ELSEVIER

Spatial representation of pitch height: the SMARC effect

Elena Rusconi[a,b], Bonnie Kwan[a], Bruno L. Giordano[b],
Carlo Umiltà[b], Brian Butterworth[a,*]

[a]*Institute of Cognitive Neuroscience, University College London, 17 Queen Square, WC1N 3AR, London, UK*
[b]*Università degli Studi di Padova, Padova, Italy*

Abstract

Through the preferential pairing of response positions to pitch, here we show that the internal representation of pitch height is spatial in nature and affects performance, especially in musically trained participants, when response alternatives are either vertically or horizontally aligned. The finding that our cognitive system maps pitch height onto an internal representation of space, which in turn affects motor performance even when this perceptual attribute is irrelevant to the task, extends previous studies on auditory perception and suggests an interesting analogy between music perception and mathematical cognition. Both the basic elements of mathematical cognition (i.e. numbers) and the basic elements of musical cognition (i.e. pitches), appear to be mapped onto a mental spatial representation in a way that affects motor performance.
© 2005 Elsevier B.V. All rights reserved.

1. Introduction

"…The results are clear-cut and unequivocal. *High tones are phenomenologically higher in space than low ones.* …The fact that on any place-theory of hearing the lowest tones would fall at the apex and the highest tones at the base of the cochlea

---

* Corresponding author. Tel.: +44 20 7679 1150; fax: +44 20 7813 2835/71 580 1100.
 E-mail address: b.butterworth@ucl.ac.uk (B. Butterworth).

opposite the oval window no more means that we hear the world upside down than the inversion of the retinal image forces us to stand on our heads to see the world right side up. The experiments were done, however, not so much with auditory theory in mind as with the query as to whether the results would throw any light on the moot question of the apparent auditory movement which is set up by tones of different pitch when presented in succession."

Pratt, 1930

Pitch, primarily determined by frequency (Moore, 2003), is classified in many languages by using terms having a spatial connotation such as "high" and "low" (e.g. Chinese, English, French, German, Italian, Polish and Spanish). Melara and Marks (1990) showed that, in speeded reaction tasks, responding to the written word "high" is faster in the presence of a high pitch, whereas responding to "low" is faster in the presence of a low pitch. They argued that crosstalk occurs between written words and pitch at a semantic level of processing, which suggests that the association between verbal labels denoting spatial positions and sound frequency is not arbitrary. In music psychology, the structural relations among perceived pitches have been described with geometrical models derived from multidimensional scaling of relative pitch judgments (Shepherd, 1982; Ueda & Ohgushi, 1987). A model which is typically regarded as an effective metaphor for pitch perception in both musically naïve people and trained musicians is in the shape of an ascending spiral having circular (chroma) and vertical rectilinear (octave) components (Ueda & Ohgushi, 1987). Also, the standard music notation system maps pitches to vertical locations, whereby notes corresponding to higher pitches are represented with higher spatial positions on the staff.

Stumpf (1883) was looking for the origin of the association between pitch height and the vertical spatial dimension which emerged consistently across languages. He argued that cross-modality associative bonds and relations of similarity are used as effective metaphors but, in fact, no real spatial characterization is intrinsic to the tonal sensation. It was only in 1930 that Pratt put forward the hypothesis of a real pitch-space correspondence after observing that the specific succession of the tones in a musical phrase can generate a sensation of apparent movement (e.g. by presenting successively the notes of the diatonic scale from *C3* to *C4*, almost everybody perceives an upward movement). He asked participants to locate on a numbered scale running from the floor to the ceiling the position of tones sounded by a hidden loudspeaker having variable location. Five tones were used which bore the octave-relation to each other and had frequencies of 256, 512, 1024, 2048, and 4096 Hz. Participants were asked to indicate by one of the numbers on the scale the region from which the tone seemed to come. Every participant consistently judged the tones to be placed in the order, from top to bottom, 4096, 2048, 1024, 512, and 256 Hz. Thus, Pratt concluded that "…of two tones of different pitch the one of greater frequency is called higher, not because of any extraneous associations with altitude, but simply because it is perceived as occupying a higher position in phenomenological space" (Pratt, 1930, p. 283). However, as Pratt, (1930) himself and Trimble (1934) noticed, it is possible to interpret those results in the light of habits in pitch discrimination, in which case localization would be a matter of discrimination on a pitch continuum rather than discrimination on a spatial continuum

(i.e. participants systematically assigned locations on the basis of the pitch, even if not instructed to do so, for sound localization is very poor in the vertical dimension). Moreover, the only expedient aimed at preventing the use of words such as "high" and "low" to mean perceived spatial location, but which could also evoke the associated concept of pitch height, was labelling locations with numbers, and the number-to-location assignment was not counterbalanced between participants (small numbers were always assigned to low and large numbers to high locations). Trimble (1934; Experiment 2) replicated Pratt's (1930) results with a fixed sound source and nine different tones, which did not bear the octave relation (500, 600, 700, 850, 1200, 1750, 1900, 2250, and 3950 Hz). He asked participants to record the apparent displacement of a phantom sound source from the horizontal plane both at the beginning and at the end of an "ascending" and a "descending" series of pitches in rapid succession. In addition, participants were instructed to draw on a chart the apparent course made by the phantom sound. As expected, higher frequency tones were localized higher in space than lower frequency tones and musically "ascending" series followed a bottom-to-up trajectory, whereas "descending" series followed an up-to-bottom trajectory. More than 30 years later, Roffler and Butler (1968) replicated again Pratt's (1930) results with nine tones (250, 400, 600, 900, 1400, 2000, 3200, 4800 and 7200 Hz), and showed that the association between pitch height and vertical locations persisted even when the distance of the viewer from the panel and the viewer's orientation were manipulated. They also found similar results in congenitally blind people and in 4–5 year-old children who, according to their self report, were unaware of the use of the words "high" and "low" to describe differently pitched sounds.

