



Using spatial manipulation to examine interactions between visual and auditory encoding of pitch and time

Neil M. McLachlan, Loretta J. Greco, Emily C. Toner and Sarah J. Wilson*

Psychological Sciences, The University of Melbourne, Melbourne, VIC, Australia

Edited by:

Tim Griffiths, Newcastle University, UK

Reviewed by:

Catherine (Kate) J. Stevens, University of Western Sydney, Australia

Robert J. Zatorre, McGill University, Canada

***Correspondence:**

Sarah J. Wilson, Psychological Sciences, The University of Melbourne, Parkville, Melbourne, 3010 VIC, Australia.

e-mail: sarahw@unimelb.edu.au

Music notations use both symbolic and spatial representation systems. Novice musicians do not have the training to associate symbolic information with musical identities, such as chords or rhythmic and melodic patterns. They provide an opportunity to explore the mechanisms underpinning multimodal learning when spatial encoding strategies of feature dimensions might be expected to dominate. In this study, we applied a range of transformations (such as time reversal) to short melodies and rhythms and asked novice musicians to identify them with or without the aid of notation. Performance using a purely spatial (graphic) notation was contrasted with the more symbolic, traditional western notation over a series of weekly sessions. The results showed learning effects for both notation types, but performance improved more for graphic notation. This points to greater compatibility of auditory and visual neural codes for novice musicians when using spatial notation, suggesting that pitch and time may be spatially encoded in multimodal associative memory. The findings also point to new strategies for training novice musicians.

Keywords: spatial manipulation, visual, auditory, encoding, pitch, time, music, notation

INTRODUCTION

Rapid variations of pitch over time are a feature of many biological signals. Animals and humans learn to recognize patterns of pitch over time and attribute meaning to them. Species specific calls are an example of this behavior. Deutsch (1999) has documented recognition mechanisms that interact with memory traces of pitch over durations of up to 10 s when people listen to music. McLachlan and Wilson (2010) have proposed that recognition and identification mechanisms are likely to be central to most auditory processing tasks as these mechanisms rapidly form gestalts based on long-term memory to integrate and stream afferent information. Their object-attribute model (McLachlan and Wilson, 2010) extends “what” and “where” models of auditory processing (Rauschecker and Tian, 2000; Maeder et al., 2001; Griffiths and Warren, 2002; Zatorre et al., 2002; Arnott et al., 2004; Lomber and Malhotra, 2008) by conceptualizing pitch and loudness as spatial attributes that are bound with sound identities in auditory short-term memory (ASTM). These attributes are spatially represented by place-rate encoding in a multidimensional array (Gomes et al., 1995) that preserves sequential temporal order and corresponds to the subjective experience of sound. Subsequent to this, spatial encoding of pitch height has been demonstrated in humans and monkeys by depth electrode recordings from spatial arrays of sharply tuned neurons in the auditory core with response functions that are consistent with human subjective pitch responses (Bendor and Wang, 2005; Bitterman et al., 2008). This form of encoding will be referred to as spatial encoding in this paper. Given that ASTM is a buffer of pitch and other auditory feature information over time it follows that the spatial dimension of pitch height is sequentially encoded along a separate neural dimension of time, so that melodic information is preserved and may interact with long-term memory systems.

The object-attribute model proposes a dual pathway for higher level cortical processing of pitch, where one pathway involves categorical (symbolic) pattern recognition, and the other involves the spatial encoding of pitch strength along the continuous dimensions of pitch height and time. Vision has been shown to influence auditory processing at various stages of the auditory pathway (McGurk and MacDonald, 1976; O’Leary and Rhodes, 1984; Shamsa et al., 2002; Sherman, 2007; McLachlan and Wilson, 2010), and since visual fields are well known to be spatially encoded (Rolls and Deco, 2002), it is possible that spatial encoding of pitch and time by the auditory pathway may be influenced by the visual representation of music. Visual representations can vary in their extent of spatial (graphic) or symbolic notation (Larkin and Simon, 1987; Bauer and Johnson-Laird, 1993; Scaife and Rogers, 1996). For example, a written sequence of note names and duration symbols was commonly used in East Asia (Nelson, 2008). Such notations are efficient mnemonic aids for highly trained performers, since the complete spatial field for each music dimension does not need to be represented. This means that instrumental training in a particular music culture, and with a particular music notation involves increasing reliance on categorical and symbolic encoding strategies, as these reduce working memory load during performance. To date, music notation has received limited attention in cognitive neuroscience, despite its importance to music performance and education, and its potential to elucidate the relationships between fundamental perceptual mechanisms. This study will investigate whether music notations with linear spatial relationships to pitch and time facilitate performance on melodic and rhythmic processing tasks in novice musicians who have not formed symbolic associations for musical categories.

GRAPHIC AND MUSIC NOTATION SYSTEMS

Western music notation evolved over many centuries into a composite symbolic-spatial notation system. Pitch height is nominally spatially represented (proportional to vertical position) but is conditioned by symbolic information in the form of a key signature, sharps, and flats (**Figure 1**). Similarly, note duration is nominally spatially represented along a time-line, but is predominantly symbolically represented in the form of shading, stems, dots, and tails (**Figure 1**). Text is also commonly used to define tempo and loudness changes.

Gaare (1997) raised concerns about the limitations of western music notation. He suggested that the use of key signatures, sharps, and flats results in an additional processing step for the reader because their meanings must be applied to a note before it can be recognized, placing greater demands on working memory. Furthermore, the pitch intervals in the most commonly used western music scale (“C major”) are irregular, but their representation is equally spaced in western notation. This may produce incongruence between the notation and the corresponding auditory experience. Finally, rapid mathematical calculations must be performed to decipher note durations, which can be extremely taxing on working memory (Gaare, 1997). Sloboda (2005) proposed that many of the difficulties experienced in learning to read music are due to its symbolic nature and layout.

Graphic notation may facilitate problem solving to a greater extent than symbolic notation for a variety of reasons. Principal amongst these is the reduction in cognitive effort associated with the use of fewer symbols to search, recognize, and retain in working memory (Larkin and Simon, 1987; Bauer and Johnson-Laird, 1993; Scaife and Rogers, 1996). Koedinger and Anderson (1990) highlighted the facilitatory effects of perceptual congruence on cognitive performance associated with the use of graphic notation. With the advent of digital music systems it is now possible to tailor the extent of symbolic or graphic notation used to scaffold an individual’s learning relative to previous training. This, in turn, may support the development of representations to aid memory in performance or analyze musical structures.

STUDY RATIONALE AND HYPOTHESES

Symbolic notation may activate long-term memory for music identities, and so prime specific patterns of activation in ASTM. In contrast, graphic notation may prime novel spatial patterns of activation by direct spatial congruence between visual and auditory neural representations. In both cases the presence of a stable visual notation over time may help maintain activation of the memory trace in ASTM after the auditory stimulus has ceased. Since novice musicians do not have well developed long-term memories of music identities such as chords, notation is only likely to facilitate auditory processing of music if it contains congruent spatial information.

The primary aim of this study was to evaluate whether graphic notation of pitch and time facilitated cognitive processing of novel stimuli by novice musicians, as compared to western symbolic-spatial notation. An auditory analog of the visual mental rotations test (MRT) of Vandenberg and Kuse (1972) was developed to test spatial working memory (SWM) for auditory stimuli that

were presented with or without visual notation. Mental spatial manipulation requires multiple processing steps. First, a stable mental representation of the stimulus must be formed, followed by its mental spatial manipulation in SWM, and finally comparison with a referent stimulus (Hooven et al., 2004). We also used discrimination tasks to investigate the effect of notation on encoding *per se*. In discrimination with auditory only stimuli (AO) the representation is held in ASTM, whereas discrimination with audio-visual stimuli (AV) likely requires integration of multimodal inputs in multimodal associative memory (MAM, Pandya and Yeterian, 1985). Overall we expected the discrimination tasks to be easier than the manipulation tasks, and that performance of both discrimination and manipulation would improve in the presence of notation.

Two separate experiments were undertaken to explore the cognitive processing of spatially transformed melodic and rhythmic phrases. Inversion, transposition, and retrograde transforms were used for melodic stimuli, but only transposition and retrograde were possible for rhythmic stimuli. Participants were presented with a pair of stimuli and asked to identify the type of transformation that had been applied to the reference stimulus to generate the target. As baseline tasks, melodic, and rhythmic discrimination without notation were used to ensure there were no differences in the ability of participants to encode and use auditory information to perform cognitive manipulations. Discrimination and manipulation tasks were also administered with notation to measure any differences between participants exposed to graphic or western notation. Western music notation requires training to decode the meaning of many symbols, and is not visually symmetric for most of the transformations used in experiments 1 and 2. This means that novice musicians would be unable to perform the tasks using western notation in a visual only condition and thus, the MRT was used as a baseline task to measure visuo-spatial mental rotation ability (Douglas and Bilkey, 2007).

A second aim of our study was to compare the rates of learning between the notations. In keeping with Virsu et al. (2008), adaptation of visual and auditory neural encoding mechanisms to coordinate neural representations should improve performance for both notation types. A slower rate of learning for western notation would indicate the presence of additional learning processes involved in decoding symbolic meaning. Conversely, the ability to perform the manipulation tasks without evidence of learning over time would suggest reliance on one dominant modality.

