

The effect of expertise in music reading: cross-modal competence

Véronique Draï-Zerbib
CHART/LUTIN
University of Pierre & Marie Curie
Paris, France

Thierry Baccino
CHART/LUTIN
University of Paris 8
Paris, France

We hypothesize that the fundamental difference between expert and learner musicians is the capacity to efficiently integrate cross-modal information. This capacity might be an index of an expert memory using both auditory and visual cues built during many years of learning and extensive practice. Investigating this issue through an eye-tracking experiment, two groups of musicians, experts and non-experts, were required to report whether a fragment of classical music, successively displayed both auditorily and visually on a computer screen (cross-modal presentation) was same or different. An accent mark, associated on a particular note, was located in a congruent or incongruent way according to musical harmony rules, during the auditory and reading phases. The cross-modal competence of experts was demonstrated by shorter fixation durations and less errors. Accent mark appeared for non-experts as interferences and lead to incorrect judgments. Results are discussed in terms of amodal memory for expert musicians that can be supported within the theoretical framework of Long-Term Working Memory (Ericsson and Kintsch, 1995).

Keywords: expert memory, cross-modality, sight-reading, music, eye movements

Musical sight-reading and multisensory integration

The composer Robert Schumann (1848) advised young musicians to sing at first sight a score without the help of piano and to train hearing music from the page in order to become an expert in music. Nowadays, to be successful at the written test of the Superior Academy of Music entrance examination in Paris, the candidate must be able to analyze a written piece of music, without listening it. For instance, when reading without listening the JS Bach's Sonata en trio, the prospective student should be able to find out and hierarchize the different musical elements (musical form, thematic parts, dynamic, rhythmic, harmonic specifications). Although this idea of mixing different sources of information (visual and auditory)

may seem obvious to most of us and is generally thought to result from extensive learning and many years of practice on a musical instrument, there are still no satisfactory scientific explanations of this expertise. This ability to handle concurrently several sources of information represents the *cross-modal competence* of the expert musician. Music sight-reading, which consists to read a score and play concurrently an instrument without having it seen before, relies typically on this cross-modal competence. The underlying processes consist in extracting the visual information, interpreting the musical structure and performing the score by simultaneous motor responses and auditory feedback. This suggests that at least three different modalities may be involved: vision, audition and motor processes linked themselves by a knowledge musical structure if available in memory and only very few studies have investigated these interactions in a real musical context (Williamon & Egner, 2004; Williamon &

Valentine, 2002a). Furthermore, although eye-movements in music reading is a rising topic in the fields of psychology, education and musicology (Madell & Hébert, 2008, Wurtz, Mueri & Wiesendanger, 2009, Penttinen & Huovinen, 2011, Penttinen, Huovinen & Ylitalo, 2013, Lehmann & Kopiez, 2009, Gruhn, 2006, Kopiez, Weihs, Ligges & Lee, 2006), there are also only few papers investigating expertise, cross-modality and music reading using eye tracking (Drai-Zerbib & Baccino, 2005; Drai-Zerbib, Baccino & Bigand, 2012).

In recent years, we explored the role of an amodal memory in musical sight-reading by crossing visual and auditory information (cross-modal paradigm) while eye movements were recorded (Drai-Zerbib & Baccino, 2005; Drai-Zerbib, Baccino & Bigand, 2012). In our former study, expert and non-expert musicians were asked to listen, read and play piano fragments. Two versions of scores, written with or without slurs (symbol in Western musical notation indicating that the notes it embraces are to be played without separation) were used during listening and reading phases. Results showed that skilled musicians had very low sensitivity to the written form of the score and reactivated rapidly a representation of the musical passage from the material previously listened to. In contrast, less skilled musicians, very dependent on the written code and of the input modality, had to build a new representation based on visual cues. This relative independence of experts from the score was partly replicated in our latter study (Drai-Zerbib et al, 2012). For example, experts ignored difficult fingerings annotated on the score when they had prior listening. All these findings suggested that experts have stored the musical information in memory whatever the input source was.

The cross-modal competence of experts may also be illustrated at the brain level by studies using brain imagery techniques. It has been shown that the same cortical areas were activated both during reading or playing piano scores (Meister, Krings, Foltys, Boroojerdi, Muller, Topper & Thron, 2004), which is consistent with the idea that music reading involves a sensorimotor transcription of the music's spatial code (Stewart, Henson, Kampe, Walsh, Turner & Frith, 2003). Moreover, when a musical excerpt is presented for reading, the auditory imagery of the future sound is activated (Yumoto, Matsuda, Itoh, Uno, Karino & Saitoh, 2005) and while sight reading, the musician seems first translate the visual score into an auditory cue, starting around 700 or 1300 ms, ready for storage and delayed comparison with the auditory feedback (Simoens & Tervaniemi, 2013). Wong & Gauthier (2010) compared brain activity during perception of musical notation, Roman letters and mathematical symbols. They found selectivity for musical notation for expert musicians in a multimodal network of areas compared to

novices. The correlation between the activity in several of these areas and behavioral measures of perceptual fluency with musical notation suggested that activity in nonvisual areas can predict individual differences in the expertise. Thus, expertise in sight-reading, as in all complex activities, seems heavily related to multisensory integration. But how this multisensory integration may be done?

