Seeing, listening, drawing: interferences between sensorimotor modalities in the use of a tablet musical interface

Olivier Perrotin, Christophe d'Alessandro

▶ To cite this version:

Olivier Perrotin, Christophe d'Alessandro. Seeing, listening, drawing: interferences between sensorimotor modalities in the use of a tablet musical interface. ACM Transactions on Applied Perception, 2016, 14 (2), pp.1 - 19. 10.1145/2990501 . hal-01672241

Seeing, Listening, Drawing: Interferences between Sensorimotor Modalities in the Use of a Tablet Musical Interface

OLIVIER PERROTIN and CHRISTOPHE D'ALESSANDRO, LIMSI, CNRS, Université Paris-Saclay

---

Audio, visual, and proprioceptive actions are involved when manipulating a graphic tablet musical interface. Previous works suggested a possible dominance of the visual over the auditory modality in this situation. The main goal of the present study is to examine the interferences between these modalities in visual, audio, and audio-visual target acquisition tasks. Experiments are based on a movement replication paradigm, where a subject controls a cursor on a screen or the pitch of a synthesized sound by changing the stylus position on a covered graphic tablet. The experiments consisted of the following tasks: (1) a target acquisition task that was aimed at a visual target (reaching a cue with the cursor displayed on a screen), an audio target (reaching a reference note by changing the pitch of the sound played in headsets), or an audio-visual target, and (2) the replication of the target acquisition movement in the opposite direction. In the return phase, visual and audio feedback were suppressed. Different gain factors perturbed the relationships among the stylus movements, visual cursor movements, and audio pitch movements. The deviations between acquisition and return movements were analyzed. The results showed that hand amplitudes varied in accordance with visual, audio, and audio-visual perturbed gains, showing a larger effect for the visual modality. This indicates that visual, audio, and audio-visual actions interfered with the motor modality and confirms the spatial representation of pitch reported in previous studies. In the audio-visual situation, vision dominated over audition, as the latter had no significant influence on motor movement. Consequently, visual feedback is helpful for musical targeting of pitch on a graphic tablet, at least during the learning phase of the instrument. This result is linked to the underlying spatial organization of pitch perception. Finally, this work brings a complementary approach to previous studies showing that audition may dominate over vision for other aspects of musical sound (e.g., timing, rhythm, and timbre).

CCS Concepts: • **General and reference → Measurement**; **Evaluation**; • **Human-centered computing → Displays and imagers**; **Pointing devices**; **Auditory feedback**; **Pointing**; **Gestural input**; *Laboratory experiments*; • **Applied computing →** *Sound and music computing*; Performing arts;

Additional Key Words and Phrases: Sensorimotor modalities, target acquisition, graphic tablet

ACM Reference Format:

Olivier Perrotin and Christophe d'Alessandro. 2016. Seeing, listening, drawing: Interferences between sensorimotor modalities in the use of a tablet musical interface. ACM Trans. Appl. Percept. 14, 2, Article 10 (October 2016), 19 pages.
DOI: http://dx.doi.org/10.1145/2990501

This work was supported by the Agence Nationale de la Recherche, under grant ANR-13-CORD-0011.
Authors' addresses: O. Perrotin and C. d'Alessandro, LIMSI, CNRS, Bâtiment 508, rue John Von Neumann, 91405 Orsay, France; emails: olivier.perrotin@limsi.fr, cda@limsi.fr.

ACM 1544-3558/2016/10-ART10 $15.00
DOI: http://dx.doi.org/10.1145/2990501

1. INTRODUCTION

1.1 Motivation of the Work

This research originated from a question raised by the practice of using a graphic tablet for pitch control in speech and music. The tablet musical interface is quite widely used in various research and musical groups today [Kessous 2004; d'Alessandro et al. 2006; Zbyszynski et al. 2007; d'Alessandro and Dutoit 2009] and will probably become a major musical interface in the future because of the high quality of musical control allowed. Chironomic control of pitch such as manual pitch control on a tablet has been studied in previous works by the authors [d'Alessandro et al. 2011, 2014] in which graphic tablets demonstrated a high level of accuracy and precision when used as an interface for voice fundamental frequency (pitch) control. The tablet was shown to be effective both for mimicking speech intonation and for singing with synthesized voices. In a musical application, visual cues printed on the tablet dramatically enhanced the instrument's usability. This raises an important issue concerning the relative weight and interactions of the three sensorimotor modalities involved in the task: seeing, listening, and drawing. This question is addressed in the present work with the help of a simplified audio-visual target acquisition paradigm, where the cursor position on a screen and/or the pitch of a synthesized vocal sound vary accordingly to the position of a stylus on a graphic tablet. Interferences between distal action effects (cursor position and/or pitch of vocal sound) and proximal action effects (stylus motion on a graphic tablet) are analyzed.

In d'Alessandro et al. [2011], 10 subjects were asked to mimic the (audio) intonation pattern of sentences both with their own voices and with the help of a stylus on a graphic tablet, using a system for real-time manual intonation modification. In this experiment, only the y-axis of the graphic tablet was used for pitch control, allowing participants to draw various two-dimensional (2D) patterns on the tablet surface yet create the same pitch (y-axis) contour. There were no visual cues on the tablet to help the intonation control. Similar intonation patterns were achieved using a variety of hand motions: circular, sinusoidal, steplike, vertical gestures, and so on. The task involved both rhythmic hand motion planning and melodic targeting. It seemed that subjects tended to imitate pitch contours rather than aiming at precise pitch targets. In conclusion, the authors stated that the performance levels reached through the graphic tablet drawing process and through vocal intonation imitation were comparable. A perceptual test showed that the pitch contours resulting from voice imitation and chironomic imitation were, in many cases, indistinguishable. This result suggested that intonation, both on its perceptual and motor production aspects, seems essentially independent of the modality used to reproduce it. Note that the role played by vision has not been explicitly studied, but it appeared to be of secondary importance in this study because no visual cues (except the tablet itself) were available—nor did they seem needed.

The situation differs when using the tablet for musical control: In this case, visual cues are clearly needed to improve performance. The acquisition of musical notes (accurate pitch targets, in contrast with pitch contours) was studied in a subsequent work [d'Alessandro et al. 2014] in which two groups of 20 and 28 subjects were asked to sing musical intervals and short melodies at various tempi, using chironomy (hand-controlled singing), mute chironomy (without audio feedback), or their own voices. The "mute chironomy" condition, that is, visual note acquisition without audio feedback, was introduced to observe the role of vision in pitch target acquisition. Accuracy (mean) and precision (standard deviation) of pitch target acquisition errors were measured. The results showed that the mute chironomy condition degraded pitch target acquisition (i.e., audio feedback does play a role in the task); nevertheless, the results were very precise and accurate (an average error of 9 cents of semitones), highlighting a strong dependency on the visual feedback.

For the sake of comparison between the two studies, the results of d'Alessandro et al. [2011] were reinterpreted as a condition without visual cues. Syllables constituted the basic melodic units, analogous to notes; Root Mean Square distance between the trial and the target pitch values defined pitch target acquisition errors. Accuracy and precision were computed as the mean and standard deviation of these errors, respectively. It appeared that accuracy and precision were better by a factor of 1.5 and 1.2, respectively, for the voice condition than for the blind chironomy condition but were not perceptually distinguishable for this speech intonation experiment. Note that the accuracy of speech intonation perception is much lower than of musical intonation perception; the former compares only contours, while accurate and stable notes are expected in the latter. Overall, these results indicated that a formal detailed study of the influence of vision on chironomic control of pitch was clearly needed. This is the question addressed in the present work.

1.2 Visual, Auditory, and Kinesthetic Interferences

Visual, auditory, and kinesthetic interferences have been studied along various lines. The earliest works addressed the question of visual vs. audio dominance in the guidance of rhythmic tasks. In a synchronization task using a taping paradigm, Repp and Penel [2004] showed the dominance of audition over vision. Specifically, while discrete auditory cues provide higher temporal accuracy for the guidance of various movements, a continuous auditory feedback leads to a more consistent synchronization of subjects' gesture velocity with the target [Rodger and Craig 2011; Varlet et al. 2012]. Overall, audio has a better resolution than vision for duration perception [Ortega et al. 2014], and the muscular response to auditory stimuli is faster (80 ms) than to visual stimuli (125 ms) [Luschei et al. 1967]. Moreover, the auditory modality, with its higher temporal resolution, seemed predominant over the visual modality independently of its weighting and the subjects' attention. From this property, it has been shown that auditory feedback can be substituted for kinesthetic feedback in the production of repeated movements among people lacking in proprioception [Ghez et al. 2000].

