# A grammar of action generates predictions in skilled musicians

Giacomo Novembre [a,b,*], Peter E. Keller [a]

[a] Music Cognition and Action Group, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[b] International Max Planck Research School "Neuroscience of Communication: Function, Structure and Plasticity" (IMPRS NeuroCom), Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

### ABSTRACT

The present study investigates shared representations of syntactic knowledge in music and action. We examined whether expectancy violations in musical harmonic sequences are also perceived as violations of the movement sequences necessary to produce them. Pianists imitated silent videos showing one hand playing chord sequences on a muted keyboard. Results indicate that, despite the absence of auditory feedback, imitation of a chord is fastest when it is congruent with the preceding harmonic context. This suggests that the harmonic rules implied by observed actions induce expectations that influence action execution. As evidence that these predictions are derived at a high representational level, imitation was more accurate for harmonically incongruent chords than for congruent chords executed with unconventional fingering. The magnitude of the effects of context and goal prioritization increased with musical training. Thus, musical training may lead to a domain-general representation of musical grammar, i.e., to a grammar of action.

© 2011 Elsevier Inc. All rights reserved.

## 1. Introduction

The same sensory event can acquire markedly different significance if perceived in different *local* contexts (depending on the events that preceded it) and, as a consequence, elicit very different responses in the brain and behavior. Evidence in favor of this general principle of perception, which has been referred to as "situatedness" (Clark, 1999; Markman & Dietrich, 2000), has been observed in studies investigating the perception of complex auditory stimuli such as speech (Friederici, Rüschemeyer, Hahne, & Fiebach, 2003; Hahne & Friederici, 2002), musical harmonic sequences (Koelsch, Fritz, Schulze, Alsop, & Schlaug, 2005; Maess, Koelsch, Gunter, & Friederici, 2001; Tillmann, Janata, & Bharucha, 2003) and, more generally, the visual perception of sequences of movements forming goal-directed actions (Iacoboni et al., 2005; Kilner, Vargas, Duval, Blakemore, & Sirigu, 2004).

A key property shared by stimuli in the speech, music and action domains is that they consist of discrete items (i.e. words, musical sounds, and movements) strung together in such a way that meaning builds up over the course of a sequence. Specific rules determining how these chains of items are combined into structures help us to generate expectations for up-coming events. Thus, one function of language-specific grammars is to dictate how words need to be arranged in a meaningful sentence, culture-dependent tonal systems determine the degree of harmonic pleasantness and tension in a succession of sounds (Krumhansl, 2004; Patel, 2008), and task-specific motor constraints govern how sequences of movements need to be ordered to successfully complete a goal-directed action (Bernstein, 1996; Grafton & Hamilton, 2007).

---

\* Corresponding author. Address: Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstrasse 1a, 04103 Leipzig, Germany. Fax: +49 341 9940 113.

*E-mail address:* novembre@cbs.mpg.de (G. Novembre).

These rules are implicitly learned through individual experience (e.g. the grammar of a specific language, the tonal system of a certain culture, proficiency in a motor task) and contribute to the establishment of internalized knowledge structures that make online prediction possible (Castellano, Bharucha, & Krumhansl, 1984; Elman et al., 1996; Loui, Wessel, & Hudson Kam, 2010; Meyer, 1956; Underwood & Bright, 1996). In this article, we will refer to these rules as "syntactic-structures". We do so not in a language specific sense, but rather in terms of logical connotation: the branch of logic that deals with how the meaning of individual signs can be altered depending on contextual factors such as their position within a sequence (Bunnin & Jiyuan, 2004).

For some time, there has been interest in shared syntactic structures in language and music (Lerdahl & Jackendoff, 1983; Patel, 2003). In support of such overlap, neuroimaging studies have found that the brain areas involved in syntactic analysis during auditory language comprehension – including Broca's area – are also recruited during the perception of harmonic sequences in music (Koelsch et al., 2005; Maess et al., 2001). Furthermore, lesions of these areas cause impaired syntactic processing in both language and music (Patel, Iversen, Wassenaar, & Hagoort, 2008; Sammler, Koelsch, & Friederici, in press). Such findings are consistent with the recent proposal that the brain regions involved in processing linguistic and musical syntactic structures may be domain-general and therefore devoted to the processing of syntactic structures that determine the meaning of all varieties of action (Fadiga, Craighero, & D'Ausilio, 2009; Pulvermüller & Fadiga, 2010; see also Janata & Grafton, 2003). This leads to the intriguing hypothesis of a general "grammar of action", which constrains how a succession of movements needs to be ordered if they are to be perceived as a goal-directed action.

The hypothesis of a grammar that controls movements had been previously advanced in relation to specific motor tasks, such as drawing (Goodnow & Levine, 1973), music production (Baily, 1977) and dance performance and observation (Opacic, Stevens, & Tillmann, 2009). However, interest in the domain-generality of this grammar has heightened only recently (Clerget, Winderickx, Fadiga, & Olivier, 2009; Fazio et al., 2009), motivating questions concerned with, for example, "how and to what extent action, music and language share similar syntactic structures and mechanisms" (Fadiga, Craighero, & D'Ausilio, 2009, p. 456).

The present study aimed at providing behavioral evidence in favor of shared syntactic structures in action and music. We assumed that such overlap should be dependent on expertise to the extent that long-term musical experience establishes internalized knowledge structures (i.e. knowledge of musical harmony and proficiency in particular motor skills) that contribute to the derivation of strong expectations. If a musical context can induce specific expectations in the auditory domain (i.e. a specific tone or combination of simultaneous tones in a chord), then establishing the same context in the action domain should generate comparable expectations (i.e. the specific movement required to produce the expected tone(s)). Given the growing body of evidence for especially well developed audio-motor coupling in musicians' brains (Bangert & Altenmüller, 2003; D'Ausilio, Altenmüller, Olivetti Belardinelli, & Lotze, 2006; Drost, Rieger, Brass, Gunter, & Prinz, 2005; Haueisen & Knösche, 2001), we hypothesized that violations of syntactic rules in the musical domain are also violations of syntactic rules in the action domain for experienced performers.

In order to investigate this question, we employed a task that required pianists to imitate actions shown in silent videos depicting the right hand of a pianist producing chord sequences on a piano keyboard. The chord sequences were based on those previously used to induce harmonic expectations in the auditory domain (Koelsch et al., 2005; Maess et al., 2001). In our first experiment, the action sequences established a local musical context over the course of four chords, and then a fifth chord was presented that could be either congruent or incongruent with this context according to the rules of classical harmony (Bharucha & Krumhansl, 1983; Krumhansl & Kessler, 1982). As the sequences were presented in the action domain, we termed the Incongruent condition "Goal-Incongruent". In fact, this violation did not explicitly involve auditory representations of the harmonic sequences of chords (i.e. as the videos were silent), but rather the movements necessary for producing them. In addition, to assess the extent to which context-based motor expectations are goal-directed and independent of the movements employed to achieve the goal, we included a third condition in which a harmonically congruent final chord was executed using unconventional fingering (Manner Incongruent, i.e., the performance manner was unconventional) (see Fig. 1). A 'double violation' condition (in which both the goal and manner are incongruent) was not included because we sought to use ecologically valid action sequences that are consistent with those commonly encountered in musical contexts that abide by the rules of Western harmony. This design was intended to allow us to ascertain the level(s) of the motor control hierarchy at which local context is established. We were specifically interested in the distinction between high-level goal representations of the target keys and lower level motor representations of the specific fingers used to press these keys.