The major shortcoming of Pratt, (1930), Trimble (1934) and all the following studies claiming for a spatial representation of pitch height (e.g. Mudd, 1963; Roffler & Butler, 1968) was that participants were explicitly instructed to estimate the spatial position of sounds. For example, Mudd (1963) tested pitch height for "associative spatial stereotype", following Fitts and Deininger (1954) who had defined the operational measurement of population stereotype as the determination of the relative frequency with which each permissible response is made to a stimulus, when instructions do not specify what response is considered as appropriate. On each trial, Mudd's participants listened to two stimuli of different frequency. After listening to the stimulus pair, they were asked to move a plug from the centre reference hole in a pegpanel, whose position was assumed to correspond to the first stimulus, and plug the peg into another hole in the panel to represent the second stimulus. The spatial difference between the two pegholes was taken to represent the difference between the auditory stimuli being compared. Interestingly, Mudd found frequency to have both a horizontal and a vertical associative spatial stereotype (i.e. it was multidimensional), the latter being considerably stronger. Higher-frequency pitches were assigned to right/up locations and lower-frequency pitches were assigned to left/down locations. The procedure used to measure the "associative spatial stereotype" was considered adequate by Mudd (1963) on the basis of two arguments: (a) the verbal instructions emphasised that participants could and should plot anywhere on the panel and (b) the cartoon illustrations used to facilitate understanding of the task were designed to avoid, or at least to hold constant, any such tendency to establish any response bias.

However, the instructions requested that pitch height be intentionally represented in terms of visuo-spatial metrics. As a consequence, one cannot say whether pitch height is spontaneously or mandatorily mapped onto space, and whether the assignment of high-frequency pitches to high (or right) locations and low-frequency pitches to low (or left) locations generalizes to a context where participants' main concern is not to locate or to represent pitches in space.

In addition to the theoretical relevance of the question as to whether pitch height possesses intrinsic spatial characteristics, knowledge of the mental representation of pitch may have applications in ergonomics. A general rule for interface design, indeed, should be to maintain correspondence when possible. Moreover, although compatibility effects decrease in magnitude with practice, significant effects remain that do not disappear even with extended practice and stimulus-response incompatibility cannot be overcome by training (Dutta & Proctor, 1992). In the present study, we tested the spatial representation of pitch height in more restrictive conditions than previous works did. We asked whether pitch height would prime and speed up manual spatial responses, when the task does not require consideration of spatial location of the stimuli.

We explored the spatial representation of pitch height through the pairing of pitch to different response positions. In speeded choice reaction tasks, whenever *dimensional overlap* occurs between stimuli and responses, the mapping of stimuli to responses influences speed and accuracy of performance (Kornblum, Hasbroucq, & Osman, 1990), a phenomenon known as stimulus-response compatibility (SRC), and which has been a matter of relevant interest both in ergonomics and in experimental psychology since the pioneering studies of Fitts and colleagues (Fitts & Deininger, 1954; Fitts & Seeger, 1953). Dimensional overlap between stimuli and responses can occur in their spatial location, so that stimuli appearing on the right are responded to faster with a right than a left response key (Fig. 1A), and stimuli appearing up are responded to faster with an upper than a lower response key, irrespective of whether the keys lie on a frontal or on a transverse plane (Cho & Proctor, 2003; Vu, Proctor, & Pick, 2000).

Spatial SRC effects with stimuli accessing a mental spatial representation and responses mapped in the physical space have also been reported. Dehaene, Bossini, and Giraux (1993) and Dehaene, Dupoux, and Mehler (1990) showed that, with parity or magnitude judgments on centrally presented numbers, responses to large numbers are faster with the right than the left key, whereas responses to small numbers are faster with the left than the right key, irrespective of the responding hand (*Spatial–Numerical Association of Response Codes* or SNARC effect; Fig. 1B). Therefore, number magnitude is thought to be mapped onto a spatial representation, each number being represented as a point on a left-right mental line.

We reasoned that mapping sets of elements onto spatial positions might represent a useful cognitive tool across different domains (see also Gevers, Reynvoet, & Fias, 2003), and if spatial codes were assigned to pitch height, performance would result better when pitch (cognitive) location corresponded to response location than when it did not (Fig. 1B).
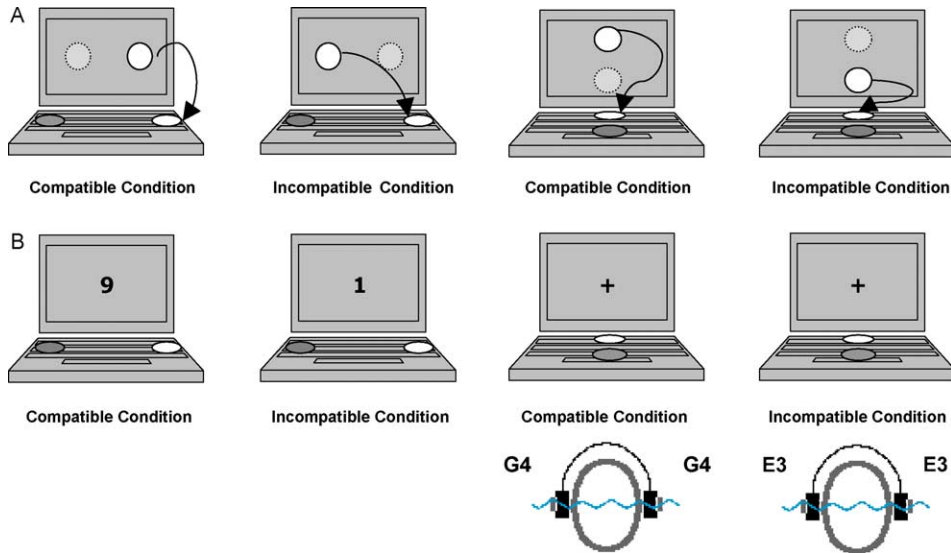
Fig. 1. SRC effects when stimuli and responses share a spatial dimension. Stimuli and correct responses are shown in white. (A) Physical spatial SRC: horizontal alignment is shown on the left and vertical alignment is shown on the right. (B) Representational spatial SRC: Spatial–Numerical Association of Response Codes (SNARC) on the left and Spatial–Musical Association of Response Codes (SMARC) on the right.

