

## Do you see what I'm singing? Visuospatial movement biases pitch perception

Louise Connell<sup>a,\*</sup>, Zhenguang G. Cai<sup>a</sup>, Judith Holler<sup>a,b</sup>

<sup>a</sup> School of Psychological Sciences, University of Manchester, Manchester, UK

<sup>b</sup> Max Planck Institute for Psycholinguistics, Nijmegen, Netherlands

### ARTICLE INFO

#### Article history:

Accepted 28 September 2012

#### Keywords:

Mental representation  
Pitch perception  
Music  
Space  
Spatial representation

### ABSTRACT

The nature of the connection between musical and spatial processing is controversial. While pitch may be described in spatial terms such as “high” or “low”, it is unclear whether pitch and space are associated but separate dimensions or whether they share representational and processing resources. In the present study, we asked participants to judge whether a target vocal note was the same as (or different from) a preceding cue note. Importantly, target trials were presented as video clips where a singer sometimes gestured upward or downward while singing that target note, thus providing an alternative, concurrent source of spatial information. Our results show that pitch discrimination was significantly biased by the spatial movement in gesture, such that downward gestures made notes seem lower in pitch than they really were, and upward gestures made notes seem higher in pitch. These effects were eliminated by spatial memory load but preserved under verbal memory load conditions. Together, our findings suggest that pitch and space have a shared representation such that the mental representation of pitch is audiospatial in nature.

© 2012 Elsevier Inc. All rights reserved.

### 1. Introduction

Musical and spatial processing are interlinked, but the exact nature and extent of the connection is controversial. People with amusia (i.e., an impaired ability to discriminate pitch) have corresponding spatial deficits in some reports (Douglas & Bilkey, 2007; Särkämö et al., 2009), but others have failed to replicate the association (Tillmann et al., 2010; Williamson, Cocchini, & Stewart, 2011). People have been found to map musical pitch to vertical spatial locations (Melara & O'Brien, 1987; Pratt, 1930; Rusconi, Kwan, Giordano, Umiltà, & Butterworth, 2006), but they are also willing to map it to psychophysical luminosity and loudness (Hubbard, 1996; McDermott, Lehr, & Oxenham, 2008), and to words denoting emotion, size, sweetness, texture and temperature (Eitan & Timmers, 2010; Nygaard, Herold, & Namya, 2009; Walker & Smith, 1984). Thus, while pitch may be described in spatial terms such as “high” or “low”, it remains unclear whether pitch and space are merely two amongst many associated dimensions or whether the representation of pitch is fundamentally spatial.

As a psychoacoustic property corresponding to waveform frequency, the representation of pitch involves the primary auditory cortex, but the full neural specification of pitch processing is still not well understood (Bendor, 2012; Zatorre, Belin, & Penhune, 2002). Both medial and lateral Heschl's gyrus have been implicated

in pitch processing (e.g., Krumbholz, Patterson, Seither-Preisler, Lammertmann, & Lutkenhoner, 2003; Patterson, Uppenkamp, Johnsrude, & Griffiths, 2002), but so too has the planum temporale (Hall & Plack, 2009; Warren & Griffiths, 2003), within which overlapping areas respond to both pitch and spatial motion (Hart, Palmer, & Hall, 2004). Space is a physical property of the three-dimensional body we occupy and the world through which we move, and several researchers have argued that it is represented in a multimodal or supramodal system that takes input from vision, touch, sound, and other perceptual modalities in order to create a common spatial code (Bryant, 1992; Giudice, Betty, & Loomis, 2011; Lacey, Campbell, & Sathian, 2007; Renier et al., 2009; Struiksmá, Noordzij, & Postma, 2009). Numerous behavioural studies have shown that activating pitch also activates space along the vertical axis (with some cultural variation: Dolscheid, Shayan, Majid, & Casasanto, 2011; Eitan & Timmers, 2010). A high-pitch prime leads people to explicitly relate it to a high spatial location (Pratt, 1930), and to implicitly attend to a visual or tactile target in a high spatial location (Mossbridge, Grabowecky, & Suzuki, 2011; Occelli, Spenc, & Zampini, 2009; Walker et al., 2010), or make a manual response in a high spatial location (Lidji, Kolinsky, Lochy, & Morais, 2007; Rusconi et al., 2006). However, the above findings cannot distinguish between an *associative mapping* explanation, where representations of pitch and space are separate but linked, and a *shared representation* explanation, where pitch and space share common representational and processing resources.

According to an associative mapping explanation, the representation of musical pitch is purely auditory in nature. An

\* Corresponding author. Address: School of Psychological Sciences, University of Manchester, Oxford Road, M13 9PL, UK.

E-mail address: [louise.connell@manchester.ac.uk](mailto:louise.connell@manchester.ac.uk) (L. Connell).

individual's perception of a note's pitch would essentially comprise a modality-specific auditory representation of its sound frequency, and one would recall its pitch as a simulation (i.e., a partial replay of the neural activation that arose during experience: Barsalou, 1999) of that frequency. Perceiving a high pitch note rapidly activates a high spatial location because the two representational dimensions are directly associated, as are the dimensions of pitch and loudness (McDermott et al., 2008), or pitch and happiness (Eitan & Timmers, 2010). Notwithstanding these associations, pitch perception and discrimination itself remains an exclusively auditory matter. Conversely, a shared representation explanation for pitch/space effects would hold that the representation of musical pitch is audiospatial in nature. Here, an individual's perception of a note's pitch would comprise an audiospatial representation of both its sound frequency and its height on the vertical axis. One would then recall its pitch as an auditory and spatial simulation of that frequency and height. People may therefore be willing to map musical pitch to other dimensions because they all share a common spatial grounding (i.e., are mediated by space): for example, both loudness (Eitan, Schupak, & Marks, 2008) and emotional valence (Meier & Robinson, 2004) show similar effects to pitch in vertical space. Pitch perception and discrimination, therefore, is obligatorily audiospatial.