If we follow a bottom-up view, we have at first level a visual and auditory recognition implying some pattern-matching processes (Waters, Underwood & Findlay, 1997; Waters & Underwood, 1997) and information retrieval (Gillman, Underwood & Morehen, 2002). Waters et al. (1997, 1998) gave some evidence that experienced musicians recognized notes or groups of notes very quickly by a simple gaze. They assume that experts develop a more efficient encoding mechanism for identifying the shape or patterns of notes rather than reading the score note by note. Thus, they may identify salient locations on the score that facilitate a rapid identification such as interval between notes or global form (chromatic accent...). In another context, Perea, Garcia-Chamorro, Centelles & Jiménez (2013) showed that the visual encoding of note position was only approximated at low-level processes and was modulated by expertise. However, musical recognition is not only a matter of efficient pattern-matching processing, but also for inference-making processes (Lehmann & Ericsson, 1996). Hence, recognition needs also to activate some musical rules that are not mandatory written on the score but memorized for a long time by a repetitive learning.

One efficient way to represent the role of this musical memory may be carried out through the theory of Long Term Working Memory (LTWM) which proposes: "To account for the large demands on working memory during text comprehension and expert performance, the traditional models of working memory involving temporary storage must be extended to include working memory based on storage in long-term memory" (Ericsson & Kintsch, 1995, p. 211). LTWM is a model of expert memory in which structures of knowledge are built by intensive learning and years of practice. The retrieval of information relies on an association between the encoded information and a set of cues available in Long Term Memory (called retrieval cues). These retrieval cues, once stabilized, are integrated together for forming a retrieval structure. These retrieval structures contain different kinds of cues (perceptual, contextual and linguistic) allowing to reinstate rapidly the information under the focus of attention. So, the LTWM theory supposes two different types of information encoding. Firstly, a hierarchical organization of retrieval cues associated with units of encoded information (e.g. retrieval structures) and sec-

only a knowledge base that link these encoded information to schemas or patterns built and stored in memory by deliberate practice (Ericsson, Krampe & Tesch-Ramer, 1993) (see figure 1 for a representation of the LTWM).

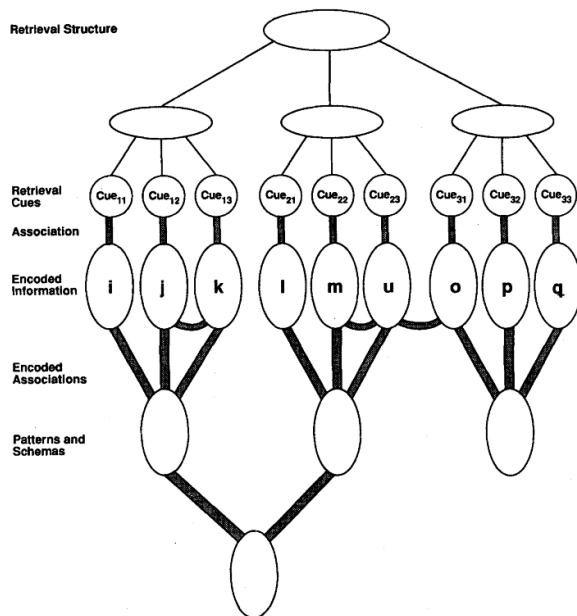


Figure 1: Two different types of encodings of information stored in long-term working memory. On the top, a hierarchical organization of retrieval cues associated with units of encoded information. On the bottom, knowledge-based associations relating units of encoded information to each other along with patterns and schemas establishing an integrated memory representation of the presented information in LTWM (adapted from Ericsson & Kintsch, 1995).

For the retrieval structures, several studies have suggested that expert musicians used hierarchical retrieval systems to recall encoded information (Halpern & Bower, 1982; Aiello, 2001; Williamon & Valentine, 2002). For example, Aiello (2001) showed that classical concert pianists memorize a work they have to play by analyzing the musical score in detail and taking more notes about the musical elements of the piece than non-experts, who simply learn the piece by heart without analyzing it. Accordingly, expert musicians rapidly index and categorize musical information to form meaningful units (Halpern & Bower, 1982) which they use later when practicing or during a performance (Williamon & Valentine, 2002). One can assume that different harmonic rules and codification may belong to these retrieval structures (Williamon & Valentine, 2002) and allow experts musicians to perform more efficiently a cross-modal task. This kind of task is supposed to be amodal at a certain level of the hierarchy and this amodality reflects the competence

of expert (Drai-Zerbib & Baccino, 2005). Some other evidence of this amodality involved in a cross-modal task has been recently shown at a brain level (Fairhall & Caramazza, 2013).

In a recognition task simpler than sight-reading, this study proposes to investigate the cross-modal competence in experts and their ability to detect harmony violations (mislocated accent mark). An accent mark ('*Marcato*') is an emphasis placed on a particular note that contributes to the dynamic, the articulation and the prosody of a musical phrase during performance (Figure 2).



Figure 2: Five types of accent mark and the most common is the 4th which indicates the emphasis of a note. We used that 4th accent mark in the experiment.