2. Experiment 1

With Experiment 1 we aimed at testing whether any SRC effect was present when pitch height and response position were varied orthogonally, in a task where nonmusicians were asked to compare the frequency (and not the spatial location) of two pure tones. Higher/lower judgments, when referred to the perceived location of a sound source as in Pratt's original work, call into question a different concept than higher/lower judgements referring to pitch height. Polysemantic words, such as *bank*, refer to different concepts depending on the context (e.g. an "institution offering services such as the safekeeping and lending of money" or "sloping ground at the side of a river"), thus a superficial similarity (the use of the same word) does not imply the existence of a deep similarity (the same meaning). Moreover, context can select for one of the meanings (e.g. Marcel, 1983). If spatial coordinates were spontaneously assigned to pitch height, we would expect performance to be significantly faster when participants responded to high-frequency pitches with an upper key and to low-frequency pitches with a lower key. We also investigated the mapping of pitches horizontally in cognitive space with left and right responses to test previous claims of multidimensionality (see Section 1 and Mudd, 1963). Mappings and responding hands were varied orthogonally within participants; that is, each mapping was performed with both uncrossed and crossed arms. It is worth recalling that the task did not require participants to locate a sound source in space nor to provide an explicit visuo-spatial representation for pitch height.

2.1. Method

2.1.1. Participants

Twenty participants (15 females and 5 males; median age: 21) took part in the experiment as paid volunteers and were recruited either in Hong-Kong or in London. All of them were Chinese, right-handed and had no experience of reading or playing music; none of them was tone deaf or reported hearing problems. Most of them spoke Cantonese as their first and English as their second language.

2.1.2. Apparatus, stimuli and procedure

Participants were tested individually in an undisturbed room. Their task was to compare the pitch of variable frequency tones (164.81, 185.00, 207.65, 233.08, 293.66, 329.63, 369.99, and 415.30 Hz, corresponding to *E3*, *F3#*, *G3#*, *A3#*, *D4*, *E4*, *F4#*, *G4#*, respectively) with that of a fixed reference (261.63 Hz, corresponding to *C4*). Both comparison and reference tones had a duration of 1000 ms, and had a constant amplitude. Stimuli were presented through headphones connected to a laptop computer, at a comfortable listening level.

On each trial participants fixated a cross in the centre of the computer screen and the probe tone was always preceded by the reference tone. Participants had to respond whether it was higher or lower than the reference by pressing one of two keys (Q and P or spacebar and 6[1]) with their left or right index finger. The deadline for response was fixed at 1300 ms; at the end of each trial visual feedback on accuracy was provided for 750 ms. Then the next trial began and the reference was played 250 ms after fixation onset.

The experiment was divided into two sessions corresponding to the two response alignments. Participants were tested with both their arms uncrossed and crossed. This manipulation was introduced to distinguish between pitch height-response location compatibility and pitch height-responding hand compatibility (an effect which might indicate a difference in the processing of pitch height between cerebral hemispheres; see Deutsch, 1995, 1999). In the horizontal alignment, the crossed-arm condition was assumed to occur when participants were to press the left response key with their right index finger and the right response key with their left index finger; in the vertical alignment, it was assumed to occur when participants were to press the lower response key with their right index finger and the upper response key with their left index finger. The order of mappings and arm conditions was counterbalanced within each alignment and each participant received the same order of conditions in the horizontal as in the vertical alignment. Half the participants started with the vertical and half with the horizontal alignment. There were four blocks (2 mappings×2 arm conditions) of 80 trials each for each alignment and participants were encouraged to take short breaks between blocks.

---

[1] *Spacebar* and *6* (on a standard British qwerty keyboard) will be, respectively, referred to as the lower and the upper response key. However, note that their verticality is only metaphorical, for they lie on a transverse plane relative to participants. Vu et al. (2000), showed that, in tasks with vertically aligned stimuli, spatial SRC effects are found irrespective of whether the response keys are aligned along the actual (frontal plane) or the metaphorical (transverse plane) vertical.

2.2. Results and discussion

RTs faster than 200 ms (1.25% of the total trials) were excluded from subsequent analyses. Both mean correct RTs ($M = 657$ ms; SE$= 12$) and number of errors (8.09% of the total trials) in each alignment were submitted to a $2 \times 4 \times 2 \times 2$ repeated measures ANOVA with Pitch Height (lower, higher), Distance from the reference (1, 2, 3, 4 tones), Responding Hand (left, right) and Response Location (up, down or left, right) as factors.

Globally, in the vertical alignment the absolute mean RT was 664 ms (SE$= 13$) and the error rate was 8.65%. A significant main effect of Distance was found, with faster RTs [$F(3,57) = 104.83$, MSE$= 5021$, $P < .001$] and higher accuracy [$F(3,57) = 55.17$, MSE$= 1.245$, $P < .001$] for larger distances. The Pitch Height$\times$Response Location interaction was significant [$F(1,19) = 7.10$, MSE$= 26,438$, $P = .015$], revealing a 34.5-ms advantage for up responses to high-frequency and down responses to low-frequency pitches (compatible mapping) relative to down responses to high-frequency and up responses to low-frequency pitches (incompatible mapping; see Fig. 2, left panel). The same interaction was significant in the error analysis [$F(1,19) = 5.27$, MSE$= 1.495$, $P = .033$], showing that the compatible was more accurate than the incompatible mapping. No main effects or interactions involving Responding Hand were significant.

Globally, in the horizontal alignment the absolute mean RT was 651 ms (SE$= 12$) and the error rate was 7.53%. A significant main effect of Distance was found also in the horizontal alignment [mean correct RTs: $F(3,57) = 121.49$, MSE$= 4510$, $P < .001$; number of errors: $F(3,57) = 54.71$, MSE$= 1.598$, $P < .001$], with faster RTs and higher accuracy for larger distances. The Pitch Height$\times$Response Location interaction only approached significance in the RTs analysis [$F(1,19) = 4.19$, MSE$= 10,629$, $P = .054$], revealing a 16.5-ms advantage for right responses to high-frequency and left responses to
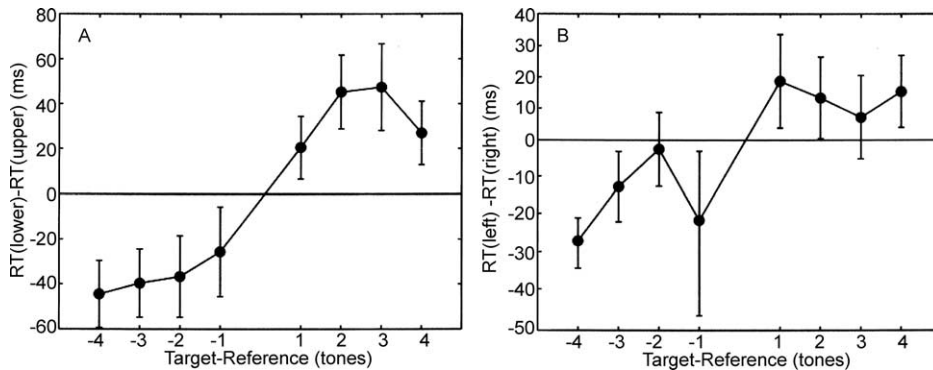