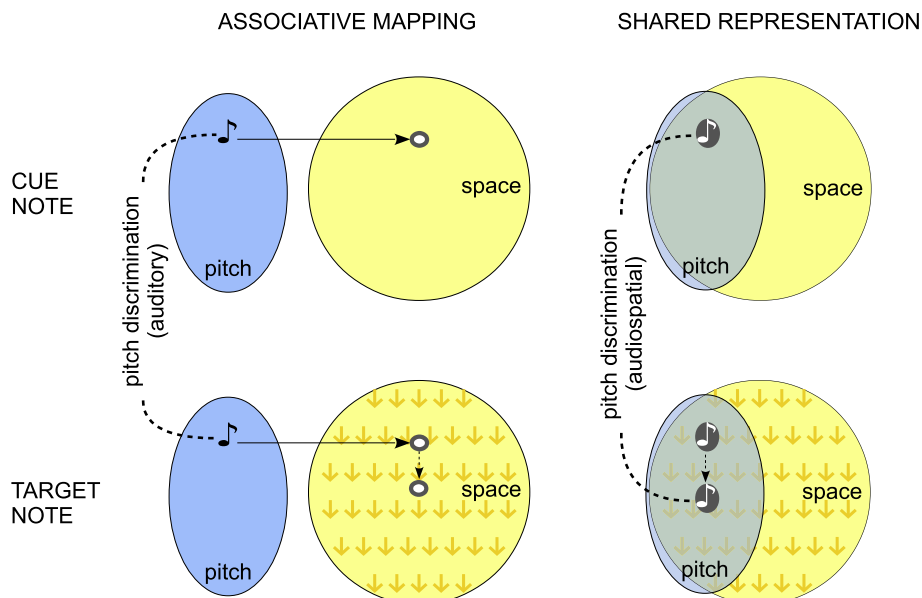
In the present studies, we aimed to distinguish between these two explanations by using a basic psychophysical task of pitch discrimination, where participants must judge whether a target vocal note is the same as (or different from) a preceding cue note. Importantly, target trials were presented as video clips where a singer sometimes gestured upward or downward while singing that target note, thus providing an alternative, concurrent source of spatial information. Signal detection analysis then allowed us to isolate the response criterion of pitch discrimination (i.e., the underlying bias towards the belief that pitch has or has not changed), for which the two accounts produced differing predictions. An associative mapping explanation of the pitch/space relationship would predict that a concurrent spatial stimulus should have no effect on response criterion. Because pitch representations are purely auditory (see Fig. 1), people discriminate pitch on the basis of audi-

tory frequency alone, meaning that pitch discrimination has an auditory response criterion. Hence, regardless of what other processing might be taking place in the spatial system, spatial movement has no power to bias pitch responses. Only in the shared representation account, where the audiospatial representation of musical pitch cannot be disentangled from the visuospatial representation of vertical gesture, would a criterion shift emerge. Because pitch representation is audiospatial (Fig. 1), pitch discrimination has a spatial response criterion, and people cannot discriminate pitch without being biased by concurrent spatial movement.

## 2. Experiment 1: Biasing pitch

In this and the following experiments, participants watched target trials of an actor gesturing while singing a particular musical note. Gestures frequently and effectively communicate spatial information to recipients that goes beyond what is conveyed in speech (Graham & Argyle, 1975; Holler, Shovelton, & Beattie, 2009; Ping & Goldin-Meadow, 2008), and may even be considered as explicit expressions of spatial action simulations (Hostetter & Alibali, 2008). Our nonlinguistic combination of gesture and pitch stimuli therefore allowed us to embed spatial information in a naturalistic context to which people are sensitive, but in a less obtrusive manner than pairing pitch with (for example) geometric shapes (see also Cai, Connell, & Holler, 2012).

Our hypotheses were simple. If the shared representation explanation is correct and pitch representations are audiospatial, then spatial information in concurrent gesture should influence pitch discrimination in two specific ways. First, the spatial movement of gesture should bias participants towards the belief they had perceived a pitch movement (i.e., that the target note was different to the cue). Furthermore, participants should be sensitive to the direction of spatial movement, where downward gestures would make pitch appear lower in frequency, and upward gestures would make pitch appear higher in frequency. On the other hand, if the associative mapping explanation is correct, then none of these effects would appear.



**Fig. 1.** Schematic of representations involved in pitch discrimination, under the associative mapping and shared representations accounts, for a trial involving same-pitch notes and downward gesture. The cue note is presented as a plain sound file while the target note is presented in a video with concurrent spatial movement from the singer's gesture.

## 2.1. Method

### 2.1.1. Participants

Thirty-two native speakers of English from the University of Manchester took part in the experiment. Five were replaced when funnel debriefing indicated they were aware of the potential effect the gestures could have had on their pitch discrimination judgments. All were right-handed, had no hearing impairment, and were non-musicians (i.e., not musically trained). They received course credits or £4 for participation.

### 2.1.2. Materials

Target notes consisted of 16 vocal notes, sung by professional actors/singers on a major scale from A2 (110 Hz) to A3 (220 Hz) for the male actor, and from A3 (220 Hz) to A4 (440 Hz) for the female actor. The fundamental frequency of these vocal notes was a maximum of 17 cents (17% of a whole tone) away from the intended pitch. Each actor was filmed while singing and moving the right hand downward or upward for the duration of the note, (i.e., downward or upward gesture), or resting their hands naturally on the lap (i.e., no gesture) (see Fig. 2). In order to ensure stimulus consistency in gestural and vocal behaviour, we separated the audio and video tracks and overdubbed the best gesture videos with the best target notes, and ensured each final stimulus was a seamless synchronisation of mouth movement, gesture movement, and sung vocal. All 48 target videos lasted 1.4 s.