Two experiments were conducted to address the above issues. In both, skilled pianists were asked to watch and simultaneously imitate the chord sequences on a muted keyboard. The first experiment employed sequences of five chords, while the second experiment used shorter sequences comprising two chords in order to reduce contextual information for control purposes (see below). The imitation task, which required direct matching of the observed action onto a corresponding internal motor representation (cf. Buccino et al., 2004; Haslinger et al., 2005; Iacoboni et al., 1999), was performed in the total absence of sound: both the videos and the keyboard were silent. To the extent that the harmonic contexts defined by the chord sequences are converted into action contexts, the process of observation and imitation should lead to expectations in the action domain that are analogous to those that occur when listening to auditory chord sequences. Therefore, we expected imitation of congruent chords to be faster than imitation of (Goal and Manner) incongruent chords, implying the existence of a context-dependent action prediction mechanism.

In addition, we were interested in comparing pianists' accuracy (both in terms of keys pressed and fingering used) by examining errors produced when imitating chords in the Goal and Manner Incongruent conditions. This comparison was

expected to shed light on the nature of the hypothesized prediction mechanism by revealing whether the goal of the actions had priority over specific movements.

Finally, we expected participants' musical expertise to modulate the magnitude of the above effects. This hypothesis was based on previous studies indicating that musical enculturation and training leads to richer mental representations of hierarchical structure in musical harmony (Hannon & Trainor, 2007; Jentschke & Koelsch, 2009; Koelsch, Schmidt, & Kansok, 2002; Loui & Wessel, 2007; Regnault, Bigand, & Besson, 2001), and that the strength of auditory-motor integration can vary as a function of musical training (Bangert & Altenmüller, 2003; D'Ausilio, Altenmüller, Olivetti Belardinelli, & Lotze, 2006; Drost et al., 2005; Haueisen & Knösche, 2001; Keller & Koch, 2008; Lahav, Saltzman, & Schlaug, 2007). Therefore, we expected that disruption by incongruent chords, and the prioritization of goal representations, should be relatively strong in participants with large amounts of musical experience.

## 2. Experiment 1

### 2.1. Methods

#### 2.1.1. Participants

Thirteen right-handed pianists (seven females), aged 19–34 years (mean = 23.92; SD = 3.66), participated in Experiment 1. All had at least 10 years of formal training in classical music. A questionnaire assessed the age at which piano studies commenced (mean = 9.08 years; SD = 3.61), years of piano training (mean = 14.84; SD = 2.91), and weekly amount of practice (mean = 1.98 h; SD = 2.78). The pianists were naïve with regard to the purpose of the study and were not debriefed after the experiment was completed.

#### 2.1.2. Stimuli

Stimuli were videos showing a female pianist's right hand playing sequences of five chords on a muted keyboard (Yamaha EZ200). We used a keyboard equipped with red light-emitting diodes (LED) positioned at the top of each key because, in a pilot study, we found that the identity of chords executed on a muted keyboard was occasionally unclear to participants. The LEDs were illuminated for the duration of each key press and made the identity of the pressed keys clear to the participant. Although the original keyboard already housed this LED feature, we modified it by moving the position of the LEDs approximately 3 cm higher, towards the back of the keyboard, to avoid occlusion by the hand. LEDs on black and white keys appeared to differ slightly in intensity due to the differences in the luminance of the keys' plastic surfaces (see Fig. 1, right panel).

A total of 120 different chord sequences were created for the experiment: 40 were regular in the sense that they came to a conventional harmonic resolution (Congruent), 40 were irregular in that they ended in an uncommon and unresolved harmony (Goal Incongruent), and 40 were irregular in terms of fingering but not harmony (Manner Incongruent) (see Fig. 1, right panel).[1] The chord sequences in each condition were in the key of C, D, F, or G major (10 sequences per key). All chords included in the sequences consisted of three piano keystrokes. The first chord was always the tonic of the given tonal context. Chords at the second position were the tonic, supertonic, or subdominant. Chords at the third position were the tonic, subdominant, supertonic or submediant. Chords at the fourth position were dominant seventh chords presented in root position, or in first or third inversion. The chord at the fifth position differed between three conditions: a tonic chord (Congruent), a Neapolitan chord (Goal Incongruent), or a tonic chord played with unconventional fingering (Manner Incongruent). A Neapolitan chord is a minor subdominant with a diminished sixth instead of a fifth. Although classical Neapolitan chords most often occur in first inversion, they were presented in both first and second inversion, which is permissible according to the principles of Western classical harmony (Schoenberg, 1969). Chord sequences were constructed in such a way that they had different 'melodic contours' (e.g. starting with the first, third, or fifth degree of the tonic chord) in the top voice. Tonic chords (i.e. the first chord of all sequences and the last chord of congruent sequences) were presented in root position, first, and second inversions. Examples of the chord progressions are provided in Fig. 1 (left panel).

Because the data of interest were participants' imitations of the fifth chord in each sequence, the time at which this chord was presented was locked to the video onset. Specifically, every video was individually examined and edited in such a way that the fifth chord occurred precisely 11 s after video onset. This manipulation was performed by decomposing each video into its constituent frames (of which there were 30 per second), and then ensuring that the first frame showing the hand producing the fifth chord was presented 11 s following video onset (movies were edited using the software iMovie HD 6.0.3, Apple Computer, Inc.). With the exception of the first chord duration (which was extended or shortened so that the fifth chord occurred precisely 11 s after video onset), videos depicted real motion.[2]

---

[1] Twenty pianists were asked to watch the videos and to judge whether the last chord was congruent or incongruent with the rest of the chord sequence. The pianists did so by pressing one of two buttons: one for "correct sequences" and the other for "incorrect sequences". Judgments confirmed that our classification was valid. Accuracy was high for Goal Incongruent sequences (judged to be incorrect, mean = 90.65%, SE = 2.95), Manner Incongruent sequences (judged to be incorrect, mean = 79.90%, SE = 4.95) and congruent sequences (judged to be correct, mean = 89.95%, SE = 2.52).

[2] An anonymous reviewer raised the point of whether preparatory (hand or finger) movements made before initiating an incongruent sequence might have been informative about how the sequence would end. Although we cannot exclude this possibility, we can note that the pianist who performed the sequences was highly familiar with them and all performances appeared fluent.

**Fig. 1.** Examples of musical scores of the stimulus sequences (left) and visual examples of sequences' last chord from the sequences (right). The figure shows a right hand performing (A) a congruent chord, (B) a Goal Incongruent chord and (C) a Manner Incongruent chord. Chord Sequence A consists of the following five chords (from left to right): tonic; tonic; subdominant; dominant; tonic. Chord Sequence B consists of: tonic; subdominant; supertonic; dominant; Neapolitan. Chord Sequence C consists of: tonic; subdominant; tonic; dominant; tonic. Numbers indicate which fingers are used to play each chord. Videos in all conditions were presented without sound.

### 2.1.3. Procedure

Participants were asked to watch and simultaneously imitate the video chord sequences, which were presented on a computer monitor placed on a MIDI piano (Yamaha Clavinova CLP150). Participants were instructed to imitate the key presses and fingerings as fast as possible with the right hand. Stimulus presentation and response registration were controlled by Presentation Software (Version 14.2, Neurobehavioral Systems, Inc.) while the performed fingering was recorded with a video camera (Sony, Digitale HD-VideoKamera, HDR-HC9E) placed above the piano, giving an aerial view of the participant's hand.

