

# Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity

VICTORIA J. WILLIAMSON

*Goldsmiths, University of London, London, England*

AND

ALAN D. BADDELEY AND GRAHAM J. HITCH

*University of York, York, England*

Language–music comparative studies have highlighted the potential for shared resources or neural overlap in auditory short-term memory. However, there is a lack of behavioral methodologies for comparing verbal and musical serial recall. We developed a visual grid response that allowed both musicians and nonmusicians to perform serial recall of letter and tone sequences. The new method was used to compare the phonological similarity effect with the impact of an operationalized musical equivalent—pitch proximity. Over the course of three experiments, we found that short-term memory for tones had several similarities to verbal memory, including limited capacity and a significant effect of pitch proximity in nonmusicians. Despite being vulnerable to phonological similarity when recalling letters, however, musicians showed no effect of pitch proximity, a result that we suggest might reflect strategy differences. Overall, the findings support a limited degree of correspondence in the way that verbal and musical sounds are processed in auditory short-term memory.

There is a long scientific tradition of examining the similarities between language and music (Darwin, 1871; Patel, 2008). Studying commonalities between how language and musical sounds are processed affords more opportunities to learn about how we interact with the auditory environment than are available from studying either domain in isolation (Callan et al., 2006; Koelsch et al., 2009; Kraus & Banai, 2007; Patel, 2009; Peretz & Zatorre, 2005; Wong, Skoe, Russo, Dees, & Kraus, 2007). In the present article, we compare the serial recall of speech and music from short-term memory.

The first reason for choosing short-term memory for a language–music comparison is that much is known about the operation of auditory–verbal serial recall. For simplicity, we present our studies within a single broad theoretical framework—the multicomponent working-memory model (Baddeley, 2000; Baddeley & Hitch, 1974)—although we acknowledge that they could be fitted into other frameworks developed to explain processing in short-term memory. The working-memory model comprises an attention-controlling central executive and three subsystems: the phonological loop, the visuospatial sketchpad, and the episodic buffer. The phonological loop is associated with the processing of speechlike information. It comprises a passive phonological store, which acts

as a temporary holding center for speech-based information, and an articulatory rehearsal process, during which incoming visual information can be recoded and rehearsed using a phonological code (Baddeley, 2007). In the present research, we question whether the phonological loop also may be capable of processing musical stimuli.

A second reason for focusing the language–music comparison within short-term memory is the importance of verbal short-term memory for a range of higher cognitive abilities. These include the acquisition of vocabulary and grammar (Baddeley, Papagno, & Vallar, 1988; Gathercole & Baddeley, 1990), reading (Baddeley, Gathercole, & Papagno, 1998), and action control (Baddeley, Chincotta, & Adlam, 2001; Liefooghe, Barouillet, Vandierendonck, & Camos, 2008; Miyake et al., 2000). An empirical investigation of language and music processing in short-term memory therefore has the potential to encourage investigations into the role of musical short-term memory in equivalent domain cognitive abilities (e.g., reading music). There is also the possibility, if evidence for resource sharing can be demonstrated, that music may have a role in a number of important verbal processes, such as those listed above. A link between music skills and phonological awareness has been demonstrated in a number of different populations (Anvari, Trainor, Woodside, & Levy,

---

V. J. Williamson, v.williamson@gold.ac.uk

---

2002; J. L. Jones, Lucker, Zalewski, Brewer, & Drayna, 2009; Overy, 2003).

A third reason for our empirical focus is that neuroimaging evidence suggests that a degree of resource sharing occurs during verbal and musical short-term memory tasks. Studies have demonstrated that areas of common activation in such tasks include, but are not exclusive to, Broca's area, premotor cortex, the supplementary motor area, the intraparietal sulcus, and the supramarginal gyrus (Brown & Martinez, 2007; Brown, Martinez, Hodges, Fox, & Parsons, 2004; Brown, Martinez, & Parsons, 2006; Gaab et al., 2005; Hickok, Buchsbaum, Humphries, & Muftuler, 2003). Koelsch et al. (2009) conducted the first fMRI study to directly compare the neural architecture active during both a verbal and musical short-term memory task. Overall, the authors identified a pattern of activation that was remarkably similar.

