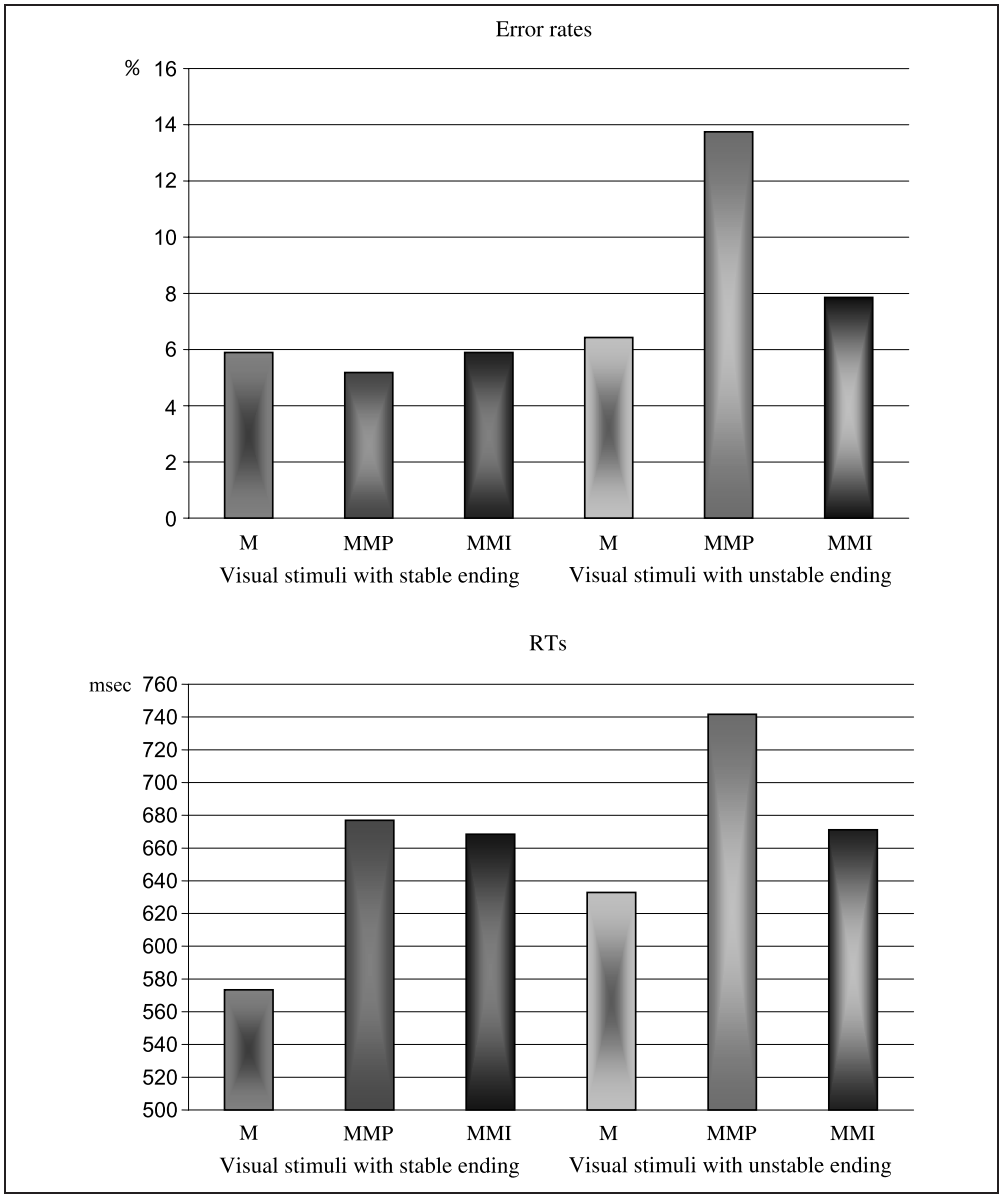

Figure 3. Error rates (in percentage) and RTs (in milliseconds). M = match; MMP = plausible mismatch; MMI = implausible mismatch.

this result. It may be that it is more difficult to anticipate unstable than stable matching endings. Alternatively, musicians may be able to anticipate visual unstable endings to the same degree as stable endings, but the pitch of the unstable endings may remain somewhat surprising (it is a ''wrong'' note after all), thus causing a delay in the response. This latter explanation builds on the hypothesis that a wrong note remains incongruous even when one knows it in advance (see also general discussion). Finally, the finding that plausible mismatches were associated with more errors and slower RTs than implausible mismatches, when the visual endings were unstable, may reflect some inhibition process. Indeed, within unstable-ending sequences, a plausible mismatch provides a better musical closure than the matching ending. Therefore, musicians may need to inhibit a positive response (i.e., match) when presented with a plausible mismatch, thus increasing RTs and error rates.