We hypothesized that:

1. Discrimination performance would be better than manipulation regardless of the presence of notation.
2. The presence of notation would facilitate performance across discrimination and manipulation tasks.
 3. a. Graphic notation would facilitate melodic and rhythmic discrimination more than western notation.
 - b. Graphic notation would facilitate melodic and rhythmic manipulation more than western notation.
4. Melodic and rhythmic manipulation would improve over weekly sessions for both notation types, with graphic notation facilitating learning to a greater extent.

EXPERIMENT 1: IDENTIFICATION OF RHYTHMIC TRANSFORMATIONS

This experiment tested whether spatially mapped notations of inter-onset interval (IOI) patterns facilitated the identification of rhythmic transformations compared to western notation. Pairs of novel rhythms were presented and participants were asked to identify whether a transformation had been applied to the second rhythm in the stimulus pair. IOI patterns can only be transformed by reflecting their temporal order or shifting their phase, making chance performance high in a forced choice paradigm. Therefore the discrimination trials were included in the design to provide four alternate responses for each trial. To compare the efficacy of the two notation types, novice musicians were pseudo-randomly assigned to a western or graphic notation condition, keeping age and gender approximately equal.

METHODS

Participants

Twenty-eight participants took part in the experiment (see Results for demographic details). All participants considered themselves novice musicians. Four participants were recruited from the wider community and 24 were first year psychology students at The University of Melbourne who received course credit for participation. All were screened for music experience and music reading ability before assignment to one of the notation conditions. Seven participants understood written music but rated themselves as “not proficient” music readers and were evenly distributed between the groups. No participants reported experience with digital music systems such as musical instrument digital interface (MIDI) sequencers.

Stimuli

Ecologically valid rhythms were created according to the criterion of Pressing (1983) of approximately 50% density of accents to rests (that is, 50% of the musical beats are played). It was also important to prevent stimuli from being biased by properties of the notations themselves. For example, western rhythms have symbols that denote the iterative halving of beat durations whereas additional text denotes the division of beats into thirds (Figure 1), potentially facilitating rhythmic processing of phrases divisible by two. This problem was addressed by first generating a set of all possible IOI patterns (where accents and rests were denoted by a binary code for each beat). A computer program then selected plausible rhythmic stimuli according to the set of culturally non-specific criteria detailed in Table 1. A second set of criteria, detailed in Table 1, were developed according to the constraints of the manipulation task. IOI patterns that did not fulfill the specified criteria were eliminated, and the required number of patterns was then pseudo-randomly selected to ensure that rhythms ranged in complexity.

Ninety IOI patterns were selected for use in the experimental tasks and practice trials. A standard MIDI “woodblock” instrumental sound was used because of its sharp attack and familiarity in musical contexts. Four types of transformations were employed. They were; “same,” “different,” “backward,” or “shifted.” Shifted involved beginning the pattern at a different temporal location, backward involved reversing the temporal order of the pattern from

FIGURE 1 | An example of western music notation.

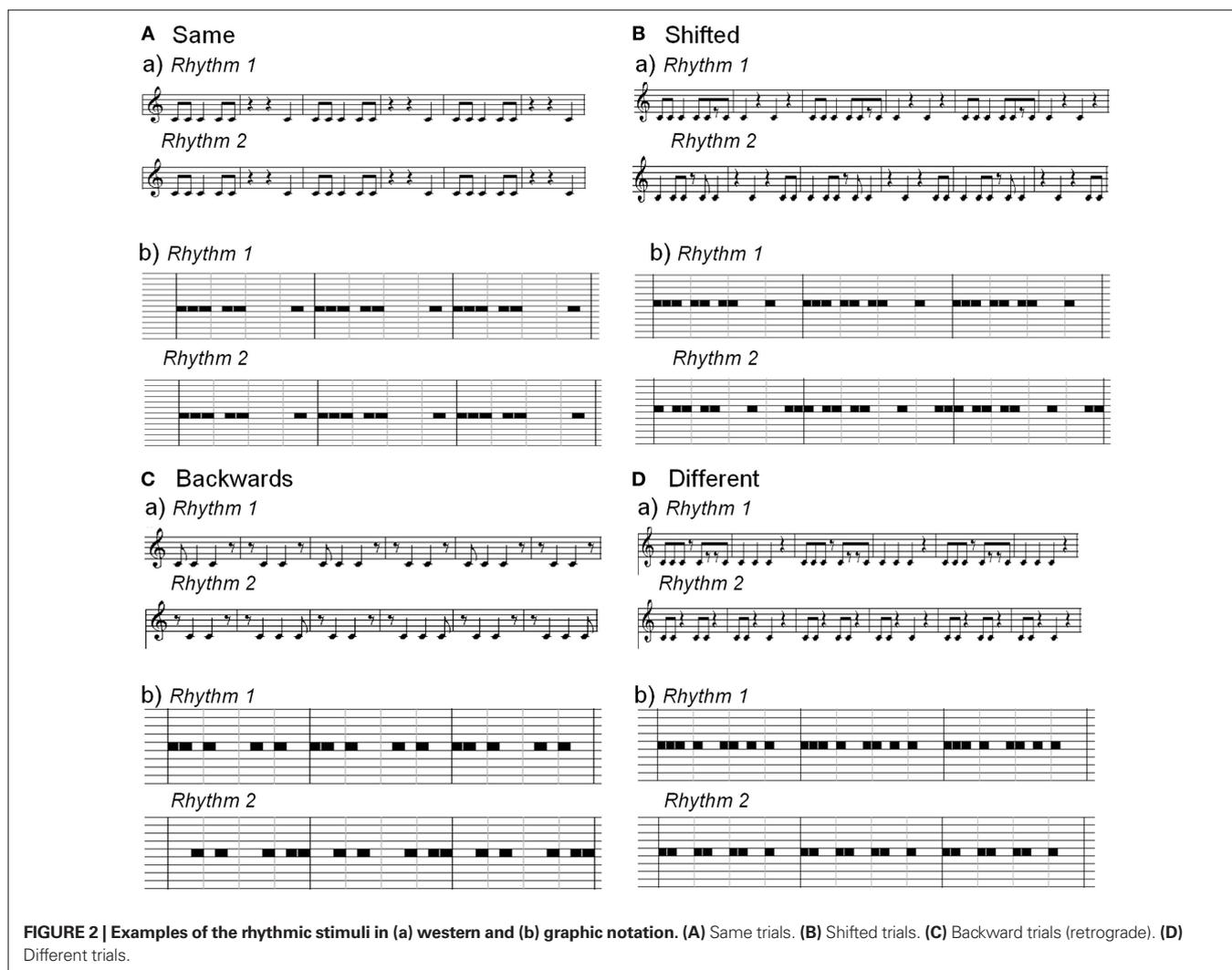
Table 1 | Criteria used to select the rhythmic stimuli.

CRITERION RULES	
1.	Rhythms must be 12- or 16-beats long
2.	Rhythms must have varying inter-onset intervals
3.	Rhythms must always have an accent on the first beat
4.	Rhythms must have no more than three accents in succession
5.	Rhythms must have no more than four rests in succession
6.	Rhythms must have approximately 50% density of accents, so: 12-beat patterns have 5, 6, or 7 accents 16-beat patterns have 7, 8, or 9 accents
7.	Rhythms must be excluded if they add an extra accent to an otherwise well defined pattern of groupings. That is, unless this accent generates a new sub-pattern of groupings
MANIPULATION TASK RULES	
1.	Rhythms must not be the same when shifted
2.	Rhythms must not be the same when backward
3.	Rhythms must not generate the same pattern when shifted and backward

For the criterion rules, rule 7 could not be specified in mathematical code, so the researchers removed these rhythms according to perceptual judgments. For the manipulation task rules, rule 3 was used to avoid a correct answer that could be both backward and shifted.

the last beat, and different involved selecting at random another rhythm from the pool of stimuli. Figure 2 gives an example of each manipulation represented in western and graphic notation. An equal number of 12- and 16-beat rhythms was used in each notation condition. In total there were 48 trials, with the same stimulus pairs randomly presented in the western and graphic conditions. Equal numbers of each response type were used for the AO and AV tasks.

Pilot studies indicated optimal performance of the manipulation task occurred when rhythms were presented three times in succession. It is likely that the first two repetitions established the length of the rhythmic cycle, and the final repetition allowed consolidation in ASTM. Thus, for each stimulus pair a minimum IOI of 250 ms was used as this is easily perceived, and allowed three repetitions of each rhythmic phrase to be played within the 10–15 s upper limit of ASTM (McLachlan and Wilson, 2010). The two rhythms that comprised each pair were separated by an interval of 1.8 s followed by 5 s response time. A chime was presented before each new trial.



The visual representations were made as simple as each notation system would allow. For western notation, 12- and 16-beat rhythms were spread over two bars so that only crotchet (1/4) and quaver (1/8) notes and rests were used. Quavers were made into crotchets wherever possible. The time signature of 3 crotchets per bar was used for 12-beat rhythms, and 4 crotchets per bar for 16-beat rhythms. For graphic notation, 12-beat rhythms were divided into 12 units per bar, and then further subdivided into four sections of three units each. Similarly, 16-beat rhythms were divided into 16 units per bar, and subdivided into four sections of four units each. Western music notation was created using Finale software and graphic notation was created using a standard piano roll MIDI notation (Figure 2).