Fig. 2. Pitch comparison task (nonmusicians). On the left, mean differences between lower and upper key RTs as a function of distance from the reference are shown along with their standard error. On the right, mean differences between left and right key RTs as a function of distance from the reference are shown along with their standard error. In Experiment 1, the reference was *C4* and preceded on every trial a probe tone (*E3*, *F3#*, *G3#*, *A3#*, *D4*, *E4*, *F4#*, or *G4#*). Participants were faster to respond "lower" with the lower and the left response key, whereas they where faster to respond "higher" with the upper and the right response key. However, the interaction Pitch Height$\times$Response Location was fully significant in the vertical alignment and showed only a trend toward significance in the horizontal alignment.

low-frequency pitches (see Fig. 2, right panel). The same interaction was not significant in the error analysis ($F < 1$). No main effects or interactions involving Responding Hand were significant.

Experiment 1 revealed the presence of a clear-cut SRC effect with vertically aligned responses to pitch height. Performance was faster and more accurate when participants responded to high-frequency pitches with the upper key and to low-frequency pitches with the lower key, which indicates dimensional overlap (Kornblum et al., 1990) and a preferential mapping between tones having variable pitch and vertically aligned responses. Also, performance tended to be faster when participants responded to high-frequency pitches with the right key and to low-frequency pitches with the left key, which partially confirms that pitch height has a multidimensional spatial representation (Mudd, 1963). However, it might also suggest the presence of a vertical-to-horizontal (orthogonal) SRC effect. Indeed, when stimuli vary along the vertical and responses vary along the horizontal dimension in physical space, there is an advantage for the up/right–down/left relative to the up/left–down/right mapping, so that spatially high stimuli are responded to faster and more accurately with right responses, whereas spatially low stimuli are responded to faster and more accurately with left responses (for a review see Cho & Proctor, 2003). In the case of pitch height, the orthogonal SRC effect hypothesis would exactly predict better performance with right responses to high-frequency pitches and left responses to low-frequency pitches than the opposite.

By asking participants to compare the height of probe pitches to a reference, the obvious advantage was that pitch height had to be processed attentively before selecting a response. The drawback consisted of the use of spatial words ("higher" and "lower") as response alternatives that, mapped onto response locations, may induce SRC-like effects. Thus, in Experiments 2 and 3 we employed an indirect task, which has the advantage of eliminating any systematic confounds related to response labels.

3. Experiment 2

In Experiment 2, nonmusicians required to classify sounds as being produced by wind or percussion instruments (i.e. they had to perform a musical instrument identification task). To proceed with the analogy between music psychology and mathematical cognition, our identification task can be regarded as the equivalent of a parity judgment task. In a parity judgment task participants are required to decide whether a number is even or odd, instead of processing number magnitude. Nonetheless they are faster in responding to large numbers with the right key and to small numbers with the left key (Dehaene et al., 1993), which implies that a spatial response can be primed by magnitude even when it is irrelevant to the task. Numbers are classified as large or small relative to the contingent central value (implicit reference). For example, if participants are to judge the parity of numbers from 1 to 9, 5 is the implicit reference, numbers from 1 to 4 are small, and numbers from 6 to 9 are large. In the identification task we used equally spaced pitches between *F3#* and *A4#*, and *D4* was our implicit reference. Thus, pitches from *F3#* to *C4* were low frequency and pitches from *E4* to *A4#* were high frequency. No SRC effects should be present in Experiment 2, if (a) correspondence between response labels and key

positions was the sole explanation for the results of our Experiment 1 and (b) the spatial representation of pitch height was an artefact induced by experimental instructions in all the previous works (e.g. Mudd, 1963; Pratt, 1930).

3.1. Method

3.1.1. Participants

Twenty participants (9 females and 11 males; median age: 21) took part in the experiment as paid volunteers and were all recruited in Hong-Kong. All but two of them were right-handed. None of them was tone deaf or reported hearing problems. They had no experience of reading or playing music and had not participated in Experiment 1. Most of them spoke Cantonese as their first and English as their second language.

3.1.2. Apparatus, stimuli and procedure

Participants were required to identify the instrument family of variable frequency tones (174.61, 207.65, 233.08, 261.63, 329.63, 349.23, 369.99, 415.30 and 466.16 Hz, corresponding to *F3#*, *G3#*, *A3#*, *C4*, *E4*, *F4#*, *G4#* and *A4#*, respectively; implicit reference: 293.66 Hz, corresponding to *D4*). For each level of frequency, we used four sample sounds played by different musical instruments, making up a total of 36 stimuli. Two sets of stimuli were played by wind instruments (French Horn and Tenor Trombone) and the other two sets were played by percussions (Marimba and Vibraphone). Stimuli duration was equalized to 1000 ms by applying a linear amplitude ramp of 10 ms. Stimuli were equalized in loudness by one of the authors and samples were extracted from the McGill University Master Samples CD (Opolko & Wapnick, 1987). Stimuli were presented through headphones connected to a laptop computer, at a comfortable listening level.

On each trial, participants fixated a cross in the centre of a computer screen while the stimulus was presented. They responded whether it was played with a percussion, or wind instrument by pressing one of two keys (Q and P or spacebar and 6). The deadline for response was fixed at 1300 ms and at the end of each trial feedback on accuracy was provided for 750 ms. Then, the next trial began with a sound played 250 ms after fixation onset. The experiment was divided into two sessions corresponding to the two response alignments and participants were tested both with their arms parallel and crossed. Each of the eight experimental blocks comprised 160 trials. Everything else was identical to Experiment 1.

