

Journal on Multimodal User Interfaces

Exploring crossmodal perceptual enhancement and integration in a sequence-reproducing task with cognitive priming

--Manuscript Draft--

Manuscript Number:	JMUI-D-18-00093R2
Full Title:	Exploring crossmodal perceptual enhancement and integration in a sequence-reproducing task with cognitive priming
Article Type:	Original Paper
Corresponding Author:	Feng Feng Queen Mary University of London UNITED KINGDOM
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Queen Mary University of London
Corresponding Author's Secondary Institution:	
First Author:	Feng Feng
First Author Secondary Information:	
Order of Authors:	Feng Feng Puhong Li Tony Stockman
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	<p>Crossmodal correspondence, a perceptual phenomenon which has been extensively studied in cognitive science, has been shown to play a critical role in people's information processing performance.</p> <p>However, the evidence has been collected mostly based on strictly-controlled stimuli and displayed in a noise-free environment.</p> <p>In real-world interaction scenarios, background noise may blur crossmodal effects that designers intend to leverage. More seriously, it may induce additional crossmodal effects, which can be mutually exclusive to the intended one, leading to unexpected distractions from the task at hand.</p> <p>In this paper, we report two experiments designed to tackle these problems with cognitive priming techniques. The first experiment examined how to enhance the perception of specific crossmodal stimuli, namely pitch-brightness and pitch-elevation stimuli.</p> <p>The second experiment investigated how people perceive and respond to crossmodal stimuli that were mutually exclusive.</p> <p>Results showed that first, people's crossmodal perception was affected by cognitive priming, though the effect varies according to the combination of crossmodal stimuli and the types of priming material.</p> <p>Second, when two crossmodal stimuli are mutually exclusive, priming on only the dominant one (Pitch-elevation) lead to improved performance.</p> <p>These results can help inform future design of multisensory systems by presenting details of how to enhance crossmodal information with cognitive priming.</p>

Response to Reviewers:

Please see the uploaded 'jomui_list_of_corrections_2nd' document, which is in pdf format. We are sorry for the inconvenience since the submission system does not support table format. Many thanks.

Exploring crossmodal perceptual enhancement and integration in a sequence-reproducing task with cognitive priming

Received: date / Accepted: date

Abstract Crossmodal correspondence, a perceptual phenomenon which has been extensively studied in cognitive science, has been shown to play a critical role in people's information processing performance. However, the evidence has been collected mostly based on strictly-controlled stimuli and displayed in a noise-free environment. In real-world interaction scenarios, background noise may blur crossmodal effects that designers intend to leverage. More seriously, it may induce additional crossmodal effects, which can be mutually exclusive to the intended one, leading to unexpected distractions from the task at hand. In this paper, we report two experiments designed to tackle these problems with cognitive priming techniques. The first experiment examined how to enhance the perception of specific crossmodal stimuli, namely pitch-brightness and pitch-elevation stimuli. The second experiment investigated how people perceive and respond to crossmodal stimuli that were mutually exclusive. Results showed that first, people's crossmodal perception was affected by cognitive priming, though the effect varies according to the combination of crossmodal stimuli and the types

of priming material. Second, when two crossmodal stimuli are mutually exclusive, priming on only the dominant one (Pitch-elevation) lead to improved performance. These results can help inform future design of multisensory systems by presenting details of how to enhance crossmodal information with cognitive priming.

Keywords Crossmodal interaction · Cognitive priming · Crossmodal perception and integration · Sensory-motor performance

1 Introduction

Understanding how people integrate multimodal information to optimise sensory-motor activity has implications for multisensory system design. The multimodal perceptual phenomenon of crossmodal correspondence (CC) has long been investigated in the cognitive science research field. It refers to the associated perceptual relationship between two or more sensory modalities, from which information perceived through visual, auditory and other sensory channels are combined to form a stable output [41]. For example, an upward visual stimulus perceptually corresponds with a high pitch sound [39, 27]. Multisensory information presented with this regularity enables people to have better perceptual accuracy than that displayed in a reversed polarity, i.e. incongruent correspondence between, for instance, downward visual stimulus with a high pitch sound.

Despite mounting evidence showing that congruent CCs play a positive effect on sensory-motor performance [30, 13, 8], implementations with an effort from Human-computer interaction (HCI) have certain limitations due to the following reasons. On the one hand,

E-mail:

E-mail:

E-mail:

an environment that was controlled to maximise the internal validity of the experiment is different from a real-world scenario where interaction activities take place. In this latter scenario, streams of interactive information unavoidably mingle with physical and attentional background noise, which may act as visual, auditory or multisensory distractors. As a result, the CCs observed from the strictly-controlled, noise-free experiments may not function well when applied in a real-world scenario. On the other hand, many cases of crossmodal interaction, such as spatial localisation [19], graph sonification [3], and motor skill learning [17], involve consistent sensory-motor responses with continuous, and graded crossmodal feedback. We lack empirical knowledge on how such crossmodal stimuli influence interaction that requires consistent sensory-motor engagement.

Thirdly, real-world interaction requires not only a bottom-up perception of the crossmodal feedback, but top-down perception influenced by subjective knowledge, experiences and specific interaction situations [2, 22]. For example, in a target-searching task, the crossmodal association between colour and euclidean distance, i.e. an increase in R value (RGB scheme) associated with a decrease in the distance towards the target, was occasionally perceived as, on a semantic level, an association between a colour signifier and motion correctness, i.e. red indicated wrong moves, therefore associated with moving away from the target [16]. These two crossmodal associations raised a certain level of confusion for some participants, due to the mutually exclusive congruency.

To our knowledge, there has been little research into presenting crossmodal information that can be unequivocally perceived by participants without requiring extra cognitive effort. More specifically, we lack an understanding of whether the perception of specific crossmodal information can be enhanced or inhibited subliminally for an interaction task. Furthermore, when information contains mutually exclusive CCs (i.e. the perception of the congruency of one CC excludes the perception of the congruency of the other), we lack understanding of how people integrate such information.

Interestingly, the role of subliminal information has drawn increased attention in the HCI community, especially in the field of pervasive technology and mindless computing [4, 1, 37]. Cognitive priming, used for conveying subliminal information, refers to a technique which enhances cognitive performance without overloading cognitive channels. [26, 12, 4]. It showed an enhancement effect on various cognitive functions including perception [4, 37], judgement [21], and affectiveness [26] over the course of interaction. Given the effective-

ness of this technique, in this paper, we aim to investigate two research questions using cognitive priming:

RQ1: Whether people’s crossmodal perception can be enhanced by cognitive priming, and if it can be, to what extent it modulates small and fast sensory-motor responses in an interactive task?

RQ2: How do people integrate crossmodal information in which two CCs are mutually exclusive? Furthermore, how, if at all, does the integration of such information change in the presence or absence of cognitive priming?

We addressed these two questions by using the approach of cognitive priming as a manipulation factor [21, 25], with the purpose of engaging people with two CC pairs respectively, and investigating the influence of this upon subsequent interaction behaviour. The detailed research questions and the investigation scope is presented in section 3. Section 4 presents our investigation approach and experimental design. This is followed by the results of the experiments and a discussion in section 5 and 6 respectively.

The contributions of the present study include: firstly, a systematic investigation of two CCs, pitch-brightness and pitch-elevation, with graded stimuli in interactive tasks rather than two polarised stimuli for binary choices; secondly, an evaluation of the role of cognitive priming in crossmodal perceptual enhancement; thirdly, the paper employs two types of priming materials and investigates their effect on people’s sensory-motor performance. Lastly, we explored how people integrate mutually exclusive CCs under cognitive priming.

2 Related work

2.1 A view from cognition research and behavioural studies

Our perception of certain physical features or attributes are naturally associated with each other, such as the auditory perception of an increase and decrease in pitch is associated with the visual or haptic perception of rising and falling in vertical position [14, 35, 42]. These kinds of associations can be found in other cross-sensory features as well, including the sound-brightness association, sound-quantity association etc. [46, 14, 39, 18]. As such, the term crossmodal correspondence (CC) in this paper refers to ‘a compatibility effect between attributes or dimensions of a stimulus (i.e., an object or event) in different sensory modalities (be they redundant or not)’ [41]. Accumulating behavioural studies show that when values of these associated attributes are combined congruently, for example, increased pitch combined with increased vertical elevation, the accuracy

1 and efficiency of people’s sensory-motor response will
2 be improved (for a detailed review, see [41]). However,
3 if the attributes are combined incongruently (i.e. the
4 direction of change of one of the involved attributes is
5 reversed, for instance, when an increased pitch is as-
6 sociated with downward visual stimuli), sensory-motor
7 responses will be disturbed.