Cue notes consisted of synthesized notes at the same fundamental frequencies as the target notes, created with Garageband software with the Classical Ensemble voice (which sounded like a mixed choir of male and female vocalists). We chose to use synthesized human voices in order to avoid the spatial pitch characteristics associated with musical instruments (e.g., horizontal for a piano, vertical for a clarinet), and to give cue notes a similar timbre to target notes while still allowing us to use the same type of cue for male and female actors' notes. We then edited the synthesized cue notes in Audacity to replicate the target sung notes' frequency exactly (same pitch), or to shift them one semitone up (higher pitch) or down (lower pitch).

Cue and target stimuli were paired so that each cue note was followed by a target note of the same pitch, higher pitch, or lower pitch (accompanied by a downward gesture, no gesture, or upward gesture), resulting in 144 cue-target pairs. We divided these 144 pairs into two materials lists, where both lists included all 48 same-pitch pairs, and the remaining stimuli were distributed so each list had 24 higher-pitch and 24 lower-pitch pairs (i.e., an equal number of same and different pitch).

### 2.1.3. Procedure

Participants were instructed that they should watch videos of professional actors singing musical notes, and, in each case, judge

as quickly and accurately as possible whether the actor's sound was the same pitch as an earlier musical note. The experiment was run with Superlab 4.0 on a MacBook laptop, with videos displayed onscreen at approximately  $14 \times 10.5$  cm. Participants were randomly assigned to one of the two material lists and were tested individually in a lab cubicle. In each trial, they first saw a fixation cross for 500 ms, then heard the synthesized note, and immediately afterwards saw the target note video. After the video, a screen appeared with the prompt "SAME DIFFERENT", and participants were asked to press the left-hand key on a response box if they thought the actor's sound was the same pitch as the earlier musical note, or the right-hand key if they thought it was a different pitch to the earlier musical note (left/right mapping to same/different responses counterbalanced across participants). If participants pressed the "different" key, another screen appeared with the prompt "HIGHER LOWER", and participants were asked to press the left-hand key if they thought the actor's sound was a higher pitch than the earlier musical note, or the right-hand key if they thought it was a lower pitch (left/right mapping to higher/lower responses counterbalanced across participants). There was a blank of 500 ms between trials.

Within each materials list, stimuli were arranged into six blocks so that each of the 16 target notes appeared once per block (gestures counterbalanced). The order of blocks was fixed but presentation of trials within a block was randomized per participant. Participants performed four practice trials before the main experiment (two of which later re-appeared as experimental trials), and the whole procedure lasted for about 15 min.

### 2.1.4. Design & analysis

We ran two stages of analysis of variance, each with a single within-participants factor of gesture (downward, no-gesture, upward) and effect sizes reported as partial eta-squared ( $\eta_p^2$ ). First, signal detection analysis examined performance on the same/different judgments to determine if gesture affected people's response bias and sensitivity in pitch discrimination. "Different" responses to different-pitch targets constituted hits, and those to same-pitch targets constituted false alarms. For each participant, we then calculated criterion  $c$  (criterion or bias) and  $d'$  (sensitivity) statistics for each gesture condition (e.g., Stanislaw & Todorov, 1999).

Second, we examined the trajectory of error to determine whether downward gestures made notes seem lower than they really were (and upward gestures higher). Each error in the same/different and higher/lower judgments represented an upward or downward response trajectory: for example, a downward trajectory was one where (1) a same-pitch target was judged to be lower in pitch, (2) a higher-pitch target was judged to be the same pitch, or (3) a higher-pitch target was judged to be lower in pitch. For each participant, we calculated the proportion of downward errors out of all errors in each gesture condition. Four participants