Event-related Brain Potential Results

As can be seen in Figure 4, the visuoauditory music task used in this experiment elicited several ERP components in the different experimental conditions. A N1 component is clearly evident at all electrode sites, peaking around 110 msec. A rapid positive (190 msec)–negative (220 msec) complex (P2–N2) is then elicited, with larger negative components in the mismatching conditions. Although a positivity then develops at posterior sites (P3), a negative-going component is present at frontal sites (N5). To analyze in detail how these components were modulated by the independent variables manipulated in this experiment, we first computed general

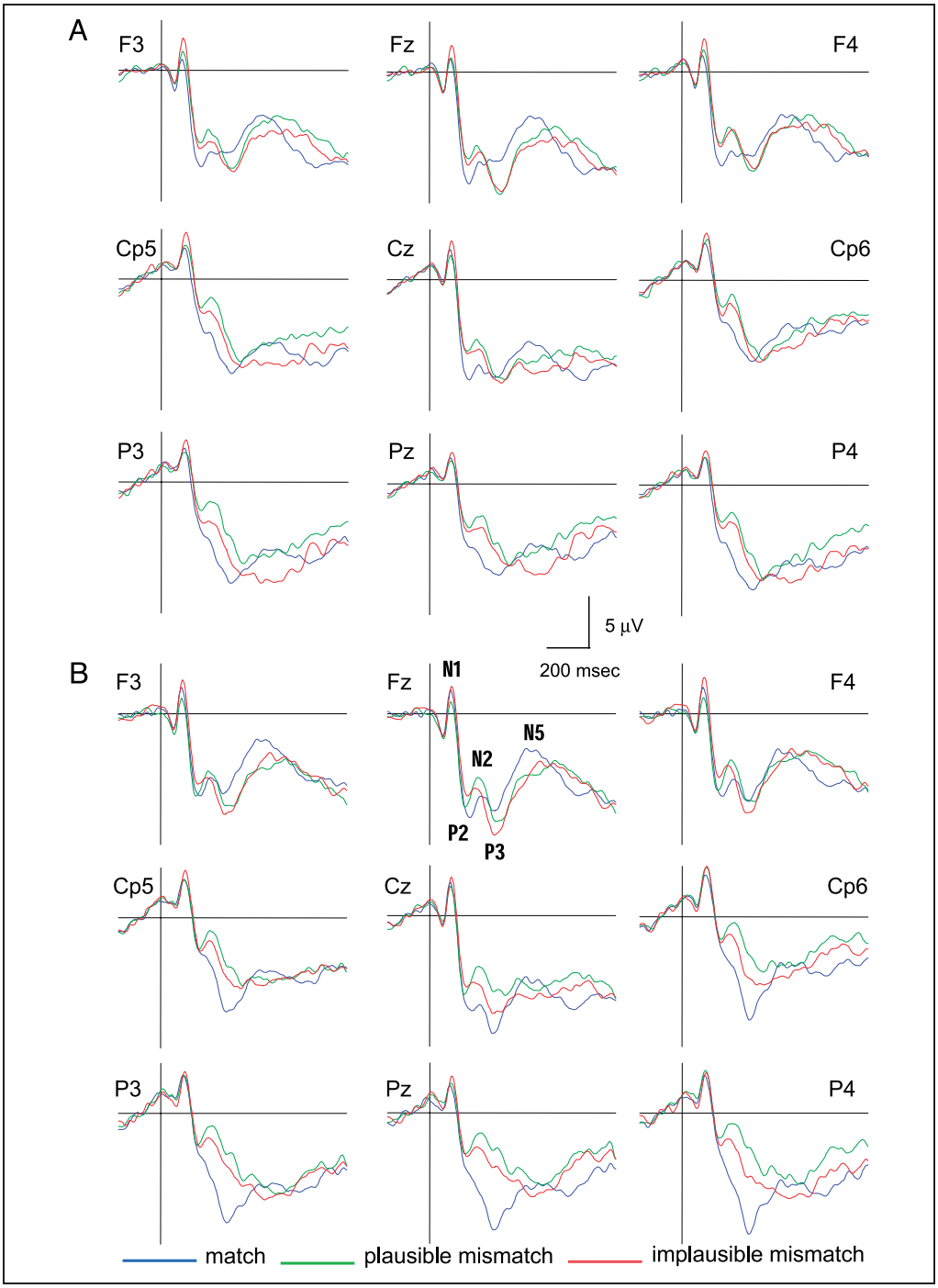

Figure 4. Illustration of the variations in brain electrical activity time-locked to the final note when expecting a stable ending (top, A) or an unstable ending (bottom, B) for matching (blue trace) or plausible (green trace) and implausible mismatching (red trace) conditions. Each trace represents an average of electrophysiological data recorded from 14 musicians. Although EEG was recorded from 28 electrodes, only the most representative clusters of electrodes (12 electrodes) were analyzed using ANOVAs and selected traces from 9 electrodes are presented. On this figure, as on the following ones, the amplitude (in microvolts) is plotted in ordinate (negative up) and the time (in milliseconds) is in abscissa.

analyses including all conditions and then separate analyses for the matching and mismatching conditions. In line with the general aims of this experiment, and in order not to complicate further the presentation of the results, we describe in detail only the results of separate analyses for the matching and mismatching conditions. For the global analyses, it is sufficient to note that they showed a main effect of visual expectancy in the 350- to 800-msec latency range, $F(1,13) = 6.9$, $p = .02$, a main effect of auditory match/mismatching in the 100- to 300-msec latency range, $F(1,13) = 9.0$, $p = .002$, then interacting with the anterior/posterior dimension in the 300- to 600-msec window, $F(2,26) = 30.5$, $p < .0001$, and finally, a visual by auditory match/mismatch by anterior/posterior interaction in the 250- to 400-msec latency range, $F(2,26) = 9.6$, $p = .007$.

Matching Events

An ANOVA was computed with visual expectancy (2), hemisphere (2), anterior/posterior (2), and electrodes (3) as factors. A visual expectancy by anterior/posterior

interaction was found significant in both the 200- to 350-msec and 500- to 700-msec latency ranges [$F(1,13) = 8.0$, $p = .01$, and $F(1,13) = 5.7$, $p = .03$, respectively]. Precise analysis of the time course of these effects (50-msec windows) revealed that the amplitude of early (N2: 200–300 msec latency range) and late (N5: 500- to 700-msec latency range) anteriorly distributed negative components was larger for unstable than stable endings (Figure 5). Moreover, a posteriorly distributed positive component (P3: 250- to 350-msec latency range) was also larger in unstable than stable matching endings (Figure 5).