8
9 Some research studies have pointed out that many
10 CCs may occur due to the repeated exposure to the
11 natural environment, from which people learned per-
12 ceptual regularities of associated crossmodal physical
13 features [41,43]. Praise et al. empirically investigated
14 the statistical mapping between sound frequency and
15 the perceived elevation of the sound source [35]. With
16 a wearable two-directional headphone, a large sample
17 of natural sound recordings was collected by partici-
18 pants who move freely indoors and outdoors. The anal-
19 ysis of recordings reveals a consistent mapping between
20 the sound frequency and the physical elevation of the
21 source of the sound. Furthermore, when participants
22 were asked to do a sound localisation task, researchers
23 obtained a consistent observation concerning the frequency-
24 elevation correspondence. Even though the correlation
25 between the statistical environmental mapping and the
26 perceptual judgement performance cannot be fully ex-
27 plained for now, the above investigation provides exper-
28 imental ground for an ecological linkage between CCs
29 and daily experience [2,43].

30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
Recent research also suggests that perception of cross-
modal information can be influenced by task-irrelevant
contextual features [47,9]. Walker and Walker conducted
an experiment to address the relative perception of cross-
modal feature values. In their experiment, participants
were presented with six circles with different luminance.
Three of the circles were brighter than the background
colour, and the other three appeared darker. Partici-
pants were asked to classify whether each of the circles
was brighter or darker than the background, and
confirm their answer by pressing one of two differently-
sized keys, though participants neither visually nor hap-
tically aware of the difference in size. Results showed
that participants classify brighter circles more quickly
with the smaller key in hand, and classified darker cir-
cles more rapidly with the larger key in hand. This find-
ing suggests that CCs are not always absolute, and that
sensory-motor responses based on them could be influ-
enced subliminally by contextual stimuli, which in this
case, the haptic perception of the size of the keys. [47].

Moreover, the relativity of crossmodal perception
points to further investigation, with a focus on inter-
mediate crossmodal values that sit between pairs of
polarised stimuli [9,42]. Without fully understanding
the sequential perception of intermediate values, i.e.

graded crossmodal stimuli, the implementation of CCs
in real-world scenarios remains limited. After all, a va-
riety of crossmodal interaction cases involve consistent
perception-action loops, either for spatial localization
[19], graph sonification [3], or motor skill learning [17].

2.2 A view from HCI design implementation

From an HCI perspective, we are aware of two studies
that have directly investigated the design possibilities
with graded crossmodal stimuli. Metatla et al. evalu-
ated interaction performance with different crossmodal
congruency levels between shape, size and elevation [29].
They adopted a game mechanism, using a tablet to dis-
play a sequence of visual, auditory, or visual-auditory
stimuli. Participants were asked to tap in the perceived
sequence as quickly and accurately as possible. Results
showed that the visual condition produced better per-
formance than both the auditory and visual-auditory
conditions. However, in the bimodal condition where
two CCs co-existed, participants tended to rely on pitch-
elevation rather than visual stimuli as the primary CC.
This result implies that different CCs may have a dif-
ferent level of influence on interactive performance.

In another experiment, Feng and Stockman tried
to replicate the crossmodal effect through augmented
physical features of a tangible object on an interactive
tabletop [16]. Participants were required to discover a
hidden target by moving the object around the table,
and by observing the concurrent crossmodal feedback to
determine the distance to the target. Visual, auditory
and haptic modalities were combined into unimodal,
bimodal and trimodal feedback. Results revealed that
more accurate movement and efficient corrections were
achieved with bimodal and trimodal feedback. How-
ever, in the crossmodal condition which implemented
the mapping between colour and euclidean distance, i.e.
an increase in R value (RGB scheme) associated with
a decrease in the distance towards the target, one-sixth
participants appeared to have opposite mapping polar-
ity. This was due to their familiarity with a different
crossmodal association, the colour signifier and motion
correctness, i.e. red indicated wrong moves, therefore
associated with moving away from the target. These
two associations happened to have a mutually exclu-
sive mapping polarity. This result suggested that peo-
ple’s sensory-motor reaction may not solely be deter-
mined by crossmodal stimuli in a bottom-up manner,
but sometimes rather influenced by previous experience
and current interaction goals [2,22].

From a human-centred design perspective, researchers
have shown that CCs can be implicitly activated in
a natural interactive scenario. Bakker et al. applied a

1 human-centred iterative design approach during several
2 music education workshops, where groups of children
3 were encouraged to use body movements to express
4 sound features. Several auditory-haptic crossmodal as-
5 sociations have been identified from self-generated move-
6 ments, such as increased volume associated with in-
7 creased speed of movements, as well as rising postures
8 [5]. When implemented these identified CCs on hand-
9 held musical interfaces and applied to a subsequent mu-
10 sic learning session, children showed improved music
11 reproduction performance. However, we do not have
12 enough empirical evidence to determine whether the
13 improved performance was due to self-activated cross-
14 modal mappings or due to an improved familiarity pro-
15 duced by a series of iterative design activities.
16
17

20 2.3 Subliminal cueing and cognitive priming in HCI

21 Priming, in the fields of cognitive and social psychology,
22 generally refer to an experimental technique whereby
23 exposure to one stimulus influences a response to a sub-
24 sequent stimulus, without conscious guidance or inten-
25 tion [6]. It includes different priming mechanisms: the
26 conceptual priming activate an internal mental repre-
27 sentation in one context in such a way that the partic-
28 ipant does not realise the relation between that acti-
29 vation and the later influence in an unrelated task [6,
30 15]; the perceptual priming shares the same activation
31 mechanism, but the contextual features of priming have
32 direct indications for later tasks [6,37].
33
34

35 In the field of HCI, researchers applied different
36 priming techniques to facilitate interaction [1,25] and
37 enhance cognitive performance [26,12,21]. Two major
38 categories are implemented with empirical investiga-
39 tion. The technique of using subliminal stimuli to trig-
40 ger fast and automatic responses is referred to as sub-
41 liminal cueing, which adopts strict experimental control
42 over the factors of cueing time and content. The cueing
43 time is less than 50 ms, within which the perception is
44 assumed to be subliminal, and the cueing is commonly
45 an external stimulus presented as a visual image or ge-
46 ometric shape, which exposes information that would
47 appear during subsequent tasks. This technique was in-
48 herited from the masking paradigm from experimental
49 psychology, for the purpose of investigating the effect
50 of priming on selection behaviour [4,37]. Thus sublim-
51 inal cueing is also known as subliminal visual masking.
52 For example, Aranyi et al. conducted an experiment to
53 investigate the effect of masked indices on subsequent
54 item selection behaviour in a virtual environment. Re-
55 sults showed that a short-lived impact (within one sec-
56 ond) of masked cues was indeed influence participants
57 selections subliminally [4]. Recently, a similar masking
58
59
60
61
62
63
64
65

paradigm with three types of visual cues has been ap-
plied in a mobile application. Evidence showed that the
priming stimuli presented with a time window of 17ms
were not fully subliminal [37]. In summary, investiga-
tions of cueing effects using the masking paradigm that
employing different time windows lead to inconsistent
observations. The reliability of the results needs to be
further tested. Therefore, this cueing technique is not
suitable for tackling the present research question: un-
derstanding whether crossmodal perception can be en-
hanced subliminally for a specific interaction task.

The technique of cognitive priming has been applied
with priming material either in the form of visual im-
ages, video, or textual stories [26,25,21]. The material
does not always have a direct relationship with the in-
teraction tasks, and it can be presented either before or
during interaction without restriction on time windows.
The purpose of this technique is to enhance cognitive
function or render affective states for specific interac-
tion tasks. Harrison et al. employed text-based stories
as the priming material to investigate the influence on
subsequent visual judgment performance on different
types of charts. Results showed that positive priming
improved participants' visual judgment accuracy [21]. In
another case, Lewis et al. applied affective computa-
tional priming in the form of background pictures on
a creativity-support design tool [26]. With the back-
ground priming, participants showed an improvement
in the quality but not the quantity of their design. The
effect of cognitive priming was observed even after the
priming has been removed in the latter design tasks.

In conclusion, in the existing HCI literature, the
reliability of applying the subliminal cueing needs to
be further tested. In comparison, cognitive priming has
been used to invoke previous experience and mental
states for goal-oriented tasks. Following this line of in-
vestigation, in the present study, we used a cognitive
priming technique to invoke crossmodal experience and
measured its effect on subsequent sensory-motor re-
sponses.

3 Research aim and scope

Previous studies have investigated crossmodal effects
by manipulating physical values that were varied ei-
ther congruently or incongruently. While in most cases,
CCs have been tested with polarised values such as high
and low in pitch or big and small in size [41,9]. However,
crossmodal stimuli switching between two polarised val-
ues are rarely encountered in everyday interactive en-
vironments. More often, the question is first, whether
and how people attend to available crossmodal infor-
mation which has graded feature values; and second,

1 how people integrate several crossmodal information
2 streams that are mutually exclusive? Specifically, the
3 present paper seeks to address the following research
4 questions:

5
6 RQ1: whether people’s crossmodal perception can
7 be enhanced by cognitive priming, and if it can be, to
8 what extent it modulates small and fast sensory-motor
9 responses in an interactive task? Experiment 1 was de-
10 signed to tackle this question by using different types of
11 priming material as a manipulation factor. We postu-
12 late that if crossmodal perception cannot be enhanced
13 by subliminal priming, we would observe a similar level
14 of performance across conditions. However, if the cross-
15 modal perception indeed can be enhanced, there would
16 be a perceptual reinforcement of a particular CC. Based
17 on previous empirical findings [21,25], we hypothesise
18 that an improved performance should be observed in
19 conditions that include cognitive priming.