Another meeting point of audio, visual, and kinesthetic control is the sonification of the cursor position on a screen to assist in target acquisition tasks. Nevertheless, the auditory feedback used for this sake was often discrete, that is, associated to a punctual event (cursor on a target [Akamatsu et al. 1995; Cockburn and Brewster 2005; Charoenchaimonkon et al. 2010] or cursor out of a defined path [Sun et al. 2010]). In these cases, the rhythmic feature dominated, as discussed above. Few studies have dealt with the influence of continuous feedback related to gesture movement. Andersen and Zhai [2010] investigated the sonification of a stylus trajectory on a graphic tablet in the drawing of shapes, using pitch, amplitude, and timbre changes of a synthesized sound. Whereas a given visual feedback improved shape reproduction, the auditory feedback had little impact. It was exerted here that auditory information processing was too slow to influence motor control and that the latter may rely more on strong internal representations rather than feedback [Wright 1990]. Nevertheless, the subjects' answers to a modified version of the Questionnaire for User Interface Satisfaction [Chin et al. 1988], collected in a second experiment, drew attention to the greater attractiveness of the task when auditory feedback was provided.

Further attention has been drawn to the use of auditory guidance in shape drawing by the sonification of movement kinematics. Metaphors relating gesture kinematics to sound features (e.g., amplitude and brightness) were introduced by Caramiaux et al. [2010]. Later, Boyer et al. [2013] showed that changes in the timbre of a synthesized sound used as auditory feedback helped users match a given angular velocity profile in the displacement of an object. In the case of natural sounds related to gestures (such as rubbing a stylus on a sheet of paper), some gestures can be identified solely from the sound they produce [Thoret et al. 2014a], providing evidence of a strong cognitive link between a gesture and its natural auditory feedback. The same authors investigated the influence of artificial visual and auditory feedback on the drawing of shapes [Thoret et al. 2014c, 2014b]. As participants drew ellipses or circles on a hidden graphic tablet, a cursor moved accordingly on a screen, and a corresponding rubbing sound was synthesized and played in headsets. The introduction of interferences by changing the ellipses' eccentricity between action and audio and visual feedback proved to have a large influence on the kinematic trajectories of the drawings, emphasizing the strong dependence between gesture and natural sounds. Additionally, the same team proved that real-time sonification of trajectory kinematics with timbre variation of a synthesized sound considerably improved the learning of new written characters [Danna et al. 2014].

The present study focuses on a melodic target acquisition task: pointing to an accurate pitch target. This music-related specific task differs considerably from melodic contour imitation, as in speech intonation, and it is neither a rhythmic or synchronization task nor a shape or contour task. Therefore, a new experimental paradigm is needed.

1.3 Spatial Representation of Pitch

Melodic control using a graphic tablet can be considered in the framework of the ideomotor principle. An action is planned to obtain the results for which the perception is anticipated [Greenwald 1970]. After a learning period, the results of an action can be predicted; consequently, the action is planned accordingly. According to the theory of event coding, action and perception share the same cognitive domain of representation [Hommel et al. 2001]. Therefore, action and perception are tightly linked, and the perception of an event strongly affects a simultaneous planned action.

Spatial organization is, of course, inherent to visual targeting tasks. It is important to point out that spatial organization also seems to underlie melodic targeting. Cognitive spatial pitch representation has been studied by Rusconi et al. [2006]. In one task, participants compared pairs of sounds according to their relative melodic pitches by pressing left and right, or up and down, keys on a keyboard. The results showed that higher and lower pitches were identified faster when participants were asked to press right/up or left/down keys, respectively. The subsequent Spatial-Musical Association of Response Codes (SMARC) effect suggests a spatial representation of pitch that was also confirmed by Kadosh et al. [2008], who compared response times depending on interval sizes. If both visual and auditory modalities share a spatial representation in targeting tasks that are likely to influence motor action, then this raises the question of which perceived modality is more likely to influence motor planning. Therefore, methodologies developed for the study of spatial, visual, and kinesthetic interferences seem adapted to study spatial, audio, and kinesthetic interferences and are presented below.

1.4 Spatial Visual and Kinesthetic Interferences

Dominance of the visual modality over the kinesthetic modality has been reported in the literature [Sutter et al. 2013]. Rieger et al. [2005] studied the adaptation to changes in the environment. While subjects were asked to reach targets on a screen using a cursor controlled by a covered graphic tablet, gain changes were applied to modify the amplitude of either the cursor or the stylus only. Subjects' compensations for cursor perturbations were faster than for stylus perturbations, suggesting that adaptation to the environment relies mostly on visual feedback. More evidence on visual modality dominance emerges from perception of shapes drawn by hand motions [Müsseler and Sutter 2009; Wang et al. 2012]. In these studies, participants were asked to draw shapes with hand gestures while perceiving conflicting visual feedback on a screen. They appeared mostly unaware of their hand movements and were strongly influenced by the conflicting visual feedback.

Following these studies, a series of experiments aimed at measuring and quantifying the aftereffects observed in hand movements by introducing interferences between the visual and kinesthetic modalities was developed [Ladwig et al. 2012, 2013; Wendker et al. 2014]. Subjects were asked in a first phase to aim at a visual target on a screen by moving a cursor controlled by a stylus on a covered graphic tablet. In a second phase, the visual feedback disappeared, and the subjects' task was to produce the reverse stylus motion, with the same stylus motion amplitude. Different gains were applied between stylus and cursor amplitude, either by changing the cursor amplitude only or by changing the stylus amplitude only. The results showed that subjects overshot or undershot the stylus movement when the cursor amplitude was larger or shorter, respectively, than the stylus amplitude. These results were emphasized when the direction of the stylus motion matched the direction of the cursor motion [Ladwig et al. 2012] while no aftereffects were observed for perturbed cursor amplitude conditions when the cursor trajectory (elliptic) did not match the stylus trajectory (straight line) [Wendker et al. 2014]. Furthermore, Ladwig et al. [2013] showed that when the second phase consisted of reversing the cursor motion instead of the stylus motion, no aftereffects were observed following perturbed stylus amplitude.

1.5 Aims of the Work

In this study, interferences among auditory, visual, and motor modalities are investigated in the context of a multimodal target acquisition task with a graphic tablet, using an experimental paradigm derived from Ladwig et al. [2012]. The experiments were designed to investigate the following: (1) the impact of visual feedback on pointing target acquisition tasks, (2) the impact of auditory feedback on pitch target acquisition tasks, and (3) the interferences between visual and auditory feedbacks and their impacts on audio-visual target acquisition tasks. Three experimental conditions were tested: a visual condition (V condition) using visual feedback only (this is a replication of Ladwig et al. [2012] for the sake of comparison); an auditory condition (A condition) using auditory feedback only; and an audiovisual condition (AV condition) that presented visual and audio at the same time, thus, causing interferences with each other. The first aim of the experiment was to study interferences between audio and motor effects, in comparison with interferences between visual and motor effects. The second goal, in the audio-visual condition, was to investigate whether visual and auditory modalities have comparable effects or whether vision dominates audio or vice versa.

2. EXPERIMENTS

2.1 Procedure

An experimental paradigm inspired by Ladwig et al. [2012] was constructed, consisting of a target acquisition task guided by a visual, an auditory, or an audiovisual feedback, followed by a return movement. The audio and/or visual target is acquired with a straight and horizontal stylus stroke on a graphic tablet. The tablet is kept out of sight and the visual information is displayed by a cursor on a screen to spatially separate the visual and motor events. Pitch is controlled by the horizontal stylus position and played through headphones. Both visual and auditory feedback are distal actions controlled by the proximal motor action. The experiment setting is shown in Figure 1. One stimulus acquisition is divided into two phases. Phase 1 is a closed-loop target acquisition task, where the subject moves the stylus in one horizontal direction to reach a distal target provided by visual or/and auditory feedback. After the target is reached, Phase 2 begins, which consists of returning to the stylus starting point by reversing the stylus movement without the aid of the visual or/and auditory feedback and as accurately as possible. This is an open-looped replication task. An illustration of the different phases for each condition is shown in Figure 2. The distance between the final and initial stylus positions, called stylus deviation, is measured to quantify the influence of the feedback on the perception of movement.

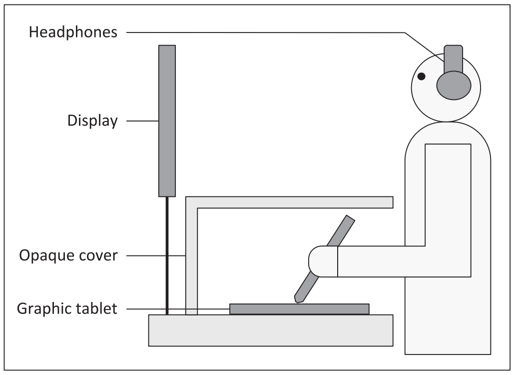

Fig. 1. Setting of the experiment.