Mismatching Events

An ANOVA was computed with visual expectancy (2), mismatch type (2), hemisphere (2), anterior/posterior (2), and electrodes (3) as factors. The main effect of visual expectancy was significant between 300 and 450 msec, $F(1,13) = 10.7$, $p = .006$, and interacted with the anterior/posterior dimension, $F(1,13) = 19.6$, $p = .0007$: The amplitude of posteriorly distributed positivities was larger for visual stable than unstable endings. The main effect of mismatch type was also significant in the 100- to 150-msec (N1), 200- to 350-msec, and 400- to 700-msec latency bands [$F(1,13) = 9.4$, $p = .009$; $F(1,13) = 10.2$, $p = .007$; and $F(1,13) = 14.2$, $p = .002$, respectively]: The N1, the P3a, and a later positive component were larger in implausible than in plausible mismatches (see Figure 6).

Interestingly, the visual expectancy by auditory mismatch interaction was significant between 450 and 550 msec, $F(1,13) = 4.6$, $p = .05$: The positivity elicited by the implausible mismatches was larger when the visual stimuli were musically stable endings (see Figure 7).

Discussion

Overall, results showed that the mismatching conditions differed from the matching conditions both when the visual stimuli had stable and unstable endings. Although this speaks in favor of the capacity of musicians to anticipate or prepare themselves for unstable endings, detailed analyses of our results show that such a strong conclusion cannot be arrived at without precautions. In the following discussion, we will first consider the results in the auditory matching conditions and then those in the auditory mismatching conditions.

Matching. When the auditory stimuli match the note expected based on the visual score (matching conditions), the amplitude of early and late negative components (N2 and N5), as well as the amplitude of a positive component (P3), is larger in the unstable than in the stable-ending condition. Thus, the answer to the first question asked in the Introduction (i.e., ''Are musicians able to anticipate an unstable matching auditory event based on the visual score?'') is rather negative because unstable matching events seem to be processed differently from stable matching events.

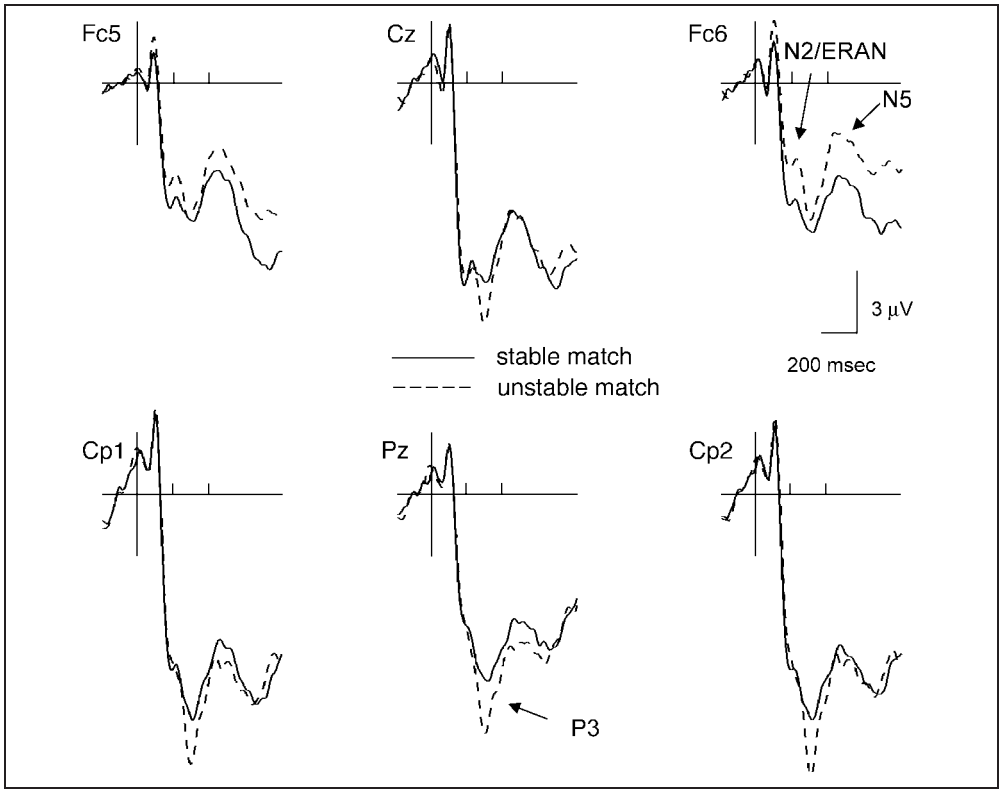

Figure 5. Illustration of the variations in brain electrical activity time-locked to the final note when it matches either the visual stable endings (solid line; Condition 1) or the visual unstable endings (dashed line; Condition 4). A larger late positivity (P3) is visible when the subjects expect an unstable ending.