20
21 RQ2: How do people integrate crossmodal informa-
22 tion in which two CCs are mutually exclusive? Fur-
23 thermore, how, if at all, does the integration of such
24 information change in the presence or absence of cogni-
25 tive priming? Experiment 2 was designed to tackle this
26 question by employing cognitive priming. We could de-
27 duce the integration process by observing whether the
28 perception and corresponding task performance with
29 primed crossmodal information is enhanced or reduced.
30 If the perception of a primed stimulus is distracted by
31 another unprimed crossmodal stimulus, it is likely that
32 the crossmodal cues have all been taken into account
33 additively [11]; thus the subsequent performance should
34 be lowered by the distractor cue. If the perception of
35 the primed stimulus is enhanced by the priming mate-
36 rial, attention is likely drawn to the correlated cross-
37 modal information selectively [11], i.e. people are less
38 susceptible to the distractor cue, the performance in
39 the subsequent task should be better.

40
41 There are two CCs and two types of priming materi-
42 als involved. Based on previous empirical findings con-
43 cerning CCs, we chose two pairs of crossmodal stimuli
44 that have been frequently referenced in the literature
45 for our experiments: the pitch-brightness and pitch-
46 elevation crossmodal mappings [14,41,35]. We also in-
47 troduced two types of priming as the manipulation fac-
48 tor. One type used physical features of brightness, el-
49 evation and pitch, to imply the CCs, namely percep-
50 tual priming; and another type used videos in which
51 the brightness, elevation and pitch were implied with
52 meaningful contents, namely the conceptual priming.
53 Materials used in the conceptual priming do not have
54 any direct connection with the crossmodal stimuli em-
55 ployed in subsequent tasks, while perceptual priming
56 emphasises the physical values of the stimuli which con-

tain similarities to stimuli presented later. Specifically,
the conceptual priming context uses naturalistic sound
and corresponding video clips to represent the corre-
spondence between auditory pitch and visual bright-
ness, as well as auditory pitch and visual elevation [35,
2]; while perceptual priming uses musical notes with
a set of congruent visual icons to implement the same
CCs. Further details about the priming materials will
be explained in the apparatus section of the paper.

4 Methods

The same design and procedure were used for two ex-
periments. In this section, we report the general exper-
imental protocol, followed with experiment 1. Then in
section 6, we explain the change in the manipulation
factor and report experiment 2.

4.1 Experimental design

The experiments have to be a between-subjects design
to eliminate the learning effect of priming. There were
two independent variables. The first was the type of
priming: the perceptual priming (P-prime) and the con-
ceptual priming (C-prime). The second variable was the
two crossmodal mappings, the pitch-brightness corre-
spondence and the pitch-elevation correspondence. The
two experiments have six conditions with four manip-
ulation groups and two control groups. The manip-
ulation groups were: P-prime and C-prime on pitch-
brightness mapping respectively (condition 1, 2), the
P-prime and C-prime on another crossmodal mapping,
pitch-elevation respectively (condition 3, 4). The con-
trol groups were pitch-brightness mapping without prim-
ing (condition 5) and the pitch-elevation mapping with-
out priming (condition 6). The manipulation factors
and experimental conditions are listed in table 1.

4.2 Task and general procedure

The task in both experiments is learning and repro-
ducing crossmodal sequences. The sample sequence was
made up of two parts, the auditory melody with five-
pitch values equal to musical notes C, D, E, G, A; and
the visual counterpart with either five levels of
brightness or vertical position displayed in five circles
on the screen (Figure 1 Experiment 1a). The detailed
information about crossmodal mappings is presented in
the experimental platform section. During each trial,
a randomly generated sample sequence was displayed
once, which has melody matched with the concurrent

Table 1: Experimental design for experiment 1.

Conditions	Manipulation groups				Control groups	
	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6
Priming factor	P-prime for pitch-brightness	C-prime for pitch-brightness	P-prime for pitch-elevation	C-prime for pitch-elevation	Control group for pitch-brightness	Control group for pitch-elevation
CC factor	Pitch-brightness mapping	Pitch-brightness mapping	Pitch-elevation mapping	Pitch-elevation mapping	Pitch-brightness mapping	Pitch-elevation mapping

Colour code applied in this paper: green indicate perceptual priming groups and red indicate conceptual priming groups.

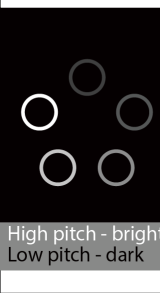
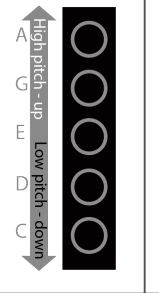
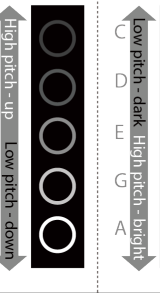
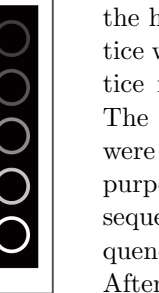

Exp	Experiment 1		Experiment 2			
CCs	a Pitch - brightness	b Pitch - elevation	Pitch - elevation	Pitch - brightness		
Interface layout						
Visual	HSL display					
	Hue	0				
	Saturation	100				
	Lightness	12	34	56	78	100
Audio	Notes	C4	D4	E4	G4	A4
	Frequency	261.63	293.66	329.63	392.00	440.00

Fig. 1: Interaction trial layout. Experiment 1a implement the pitch-brightness mapping, the bright circle paired with high pitch sound and the dark circle paired with low pitch sound. Experiment 1b implement the pitch-elevation mapping. The arrow indicates cross-modal mapping directions. Experiment 2 implement the two crossmodal mappings arranged in a mutually exclusive manner, i.e. high pitch (note A) either congruent with the ‘up’ as shown on the left or ‘bright’ as shown on the right.

visual stimuli. Participants are required to reproduce the melody by clicking the circles on the screen to reproduce the melodic sequence.

The general procedure is as follows. First, participants were given a consent form with details of the experiment to read and sign. Then they were introduced to the priming session based on which group they were assigned. Participants who were in manipulation groups were told to watch a short video or an animation, while control groups had no priming session

and simply moved to the next session directly (Figure 2). After priming, participants move to the warm-up session to get familiar with the task procedure and the hardware setup. To ensure everyone received practice without overtraining, they were instructed to practice no less than twice and no more than six times. The crossmodal stimuli used in the warm-up session were different from those in the task session, for the purpose of minimising learning effects. Specifically, the sequences in the warm-up session used three-note sequences, while the task session used five-notes sequences. After practice, participants moved to the task session, which contained 16 trials (Figure 2). All the sessions were programmed in a single piece of software, participants moved to the following section by pressing a “next” button on the screen.

After participants finished all the trials, they were asked to complete a post-experiment questionnaire. The first section of this questionnaire collected participants’ demographic information, music training history, and whether they had any visual or auditory disorder recently. The second section collected subjective evaluation data, including their interpretation of the priming material, and the interaction strategy they used during the game. The entire experiment lasted for 7 to 13 minutes.