The MRT is a 20-item multiple-choice test that involves the mental rotation of drawn block constructions. Each item consists of a criterion construction and four alternatives; two identical to the criterion but rotated in orientation, and two distractors. The test is divided into two sections (10 items each), each of which was time limited to 3 min in accordance with standardized administration. The MRT has high internal consistency (split half correlation, $r = 0.88$) and substantial retest reliability ($r = 0.83$) after at least a 1 year interval (Vandenberg and Kuse, 1972).

The previous music training of the participants was assessed using the Survey of Musical Experience (Wilson et al., 1999). This canvasses the study of music as part of general schooling, music preferences and amount of music listening, any formal training in instrumental or vocal performance, music theory or composition, previous and current engagement in music activities, and the ability to read or write music. Demographic details were also collected for each participant, as well as any strategies reported to perform the tasks.

Procedure

Ethics approval was obtained from the Human Research Ethics Committee of The University of Melbourne, and all participants gave written informed consent. The two experimental sessions outlined in Table 2 were run 1 week apart in a quiet room in groups of no more than 15 participants. The notation was presented with a computer and projected onto a large screen (1.5 m wide \times 1 m high). Participants were seated 1.5–2.5 m from the screen. MIDI music software was used to present the music with western, graphic, or no notation. A speaker in free field was placed near the screen facing the participants and the volume was adjusted to allow comfortable

listening conditions. In the first session, a brief introduction to the notation was given with a similar level of detail provided for each group. Each task included practice items with feedback to ensure participants fully understood the tasks. Participants recorded their responses via pen and paper using a multiple-choice format. Scores were calculated as percentage items correct for each task.

Sixteen AO trials were run in session one (eight discrimination and eight manipulation) and 16 AV trials were run in each of sessions one and two (eight discrimination and eight manipulation; **Table 2**). Only two sessions were undertaken in this experiment as participants in the graphic notation condition reached ceiling performance.

Design

Table 3 shows the design of this experiment. The baseline tasks of AO discrimination and manipulation were presented to participants in the two notation conditions in session one. This was followed by the experimental AV discrimination and manipulation trials. This allowed hypotheses 1 (discrimination > manipulation) and 2 (notation improves performance) to be analyzed using a multivariate repeated measures analysis of variance with factors of task and modality (AO or AV) evaluated for data from session one pooled over the two conditions. We planned to analyze hypotheses 3 (graphic > western notation) and 4 (learning) separately for

the discrimination and manipulation tasks using omnibus, mixed between and within-subjects analyses of variance with session number as the within-subjects factor and notation condition as the between-subjects factor. The MRT was administered in session one to ensure there was no difference in visuo-spatial mental rotation ability of the group assigned to the graphic or western notation condition.

RESULTS

The criterion of statistical significance was set at an alpha level of 0.05 for all tests. Independent samples *t*-tests showed no significant differences between the groups for gender, age, handedness, years of education, years of music experience, AO rhythmic discrimination, AO rhythmic manipulation, or mental rotation ability (**Table 4**).

Hypothesis 1, that discrimination performance ($M = 84.32$, $SD = 14.87$) would be better than manipulation performance ($M = 59.59$, $SD = 25.16$), was supported by a main effect for task [$F(1,27) = 83.38, p < 0.001$]. Hypothesis 2, that the performance of discrimination and manipulation tasks with notation ($M = 84.37$, $SD = 18.07$) would be better than without notation ($M = 59.07$, $SD = 22.72$), was also supported by a main effect of notation [$F(1,27) = 43.24, p < 0.001$]. There was an interaction between task and the presence of notation [$F(1,27) = 16.32, p < 0.001$] with the facilitatory effect of notation greater for manipulation than discrimination performance.

Table 5 compares AV discrimination scores with manipulation scores for the western and graphic notation groups over sessions one and two. Ceiling effects were observed for all AV tasks excepting the manipulation performance of the western notation group. These effects precluded the use of omnibus analyses of variance to assess hypotheses 3 and 4. As there were no useful transformations to correct the data distributions, we used Mann–Whitney *U*-tests to assess the effect of notation type on AV discrimination and manipulation performance (hypothesis 3). For hypothesis 4 learning effects could only be tested for the western notation group, using a paired-samples *t*-test.

Table 2 | Structure of the two sessions of Experiment 1.

Session	Tasks	Presentation		
		Trials/block	Blocks	Time/block (min:s)
1	Mental rotation test	20	1	6
	AO discrimination and manipulation	16	1	7:25
	AV discrimination and manipulation	16	1	7:25
2	AV discrimination and manipulation	16	1	7:25
	Music experience questionnaire			10

AO, auditory only; AV, audio-visual.

Table 3 | Stimuli and conditions for the baseline and experimental rhythmic tasks.

Task	Modality	Notation condition	Session	Stimulus category (n)	
BASELINE					
Discrimination	AO	None	1	Same (4)	Diff (4)
Manipulation	AO	None	1	Retro (4)	Shift (4)
Mental rotation test	VO	None	1	Same (40)	Diff (40)
EXPERIMENTAL					
Discrimination	AV	West	Graph	1–2	Same (8) Diff (8)
Manipulation	AV	West	Graph	1–2	Retro (8) Shift (8)

AO, auditory only; AV, audio-visual; West, western notation; Graph, graphic notation; Diff, different; Retro, retrograde (“backward”); Shift, shifted.

Table 4 | Demographic and baseline characteristics of the groups in Experiment 1 (±SD).

Characteristic	Graphic notation (n = 15)	Western notation (n = 13)
Number of males	2	5
Mean age (years)	20.50 (2.82)	21.53 (6.11)
Percentage right handed	86	92
Mean education (years)	14.07 (1.64)	13.60 (1.60)
Mean music experience (years)	1.71 (1.86)	1.40 (1.91)
Mean AO discrimination (% correct)	70.09 (18.11)	76.33 (13.48)
Mean AO manipulation (% correct)	41.90 (23.97)	42.56 (16.45)
Mean mental rotation ability (% correct)	37.86 (15.41)	46.00 (17.65)

AO, auditory only; AV, audio-visual. $p > 0.05$ for all group comparisons. Chance performance for AO discrimination and AO manipulation both = 25%.

Table 5 | Experiment 1 audio-visual (AV) discrimination and manipulation performance (\pm SD).

Task	Graphic notation (% correct)		Western notation (% correct)	
	Session 1	Session 2	Session 1	Session 2
Discrimination	88.75 (12.44)	92.92 (12.14)	87.98 (14.35)	90.38 (15.13)
Manipulation	92.92 (8.47)***	92.50 (10.17)***	57.21 (17.74)	69.23 (21.28)*

*** $p < 0.001$ * $p < 0.05$. Chance performance for discrimination and manipulation both = 25%.

For hypothesis 3b the results showed that graphic notation manipulation scores were significantly higher than western notation scores, both in session one ($U = 17.5, n_1 = 25, n_2 = 30, p < 0.001$, two-tailed) and session two ($U = 122.5, n_1 = 25, n_2 = 30, p < 0.001$, two-tailed). In contrast, for hypothesis 3a the discrimination scores did not differ significantly between the notation groups. For hypothesis 4, a significant increase in western notation manipulation scores was evident from session 1 to session 2 [$t(24) = -2.12, p < 0.05$, two-tailed].

DISCUSSION

As expected, rhythmic manipulation was shown to be more difficult than discrimination (hypothesis 1). The results also provide strong support for the facilitatory effect of music notation on the auditory discrimination and manipulation performance of novice musicians (hypothesis 2). This effect was most pronounced for auditory manipulation as evident from the interaction between task and notation. It suggests that visual information plays an increasingly important role as task difficulty increases. More generally, this finding is consistent with the facilitatory effects of visual cues on auditory processing, such as the benefits of lip reading for recognizing phonemes in noisy conditions (McGurk and MacDonald, 1976; O'Leary and Rhodes, 1984).

The spatially congruent graphic notation facilitated manipulation performance more than the composite symbolic-spatial western notation (hypothesis 3b), whereas no difference was observed for discrimination performance between the notation groups (hypothesis 3a). This likely reflects the ease with which all participants performed the discrimination task, lending support to the notion that the difference in manipulation performance was due to decay of the memory trace while performing the spatial manipulation. Moreover, the results provide evidence of learning over the two sessions for the western notation group (hypothesis 4). Learning effects were not observed for graphic notation because performance reached ceiling in session one and plateaued in session two. This suggests that graphic notation provided highly effective visual scaffolding for novice musicians, minimizing demands on ASTM. This, in turn, likely enabled rapid learning of the ability to spatially manipulate AV information, and points to the potential benefits of graphic notation to assist music learning. In contrast, slower learning with western notation reflected the need to establish complex rules between the symbols and their associated auditory stimulus durations.

EXPERIMENT 2: IDENTIFICATION OF MELODIC TRANSFORMATIONS

This experiment tested whether spatially mapped notations of melodic contours facilitated the identification of melodic transformations compared to western notation. Pairs of novel melodies were presented and participants were asked to identify whether a transformation had been applied to the second melody in the stimulus pair. Melodies can be transformed by reflecting their temporal order and pitch height or by shifting their starting pitch. Initial testing showed ceiling performance for manipulation trials using these transformations in some conditions and so another set of stimuli were created in which two transformations were applied to the referent melody. The participant was informed of the first transformation and was required to select one of the two possible second transformations. Discrimination trials were run in a separate block in which the participant was required to choose whether the melodies were the same or different. To compare the efficacy of the two notation types, novice musicians were pseudo-randomly assigned to a western or graphic notation condition, keeping age and gender approximately equal.