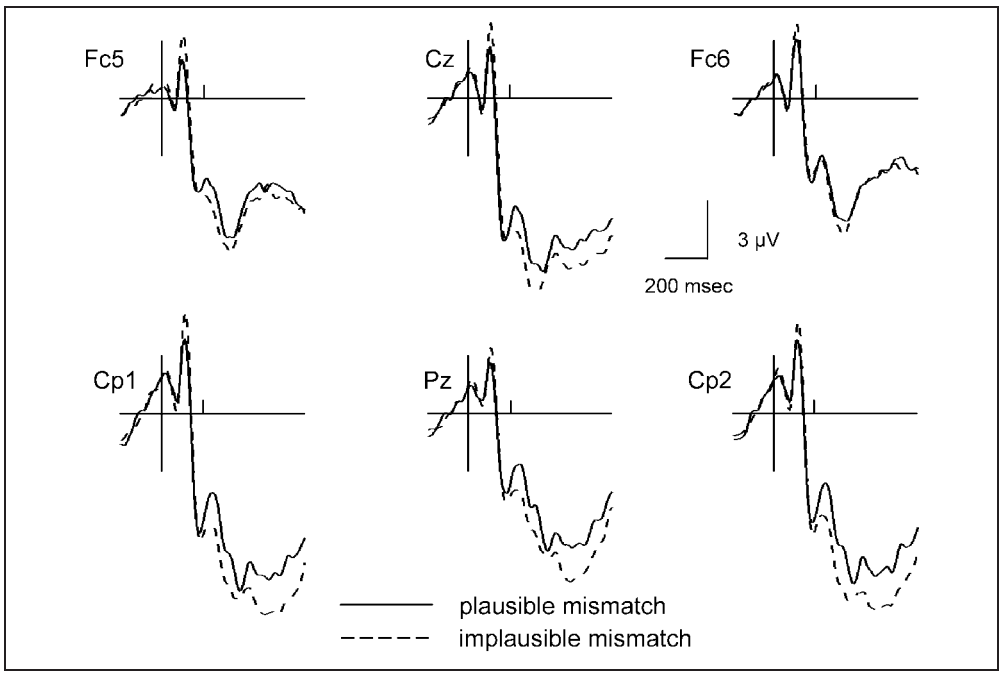

Figure 6. Zoom of the variations in brain electrical activity time-locked to the final note for both plausible mismatches (solid line, average of Conditions 2 and 5) and both implausible mismatches (dashed line, average of Conditions 3 and 6). The N1 and P3 are larger in implausible endings than in plausible endings.

In other words, the expectancy that musicians develop based on the visual score is not as strong for unstable- than for stable-ending events. As mentioned previously, musicians can anticipate a specific expected (i.e., stable) event, but cannot fully anticipate unexpected (i.e., unstable) events, may be because of the partly automatic manner in which the music structure is parsed (Besson & Schön, 2001; Jackendoff, 1991). The larger positive component (P3) to unstable matching endings seems to indicate that, independently of the fact that musicians are ''expecting'' an unstable (i.e., odd) ending based on the visual stimulus, the auditory ending remains somewhat unexpected.

Although our hypothesis was mainly based upon potential modulations in the amplitude of LPCs, results showed that the amplitude of an early negative component (150- to 250-msec latency band) was also sensitive to stable versus unstable tonal endings. Interestingly, the analysis of the scalp distribution of this early negativity revealed a right frontal lateralization (Figure 8) that is reminiscent of the right anterior temporal negativity (RATN) described by Patel et al. (1998) and of the early right anterior negativity (ERA) described by Koelsch and colleagues (Koelsch, Schroger, & Gunter, 2002; Koelsch, Gunter, Schroger, et al., 2001; Koelsch et al., 2000). More precisely, the negative component described here

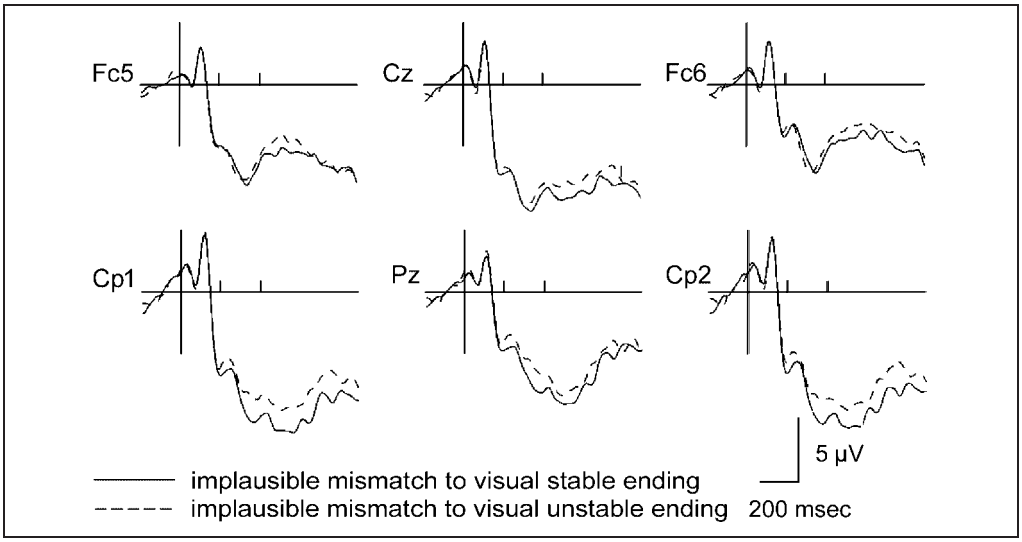

Figure 7. Illustration of the variations in brain electrical activity time-locked to the implausible mismatches in visual stable endings (solid line, Condition 3) and in visual unstable endings (dashed line, Condition 6). A larger late positivity is visible when the score contains a stable ending.