4.3 Apparatus

4.3.1 Conceptual priming material

Early studies claim that our experience of nature can form the basis of a CC. For example, ‘thin, small, light and airy’ things tend to be found at relatively high altitude, compared with ‘dark, heavy, gloomy’ things that tend to be observed near or on the ground [38]. More recently, the results of empirical investigations [41, 35, 40] suggest that some CCs do indeed arise from the experience of correlated physical properties in the natural world. Hence, the conceptual priming in the present study was designed based on naturalistic sound and vi-

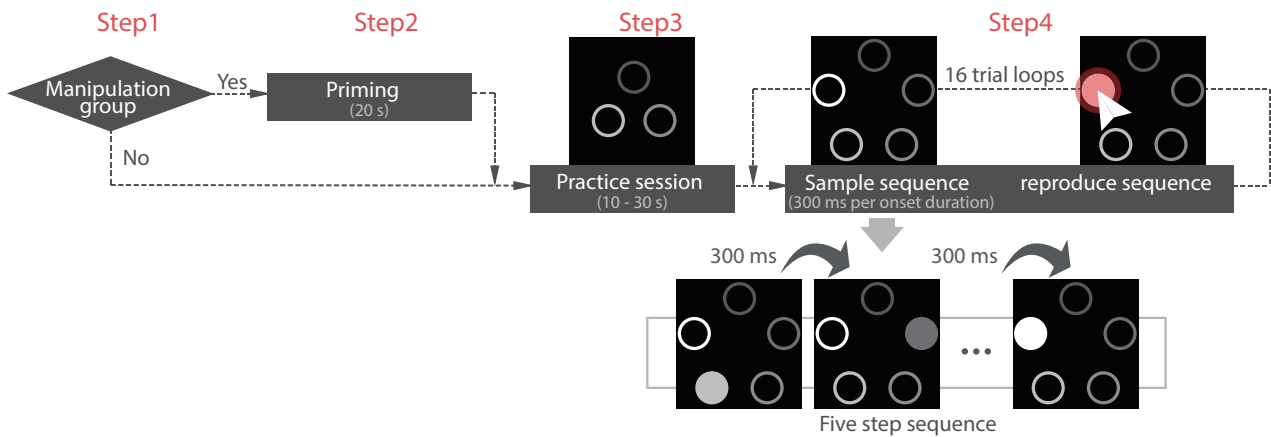


Fig. 2: Experimental procedure. Step 1, assign groups; Step 2, manipulation groups will watch priming video or animation, and control groups will go to S3 directly; Step 3, practice; Step 4, do 16 trials with each one contains a 5 step sequence to learn.

sual representations. Instead of using an image or picture as the priming material, which contains only visual information [4,36], we used video or animation to display crossmodal information. For the pitch-brightness crossmodal mapping, we composed sound clips of birds singing and land-based carnivores roaring to represent high pitch sound and low pitch sound respectively; and in the visual mode, a mute time-lapse video of night-to-day was used to represent the change of brightness (Figure 3 a). This approach was developed based upon the sensory experience that birds are usually active in the daylight and land-based carnivores are usually active at night, thus high pitch sounds tend to be associated with brightness and low pitch sounds with darkness. The composed video are available to review from ([Hyperlink: Pitch-brightness](#)) and from ([Hyperlink: Pitch-elevation](#)).

For the pitch-elevation crossmodal correspondence, we used the same sound clips to represent the high and low pitch, and composed a mute video of birds flying in the sky and carnivores running on the ground to represent the vertically high or low elevation (Figure 3 b). The sensory experience that we build upon is that high pitch sound corresponds to high in the sky and low pitch sound corresponds to being on the ground [38,41].

4.3.2 Perceptual priming material

Based on cognitive studies of crossmodal correspondence [41,35], we used piano sample sounds to represent pitch values. For the priming of pitch-brightness, the notes were presented in the order C-D-E-G-A-A-G-E-D-C, and were paired with a circle which has synchronized changing in brightness from dark to bright and

back to dark. The same sound samples with the same playing sequence were used in the priming for the pitch-elevation matching, with a synchronized visual display of a circle moving from the bottom to the top on the same background.

4.3.3 Experimental platform

The platform for experiment 1 has the pitch-brightness and the pitch-elevation mapping arranged separately in different conditions. Both of the stimuli are congruent without interfering with each other, e.g. increasing pitch corresponded with upward position or brighter visual appearance (Figure 1, Experiment 1a and 1b). The audio was presented at a constant range between 50-55dB (displayed through Mac Pro device, which was tested for comfortable hearing) with a WH-CH500 headphone.

As a screen-based study [45,28], the HSL (hue-saturation-lightness) colour scheme was used in the visual representation. We keep the $H = 250$, $S = 0$, and the lightness value ranging from 12 to 100 with the interval of 22. The corresponding auditory stimuli were notes C, D, E, G, A. Each of the five notes lasted for 300ms with 100ms intervals in-between. All the elements were displayed on a black background ($H = 0$, $S = 100$, $L = 0$) with a screen resolution of 2560*1600 dpi. In order to avoid the potential confound of stimulus-response compatibility effect [39] (e.g. high pitch naturally correlated to up/right and low pitch to down/left), instead of a linear arrangement, the circles were organized in a pentagon formation (Figure 1 a), during each trial a randomly generated sequence is displayed. The position of each circle placed in the pentagon was also randomised.

4.4 Measurements

Two depended variables were time intervals between each input and task error rate. Time intervals were counted as the length of time between adjacent clicks. The difference in time intervals between the manipulation groups and the baseline groups was calculated as an indicator of the priming effect on sensory-motor reaction efficiency. The interaction error rate was calculated by dividing the number of wrongly produced notes by the total number of sample notes.

In addition, the qualitative data that was included in the analysis was as follows: 1) the subjective interpretations of priming material collected from the post-experiment questionnaire, and 2) Plots of the overall task accuracy which was calculated not by summing the number of wrongly produced notes, but by the number of incorrectly aligned CCs during reproduction. For example, if one step in a sample sequence change from note C to note E with the brightness going up 2 levels, both the reproduced move from C to E with brightness going up 2 levels and D to E with brightness going up 1 level would be counted as correct, since the increased pitch value corresponded with brightened visual display. However, the step from A to E with brightness level going down would be counted as an incorrect step, as the decreased pitch sound should not be aligned with the increased brightness level.

4.5 Hypotheses

Since the pitch-brightness and pitch-elevation correspondences have been repeatedly evaluated in the literature as working for most people [14,39,18], we assumed that most participants would be able to do the task without perceptual discrepancy on those crossmodal stimuli. Thus we hypothesized that:

H1: Following previous studies [32,25,26], we predict that people’s crossmodal perception can be enhanced by cognitive priming. Specifically, that both the P-prime and the C-prime for the pitch-brightness mapping and the pitch-elevation mapping will support faster sensory-motor responses and produce better task accuracy than would be seen in the two baseline groups.

H2: The type of priming will have different sensory-motor modulation effects on task performance. However, we chose not to predict the direction of the difference.





Con	Visual priming	Auditory priming	CC implied
a. Condition 1 and 2	 <p>Video frames of night to day</p>	 <p>Audio stream from carnivore roaring to birds singing</p>	<p>Pitch-brightness correspondence</p> <p>Night to day implies the change of visual brightness level, carnivore roaring to birds singing implies the change on pitch levels</p>
b. Condition 3 and 4	 <p>Video frames of sky and ground</p>	 <p>Audio stream from birds singing to carnivore roaring</p>	<p>Pitch-elevation correspondence</p> <p>The view from sky to ground implies the change of visual elevation, birds singing to carnivore roaring implies the change on pitch levels</p>

Fig. 3: Video screenshots and partial waveform intercepted from conceptual priming material for pitch-brightness crossmodal mapping (a) and pitch-elevation mapping (b).

5 Experiment 1: Study on congruent crossmodal mappings

One hundred and twenty participants (67 male, 53 female, aged 18-55 years, mean = 26.17, SD = 5.56) were involved in the first experiment. All participants confirmed that they have no visual and auditory disorders before trials (after correction). Participants were balanced across groups according to their age and gender. In order not to introduce a potential confounding factor of cultural background on CC perception or interpretation, a mixture of volunteers of different nationalities, professions and music training experience were recruited through the universities’ e-mail list and social network sites.

5.1 Results

A Kolmogorov-Smirnov test showed that the time intervals data can be assumed to be normally distributed. We ran a one-way ANOVA to compare the results of P-prime, C-prime and baseline conditions on pitch-brightness mapping, as well as the P-prime, C-prime and the baseline condition on pitch-elevation mapping. Fisher’s LSD test was used for posthoc tests of main effects. We used a confidence level of $\alpha = 0.05$ for the tests.

Table 2: Statistical analysis based on time intervals in experiment 1.

	F(2, 3837)	p	r	Note
Pitch-brightness mapping	76.649	.000	.17	P-prime (517.57 ms) < Control group (555.04 ms), C-prime (500.70 ms) < Control group
Pitch-elevation mapping	22.246	.000	.10	P-prime (578.43 ms) < Control group (558.48 ms), C-prime (545.98 ms) < Control group