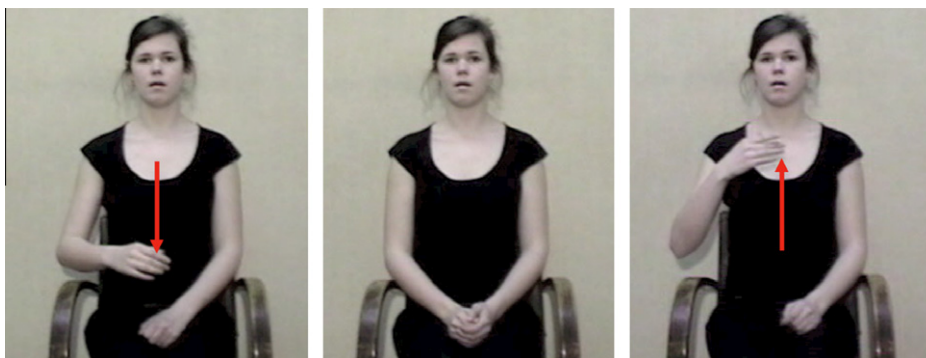
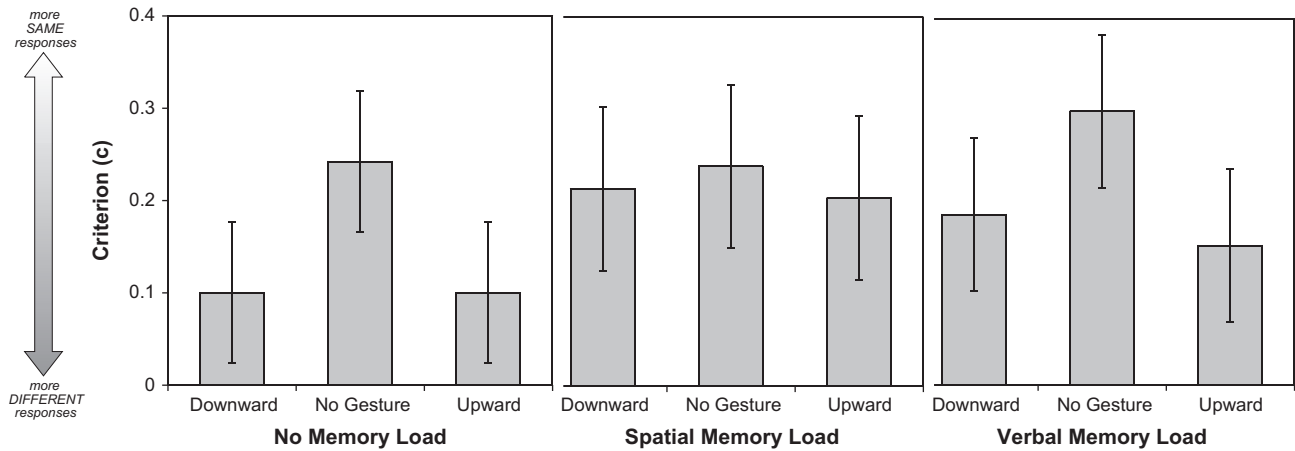
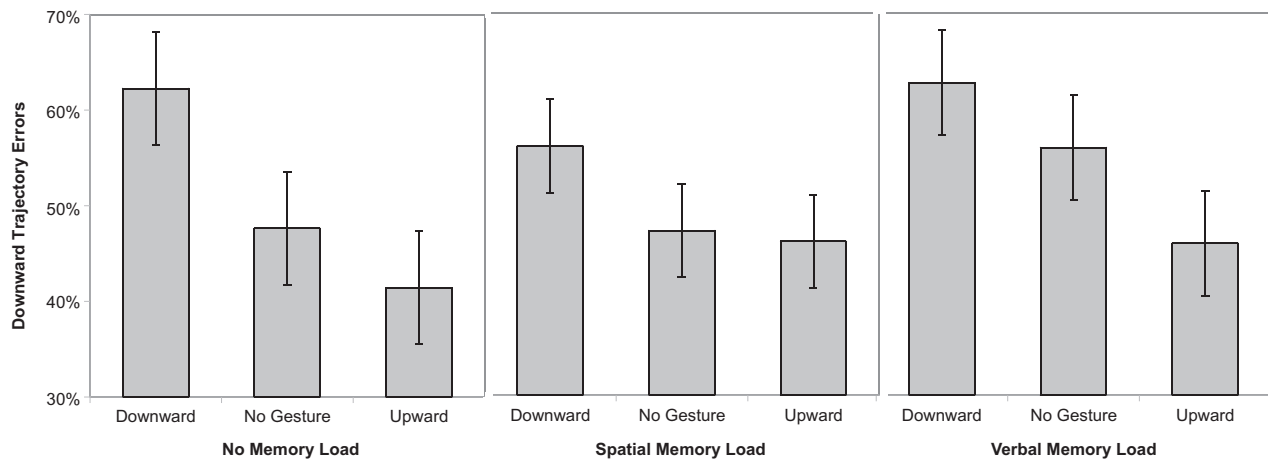


Fig. 2. Stills from video stimuli, showing a singer gesturing downward, at rest with no gesture, and gesturing upward. Arrows indicate extent and direction of movement.



**Fig. 3.** Response criterion in pitch discrimination (i.e., bias towards belief that target note was same as/different to cue) per gesture condition in Experiment 1–3. Error bars show within-participants 95% confidence intervals (Loftus & Masson, 1994).



**Fig. 4.** Proportion of pitch discrimination errors that expressed a downward trajectory (i.e., where participants thought the note was lower in pitch than reality) per gesture condition in Experiments 1–3. Error bars show within-participants 95% confidence intervals (Loftus & Masson, 1994).

with empty cells (i.e., perfect accuracy in one or more conditions) were excluded from trajectory analysis.

## 2.2. Results & discussion

People found the pitch discrimination task moderately difficult, with overall accuracy of 71.1%. Signal detection analysis supported the shared representation prediction that the spatial movement in gesture would affect pitch discrimination. There was a criterion difference between gesture types,  $F(2, 62) = 4.57$ ,  $p = .014$ ,  $\eta_p^2 = .129$ , as shown in Fig. 3 (left panel). Most trials showed a bias towards “same” responses (i.e.,  $c > 0$ ), but planned comparisons showed this bias was weaker for downward ( $p = .006$ ,  $\eta_p^2 = .187$ ) and upward ( $p = .011$ ,  $\eta_p^2 = .156$ ) gestures compared to when notes were unaccompanied by gesture. Upwards and downward gestures had the same response bias ( $p = .999$ ). Participants’ increased propensity to make “different” responses in the presence of gesture did not affect their overall sensitivity in pitch discrimination,  $F(2, 62) = 2.04$ ,  $p = .139$ ,  $\eta_p^2 = .062$ , with equivalent performance in no-gesture ( $d' = 1.79$ ), downward ( $d' = 1.99$ ) and upward ( $d' = 1.83$ ) gesture conditions.

Analysis of error trajectory also followed predictions (see Fig. 4, left panel). The nature of errors that people made was influenced by gesture,  $F(2, 54) = 9.23$ ,  $p < .001$ ,  $\eta_p^2 = .255$ . Specifically, planned comparisons showed that, relative to the no-gesture condition,

downward gestures increased the number of downward trajectory errors ( $p = .007$ ,  $\eta_p^2 = .205$ ) while upward gestures reduced them ( $p = .043$ ,  $\eta_p^2 = .105$ ).<sup>1</sup>

## 3. Experiment 2: Spatial memory load

If the shared representation explanation of pitch/space effects is correct, then the criterion shift and error trajectory in Experiment 1 emerge from an overlapping spatial representation of gestural movement and pitch. A spatial memory load should therefore attenuate these effects by occupying resources required for audio-spatial pitch discrimination. Holding a spatial load in memory should remove the biasing effect of spatial movement on pitch discrimination, meaning that people will remain quite liberal in their tendency to assume that notes are the same. Consequently, the direction of spatial movement should no longer drive the trajectory of error to the same extent.