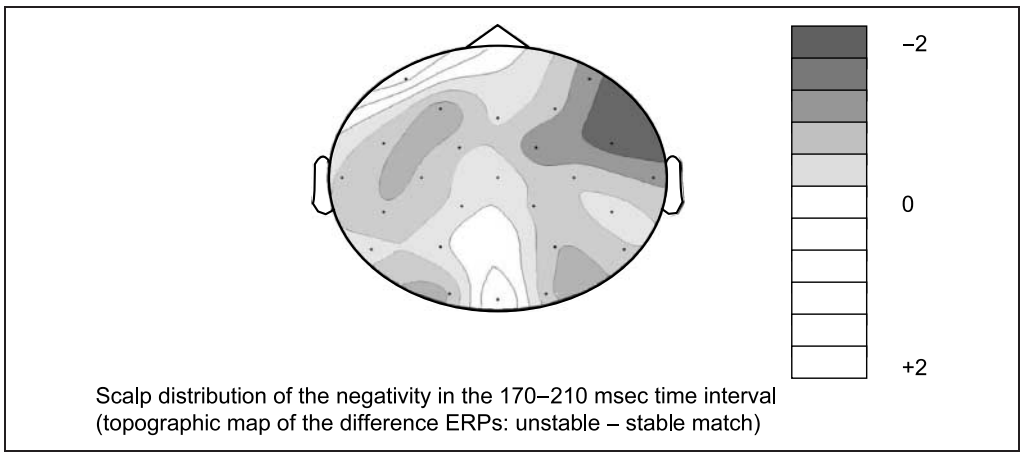

Figure 8. Topographic maps computed as an integration of mean amplitude values across time in the 170- to 210-msec time windows, from the difference waves (unstable–stable match).

resembles the ERAN more closely than the RATN, because the latter peaks around 350 msec. The ERAN has been shown to be elicited by deviant chords relative to a harmonic context, maximal around 150–250 msec. It has been taken to reflect the violation of the listener musical expectancy and, more generally, music-syntactic processing (Maess, Koelsch, Gunter, & Friederici, 2001). So far, this component has been described in studies using an auditory presentation of chords, but not when melodies were presented. It may be that a harmonic context creates stronger expectancies than a purely melodic context, because the rules governing harmony are less flexible than those governing melody. In our study, however, an ERAN may have developed because melodic expectancies were possibly enhanced by the presentation of the visual score. Note also that in previous studies using simple melodies, such early differences, which are usually small in amplitude, were possibly masked by the occurrence of larger LPCs. As the authors were generally mainly interested in the LPCs, such early differences may have gone unnoticed.

Finally, the amplitude of a late negative component (N5) was also larger in the unstable matching condition than in the stable matching conditions. This negativity, peaking around 500 msec, resembles the N5 described by Koelsch et al. (2000, 2002). The authors proposed that this N5, which was larger in Neapolitan chords compared with tonic chords, reflects musical integration processes. Indeed, Neapolitan chords are more difficult to integrate into the preceding harmonic context than are tonic chords. Thus, according to this interpretation, unstable matching endings in the present experiment could be analogous to the Neapolitan chords, in that they are more difficult to integrate. As hypothesized for the ERAN, the fact that no such a late negative component was described in previous studies using melodies rather than chord sequences may be because of an enhancement of melodic expectancy by the visual score (see also Gunter, Schmidt, & Besson, 2003, for a discussion of the ERAN in music reading). Further research will be needed to directly compare the late negativities elicited by chords and melodies.

Mismatch. When the auditory stimulus mismatches the note expected based on the visual score (mismatching conditions), several interesting results emerge. First, independently of whether musicians are anticipating stable- or unstable-ending notes, implausible auditory mismatches are associated with larger N1, P3, and LPCs than plausible auditory mismatches (Figure 6). Thus, these results are in line with previous results in the ERPs and music literature (Hashimoto, Hirata, & Kuriki, 2000; Besson & Macar, 1987; Schön, Magne, & Besson, 2004) when purely auditory musical stimuli are used. However, in these studies, expected (matching) events were always compared with unexpected (mismatching) events. Therefore, it was not possible to determine whether the effects reported on the N1, P3, and late positivities were linked to a general mismatch process or to the violation of specific musical expectancies. By contrast, in the present study, the comparison is between 2 mismatching events: one that is implausible within the melodic context and the other that is highly plausible. Therefore, the fact that we found differential effects, with larger differences for implausible than plausible auditory mismatches, clearly show that these different ERP components are sensitive to specific musical expectancies. Note, furthermore, that in previous studies, the number of trials included in the averages for congruous and incongruous trials were often different, thus possibly influencing the amplitude of the early components such as the N1.

It should be mentioned that the N1 results differ from the results of a previous experiment (Regnault et al., 2001), in which the N1 component was found to be larger in consonant than in dissonant chords. Insofar as notes were used in the present experiment, although chords were presented in the study of Regnault et al. (2001), it may be that these differences reflect the fact that harmonic processing differs to a certain extent from

melodic processing. Further research is needed to directly compare single notes and chords and to understand the effect of the hierarchic organization of tonal music on the different musical processing stages. Second, and very interestingly, the amplitude of the late positivity associated to implausible auditory mismatches was larger when musicians expected stable endings than when they expected unstable endings (Figure 7). This result supports the view that visual stable endings induce a strong musical expectancy for a plausible event, thereby increasing the auditory mismatch effect of the implausible ending. By contrast, unstable endings create expectancy for an implausible event, thereby decreasing the auditory mismatch effect of the implausible ending.