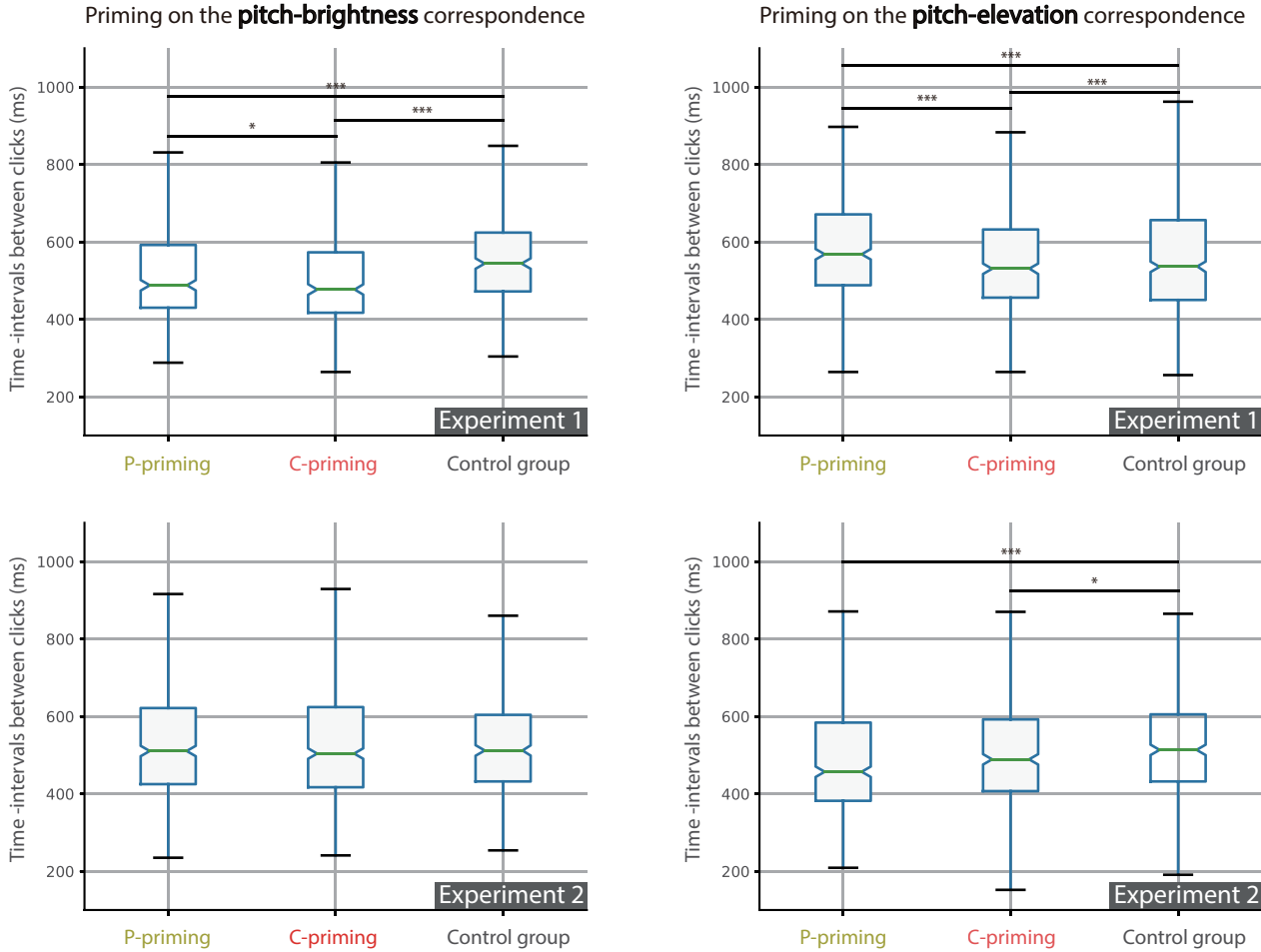


Fig. 4: Box plot of the time intervals under the priming for the pitch-brightness correspondence and for the pitch-elevation correspondence in experiment 1 (top two panels) and experiment 2 (bottom two panels). $*P < 0.05$, $***P < 0.001$.

5.1.1 Results on time intervals between inputs of a sequence

During experimental trials, two of the priming groups with the pitch-brightness mapping had statistically significantly smaller time intervals between clicks than the control group, and the C-prime group also produced smaller time intervals than the P-prime group (Figure 4). However, for the pitch-elevation mapping, only the C-prime groups had statistically significant faster inputs than the control group, while the P-prime group

produced statistically significant slower inputs than the control group (Figure 4). The detailed statistical results are listed in table 2.

5.1.2 Results on task error rate

For the pitch-brightness mapping, C-prime group (11.13%) produced more accurate sequences than the P-prime group (14.25%) and baseline group (14.44%) (Figure 5 a (pitch-brightness)). For the pitch-elevation mapping, the P-prime group (21.44%) produced a higher

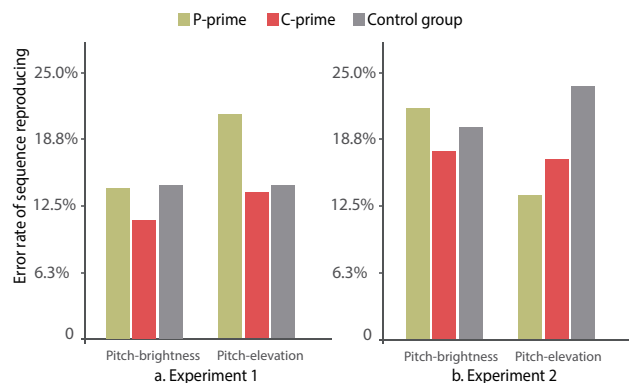


Fig. 5: Error rate of sequence re-producing. Figure a shows results of error rate of experiment 1. Figure b shows results of error rate of experiment 2.

error rate than the control group (14.50%) and C-prime group (13.88%) as shown in figure 5a (pitch-elevation).

5.2 Discussion

The hypothesis H1 that crossmodal perceptual preference can be induced by cognitive priming has been confirmed for the pitch-brightness correspondence. With the pitch-elevation correspondence however, only the C-prime group showed a positive effect on task performance, while the P-prime group had slower motor responses (Figure 4 (up-right chart)) and poorer accuracy (Figure 5a (pitch-elevation)).

These inconsistent outcomes between the two crossmodal correspondences may result from different interaction strategies afforded by the task paradigm. Previous research mainly used a speeded classification paradigm [41], which required participants to react to stimuli with a one-shot key press. This process requires little working memory or cognitive resource. In the current experiment, however, participants had a clear interaction goal and engaged with a sequence of input actions. In this regard, the P-prime, which primed a direct association between the auditory and visual stimuli, possibly functioned as explicit instruction, and encouraged participants to recall and check during the interaction constantly. This extra cognitive load could explain the increase in response time. Subjective evaluation support this explanation, which will be discussed in the latter part of the section.

Surprisingly, such degraded performance was less evident in the P-prime group with the pitch-brightness correspondence. According to self-report, the interaction strategy used by participants might compensate the inferior effect. In responses to a question about interaction strategies for reproducing the sequences, 67%

of participants in the pitch-brightness condition reflected that they used visual tracking strategy to follow the flash pattern of the sequence. In contrast, none of the participants in the pitch-elevation mapping group recalled they had used any strategy. Indeed, the visual pattern for the pitch-elevation correspondence has a vertical arrangement, and the moving path of the sequence is overlapped. While the visual pattern for the pitch-brightness correspondence has a planar arrangement, and the moving path was more likely to be perceived as a trackable trajectory [20]. Since visual perception is more sensitive to spatial arrangement, while auditory perception is more sensitive to temporal information, the combination of a 2-D visual stimulus with a 1-D auditory stimulus can produce a multisensory enhancement effect [24]. As a result, the time delay and the error were less salient in the conditions with a pitch-brightness mapping.

The hypothesis H2 that priming types will have different effects on sensory-motor modulation has been confirmed. Results showed that the C-prime condition did improve participant's crossmodal perception, in both the pitch-brightness and the pitch-elevation correspondences, which in turn supported faster motor response as well as improved accuracy of crossmodal sequence reproduction. In comparison, the P-prime condition had a positive effect only on the pitch-brightness crossmodal mapping. This effect may be due to the compensatory interaction strategy afforded by the task paradigm. Previous studies have shown that the priming process without conscious awareness can have a positive effect on people's affective reaction and decision making [32,25], thus we deduce that the perceptual enhancement with C-priming in our case could also have functioned in a subliminal way. It enabled participants to have a stronger perceptual alignment with a specific CC, which facilitated fast and instantaneous motor response. While the P-prime material may rise to the level of conscious awareness during the priming process, thus involving working memory allocation while doing the task, which lowered the overall performance to some extent.

A supportive finding of the above assumption can be gleaned from the post-experimental questionnaire. Most of the participants in the C-priming groups were not aware of the CC presented during priming, and could not consciously recall the correlation between the priming and the interaction task. In all of the 40 subjective interpretations from the C-prime groups (2 C-prime groups with 20 participants in each), only 2 participants reported that 'it was about how the sound changes with visual information', and that 'they are there to focus your concentration on the screen and the sound'. The other 28 participants' answers were ex-

clusively focused on the detailed contents of the prime context. Such as ‘It’s showing the sunrise’ for the pitch-brightness priming and ‘flying birds, running animal’ for the pitch-elevation priming. Meanwhile, only 4 of the 40 participants rated that the C-prime was helpful for the interaction task. In contrast, 36 of 40 participants in the P-priming groups were fully aware of the purpose of the priming. One of the typical answers, for instance, was that ‘information that helps you prepare for the game’ for the pitch-brightness priming, and that ‘different notes of music vertically spaced’ for the pitch-elevation priming. 36 of the 40 participants rated that the priming was helpful. From the usability perspective, this evaluation can explain the assumption of extra cognitive load that was discussed for the hypothesis H1.