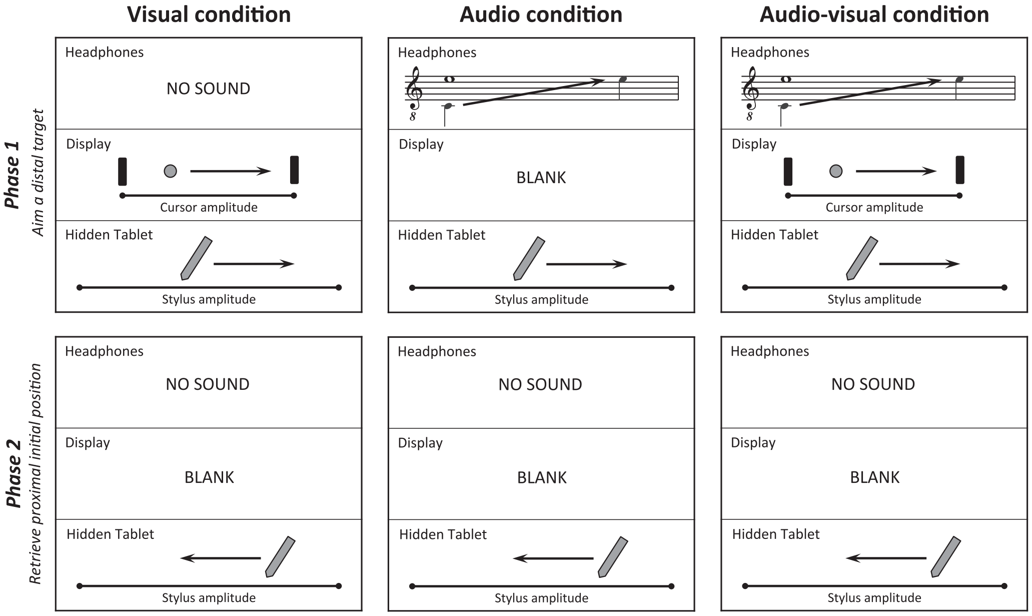

Fig. 2. The two phases of the experiment for each condition. The relationships among the cursor, the stylus, and the pitch amplitudes are given in Section 2.3.

The experiment was conducted in an acoustically insulated room and run on an Apple Macintosh computer in the Max/MSP environment (Cycling 74'). Subjects sat in front of a 75cm-high table. A Wacom Intuos 5M graphic tablet lying on the table at 10cm from the subjects and its corresponding stylus were used to record the subjects' horizontal strokes. A panel was placed above the graphic tablet to hide the latter from the subjects' view, as shown in Figure 1. The visual feedback was displayed on a DELL 2007FP screen with a 1600 × 1200 resolution and a 60Hz refresh rate, located at 50cm from the subject. The auditory feedback was played through AKG-K271 headphones with a sampling rate of 44,100Hz.

Table I. Summary of the Nine Gains Used for the *V* and *A* Conditions

	Cursor		
Stylus	Small (6 cm)	Medium (12 cm)	Large (18 cm)
Small (6 cm)	Unperturbed	Perturbed stylus	Opposite perturbation
Medium (12 cm)	Perturbed cursor	Unperturbed	Perturbed cursor
Large (18 cm)	Opposite perturbation	Perturbed stylus	Unperturbed

a. Visual condition

	Pitch		
Stylus	Small (8 ST)	Medium (16 ST)	Large (24 ST)
Small (6 cm)	Unperturbed	Perturbed stylus	Opposite perturbation
Medium (12 cm)	Perturbed pitch	Unperturbed	Perturbed pitch
Large (18 cm)	Opposite perturbation	Perturbed stylus	Unperturbed

b. Auditory condition

2.2 Participants

Fifteen participants took part in the experiment (five were female), recruited either from the lab or the university. The participants were between 18 and 42 years of age (average 23 years); none of them reported auditory or motor issues or non-corrected vision, and all were right handed. Twelve had musical training (average 11.7 years), and six either regularly or occasionally used a graphic tablet. All the participants were familiar with controlling a cursor on the screen: eleven from using a trackpad and four from using a mouse exclusively. All the participants were naive to the experiment and received compensation (20 euros).

2.3 Experiment Design

2.3.1 *Visual (V) Condition: Visual and Kinesthetic Interferences.* In the *V* condition, two rectangular bars of size 0.8×0.2cm along with a 0.4cm-diameter circular cursor are displayed on the screen (Figure 2). Phase 1 commences when the cursor is on the start bar, whereas phase 2 begins when the cursor reaches the second bar. Neither the bars nor the cursor are displayed in Phase 2. The relationship between the stylus and the cursor amplitude is perturbed by nine different gains: three stylus amplitudes (6, 12, and 18cm) combined with three cursor amplitudes (6, 12, and 18cm) are tested. Identical values as reported in Ladwig et al. [2012] are chosen for the sake of comparison. The combinations (stylus = 12cm × cursor = (6 or 18cm)) are called perturbed cursor amplitudes, whereas the combinations (stylus = (6 or 18cm) × cursor = 12cm) are called perturbed stylus amplitudes, following the terminology in Ladwig et al. [2012]. The combinations where stylus amplitude is large or small and the cursor amplitude is small or large, respectively, are called opposite perturbations. The gains and the associate terminologies are listed in Table I(a).

2.3.2 *Auditory (A) Condition: Auditory and Kinesthetic Interferences.* In the *A* condition, a voice-like sound with an /a/ vowel is generated with *Cantor Digitalis* [Feugère 2013] and is played continuously. The pitch changes as the stylus moves horizontally on the tablet, the lowest pitch is on the left and the highest pitch on the right. Phase 1 starts with an initial pitch of (C3 = 131Hz) associated to the stylus position, along with a reference pitch played with a similar sound. Phase 2 begins when the stylus pitch reaches the reference pitch. No sounds are played during Phase 2 (Figure 2). In the *V* condition, the target presented on the screen has a width that introduces a certain tolerance in reaching the target. To simulate this width in the audio domain, an automatic pitch correction procedure adjusts the pitch to the closest semitone if its variation is low enough to be considered stable [Perrotin and d'Alessandro 2016]; in other words, the pitch is adjusted only when the stylus gets close to the target. The same three stylus amplitudes (6, 12, and 18cm) are tested along with three pitch amplitudes (8, 16, and 24 semitones (ST)), giving nine stimuli for this condition. These three pitch amplitudes are chosen to be equally spaced while avoiding intervals of a perfect fifth (which are highly consonant and might lead to pitch-matching errors) and to limit the largest interval to two octaves. This corresponds to a distance of 1.33ST/cm on the tablet for unperturbed conditions. The combinations (stylus = 12cm × pitch = (8 or 24ST)) are called perturbed pitch amplitudes, while the combinations (stylus = (6 or 18cm) × pitch = 16ST) are called perturbed stylus amplitudes. The combinations where the stylus amplitude is large or small and the pitch amplitude is small or large, respectively, are called opposite perturbations. Table I(b) shows a summary of the different gains and the associated terminologies.

2.3.3 *Audio-visual (AV) Condition: Visual, Auditory, and Kinesthetic Interferences.*

In the AV condition, both the visual target and the reference pitch are provided simultaneously. The three stylus amplitudes, the three cursor amplitudes, and the three pitch amplitudes described for the previous conditions are also tested here, resulting in 27 stimuli for this condition. The perturbed stylus amplitude combinations are (stylus = (6 or 18cm) × cursor = 12cm × pitch = 16ST); the perturbed cursor combinations are (stylus = 12cm × cursor = (6 or 18cm) × pitch = 16ST); and the perturbed pitch combinations are (stylus = 12cm × cursor = 12cm × pitch = (8 or 24ST)). Conditions where one amplitude is medium and both the others are large or small, respectively, are called double perturbations. Conditions where one of the amplitudes is medium and the others are large or small, respectively, are called opposite perturbations. Conditions where no amplitude is medium are called triple perturbations. The two opposite perturbation conditions of interest here are (stylus = 12cm × cursor = 6cm × pitch = 24 ST) and (stylus = 12cm × cursor = 18cm × pitch = 8 ST) where the cursor and pitch, respectively, have larger and shorter (or shorter and larger) amplitudes than the stylus and, therefore, interfere with each other. The gains and associated terminologies are listed in Table II.