By comparison, the majority of the evidence from the neuropsychological literature has emphasized separate processing domains for speech and for music (Peretz & Zatorre, 2005). An often-cited double dissociation is that which exists between amusia (a music perception deficit in the absence of difficulties with language) and aphasia (language processing difficulty in the absence of amusia; Ayotte, Peretz, & Hyde, 2002; Luria, Tsvetkova, & Futer, 1965; Marin & Perry, 1999; Stewart, von Kriegstein, Warren, & Griffiths, 2006). Patel, Peretz, Tramo, and Labreque (1998) examined a case of acquired amusia in the absence of aphasia by testing the ability of participant IR to discriminate between pairs of sentences in which the prosodic focus was shifted (i.e., "Take the *train* to Bruges, Anne" or "Take the train to *Bruges*, Anne") or analog melodies in which fundamental frequency was the only salient cue for discrimination. IR performed this task at chance level, compared with controls in both tasks. The authors proposed that IR had a problem maintaining pitch patterns in short-term memory and that this difficulty had the potential to affect both her verbal and musical processing. They argued that additional syntactic cues were sufficient to prevent the problem from affecting IR's everyday comprehension of language. However, being unable to maintain pitch had resulted in a profound deficit in her ability to process music.

Taken together, these studies indicate a degree of common neural resource sharing for short-term memory tasks that use speech- or pitch-based materials. However, many of the brain areas identified are large enough to encompass two distinct processing systems (Marcus, Vouloumanos, & Sag, 2003). There are also issues of establishing causality from neuroimaging results and of generalizing findings from neuropsychological cases. To effectively address the question of resource sharing, therefore, we must also compare verbal and musical short-term memory using analogous behavioral paradigms (Peretz & Zatorre, 2005).

There are many models of short-term memory that could be used to test whether speech and music stimuli show similar processing characteristics in a behavioral paradigm—for example, feature model (Nairne, 1990; Neath, 2000), scale-invariant memory and perceptual learning (SIMPLE; Neath & Brown, 2006), and object-oriented episodic record (O-OER; D. M. Jones, 1993).

We chose to focus on the working-memory model for two reasons. The first was that this framework had already been modified in a theoretical attempt to explain music processing. Although Berz (1995) proposed a link between the phonological loop and a new "musical loop," that study presented little detail regarding the nature of the new loop or the link with speech processing, so the model cannot be developed to make testable predictions. The present article contributes to the debate over domain specificity in the working-memory model.

A second reason why we selected the working-memory model as an empirical framework was because it has had a degree of success in explaining the results of a number of behavioral paradigms within which short-term memory for verbal and musical materials can be compared. One that already had been employed is the irrelevant sound paradigm (Colle & Welsh, 1976). The premise for such an experiment is that shared processing in short-term memory should result in irrelevant background music interfering with memory for speech and vice versa. Salamé and Baddeley (1989) found that serial recall of verbal stimuli was impaired more significantly by irrelevant speech and by instrumental music compared with the presence of amplitude-modulated noise (a control for the general distraction effect of sound) or with silence. Similar irrelevant music/tone effects on visual and auditorily presented items have since been demonstrated in a number of studies (Hadlington, Bridges, & Darby, 2004; D. M. Jones & Macken, 1993; D. M. Jones, Macken, & Murray, 1993; Macken, Tremblay, Houghton, Nicholls, & Jones, 2003; Schlittmeier, Hellbrück, & Klatte, 2008).

In the reverse situation (disruption of music memory by language), results have been more varied. Deutsch (1970) found that irrelevant tones disrupted performance on tone recognition, but that irrelevant speech materials had little if any detrimental effect. Since that study however, several authors have reported a significant effect of background speech and music on memory for musical sounds (D. M. Jones, Macken, & Harries, 1997; Pechmann & Mohr, 1992; Semal & Demany, 1991, 1993; Semal, Demany, Ueda, & Hallé, 1996). Semal et al. concluded that the pitch dimension, whether it be music based or speech based, is the element that is commonly disruptive to both verbal and musical processing in short-term memory.