3.2. Results and discussion

RTs faster than 200 ms (2.8% of the total trials) were excluded from the analyses. Both mean correct RTs (global $M = 613$ ms; SE $= 25$) and the number of errors (5.54% of the total trials) in each alignment were submitted to a $2 \times 2 \times 4 \times 2 \times 2$ repeated measures ANOVA with Instrument (percussion, wind), Pitch Height (lower, higher), Distance from the reference (1, 2, 3, 4 tones), Responding Hand (left, right) and Response Location (up, down or left, right) as factors.

Globally, in the vertical alignment the absolute mean RT was 615 ms (SE $= 25$) and the error rate was 6.12%. In the RTs analysis for vertical alignment the main effects of
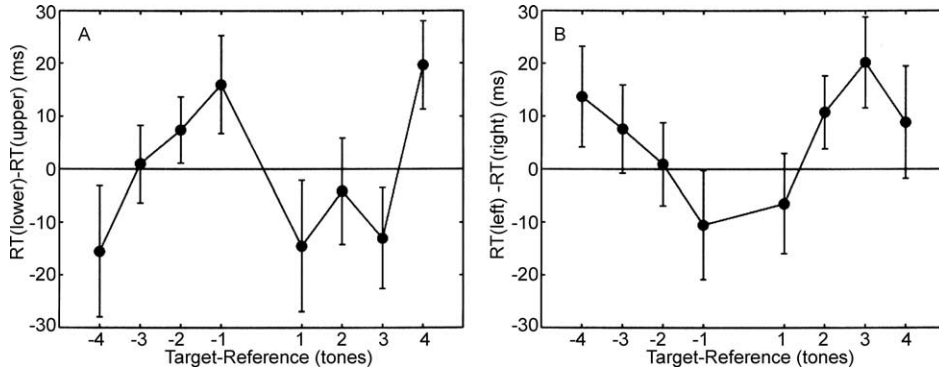
Fig. 3. Instrument identification task (nonmusicians). On the left, mean differences between lower and upper key RTs as a function of distance from the implicit reference are shown along with their standard error. On the right, mean differences between left and right key RTs as a function of distance from the implicit reference are shown along with their standard error. In Experiment 2, the implicit reference was *D4* and was never heard by participants. Targets (*F3#*, *G3#*, *A3#*, *C4*, *E4*, *F4#*, *G4#*, and *A4#*) were classified on the basis of their timbre whereas pitch height was task irrelevant and varied orthogonally to timbre categories. Nevertheless, in the vertical alignment participants showed, at the largest distance, an SRC effect consistent with that found in Experiment 1 and, at the shortest distance, a trend toward a reversed SRC.

Instrument [$F(1,19)=23.66$, MSE$=17,329$, $P<.001$] and Distance [$F(3,57)=4.09$, MSE$=4465$, $P=.011$] were significant, as the Instrument×Distance two-way interaction [$F(3,57)=4.95$, MSE$=5283$, $P=.004$]. Percussion instruments tones were identified 36-ms faster than wind instruments tones, and the difference between instruments increased with distance. However, the significant main effect of Instrument [$F(1, 19)=12.51$; MSE$=.649$, $P=.002$] and the Instrument×Distance interaction that approached significance in the error analysis [$F(3, 57)=2.18$; MSE$=.577$, $P=.099$] pointed in the opposite direction, suggesting the presence of a speed-accuracy trade-off. The Pitch Height×Distance×Response Location interaction was significant in the RTs analysis [$F(3,57)=4.85$, MSE$=3204$, $P=.044$], with a 17.5-ms SRC effect at the 4-tone distance only [$t_{(19)}=2.78$, $P=.012$]. Even though not significant, the SRC effect was reversed at the 1- to 3-tone distances (see Fig. 3, left panel), which explains why the Pitch Height× Response Location interaction was far from significance ($F<1$).

Globally, in the horizontal alignment the absolute mean RT was 611 ms (SE$=25$) and the error rate was 4.94%. No significant interactions between Pitch Height and Response Location emerged for horizontal alignment in either the RTs or the error analysis (see Fig. 3, right panel).

Thus, in the absence of spatial words a SRC effect in the same direction as that found in Experiment 1, and consistent with previous studies (e.g. Mudd, 1963; Pratt, 1930), persisted with vertical responses only at the 4-tone distance from implicit reference (i.e. *F3#* primed low responses and *A4#* primed up responses, but the intermediate pitches were not reliably associated to spatial representations). This indicates that a vertical representation of pitch height influences behaviour in nonmusicians only when pitches are separated by more than an octave (*F3#* and *A4#*, indeed, are separated by 16 pitches from

each other and by eight pitches from the implicit reference *D4*). Presumably, very large differences in pitch height are more likely than small ones to be processed automatically during a musical instrument identification task. Our hypothesis does not imply that participants were unable to perceive smaller differences (also in consideration of the fact that their mother tongue is a tone language; see Deutsch, 2002; Deutsch, Henthorn, & Dolson, 1999, 2004), rather only that pitch height could interfere with the task at hand only when having extreme values. Therefore, we included in the study a group of musicians to ascertain whether the effect depended on musical expertise and processing automaticity for pitch height.

4. Experiment 3

Rationale and stimuli were identical to Experiment 2, but participants were trained musicians (i.e. they could sight-read scores).

4.1. Method

4.1.1. Participants
Twenty participants (14 females and 6 males; median age: 21) took part in the experiment as paid volunteers and were all recruited in London. All of them were Chinese and right-handed. They spoke Cantonese as their first and English as their second language.

4.1.2. Apparatus, stimuli and procedure
They were identical to those in Experiment 2.

4.2. Results and discussion

RTs faster than 200 ms (1.54% of the total trials) were excluded from the analyses. Both mean correct RTs (global $M=589$ ms; SE$=25$) and the number of errors (4.41% of the total trials) in each alignment were submitted to a $2\times2\times4\times2\times2$ repeated measures ANOVA with Instrument (percussion, wind), Pitch Height (lower, higher), Distance from the reference (1, 2, 3, 4 tones), Responding Hand (left, right) and Response Location (up, down or left, right) as factors.