Each trial consisted of a visual fixation cross presented on the computer monitor for 500 ms followed by the visual presentation of one chord sequence video with 13 s duration. Each presentation began with a frozen frame lasting 2 s followed by 1 s of video showing the stationary model hand about to press the three keys associated with the first chord. This gave the participant enough time to match the initial position of his or her own hand with the position of model hand in the video. After that, the model hand performed a progression of five chords, with each chord lasting approximately 2 s. When a trial was over, participants could initiate the next trial by pressing the leftmost key of the piano keyboard with their left hand. The order of stimulus presentation was randomized individually for each participant. All participants saw all 120 videos (40 Congruent; 40 Goal Incongruent; 40 Manner Incongruent).

A MIDI interface connected the piano keyboard used for imitation and the program controlling the experiment. This device converted the MIDI key values received from the piano keyboard into a serial signal that was compatible with Presentation Software. This permitted us to compute time and codes of the pressed keys in relation to event timing in the video.

### 2.1.4. Data analysis

Errors and response times for imitation of only the last chord of each trial were analyzed. Error rates were calculated with respect to the *goal* (keys pressed) or the *manner* (fingers used). The fingering performed by each participant was examined visually (off-line) by inspecting the video camera recordings and comparing them with the fingering in the stimulus videos. If the last chord was correctly imitated in terms of the keys pressed, but incorrectly in terms of the fingering employed, then we counted a manner error. Conversely, if it was correctly imitated in terms of the fingering, but incorrectly in terms of the

goal, then we counted a goal error. Goal errors and manner errors were counted only if the previous chord (i.e. the second last chord) had been correctly imitated in terms of both goal and manner. Trials in which keystrokes within the last two chords of the sequence were not synchronous (i.e., when more than 150 ms intervened between the first and the last keystroke of either chord) were excluded from analysis (cf. Drost et al., 2005).

Response times (RT) were measured by calculating the time elapsed between the presentation of the last chord and the participant's execution of the same chord (i.e. mean of the three keystroke times composing the chord) in correct trials. Response times exceeding 3000 ms were not analyzed (cf. Drost et al., 2005). Data from two participants were excluded due to an insufficient number of correct trials (i.e. >66% missing).

## 2.2. Results and discussion

### 2.2.1. Errors

Fig. 2 (panel A) displays the mean number of errors — summed across goal errors and manner errors — produced in the three conditions (Congruent, Goal Incongruent, Manner Incongruent). A repeated measures ANOVA on these data yielded a main effect of Condition, $F(2, 24) = 17.83$, $(P < .001)$. The first of two planned orthogonal contrasts revealed that more errors were produced for imitation of incongruent chords (Goal Incongruent and Manner Incongruent combined) than for congruent chords, $F(1, 12) = 49.44$, $(P < .001)$. A second contrast indicated that participants performed more errors in the Manner Incongruent condition than in the Goal Incongruent condition, $F(1, 12) = 5.06$, $(P < .05)$.

A second analysis was conducted to examine the likelihood of producing different types of errors across conditions. Fig. 2 (Panel B) shows the distribution of errors (goal errors vs. manner errors) expressed as a percentage of total errors in each condition. Here it can be seen that, although manner errors were overall more frequent, relatively more goal errors were committed in the Congruent condition than in the two Incongruent conditions. The significance of this pattern of results was confirmed in a 3 × 2 chi-square test assessing the effects of the three experimental conditions (Congruent, Goal Incongruent, Manner Incongruent) on the two specific types of errors committed (goal errors, manner errors). This test reached
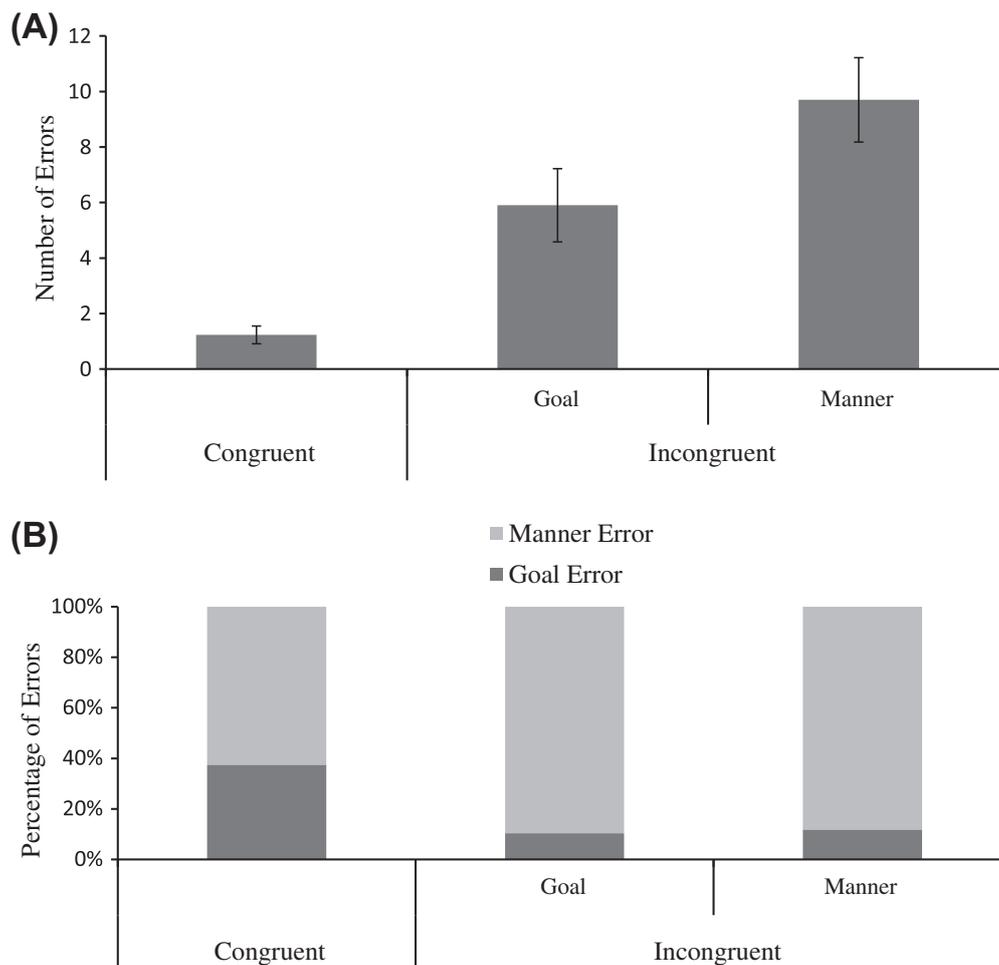


**Fig. 2.** Panel A shows mean number of errors for each condition. Error bars are 1 standard error of the mean. Panel B shows the distribution of errors (goal errors or manner errors) expressed as percentages in each condition.

significance: $\chi^2 (2) = 9.00$ ($P < .05$), indicating that different conditions were associated with different proportions of goal and manner errors. Specifically, goal and manner errors were more equally distributed when imitating congruent chords, while manner errors were more prevalent during the imitation of incongruent chords. An additional $2 \times 2$ chi-square test, with data representing the two types of errors in only the two Incongruent conditions, yielded a non-significant result, $\chi^2 (1) = .09$ ($P > .05$). This indicates that the relative percentage of manner and goal errors was constant across Incongruent conditions: more manner errors than goal errors were committed irrespectively of whether the video displayed a Manner Incongruent or Goal Incongruent final chord.