Finally, for each condition, the two directions for target acquisition are investigated: starting with a leftward movement and/or higher pitch in phase 1 or conversely. In the end, 3 stylus amplitudes × 3 cursor amplitudes × 2 directions = 18 stimuli are proposed for the V condition, 3 stylus amplitudes × 3 pitch amplitudes × 2 directions = 18 stimuli are proposed for the A condition, and 3 stylus amplitudes × 3 cursor amplitudes × 3 pitch amplitudes × 2 directions = 54 stimuli are proposed for the AV condition.

2.4 Task

For all experiments, a starting bar was displayed on the screen along with a red cursor, at the beginning of each trial, for all conditions. Once the bar was reached, the cursor became orange and the subject had to press one of the tablet buttons to start Phase 1. At that point, both target bars were displayed and the cursor became green for the V and AV conditions; for the A condition, the starting bar and cursor were removed. The stylus and reference pitches were played together for the A and AV conditions. When the cursor reached the target bar and/or the reference pitch, the participants pressed the tablet button a second time to start Phase 2, in which they were asked to replicate their previous stylus movement in reverse and press the tablet button a third time when they thought they had reached their initial stylus position. Finally, the subjects were asked to press the tablet button one more time to move on to the next trial. Subjects were instructed to monitor their stylus movement during Phase 1 and to reach the target and their initial stylus position as efficiently and accurately as possible, without

Table II. Summary of the 27 Gains Used for the *AV* Condition

Small Stylus (6 cm)	**Cursor**		
Pitch	Small (6 cm)	Medium (12 cm)	Large (18 cm)
Small (8 ST)	Unperturbed	Double perturbation	Triple perturbation
Medium (16 ST)	Double perturbation	Perturbed stylus	Opposite perturbation
Large (24 ST)	Triple perturbation	Opposite perturbation	Triple perturbation

Medium Stylus (12 cm)	**Cursor**		
Pitch	Small (6 cm)	Medium (12 cm)	Large (18 cm)
Small (8 ST)	Double perturbation	Perturbed pitch	**Opposite perturbation**
Medium (16 ST)	Perturbed cursor	Unperturbed	Perturbed cursor
Large (24 ST)	**Opposite perturbation**	Perturbed pitch	Double perturbation

Large Stylus (18 cm)	**Cursor**		
Pitch	Small (6 cm)	Medium (12 cm)	Large (18 cm)
Small (8 ST)	Triple perturbation	Opposite perturbation	Triple perturbation
Medium (16 ST)	Opposite perturbation	Perturbed stylus	Double perturbation
Large (24 ST)	Triple perturbation	Double perturbation	Unperturbed

stopping or reversing the movement during each phase and without time constraints. Participants were holding the stylus in their dominant hands, while they were asked to press the tablet button with their other hands. Additionally, they were allowed to rest between each trial. Stimuli for the three experiments were randomly mixed within blocks of 90 stimuli. Five blocks preceded by a training session of 54 stimuli were presented to each subject. The experiment lasted approximately 120 min and the participants could rest as much as they liked between blocks.

3. RESULTS

For all conditions, the stylus deviations between Phases 1 and 2 were analyzed. These deviations (i.e., the difference between final and initial stylus positions) were extracted from error-free trajectories. Some trials were considered as errors and discarded. These errors occurred when the second click on the tablet button (the end of Phase 1) occurred but the target had not been reached, when subjects overshot the target, when the direction of movement changed within Phase 1 or 2, or when the movement was interrupted. The error rate was 2.89% under the *V* condition, 30.1% under the *A* condition and 2.10% under the *AV* condition. Note that the error rate is approximately one order of magnitude higher for the *A* condition; this result indicates that the *A* condition was more difficult for the subjects because it required more concentration.

In the following sections, the results are analyzed with respect to the different sets of conditions. In each case, mixed-effect models are used, where the influence of subjects and repetitions throughout each block are modeled as random factors. A simplifying procedure for each model is achieved by iteratively suppressing non-significant effects of factors, provided that the initial model and the simplified model do not differ significantly [Crawley 2013]. Analyses of variance give the deviances explained by each factor for each model and post hoc Tukey's HSD tests are run to quantify the influence of each factor. The *lme4*, *car*, and *multcomp* libraries from the R language were used for the analyses.

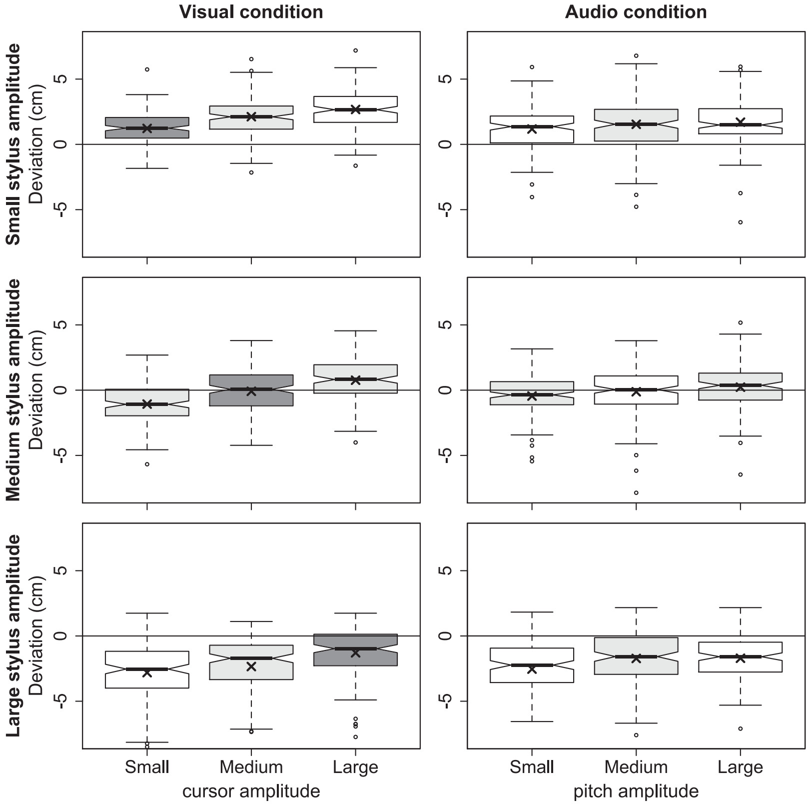

Fig. 3. The stylus deviation (cm) between guided and replicated stylus movements for all subjects and trials in the V condition (left) and the A condition (right) depending on the stylus amplitude (top to bottom) and the cursor or pitch amplitude (in each panel). Each box contains 50% of the values; the thick bars are medians and the crosses are means. The light grey boxes are the perturbed stylus, cursor, and pitch conditions, while the dark grey boxes are the unperturbed visual conditions.

3.1 Separate Auditory and Visual Modalities

Figure 3 relates the deviations of all subjects and trials for the V (left) and A (right) conditions depending on the stylus amplitude (top: small; middle: medium; bottom: large) and the cursor or pitch amplitude, respectively (on each panel, left: small; center: medium; right: large). Each box contains 50% of the values; the thick bars are medians and the crosses are means. The light grey boxes are the perturbed stylus, cursor, or pitch conditions as described by Ladwig et al. [2012], and the dark grey boxes are the unperturbed visual conditions (same stylus and cursor amplitude). Table III summarizes all mean deviations values in the motor domain (in cm), in the visual domain (in cm), and in the auditory domain (in ST).

3.1.1 *Effect of the Visual Modality.* The V condition is described by two fixed factors: the stylus amplitude (stylus: three levels) and the cursor amplitude (cursor: three levels). The first row of Table IV lists the deviances explained by each factor of the model after simplification. The first row of Table V gives the overall mean deviations for each factor and movement amplitude and their differences relative to the medium amplitude.