Globally, in the vertical alignment the absolute mean RT was 600 ms (SE$=25$) and the error rate was 4.48%. The main effects of Instrument [$F(1,19)=23.66$, MSE$=17,329$, $P<.001$], Pitch Height [$F(1,19)=34.93$, MSE$=2138$, $P<.001$] and Distance [$F(3,57)=4.09$, MSE$=4465$, $P=.011$] were significant. However, the most interesting result was the significant Pitch Height$\times$Response Location interaction both in the RTs analysis [$F(1,19)=7.67$, MSE$=4003$, $P=.012$] and in the error analysis [$F(1,19)=6.57$, MSE$=.616$, $P=.019$]. The RTs analysis revealed a 10-ms advantage for low-frequency pitches responded to with the lower key and high-frequency pitches responded to with the upper key (see Fig. 4, left panel). The error analysis pointed in the same direction.
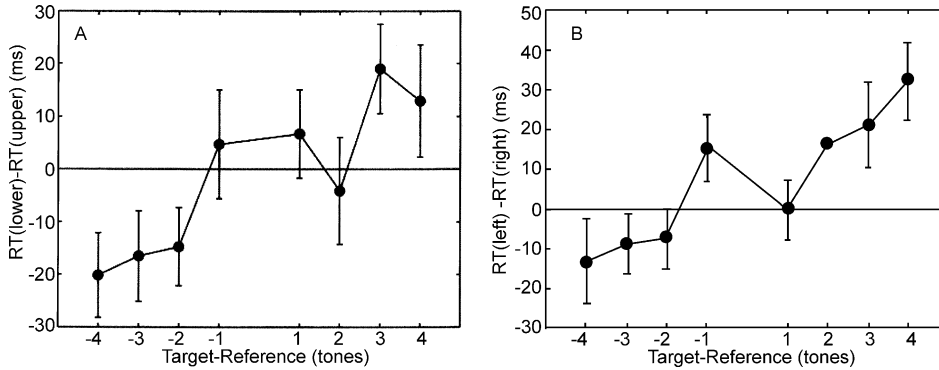
Fig. 4. Instrument identification task (musicians). On the left, mean differences between lower and upper key RTs as a function of distance from the implicit reference are shown along with their standard error. On the right, mean differences between left and right key RTs as a function of distance from the implicit reference are shown along with their standard error. In Experiment 2, the implicit reference was *D4* and was never heard by participants. Targets (*F3#*, *G3#*, *A3#*, *C4*, *E4*, *F4#*, *G4#*, and *A4#*) were classified on the basis of their timbre whereas pitch height was task irrelevant and varied orthogonally to timbre categories. Nevertheless, Pitch Height, Distance from the reference and Response Location interacted in the horizontal alignment, showing that the SRC effect emerged at the two largest distances; Pitch Height and Response Location interacted in the vertical alignment, showing that participants were faster to respond with the upper key to high-pitch targets and they were faster to respond with the lower key to low-pitch targets.

Globally, in the horizontal alignment the absolute mean RT was 577 ms (SE = 25) and the error rate was 4.34%. The main effect of Instrument was significant in the RTs analysis [$F(1,19) = 20.65$, MSE = 25,665, $P < .001$], and the main effect of Distance was significant in the error analysis [$F(3,57) = 5.17$, MSE = 2138, $P = .003$]. Moreover, the Pitch Height × Response Location interaction was significant in the RT analysis [$F(1,19) = 6.76$, MSE = 5231, $P = .018$], with a 10.5-ms advantage for high-frequency pitches responded to with the right key and left-frequency pitches responded to with the left key. The interaction pointed in the same direction and approached significance in the error analysis [$F(1,19) = 3.69$, MSE = .573, $P = .070$]. The effect was further qualified by a significant Pitch Height × Distance × Response Location interaction [$F(3,57) = 5.06$, MSE = 2873, $P = .003$] in the RTs analysis, which approached significance in the error analysis [$F(3,57) = 3.03$, MSE = .473, $P = .051$]. Planned comparisons on mean correct RTs revealed that the horizontal SRC effect was significant at the 3-tone [10 ms; $t_{(19)} = 2.26$, $P = .036$] and 4-tone distances [23.5 ms; $t_{(19)} = 3.29$, $P = .004$] only (see Fig. 4, right panel).

Because Experiment 3 suggested the importance of expertise in producing SRC effects with pitch height, Musical Expertise was then entered as a between-participant factor into an ANOVA, having Distance from the reference (1, 2, 3, 4 tones), Pitch Height (lower, higher) and Response Location (up, down or left, right) as within-participant factors, on the combined RTs and error data from Experiments 2 and 3. If Musical Expertise did play a role, it should interact with the Pitch Height × Response Location effect.

In the vertical alignment, the main effect of Musical Expertise was far from significance (Fs < 1) both in the RTs and in the error analysis. Musical Expertise was involved in a significant three-way interaction with Pitch Height and Response Location in the RT

analysis [$F(1,38)=5.616$, MSE$=1158$, $P=.023$], thus confirming the presence of a causal relation between expertise and the vertical SRC effect (global SRC effect for musicians: $t_{(38)}=2.66$, $P=.011$; global SRC effect for nonmusicians: $t<1$; see analyses above). In addition, the Distance$\times$Pitch Height$\times$Response Location three-way interaction was significant [$F(3,114)=4.754$, MSE$=.881$, $P=.004$], indicating that the vertical SRC effect was present in both groups at a distance of four tones from the implicit reference ($t_{(38)}=3.46$, $P=.001$). The error analysis revealed a significant Pitch Height$\times$ Response Location two-way interaction [$F(1,38)=5.56$; MSE$=2.924$, $P=.024$] and a significant Distance$\times$Pitch Height$\times$Response Location interaction [$F(3,114)=4.22$; MSE$=1.218$, $P=.007$], which mirrored the RTs analysis (significant SRC effect at the 4-tone distance: $t_{(38)}=3.50$, $P=.001$), whereas Musical Expertise did not interact significantly with any of the other factors.