Our aim for the following experiments was to carry out a comparison of short-term memory for verbal and musical pitch materials with two new developments. First, we designed an analogous response method that could be used to measure immediate serial recall of both verbal and musical materials. Second, we moved on from the irrelevant sound experiments and selected another manipulation known to affect immediate serial recall of verbal materials—phonological similarity. We used a musical equivalent to this manipulation in order to directly compare similarities in the features of recall across verbal and musical materials.

### The New Method

The majority of music serial-recall tasks require the ability to sing (Sloboda & Parker, 1985) or a knowledge of

	Tone 1	Tone 2	Tone 3	Tone 4
High				
Medium				
Low				

Figure 1. An example of the response grid used by participants to recall a four-tone sequence.

music notation (Roberts, 1986; Schendel & Palmer, 2007). Such tasks are unsuitable for the majority of people, who are not trained musicians. The new method used a simple visual grid response. To-be-remembered sequences were composed from a pool of three tones. Three tones were chosen, because pilot work indicated that nonmusicians' performance suffered with longer sequences. After hearing each sequence, participants entered their immediate serial-recall response into the grid (see Figure 1). Keller, Cowan, and Sauls (1995) used a similar system in their study of auditory memory, but we believe our study to be its first use in a serial-recall paradigm.

### The Similarity Manipulation

Phonological similarity describes the phenomenon by which immediate serial recall of a sequence of visual-verbal items is detrimentally affected when the items are acoustically similar as opposed to when they are dissimilar (e.g., B, V, G vs. F, K, R; Conrad, 1964; Conrad & Hull, 1964). Phonological similarity has also been shown to detrimentally affect performance when presentation of to-be-remembered materials is auditory (Baddeley, Lewis, & Vallar, 1984; Surprenant, Neath, & LeCompte, 1999).

Phonological similarity is a robust and highly replicable effect that is regarded as a characteristic of immediate serial recall from auditory-verbal short-term memory. This makes it a suitable manipulation to compare memory for verbal and musical sequences. If the same short-term memory system is processing speech and pitch sounds, the following testable prediction can be made: Tonal similarity, a manipulation akin to phonological similarity, will detrimentally impair performance when participants recall sequences of tones.

In the following experiments, we manipulated the distance between the pitches of notes within to-be-remembered sequences in order to increase tonal similarity. Sequences in which tones were close together were termed *pitch proximal*, and sequences in which the tones were further apart were termed *pitch distal*. One issue with using pitch proximity as a manipulation is that of perceptual confusion. In the case of verbal material, confusion at encoding does not appear to contribute to the phonological similarity effect (Baddeley, 1966). However, if participants were less able to discriminate among the tone pitches, then any proximity effect could represent an encoding effect. We took a number of steps to reduce the impact of this factor. The

first was to use only tone intervals that were larger than a whole tone. Research has suggested that this interval can be discriminated by the majority of people (Kishon-Rabin, Amir, Vexler, & Zaltz, 2001; Levitin, 2006; Moore, 2003). Second, we introduced a pitch-recognition training task designed to identify participants who were unable to reliably identify a difference between the tones.

Surprenant, Pitt, and Crowder (1993, Experiment 4) manipulated the distance between pitches in a serial-recall task to test musicians' and nonmusicians' recall of four-tone eight-item sequences that were close [C4 (262 Hz), C#4 (277 Hz), D4 (294 Hz), D#4 (311 Hz)] or distant [C3 (130 Hz), B3 (247 Hz), A4 (440 Hz), G5 (784 Hz)] in pitch. Participants learned to associate the four tones with the numbers 1–4 and used buttonpresses to record their response. They found a significant effect of proximity in nonmusicians but not in musicians, which suggests an interaction between proximity and expertise that we have attempted to replicate in the present study. Although this result is promising, there are a number of potential limitations, including the use of a direct association verbal response, random generation of tone sequences, and a floor effect in the nonmusicians' data. There was also no direct comparison of a verbal equivalent using an analogous response.