<sup>1</sup> Although our participants were not musically trained, this fact did not preclude some level of knowledge about music; at the end of the experiment, we therefore gave participants a questionnaire to probe their exposure to music instruction (e.g., experience of playing a musical instrument, ability to distinguish pitch differences in staff notation). Musical knowledge was unrelated to either global response criterion  $r(30) = -.055$ ,  $p = .765$ , or downward error trajectory  $r(26) = -.182$ ,  $p = .354$ , though it did correlate positively with overall sensitivity,  $r(30) = .597$ ,  $p < .001$ .



### 3.1. Method

#### 3.1.1. Participants

Thirty-two new participants took part under the same criteria as Experiment 1. Five participants were replaced for awareness of the gesture effect. All had adequate recall of the spatial memory load (i.e., correctly recalled four or more out of six grids, see Section 3.1.2).

#### 3.1.2. Materials

Stimuli were as per Experiment 1. In addition, items in the spatial memory task consisted of six different 3-by-3 grids (plus one for practice) in which five random cells had been filled with an X.

#### 3.1.3. Procedure

Instructions were identical to Experiment 1 except that participants were asked to hold in memory a visually-presented spatial grid during each block of the pitch discrimination task. Before each of the six blocks, participants saw a spatial grid onscreen and could study it until they were satisfied they had memorised it. At the end of the block, participants were asked to recall the grid by drawing the positions of the Xs on a blank grid; these were later coded for accuracy (a grid must be perfectly recalled to qualify as an accurate response). The experiment lasted approximately 20 min.

#### 3.1.4. Design & analysis

As in Experiment 1. Six participants with perfect accuracy in one or more conditions were excluded from trajectory analysis.

### 3.2. Results & discussion

Overall accuracy was similar to Experiment 1 at 73.8%. Accuracy in the spatial memory task was also high ( $M = 93.8\%$ ,  $SD = 12.5\%$ ). Signal detection analysis confirmed the predictions of the shared representation account that a spatial memory load would eliminate the biasing effect of spatial movement on pitch discrimination. There was no longer any criterion difference between gesture types,  $F(2, 62) = 0.15$ ,  $p = .856$ ,  $\eta_p^2 = .005$  (see Fig. 3, centre panel): a similar bias towards “same” responses appeared for downward, upward and no-gesture conditions (all  $ps > .3$ ,  $\eta_p^2 s < .009$ ). Sensitivity of pitch discrimination was unaffected by gesture,  $F(2, 62) = 1.11$ ,  $p = .176$ ,  $\eta_p^2 = .054$ : no-gesture  $d' = 2.01$ , downward gesture  $d' = 1.92$ , upward gesture  $d' = 1.88$ .

Analysis of error trajectory showed attenuated effects compared to Experiment 1 (see Fig. 4, centre panel). Spatial movement in gesture had an influence on the direction of error,  $F(2, 50) = 3.61$ ,  $p = .034$ ,  $\eta_p^2 = .191$ . Downward gestures led to more downward trajectory errors than no gesture ( $p = .034$ ,  $\eta_p^2 = .127$ ), but upward gestures did not reduce their occurrence relative to no gesture ( $p = .406$ ,  $\eta_p^2 = .002$ ).<sup>2</sup>

### 4. Experiment 3: Verbal memory load

While the results of Experiment 1 support the shared representation account of pitch/space effects, it is possible that participants were silently labelling the pitch of the target notes as “higher” or “lower” in preparation for the discrimination task. The spatial movement in gesture could then have interacted with the representation of this verbal label rather than inducing a bias in pitch discrimination itself. We therefore examined the origin of the cri-

terion shift by replicating the task while participants held a verbal load in memory to block a linguistic labelling strategy. If the shared representation explanation is correct, then the criterion shift of Experiment 1 should emerge unscathed. Furthermore, if the criterion shift re-emerges under a verbal memory load, it will verify that the cancelled effects in Experiment 2 were not due to generic processing difficulties under memory load conditions but rather were specific to spatial content.

### 4.1. Method

#### 4.1.1. Participants

Thirty-two new participants took part under the same criteria as Experiment 1. Three participants were replaced for inadequate recall of the verbal memory load (i.e., anyone who recalled fewer than four out of six diphone sequences, see Section 3.1.2).

#### 4.1.2. Materials

Stimuli were as per Experiment 1. In addition, items in the verbal memory task consisted of six different sequences of three nonsense diphones (e.g., [tɛ kæ vo]); one further sequence was used for practice. Each sequence was recorded by a male speaker of British English with clear enunciation.

#### 4.1.3. Procedure

Instructions were identical to Experiment 1 except that participants were asked to hold in memory an auditorily-presented diphone sequence during each block. Before each of the six blocks, participants listened to a diphone sequence three times and repeated it back to the experimenter (second author); if there were any errors in repetition, the experimenter enunciated the sequence again until participants got it right. At the end of each block, participants recalled aloud the memorised diphone sequence to the experimenter who transcribed it and later coded it for accuracy (a sequence must be perfectly recalled to qualify as an accurate response). Participants were familiarised with diphone recall during the practice session. The experiment took approximately 20 min to complete.

#### 4.1.4. Design & analysis

As in Experiment 1. Three participants with perfect accuracy in one or more conditions were excluded from trajectory analysis.