Significant aftereffects emerge due to the stylus and cursor amplitudes for the V condition. Compared to the medium stylus amplitude, the small stylus amplitude induces larger replication

Table III. Mean Deviations between Guided and Replicated Stylus Movement for the *V* and *A* Conditions in the Motor Domain and in the Visual/Auditory Domain (Between Brackets). Values in the Motor Domain Are Represented by the Crosses in Figure 3

	Cursor		
Stylus	Small (6cm)	Medium (12cm)	Large (18cm)
Small (6cm)	1.22cm (1.22cm)	2.12cm (4.24cm)	2.67cm (8.02cm)
Medium (12cm)	−1.07cm (−0.53cm)	−0.09cm (−0.09cm)	0.76cm (1.14cm)
Large (18cm)	−2.81cm (−0.94cm)	−2.34cm (−1.56cm)	−1.29cm (−1.29cm)

a. Visual condition

	Pitch		
Stylus	Small (8ST)	Medium (16ST)	Large (24ST)
Small (6cm)	1.18cm (1.56ST)	1.54cm (4.10ST)	1.70cm (6.80ST)
Medium (12cm)	−0.45cm (−0.30ST)	−0.12cm (−0.16ST)	0.23cm (0.46ST)
Large (18cm)	−2.54cm (−1.13ST)	−1.72cm (−1.53ST)	−1.71cm (−2.28ST)

b. Auditory condition

Table IV. Deviances Explained by Each Factor for *V*, *A*, and $V + A$ Sets of Conditions After Model Simplification. Significance Was Determined by a χ^2 Distribution

Condition	**Factor**	χ^2	**d.f.**	**p**
V	stylus	1,491	2	$<10^{-15}$
	cursor	224	2	$<10^{-15}$
A	stylus	775	2	$<10^{-15}$
	pitch	32.1	2	$<10^{-7}$
V and A	stylus	2,021	2	$<10^{-15}$
	perturbation	183	2	$<10^{-15}$
	stylus:modality	24.0	3	$<10^{-5}$
	perturbation:modality	31.6	2	$<10^{-7}$

Table V. Overall Mean Deviations for Each Factor between Guided and Replicated Stylus Movement for the *V* and *A* Conditions in the Motor Domain. Δ Values Are the Difference between the Overall Mean Deviation for Large and Medium Movement (Small and Medium Movement, Respectively). Significances Are Assessed by a Post-Hoc Tukey's HSD Test

		Small	Medium	Large
V	stylus	2.00cm ($\Delta = +2.13$cm, $p < 10^{-15}$)	−0.13cm	−2.15cm ($\Delta = -2.02$cm, $p < 10^{-15}$)
	cursor	−0.89cm ($\Delta = -0.79$cm, $p < 10^{-10}$)	−0.10cm	0.72cm ($\Delta = +0.82$cm, $p < 10^{-10}$)
A	stylus	1.46cm ($\Delta = +1.58$cm, $p < 10^{-15}$)	−0.12cm	−1.99cm ($\Delta = -1.87$cm, $p < 10^{-15}$)
	pitch	−0.60cm ($\Delta = -0.50$cm, $p < 10^{-3}$)	−0.10cm	0.07cm ($\Delta = +0.17$cm, $p = 0.34$)

movements (+2.13cm, $p < 10^{-15}$) while the large stylus amplitude leads to smaller replication movements (−2.02cm, $p < 10^{-15}$). Conversely, the replication movement is significantly shorter (−0.79cm, $p < 10^{-10}$) with small cursor amplitudes and longer with large cursor amplitudes (+0.82cm, $p < 10^{-10}$) compared to medium cursor amplitudes. No interactions between stylus and cursor are observed.

The impact of visual feedback on motor control has been demonstrated in the use of the graphic tablet [Müsseler and Sutter 2009; Sutter et al. 2011; Ladwig et al. 2012, 2013; Sutter et al. 2013; Wendker et al. 2014]. The V condition is a reference test to compare this study to previous articles. The results suggest that the aftereffects obtained in this condition are homologous and have the same order of magnitude as in previous studies [Ladwig et al. 2012]. A large cursor amplitude presented to the subject tends to make them overestimate the size of their hand movement, and the reverse is also true. Therefore, interferences between distal perception and proximal action can be reported.

A major effect observed here and not measured by Ladwig et al. [2012] is the influence of the stylus amplitude. The dark grey boxes in Figure 3 represent the unperturbed conditions, where the stylus and cursor amplitude are equal. While the medium stylus and cursor amplitudes result in a deviation centered on 0, the replicated movements systematically overshoot or undershoot the initial position for small or large stylus and cursor amplitudes, respectively. A 6cm or 18cm gesture on the tablet is always overestimated or underestimated, respectively, for any cursor amplitude. In other words, when the stylus amplitude varies, a constant bias specific to the stylus movement itself and independent from the visual perception is added to the cursor amplitude variation effects. This could suggest that the medium stylus amplitude (12cm) results in a more comfortable movement that subjects find easier to memorize. When confronted with smaller or larger amplitudes, the subjects tend to replicate a movement closer to this medium amplitude, expanding small amplitudes and contracting large ones. This effect has also been observed by Wendker et al. [2014] and is called regression effect [Teghtsoonian and Teghtsoonian 1978], which suggests the existence of a linear relation between the log ratio of the range of the produced stimuli (the stylus movements in phase 2) and the log ratio of the range of proposed stimuli (6 to 18cm). Taking the errors of the unperturbed conditions from Table III, we find a coefficient of $n_V = \log\left(\frac{18-1.29}{6+1.22}\right) / \log\left(\frac{18}{6}\right) = 0.76$, with the same order of magnitude than coefficients computed in Teghtsoonian and Teghtsoonian [1978].

The stronger effects of stylus amplitude over cursor amplitude are reflected in the asymmetry of the aftereffects observed between the perturbed cursor conditions (medium stylus amplitude and variable cursor amplitudes) and the perturbed stylus conditions (medium cursor amplitude and variable stylus amplitudes), which are highlighted by the light grey boxes in Figure 3 and have been identified in previous experiments [Ladwig et al. 2012; Wendker et al. 2014].

3.1.2 *Effect of the Auditory Modality.*

The A condition is described by two fixed factors: the stylus amplitude (stylus: three levels) and the pitch amplitude (pitch: three levels). The second row of Table IV lists the deviances explained by each factor of the model after simplification. The second row of Table V gives the overall mean deviations for each factor and movement amplitude and their differences relative to the medium amplitude.

Similar effects arise for the A condition. The main effect is the influence of the stylus amplitude, where replication movements become larger for short stylus amplitudes (+1.58 cm, $p < 10^{-15}$) and smaller for large stylus amplitudes (−1.87 cm, $p < 10^{-15}$) in comparison with the medium stylus amplitude. The pitch amplitude has a rather smaller effect; the mean of replication movements is reduced by only (−0.50 cm, $p < 10^{-3}$) for small pitch amplitudes, and it is not significant for large pitch amplitudes.

Audio interferences show a significant effect but much less impact than visual interferences on the stylus replications. The aftereffects are twice less pronounced than in the V condition. Additionally, one can note a clear asymmetry between perturbed pitch conditions (medium stylus amplitude and variable pitch amplitudes) and the perturbed stylus conditions (medium pitch amplitude and variable stylus amplitudes). Similarly to the visual case, a bias results from the different stylus amplitudes. The regression effect coefficient $n_A = \log\left(\frac{18-1.71}{6+1.18}\right) / \log\left(\frac{18}{6}\right) = 0.75$ is close to the one of the V condition, emphasizing the consistency of stylus movements across the different conditions.

3.1.3 *Comparison between Visual and Auditory Modalities.* To compare the relative influence of visual and auditory modalities on motor movement, the results are analyzed for a single (i.e., either visual or auditory) perturbation. For this purpose, the V and A conditions are merged and described by three fixed factors: the stylus amplitude (stylus: three levels), the perturbation amplitude (perturbation: three levels; small, medium, and large amplitude), and the modality (modality: two levels; visual and auditory). The last row of Table IV lists the deviances explained by each factor of this model after simplification.

The results show a significant main effect of the stylus amplitude on the stylus deviation. A significant but smaller effect of perturbation is also observed. Stylus-modality and perturbation-modality interactions show that the stylus and perturbation amplitudes have a significantly different impact on stylus deviation for the V and A conditions.

The magnitude difference between visual and audio effects can be explained by two main factors. First, vision is more strongly and naturally associated to spatial motion than pitch perception. It is a common experience to link visual feedback to a stylus motion of comparable magnitude and to the associated kinesthetic perception. Pitch spatial representation has been demonstrated but is, nevertheless, weaker, and there is no straightforward association between the magnitude of a stylus motion and the magnitude of a pitch movement. However, a consistent association between motion and pitch movement can be acquired when learning audio-motor skills such as playing the piano or violin. Second, pitch interval perception is a demanding task and uncommon in daily life compared to the perception of visual distances. Even though most of the subjects had some musical training, they all found the A condition more difficult than the others, which is reflected by the error rate observed for each condition (30% for A condition compared to less than 3% for the other conditions). Nevertheless, the results show the influence of audio on motions: a small pitch interval induces slightly shorter movements, whereas a large pitch interval induces larger movements. Even though the association is less important, auditory perception does indeed interfere with the proximal movements.

To conclude, visual and audio distractions and stylus amplitude all have a significant influence on stylus deviation. Stylus amplitudes are the main influence. Then, cursor amplitudes induce wider stylus deviations than pitch. Analysis of both the V and A conditions simultaneously showed that the difference of influence between visual and audio perturbation is significant. Thus, this experiment conducted in this article provides information on the impact of the auditory modality on motor control and its weight compared to the impact of the visual modality.