In the horizontal alignment, the main effect of Musical Expertise was not significant ($F<1$) in either analysis. The Pitch Height$\times$Response Location interaction was significant in the RTs analysis [$F(1,38)=7.638$, MSE$=914$, $P=.009$] and approached significance in the error analysis [$F(1,38)=3.501$, MSE$=2.785$, $P=.069$]. Musical Expertise was involved in a significant four-way interaction with Distance, Pitch Height and Response Location in the error analysis [$F(3,114)=3.423$, MSE$=2.105$, $P=.020$], which approached significance in the RTs analysis [$F(3,114)=2.535$, MSE$=799$, $P=.060$]. In both cases, the SRC effect resulted significant at large distances and for musicians only (error analysis, 4-tone distance: $t_{(38)}=2.64$, $P=.012$; RT analysis, 3-tone distance: $t_{(38)}=2.13$, $P=.040$, and 4-tone distance: $t_{(38)}=3.07$, $P=.004$).

In summary, musical expertise interacted significantly with the SRC effect, both with a vertical (in the RTs analysis) and with a horizontal (in the error analysis) alignment of response keys. In either case, musical expertise favoured the presence of the SRC effect, to which henceforth we will refer as the SMARC (*Spatial–Musical Association of Response Codes*) effect. Our initial hypothesis concerned the mapping of pitch height onto a vertical spatial dimension, as suggested by linguistic clues, however, the presence of a horizontal SMARC effect was also detected in musicians, which might either be attributed to orthogonal mapping (see Cho & Proctor, 2003, for a review) or to musicians' familiarity with keyboards (on which low-frequency pitches are played by pressing leftward keys and high-frequency pitches are played by pressing rightward keys). Recently, a study by Stewart, Walsh, and Frith (2004) showed that numbers associated with manual responses lead to Stroop like interference in trained musicians, when numbers are either horizontally or vertically displayed on a computer screen. In two experiments they asked participants to respond to the numbers 1, 2, 3, 4, and 5, respectively, with the thumb, index, middle, ring and little finger of their right hand, consistently with the convention used in keyboards music notation. In Experiment 1, pianists performed better when small numbers were superimposed to notes on lower locations of the staff (which is the set of horizontal lines used in sheet music) and large numbers were superimposed to notes on higher locations of the staff. In Experiment 2, the association of small numbers with low or left locations and large numbers with high or right locations improved pianists' performance, although staff and notes did not appear on the display. Stewart et al. (2004) concluded that pianists possess a specific vertical-to-horizontal stimulus-response mapping and attributed it to musical training. "Pianists were faster when stimuli specifying a leftward response were

presented in vertically lower locations and stimuli specifying a rightward response were presented in vertically higher locations. Nonmusicians showed the reversed pattern. No group differences were found on a task that required horizontal-to-horizontal mappings" (Stewart et al., 2004, p. 183). Although the hypothesis is reasonable and intuitively appealing, in that study only the association between numbers and locations was systematically varied, whereas the mapping of numbers to response positions was kept constant (i.e. 1 was always responded to with the leftmost finger and 5 with the rightmost finger; never the opposite). As a consequence Stewart et al. (2004) documented a vertical stimulus-stimulus correspondence effect (Stroop-like effect; see also the classification of Kornblum et al., 1990) between number magnitude and visual locations, rather than a SRC effect between the height of visual stimuli and horizontal response locations.

In our Experiment 3, only half of the participants had piano as their primary musical specialization, the other half being specialized in instruments on which high-frequency tones are not mapped high, far or right in space (e.g. cello, viola, violin, guitar or flute players), and with only a secondary experience with piano. Thus, on the basis of the current literature, we will maintain that our horizontal SMARC effect is a variant of the well-known orthogonal SRC effect, that is a preferential mapping of spatially lower stimuli on left responses and higher stimuli on right responses (Cho & Proctor, 2003).

5. Conclusion

We asked a group of musically naïve participants to perform a pitch comparison task, a different group of musically naïve participants and a group of musicians to perform a musical instrument identification task on sounds having different pitch. A SMARC effect (i.e. high-frequency pitches favouring up responses and low-frequencies pitches favouring down responses) was present both when pitch was task relevant, and when it was task irrelevant. Moreover, when pitch height was task irrelevant, a horizontal mapping of pitches appeared for musicians only.

In conclusion, we found that a representational dimension (pitch height) influences performance with vertically aligned responses irrespective of its relevance to the task. This suggests that our cognitive system maps pitch onto a mental representation of space. Our results are thus consistent with studies pointing to the integral nature of spatial and spectral processing of auditory stimuli. For example, responding to the location of a tone is affected by irrelevant variation in its pitch—and vice versa—when a Garner's paradigm is used (Mondor, Zatorre, & Terrio, 1998). This would demonstrate that the selection of auditory information may not be guided independently via location and frequency channels and that the integration of spatial and spectral attributes may occur at an early stage of processing (for neuroanatomical correlates see the PET study of Zatorre, Bouffard, Ahad, & Belin, 2002). In addition, we showed that even when the physical location of a sound source is kept constant, pitch still interacts with space by affecting speed and accuracy of manual responses through a preferential pairing with response locations. We called this phenomenon the SMARC effect, in analogy to the SNARC effect, which has been first, described with numbers (Dehaene et al., 1993) and subsequently with other abstract sequences (Gevers et al., 2003). However, the fact that both pitch height and number

magnitude produce spatial SRC effects does not imply that numbers and tones follow identical sensorimotor transformations in the cognitive system. In fact, the link between musical, spatial and numerical skills -although popular with the general public- is an open question (e.g. Schellenberg, 2001); so the proposed analogy between the SMARC and the SNARC effects may indicate a research direction that draws them together in a more rigorous way.