### 2.2.2. Response times

Fig. 3 shows mean RT for correctly produced target chords in each condition. A repeated measures ANOVA on these data yielded a main effect of Condition, $F (2, 20) = 16.59$, ($P < .001$). Planned contrasts revealed that RTs were faster in the Congruent condition than in both Incongruent conditions combined, $F (1, 10) = 98.03$ ($P < .001$), while RTs did not differ significantly between the two Incongruent conditions, $F(1, 10) = 1.53$ ($P > .05$). Thus, imitation was faster when target chords were congruent than when they were incongruent with the preceding harmonic context.

In order to investigate individual differences in the prioritization of high level (goal) and low level (manner) action representations, we computed the difference between mean RT for Goal Incongruent trials and Manner Incongruent trials for each participant. For this "Goal-Manner" index, positive difference scores indicate that imitation of Goal Incongruent trials took longer than Manner Incongruent trials, while negative values indicate the opposite. We assumed that a positive value indicates that action prediction is disturbed to a greater degree in the Goal Incongruent condition, while a negative value indicates that the Manner Incongruent condition is more disturbing. We ran a correlation analysis to examine the relationship between this Goal-Manner index and participants' years of piano training, which was quantified by subtracting the age at which piano studies commenced from each participant's age at the time of the experiment. (One participant's data were excluded because his Goal-Manner index was greater than two times the SD of the residuals.) This analysis revealed a significant positive correlation ($r(8) = 0.687$, $P < .05$), indicating that—as can be seen in Fig. 4—the size of the RT difference for imitation of goal-violated and manner-violated trials increased with the amount of musical training. Thus, experienced pianists prioritized goal over manner while less experienced pianists prioritized manner over goal.

Our main finding—that imitation was faster for congruent than incongruent chords—is consistent with the notion that expectancies build up over the course of each observed (silent) movement sequence, resulting in relatively fast responses to strongly expected endings and slower responses to less expected endings. There is, however, an outstanding issue that needs to be resolved before this interpretation is warranted. This issue relates to the distance that the pianist's hand was required to travel between the penultimate and ultimate chords in the different experimental conditions. Specifically, because chord sequences in the Goal Incongruent condition ended with a Neapolitan chord, the trajectory that the hand was required to perform when moving from the fourth chord to the last chord was longer in this condition than in the Congruent and Manner Incongruent conditions. We quantified this difference by computing the spatial distance between the fourth chord and the last chord of each sequence in terms of the number of semitones separating the piano keys associated with the two chords. Mean spatial distance was identical in congruent (mean = 2.00, SD = .00) and manner-violated sequences (mean = 2.00, SD = .00), while goal-violated sequences had higher and more variable spatial distance (mean = 4.95, SD = 2.73). Because of this difference, one could argue that the stimuli belonging to the Goal Incongruent condition would require more time in order to be imitated independently of the preceding context. We addressed this potential confound in a second experiment that aimed to disentangle these two different components (i.e. context and spatial distance).
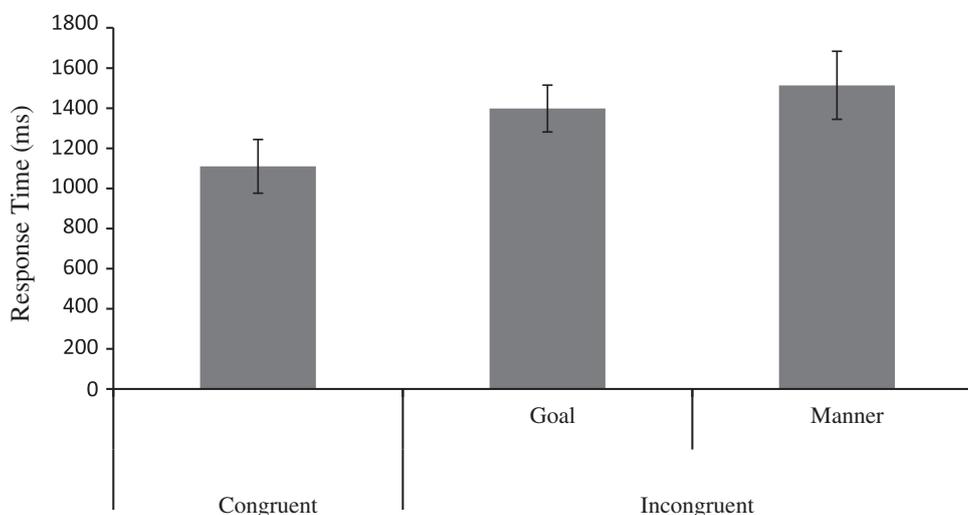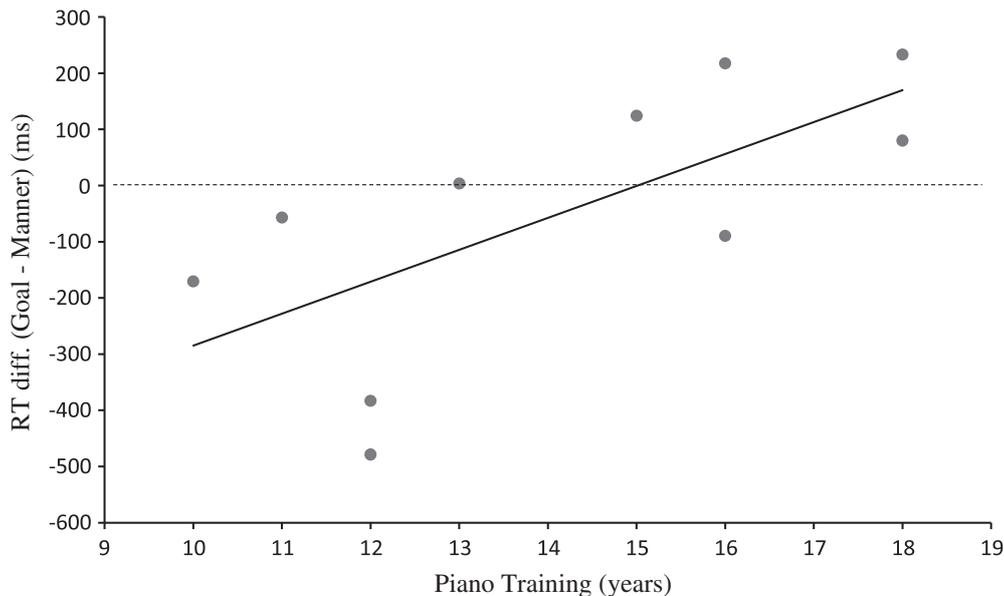


**Fig. 3.** Mean Response Time (RT) during imitation of Congruent, Goal Incongruent, and Manner Incongruent chords. Error bars are 1 standard error of the mean.

**Fig. 4.** Scatter plot showing the relationship between years of piano training (on the *x*-axis) and the "Goal-Manner" index (on the *y*-axis, where positive values indicate that imitation of Goal Incongruent chords took longer than Manner Incongruent chords, while negative values indicate the opposite). Each data point represents an individual pianist. The diagonal indicates the line of best fit. The dotted line indicates zero (i.e. equal RT between the two Incongruent conditions).

## 3. Experiment 2

Experiment 2 was conducted to confirm that the congruency effect found in Experiment 1 was attributable to context based expectancies (as we hypothesized) rather than greater spatial distance between penultimate and ultimate chords in the Goal Incongruent than the Congruent condition (a potential confound). To this end, the same pianists who took part to Experiment 1 were invited back to the lab. They were presented with stimuli consisting of only the last two chords of the sequences from the Congruent condition and the Goal Incongruent condition in Experiment 1. This manipulation reduced the musical context and was therefore expected to dilute or eliminate the contextual congruence effect found in Experiment 1. We were thus able to estimate the magnitude of the effect of spatial distance on RT. We expected that this would account for only a small portion of the large RT differences between Congruent and Goal Incongruent conditions observed in Experiment 1.