3.2 Joint Auditory and Visual Modalities

Figure 4 depicts the stylus deviations in the V (left column), A (first row), and AV (thick panels) conditions. The stylus amplitude effects are plotted from the left to the right in the boxes of each panel, the cursor amplitude effects are displayed from the top to bottom rows, and the pitch amplitude effects from the left to the right columns. The light grey boxes represent the perturbed pitch, cursor, and stylus conditions, while the dark grey boxes represent the unperturbed conditions. Table VI summarizes all mean deviations values in the motor domain (in cm), in the visual domain (in cm), and in the auditory domain (in ST).

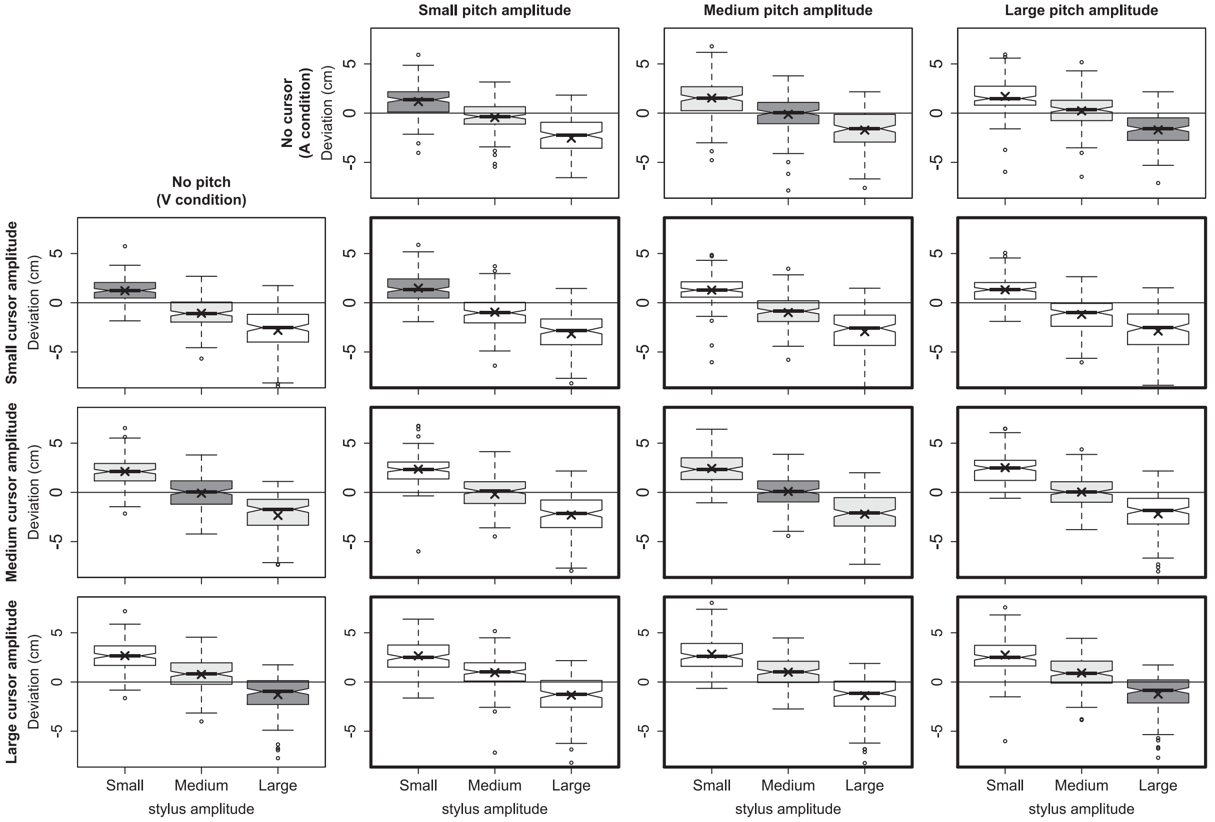

Fig. 4. The stylus deviation (cm) between guided and replicated stylus movements for all subjects and trials in the *V* condition (left column), *A* condition (first row) and *AV* condition (thick panels) depending on cursor amplitudes (top to bottom rows), pitch amplitudes (left to right columns), and stylus amplitudes (in each panel). Each box contains 50% of the values; the thick bars are medians and the crosses are means. The light grey boxes are the perturbed stylus, cursor, and pitch conditions while the dark grey are the unperturbed conditions.

3.2.1 *Interferences between Visual and Auditory Modalities.* The results are first analyzed by considering the *AV* condition only, with three fixed factors: the stylus amplitude, the cursor amplitude, and the pitch amplitude. The first row of Table VII lists the deviances explained by each factor of this model after simplification. Table VIII gives the overall mean deviations for each factor and movement amplitude and their differences relatively to the medium amplitude.

As for the *V* and *A* conditions, the stylus amplitude has the strongest effect, with replication movements larger (+2.22 cm, $p < 10^{-15}$) for small stylus amplitudes and smaller (−2.14 cm, $p < 10^{-15}$) for large stylus amplitudes in contrast to medium stylus amplitudes. The cursor amplitude has the inverse effect, as the replication movements are shorter (−0.95 cm, $p < 10^{-10}$) for smaller amplitudes and longer (+0.73 cm, $p < 10^{-10}$) for larger amplitudes. No significant effect of pitch amplitude is observed. On the other hand, a significant interaction between stylus and cursor amplitudes is evident, revealing no differences between medium and large cursor amplitudes for small stylus movements while an expansion of the replication movement appears for the other stylus amplitudes.

The *AV* condition confronts both modalities. It appears that the stronger influence of the visual feedback inhibits the auditory feedback effects because the pitch amplitude factor has no significant effect

Table VI. Mean Deviations between Guided and Replicated Stylus Movement for the AV Condition in the Motor Domain and in the Visual/Auditory Domain (between Brackets). Values in the Motor Domain Are Represented by the Crosses in Figure 4

Small Stylus (6 cm)	**Cursor**		
Pitch	Small (6cm)	Medium (12cm)	Large (18cm)
Small (8ST)	1.49cm (1.49cm/1.98ST)	2.35cm (4.71cm/3.14ST)	2.66cm (7.98cm/3.55ST)
Medium (16ST)	1.28cm (1.28cm/3.42ST)	2.43cm (4.87cm/6.49ST)	2.84cm (8.51cm/7.57ST)
Large (24ST)	1.32cm (1.32cm/5.27ST)	2.51cm (5.02cm/10.0ST)	2.74cm (8.22cm/11.0ST)

Medium Stylus (12 cm)	**Cursor**		
Pitch	Small (6cm)	Medium (12 cm)	Large (18 cm)
Small (8ST)	−0.95cm (−0.48cm/−0.64ST)	−0.19cm (−0.19cm/−0.13ST)	0.95cm (1.43cm/0.63ST)
Medium (16ST)	−1.00cm (−0.50cm/−1.33ST)	0.09cm (0.09cm/0.12ST)	1.00cm (1.50cm/1.33ST)
Large (24ST)	−1.18cm (−0.59cm/−2.37ST)	0.03cm (0.03cm/0.06ST)	0.92cm (1.38cm/1.85ST)

Large Stylus (18 cm)	**Cursor**		
Pitch	Small (6cm)	Medium (12cm)	Large (18cm)
Small (8ST)	−3.16cm (−1.05cm/−1.40 ST)	−2.30cm (−1.53cm/−1.02 ST)	−1.34cm (−1.34cm /−0.60ST)
Medium (16ST)	−2.94cm (−0.98cm/−2.61 ST)	−2.21cm (−1.47cm/−1.96 ST)	−1.39cm (−1.39cm /−1.23ST)
Large (24ST)	−2.87cm (−0.96cm/−3.83 ST)	−2.19cm (−1.46cm/−2.92 ST)	−1.24cm (−1.24cm /−1.65ST)

Table VII. The Deviances Explained by Each Factor for the AV and $V + AV$ Conditions After Model Simplification. Significance Was Determined by a χ^2 Distribution

Condition	**Factor**	χ^2	**d.f.**	**p**
AV	stylus	4,972	2	$<10^{-15}$
	cursor	746	2	$<10^{-15}$
	stylus:cursor	30.3	4	$<10^{-5}$
V and AV	stylus	6,447	2	$<10^{-15}$
	cursor	966	2	$<10^{-15}$
	stylus:cursor	35.1	4	$<10^{-7}$

Table VIII. Overall Mean Deviations between Guided and Replicated Stylus Movement for the AV Condition in the Motor Domain. Δ Values Are the Difference between the Overall Mean Deviations for Large and Medium Movement (Small and Medium Movement, Respectively). Significances Are Assessed by a Post-Hoc Tukey's HSD Test

		Small	Medium	Large
AV	stylus	2.18cm ($\Delta = +2.22$cm, $p < 10^{-15}$)	−0.04cm	−2.18cm ($\Delta = -2.14$cm, $p < 10^{-15}$)
	cursor	−0.89cm ($\Delta = -0.95$cm, $p < 10^{-15}$)	0.06cm	0.79cm ($\Delta = +0.73$cm, $p < 10^{-15}$)

on replication movements. As the visual feedback is easier to follow than the auditory feedback, in this competing situation, the subjects seem to have essentially relied on the cursor movement rather than on the pitch movement to complete the target acquisition task. They appeared to have been influenced by the visual feedback only. For the opposite perturbation conditions, where either the cursor amplitude is larger than the stylus amplitude and the pitch amplitude smaller or, conversely, auditory and visual feedback were in these cases supposed to influence the movement in the opposite direction. Nevertheless, the results show no dependence on the pitch, indicating that the visual feedback was dominant over the auditory feedback.