### 4.2. Results & discussion

Overall accuracy was similar to Experiment 1 at 69.3%. Accuracy in the present verbal memory task ( $M = 89.6\%$ ,  $SD = 12.5\%$ ) was equivalent to Experiment 2's spatial memory task,  $t(62) = 1.33$ ,  $p = .188$ , confirming the memory load was comparable in difficulty. Signal detection analysis replicated the findings of Experiment 1, and confirmed that the biasing effect of spatial movement on pitch discrimination was not due to a verbal labelling strategy. Fig. 3 (right panel) shows the criterion difference emerged between gesture types,  $F(2, 62) = 3.39$ ,  $p = .040$ ,  $\eta_p^2 = .098$ . As before, the bias towards “same” responses was weaker for downward ( $p = .040$ ,  $\eta_p^2 = .095$ ) and upward ( $p = .009$ ,  $\eta_p^2 = .168$ ) gestures compared to when notes were unaccompanied by gesture. Upwards and downward gestures had the same response bias ( $p = .549$ ). Since these “different” responses were distributed across both correct and incorrect trials, sensitivity of pitch discrimination did not change with gestural movement,  $F(2, 62) = 1.78$ ,  $p = .176$ ,  $\eta_p^2 = .054$ , with equivalent performance in no-gesture ( $d' = 1.68$ ), downward ( $d' = 1.89$ ) and upward ( $d' = 1.84$ ) gesture conditions.

Analysis of error trajectory again replicated Experiment 1 (see Fig. 4, right panel). The nature of errors in pitch discrimination was influenced by the spatial movement in gesture,  $F(2,$

<sup>2</sup> Musical knowledge was again unrelated to response criterion  $r(30) = -.196$ ,  $p = .282$ , and error trajectory  $r(24) = .084$ ,  $p = .984$ , and correlated with overall sensitivity,  $r(30) = .546$ ,  $p = .001$ . Furthermore, accuracy in the spatial memory task was not reliably correlated with sensitivity,  $r(30) = .158$ ,  $p = .388$ , response criterion,  $r(30) = -.167$ ,  $p = .361$ ], nor error trajectory,  $r(24) = .214$ ,  $p = .294$ .

56) = 6.62,  $p = .003$ ,  $\eta_p^2 = .191$ . Downward gestures marginally increased the frequency of downward trajectory errors compared to no gesture ( $p = .052$ ,  $\eta_p^2 = .092$ ) while upward gestures reduced their occurrence ( $p = .034$ ,  $\eta_p^2 = .115$ ).<sup>3</sup>

## 5. General discussion

In the present paper, we show that concurrent visuospatial movement biases pitch discrimination. Viewing upward and downward gestures biased people towards believing they had perceived a change in pitch, despite an underlying tendency to assume that all notes were the same. Indeed, when we examined the pattern of errors that people made, we found that the direction of gesture was also driving the direction of error: downward gestures made notes seem lower in pitch than they really were, and upward gestures made notes seem higher in pitch than they really were. These effects were not due to a verbal labelling strategy as they were preserved under verbal memory load. However, their disappearance under spatial memory load conditions indicates that the biasing effect is spatial in origin. Together, these findings support the shared representation explanation for the relationship between pitch and space.

When people hear a musical note, its pitch is not just represented in the auditory modality. Rather, its representation is audiospatial, in that it comprises both an auditory and a spatial representation of the note's frequency. However, things become more complicated when people watch someone singing a note. On the one hand, if the singer remains still, then the same story applies: the audiospatial representation still reflects the note's pitch. But, on the other hand, if the singer gestures with an upward or downward movement, then both the visual gesture and auditory note require representational resources in the vertical spatial axis. Hence, the spatial information in the gesture is co-perceived with that in the note, and results in an audiospatial representation of the note's pitch that has been modulated by the direction of visuospatial movement. While many previous studies have examined the relationship between pitch and vertical space, they could not determine the nature of pitch representation because both associative mappings and shared representations would lead a pitch stimulus to prime its corresponding spatial location and facilitate motor responses to that location (e.g., Rusconi et al., 2006). However, a mapping from high pitch to high spatial location would be static, and could not explain why the spatial movement in gesture biased participants towards believing they had perceived a movement in pitch. It could be argued that participants merely attended to the end point of gestural movement, and hence that our effects are still location-based (i.e., low spatial position biases judgement towards low pitch) rather than movement-based (i.e., downward movement biases judgement towards lower pitch). If this were the case, then our no-gesture condition, where the actors rested their hands in a clearly-visible low spatial location (i.e., their laps: see Fig. 1), would have produced similar effects to downward gestures, but this did not occur. Rather, downward gestures significantly differed from no gestures in shifting response criterion and error trajectory as we predicted. A dynamic, shared representation of pitch and space, where pitch is represented not only in terms of spatial position but also movement and direction, is consistent with our results.

Indeed, this notion of dynamic pitch representations equating to spatial movement is also consistent with prior research regard-

ing the function of the planum temporale. As well as its involvement in pitch processing (Hall & Plack, 2009; Hart et al., 2004; Warren & Griffiths, 2003), the planum temporale is also activated by auditory motion (i.e., when the point of origin of a sound appears to change position: Alink, Singer, & Muckli, 2008; Deouell, Heller, Malach, D'Esposito, & Knight, 2007; Hart et al., 2004; Warren, Zielinski, Green, Rauschecker, & Griffiths, 2002) and – critically – even by visuospatial motion (Howard et al., 1996; see also Griffiths & Warren, 2002). The potential role of the planum temporale in a common spatial processing system (e.g., Bryant, 1992; Giudice et al., 2011; Lacey et al., 2007; Renier et al., 2009; Struikma et al., 2009), and the precise neural mechanisms underlying our present findings, should be further explored.