Third, an early negative component, peaking around 200 msec, was elicited by both plausible and implausible mismatches, although it was somewhat larger for plausible mismatches (especially in the unstable-ending condition, see Figure 4). Because this negative component is followed by a clear positivity at frontal sites, we tend to interpret this succession of negative positive components as an N2–P3 complex, which would develop in response to surprising unexpected events. However, the finding that the amplitude of this N2-like component is overall larger in plausible than implausible mismatches seems to indicate that surprise is not the only factor influencing the results. As proposed in the discussion of the RT results, it might be that some inhibition processes are at play when plausible mismatches are presented at the end of unstable sequences because they match the preceding tonal context. Following this line of reasoning, the increased N2 amplitude in plausible mismatches (most evident in the unstable conditions) would reflect some inhibition linked with the suppression of a ''match response.'' This interpretation is in line with the increased N2 amplitude to no-go responses reported in the literature (Simson, Vaughan, & Ritter, 1977). Finally, it is interesting to note that the distribution of this negativity presents some differences with respect to the one found for the ERAN. In particular, although the ERAN cannot be seen at posterior sites and has a quite localized right frontal distribution, the negativity to mismatching endings can clearly be seen at fronto-central and parietal electrodes. It might be that the greater saliency of the mismatch and inhibition effects hides the smaller effect sensitive to musical stability (syntax).

GENERAL DISCUSSION

The general aim of this experiment was to investigate the processes involved in reading musical notation and, more precisely, to study the relationship between written music and its auditory representation. It was of main interest to determine whether musicians can develop expectancies for specific stable or unstable auditory events based on the visual score alone. Taken together, results showed that this seems to be the case, at least to some extent. Indeed, musicians are clearly able to anticipate stable endings based on visual information and are also able to do so for unstable endings, although to a smaller extent. Moreover, error rates, RTs, and electrophysiological data showed that the auditory mismatch effects found in the present experiment were clearly modulated by the musical expectations built from the score. Thus, strong interactions seem to exist between visual and auditory musical codes, in that the representations built from visual stimuli influence the way auditory musical sequences are perceived. Indeed, although N1 amplitude was significantly different between plausible and implausible mismatches (i.e., between notes that differed in terms of musical expectancy but that could not be, being mismatches, anticipated based on visual information, see Figure 6), such an N1 difference was not significant between stable and unstable matches (i.e., between notes that differed in terms of musical expectancy in the same way as plausible and implausible mismatches, but that, by contrast, could be anticipated based on visual information, see Figure 5). Therefore, we interpret this lack of difference on the N1 as an influence of visual processing on auditory processing, because, what distinguishes the matches from the mismatches is that the matches are expected based on visual information. In other words, being able to anticipate a precise target note based on visual information seems to influence auditory processing as early as 100 msec.

In support of the hypothesis of top–down influences on the N1, some studies have shown that the auditory cortex (where the N1 would be generated, Pantev et al., 1995; Liegeois-Chauvel, Musolino, Badier, Marquis, & Chauvel, 1994) is sometimes involved in imagery tasks. For example, the auditory association areas (left) are active during a simple auditory imagery task in which musicians had to imagine the sound of a single note presented visually (Schürmann et al., 2002). In addition, imagery for familiar tunes showed activation in the right auditory association cortex of subjects imaging the continuation of a tune cued by its first few notes (Halpern & Zatorre, 1999). In the present experiments, insofar as subjects had to imagine in advance the sound of the last note of a visual sequence, this may have reduced the effect of syntactic incongruity. However, note that differences on later components (ERAN and P3, N5) are still visible (see Figure 5). It may be the case that although the N1 difference is a rather small difference, and thus when attenuated it is no longer significant, the differences on the ERAN and on the P300 remain significant. In other words, although there is an early effect of visual information on auditory processing, the effect of syntactic incongruity is only reduced and does not disappear. This is in line with the hypothesis proposed by Jackendoff (1991) that a processor uses the rules of music to analyze musical surface structure and

operates in a rather autonomous manner, with a limited access to musical memory: ''No matter how well one knows a piece, expectation, suspense, satisfaction, and surprise still occurs *within the parser*'' (p. 228). Such activity of deriving in real time the detailed and complex structure of music, with its points of tension and instability, can partly explain why a piece of music can create strong emotions despite repeated hearings. Further experiments using fMRI will help in defining to what extent and in which areas the visual presentation of notes can modulate their auditory perception. From the present findings, this modulation should be present in auditory areas and also in the (right) prefrontal regions considered to be important for musical syntax processing (Maess et al., 2001) and pitch retention.

Another interesting result of the present study concerns the functional interpretation of the P300 (P600). Several studies, in the literature on music processing and ERPs, have demonstrated that, in the auditory modality, the amplitude of LPCs, also called the P300 or the P600, is larger in unexpected than expected endings. Thus, the aim was to determine whether the modulations in amplitude of this LPC are because of the call into play of a general mismatch detection process or are also linked with specific musical expectancies (stability/plausibility). In the present experiment, insofar as we used two types of mismatching conditions, plausible and implausible, we hypothesized that the differences between the two would not be due to a mismatching process, but rather to musical expectancy. Results clearly showed that the amplitude and latency of the LPC were sensitive to specific musical expectations. However, it is interesting to note that, overall, we found smaller differences between conditions than expected based on previous findings (e.g., Besson & Faïta, 1995). Therefore, it may be that in previous studies on music processing, the amplitude (and latency) of the positivity were modulated by several factors such as a general mismatch detection process, task-relevant expectations (probability of occurrence), and expectations based on the musical structure. It is therefore important, in future studies on musical expectancy, to consider that these different factors may be at work simultaneously.

METHODS

Participants

Eighteen volunteer musicians were tested in the experiment. However, because of artifacts in the ERP data of 4 participants, only 14 (years of training: $M = 17$; range = 11–30) were retained for analyses. All were right-handed, neurologically normal, had normal or corrected-to-normal vision, normal audition, and were native French speakers (age: $M = 25$ years, 6 women). None of them had absolute pitch. All participants were paid for their participation to the experiment.