METHOD

Participants

A new group of 26 participants took part in this experiment (see Results for demographic details). All were classified as novice musicians, had limited music experience, and described themselves as "not proficient" in reading western music notation. None were currently practicing music. The majority of participants ($n = 22$) were community volunteers and four were first year psychology students from The University of Melbourne who received course credit for participation. As for Experiment 1, participants were pseudo-randomly assigned to the western or graphic notation conditions, keeping age and gender approximately equal.

Stimuli

The stimuli comprised pairs of novel isochronous pitch height contours (melodic phrases) developed by Wilson et al. (1997). They consisted of 40 novel melodic pairs, with pitches ranging from A#3 to D#5. There were identical numbers of tonal and non-tonal, and five- and eight-note melody pairs that were counterbalanced across the conditions for each task. The stimuli were created using a grand piano instrumental timbre in MIDI computer software and were played at a tempo of 120 crotchets/min. The added dimension of pitch height to temporal order enabled more transformations than was possible in Experiment 1. The second melody of each pair underwent one of the following transformations: (1) Inversion, by inverting pitch heights around a central pitch, (2) Retrograde, by presenting the melody backward, or (3) Transposition, by shifting the melody up or down by three or four semitones in pitch.

Figure 3 illustrates the two notation conditions for the melodic discrimination and manipulation trials. Pilot testing indicated that participants in both notation groups performed at ceiling levels in the AV manipulation trials, and so a set of double transformation trials was included. In these trials the first transformation that had been applied to the reference stimulus was given, and participants were asked to identify the second transformation. The AO manipulation task involved single manipulations only since pilot testing indicated that double manipulation was too difficult for this task.

Different stimulus pairs were used in the AO discrimination and manipulation tasks and in the AV discrimination task. Ten of the melodic pairs were used in each of the AO tasks, and the remaining 20 were used for the AV discrimination task. All 40 stimulus pairs were used in the AV manipulation task and thus stimuli were presented so that 20 of the stimulus pairs were used in the first session and the remaining 20 in the second session. The third and fourth sessions used the stimuli of the first two sessions respectively but in a different order to minimize memory effects. The double transformations were generated from the same 40 melodies by applying two transformations to the second melody. **Table 6** outlines the response categories for each of the tasks.

Procedure

The experimental procedure replicated Experiment 1 except for the ordering of tasks (see **Table 7**). To prevent fatigue in session one, the MRT was run in session four. In the discrimination and manipulation tasks, one trial consisted of the presentation of a

pair of melodies separated by an interval of 1.8 s, followed by 5 s response time. A chime tone was presented before each new trial. Five- and eight-note melody pairs took 8 and 11 s to complete respectively. The total duration of each block of trials and the order of the experimental tasks over the four weekly sessions are shown in **Table 7**. In the first session, a brief introduction to the notation was given with a similar level of detail provided for each group.

Participant responses were recorded via pen and paper using a multiple-choice format. In discrimination trials participants were required to select “same” or “different.” In the single manipulation trials, multiple-choice responses reflected the different transformation types; “upside down” (inversion), “backward” (retrograde), or “shifted” (transposition). For the double manipulation trials, the first transformation was provided and participants were required to correctly identify the second transformation from the two remaining alternatives. Scores were calculated as the percentage of items correct for each task.

Design

Table 6 shows the design of this experiment. The discrimination trials were run separately from the manipulation trials due to the increased number of response categories in the manipulation task. As in Experiment 1, we analyzed hypotheses 1 (discrimination > manipulation) and 2 (notation improves performance) using a multivariate repeated measures analysis of variance with factors of task and modality (AO or AV) evaluated for data from session one pooled over the two conditions. We then planned to analyze hypotheses 3 (graphic > western notation) and 4 (learning) using an omnibus, mixed between and within-subjects analysis of variance with session number as the within-subjects factor, and notation condition and task as the between-subjects factors. This experiment spanned four sessions because ceiling performance for the manipulation task was not reached as quickly as in Experiment 1.

RESULTS

The criterion of statistical significance was set at an alpha level of 0.05 for all tests. Independent samples *t*-tests revealed no significant differences between the two groups for age, years of education, handedness, and years of previous music experience (**Table 8**). Hypothesis 1, that discrimination performance ($M = 86.82, SD = 14.65$) would be better than manipulation performance ($M = 70.48, SD = 23.81$),

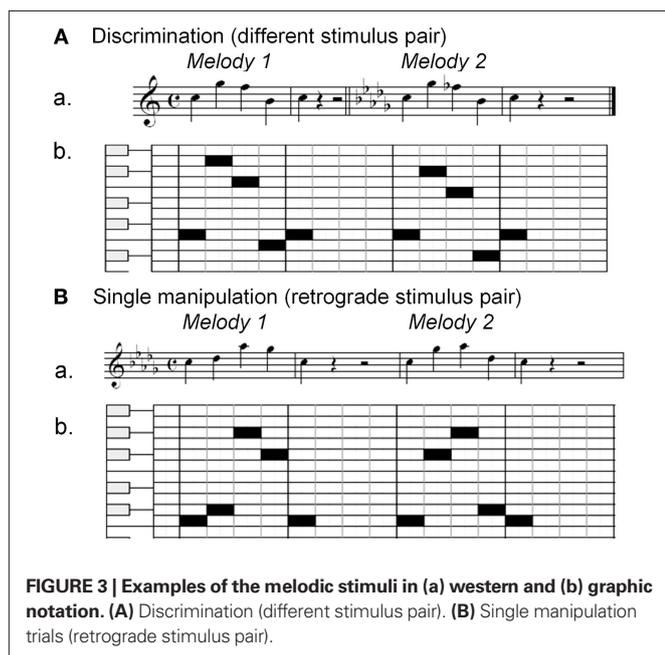


FIGURE 3 | Examples of the melodic stimuli in (a) western and (b) graphic notation. (A) Discrimination (different stimulus pair). (B) Single manipulation trials (retrograde stimulus pair).

Table 6 | Stimuli and conditions for the baseline and experimental melodic tasks.

Task	Modality	Notation condition		Session	Stimulus category (n)		
BASELINE							
Discrimination	AO	None		1	Same (4)		Diff (6)
Manipulation	AO	None		1	Inv (3)	Ret (3)	Tran (4)
Mental rotation task	VO	None		4	Same (40)		Diff (40)
EXPERIMENTAL							
Discrimination	AV	West	Graph	1	Same (8)		Diff (12)
Single manipulation	AV	West	Graph	1–4	Inv (13)	Retro (13)	Tran (14)
Double manipulation	AV	West	Graph	1–4	Inv + retro (14)	Tran + inv (13)	Tran + retro (13)

AO, auditory only; AV, audio-visual; West, western notation; Graph, graphic notation; Diff, different; Inv, inversion; Retro, retrograde (“backward”); Tran, transposition.

was supported by a main effect for task [$F(1,26) = 17.00, p < 0.001$]. Hypothesis 2, that the performance of discrimination and manipulation tasks with notation ($M = 88.27, SD = 16.77$) would be better than without notation ($M = 69.04, SD = 21.17$), was supported by a main effect for notation [$F(1,26) = 51.18, p < 0.001$]. No interaction effect was observed.

Despite pseudo-random assignment of participants to groups, the mean MRT score for the graphic notation group was significantly higher than the western notation group [Table 8, $t(24) = -2.15, p < 0.05$, two-tailed]. This was likely due to the larger number of male participants in the graphic notation group (Hooven et al., 2004). Thus, Pearson product-moment correlations were used to investigate the influence of MRT scores on melodic

manipulation performance. A significant positive correlation was found between mental rotation ability and melodic manipulation in the presence of notation ($r = 0.56, p < 0.01$). Using partial correlation, the strength of this relationship remained similar when controlling for sex ($r = 0.55, p < 0.05$). Given these findings, MRT scores were included as a covariate in the analyses examining melodic manipulation performance. There was no significant correlation between MRT and AO melodic manipulation scores ($r = 0.181, p > 0.05$).

Hypothesis 3a, that graphic notation would facilitate discrimination performance more than western notation, could not be tested as scores for discrimination in the presence of graphic notation were close to ceiling (see Table 9). Instead, an omnibus analysis of covariance was used to assess hypothesis 3b in conjunction with hypothesis 4.

Hypothesis 3b, that graphic notation would facilitate manipulation performance ($M = 90.54, SD = 9.55$) more than western notation ($M = 83.62, SD = 7.74$), was supported by a main effect for group [$F(1,23) = 4.60, p < 0.05$] while covarying for MRT ability. The effect of the covariate was also significant [$F(1,23) = 6.16, p < 0.05$, see Table 8]. Hypothesis 4, that manipulation performance would improve over the four sessions for both notation types, was supported by a main effect of time [$F(3,14) = 5.16, p < 0.05$]. There was no significant interaction between time and group [$F(3,14) = 1.05, p > 0.05$] indicating that there was no difference in the average rate of learning across the four sessions for the two notation conditions (see Figure 4). In other words, the mean manipulation scores across notation groups significantly increased with each session. Closer inspection of Figure 4 shows that the pattern of performance of the two notation groups differed for melodic manipulation over the four sessions, with variable rates of learning evident between different sessions for the two groups. To assess this, planned repeated contrasts were performed and revealed significant group differences between mean manipulation scores (graphic > western notation) when comparing sessions 1 and 2 [$F(1,16) = 4.80, p < 0.05$], sessions 2 and 3 [$F(2,16) = 113.95, p < 0.001$], and sessions 3 and 4 [$F(2,16) = 4.80, p < 0.05$].