An accent mark belongs to the harmonic rules (Danhauser, 1996). If these harmony rules, represented as retrieval structures, are well grounded (as assumed in expert musicians), we hypothesize that the note retrieval will be easier when associated to a mislocated accent mark (harmonic violation). Experts, having assimilated the rules of harmony, should detect easily this bad relationship between a note and a mislocated accent and consequently produce less recognition errors. Conversely, less expert musicians should be less affected by this harmonic violation. Using a cross-modal design, low-expert and high-expert musicians have successively to hear a sequence of notes, to read the corresponding stave and recognize whether a note has been changed (same/different) and where. We hypothesize that expert should perform better due to their cross-modal competence (ability to switch from one code to another). An accent mark, associated on a particular note, was located in a congruent or incongruent way according to musical harmony rules, during the auditory and reading phases.

Methods

Participants

Sixty-four participants gave their oral consent after presentation of the global process of the study. They all have normal or corrected-to-normal vision. They were divided into two groups of expertise on the basis of their musical skill according their position in the institution. They were music students or teachers at the National Conservatory of Music in Nice, France: 26 experts (more

than 12 years of academic musical practice) and 38 non-experts (5 to 8 years of academic musical practice).

Material

The musical material consisted of 48 excerpts from the classical tonal repertoire of music. For the reading phase (visual presentation), each stave was written in treble clef, 4 bars long, using Finale software™. An accent mark placed on one specific note and contributing to the prosody of the musical phrase, was located in a congruent or incongruent position (according to harmony rules) both on visual and auditory presentation. The accent mark was randomly distributed across the four bars. Six versions of each excerpt were generated according to the crossing of 3 experimental factors: *note modification* (Original vs Modified stave), *visual congruency* (Congruent vs Incongruent vs No accent mark) and *auditory congruency* (Congruent vs Incongruent accent mark), see Figure 3a,b. For the auditory phase, the 48 original excerpts have been played by a piano teacher of the conservatory in Nice on a Steinway™ piano and recorded with a recorder mini disc Sony™. Two versions of the 48 excerpts were generated with Sound Forge software™ enhancing the sound of the note that have been cued congruently or incongruently. The experiment displaying auditory and visual stimuli was designed using E-prime™ (Psychology Software Tools Inc., Pittsburgh, USA).



Figure 3a: Example of musical stimulus: « Duo » Bruni/Van de Velde segmented into 4 AOIs (each bar). The red rectangles indicates the target bars on which the modifications may occur (for this example).

	Auditory	
	Audit. Congruent	Audit. Incongruent
Visual Congruency	Original Stave	Note Modified
	Congruent	Incongruent
No Accent Mark	Congruent	Incongruent
	Congruent	Incongruent

Figure 3b: Description of the different versions designed from an instance of the musical material (figure 2a). The versions are built according to the crossing of the 3 experimental factors (Auditory Congruency, Visual Congruency and Note modification). Firstly, during the auditory phase, the accent mark is located in a congruent/incongruent way on the target bar (top of the figure). Secondly, during the reading phase, crossing both the location of the accent mark and the note modification led to 6 different versions of the target bar (Below).

Apparatus and calibration

During recordings, participants were comfortably seated. After a nine point calibration procedure, their eye movements were sampled at a frequency of 50 Hz using the TOBII Technology 1750™ eye-tracking system. We detected the onset and offset of each saccade offline with the following velocity-based algorithm (Stampe, 1993): at each sampling point in time (t) we tested two logical conditions on the actual eye position signal $S(t_i)$ in millivolts, i.e. $\text{abs}[S(t_{i-3}) - S(t_i)] > T_{\text{sacc}}$ and $\text{abs}[S(t_{i-1}) - S(t_i)] < T_{\text{fix}}$. If both conditions were fulfilled, a sampling point is assumed to fall within the course of a saccade. The first condition is fulfilled when the voltage exceeds a chosen threshold value of $T_{\text{sacc}} = 12 \text{ mV}$; the second condition with $T_{\text{fix}} = 10 \text{ mV}$ prevents stretching of saccades and erosion of the fixation following the saccade; this resulted in a threshold velocity of about $30^\circ/\text{s}$. The music staves stimuli were presented on a 17" display; image resolution of 1024×768 pixels, from a viewing distance of 60 cm. Sony Plantronics™ headphones were connected to the computer to display auditory excerpts stimuli.

Procedure

Example trials are illustrated in Fig. 4. Participants were instructed to report whether a fragment of classical music, successively displayed both auditorily and visually on a computer screen (cross-modal presentation) was judged same or different. The order of this cross-modal presentation was always identical (firstly hearing, secondly reading). The goal was to detect whether a note has been modified between the auditory and reading phases. The participant fixated a cross during the listening of the musical fragment lasting 12 s. on average (range from 7 to 17 s.) and afterwards the fragment to read appeared on the computer display. Reading began and the participant had to click left on the mouse, if the fragment was modified from the auditory presentation, or clicked right if not modified. If he clicked left (modified staff), then an additional display asked him to choose on which of the 4 bars the note has been modified. If he clicked right, then the next trial was displayed and so on for the 48 randomized excerpts of music. The session lasted approximately 40 min.

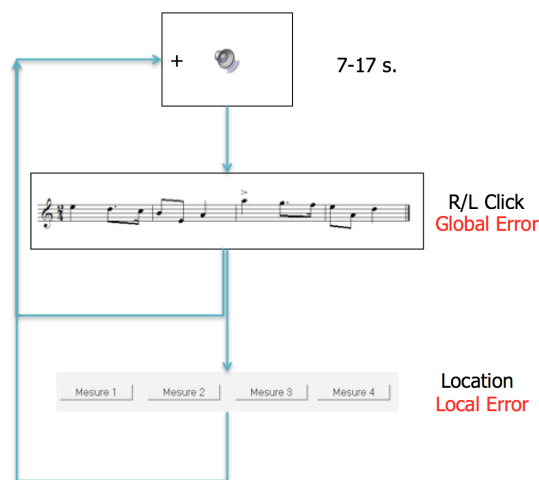


Figure 4: Schema of the procedure: 1) Listening phase (7-17 sec.); 2) Reading phase and global judgment (detecting whether a modification occurred); 3) localization judgment (on which bar).