Stimuli

Three hundred twenty pairs of visual/auditory stimuli were used in the experiment. All stimuli comprised 5 notes of equal duration (quarter notes of 750 msec). White visual stimuli on a dark background were presented at the center of a computer screen (placed 60 cm in front of the participants). Auditory stimuli were generated using the piano sound produced by a general MIDI (Korg, X5DR) under computerized control. The visual 5 notes were presented all at once on the screen, and the stimulus onset was synchronized with the first note of the 5-note auditory sequence. The visual stimulus remained on the screen for the entire duration of the auditory stimulus (i.e., 3750 msec). A row of X was presented 1250 msec after auditory final note offset for 2000 msec to allow blinking. A further 1000 msec interval followed the row of X before the next trial (see Figure 9).

Visual stimuli either had stable or unstable endings (see Figure 2). To increase variability, several keys were used (major keys: C, F, B♭, E♭, G, D, A, B; minor keys: D, F, F#, G, A, B). The 40 stable stimuli were composed of 5 notes within the same key. The last note was always an in-key note (tonic, supertonic, mediant, subdominant, dominant, submediant, or leading tone). The 40 unstable-ending stimuli were similar to the stable-ending stimuli in that they were built by transposing them (the original stable-ending stimulus could be transposed by a minor or major third, by a fourth and by a fifth, above or below). After transposition, the last note was modified (by a chromatic semitone or by a tone) to become an unstable ending (mostly out-of-key tones, very rarely an unstable degree of the key).

Auditory stimuli were of 3 types. The final note of the auditory stimulus either matched (Conditions 1 and 4) or mismatched with the last note of the visual stimulus. Mismatches were either plausible or implausible endings. Whereas in the plausible mismatch conditions the last (auditory) note was always a diatonic note (Conditions 2 and 5), in the implausible mismatch conditions,

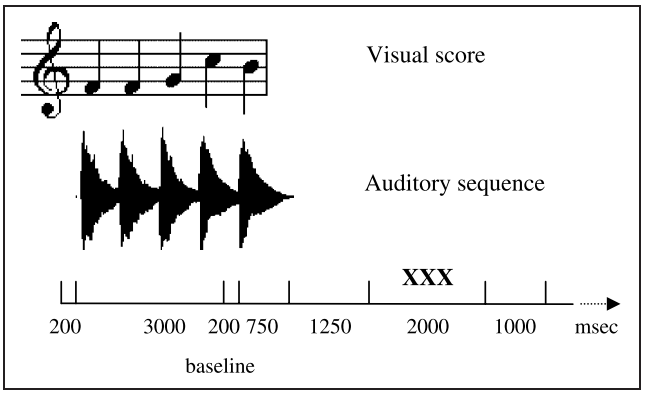

Figure 9. Illustration of the experimental design.

the last note was most often out of key and, in some rare cases, corresponded to an unstable degree of the key (Conditions 3 and 6). All mismatches did, however, preserve the contour of the visual sequence, and the change was only in the size of the interval between the fourth and the fifth note. To test for the psychological validity of the stimuli, we asked 6 musicians, naïve with respect to the aims of the experiment, to judge on a 5-point scale how much the last note they heard was stable/expected. Results showed that ending notes in Conditions 1, 2, and 5 (stable/plausible) were judged as more stable than ending notes in Conditions 3, 4, and 6 (unstable/implausible), $F = 145.35$, $p < .0001$. By contrast, comparisons between conditions to test ''stability'' within the stable/plausible conditions (e.g., comparing visual stable match with plausible/stable mismatch, i.e., Condition 1 vs. Condition 2) and unstable/implausible conditions (comparing visual unstable match with implausible/unstable mismatch, i.e., Condition 4 vs. Condition 6) showed no significant differences.

To summarize, 6 conditions were used in the experiment (see Figure 2): visual stable endings with (1) auditory (stable) match, (2) auditory plausible mismatch, (3) auditory implausible mismatch; visual unstable endings with (4) auditory (unstable) match, (5) auditory plausible mismatch, and (6) auditory implausible mismatch. Each condition comprised 40 items. Moreover, to balance the proportion of matches and mismatches, 40 fillers were added in each of the two matching conditions (expected and unexpected). Stimuli were presented in a pseudorandom order.

Procedure

Participants, comfortably seated in a Faraday box, read a 5-note score while a computer simultaneously played the notes. They were instructed to decide, as quickly and as accurately as possible, whether the final auditory note matched or mismatched with the final visual note on the score. The association between hand side (left or right) and responses (yes or no) was balanced across participants.

Data Acquisition and Analysis

Electroencephalogram (EEG) was recorded for 5200 msec starting 200 msec before the onset of the stimuli (baseline), from 28 scalp electrodes located at standard left and right hemisphere positions over frontal, central, parietal, occipital, and temporal areas (International 10/20 system sites: Fz, Cz, Pz, Oz, Fp1, Fp2, F3, F4, C3, C4, P3, P4, O1, O2, F7, F8, T3, T4, T5, T6, Fc5, Fc1, Fc2, Fc6, Cp5, Cp1, Cp2, and Cp6). These recording sites plus an electrode placed on the right mastoid were referenced to the left mastoid electrode; the data were then re-referenced offline to the algebraic average of the left and right mastoids. Impedances of these electrodes never exceeded 3 kΩ. The horizontal electrooculogram (EOG) was recorded from bipolar electrodes placed 1 cm to the left and right of the external canthi, and the vertical EOG was recorded from an electrode beneath the right eye, referenced to the left mastoid, to detect blinks and vertical eye movements. Trials containing ocular artifacts, movement artifacts, or amplifier saturation were excluded from the averaged ERP waveforms. The EEG and EOG were amplified by an SA Instrumentation amplifier with a band-pass of 0.01–30 Hz and were digitized at 250 Hz.