Table 7 | Structure of the four melodic experimental sessions.

Session	Tasks	Presentation		
		Trials/ block	Blocks	Time/ block (min:s)
1	AV manipulation	20	1S and 1D	5:00
	AO discrimination	10	1	2:30
	AO manipulation	10	1	2:30
	AV discrimination	20	1	5:00
2 and 3	AV manipulation	20	1S and 1D	5:00
4	AV manipulation	20	1S and 1D	5:00
	Mental rotation	20	1	6:00
	test			
	Music experience questionnaire			10:00

AO, auditory only; AV, audio-visual; S, single manipulation trials; D, double manipulation trials.

Table 8 | Demographic and baseline characteristics of the groups in Experiment 2 (±SD).

Characteristic	Graphic notation (n = 13)	Western notation (n = 13)
Number of males	9	5
Mean age (years)	25.31 (8.24)	23.23 (2.09)
Percentage right handed	92	85
Mean education (years)	16.23 (1.23)	17.23 (1.36)
Mean music experience (years)	2.35 (2.08)	1.77 (1.59)
Mean AO discrimination (% correct)	76.15 (15.02)	79.23 (15.25)
Mean AO manipulation (% correct)	56.92 (22.50)	63.85 (23.99)
Mean mental rotation ability (% correct)	59.46 (18.45)	47.31 (8.75)*

AO, auditory only; AV, audio-visual. $p > 0.05$ for all group comparisons except the mental rotations test. * $p < 0.05$. Chance performance for AO discrimination = 50%, and AO manipulation = 33%.

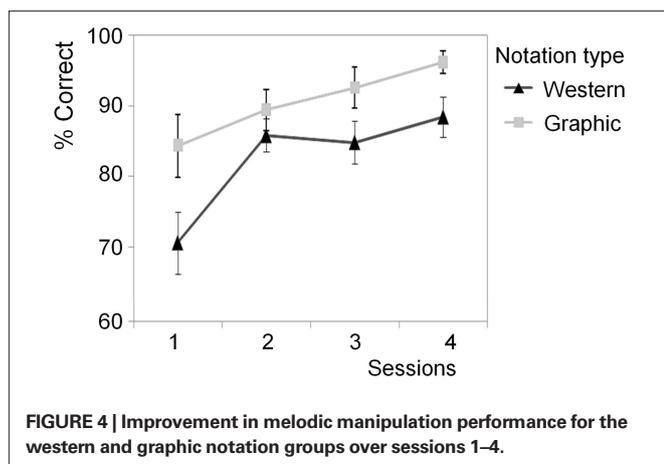
DISCUSSION

These results confirm that discrimination of melodic contours was better than their manipulation (hypothesis 1). They also support the facilitatory effect of music notation on melodic discrimination and manipulation in novice musicians (hypothesis 2). The spatially congruent graphic notation facilitated manipulation performance more than the composite symbolic-spatial

Table 9 | Melodic discrimination and manipulation performance with notation for session 1 (±SD).

Task	Graphic notation (% correct)	Western notation (% correct)
Discrimination	98.08 (3.84)	93.85 (7.40)
Manipulation	84.31 (16.04)	70.54 (15.83)

Chance performance for discrimination = 50%, and manipulation = 42%.



western notation (hypothesis 3b). However ceiling effects prevented the evaluation of any differences in discrimination between the notation conditions (hypothesis 3a). This is a similar pattern of results to those described in Experiment 1 and again reflects the greater ease with which participants performed the discrimination task.

The results show evidence of learning across the four sessions of the experiment irrespective of notation type (hypothesis 4). Despite no difference in the average rate of learning for the two notation conditions, the learning curves suggested that western notation provided a less effective scaffold in session 1. Learning in this condition then showed more rapid acceleration in session 2, possibly reflecting that participants had gained a sense of confidence and familiarity with the notation. Subsequent to this, the performance of the western notation group tended to plateau across sessions 2–4. In contrast, the learning curve of the graphic notation group showed a steady rate of increase across the four sessions with performance starting at a higher level in session 1. This learning curve supports the generally superior facilitation effect of graphic notation shown by the main effect. These findings were not attributable to differences in melodic discrimination or manipulation ability in the absence of notation and remained after visual mental rotation ability had been accounted for.

GENERAL DISCUSSION

Taken together the results of the two experiments show facilitation of the processing of patterns of auditory information by novice musicians when presented with concurrent, spatially coherent visual information. Facilitation was greater when the visual information was spatially linear, and contained less symbolically encoded information. Since visual fields are well known to be spatially encoded (Rolls and Deco, 2002), this supports the proposition that pitch and time are spatially encoded in ASTM. The greatest effect of visual information was in the rhythmic manipulation task with graphic notation where most participants were able to successfully apply visual processing skills to the AV information in the first session. Western notation of rhythm is largely symbolic. Since symbolic information was unlikely to be learnt without guidance this likely created limits to task performance in the western notation condition. For the melodic manipulation task with graphic notation performance improved steadily and approached ceiling

after four sessions. Compared to western notation it began from a higher baseline and continued to improve in each session. Western notation contains both spatial and symbolic pitch height information, leading to more limited learning in this condition whereas improved performance in the graphic condition likely reflects participants learning to associate spatially coherent melodic and visual contours. The dimensions of pitch and time were studied independently to ensure that the effect of graphic notation was facilitatory in both dimensions. In real-world music, these dimensions are combined so the value of graphic notation for novices could be expected to be greater than shown in these experiments.

It was not possible to compare manipulation performance to a visual only condition to ensure that participants did not simply use the visual notation to perform the task. However this is unlikely because participants were exposed to both the visual and auditory information simultaneously, and must have processed the auditory information at least at a subconscious level. Furthermore gradual learning effects were observed in most conditions of the manipulation tasks. If performance was based purely on the visual representation then previously learnt skills in visual manipulation would have led to ceiling performance in graphic notation conditions, or floor performance in western notation conditions as the participants could not decode the symbolic information. These patterns were not observed in the data, supporting our interpretation that the participants undertook a multimodal (AV) processing task.