3.2.2 *Influence of the Presence of an Auditory Distraction.* Because pitch has no influence in the AV condition, a last analysis is conducted, merging the V and AV conditions to test the influence of the addition of an audio modality to a visual distraction. The same three fixed factors are used: the stylus amplitude, the cursor amplitude, and the pitch amplitude (four levels: none, small, medium, and large amplitude). The level "none" of the pitch amplitude corresponds to the V condition. The second row of Table VII lists the deviances explained by each factor of this model after simplification. Again, no influence of either pitch alone or pitch interaction with stylus and cursor appears. This indicates that stylus and cursor have similar effects both with and without audio feedback in the presence of visual distractions.

Overall, the magnitudes of the effects are strongest for the stylus, moderate for the cursor, and null for the audio. A clear asymmetry can be observed between the perturbed cursor (medium stylus and pitch amplitudes and variable cursor amplitudes), perturbed pitch (medium stylus and cursor amplitudes and variable pitch amplitudes) and perturbed stylus (medium cursor and pitch amplitudes and variable stylus amplitudes) conditions, as highlighted by the light grey boxes in Figure 4. These results indicate that the visual modality dominates over the auditory modality; the stylus movements are influenced by visual perturbations but not by auditory perturbations.

4. CONCLUSION

4.1 Discussion

The major result of our study is that the presence of an auditory modality alone, a distal action effect, interferes with proprioception, a proximal action effect, in a pitch target acquisition task. This is in good agreement with the ideomotor principle and the theory of event coding because audio, visual, and audio-visual interferences with motor effects indicate a cognitive level cooperation between proximal and distal action effects for a given pitch targeting task. The audio effect is similar to previously reported effects for visual and proprioceptive interferences but with a smaller magnitude. In the absence of visual feedback, a non-negligible influence of the auditory feedback on gesture replication was also observed in Andersen and Zhai [2010]. Our results are also compatible with the spatial organization of pitch perception demonstrated in other studies (the SMARC effect).

Additionally, we observed that when visual and auditory feedback are presented simultaneously, the predominance of the visual modality could be explained by a closer association of vision with motion than of audition with motion in the representation of space. Visual spatial resolution is, of course, much more accurate than audio spatial resolution because visual spatial associations are more common experiences, although the audio spatial association is unconscious or learned in the special situation of musical training. The stronger influence of visual feedback shows in the improved control when viewing visual cues instead of listening to pitch variations. Consequently, the learning curve for the manipulation of a gesture-controlled system is linked to these effects. Because the auditory information is harder to process, some amount of training seems necessary to manipulate the system with both distal perceptions. Indeed, in the case of a musical instrument, for instance, it is well known that the visual dependency drops as the players become more familiar with that instrument through regular practice.

4.2 Applications

In this study, only target acquisition tasks were considered. The question of contour (rather than target) processing certainly differs. Gesture dynamics are more salient in contour drawing, and, therefore, rhythmic aspects are implied. Audition may play a more important role in contour processing than in target acquisition processing and should be investigated by new experiments. Additionally, apart from pitch and rhythm, timbre is a third sound dimension that is often controlled by gestures in sonification or musical tasks. This dimension shows strong analogies with gesture kinematics, and interferences between proprioception and timbre variation would certainly be worth investigating. The question of kinesthetic cues is also left open. In many real-world situations, hand motions are also guided by kinesthetic discontinuities such as keyboards, cranked wheels, uneven tactile surfaces, and so on. The influence of kinesthetic cues on gesture control in audio-visual context would also require new experiments.

On the practical side, fine spatial and temporal resolutions for visual and auditory modalities, respectively, are complementary and can be exploited in gesture-controlled systems (e.g., digital musical instruments). Particularly for novice users, a decoupling of visual/spatial and auditory/temporal perception could help in processing distal feedback information. Eventually, a full understanding of the links among each modality (visual, auditory, kinesthetic) and gesture characteristics (position, timing, kinematics) might lead to the design of more sophisticated controls, which have applications ranging from musical applications (sound-oriented systems) to medical assistance applications (motion-oriented systems).

REFERENCES

Motoyuki Akamatsu, I. Scott MacKenzie, and Thierry Hasbrouc. 1995. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics* 38, 4 (1995), 816–827. DOI:http://dx.doi.org/10.1080/00140139508925152 PMID: 7729406.

Tue Haste Andersen and Shumin Zhai. 2010. "Writing with music": Exploring the use of auditory feedback in gesture interfaces. *ACM Trans. Appl. Percept.* 7, 3, Article 17 (Jun. 2010), 24 pages. DOI:http://dx.doi.org/10.1145/1773965.1773968

Eric O. Boyer, Quentin Pyanet, Sylvain Hanneton, and Frédéric Bevilacqua. 2013. Learning movement kinematics with a targeted sound. In *Sound, Music, and Motion: 10th International Symposium, CMMR*, Mitsuko Aramaki, Olivier Derrien, Richard Kronland-Martinet, and Sølvi Ystad (Eds.). Springer, Marseille, France, 218–233. DOI:http://dx.doi.org/10.1007/978-3-319-12976-1_14

Baptiste Caramiaux, Frédéric Bevilacqua, and Norbert Schnell. 2010. Towards a gesture-sound cross-modal analysis. In *Gesture in Embodied Communication and Human-Computer Interaction*, Stefan Kopp and Ipke Wachsmuth (Eds.). Lecture Notes in Computer Science, Vol. 5934. Springer, Berlin, 158–170. DOI:http://dx.doi.org/10.1007/978-3-642-12553-9_14

Eakachai Charoenchaimonkon, Paul Janecek, Mattew N. Dailey, and Atiwong Suchato. 2010. A comparison of audio and tactile displays for non-visual target selection tasks. In *Proceedings of the International Conference on User Science and Engineering (i-USEr)*. IEEE, Shah Alam, Malaysia, 238–243. DOI:http://dx.doi.org/10.1109/IUSER.2010.5716759

John P. Chin, Virginia A. Diehl, and Kent L. Norman. 1988. Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'88)*. ACM, New York, NY, 213–218. DOI:http://dx.doi.org/10.1145/57167.57203

Andy Cockburn and Stephen Brewster. 2005. Multimodal feedback for the acquisition of small targets. *Ergonomics* 48, 9 (Jul. 2005), 1129–1150. DOI:http://dx.doi.org/10.1080/00140130500197260

Michael J. Crawley. 2013. *The R Book–Second Edition*. Wiley, Chichester, West Sussex, United Kingdom. Retrieved from http://www.bio.ic.ac.uk/research/mjcraw/therbook/index.htm.

Christophe d'Alessandro, Lionel Feugère, Sylvain Le Beux, Olivier Perrotin, and Albert Rilliard. 2014. Drawing melodies: Evaluation of chironomic singing synthesis. *Acoust. Soc. Am.* 135, 6 (June 2014), 3601–3612. DOI:http://dx.doi.org/10.1121/1.4875718

Christophe d'Alessandro, Albert Rilliard, and Sylvain Le Beux. 2011. Chironomic stylization of intonation. *Acoust. Soc. Am.* 129, 3 (2011), 1594–1604. DOI:http://dx.doi.org/10.1121/1.353180

Nicolas d'Alessandro, Christophe d'Alessandro, Sylvain Le Beux, and Boris Doval. 2006. Real-time CALM synthesizer new approaches in hands-controlled voice synthesis. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*. ACM, 266–271.

Nicolas d'Alessandro and Thierry Dutoit. 2009. *Handsketch: Bi-Manual Control of Voice Quality Dimensions and Long Term Practice Issues*. Quarterly Progress and Status Report 2. Numediart Research Program.