ERP data were analyzed by computing the mean amplitude, starting 200 msec before the last note was played and ending 2000 msec after. Only trials in which performance was correct were taken into account in the analyses. Analysis of variance (ANOVA) was used for statistical evaluation. To test the distribution of the effects, 4 regions of interest (ROIs) were selected for statistical evaluation: left anterior (F3, Fc1, and Fc5), left posterior (Cp1, Cp5, and P3), right anterior (F4, Fc2, and Fc6), right posterior (Cp2, Cp6, and P4). Note that ANOVAs that included the midline electrodes were also performed. However, because no major differences were found between these two types of analyses, we only report those including the ROIs. We used latency windows of 50 msec in the 0- to 500-msec range and windows of 100 msec in the 500- to 900-msec range. All p values reported below were adjusted with the Greenhouse–Geisser epsilon correction for nonsphericity, when appropriate. Tukey tests were used in post hoc comparisons unless specified otherwise.

Acknowledgments

This work was completed while Daniele Schön was a postdoctoral student. This work was first supported by a grant from the International Foundation for Music Research to Mireille Besson (RPA 194) and later by a grant from the Human Frontier Science Program (RG P0053/2002). We thank Andras Semjen and 2 anonymous reviewers for helpful comments on a previous version of the manuscript.

Reprint requests should be sent to Daniele Schön, INCM-CNRS, 31 Ch J. Aiguier, 13420 Marseille, France, or via e-mail: schon@incm.cnrs-mrs.fr.

Note

1. Note that to distinguish visually presented notes from auditory presented notes, we will use the terms *stable*/*unstable* for visual presentation and *plausible*/*implausible* for auditory presentation. However, the meaning of these terms is the same.

REFERENCES

Besson, M., & Faïta, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparisons of musicians with non-musicians. *Journal of Experimental*

Psychology: Human Perception and Performance, 21, 1278–1296.
Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology, 24,* 14–25.
Besson, M., & Schön, D. (2001). Comparison between language and music. *Annals of the New York Academy of Science, 930,* 232–258.
Gunter, T. C., Schmidt, B. H., & Besson, M. (2003). Let's face the music: A behavioral and electrophysiological exploration of score reading. *Psychophysiology, 40,* 742–751.
Halpern, A., & Bower, G. (1982). Musical expertise and melodic structure in memory for musical notation. *American Journal of Psychology, 95,* 31–50.
Halpern, A., & 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.
Hashimoto, T., Hirata, Y., & Kuriki, S. (2000). Auditory cortex responds in 100 ms to incongruity of melody. *NeuroReport, 11,* 2799–2801.
Jackendoff, R. (1991). Musical parsing and musical affect. *Music Perception, 9,* 199–230.
Koelsch, S., Gunter, T., Friederici, A. D., & Schroger, E. (2000). Brain indices of music processing: ''Nonmusicians'' are musical. *Journal of Cognitive Neuroscience, 12,* 520–541.
Koelsch, S., Gunter, T. C., Schroger, E., Tervaniemi, M., Sammler, D., & Friederici, A. D. (2001). Differentiating ERAN and MMN: An ERP study. *NeuroReport, 12,* 1385–1389.
Koelsch, S., Schroger, E., & Gunter, T. C. (2002). Music matters: Preattentive musicality of the human brain. *Psychophysiology, 39,* 38–48.
Krumhansl, C. (1990). *Cognitive foundation of musical pitch.* New York: Oxford University Press.
Liegeois-Chauvel, C., Musolino, A., Badier, J. M., Marquis, P., & Chauvel, 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.
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.
McGurk, H., & McDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746–748.
Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology, 94,* 26–40.
Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10,* 717–733.
Regnault, P., Bigand, E., & Besson, M. (2001). Different brain mechanisms mediate sensitivity to sensory consonance and harmonic context: Evidence from auditory event related brain potentials. *Journal of Cognitive Neuroscience, 13,* 241–255.
Schön, D., & Besson, M. (2002a). Processing pitch and duration in music reading: A RTs and ERPs study. *Neuropsychologia, 40,* 868–878.
Schön, D., & Besson, M. (2002b). Musica, maestro! In L. Bäckman & C. von Hofsten (Eds.), *Psychology at the turn of the millennium* (pp. 119–152). Hove, UK: Psychology Press.
Schön, D., Magne, C., & Besson, M. (2004). The music of speech. ARTs and ERPs study of pitch contour in prosody and music. *Psychophysiology*, *41,* 341–349.
Schön, D., Semenza, C., & Denes, G. (2001). Naming of musical notes: A selective deficit in one musical clef. *Cortex, 37,* 407–421.
Schürmann, M., Raij, T., Fujiki, N., & Hari, R. (2002). Mind's ear in a musician: Where and when in the brain. *Neuroimage, 16,* 434–440.
Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature, 385,* 308.
Simson, R., Vaughan, H. G., & Ritter, W. (1977). The scalp topography of potentials in auditory and visual Go/NoGo tasks. *Electroencephalography and Clinical Neurophysiology, 43,* 864–875.
Sloboda, J. A. (1976). Visual perception of musical notation: Registering pitch symbols in memory. *Quarterly Journal of Experimental Psychology, 28,* 1–16.
Sloboda, J. A. (1978). Psychology of music reading. *Psychology of Music, 6,* 3–20.
Sloboda, J. A. (1984). Experimental studies of musical reading: A review. *Music Perception, 2,* 222–236.