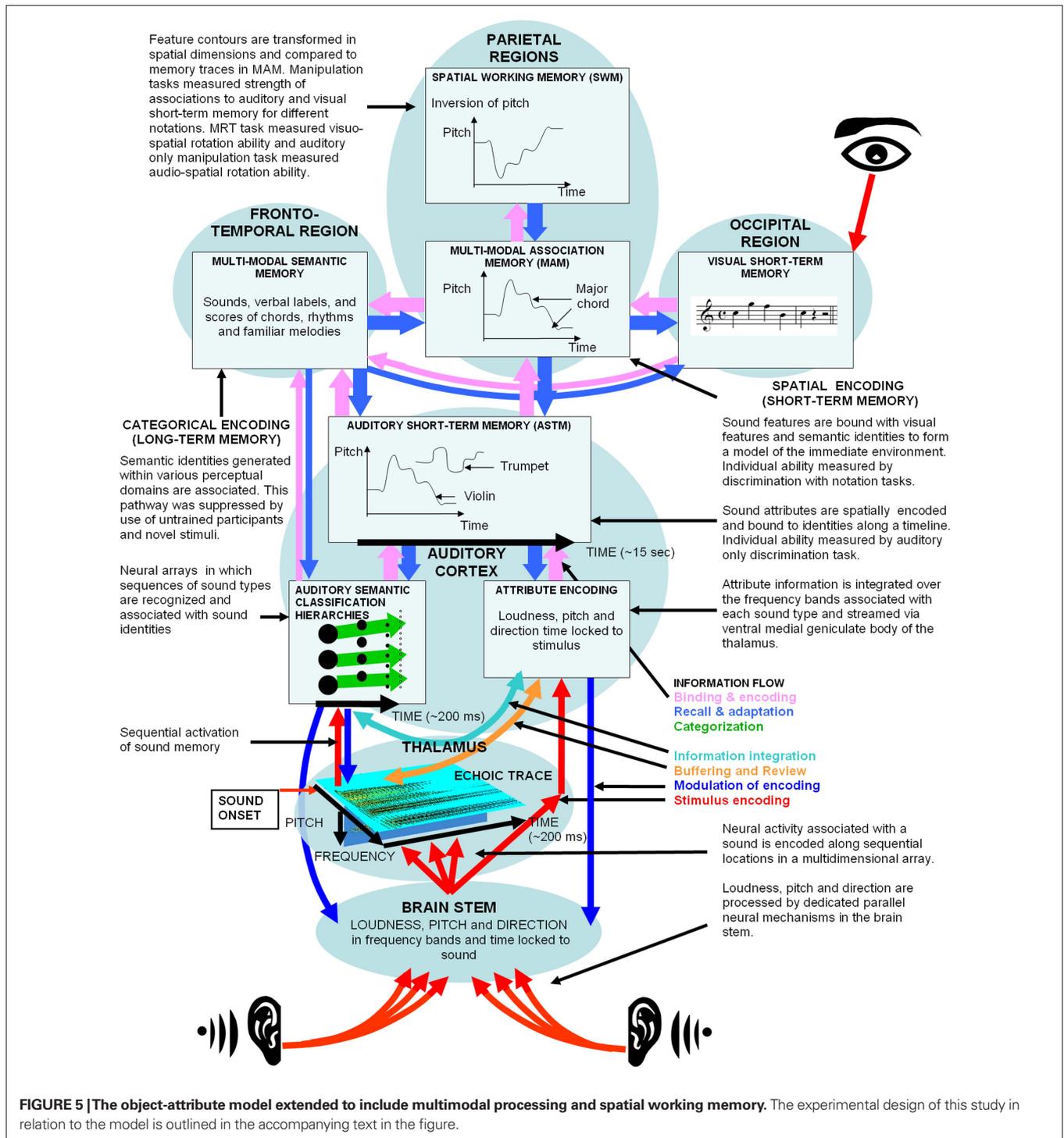
EXTENSION OF THE OBJECT-ATTRIBUTE MODEL TO INCLUDE MULTIMODAL PROCESSING

The object-attribute model of auditory processing (McLachlan and Wilson, 2010) informed the design of the experiments undertaken in this paper. To interpret the results we first need to introduce an expanded version of this model. Figure 5 shows a simplified representation of the object-attribute model (McLachlan and Wilson, 2010) with the addition of higher cortical processes involved in multimodal information processing, and semantic and SWM. Figure 5 also outlines how the experimental design of this study should activate specific processes of the model. The following section provides an explanation of the model in light of previous neurophysiological, behavioral, and imaging findings. The final section then interprets the results of this study in relation to the model.

The echoic trace and auditory short-term memory

Models of auditory stimulus representation have distinguished two phases of auditory information processing (Cowan, 1988; Näätänen and Winkler, 1999; McLachlan and Wilson, 2010). In the first pre-representational phase, the stimulus driven afferent activation pattern is processed to generate auditory features such as pitch and loudness. This phase of processing occurs approximately over the first 200 ms (Näätänen and Winkler, 1999) and is represented by the echoic trace at the level of the thalamus in Figure 5.

In the object-attribute model patterns of neural activation in the echoic trace are correlated through time with information stored in patterns of synaptic connectivity in sound classification hierarchies. While this circuit leads to the recognition of specific sound types in the cortex, it also regulates the potentiation of neurons in the echoic trace according to a fixed temporal sequence of spectral patterns that are associated with the particular activated sound type. Pitch



information in the echoic trace is shown to be distributed across a periodotopic dimension in each critical band (Z-axis of the echoic trace in **Figure 5**) and is integrated across critical bands according to templates stored in long-term memory for specific sound types.

The second representational phase of auditory processing is shown to occur in ASTM in **Figure 5**. It can encode and sustain information for about 15 s (Deutsch, 1999; Näätänen and Winkler, 1999). Here, auditory features are integrated into a unitary percept

and projected onto a temporal dimension (Näätänen and Winkler, 1999). The object-attribute model postulates that streaming of attribute information to ASTM through the auditory core can only occur after a gestalt based on sound identity has been formed. Pitch height and temporal onset are spatially encoded along a neural multidimensional array and associated with an identity in ASTM, as shown in the central panel of **Figure 5** (Deutsch, 1999; McLachlan and Wilson, 2010).

This model is supported by data from many experimental studies. Participants tend to assign high tones to high positions and low tones to low positions in vertical space (Pratt, 1930; Trimble, 1934; Mudd, 1963; Roffler and Butler, 1968), and performance on melodic inversion and retrograde judgments was predicted by performance on spatial tasks (Cupchik et al., 2001). Spatial response coding experiments demonstrate that in a proportion of individuals performance is both faster and more accurate when judgments of pitch height match response patterns by using an “upper” or “lower” response key or the terms “higher” or “lower” (Melara and Marks, 1990; Rusconi et al., 2005, 2006; Beecham et al., 2009). This has been referred to as the spatial music association of response codes (SMARC) effect. Studies of amusia (disorders of music perception or production) also provide neuropsychological support for associations between disorders of cortical pitch and spatial processing (Wilson et al., 2002; Douglas and Bilkey, 2007).

A range of experimental paradigms has been used to investigate the properties of ASTM for time varying pitched stimuli. These include distortion of temporal judgments by systematic deviations of pitch for stimuli that vary in pitch at a constant rate (Henry and McAuley, 2009), multidimensional similarity ratings and unspeeded classifications (Grau and Kemler-Nelson, 1988), and mismatch negativity in evoked potential EEG (Gomes et al., 1995). These studies all support the proposition that pitch and temporal duration are interacting features encoded in a multidimensional spatial array, presumably located in ASTM. Our findings from Experiment 2 show that more spatially congruent music notation facilitated the cognitive manipulation of pitch-time patterns. These findings are consistent with spatial encoding of pitch information in ASTM, and its integration with visuo-spatial information in MAM prior to cognitive manipulation in SWM.

The presence of a categorical pitch encoding pathway linked to ASTM (represented on the left of **Figure 5**) is supported by research showing that music is processed in relation to learnt schema (Deutsch, 1999). Deutsch (1999) proposed a model of hierarchical networks that operate on a spatiotemporal array of pitch information to represent pitch sequences in the form of grouped operations, some of which form melodic archetypes that are stored in long-term memory. This would enable the streaming and segmentation of music information based on rules acquired through previous music exposure (McLachlan and Wilson, 2010). Recognition of pitch intervals and melodies is a commonly occurring form of categorical perception in humans (Shepard and Jordan, 1984). Musically trained listeners are also able to attach verbal and symbolic (textual) labels to music identities such as chords and rhythms. A small proportion of musicians possess absolute pitch, which is the ability to apply verbal labels to pitch (Levitin and Rogers, 2005). This means that categories of pitch, pitch intervals, and patterns of duration intervals can be stored in long-term memory and may be associated with predominantly symbolic music notation.

The time between the onsets of successive sounds (the IOI) determines how a rhythm is perceived (Krumhansl, 2000). Dessain (1992) and McLachlan (2000) have proposed expectation–realization models for rhythmic perception that could generate temporal expectancies in ASTM. Strong neural responses have been reported in the auditory association cortex when expected tones were omitted from an auditory stream, at latencies of around 300 ms from the expected sound

onset (Hughes et al., 2001; Ioannides et al., 2003). These responses are thought to involve increased processing in ASTM associated with the generation of new expectancies, and support the proposition that temporal events are encoded along a spatial time-line in ASTM as shown in the central panel of **Figure 5** (McLachlan and Wilson, 2010). Support also comes from the recent findings of Frassinetti et al. (2009) who proposed that temporal intervals are represented as horizontally arranged in space, and that spatial manipulation of temporal processing likely occurs via cuing of spatial attention.

Griffiths and Warren (2002) suggested that the planum temporale is a “computational hub” involved in most aspects of auditory processing. This means that in the object-attribute model the planum temporale may be the location of the first processing stage of ASTM, where associations between sound identity and features occur. Consistent with this, Hall and Plack (2009) found fMRI activation in the planum temporale for a wide variety of pitch eliciting stimuli. They also showed activation of the prefrontal cortex consistent with categorical encoding in anterior temporal regions (the “what” pathway on the left of **Figure 5**), as well as activation of the temporo-parieto-occipital junction that may be associated with the integration of visual and auditory information in MAM. Hall and Plack (2009) observed considerable individual differences in patterns of fMRI activation consistent with interactions of long-term memory templates in the categorical pathway with pitch processing in the auditory core and planum temporale. In the object-attribute model, these interactions are shown extending from the visual input at the top right of **Figure 5**.

Lesion studies report severe deficits for judging the order of appearance of verbal stimuli in patients with left frontal lobe lesions, and visual stimuli in patients with right frontal lobe lesions (Milner, 1972). A similar pattern of deficits was observed for participant-ordered tasks involving verbal and visual stimuli in patients with frontal lobe and posterior temporal lobe lesions (Petrides and Milner, 1982). These findings suggest that the anterior categorical pathway shown on the left of **Figure 5** may facilitate the manipulation of information contained in SWM. Using diffusion fiber tractography, Frey et al. (2008) have shown distinct connections between Brodmann areas 44 and 45 in the ventrolateral frontal cortex, and the inferior parietal lobule and anterior superior temporal gyrus respectively. Activation in neighboring Brodmann area 47 in the mid-ventrolateral prefrontal cortex has also been associated with recently learnt memories of acoustical features (Bermudez and Zatorre, 2005; Schönwiesner et al., 2007). Bor et al. (2003) showed increased activation bilaterally in the lateral frontal cortices (Brodmann areas 44 and 45) and in the fusiform gyrus when a visual working memory task was facilitated by recognizable visual structures. This pattern of activation was interpreted as representing the use of object-based information stored in the fusiform gyrus to facilitate chunking of visual information. Taken together, these data suggest that multimodal semantic memory processes are mediated by these regions, facilitating the manipulation of spatially encoded music information as shown in **Figure 5**.