There are several possibilities as to how and why musical pitch is represented in vertical space, and not in some other spatial dimension. When speaking, producing a pitch higher than normal voice frequency moves the larynx upward in the throat, and producing a lower pitch moves it downward. Furthermore, breathing from the top of the lungs by raising and lowering the shoulders tends to produce higher-pitch vocal notes, while breathing from the bottom of the lungs by tensing and relaxing the thoracic diaphragm tends to produce lower-pitch, resonant notes. Thus, cumulative experience with our own voices provides a possible vertical grounding for vocal pitch, which could then generalise to pitch of other people's voices, musical instruments, and so on. Indeed, watching other people exhibit such behaviours could help support this bodily grounding of pitch in vertical space, given that facial expressions and head movements can provide useful cues to vocal pitch (e.g., Munhall, Jones, Callan, Kuratate, & Vatikiotis-Bateson, 2004; Thompson & Russo, 2007). The idea that spatial representations of pitch are based in bodily experience is consistent with findings that tilting the head 90° eliminates the usual effect of high pitch activating a high spatial location (Mossbridge et al., 2011). Research on pitch and space tends to focus on listening to auditory pitch from an outside source rather than producing pitch with one's own voice, and the present studies are no exception; however, an audiospatial representation of pitch that is grounded in bodily experience would predict similar findings should emerge in pitch production.

While some have claimed the appearance of pitch/space effects in young infants means the connection between domains is innate (Walker et al., 2010), even 3–4 month old babies have considerable experience of vocalisation. A conservative estimate of 1 h per day crying, fussing and other vocalising (e.g., Michelsson, Rinne, & Paajanen, 1990) provides a 4-month old infant with over 100 h experience of vocal pitch under various body configurations. Since infants of that age can learn statistical regularities in the environment with only a few minutes' exposure (Kirkham, Slemmer, & Johnson, 2002), it seems premature to assume they could not have learned to represent pitch spatially. Indeed, other sources of experience, such as language (Dolscheid et al., 2011) or training with a horizontal musical instrument like a piano (Lidji et al., 2007), appear to offer opportunities for people to learn alternative spatial representations of pitch.

Future research will need to determine whether pitch/space effects emerge from a learned or innate mechanism, but, whatever their origin, the present paper demonstrates that pitch perception is fundamentally audiospatial. The nature of the link between musical and spatial processing is one of shared representation.

## Acknowledgments

This work was supported by a research project grant from the Leverhulme Trust (F/00 120/CA) to the first and last authors. Thanks to Dermot Lynott for comments. Correspondence concerning this article should be addressed to Louise Connell, School of

<sup>3</sup> As before, musical knowledge was unrelated to response criterion  $r(30) = -.040$ ,  $p = .828$ , or error trajectory  $r(27) = .044$ ,  $p = .820$ , and correlated with overall sensitivity,  $r(30) = .394$ ,  $p = .026$ . Accuracy in the verbal memory task did not reliably correlate with sensitivity,  $r(30) = .021$ ,  $p = .909$ , response criterion,  $r(30) = -.246$ ,  $p = .175$ , nor error trajectory,  $r(27) = -.251$ ,  $p = .189$ .

Psychological Sciences, University of Manchester, Oxford Road, M13 9PL, UK. Email: louise.connell@manchester.ac.uk.