A pilot was carried out to compare 24 nonmusicians' serial recall of six-item verbal sequences as a function of phonological similarity with their recall of four-item tone sequences as a function of pitch proximity. There was a significant effect ( $p < .001$ ) of similarity in the predicted direction (letters,  $M = 87.44\%$  vs.  $63.83\%$ ; tones,  $M = 84.11\%$  vs.  $69.79\%$ ). This was qualified by a significant interaction, reflecting a larger similarity effect for letter recall. Encouraged by this result, we set out to further examine the pitch-proximity effect across three experiments. In the first, we refined the methodology and explored the effect of increasing sequence length. In the second, we replicated the pitch-proximity effect using a serial-recognition paradigm. In the third, we compared the effects of phonological similarity and pitch proximity in both musicians and nonmusicians.

## EXPERIMENT 1

In Experiment 1, we sought to replicate the pitch-proximity effect. Our hypothesis was that pitch-distal sequences would be recalled more successfully than would

pitch-proximal sequences. We focused solely on memory for tones at this point, deferring further comparisons with memory for verbal stimuli until we had learned more about the pitch-proximity effect.

Along with a replication of the pitch-proximity effect, we were interested in whether short-term memory for tones shared characteristics with verbal short-term memory. Research on verbal short-term memory suggests that increasing list length detrimentally influences performance (Baddeley & Larsen, 2003; Hanley & Bakopoulou, 2003; Salamé & Baddeley, 1986). In Experiment 1, therefore, we examined the robustness of the pitch-proximity effect across changes in sequence length. A parallel with verbal short-term memory would suggest that pitch proximity might have smaller or less reliable effects for longer sequences.

## Method

**Participants.** Twenty-four University of York undergraduates (6 male, 18 female) aged from 18 to 35 ( $M = 21.38$ ,  $SD = 4.14$ ) completed the experiment for course credit or £4. None had formal music training and all reported normal hearing and were naive with respect to the experiment.

**Design.** The experiment used a  $2 \times 5$  within-subjects design with tone similarity (proximal vs. distal) and sequence length (3–7) as factors. The dependent variable was the proportion of items recalled in the correct serial position.

**Materials.** The musical stimuli consisted of three-tone sequences. Pitch-proximal tones were C4, D4, and E4; pitch-distal tones were C4, G4, and B4. Tones were chosen to be of similar tonal strength according to Krumhansl's (1990) theory of tonal hierarchy in the context of C major. Tones were played on a Disklavier and were recorded in stereo using ProTools LE with MBox hardware and two condenser microphones (AKG 414). The recordings were edited as .wav files in Adobe Audition, so that each tone lasted 800 msec, and signal levels were normalized. Each tone had a maximum level of 74 dB at the ear. A 200-msec pause of generated silence was inserted at the end of each tone. A C-major chord (C4, E4, G4), used as an auditory cue, was compiled as a multitrack waveform.

Four sequences (one practice and three experimental) were randomly generated for each sequence length condition. Any sequence that involved an immediate repetition of a tone was rejected. An equal number of sequences began with each of the three tones.

**Procedure.** Presentation of similarity conditions was blocked and counterbalanced across participants. Previous studies of musical memory have used pitch training to familiarize participants with tones (Greene & Samuel, 1986) and to reduce error variance by identifying individuals who have problems discriminating them (D. M. Jones et al., 1997). Before each block in the present experiment, participants received separate pitch training, which consisted of brief equal exposure to the tones used (*exposure phase*) followed by practice in pitch identification (*discrimination phase*).

During the exposure phase, participants were played the sequence of the three tones in ascending pitch height 10 times. In the discrimination phase, participants heard a C-major chord cue followed, after 2 sec, by one of the tones. The selection of tones was random, but was the same for each participant. Each tone was presented four times, making a total of 12 trials. Participants identified each tone by marking the relevant box on the grid.

After the training was completed, the recall task began. Within each block, there were three trials at each sequence length, starting with three-item sequences and stepping up to seven-item sequences. Immediately before each block, participants completed one practice trial at each sequence length starting with a three-item sequence and stepping up to a seven-item sequence. A 5-min rest interval was given between blocks. No feedback was given until the end of the experiment.

Before the experiment began, participants were instructed that they should try to maintain the sounds of the tones in mind, but that their rehearsal must be silent. Participants were also required to respond in the correct serial order. Responses were observed to ensure compliance with instructions. On each trial, a cross appeared on the screen for 2 sec and a C-major chord was sounded. After a 1-sec silence, the to-be-remembered sequence was played. Participants performed their recall immediately after the end of the sequence by marking the relevant boxes on the response grid. They triggered the next trial when they were ready. Each session lasted approximately 45 min.