Multimodal associative memory and spatial working memory

There is an expanding literature on the facilitatory effect of cross-modal processing of vision and audition. These senses have a mutual influence in perception; if one sense receives ambiguous

information, the other can aid in decision making (McGurk and MacDonald, 1976; O'Leary and Rhodes, 1984). In other words, predictions based on one sensory modality can influence expectations in the short-term memory of another modality, which in turn can influence sub-cortical response fields (Shamsa et al., 2002; Sherman, 2007; McLachlan and Wilson, 2010). In the case of music cognition, information from visual, auditory, and kinesthetic perceptual processes is available to MAM. Hasegawa et al. (2004) reported fMRI activation that correlated with the participant's level of piano training when they observed video footage of hands playing piano keys. The activation occurred in the parietal lobules, premotor cortex, supplementary motor area, occipito-parietal junction, thalamus, inferior frontal gyrus, and planum temporale. Molholm et al. (2006) reported changes in human intracranial electrophysiological recordings associated with the integration of audio and visual signals in the superior parietal lobule. In keeping with Hall and Plack (2009) these data place the likely location of MAM in the parietal region near the temporo-parieto-occipital junction. This is the ideal location to integrate information from auditory and visual cortical regions (Figure 5), and is consistent with long standing observations of primary, secondary, and tertiary cortical zones, with multimodal integration occurring later in brain development as myelination increases in tertiary zones (Luria, 1980). The temporal processing of multimodal information is complicated by differing latencies of the arrival of auditory and visual information from simultaneous events, and different neural processing speeds in the two modalities (Molholm et al., 2006; Virsu et al., 2008). However Virsu et al. (2008) showed that training can improve the ability to judge simultaneity, and that such learning effects were preserved for over 5 months without practice.

Spatial working memory shown at the top of Figure 5 receives spatially encoded information from multiple sensory modalities that is integrated in MAM, as well as categorically encoded information from semantic networks. Foster and Zattore (2009) reported fMRI activation in the right intraparietal sulcus that correlated with performance on melodic transposition tasks. Activation was also observed in the auditory and premotor areas, inferior parietal cortex, ventrolateral frontal cortex, and primary visual cortex, which did not correlate with performance of the transposition task and so is likely associated with stimulus encoding. Zatorre et al. (2009) reported a similar pattern of activation for melodic retrograde tasks. This pattern of activation is consistent with a network involving multimodal and categorically encoded inputs and feedback as shown in Figure 5 (Rauschecker and Scott, 2009), with SWM located in the intraparietal sulcus and MAM located in the inferior parietal cortex.

INTERPRETATION OF FINDINGS IN RELATION TO THE EXPANDED OBJECT-ATTRIBUTE MODEL

According to the model shown in Figure 5, discrimination performance in the AO condition was an index of the efficiency of encoding in ASTM, while in the AV condition it was an index of the integration of the auditory and visually encoded information to form a stable representation. The manipulation task required spatial manipulation of memory traces in SWM and comparison to referent traces stored in MAM (Hooven et al., 2004). By accounting

for visual encoding and spatial manipulation ability (measured by the MRT), the auditory manipulation task serves as an index of the stability of the memory traces stored in MAM.

Overall, discrimination performance was better than manipulation performance, which supports the proposition that discrimination places less demands on MAM to maintain a stable representation over time. Discrimination and manipulation generally improved in the presence of notation. This suggests that the notation served as a scaffold for the memory trace in MAM. The interaction observed between task and the presence of notation for rhythmic stimuli further suggests that visual information was more important when the demands on MAM increased.

We propose that the pattern of increased facilitation of manipulation performance with increased spatial coding in the notation provides evidence for spatial dimensions in neural representations of pitch and time in ASTM. The facilitatory effect was most striking when visual information was spatially congruent with a linear dimension for time, as revealed by ceiling effects for the rhythmic stimuli. The facilitatory effect of a linear pitch notation was less striking, likely due to western notation employing more spatial information for pitch height than for beat duration. This is supported by the observation that melodic manipulation performance of the *western notation* group was better than rhythmic manipulation performance, despite more manipulation options and the presence of double manipulation trials.

Learning effects were observed for all AV manipulation tasks except for rhythmic manipulation trials with graphic notation. This provides strong evidence that participants learnt to assimilate visually encoded information with auditory memory traces, and that in all but this one condition, neither modality overly dominated in the performance of the task. Near ceiling performance in the rhythmic manipulation trials with graphic notation suggests that this notation is highly efficient for the spatial manipulation of rhythmic information, and that the visual modality dominated in the performance of the task. Conversely, learning rhythmic manipulation with western notation was the slowest overall, and likely required additional learning processes to decode the largely symbolic notation.

Most people in the developed world implicitly learn to recognize western music scales, intervals, and common rhythmic structures by early to mid-childhood (Cuddy and Badertscher, 1987; Wilson et al., 1997; Hannon and Trainor, 2007). However it is well known that extensive training is required to master western music notation. This type of training is likely to reflect additional processes of forming symbolic associations for auditory music identities and their visual notations in long-term memory. Given extensive training, composers may use symbolically encoded notation to undertake manipulations of music information in logical operations, thereby bypassing MAM and ASTM. In other words, they compose music without imagining the sounds they are symbolically manipulating. This pathway is identified as the categorical encoding pathway shown on the left of Figure 5.

CONCLUSIONS

The object-attribute model (McLachlan and Wilson, 2010) proposes that sound attributes such as pitch, loudness, duration, and location are spatially encoded in a multidimensional array in ASTM, and bound with identities that have been categorically encoded

in a parallel pathway. Rapid and accurate mapping of information between auditory and visual neural systems for complex and novel stimuli suggests a high level of compatibility between the neural codes. It is well established that visual information is spatially encoded (Rolls and Deco, 2002). The present data support the proposition that pitch and time are also spatially encoded in ASTM, and then combined with visual information in MAM.

A limitation of the study was the lack of a visual only control condition. To address this, the MRT was used to control for differences in visuo-spatial rotation ability. Furthermore, visual only processing was only likely to have occurred in the graphic notation condition for rhythmic stimuli. The presence of learning effects in all but this one condition suggests that the tasks generally involved learning to integrate auditory and visual neural representations, reflecting multimodal processing. Other possible limitations include differences in visual complexity or novelty between the western and graphic notations that may have influenced task performance. However, small differences in visual complexity alone are unlikely to account for the dramatic effects reported between conditions, particularly for rhythmic stimuli. Similarly, novelty may add a small amount of variance to the data, but is unlikely to be a significant contributing factor in novice musicians.

We have shown that visual representations facilitate the mental manipulation of auditory patterns. More linear representations, such as graphic notation, showed greater facilitation of task performance for novel stimuli, supporting congruence between spatial neural codes for pitch and time and visual neural representations.

REFERENCES

- Arnott, S. R., Binns, M. A., Grady, C. L., and Alain, C. (2004). Assessing the auditory dual-pathway model in humans. *Neuroimage* 22, 401–408.
- Bauer, M. I., and Johnson-Laird, P. N. (1993). How diagrams can improve reasoning. *Psychol. Sci.* 4, 372–378.
- Beecham, R., Reeve, R. A., and Wilson, S. J. (2009). Spatial representations are specific to different domains of knowledge. *PLoS ONE* 4, e5543. doi: 10.1371/journal.pone.0005543.
- Bendor, D., and Wang, X. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature* 436, 1161–1165.
- Bermudez, P., and Zatorre, R. J. (2005). Conditional associative memory for musical stimuli in nonmusicians: implications for absolute pitch. *J. Neurosci.* 25, 7718–7723.
- Bitterman, Y., Mukamel, R., Malach, R., Fried, I., and Nelken, I. (2008). Ultra-fine frequency tuning revealed in single neurons of human auditory cortex. *Nature* 451, 197–202.
- Bor, D., Duncan, J., Wiseman, R. J., and Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron* 37, 361–367.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol. Bull.* 104, 163–191.
- Cuddy, L. L., and Badertscher, B. (1987). Recovery of the tonal hierarchy: some comparisons across age and levels of musical experience. *Percept. Psychophys.* 41, 609–620.
- Cupchik, G. C., Phillips, K., and Hill, D. S. (2001). Shared processes in spatial rotation and music permutation. *Brain Cogn.* 46, 373–382.
- Dessain, P. (1992). A (de)composable theory of rhythm perception. *Music Percept.* 9, 439–454.
- Deutsch, D. (1999). “The processing of pitch combinations,” in *The Psychology of Music*, ed. D. Deutsch (San Diego: Academic Press), 349–404.
- Douglas, K. M., and Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nat. Neurosci.* 10, 915–921.
- Foster, N. E. V., and Zattore, R. J. (2009). A role for the intraparietal sulcus in transforming musical pitch information. *Cereb. Cortex* 20, 1350–1359.
- Frassinetti, F., Magnani, B., and Oliveri, M. (2009). Prismatic lenses shift time perception. *Psychol. Sci.* 20, 949–954.
- Frey, S., Campbell, J. S. W., Pike, G. B., and Petrides, M. (2008). Dissociating the human language pathways with high angular resolution diffusion fiber tractography. *J. Neurosci.* 28, 11435–11444.
- Gaare, M. (1997). Alternatives to traditional notation. *Music Educ. J.* 83, 17–23.
- Gomes, H., Ritter, W., and Vaughan, H. G. Jr. (1995). The nature of preattentive storage in the auditory system. *J. Cogn. Neurosci.* 7, 81–94.
- Grau, J. W., and Kemler-Nelson, D. G. (1988). The distinction between integral and separable dimensions: evidence for integrality of pitch and loudness. *J. Exp. Psychol. Gen.* 117, 347–370.
- Griffiths, T. D., and Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353.
- Hall, D. A., and Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cereb. Cortex* 19, 576–585.
- Hannon, E. E., and Trainor, L. J. (2007). Music acquisition: effects of enculturation and formal training on development. *Trends Cogn. Sci.* 11, 451–498.
- Hasegawa, T., Matsuki, K., Ueno, T., Maeda, Y., Matsue, Y., Konishi, Y., and Sadato, N. (2004). Learned audio-visual cross-modal associations in observed piano playing activate the left planum temporale. An fMRI study. *Cogn. Brain Res.* 20, 510–518.
- Henry, M. J., and McAuley, J. D. (2009). Evaluation of an imputed pitch velocity model of the auditory kappa effect. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 551–564.
- Hooven, C. K., Chabris, C. F., Ellison, P. T., and Kosslyn, S. M. (2004). The relationship of male testosterone to components of mental rotation. *Neuropsychologia* 42, 782–790.
- Hughes, H. C., Darcey, T. M., Barkan, H. I., Williamson, P. D., Roberts, D. W., and Aslin, C. H. (2001). Responses of human auditory association cortex to the omission of an expected acoustic event. *Neuroimage* 13, 1073–1089.
- Ioannides, A. A., Popescu, M., Otsuka, A., Bezerianos, A., and Liu, L. (2003). Magnetoencephalographic evidence of the interhemispheric asymmetry in echoic memory lifetime and its dependence on handedness and gender. *Neuroimage* 19, 1061–1075.
- Koedinger, K. R., and Anderson, J. R. (1990). Abstract planning and perceptual chunks: elements of expertise in geometry. *Cogn. Sci.* 14, 511–550.
- Krumhansl, C. L. (2000). Rhythm and pitch in music cognition. *Psychol. Bull.* 126, 160–179.
- Larkin, J. H., and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cogn. Sci.* 11, 65–99.
- Levitin, D. J., and Rogers, S. E. (2005). Absolute pitch: perception coding