### 3.1. Methods

Ten of the pianists who took part in Experiment 1 participated in Experiment 2. These individuals returned 9–11 weeks after participating in Experiment 1, and were naïve with regard to the purpose of this second experiment.

Stimuli were excerpts from the videos presented in Experiment 1, showing the performance of only the last two chords of the sequences belonging to the Congruent condition and the Goal Incongruent condition described above. The Manner Incongruent condition was not included in Experiment 2 because spatial distance in this condition was the same as in the Congruent condition. Each trial began with presentation of a fixation-cross (500 ms). After that, a frozen frame presenting the hand about to press the fourth chord was shown (2 s). Finally, the video displayed the performance of the two sequential chords, with each chord lasting 2 s.

Experiment 2 was identical to Experiment 1 in terms of all other aspects concerning task, procedure and data analysis.

### 3.2. Results and discussion

We first report an analysis of data from Experiment 2 alone and then compare the results of Experiments 1 and 2. The number of errors committed (summed across key errors and manner errors) for imitation of the final chord was greater in the 'Goal Incongruent' condition than in the 'Congruent' condition, $F (1, 9) = 6.55$ ($P < .05$). Since target chords were no longer preceded by context-forming sequences, this difference is most likely due to the relatively high complexity of the Neapolitan chords—which required combinations of black and white keys to be pressed—in the 'Goal Incongruent' condition.

The distribution of key and manner errors was consistent with that observed in Experiment 1: more manner errors (82.1%) than goal errors (17.9%) were observed overall. However, the distribution of goal errors and manner errors within each condition was more even within 'Congruent' than within 'Goal incongruent' sequences. The significance of this pattern
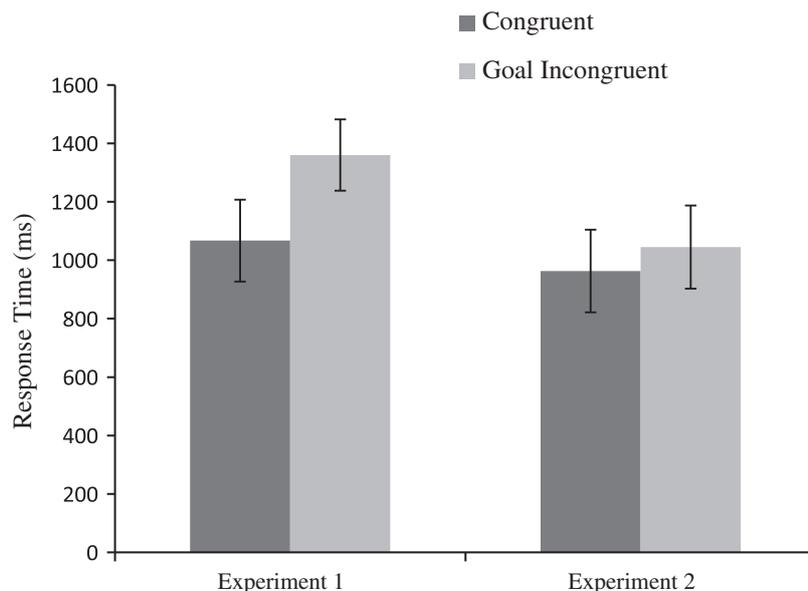
**Fig. 5.** Mean Response Time (RT) during imitation of Congruent chords (dark columns) or Goal Incongruent chords (light columns) in Experiment 1 (left columns) and Experiment 2 (right columns). Error bars are 1 standard error of the mean.

of results was confirmed in a 2 × 2 chi-square analysis with data representing the two types of errors (goal errors, manner errors) within the two conditions ('Congruent', 'Goal Incongruent'), $\chi^2 (1) = 4.94$ ($P < .05$).

The analysis of RT revealed that, as can be seen in Fig. 5 (right columns), participants were faster at imitating final chords in the 'Congruent' condition (where spatial distance between keys for the two chords was small) than in the 'Goal Incongruent' condition (where spatial distance was large), $F (1, 9) = 8.36$ ($P < .05$). This suggests that the greater spatial distance may have indeed contributed to the prolongation of RTs in the Goal Incongruent condition, independently of the preceding context, in Experiment 1.

Therefore, we conducted an inter-experiment analysis to test the degree to which this spatial distance effect accounts for the RT differences between Congruent and Goal Incongruent conditions in Experiment 1. It seemed unlikely to be fully accountable since (a) the effect observed in Experiment 2 (82 ms) was numerically much smaller than in Experiment 1 (293 ms) and (b) RTs were consistent across experiments in the Congruent condition but not in the Goal Incongruent condition (see Fig. 5). Nevertheless, to address this issue, we performed an item-analysis that examined mean RT (averaged across all participants) for each individual target chord (i.e. last chord) of the stimulus sequences from the Congruent and Goal Incongruent conditions in the two experiments.

A 2 × 2 repeated measures ANOVA with variables Experiment (Experiment 1 and Experiment 2) and Congruency (Congruent or Goal Incongruent) yielded a significant main effect of Experiment, $F (1, 39) = 80.70$ ($P < .001$), a significant main effect of Congruency, $F (1, 39) = 36.30$ ($P < .001$), and a significant Experiment × Congruency interaction effect, $F (1, 39) = 20.59$ ($P < .001$). These results were replicated in a second item-analysis where the spatial distance (i.e., the number of semitones separating penultimate and final chords in each sequence) was included as a covariate: main effect of Experiment $F (1, 38) = 42.87$ ($P < .001$), Congruency $F (1, 38) = 10.25$ ($P < .01$), Experiment × Congruency interaction, $F (1, 38) = 7.84$ ($P < .01$). Moreover, the Experiment × Congruency × spatial distance (covariate) interaction was not significant: $F (1, 38) = .063$ ($P > .05$).

These results rule out the possibility that the effect found in Experiment 1 (i.e. the difference between RT in Congruent and Goal Incongruent trials) was solely due to the higher complexity of the stimuli (measured in terms of spatial distance). Rather, the results show that this effect is attributable to the preceding context, which induced strong expectations in Experiment 1 but not in Experiment 2.[3] Nevertheless, it is worth pointing out that the higher complexity of the 'Goal Incongruent' stimuli probably introduced a delay in the RT. This was not the only source of variability, however, as confirmed by the significant Experiment × Congruency interaction in both analyses.