Combining participants’ behavioural data with their subjective rating, it can be confirmed that the C-prime subliminally enhanced crossmodal perception and led to better task performance. This fact also explains why the C-priming was not considered helpful by most of the participants. In comparison, the purpose of the P-priming was recognised in most cases and thus tended to be acknowledged to be helpful for task completion, however, the actual behavioural data of the P-primed participants pointed to the opposite.

To better understand the quantitative effects of the priming techniques on subsequent goal-oriented tasks, participants’ performance accuracy based on crossmodal alignment, as explained in section 4.4, was plotted on two temporal scales: the within-trial performance and between-trial performance (Figure 6 Experiment 1). Along the horizontal axis, which represents the steps within trials, the P-prime groups produced more accurate moves in the first two steps and were more prone to error in the last two steps. In comparison, the C-prime group and the control group do not show this pattern. These observations reflect that participants in the P-prime group relied more on working memory capacity than on crossmodal congruency. This observation is consistent with objective behavioural data as discussed previously. Along the vertical axis, which represents the between-trial performance, no improvement can be observed in the later trials in all groups. In other words, the data did not reflect learning effect between trials.

6 Experiment 2: Study on mutually exclusive crossmodal mappings

One hundred and twenty participants (62 male, 58 female, age 18-35 years, mean = 26.48, SD = 4.58) were recruited for the second experiment, with 20 participants in each condition. All the participants were newly

recruited for experiment 2, and randomised in the same way as in experiment 1.

This experiment was designed to investigate the second question: RQ2: How do people integrate crossmodal information in which two CCs are combined where the perception of the congruency of one CC excludes the perception of the congruency of the other? Furthermore, how, if at all, does the integration of such information change in the presence or absence of cognitive priming?

Participants in all conditions were introduced to the interface which contained 2 crossmodal mappings arranged in an incongruent manner (Figure 1c). Participants in conditions 1 and 2 were primed in relation to the pitch-brightness correspondence, which was applied in the subsequent task but had pitch-elevation correspondence as a distractor. Participants in conditions 3 and 4 were primed in relation to the pitch-elevation correspondence, with the pitch-brightness correspondence as the distractor (see the experimental design in table 3). In this way, participants in each of the manipulation groups can only be perceptually consistent with one of the CCs, but can not be consistent with both.

H3: For the 2 CCs arranged in a contradicted manner, the primed groups will integrate crossmodal information in a selective way, while the control groups will do in an additive way (11). Specifically, the priming groups will produce better performance, e.g. faster motor response and lower error rates than the control groups.

6.1 Results

A Kolmogorov-Smirnov test showed that the data set can be assumed to be normally distributed. The one-way ANOVA was used as the statistical method with a confidence level of $\alpha = 0.05$.

6.1.1 Results on time intervals between inputs of a sequence

When primed on the pitch-brightness correspondence with pitch-elevation as an interaction distractor, there was no significant difference between primed groups and the control group, and there was no significant difference between the two types of priming (Table 2 and figure 4 (bottom-left panel)). When primed on the pitch-elevation correspondence with pitch-brightness as an interaction distractor, both the primed groups performed significantly better than the control group. There was no significant difference between the two types of priming (Table 4 and figure 4 (bottom-right panel)).

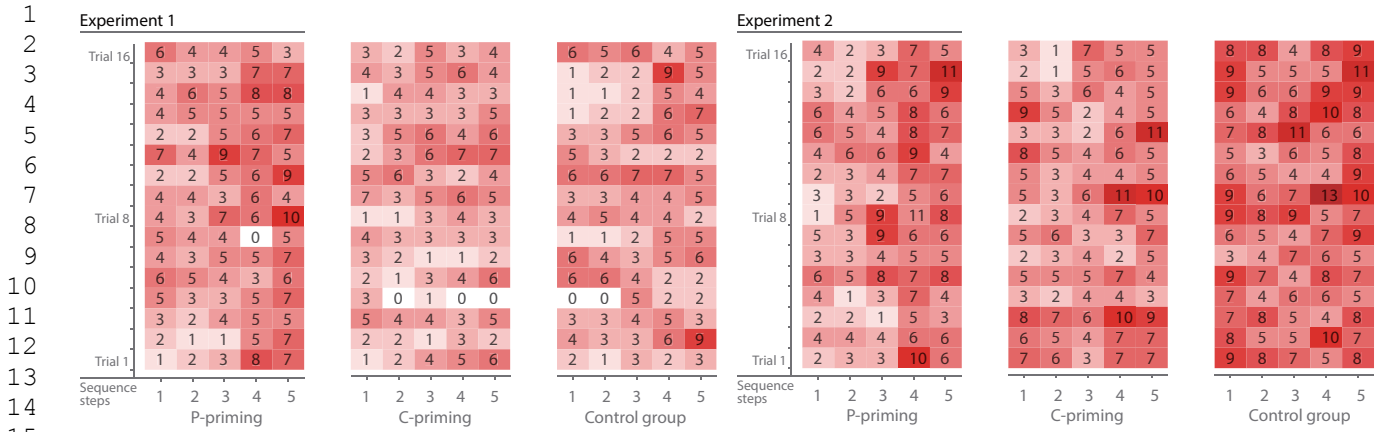


Fig. 6: Crossmodal alignment in the P-prime group, C-prime group and the control group of experiment 1 (left) and 2 (right). The horizontal axis denotes the sequence steps from 1 to 5 (within trial), and the vertical axis denotes the number of experimental trials (between trials). The digits with the red background represent the number of inputs which were not aligned with CCs at that specific step (take into account all participants' steps).

Table 3: Experimental design for experiment 2.

Conditions	Manipulation groups				Control groups	
	Condition 1	Condition 2	Condition 3	Condition 4	Condition 5	Condition 6
Priming factor	P-prime for pitch-brightness	C-prime for pitch-brightness	P-prime for pitch-elevation	C-prime for pitch-elevation	Control group for pitch-brightness	Control group for pitch-elevation
CC factor	Pitch-elevation as a distractor	Pitch-elevation as a distractor	Pitch-brightness as a distractor	Pitch-brightness as a distractor	Pitch-brightness mapping with pitch-elevation as a distractor	Pitch-elevation mapping with pitch-brightness as a distractor

Table 4: Statistical analysis based on time intervals in experiment 2.

	F(2, 3837)	p	r	Note
Prime on pitch-brightness with pitch-elevation as a distractor	.886	.412	.000	P-prime (537.17 ms) < Control group (492.21 ms), C-prime (527.74 ms) < Control group
Prime on pitch-elevation with pitch-brightness as a distractor	23.145	.000	.11	P-prime (494.23 ms) < Control group (525.79 ms), C-prime (506.54 ms) < Control group

6.1.2 Results on task error rate

With the CCs arranged incongruently, when primed on the pitch-brightness correspondence but distracted with pitch-elevation during the interaction, the C-prime group (17.88%) produced a slightly better performance in the error rate than the P-prime group (22.06%), and the control group (20.13%) (Figure 5b pitch-brightness). When primed on the pitch-elevation correspondence with the pitch-brightness as the distractor, the P-prime group (13.69%) showed a lower error rate than the C-prime

group (17.19%) and the control group (23.94%) (Figure 5b pitch-elevation).

6.2 Discussion

Hypothesis H3 has been confirmed for priming on the pitch-elevation mapping but rejected for the priming on pitch-brightness mapping. The two priming groups and the control group for the pitch-brightness correspondence have no significant difference in terms of motor response speed (Figure 4 (lower-left)), and there was no

1 obvious difference between the primed groups and the
2 control group in terms of the accuracy (Figure 5b). The
3 other two priming groups for the pitch-elevation cor-
4 respondence had significantly better performance than
5 the control group in terms of both the motor response
6 speed and sequence re-producing accuracy.

7
8 One explanation for this fact may be that the pitch-
9 elevation correspondence may have stronger perceptual
10 weight [11] than the pitch-brightness correspondence.
11 The perceptual experience of the pitch-elevation cor-
12 respondence may be encountered more frequently in daily
13 interactions than the pitch-brightness correspondence,
14 thus the neural response to pitch-elevation stimuli may
15 be stronger than that for the pitch-brightness corre-
16 spondence [7,44]. This crossmodal perceptual weighting
17 may not be easy to observe in situations where cross-
18 modal stimuli are isolated from one another, such as in
19 the case of experiment 1 and in many classification-
20 based tasks employed in previous studies [14,41,10].
21 While in the situation of experiment 2, where cross-
22 modal stimuli were overlapped and incongruent, prim-
23 ing on the relatively stronger CC, pitch-elevation cor-
24 respondence in this case, enabled participants weight-
25 ing the dominant stimulus selectively in the subsequent
26 activity. However, priming on the less dominant CC,
27 the pitch-brightness correspondence in this case, seems
28 to have little or no effect on modulating subsequent
29 activity. As a result, only those groups primed on the
30 pitch-elevation correspondence produced improved per-
31 formance. Thus we postulate that crossmodal stimuli
32 integration may occur in a selective manner rather than
33 an additive manner in the presence of priming [23].