Results

Data from the 64 participants were included in judgment and eye-movement analyses. Rate of global and local errors (proportion of incorrect responses) were judgment metrics. When the participant did not detect a

change on the staff, the error was named global error (TEG). When the error concerned on which bar the note has been modified, the error was named local error (TEL). TEL were obviously counted only when participants detected a change on the staff. First-Pass Fixation Durations (DF1), Second-Pass Fixation Durations (DF2) and Number of Fixations (Nfix) were eye movement metrics. DF1 and DF2 were calculated by dividing musical scores in four AOIs (Areas of Interest) corresponding to the different bars on the staff. DF1 is the sum of all the fixations within an AOI starting with the first fixation into that AOI until the first time the participant looks outside the AOI. All re-inspections of the AOI were labeled as DF2. All metrics were submitted to a repeated-measures analysis of variance (rmANOVA) with 3 within-subjects factors: 2 *auditory congruency* (auditory congruent/incongruent accent mark), 3 *visual congruency* (visual congruent/incongruent/none accent mark) and 2 *note modification* (original/modified staff) and 1 between-subjects factor: *musical expertise* (expert/non expert). Overall average values for eye-tracking data and errors are summarized in Table 1 and Table 2.

Table 1: Average values (mean and standard deviation) for eye-tracking data (DF1, DF2, NFix) and errors rate (TEG, TEL) according to visual congruency, note modification and musical expertise for Auditory Congruent cue.

Phase 1 - Listening with Auditory Congruent cue							
Phase 2 - Reading		Visual Congruent		Visual Incongruent		Without	
Phase 2 - Modification		M	O	M	O	M	O
DF1	E	882 (51)	834 (60)	925 (61)	878 (58)	872 (58)	872 (58)
	NE	1456 (42)	1488 (50)	1429 (51)	1453 (48)	1434 (48)	1471 (48)
DF2	E	1406 (148)	1234 (150)	1510 (181)	1427 (169)	1261 (134)	1354 (190)
	NE	1878 (122)	1944 (124)	1893 (150)	1903 (139)	1845 (112)	2004 (157)
NFix	E	5.65 (0.36)	5.55 (0.36)	6.02 (0.39)	5.84 (0.36)	5.45 (0.34)	5.73 (0.43)
	NE	6.53 (0.30)	6.74 (0.33)	6.74 (0.30)	6.67 (0.30)	6.77(0.28)	6.89 (0.36)
TEG	E	0.13 (0.03)	0.16 (0.03)	0.17 (0.17)	0.12 (0.03)	0.16 (0.02)	0.20 (0.03)
	NE	0.40 (0.03)	0.34 (0.02)	0.45 (0.02)	0.33 (0.03)	0.37 (0.02)	0.35 (0.03)
TEL	E	0.80 (0.08)		0.49 (0.02)		0.50 (0.03)	
	NE	0.72 (0.06)		0.70 (0.02)		0.73 (0.02)	

Table 2: Average values (mean and standard deviation) for eye-tracking data (DF1, DF2, NFix) and errors rate (TEG, TEL) according to visual congruency, note modification and musical expertise for Auditory Incongruent cue.

Phase 1 - Listening with Auditory Incongruent cue							
Phase 2 - Reading	Visual Congruent		Visual Incongruent		Without		
Phase 2 - Modification	M	O	M	O	M	O	
DF1	E	1018 (64)	905 (61)	897 (51)	898 (52)	840 (55)	832 (47)
	NE	1568 (53)	1520 (50)	1528 (42)	1453 (43)	1484 (45)	1441 (39)
DF2	E	1424 (183)	1383 (191)	1367 (143)	1415 (181)	1386 (169)	1205 (173)
	NE	1960 (151)	2046 (158)	1902 (118)	1894 (150)	1915 (140)	1955 (143)
NFlx	E	5.93 (0.40)	5.89 (0.38)	5.69 (0.34)	5.95 (0.42)	5.48 (0.37)	5.53 (0.45)
	NE	7.12 (0.33)	6.97 (0.31)	6.93 (0.28)	6.93 (0.34)	6.68 (0.31)	7.05 (0.37)
TEG	E	0.19 (0.03)	0.19 (0.03)	0.16 (0.03)	0.15 (0.02)	0.22 (0.03)	0.13 (0.03)
	NE	0.46 (0.03)	0.42 (0.03)	0.41 (0.02)	0.40 (0.02)	0.47 (0.03)	0.30 (0.03)
TEL	E	0.32 (0.07)		0.53 (0.02)		0.55 (0.02)	
	NE	0.71 (0.06)		0.72 (0.02)		0.75 (0.02)	

Graphs below (figure 5a and 5b) summarize the effect of expertise according to fixation durations (DF1 and DF2) and errors rate (global and local).