- and controversies. *Trends Cogn. Sci.* 9, 26–33.
- Lomber, S. G., and Malhotra, S. (2008). Double dissociation of “what” and “where” processing in auditory cortex. *Nat. Neurosci.* 11, 609–616.
- Luria, A. R. (1980). *Higher Cortical Functions in Man*. New York: Basic Books.
- Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., Pittet, A., and Clarke, S. (2001). Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage* 14, 802–816.
- McGurk, H., and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature* 264, 746–748.
- McLachlan, N. M. (2000). A spatial theory of rhythmic resolution. *Leonardo Music J.* 10, 61–67.
- McLachlan, N. M., and Wilson, S. J. (2010). The central role of recognition in auditory perception: a neurobiological model. *Psychol. Rev.* 117, 175–196.
- Melara, R. D., and Marks, L. E. (1990). Processes underlying dimensional interactions—correspondences between linguistic and nonlinguistic dimensions. *Mem. Cognit.* 18, 477–495.
- Milner, B. (1972). Disorders of learning and memory after temporal lobe lesions in man. *Clin. Neurosurg.* 19, 421–446.
- Molholm, S., Sehatpour, P., Mehta, A. D., Shpaner, M., Gomez-Ramirez, M., Ortigue, S., Dyke, J. P., Schwartz, T. H., and Foxe, J. J. (2006). Audio-visual multisensory integration in superior parietal lobule revealed by human intracranial recordings. *J. Neurophysiol.* 96, 721–729.
- Mudd, S. A. (1963). Spatial stereotypes of four dimensions of pure tone. *J. Exp. Child. Psychol.* 66, 347–352.
- Näätänen, R., and Winkler, I. (1999). The concept of auditory stimulus representation in cognitive science. *Psychol. Bull.* 125, 826–859.
- Nelson, S. G. (2008). “Court and religious music (1): history of gagaku and shōmyō,” in *The Ashgate Research Companion to Japanese Music*, eds A. McQueen Tokita and D. W. Hughes (Hampshire: Ashgate), 35–48.
- O’Leary, A., and Rhodes, G. (1984). Cross-modal effects on visual and auditory object perception. *Percept. Psychophys.* 35, 565–569.
- Pandya, D. N., and Yeterian, E. H. (1985). Architecture and connections of cortical association areas. *Cereb. Cortex* 4, 3–61.
- Petrides, M., and Milner, B. (1982). Deficits on subject-ordered tasks after frontal and temporal-lobe lesions in man. *Neuropsychologia* 20, 149–202.
- Pratt, C. C. (1930). The spatial character of high and low tones. *J. Exp. Psychol.* 13, 278–285.
- Pressing, J. (1983). Cognitive isomorphisms between pitch and rhythm in world musics: west africa, the balkans and western tonality. *Stud. Music* 17, 38–61.
- Rauschecker, J. P., and Scott, S. K. (2009). Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* 12, 718–724.
- Rauschecker, J. P., and Tian, B. (2000). Mechanisms and streams for processing “what” and “where” in the auditory cortex. *Proc. Natl. Acad. Sci. U.S.A.* 97, 11800–11806.
- Roffler, S. K., and Butler, R. A. (1968). Localization of tonal stimuli in the vertical plane. *J. Acoust. Soc. Am.* 43, 1260–1266.
- Rolls, E. T., and Deco, G. (2002). *Computational Neuroscience of Vision*. Oxford: Oxford University Press.
- Rusconi, E., Kwan, B., Giordano, B., Umiltà, C., and Butterworth, B. (2005). The mental space of pitch height. *Ann. N. Y. Acad. Sci.* 1060, 195–197.
- Rusconi, E., Kwan, B., Giordano, B. L., Umiltà, C., and Butterworth, B. (2006). Spatial representation of pitch height: the SMARC effect. *Cognition* 19, 113–129.
- Scaife, M., and Rogers, Y. (1996). External cognition: how do graphical representations work? *Int. J. Hum. Comput. Stud.* 45, 185–213.
- Schönwiesner, M., Novitski, N., Pakarinen, S., Carlson, S., Tervaniemi, M., and Näätänen, R. (2007). Heschl’s gyrus, posterior superior temporal gyrus, and mid-ventrolateral prefrontal cortex have different roles in the detection of acoustic changes. *J. Neurophysiol.* 97, 2075–2082.
- Shamsa, L., Kamitanib, Y., and Shimojo, S. (2002). Visual illusion induced by sound. *Cogn. Brain Res.* 14, 147–152.
- Shepard, R. N., and Jordan, D. S. (1984). Auditory illusions demonstrating that tones are assimilated to an internalized musical scale. *Science* 226, 1333–1334.
- Sherman, S. M. (2007). The thalamus is more than just a relay. *Curr. Opin. Neurobiol.* 17, 417–422.
- Sloboda, J. A. (2005). *Exploring the Musical Mind: Cognition, Emotion, Ability, Function*. Oxford: Oxford University Press.
- Trimble, O. C. (1934). Localization of sound in the anterior, posterior and vertical dimensions of auditory space. *Br. J. Psychol.* 24, 320–334.
- Vandenberg, S. G., and Kuse, A. R. (1972). *Mental Rotations: A New Spatial Test*. Boulder: Institute for Behavioral Genetics, University of Colorado.
- Virsu, V., Oksanen-Hennah, H., Vedenpää, A., Jaatinen, P., and Lahti-Nuutila, P. (2008). Simultaneity learning in vision, audition, tactile sense and their cross-modal combinations. *Exp. Brain Res.* 186, 525–537.
- Wilson, S. J., Pressing, J. L., and Wales, R. J. (2002). Modelling rhythmic function in a musician post-stroke. *Neuropsychologia* 40, 1494–1505.
- Wilson, S. J., Pressing, J. L., Wales, R. J., and Pattison, P. (1999). Cognitive models of music psychology and the lateralisation of musical function within the brain. *Aust. J. Psychol.* 51, 125–139.
- Wilson, S. J., Wales, R. J., and Pattison, P. (1997). The representation of tonality and meter in children aged 7 and 9. *J. Exp. Child. Psychol.* 64, 42–66.
- Zatorre, R. J., Bouffard, M., Ahad, P., and Belin, P. (2002). Where is “where” in the human auditory cortex? *Nat. Neurosci.* 5, 905–909.
- Zatorre, R. J., Halpern, A. R., and Bouffard, M. (2009). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *J. Cogn. Neurosci.* 22, 775–789.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 10 May 2010; accepted: 09 December 2010; published online: 27 December 2010.

Citation: McLachlan NM, Greco LJ, Toner EC and Wilson SJ (2010) Using spatial manipulation to examine interactions between visual and auditory encoding of pitch and time. *Front. Psychology* 1:233. doi: 10.3389/fpsyg.2010.00233

This article was submitted to *Frontiers in Auditory Cognitive Neuroscience*, a specialty of *Frontiers in Psychology*. Copyright © 2010 McLachlan, Greco, Toner and Wilson. This is an open-access article subject to an exclusive license agreement between the authors and the *Frontiers Research Foundation*, which permits unrestricted use, distribution, and reproduction in any medium, provided the original authors and source are credited.