## Results

A criterion of 10 out of 12 correct answers was set for the training task. Similar criteria have been set by previous experiments (e.g., 7/10 in D. M. Jones et al., 1997). Three participants who failed to reach criterion in the proximal condition were replaced. Proportion correct scores as a function of experimental conditions can be seen in Figure 2.

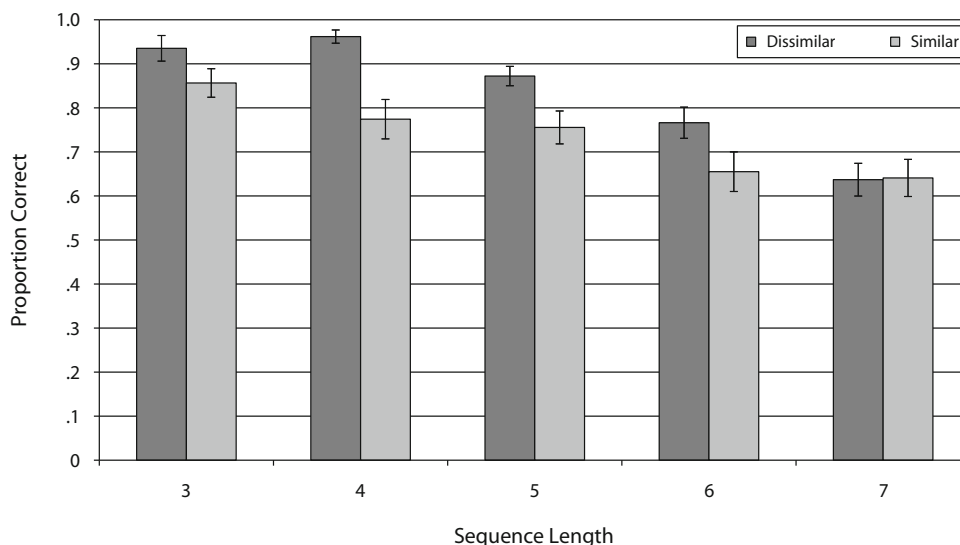
A two-way ANOVA on the arcsine-transformed data revealed significant effects of similarity [ $F(1,23) = 28.44$ ,  $MS_e = 150.494$ ,  $p < .001$ ,  $\eta_g^2 = .20$ ;  $M = 0.84$  vs.  $0.74$ ] and sequence length [ $F(4,92) = 43.05$ ,  $MS_e = 113.02$ ,  $p < .001$ ,  $\eta_g^2 = .34$ ], both in the predicted directions. There was also a significant interaction [ $F(4,92) = 4.89$ ,  $MS_e = 111.36$ ,  $p = .001$ ,  $\eta_g^2 = .05$ ], which was investigated using simple main effects. There were significant effects of similarity at all sequence lengths, except for the seven-note sequence ( $p = .58$ ).

## Discussion

Our nonmusician participants were more accurate in their immediate serial recall with pitch-distal tone sequences than with pitch-proximal sequences. This pitch-proximity effect was found alongside the standard phonological similarity effect for letters in the pilot study. Thus, tones, like letters, appear to be subject to acoustic confusability based on information that may be stored and rehearsed for serial recall: phonemes for language and pitch for tones.

This experiment also suggested a limit on the conditions under which the pitch proximity is found. As in verbal short-term memory, recall declined with increasing sequence length. This effect is consistent with a storage system that has limited capacity. A limited capacity for tone memory previously has been demonstrated in a recognition paradigm comparing 7-tone sequences with 10-tone sequences (Croonen, 1994), but, to our knowledge, this is the first demonstration of it in serial recall. The interaction between similarity and list length suggests that the proximity effect declines or is less reliable as sequence length increases, a pattern often found in verbal short-term memory when the strategy of relying on a phonological code is abandoned with increasing task difficulty (Baddeley & Larsen, 2003; Hanley & Bakopoulou, 2003; Salamé & Baddeley, 1986). However, the data by no means force this conclusion, and further evidence is needed to rule out other interpretations.