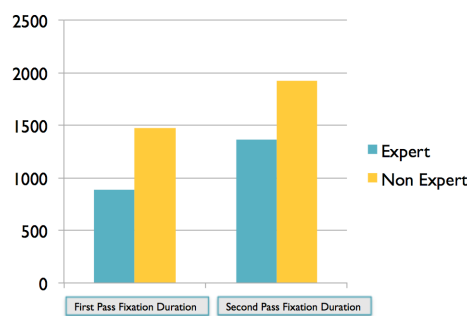


Figure 5a: Mean Fixation Durations (1st pass and 2nd pass) in ms according to the musical expertise.



Figure 5b: Error Rates (global and local errors) in percentage according to the musical expertise.

Errors analysis

Global Errors

Global errors rate was lower for expert (17%) compared to non-expert musicians (39%), showing a better multimodal representation of the same musical frag-

ment, $F(1,62)=111,40$; $p<.001$, $\eta^2 = .64$ (figure 5b). An Incongruent auditory accent mark (29%) involved more errors than a Congruent one (26%), $F(1,62)=9,88$; $p<.01$, $\eta^2 = .14$. Global error rate was lower when the staffs were original (26%) versus modified (30%); $F(1,62)=6,91$, $p<.025$, $\eta^2 = .10$. Expertise interacted significantly with Visual Congruency of the accent mark, $F(2,124)=4,83$; $p<.01$, $\eta^2 = .07$. Non expert carried out fewer errors when the accent mark was not written on the score rather than it was congruent, $F(1,62)=8,62$, $p<.01$, $\eta^2 = .14$ or incongruent $F(1,62)=4,58$, $p<.05$, $\eta^2 = .07$. This accent mark seems to disturb non expert since they are very dependent of the written code (Drai-Zerbib & Baccino, 2005). No significant difference appeared for Experts. Auditory congruency interacted significantly with Visual congruency, $F(2,124)=3,40$; $p<.05$, $\eta^2 = .05$. The 3-way interaction between Note Modification, Auditory cong. and Visual cong., $F(2,124)=7,57$; $p<.001$, $\eta^2 = .11$ emphasizes the impact of auditory and visual cues on the rate of global errors according to the modification of the staffs.

Local Errors

Local errors rate was lower for experts (53%) compared to non-experts (72%), $F(1,62)=68,57$; $p<.001$, $\eta^2 = .53$ (figure 5b). Expertise interacted significantly with Auditory congruency, $F(1,62)=5,07$; $p<.05$, $\eta^2 = .08$. Planned comparisons showed no significant difference for non-expert musicians. On the contrary, experts made fewer errors whenever they received an incongruent accent mark during listening, $F(1,62)=7,56$; $p<.01$, $\eta^2 = .11$. Incongruent accent mark may increase the attentional level during listening rendering the expert more accurate to localize the bar on which the modification occurred. The same explanation may be given for the interaction between Auditory Cong. and Visual Cong. $F(2,124)=9,24$; $p<.001$, $\eta^2 = .13$. An incongruent auditory accent mark made easier the detection of the modified note when the accent mark was congruent in the reading phase, $F(1,62)=5,32$, $p<.025$, $\eta^2 = .08$ (figure 6). But the interesting point is that effect is only true for expert musicians as emphasized by the 3-way interaction between Expertise, Auditory Cong. and Visual Cong., $F(2,124)=7,50$; $p<.001$, $\eta^2 = .11$. Non expert musicians did not make any difference between congruent or incongruent accent mark whatever the modality of presentation (figure 7). This effect may be related to the DF1 where a harmonic mismatch increased fixations duration. A time accuracy trade-off may explain these opposite effects (longer fixations and fewer errors), experts spent more time when the mismatch occurred and consequently made fewer errors for localizing the target.

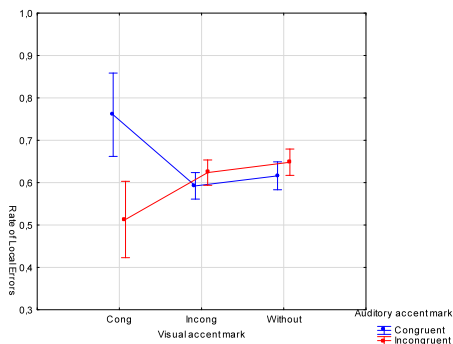


Figure 6: Rate of local errors as function of Auditory and Visual congruency of the accent mark.

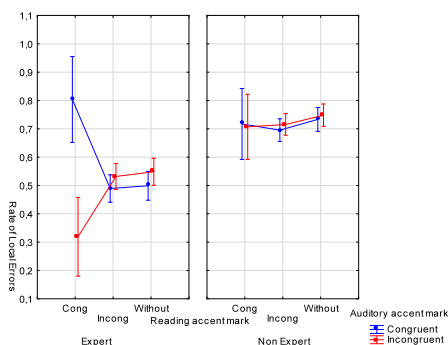


Figure 7: Rate of local errors as function of Expertise, Auditory congruency and Visual congruency of the accent mark.