In a final analysis, we explored whether the strength of expectations induced by the preceding context varies as a function of amount of piano playing experience. For each participant, we calculated a new index that we termed "contextual congruency". This variable was computed by first subtracting the mean RT in the Congruent condition from the mean RT in the Goal

---

[3] The reader might wonder whether the different percentage of congruent trials (66.66% in Experiment 1 and 50% in Experiment 2) or the temporal order of the two experiments (i.e. Experiment 1 was performed before Experiment 2) may have influenced RTs or task strategies. We can exclude that this was the case on the basis of a recently completed EEG experiment with only two conditions (Congruent and Goal-Incongruent) and two contexts (five-chord and two-chord sequences, presented in separate blocks whose order was counterbalanced). The results of this new experiment replicate the RT effect described here, demonstrating that neither the percentage of correct trials nor the temporal order of the two experiments caused it.
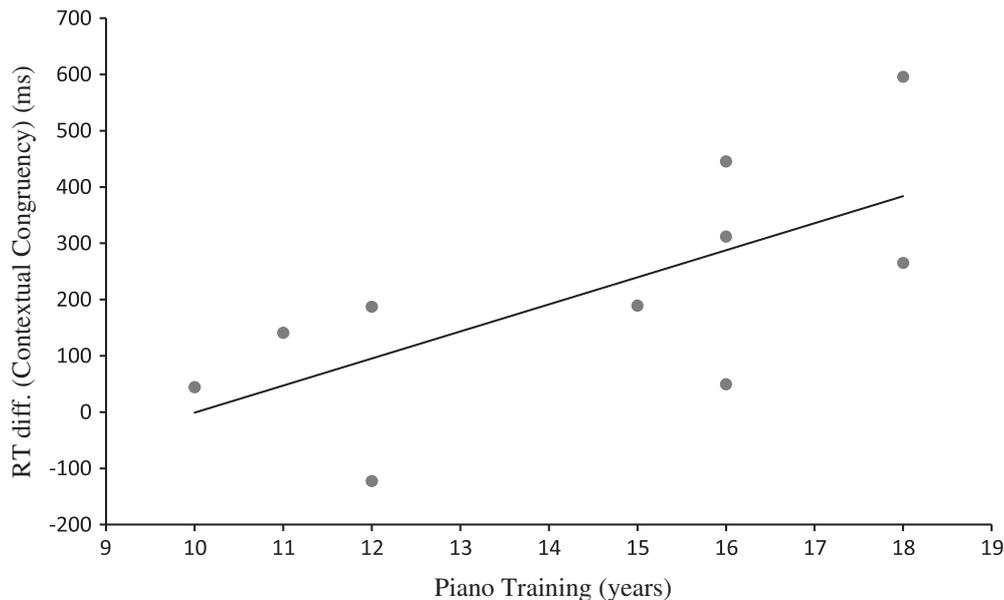
**Fig. 6.** Scatter plot showing the relationship between years of piano training (on the x-axis) and the "Contextual Congruency" index (on the y-axis, where higher values indicate greater context-dependent delay in mean RT). Each data point represents an individual pianist. The diagonal indicates the line of best fit.

Incongruent condition in each experiment, and then subtracting the resulting Experiment 2 difference score from the Experiment 1 difference score (i.e. (Exp1Goal-Exp1Congr) – (Exp2Goal-Exp2Congr)). The larger the value of this index, the higher the context-dependent delay in mean RT. We ran a correlation analysis to examine the relationship between this "contextual congruency" index and participants' years of piano training. This revealed a significant positive interaction ($r(8) = 0.675$, $P < .05$) (see Fig. 6), indicating that the strength of goal expectations and, critically, their dependence on the previous context increased with musical experience.

## 4. General discussion

The present study provides behavioral evidence in support of domain-general syntactic structures regulating the progression of motor acts associated with producing music. Specifically, our results suggest that when musical chord sequences—which have been shown to lead to harmonic expectations in the auditory domain (e.g., Koelsch et al., 2005; Maess et al., 2001) — are converted into musical action sequences, these action sequences lead to the generation of equivalent predictions in the motor domain in skilled musicians. We interpret our results in terms of four main conclusions concerning the nature of these motor predictions: (1) they depend on context, (2) they do not require auditory feedback, (3) they are derived at a high representational level in the motor control system, and (4) they vary as a function of musical experience. Each of these features will be discussed in turn.

First, the results of our two experiments, taken together, demonstrate that imitation of a motor act (i.e. the movement necessary to produce a chord) is faster when it is embedded in a well-learnt syntactic structure, i.e., into a congruent context. This finding is noteworthy in that it sheds light on the similar nature of musical perception and musical production. On the one hand, performing a musical piece involves manipulating a number of acts (i.e. movements) into an integrated motor structure (i.e. action) wherein events are ordered correctly. This process has been referred to as "sequencing" in the motor control literature (for a review see Rosenbaum, 1991), where its has been assumed to play an important role in motor skill learning (Schmidt, 1975; Willingham, 1998). On the other hand, perceiving a sequence of acts seems to be sufficient to trigger an internal representation of the motor structure, and to derive predictions about the goal of the observed action. Therefore, it seems reasonable to conclude that there is such an entity as an action context, which is used to generate predictions about up-coming motor acts when controlling our body, as well as when watching someone else performing an action we are familiar with (Schütz-Bosbach & Prinz, 2007a, 2007b).

This interpretation is consistent with evidence showing that stronger motor activations are elicited in the brain while listening to a succession of musical sounds that one can produce than when listening to a different combination of the same sounds (Lahav et al., 2007). Moreover, it is consistent with the notion of a mechanism shared by perception and motor control – such as the mirror neuron system (for a review see Rizzolatti & Sinigaglia, 2010) – which has been proposed to facilitate the process of inferring others' intentions through observation of their movements in a context-sensitive manner (Kilner, Friston, & Frith, 2007). This mechanism has been proposed to play a key role in music perception (Overy & Molnar-Szakacs, 2009), where it has been shown that visual observation of musical performance can significantly improve the communication of the performer's expressive intentions (Chapados & Levitin, 2008; Davidson, 1993; Vines, Krumhansl,

Wanderley, & Levitin, 2006). The present study suggests that a similar inference mechanism may be involved when musicians watch one another performing: visual cues may assist a performer to predict what another is going to do during joint musical interactions.

Our second main finding is that music-related motor predictions operate in the absence of auditory feedback. Although auditory perception obviously did not drive the observed effects, as the task was to imitate silent videos on a muted keyboard, auditory imagery may have played a role. A number of fMRI studies have shown that mute piano performance (Bangert et al., 2006), as well as visual observation of mute piano performance (Hasegawa et al., 2004; Haslinger et al., 2005), elicits stronger activation of auditory brain regions in musicians than in non-musicians (for a review, see Hubbard, 2010). Therefore, we cannot ascertain whether the expectations triggered by the musical context were fully derived within the motor system, or whether these expectations were the result of multimodal sensorimotor integration involving specific forms of auditory imagery. We are currently addressing this question using brain imaging methods.

Our third finding is that musical motor predictions appear to be derived at a high representational level in the control system, i.e., at the level of the movement goal. We found that, compared to manner-violated chords (with unconventional fingering), goal-violated chords elicited significantly less errors. Furthermore, participants produced significantly more fingering errors than key errors, irrespective of the condition. Taken together, these results show that, when a relatively unexpected chord was observed, imitation was more accurate with respect to the goal than to the manner independently of the type of violation. This finding has two potential explanations. On the one hand, it could suggest that the goal of an action has priority in the action control hierarchy over the specific movements necessary for achieving it (Grafton & Hamilton, 2007). This interpretation would be consistent with previous findings showing that, during imitation, aspects of an action such as the goal and the manner are hierarchically ordered, and that highest priority is given to the intended goal of a certain action, rather than the means used to achieve it (Bekkering, Brass, Woschina, & Jacobs, 2005; Bekkering, Wohlschlaeger, & Gattis, 2000; Wohlschläger, Gattis, & Bekkering, 2003; see also Grafton & Hamilton, 2007). On the other hand, goal prioritization could also be attributed to the lights on the piano keys in our stimulus videos: these may have directed participants' attention to the goal rather than the manner due to the higher salience of the lights relative to the pianist's fingers in the displays. We reject this latter interpretation for two reasons.