Jérémy Danna, Maureen Fontaine, Vietminh Paz-Villagrán, Charles Gondre, Etienne Thoret, Mitsuko Aramaki, Richard Kronland-Martinet, Sølvi Ystad, and Jean-Luc Velay. 2014. The effect of real-time auditory feedback on learning new characters. *Hum. Move. Sci.* 43 (2014), 216–228. DOI:http://dx.doi.org/10.1016/j.humov.2014.12.002

Lionel Feugère. 2013. *Synthèse Par Règles de la Voix Chantée contrôlée Par le Geste et Applications Musicales*. Ph.D. Dissertation. Université Pierre et Marie Curie (UPMC).

Claude Ghez, Thanassis Rikakis, R. Luke Dubois, and Perry R. Cook. 2000. An auditory display system for aiding interjoint coordination. In *Proceedings of the International Conference on Auditory Display (ICAD)*, Perry R. Cook (Ed.). International Community for Auditory Display, Atlanta, Georgia.

Anthony G. Greenwald. 1970. Sensory feedback mechanisms in performance control: With special reference to the ideo-motor mechanism. *Psychol. Rev.* 77, 2 (1970), 73–99. DOI:http://dx.doi.org/10.1037/h0028689

Bernhard Hommel, Jochen Müsseler, Gisa Aschersleben, and Wolfgang Prinz. 2001. The theory of event coding (TEC): A framework for perception and action planning. *Behav. Brain Sci.* 24 (2001), 849–937. http://www.psych.rwth-aachen.de/ifp-zentral/upload/muesseler/Publ/2001HoMuAsPrBBS.pdf.

Roi Cohen Kadosh, Warren Brodsky, Michal Levin, and Avishai Henik. 2008. Mental representation: What can pitch tell us about the distance effect? *Cortex* 44, 4 (2008), 470–477. DOI:http://dx.doi.org/10.1016/j.cortex.2007.08.002 Special Issue on Numbers, Space, and Action.

Loïc Kessous. 2004. Gestural control of singing voice, a musical instrument. In *Proceedings of Sound and Music Computing*. IRCAM; Centre Pompidou, Paris, France.

Stefan Ladwig, Christine Sutter, and Jochen Müsseler. 2012. Crosstalk between proximal and distal action effects when using a tool. *J. Psychol.* 220, 1 (2012), 10–15. DOI:http://dx.doi.org/10.1027/2151-2604/a000085

Stefan Ladwig, Christine Sutter, and Jochen Müsseler. 2013. Intra- and intermodal integration of discrepant visual and proprioceptive action effects. *Exp. Brain Res.* 231, 4 (2013), 457–468. DOI:http://dx.doi.org/10.1007/s00221-013-3710-2

Erich Luschei, Carol Saslow, and Mitchell Glickstein. 1967. Muscle potentials in reaction time. *Exp. Neurol.* 18, 4 (1967), 429–442. DOI:http://dx.doi.org/10.1016/0014-4886(67)90060-X

Jochen Müsseler and Christine Sutter. 2009. Perceiving one's own movements when using a tool. *Consc. Cogn.* 18 (2009), 359–365. DOI:http://dx.doi.org/10.1016/j.concog.2009.02.004

Laura Ortega, Emmanuel Guzman-Martinez, Marcia Grabowecky, and Satoru Suzuki. 2014. Audition dominates vision in duration perception irrespective of salience, attention, and temporal discriminability. *Atten. Percept. Psychophys.* 76 (2014), 1485–1502. DOI:http://dx.doi.org/10.3758/s13414-014-0663-x

Olivier Perrotin and Christophe d'Alessandro. 2016. Target acquisition vs. expressive motion: Dynamic pitch warping for intonation correction. *ACM Trans. Comput.-Hum. Interact.* 23, 3, Article 17 (Jun. 2016), 21 pages. DOI:http://dx.doi.org/10.1145/2897513

Bruno H. Repp and Amandine Penel. 2004. Rhythmic movement is attracted more strongly to auditory than to visual rhythms. *Psychol. Res.* 68, 4 (2004), 252–270. DOI:http://dx.doi.org/10.1007/s00426-003-0143-8

Martina Rieger, Günther Knoblich, and Wolfgang Prinz. 2005. Compensation for and adaptation to changes in the environment. *Exp. Brain Res.* 163, 4 (2005), 487–502. DOI:http://dx.doi.org/10.1007/s00221-004-2203-8

Matthew W. M. Rodger and Cathy M. Craig. 2011. Timing movements to interval durations specified by discrete or continuous sounds. *Exp. Brain Res.* 214 (2011), 393–402. DOI:http://dx.doi.org/10.1007/s00221-011-2837-2

Elena Rusconi, Bonnie Kwan, Bruno L. Giordano, Carlo Umiltà, and Brian Butterworth. 2006. Spatial representation of pitch height: The SMARC effect. *Cognition* 99, 2 (3 2006), 113–129. DOI:http://dx.doi.org/10.1016/j.cognition.2005.01.004

Minghui Sun, Xiangshi Ren, and Xiang Cao. 2010. Effects of multimodal error feedback on human performance in steering tasks. *Inform. Process. Soc. Jpn. J.* 51, 12 (2010), 2375–2383. DOI:http://dx.doi.org/10.11185/imt.6.193

Christine Sutter, Jochen Müsseler, and L. Bardos. 2011. Effects of sensorimotor transformations with graphical input devices. *Behav. Inform. Technol.* 30, 3 (2011), 415–424. DOI:http://dx.doi.org/10.1080/01449291003660349

Christine Sutter, Sandra Sülzenbrück, Martina Rieger, and Jochen Müsseler. 2013. Limitations of distal effect anticipation when using tools. *New Ideas Psychol.* 31, 3 (2013), 247–257. DOI:http://dx.doi.org/10.1016/j.newideapsych.2012.12.001

Robert Teghtsoonian and Martha Teghtsoonian. 1978. Range and regression effects in magnitude scaling. *Percept. Psychophys.* 24, 4 (1978), 305–314. DOI:http://dx.doi.org/10.3758/BF03204247

Etienne Thoret, Mitsuko Amaraki, Richard Kronland-Martinet, Jean-Luc Velay, and Solvi Ystad. 2014a. From sound to shape: Auditory perception of drawing movements. *J. Exp. Psychol.: Hum. Percept. Perf.* 40, 3 (Jan. 2014), 983–994. DOI:http://dx.doi.org/10.1037/a0035441

Etienne Thoret, Mitsuko Aramaki, Christophe Bourdin, Lionel Bringoux, Richard Kronland-Martinet, and Solvi Ystad. 2014b. Audio-motor synchronization: The effect of mapping between kinematics and acoustic cues on geometric motor features. In *Sound, Music and Motion*. Springer, Berlin, 234–245. DOI:http://dx.doi.org/10.1007/978-3-319-12976-1_15

Etienne Thoret, Mitsuko Aramaki, Lionel Bringoux, Richard Kronland-Martinet, and Solvi Ystad. 2014c. When acoustic stimuli turn visual circles into ellipses: sounds evoking accelerations modify visuo-motor coupling. (June 11–14 2014). Presentation at International Multisensory Research Forum (IMRF).

Manuel Varlet, Ludovic Marin, Johann Issartel, R. C. Schmidt, and Benoit G. Bardy. 2012. Continuity of visual and auditory rhythms influences sensorimotor coordination. *PLoS ONE* 7, 9 (2012), 1–10. DOI:http://dx.doi.org/10.1371/journal.pone.0044082

Lei Wang, Christine Sutter, Jochen Müsseler, Ronald Josef Zvonimir Dangel, and Catherine Disselhorst-Klug. 2012. Perceiving one's own limb movements with conflicting sensory feedback: The role of mode of movement control and age. *Front. Psychol.* 3, 289 (Aug. 2012), 1–7. DOI:http://dx.doi.org/10.3389/fpsyg.2012.00289

Nike Wendker, Oliver S. Sack, and Christine Sutter. 2014. Visual target distance, but not visual cursor path length produces shifts in motor behavior. *Front. Psychol.* 5, 225 (Mar. 2014), 1–10. DOI:http://dx.doi.org/10.3389/fpsyg.2014.00225

Charles E. Wright. 1990. Generalized motor programs: Reexamining claims of effector independence in writing. In *Attention and Performance 13: Motor Representation and Control*, Marc Jeannerod (Ed.). Lawrence Erlbaum Associates, Hillsdale, NJ, 294–320.

Michael Zbyszynski, Matthew Wright, Ali Momeni, and Daniel Cullen. 2007. Ten years of tablet musical interfaces at CNMAT. In *Proceedings of the International Conference on New Interfaces for Musical Expression (NIME) (NIME'07)*. ACM, New York, NY, 100–105. DOI:http://dx.doi.org/10.1145/1279740.1279758

Received December 2015; revised July 2016; accepted July 2016