Eye movement analysis

First-Pass Fixation Duration (DF1)

Experts musicians made shorter DF1 than non-experts, $F(1,62)=77.42$; $p<.001$, $\eta^2 = .56$ (figure 3a). DF1 was significantly shorter when the auditory accent mark was congruent rather than incongruent, $F(1,62)=14.15$; $p<.001$, $\eta^2 = .19$ and when the staffs were original than modified, $F(1,62)=6.38$; $p<.025$, $\eta^2 = .09$. Auditory congruency interacted significantly with note modification in DF1, $F(1,62)=6.31$; $p<.025$, $\eta^2 = .09$. There was a three-way interaction between auditory congruency, note modification and expertise, $F(1,62)=4.20$; $p<.05$, $\eta^2 = .06$. Planned comparisons shown that these effects are mainly due to non-expert musicians who spent more time on modified scores when the prior auditory accent mark was incongruent, $F(1,62)=9.86$; $p<.01$, $\eta^2 = .14$. The effect was only marginally significant for experts ($p=.07$, $\eta^2 = .05$). It seems that the incongruent mark attracted the attention on a specific note during listening and consequently disrupting the recognition process during reading.

This recognition is more complex when the note has been modified. Since the effect appeared on the modified score, it may suggest that pattern-matching processes involved during reading seems more difficult for non-experts. Furthermore, this suggests that retrieval structures as hypothesized are less efficient rendering the retrieval of the modified note more complex. There were also a two-way interaction between Auditory Congruency and Visual Congruency of the accent mark, $F(2,124)=14.84$; $p<.001$, $\eta^2 = .19$ and a three-way interaction between Expertise, Visual cong. and Auditory cong. which illustrates the different levels of processing according the expertise, $F(2,124)=3.14$; $p<.05$, $\eta^2 = .05$. Expert musicians made longer DF1 when they read staffs presented with a congruent accent mark only when they listened previously an incongruent accent mark, $F(1,62)=20.07$; $p<.001$, $\eta^2 = .48$. Since the accent mark belong to harmony rules, only experts seems to access to this level and they are more disrupted when a harmonic mismatch (accent mark) occurred. Probably, they attempted to solve this mismatch between listening and reading increasing as a consequence DF1. For non-experts, the mismatch is not detected.

Second pass fixation duration

Experts made shorter DF2 than non-experts, $F(1,62)=8.64$; $p<.01$, $\eta^2 = .122$ (figure 3a). No other significant effects were found on DF2, indicating that musicians had mainly proceed information during the first reading (DF1).

Number of fixations

The number of fixations was significantly lower for experts than non-experts, $F(1,62)=5.91$; $p<.025$, $\eta^2 = .29$. Moreover, auditory congruency interacted significantly with visual (reading) congruency, $F(2,124)=4.80$; $p<.01$, $\eta^2 = .07$. If during listening, musicians received an incongruent auditory accent mark, they made less fixations during reading when scores did not contain any accent mark rather than an incongruent one, $F(1,62)=5.02$, $p<.05$, $\eta^2 = .07$, or a congruent one $F(1,62)=8.37$, $p<.01$, $\eta^2 = .12$.

Discussion

This study investigated how cross-modal information was processed by expert and non-expert musicians using the eye tracking technique. The main goal was to know 1) whether the fundamental difference between expert and non-expert musicians was the capacity to efficiently integrate a cross-modal information, 2) whether this capacity relied on a kind of expert memory (i.e. musical retrieval structures according the LTWM theory) built during

many years of learning and extensive practice. For testing this latter condition, a harmonic cue (accent mark) was located in an incongruent or congruent location on the stave (congruency according to the tonal harmonic rules) both during the hearing and reading phases. These accent marks served for testing whether they can be used as retrieval cues in expert memory to recognize more efficiently a change in a note.

We found that mainly eye movements varied according to the level of expertise, experts had lower DF1, DF2, Nfix than non-experts. These results are consistent with previous studies, (Rayner & Pollatsek, 1997; Waters, Underwood & Findlay, 1997; Waters & Underwood, 1997; Drai-Zerbib & Baccino, 2005; Drai-Zerbib et al, 2012) in which expertise in musical reading was associated with both a decreasing in number of fixations and fixation durations.

Errors analysis shown that expert musicians performed better than non-experts even if crossing musical fragments from different modalities needs more cognitive resources: experts made 17% of global errors and 53% of local errors (non-experts respectively 39% and 72%). Experts and non-experts were also differently affected by an incongruent auditory accent mark. Non-experts made fewer errors when the accent mark was not written on the score. The presence of that accent mark appeared to disturb them and as indicated previously this may be related to their dependence of the written code (Drai-Zerbib & Baccino, 2005). The disruption is logically more important when an auditory incongruent accent mark was listened and when the staff was modified, rendering the recognition task highly difficult. Any change on the score both auditory and visual carried out difficulty for non-experts. They do not have probably sufficient stable musical knowledge in memory (e.g. retrieval structures in the LTWM theory) that may compensate for these musical violations.

The most interesting finding is probably the three-way interaction between expertise, auditory and visual congruency observed on DF1 and local errors. Experts gazed longer the score when there was a mismatch on the position of the accent mark between listening and reading and consequently they made fewer local errors. Firstly, this time-accuracy trade-off points out the capacity for experts to modulate their visual perception according to the difficulty encountered. Secondly, this mismatch is caused by a musical violation in harmonic rules (inappropriate accent mark) supposed to be managed in musical memory. Only experts have access to that knowledge and this effect provides evidence that accent mark is a good candidate for being a retrieval cue in musical memory. This finding is to be related to other cues that we found in prior researches (fingerings, slurs...) and we assumed

they are all integrated in retrieval structures as mentioned in the LTWM model.