34
35
36
37 Following the discussion in experiment 1, to observe
38 how the priming technique influenced performance in
39 an incongruent crossmodal situation, the crossmodal
40 alignment accuracy of participants has been plotted in
41 figure 6 (right). The within-trial comparison shows that
42 the P-priming groups produced more accurate moves
43 at the first few steps in the sequences. Compared with
44 the P-priming groups in experiment 1 (figure 6 left), a
45 similar pattern appeared regardless of how crossmodal
46 stimuli were combined. Based on this observation, we
47 can deduce that participants presented with explicit P-
48 priming material tend to do the task in a way that in-
49 volved more working memory capacity. Different from
50 the situation in experiment 1, the extra attention allo-
51 cation made participants less susceptible to perceptual
52 distractions in experiment 2. The between-trial compar-
53 ison shows that participants in the C-priming groups
54 produced better performance in the latter half of to-
55 tal trials than the earlier half. Comparing with the C-
56 priming groups in experiment 1, which shows better ac-
57 curacy in the earlier trials, it is plausible to attribute the

improvement produced in experiment 2 to the practice
effect more than to the priming effect. However, the plot
for the control groups for experiment 2 does not have a
salient observable difference between earlier trials and
later trials due to practice. This observation could in-
dicate two things. Firstly, the C-priming do play a part
in modulating crossmodal perception, but the effect is
not immediate. Secondly, without priming, participants
are susceptible to the crossmodal distractor that hap-
pened to share the same crossmodal perceptual channel.
Therefore, in the absence of priming, the integration of
mutually exclusive crossmodal stimuli more likely hap-
pened in an additive manner than a selective manner [11].

7 General Discussion

The present study examined participants' crossmodal
perception enhancement, as well as the crossmodal in-
tegration process in the conditions with and without
cognitive priming. The first contribution, to the best
of our knowledge, is a systematic investigation of the
effect of two CCs, pitch-brightness and pitch-elevation,
on an interactive task that requires consistent input
behaviour. Secondly, this study introduced and evalu-
ated the priming technique as a way to enhance cross-
modal perception for interactive tasks. Thirdly, rooted
in cognitive and social psychology, this study elucidates
two types of priming materials, i.e. conceptual priming
and perceptual priming, and investigated their effect on
people's sensory-motor performance; and last, through
cognitive priming technique, this study explored how
people integrate two CCs that happened to be mutu-
ally exclusive during the interaction.

In general, the results of experiment 1 revealed that
when two visual-auditory CCs were isolated and did
not interfere with one another, conceptual priming was
successful in enhancing crossmodal perception, which
led to faster motor responses and improved task accu-
racy. In contrast, perceptual priming operated in a more
explicit manner. With goal-oriented tasks, perceptual
priming appeared to function as an instruction, which
caused extra cognitive resource to be allocated to the
recall and comparison of the primed cue and the task
stimuli. This process diminished the task performance,
regardless of the fact that most participants regarded
the perceptual priming as being helpful. The results of
experiment 2 revealed that in the situation where two
CCs have mutually exclusive congruency, the priming
technique has little or no effect on the less dominant
CC, i.e. pitch-brightness in this case. With the more
dominant CC, i.e. pitch-elevation in this case, explicit

1 priming enhanced perception and led to faster motor
 2 responses and improved accuracy.

3 Extending previous studies on CCs with graded stim-
 4 ulti [47,29], our study further revealed perceptual phe-
 5 nomena that have not been previously observed with
 6 polarised crossmodal values. Individuals seem to pos-
 7 sess varying perceptual weights regarding different CCs.
 8 These perceptual weights can be dynamically enhanced
 9 or inhibited by different types of priming and by the
 10 way crossmodal information is organised. Moreover, we
 11 observed that priming on the dominant CC will strengthen
 12 the perceptual response on that crossmodal stimulus,
 13 while priming on the less dominant CC will have little
 14 or no facilitatory effect on fast and consecutive motor
 15 responses. Future studies are needed, however, to verify
 16 and further explore the impact of multiple CCs in in-
 17 teractive environments, with either continuous feedback
 18 [16] or graded feedback. We envision the outcomes of
 19 this line of research being applied to the scenario where
 20 information needs to be overlaid in augmented reality
 21 or virtual reality, or in situations demanding high cog-
 22 nitive load and sophisticated interactive actions.

23 One of the limitations of the present study is the
 24 small effect of the statistical analysis. The small sam-
 25 ple size involved in each experimental condition may
 26 be one of the reasons. Further verification with larger
 27 sample sizes should be conducted in the future. An-
 28 other limitation of the present investigation is that it
 29 does not indicate whether the cognitive priming effect
 30 and the perceptual weighting phenomenon would gen-
 31 eralise to other CCs. There are various crossmodal fea-
 32 ture values with different levels of intensity that need to
 33 be further tested. Although it seems unlikely that the
 34 priming contexts in the present study, i.e. video clips
 35 of natural environments, are the only two cases that
 36 exhibit priming effects. The next step in this line of
 37 research would involve systematic investigation of the
 38 priming approaches in terms of material, priming dura-
 39 tion, persistence, and interaction scenario. Last but not
 40 least, sensory channels for priming should not be lim-
 41 ited to the auditory and visual modalities. Future work
 42 could expand the scope of the priming modalities and
 43 their combinations to other sensory channels involving
 44 but not limited to haptics, olfaction, and gustation [33].

50 7.1 Implications for interactive system design

51 In terms of potential applications, the present findings
 52 could contribute to the field of mindless technology [1,
 53 36] and multisensory interaction [31]. Mindless comput-
 54 ing emphasises the subconscious mental process, and
 55 is characterized as a fast, automatic interaction that
 56 requires little or no effort. Mindless technology makes

57 use of subliminal mental state, aiming to guide peo-
 58 ple towards intended interaction behaviour without an-
 59 noying or distracting them. Previous research showed
 60 that both environmental features and the interactive
 61 system itself could impose a subtle influence on peo-
 62 ple’s choice and judgement in interactive tasks [1,4,36].
 63 The present study, built on theoretical accounts of CCs,
 64 contributes to the field by introducing cognitive priming
 65 as an approach to enhance crossmodal perception, and
 66 consequently improves interactive sensory-motor per-
 67 formance. The implications for designing multisensory,
 68 mindless interaction are first, implicit conceptual prim-
 69 ing can be used to improve interaction efficiency when
 70 information shares only one CC feature. In addition,
 71 when there are two or more streams of crossmodal in-
 72 formation which have mutually exclusive congruency,
 73 explicit perceptual priming could be used for improv-
 74 ing interactive performance.

75 The present study also provides insights for the de-
 76 sign of multisensory interactive systems [31]. As is so of-
 77 ten discussed, one of the advantages of multisensory in-
 78 teraction is that it can expand peoples’ cognitive capac-
 79 ity by presenting several streams of information through
 80 different sensory channels [34,48,24]. In order to make
 81 good use of this advantage, we need to consider the situ-
 82 ation in which two or more streams of information may
 83 occupy the same pair of crossmodal channels. To make
 84 things worse, the streams of information may contain
 85 mutually exclusive CCs, as demonstrated in experiment
 86 2. Both cases could inhibit information processing cap-
 87 acity, and as a result, negate the advantages of multi-
 88 sensory interaction. The design implication for handling
 89 this situation could be isolating information streams ei-
 90 ther spatially or temporally to avoid confusion. Another
 91 solution could be shunting information flow through
 92 different pairs of crossmodal channels, e.g. the visual-
 93 auditory channels and the visual-haptic channels, to re-
 94 duce cross-sensory distractions.