First, participants were given clear instructions to imitate both the key presses and the fingering. If, despite the instructions, participants' attention had been more focused on the goal than on the manner, this would still speak for a different priority of the two components (i.e. imitation of the keys and imitation of the fingering) of the task. The second reason is that attention to the lights would not explain the finding that pianists with longer training history were relatively more affected by goal violations than by manner violations in terms of response times. These individual differences are unlikely to be due to the salience of the lights because expert pianists do not usually play with lighted keys and, in fact, they visually fixate less on the keyboard and the hands with increasing experience. Hence, we favor the first interpretation, which implies that pianists' action prediction was driven by a high-level mechanism that inferred the goal of the observed actions relatively independently of the specific movement used to achieve it. We therefore assume that the context we introduced provided stronger expectations with respect to high-level perceptual analysis (i.e. goal of actions) rather than the specific movements needed to perform a particular action.

Finally, we found two pieces of evidence that the strength of musical motor predictions varied as a function of musical training. First, as mentioned in the preceding paragraph, we showed that pianists with longer training history were relatively more affected by goal violations than by manner violations. On average, highly trained pianists took more time in order to successfully imitate a Goal Incongruent target chord than a Manner Incongruent chord, while the opposite held for the less experienced pianists. This result implies that musicians with more expertise had stronger expectations regarding the goal of the observed actions and therefore were more perturbed by Goal Incongruent target chords. A second piece of evidence converges on a similar conclusion. By means of a comparison between the two experiments, we showed that the strength of goal predictions (as indexed by the size of the "contextual congruency" effect) and their dependence on preceding context, was positively correlated with the number of years of piano training. This suggests that the effect of context becomes stronger with increasing musical experience, and that the source of goal predictions may lie in pianists' specific motor experiences.

Although musical goals are typically auditory, our experiment was performed in the total absence of sound and in an uncommon setting, since pianists usually do not watch others' hands while performing. Therefore, the reported effects tell us more about the motor representation of musical actions in trained musicians (rather than the motor representation of sounds) and how such representations might develop with experience. Our results imply that, through training, motor experts develop an action prediction mechanism characterized by increasing predilection for goal-predictions and, perhaps, increasing flexibility with regard to the way in which a specific goal is achieved (see Palmer & Meyer, 2000). This distinctive aspect might be due to longer training history or to the fact that these musicians started training at a young age and thus benefited from especially high brain plasticity (Pantev et al., 1998; Schlaug, Jaencke, Huang, Staiger, & Steinmetz, 1995). The hypothesis of increasing goal prioritization and manner flexibility, which deserves to be further investigated, could expand and complement previous research on motor expertise (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005; Cross, Hamilton, & Grafton, 2006; Haslinger et al., 2005; Vogt et al., 2007) and its relationship to perceptual anticipation of actions performed by others (Aglioti, Cesari, Romani, & Urgesi, 2008; Farrow & Abernethy, 2003).

In conclusion, the present results are consistent with the hypothesis that – in musicians – a syntactic violation of a musical structure is also a violation of the motor structure necessary to produce it. This suggests that, throughout one's musical experience, rules determining the position of one item within a sequential structure are internalized as a form of "embodied

harmony" that is concerned with high level aspects of actions (e.g. goals rather than the mere reproduction of music related movements) in a context-dependent manner. According to this view, a long period of musical training results in equivalence between a "musical grammar", i.e. the rules determining the correct position of a chord within a sequence, and a "grammar of action". Please note that the latter term does not necessarily exclude the involvement of movement effects (such as those provided by auditory imagery). However we argue that physical sounds are not a prerequisite for the grammatical rules to be used in generating predictions. The generative character of this grammar, i.e. the property by which it can be used to predict which combinations of movements will form a musically correct (e.g. harmonic) phrase, might become particularly important for composition and improvisation (Baily, 1977, 1985; see also Johnson-Laird, 2002). We speculate that this might be one of the resources that expert players rely on when deriving predictions during music perception. A similar resource might be recruited during joint performances, when a player is required to monitor others' performance and integrate this information with one's own performance. This resource may be especially useful during musical interactions that require players to infer others' intentions under conditions of uncertainty, such as joint musical improvisation.

## Acknowledgments

## References

Aglioti, S. M., Cesari, P., Romani, M., & Urgesi, C. (2008). Action anticipation and motor resonance in elite basketball players. *Nature Neuroscience, 11*(9), 1109–1116.

Baily, J. (1977). Movement patterns in playing the Herati durar. In J. Blacking (Ed.), *The anthropology of the body* (pp. 275–330). New York: Academic Press.

Baily, J. (1985). Music structure and human movements. In P. Howell, I. Cross, & R. West (Eds.), *Musical structure and cognition* (pp. 237–258). London: Academic Press.

Bangert, M., & Altenmüller, E. (2003). Mapping perception to action in piano practice: A longitudinal DC-EEG study. *BMC Neuroscience, 4*, 26.

Bangert, M., Peschel, T., Schlaug, G., Rotte, M., Drescher, D., Hinrichs, H., et al (2006). Shared networks for auditory and motor processing in professional pianists: Evidence from fMRI conjunction. *NeuroImage, 30*(3), 917–926.

Bekkering, H., Brass, M., Woschina, S., & Jacobs, A. (2005). Goal-directed imitation in patients with Ideomotor Apraxia. *Cognitive Neuropsychology, 22*(3), 419–432.

Bekkering, H., Wohlschlaeger, A., & Gattis, M. (2000). Imitation of Gestures in Children is Goal-directed. *The Quarterly Journal of Experimental Psychology, 53*(1), 153–164.

Bernstein, N. A. (1996). On dexterity and its development. In M. L. Latash & M. T. Turvey (Eds.), *Dexterity and its development* (pp. 9–244). Mahwah, NJ: Lawrence Erlbaum Associates.

Bharucha, J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13*(1), 63–102.

Buccino, G., Vogt, S., Ritzl, A., Fink, G. R., Zilles, K., Freund, H., et al (2004). Neural circuits underlying imitation learning of hand actions: An event-related fMRI study. *Neuron, 42*(2), 323–334.

Bunnin, N., & Jiyuan, Y. (2004). *The blackwell dictionary of western philosophy*. Malden, MA: Blackwell Pub.

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex, 15*(8), 1243–1249.

Castellano, M. A., Bharucha, J. J., & Krumhansl, C. L. (1984). Tonal hierarchies in the music of North India. *Journal of Experimental Psychology: General, 113*, 394–412.

Chapados, C., & Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: Physiological correlates. *Cognition, 108*(3), 639–651.

Clark, A. (1999). An embodied cognitive science? *Trends in Cognitive Sciences, 3*(9), 345–351.

Clerget, E., Winderickx, A., Fadiga, L., & Olivier, E. (2009). Role of Broca's area in encoding sequential human actions: A virtual lesion study. *Neuroreport, 20*(16), 1496–1499.

Cross, E. S., Hamilton, A. F., & Grafton, S. T. (2006). Building a motor simulation de novo: Observation of dance by dancers. *NeuroImage, 31*(3), 1257–1267.

D'Ausilio, A., Altenmüller, E., Olivetti Belardinelli, M., & Lotze, M. (2006). Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *The European Journal of Neuroscience, 24*(3), 955–958.

Davidson, J. W. (1993). Visual perception of performance manner in the movements of solo musicians. *Psychology of Music, 21*(2), 103–113.

Drost, U. C., Rieger, M., Brass, M., Gunter, T. C., & Prinz, W. (2005). Action-effect coupling in pianists. *Psychological Research, 69*(4), 233–241.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development.* Cambridge, MA: MIT Press.