For musical knowledge structures acquired with long practice and activated by these retrieval structures, they are represented as schemas or patterns. While schemas are not clearly defined in the LTWM theory, they are used in many models for representing knowledge in memory (Gobet, 1998). A schema is a memory structure constituted both of constants (fixed patterns) and of slots where variable patterns may be stored; a pattern is a configuration of parts into a coherent structure (see also Bartlett, 1932; Kintsch, 1998). In music, it is highly plausible that conceptual knowledge such tonality and musical genre may be stored in memory as schemas (Shevy, 2008). Following linguistic formalization, musical schemas have also been represented as embedded generative structures (Clarke, 1988). The author provided instances of such musical knowledge structures based on a composition's formal structure and he claimed that "skilled musicians retrieve and execute compositions using hierarchically organized knowledge structures constructed from information derived from the score and projections from players' stylistic knowledge" (as cited by Williamon & Valentine, 2002, p. 7). So, in music reading, musical knowledge supposed to be stored as retrieval structure, can be accessed very quickly from musical cues (Drai-Zerbib & Baccino, 2005; Ericsson & Kintsch, 1995; Williamon & Egner, 2004; Williamon & Valentine, 2002b). It follows that some perceptual cues might be less important for experts since they are capable of using their musical knowledge to compensate for missing (Drai-Zerbib & Baccino, 2005) or incorrect information (Sloboda, 1984). Conversely, less-experts musicians may process perceptual cues even if they are not adapted for the performance.

While these issues are potentially important for our understanding of cross-modal competence for musicians, several points will be improved in future investigations.

Firstly, in classifying the level of expertise in music. In this experiment, we relied on the academic level reached by every musician, which is the common approach for determining musical competence. Musical abilities can also be objectively assessed using test batteries providing a profile of music perception skills (Gordon, 1979; Law & Zentner, 2012). However, it is now possible to determine automatically the level of expertise by classifying the musician as function of the performance made during a task (eye-movements, errors). Machine learning techniques (SVM, MVPA,...) may be used for this purpose and recent studies shown successfully how the difficulty of the task (Henderson, Shinkareva, Wang, Luke, & Olejarczyk, 2013) or the

level of stress (Pedrotti et al., 2013) may be classified by eye movements.

Secondly, our findings underline the ability for expert musicians to switch from one code (auditory) to another (visual). This modality switching capacity is part of the cross-modal competence for experts as it has been demonstrated in other type of expertise such as video gamers or bilinguals (Bialystok, 2006). We did not investigate the counterbalanced condition (visual → auditory) since music reading was our main goal in this experiment especially by analyzing eye movements. However, that condition is valid if we consider using the Eye-Fixation-related Potentials that we developed previously (Baccino, 2011; Baccino & Manunta, 2005). The technique allows to draw the time course of brain activity along with fixations during reading but it allows also to investigate the brain activity even if virtually no eye movements are made (as in the auditory condition). Consequently, it might be interesting to analyze the brain activity when hearing the musical sound after reading and see whether violations on the stave may be detected by experts.

Finally, reading music has been investigated for several decades but a lots of unknown factors are still pending and the development of novel on-line techniques (EFRP, fNIRS...) and appropriate statistical or computing procedures can put further light on this topic.

Acknowledgements

We would like to thank director, students and teachers of the Conservatory of music in Nice who participated in this experiment, especially J.L. Luzignant, professor of harmony, for helping at creating the material. A special thanks also for the two anonymous reviewers who help us to improve the manuscript.

References

- Aiello, R. (2001). Playing the piano by heart, from behavior to cognition. *Annals of the New York Academy of Science*, 930, 389–393.
- Baccino, T. (2011). Eye movements and concurrent event-related potentials: Eye fixation-related potential investigations in reading. In S. P. Liversedge, I. D. Gilchrist & S. Everling (Eds.), *The Oxford handbook of eye movements*. (pp. 857-870). New York, NY US: Oxford University Press.
- Baccino, T., & Manunta, Y. (2005). Eye-Fixation-Related Potentials: Insight into Parafoveal Processing. *Journal of Psychophysiology*, 19(3), 204-215.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, UK: Cambridge Univ. Press.
- Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 60(1), 68-79.
- Clarke, E. F. (1988). *Generative principles in music performance*. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 1–26). Oxford, UK: Clarendon Press.
- Danhauser, A. (1996). *Théorie de la musique*, Paris: Editions Henri Lemoine.
- Drai-Zerbib, V., & Baccino, T. (2005). L'expertise en lecture musicale: intégration intermodale. *L'année psychologique*, 3(105), 387-422.
- Drai-Zerbib, V., Baccino, T., & Bigand, E. (2012). Sight-reading expertise: Cross-modality integration investigated using eye tracking. *Psychology of Music*, 40(2), 216-235.
- Ericsson, K., & Kintsch, W. (1995). Long-Term Working Memory. *Psychological Review*, 102(2), 211-245.
- Ericsson, K. A., Krampe, R. T., & Tesch-Ramer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363-406.
- Fairhall, S. L., & Caramazza, A. (2013). Brain regions that represent amodal conceptual knowledge. *The Journal of Neuroscience*, 33(25), 10552-10558.
- Gillman, E., Underwood, G., & Morehen, J. (2002). Recognition of visually presented musical intervals. *Psychology of Music*, 30(1), 48-57.
- Gobet, F. (1998). Expert memory: a comparison of four theories. *Cognition*, 66(2), 115-152.
- Gordon, E. (1979). Developmental music aptitude as measured by the Primary Measures of Music Audiation. *Psychology of Music*, 7(1), 42-49.
- Gruhn, W., Litt, F., Scherer, A., Schumann, T., Weiß, E. M., & Gebhardt, C. (2006). Suppressing reflexive behaviour: Saccadic eye movements in musicians and non-musicians. *Musicae Scientiae*, 10(1), 19-32.
- Halpern, A., & Bower, G. (1982). Musical expertise and melodic structure in memory for musical notation. *American Journal of Psychology*, 95(1), 31–50.
- Henderson, J. M., Shinkareva, S. V., Wang, J., Luke, S. G., & Olejarczyk, J. (2013). Using Multi-Variable Pattern Analysis to Predict Cognitive State from Eye Movements. *PLoS One*, 8(5).
- Kintsch, W. (1998). *Comprehension: a paradigm for cognition*. Cambridge: Cambridge University Press.
- Kopiez, R., Weihs, C., Ligges, U., & Lee, J. I. (2006). Classification of high and low achievers in a music sight-reading task. *Psychology of Music*, 34(1), 5-26.
- Law, L. N. C., & Zentner, M. (2012). Assessing Musical Abilities Objectively: Construction and Validation of the Profile of Music Perception Skills. *PloS one*, 7(12).
- Lehmann, A. C., & Ericsson, K. A. (1996). Performance without preparation: Structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology*, 15(1), 1-29.