95 Acknowledgements

96 References

1. Alexander T Adams, Jean Costa, Malte F Jung, and Tanzeem Choudhury. Mindless computing: designing technologies to subtly influence behavior. In *Proceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing*, pages 719–730. ACM, 2015.
2. Wendy J Adams, Erich W Graf, and Marc O Ernst. Experience can change the ‘light-from-above’ prior. *Nature neuroscience*, 7(10):1057, 2004.
3. Monali Agrawal and Jacob Jorgensen. Sonify: Making visual graphs accessible. In *Human Interaction and*

- 1 *Emerging Technologies: Proceedings of the 1st International*
2 *Conference on Human Interaction and Emerging*
3 *Technologies (IHET 2019), August 22-24, 2019, Nice,*
4 *France*, volume 1018, page 454. Springer, 2019.
- 5 4. Gabor Aranyi, Sid Kouider, Alan Lindsay, Hielke Prins,
6 Imtiaj Ahmed, Giulio Jacucci, Paolo Negri, Luciano
7 Gamberini, David Pizzi, and Marc Cavazza. Subliminal
8 cueing of selection behavior in a virtual environment.
9 *Presence*, 23(1):33–50, 2014.
- 10 5. Saskia Bakker, Alissa N Antle, and Elise Van Den Hoven.
11 Embodied metaphors in tangible interaction design. *Personal and Ubiquitous Computing*, 16(4):433–449, 2012.
- 12 6. John A Bargh and Tanya L Chartrand. The mind in
13 the middle. *Handbook of research methods in social and*
14 *personality psychology*, 2:253–285, 2000.
- 15 7. Lawrence W Barsalou. Perceptions of perceptual symbols.
16 *Behavioral and brain sciences*, 22(4):637–660, 1999.
- 17 8. Lionel Brunel, Paulo F Carvalho, and Robert L Goldstone.
18 It does belong together: cross-modal correspondences
19 influence cross-modal integration during perceptual
20 learning. *Frontiers in psychology*, 6, 2015.
- 21 9. Riccardo Brunetti, Allegra Indraccolo, Claudia
22 Del Gatto, Charles Spence, and Valerio Santangelo.
23 Are crossmodal correspondences relative or absolute?
24 sequential effects on speeded classification. *Attention,*
25 *Perception, & Psychophysics*, 80(2):527–534, 2018.
- 26 10. Riccardo Brunetti, Allegra Indraccolo, Serena Mastroberardino,
27 Charles Spence, and Valerio Santangelo. The impact
28 of cross-modal correspondences on working memory
29 performance. *Journal of Experimental Psychology: Human*
30 *Perception and Performance*, 43(4):819, 2017.
- 31 11. Nicolas Bruno and James E. Cutting. Minimodularity
32 and the perception of layout. *Journal of experimental*
33 *psychology. General*, 117 2:161–70, 1988.
- 34 12. Pierre Chalfoun and Claude Frasson. Cognitive priming:
35 assessing the use of non-conscious perception to enhance
36 learner’s reasoning ability. In *International Conference*
37 *on Intelligent Tutoring Systems*, pages 84–89. Springer,
38 2012.
- 39 13. Yi-Chuan Chen and Charles Spence. When hearing the
40 bark helps to identify the dog: Semantically-congruent
41 sounds modulate the identification of masked pictures.
42 *Cognition*, 114(3):389–404, 2010.
- 43 14. Karla K Evans and Anne Treisman. Natural cross-modal
44 mappings between visual and auditory features. *Journal*
45 *of vision*, 10(1):6–6, 2009.
- 46 15. Michael W Eysenck and Mark T Keane. *Cognitive psychology: A student’s handbook*. Psychology press, 2013.
- 47 16. Feng Feng and Tony Stockman. An investigation of dynamic
48 crossmodal instantiation in tuis. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*,
49 pages 82–90. ACM, 2017.
- 50 17. Emma Frid, Jonas Moll, Roberto Bresin, and Eva-Lotta
51 Sallnäs Pysander. Haptic feedback combined with
52 movement sonification using a friction sound improves
53 task performance in a virtual throwing task. *Journal on*
54 *Multimodal User Interfaces*, 13(4):279–290, 2019.
- 55 18. Alberto Gallace and Charles Spence. Multisensory synesthetic
56 interactions in the speeded classification of visual size.
57 *Perception & Psychophysics*, 68(7):1191–1203, 2006.
- 58 19. Michele Geronazzo, Alberto Bedin, Luca Brayda, Claudio
59 Campus, and Federico Avanzini. Interactive spatial sonification
60 for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies*, 85:4–15,
61 2016.
- 62 20. Shihui Han, Glyn W Humphreys, and Lin Chen. Uniform
63 connectedness and classical gestalt principles of perceptual
64 grouping. *Perception & psychophysics*, 61(4):661–674, 1999.
- 65 21. Lane Harrison, Drew Skau, Steven Franconeri, Aidong Lu,
and Remco Chang. Influencing visual judgment through affective priming. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*,
pages 2949–2958. ACM, 2013.
22. Jörn Hurtienne. *Image schemas and design for intuitive use: Exploring new guidance for user interface design*. PhD thesis, 2011.
23. Robert A Jacobs. What determines visual cue reliability? *Trends in cognitive sciences*, 6(8):345–350, 2002.
24. Karin H James, Sophia Vinci-Booher, and Felipe Munoz-Rubke. The impact of multimodal-multisensory learning on human performance and brain activation patterns. In *The Handbook of Multimodal-Multisensor Interfaces*,
pages 51–94. Association for Computing Machinery and Morgan & Claypool, 2017.
25. Nataliya Kosmyna, Franck Tarpin-Bernard, and Bertrand Rivet. Conceptual priming for in-game bci training. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 22(5):26, 2015.
26. Sheena Lewis, Mira Dontcheva, and Elizabeth Gerber. Affective computational priming and creativity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 735–744. ACM, 2011.
27. Kelly McCormick, Simon Lacey, Randall Stilla, Lynne C Nygaard, and Krishnankutty Sathian. Neural basis of the crossmodal correspondence between auditory pitch and visuospatial elevation. *Neuropsychologia*, 112:19–30, 2018.
28. Ravi Mehta and Rui Juliet Zhu. Blue or red? exploring the effect of color on cognitive task performances. *Science*, 323(5918):1226–1229, 2009.
29. Oussama Metatla, Nuno N Correia, Fiore Martin, Nick Bryan-Kinns, and Tony Stockman. Tap the shapetones: Exploring the effects of crossmodal congruence in an audio-visual interface. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*,
pages 1055–1066. ACM, 2016.
30. Sophie Molholm, Walter Ritter, Daniel C Javitt, and John J Foxe. Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. *Cerebral Cortex*, 14(4):452–465, 2004.
31. Cosmin Munteanu, Pourang Irani, Sharon Oviatt, Matthew Aylett, Gerald Penn, Shimei Pan, Nikhil Sharma, Frank Rudzicz, Randy Gomez, Ben Cowan, et al. Designing speech, acoustic and multimodal interactions. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 601–608. ACM, 2017.
32. Sheila T Murphy and Robert B Zajonc. Affect, cognition, and awareness: affective priming with optimal and suboptimal stimulus exposures. *Journal of personality and social psychology*, 64(5):723, 1993.
33. Marianna Obrist, Elia Gatti, Emanuela Maggioni, Chi Thanh Vi, and Carlos Velasco. Multisensory experiences in hci. *IEEE MultiMedia*, 24(2):9–13, 2017.
34. Sharon Oviatt. Breaking the robustness barrier: Recent progress on the design of robust multimodal systems. In *Advances in computers*, volume 56, pages 305–341. Elsevier, 2002.
35. Cesare V Parise, Katharina Knorre, and Marc O Ernst. Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences*, page 201322705, 2014.

36. Charlie Pinder. Nonconscious behaviour change technology: Targeting the automatic. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 160–165. ACM, 2017.
37. Charlie Pinder, Jo Vermeulen, Benjamin R Cowan, Russell Beale, and Robert J Hendley. Exploring the feasibility of subliminal priming on smartphones. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*, page 21. ACM, 2017.
38. Carroll C Pratt. The spatial character of high and low tones. *Journal of Experimental Psychology*, 13(3):278, 1930.
39. Elena Rusconi, Bonnie Kwan, Bruno L Giordano, Carlo Umiltà, and Brian Butterworth. Spatial representation of pitch height: the smarç effect. *Cognition*, 99(2):113–129, 2006.
40. Nadiya Slobodenyuk, Yasmina Jraissati, Ali Kanso, Lama Ghanem, and Imad Elhajj. Cross-modal associations between color and haptics. *Attention, Perception, & Psychophysics*, 77(4):1379–1395, 2015.
41. Charles Spence. Crossmodal correspondences: A tutorial review. *Attention, Perception, & Psychophysics*, 73(4):971–995, 2011.
42. Charles Spence. On the relative nature of (pitch-based) crossmodal correspondences. *Multisensory research*, 32(3):235–265, 2019.
43. Charles Spence and Ophelia Deroy. Crossmodal correspondences: Innate or learned? *i-Perception*, 3(5):316–318, 2012.
44. Barry E Stein, Terrence R Stanford, and Benjamin A Rowland. Development of multisensory integration from the perspective of the individual neuron. *Nature Reviews Neuroscience*, 15(8):520, 2014.
45. Evan Thompson, Adrian Palacios, and Francisco J Varela. On the ways to color. *Behavioral and Brain Sciences*, 15(1):56–74, 1992.
46. Bruce N Walker and Gregory Kramer. Mappings and metaphors in auditory displays: An experimental assessment. Georgia Institute of Technology, 1996.
47. Laura Walker and Peter Walker. Cross-sensory mapping of feature values in the size–brightness correspondence can be more relative than absolute. *Journal of Experimental Psychology: Human Perception and Performance*, 42(1):138, 2016.
48. Christopher D Wickens. Multiple resources and mental workload. *Human factors*, 50(3):449–455, 2008.