Fadiga, L., Craighero, L., & D'Ausilio, A. (2009). Broca's area in language, action, and music. *Annals of the New York Academy of Sciences, 1169*, 448–458.

Farrow, D., & Abernethy, B. (2003). Do expertise and the degree of perception – Action coupling affect natural anticipatory performance? *Perception, 32*(9), 1127–1139.

Fazio, P., Cantagallo, A., Craighero, L., D'Ausilio, A., Roy, A. C., Pozzo, T., et al (2009). Encoding of human action in Broca's area. *Brain, 132*(Pt 7), 1980–1988.

Friederici, A. D., Rüschemeyer, S., Hahne, A., & Fiebach, C. J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex, 13*(2), 170–177.

Goodnow, J. J., & Levine, R. A. (1973). "The grammar of action": In children's sequence and syntax. *Cognitive Psychology, 4*(1), 82–98.

Grafton, S. T., & Hamilton, A. F. (2007). Evidence for a distributed hierarchy of action representation in the brain. *Human Movement Science, 26*(4), 590–616.

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Brain Research. Cognitive Brain Research, 13*(3), 339–356.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences, 11*(11), 466–472.

Hasegawa, T., Matsuki, K., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., et al (2004). Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Brain Research. Cognitive Brain Research, 20*(3), 510–518.

Haslinger, B., Erhard, P., Altenmüller, E., Schroeder, U., Boecker, H., Ceballos-Baumann, A. O., et al (2005). Transmodal sensorimotor networks during action observation in professional pianists. *Journal of Cognitive Neuroscience, 17*(2), 282–293.

Haueisen, J., & Knösche, T. R. (2001). Involuntary motor activity in pianists evoked by music perception. *Journal of Cognitive Neuroscience, 13*(6), 786–792.

Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin, 136*(2), 302–329.

Iacoboni, M., Molnar-Szakacs, I., Gallese, V., Buccino, G., Mazziotta, J. C., Rizzolatti, G., et al (2005). Grasping the intentions of others with one's own mirror neuron system. *PLoS Biology, 3*(3), e79.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., Rizzolatti, G., et al (1999). Cortical mechanisms of human imitation. *Science, 286*(5449), 2526–2528.

Janata, P., & Grafton, S. T. (2003). Swinging in the brain: Shared neural substrates for behaviors related to sequencing and music. *Nature Neuroscience, 6*, 682–687.

Jentschke, S., & Koelsch, S. (2009). Musical training modulates the development of syntax processing in children. *NeuroImage, 47*(2), 735–744.

Johnson-Laird, P. N. (2002). How Jazz musicians improvise. *Music Perception, 19*(3), 415–442.

Keller, P. E., & Koch, I. (2008). Action planning in sequential skills: Relations to music performance. *Quarterly Journal of Experimental Psychology, 61*(2), 275–291.

Kilner, J. M., Friston, K. J., & Frith, C. D. (2007). The mirror-neuron system: A Bayesian perspective. *Neuroreport, 18*(6), 619–623.

Kilner, J. M., Vargas, C., Duval, S., Blakemore, S., & Sirigu, A. (2004). Motor activation prior to observation of a predicted movement. *Nature Neuroscience, 7*(12), 1299–1301.

Koelsch, S., Fritz, T., Schulze, K., Alsop, D., & Schlaug, G. (2005). Adults and children processing music: An fMRI study. *NeuroImage, 25*(4), 1068–1076.

Koelsch, S., Schmidt, B., & Kansok, J. (2002). Effects of musical expertise on the early right anterior negativity: An event-related brain potential study. *Psychophysiology, 39*(5), 657–663.

Krumhansl, C. (2004). The cognition of tonality – As we know it today. *Journal of New Music Research, 33*(3), 253–268.

Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review, 89*(4), 334–368.

Lahav, A., Saltzman, E., & Schlaug, G. (2007). Action representation of sound: Audiomotor recognition network while listening to newly acquired actions. *Journal of Neuroscience, 27*(2), 308–314.

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

Loui, P., & Wessel, D. L. (2007). Harmonic expectation and affect in Western music: Effects of attention and training. *Perception and Psychophysics, 69*(7), 1084–1092.

Loui, P., Wessel, D. L., & Hudson Kam, C. L. (2010). Humans rapidly learn grammatical structure in a new musical scale. *Music Perception, 27*(5), 377–388.

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*(5), 540–545.

Markman, A., & Dietrich, E. (2000). Extending the classical view of representation. *Trends in Cognitive Sciences, 4*(12), 470–475.

Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Opacic, T., Stevens, C., & Tillmann, B. (2009). Unspoken knowledge: Implicit learning of structured human dance movement. *Journal of Experimental Psychology. Learning, Memory and Cognition, 35*(6), 1570–1577.

Overy, K., & Molnar-Szakacs, I. (2009). Being together in time: Musical experience and the mirror neuron system. *Music Perception, 26*(5), 489–504.

Palmer, C., & Meyer, R. K. (2000). Conceptual and motor learning in music performance. *Psychological Science, 11*(1), 63–68.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., Hoke, M., et al (1998). Increased auditory cortical representation in musicians. *Nature, 392*(6678), 811–814.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674–681.

Patel, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Patel, A. D., Iversen, J. R., Wassenaar, M., & Hagoort, P. (2008). Musical syntactic processing in agrammatic Broca's aphasia. *Aphasiology, 22*(7), 776–789.

Pulvermüller, F., & Fadiga, L. (2010). Active perception: Sensorimotor circuits as a cortical basis for language. *Nature Reviews. Neuroscience, 11*(5), 351–360.

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*(2), 241–255.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews. Neuroscience, 11*(4), 264–274.

Rosenbaum, D. A. (1991). *Human motor control*. San Diego, CA: Academic Press.

Sammler, D., Koelsch, S., & Friederici, A. D. (in press). Are left fronto-temporal brain areas a prerequisite for normal music-syntactic processing? *Cortex*. doi: 10.1016/j.cortex.2010.04.007.

Schlaug, G., Jaencke, L., Huang, Y., Staiger, J. F., & Steinmetz, H. (1995). Increased corpus callosum size in musicians. *Neuropsychologia, 33*(8), 1047–1055.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review, 82*(4), 225–260.

Schoenberg, A. (1969). *Structural functions of harmony* (Revised ed.). New York: Norton.

Schütz-Bosbach, S., & Prinz, W. (2007a). Perceptual resonance: Action-induced modulation of perception. *Trends in Cognitive Sciences, 11*(8), 349–355.

Schütz-Bosbach, S., & Prinz, W. (2007b). Prospective coding in event representation. *Cognitive Processing, 8*(2), 93–102.

Tillmann, B., Janata, P., & Bharucha, J. J. (2003). Activation of the inferior frontal cortex in musical priming. *Cognitive Brain Research, 16*(2), 145–161.

Underwood, G. E., & Bright, J. E. H. (1996). Cognition with and without awareness. In G. Underwood (Ed.), *Implicit cognition* (pp. 1–40). New York: Oxford University Press.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition, 101*(1), 80–113.

Vogt, S., Buccino, G., Wohlschläger, A. M., Canessa, N., Shah, N. J., Zilles, K., et al (2007). Prefrontal involvement in imitation learning of hand actions: Effects of practice and expertise. *NeuroImage, 37*(4), 1371–1383.

Willingham, D. B. (1998). A neuropsychological theory of motor skill learning. *Psychological Review, 105*(3), 558–584.

Wohlschläger, A., Gattis, M., & Bekkering, H. (2003). Action generation and action perception in imitation: An instance of the ideomotor principle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 358*(1431), 501–515.