In Section 1 we also mentioned that the influence of SRC on performance, is of practical importance in ergonomics. Our findings suggest straightforward applications to the design of interfaces where pressing one key among a series can be associated with a sound. For example, on the numerical keypad of a computer keyboard, numbered keys are ordered from left to right and from bottom to up (i.e. the 1 key is down, on the left corner of the keypad and the 9 key is up, on the right corner), whereas on home and mobile phones they are usually ordered from left to right and from up to bottom (i.e. the 1 key is up, on the left corner and the 9 key is down, on the right corner). This incompatibility between mappings might be modulated by associating high-frequency tones to higher-key pressings and low-frequency tones to lower-key pressings. In the case of the numerical keypad, this would be useful for nonmusicians, who seem to associate small numbers with high positions and large numbers with low positions in the visual field (Stewart et al., 2004); in the case of the phone, it would be useful for musicians, who have been shown to possess the opposite association between numbers and space (Stewart et al., 2004).

The results for the horizontal response alignment call for additional investigations, as the study with nonmusicians led to inconclusive findings (the only hint to a horizontal SMARC effect is a trend in Experiment 1) and the study with musicians, in which a significant horizontal SMARC effect was found at large distances, led to question whether remapping had occurred from the vertical to the horizontal dimension. Evidence favouring the hypothesis of a real multidimensional association between pitches and space can be found in the glissando illusion (Deutsch, 1995, 1999; Deutsch, Hamaoui, & Henthorn, in preparation). The stimulus pattern which generates this illusion consists of two parts: a synthesized oboe tone switching back and forth between a left and a right loudspeaker, and a sine wave tone gliding slowly up and down in pitch and switching back and forth between the two loudspeakers. Whenever the oboe tone is coming from the speaker on the right, a portion of the glissando is coming from the speaker on the left, and vice versa. Almost everybody perceives the synthesized oboe tone correctly as switching back and forth between the speakers and the glissando as joined together, even thought it is in reality switching back and forth between the speakers. Righthanders hear the glissando traversing an elliptical path from low down (in space) on the left when it is lowest in pitch, moving slowly to high up (in space) on the right as its pitch moves from low to high, and the moving back again to low down on the left as its pitch goes back from high to low (Deutsch et al., in preparation). Nonrighthanders, instead, are less consistent in the ways in which the glissando appears to move around in space, corresponding to the movement of pitch. The glissando illusion does produce a strong sense of spatial position depending on pitch height both in the vertical and in the horizontal dimension at the same time. Noticeably, like in other famous illusions relating low pitch to left space and high pitch to right space (octave and scale illusions; Deutsch, 1999), Deutsch (1995) did not find any

difference between nonmusicians and musicians, but she did find a difference between right and lefthanders.

Acknowledgements

References

Cho, Y. S., & Proctor, R. W. (2003). Stimulus and response representations underlying orthogonal stimulus-response compatibility effects. *Psychonomic Bulletin and Review*, *10*, 45–73.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*, 371–396.

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital: Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 626–641.

Deutsch, D. (1995). *Musical illusions and paradoxes*. La Jolla, CA: Philomel Records.

Deutsch, D. (1999). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music*2nd ed.. New York: Academic Press.

Deutsch, D. (2002). The puzzle of absolute pitch. *Current Directions in Psychological Science*, *6*, 200–204.

Deutsch, Hamaoui, K., & Henthorn, T. (in preparation).

Deutsch, D., Henthorn, T., & Dolson, M. (1999). Absolute pitch is demonstrated in speakers of tone languages. *Journal of the Acoustical Society of America*, *106*, 2267.

Deutsch, D., Henthorn, T., & Dolson, M. (2004). Absolute pitch, speech and tone language: Some experiments and a proposed framework. *Music Perception*, *21*, 229–256.

Dutta, A., & Proctor, R. W. (1992). Persistence of stimulus-response compatibility effects with extended practice. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *18*, 801–809.

Fitts, P. M., & Deininger, R. L. (1954). S-R compatibility: Correspondence among paired elements within stimulus and response codes. *Journal of Experimental Psychology*, *48*, 483–492.

Fitts, P. M., & Seeger, C. M. (1953). S-R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, *46*, 199–210.

Gevers, W., Reynvoet, B., & Fias, W. (2003). The mental representation of ordinal sequences is spatially organized. *Cognition*, *87*, B87–B95.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility. A model and taxonomy. *Psychological Review*, *97*, 253–270.

Marcel, A. J. (1983). Conscious and unconscious perception: An approach to the relations between phenomenal experience and perceptual processes. *Cognitive Psychology*, *15*, 238–300.

Melara, R. D., & Marks, L. E. (1990). Processes underlying dimensional interactions: Correspondences between linguistic and nonlinguistic dimensions. *Memory and Cognition*, *18*, 477–495.

Mondor, T. A., Zatorre, R. J., & Terrio, N. A. (1998). Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 66–79.

Moore, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). London: Academic press.

Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tone. *Journal of Experimental Psychology*, *66*, 347–352.

Opolko, F., & Wapnick, J. (1987). *McGill University master samples CD*. Montreal, Que.: McGill University.

Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, *13*, 278–285.

Roffler, S. K., & Butler, R. A. (1968). Localization of tonal stimuli in the vertical plane. *The Journal of the Acoustical Society of America*, *43*, 1260–1266.

Schellenberg, E. G. (2001). Music and nonmusical abilities. *Annals New York Academy of Sciences*, *930*, 355–371.

Shepherd, R. N. (1982). Geometrical approximations to the structure of musical pitch. *Psychological Review*, *89*, 305–333.

Stewart, L., Walsh, V., & Frith, U. (2004). Reading music modifies spatial mapping in pianists. *Perception and Psychophysics*, *66*, 183–195.

Stumpf, C. (1883). *Tonpsychologie*. Leipzig: Verlag von S. Hirzel.

Trimble, O. C. (1934). Localization of sound in the anterior posterior and vertical dimensions of auditory space. *British Journal of Psychology*, *24*, 320–334.

Ueda, K., & Ohgushi, K. (1987). Perceptual components of pitch: Spatial representation using a multidimensional scaling technique. *Journal of the Acoustical Society of America*, *82*, 1193–1200.

Vu, K.-P.L., Proctor, R. W., & Pick, D. F. (2000). Vertical versus horizontal spatial compatibility: Right-left prevalence with bimanual responses. *Psychological Research/Psychologische Forschung*, *64*, 25–40.

Zatorre, R. J., Bouffard, M., Ahad, P., & Belin, P. (2002). Where is "where" in the human auditory cortex? *Nature Neuroscience*, *5*, 905–909.