## References

- Alink, A., Singer, W., & Muckli, L. (2008). Capture of auditory motion by vision is represented by an activation shift from auditory to visual motion cortex. *Journal of Neuroscience*, *28*, 2690–2697.
- Barsalou, L. W. (1999). Perceptual symbol systems. *Behavior and Brain Sciences*, *22*, 577–660.
- Bendor, D. (2012). Does a pitch center exist in auditory cortex? *Journal of Neurophysiology*, *107*, 743–746.
- Bryant, D. J. (1992). A spatial representation system in humans. *Psychology*, *3*(16), space 1.
- Cai, Z. G., Connell, L., & Holler, J. (2012). Time does not flow without language: Spatial distance affects temporal duration regardless of movement or direction, submitted for publication.
- Deouell, L. Y., Heller, A. S., Malach, R., D'Esposito, M., & Knight, R. T. (2007). Cerebral responses to change in spatial location of unattended sounds. *Neuron*, *55*, 985–996.
- Dolscheid, S., Shayan, S., Majid, A., & Casasanto, D. (2011). The thickness of musical pitch: Psychophysical evidence for the Whorfian hypothesis. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd annual conference of the cognitive science society* (pp. 537–542). Austin, TX: Cognitive Science Society.
- Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature Neuroscience*, *10*, 915–921.
- Eitan, Z., Schupak, A., & Marks, L. E. (2008). Louder is higher: Cross-modal interaction of loudness change and vertical motion in speeded classification. In K. Miyazaki, Y. Hiraga, M. Adachi, Y. Nakajima, & M. Tsuzaki (Eds.), *Proceedings of the 10th international conference on music perception and cognition* (pp. 1–10). Adelaide, Australia: Causal Productions.
- Eitan, Z., & Timmers, R. (2010). Beethoven's last piano sonata and those who follow crocodiles: Cross-domain mappings of auditory pitch in a musical context. *Cognition*, *114*, 405–422.
- Giudice, N. A., Betty, M. R., & Loomis, J. M. (2011). Functional equivalence of spatial images from touch and vision: Evidence from spatial updating in blind and sighted individuals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 621–634.
- Graham, J. A., & Argyle, M. (1975). A cross-cultural study of the communication of extra-verbal meaning by gestures. *International Journal of Psychology*, *10*, 57–67.
- Griffiths, T. D., & Warren, J. D. (2002). The planum temporale as a computational hub. *Trends in Neurosciences*, *25*, 348–353.
- Hall, D. A., & Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cerebral Cortex*, *19*, 576–585.
- Hart, H. C., Palmer, A. R., & Hall, D. A. (2004). Different areas of human non-primary auditory cortex are activated by sounds with spatial and nonspatial properties. *Human Brain Mapping*, *21*, 178–190.
- Holler, J., Shovelton, H., & Beattie, G. (2009). Do iconic hand gestures really contribute to the communication of semantic information in a face-to-face context? *Journal of Nonverbal Behavior*, *33*, 73–88.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin & Review*, *15*, 495–514.
- Howard, R. J., Brammer, M., Wright, I., Woodruff, P. W., Bullmore, E. T., & Zeki, S. (1996). A direct demonstration of functional specialization within motion-related visual and auditory cortex of the human brain. *Current Biology*, *6*, 1015–1019.
- Hubbard, T. L. (1996). Synesthesia-like mappings of lightness, pitch, and melodic interval. *American Journal of Psychology*, *109*, 219–238.
- Kirkham, N. Z., Slemmer, J. A., & Johnson, S. P. (2002). Visual statistical learning in infancy: Evidence of a domain general learning mechanism. *Cognition*, *83*, B35–B42.
- Krumbholz, K., Patterson, R. D., Seither-Preisler, A., Lammertmann, C., & Lutkenhoner, B. (2003). Neuromagnetic evidence for a pitch processing center in Heschl's gyrus. *Cerebral Cortex*, *13*, 765–772.
- Lacey, S., Campbell, C., & Sathian, K. (2007). Vision and touch: Multiple or multisensory representations of objects? *Perception*, *36*, 1513–1521.
- Lidji, P., Kolinsky, R., Lochy, A., & Morais, J. (2007). Spatial associations for musical stimuli: A piano in the head? *Journal of Experimental Psychology: Human Perception and Performance*, *33*, 1189–1207.
- Loftus, G. R., & Masson, M. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, *1*, 476–490.
- McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*, *19*, 1263–1271.
- Meier, B. P., & Robinson, M. D. (2004). Why the sunny side is up: Associations between affect and vertical position. *Psychological Science*, *15*, 243–247.
- Melara, R., & O'Brien, T. (1987). Interaction between synesthetically corresponding dimensions. *Journal of Experimental Psychology: General*, *116*, 323–336.
- Michelsson, K., Rinne, A., & Paajanen, S. (1990). Crying, feeding and sleeping patterns in 1 to 12-month-old infants. *Child: Care Health and Development*, *16*, 99–111.
- Mossbridge, J. A., Grabowecy, M., & Suzuki, S. (2011). Changes in auditory frequency guide visual-spatial attention. *Cognition*, *121*, 133–139.
- Munhall, K. G., Jones, J. A., Callan, D. E., Kuratate, T., & Vatikiotis-Bateson, E. (2004). Visual prosody and speech intelligibility: Head movement improves auditory speech perception. *Psychological Science*, *152*, 133–137.
- Nygaard, L. N., Herold, D. S., & Namya, L. L. (2009). The semantics of prosody: Acoustic and perceptual evidence of prosodic correlates to word meaning. *Cognitive Science*, *33*, 127–146.
- Ocelli, V., Spenc, C., & Zampini, M. (2009). Compatibility effects between sound frequency and tactile elevation. *NeuroReport*, *20*, 793–797.
- Patterson, R. D., Uppenkamp, S., Johnsrude, I. S., & Griffiths, T. D. (2002). The processing of temporal pitch and melody information in auditory cortex. *Neuron*, *36*, 767–776.
- Ping, R. M., & Goldin-Meadow, S. (2008). Hands in the air: Using ungrounded iconic gestures to teach children conservation of quantity. *Developmental Psychobiology*, *44*, 1277–1287.
- Pratt, C. C. (1930). The spatial character of high and low tones. *Journal of Experimental Psychology*, *13*, 278–285.
- Renier, L. A., Anurova, I., De Volder, A. G., Carlson, S., VanMeter, J., & Rauschecker, J. P. (2009). Multisensory integration of sounds and vibrotactile stimuli in processing streams for “what” and “where”. *Journal of Neuroscience*, *29*, 10950–10960.
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., & Butterworth, B. (2006). Spatial representation of pitch height: The SMARC effect. *Cognition*, *99*, 113–129.
- Särkämö, T., Tervaniemi, M., Soimila, S., Autti, T., Silvennoinen, H. M., Laine, M., et al. (2009). Cognitive deficits associated with acquired amusia after stroke: A neuropsychological follow-up study. *Neuropsychologia*, *47*, 2642–2651.
- Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior Research Methods, Instruments & Computers*, *31*, 137–149.
- Struiksma, M. E., Noordzij, M. L., & Postma, A. (2009). What is the link between language and spatial images? Behavioral and neural findings in blind and sighted individuals. *Acta Psychologica*, *132*, 145–156.
- Thompson, W. F., & Russo, F. A. (2007). Facing the music. *Psychological Science*, *18*, 756–757.
- Tillmann, B., Jolicœur, P., Ishihara, M., Gosselin, N., Bertrand, O., et al. (2010). The amusic brain: Lost in music, but not in space. *PLoS ONE*, *5*(4), e10173.
- Walker, P., Brenner, G., Spring, J., Maststock, K., Slater, A., & Johnson, S. (2010). Preverbal infants' sensitivity to synaesthetic cross modality correspondences. *Psychological Science*, *21*, 21–25.
- Walker, P., & Smith, S. (1984). Stroop interference based on the synaesthetic qualities of auditory pitch. *Perception*, *13*, 75–81.
- Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *Journal of Neuroscience*, *23*, 5799–5804.
- Warren, J. D., Zielinski, B. A., Green, G. G. R., Rauschecker, J. P., & Griffiths, T. D. (2002). Perception of sound-source motion by the human brain. *Neuron*, *34*, 139–148.
- Williamson, V., Cocchini, G., & Stewart, L. (2011). The relationship between pitch and space in congenital amusia. *Brain and Cognition*, *76*, 70–76.
- Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*, 37–46.