- Lehmann, A. C., & Kopiez, R. (2009). Sight-reading. In S. Hallam, I. Cross & M. Thaut (Eds.), *The Oxford handbook of music psychology* (pp. 344-351). Oxford: Oxford University Press.
- Madell, J., & Hébert, S. (2008). Eye Movements and music reading: where do we look next? *Music Perception*, 26(2), 157-170.
- Meister, I. G., Krings, T., Foltys, H., Boroojerdi, B., Muller, M., Topper, R., & Thron, A. (2004). Playing piano in the mind--an fMRI study on music imagery and performance in pianists. *Cognitive Brain Research*, 19(3), 219-228.
- Pedrotti, M., Mirzaei, M. A., Tedesco, A., Chardonnet, J.-R., Mérienne, F., Benedetto, S., & Baccino, T. (2013). Automatic stress classification with pupil diameter analysis. *International Journal of Human-Computer Interaction*, doi: 10.1080/10447318.2013.848320.
- Penttinen, M., & Huovinen, E. (2011). The early development of sight-reading skills in adulthood: A study of eye movements. *Journal of Research in Music Education*, 59(2), 196-220.
- Penttinen, M., Huovinen, E., & Ylitalo, A.-K. (2013). Silent music reading: Amateur musicians' visual processing and descriptive skill. *Musicae Scientiae*, 17(2), 198-216.
- Perea, M., Garcia-Chamorro, C., Centelles A., Jiménez, M. (2013). Coding effects in 2D scenario: the case of musical notation. *Acta Psychologica*, 143(3), 292-297.
- 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(2), 49-53.
- Schumann, R., & Schumann, C. (2009 - Original work published 1848). *Journal intime, Conseils aux jeunes musiciens*. Paris: Editions Buchet-Chastel. (Original work published 1848).
- Shevy, M. (2008). Music genre as cognitive schema: Extramusical associations with country and hip-hop music. *Psychology of Music*, 36(4), 477-498.
- Simoens, V. L., & Tervaniemi, M. (2013). Auditory Short-Term Memory Activation during Score Reading. *PLoS One*, 8(1).
- Sloboda, J. A. (1984). Experimental Studies of Musical Reading: A review. *Music Perception* 2(2), 222-236.
- Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments & Computers*, 25(2), 137-142.
- Stewart, L., Henson, R., Kampe, K., Walsh, V., Turner, R., & Frith, U. (2003). Brain changes after learning to read and play music. *NeuroImage*, 20(1), 71-83.
- Waters, A., Underwood, G., & Findlay, J. (1997). Studying expertise in music reading : Use of a pattern-matching paradigm. *Perception & Psychophysics*, 59(4), 477-488.
- Waters, A., & Underwood, G. (1998). Eye Movements in a Simple Music Reading Task: a Study of Expert and Novice Musicians. *Psychology of Music*, 26, 46-60.
- Williamon, A., & Valentine, E. (2002). The Role of Retrieval Structures in Memorizing Music. *Cognitive Psychology*, 44(1), 1-32.
- Williamon, A., & Egner, T. (2004). Memory structures for encoding and retrieving a piece of music: an ERP investigation. *Cognitive Brain Research*, 22(1), 36-44.
- Wong, Y.K., Gauthier, I. (2010) Multimodal Neural Network recruited by expertise with musical notation. *Journal of Cognitive Neuroscience*, 22(4), 695-713.
- Wurtz, P., Mueri, R. M., & Wiesendanger, M. (2009). Sight-reading of violinists: eye movements anticipate the musical Xow. *Experimental brain research*.
- Yumoto, M., Matsuda, M., Itoh, K., Uno, A., Karino, S., Saitoh, O., et al. (2005). Auditory imagery mismatch negativity elicited in musicians. *Neuroreport*, 16(11), 1175-1178.