It is important to question whether the present pitch-proximity effect can generalize from the serial-recall



**Figure 2. Experiment 1: Proportion of tones correct across all musical sequence lengths for both similarity conditions. Error bars represent SEMs.**

paradigm. The next experiment explored the effect in an alternative task: serial recognition.

## EXPERIMENT 2

Phonological similarity detrimentally affects performance when memory is tested using a serial-recognition, in addition to a recall, paradigm (Henson, Hartley, Burgess, Hitch, & Flude, 2003; Nimmo & Roodenrys, 2005). If pitch proximity is an analogous finding, the effect should also be found in serial recognition. This logic was the premise for the present experiment comparing verbal and musical memory.

An important issue for the present experiment was how to design a serial-recognition task that would be suitable for both speech and tone stimuli. Verbal serial-recognition experiments typically test whether people can recognize an alteration to the order of items in a sequence. Music recognition experiments examine whether contour (pattern of ups and downs) or interval (the relative pitch height of tones) information is more salient in recognizing transposed melodies (in which the pitch relationships within the sequence remain the same, but the sequence begins on a different pitch). These methodological differences are driven by the need to show stimulus-specific effects. One method that is common to verbal and musical paradigms, however, is to reverse the order of two items upon re-presentation of a sequence on 50% trials.

There would be a problem with using only this order manipulation in the present experiment. Altering order in the tone sequences would always alter contour. Dowling (1991, 1994) suggests that participants store information about contour as well as about the interval of the tones in novel melodies. Therefore, altering the contour of the tone

sequences would provide information not available in the verbal condition.

In order to resolve the above concerns, we applied two manipulations. Order was manipulated in half the trials. In the other half, a single tone or a letter was replaced with another item from the pool; in the tone sequences, the contour of the original sequence was maintained.<sup>1</sup> By this method, we created a paradigm that maintained many similarities from the verbal literature, but that also had controls in place to meet the specific demands of testing with tones.

## Method

**Participants.** Twenty-four staff and students (6 male, 18 female) from the University of York or York St. John University College completed the experiment in return for course credit or £4. They were aged from 18 to 48 ( $M = 23.58$ ,  $SD = 8.82$ ) and met the criteria set down in Experiment 1.

**Design.** A  $2 \times 2$  split-plot design was employed. The between-subjects variable was stimulus type (verbal vs. musical). The within-subjects variable was similarity (similar vs. dissimilar). The dependent variable was recognition performance as measured by  $d'$ .

**Materials.** As in Experiment 1, pitch-distal tones were C4, G4, and B4 and pitch-proximal tones were C4, D4, and E4. The tones were generated using Audition software, allowing greater control over their acoustic attributes than the Disklavier recordings. Each tone was a sine wave sampled at 44100 Hz, with 16-bit resolution. Duration was set to 800 msec, followed by a 200-msec period of silence. The fundamental frequency of each tone matched the respective natural piano tone. The amplitude envelope for each tone was identical and was generated by mimicking the pattern of a piano key strike (i.e., quick attack and gradual decay). To further improve the ecological validity of the sounds, each of the first four harmonic components in the series was reduced by 12%.

Phonologically different letters were M, Q, and R, and similar letters were B, D, and G.

The letters were recorded in a sound-attenuating booth using a single microphone. A female spoke all six letters at a slow, con-



tinuous pace using a monotonous pitch (fundamental frequencies are presented in the Appendix). Signal levels were normalized and matched the average amplitude of the tone files. Each letter differed slightly in length, but no letter lasted longer than 800 msec or was briefer than 600 msec. A gap of the necessary amount of silence was added in order to ensure each item was 1 sec in total duration.

Seven-item sequences were generated for both verbal and musical conditions. In Experiment 1, sequences were generated randomly. A potential problem in this design is that different patterns of tones may be recalled with more ease than others (e.g., patterns where contour changed less frequently). In the present experiment, two lists of sequences were rotated across conditions.

A computer program produced all possible seven-item combinations of the numbers 1, 2, and 3, with the constraint that no number could immediately repeat. Sequences were then divided into 12 pools that began in a similar way (e.g., 121, . . . , 212, . . . , 312). In order to generate two 36-sequence lists, 3 sequences were selected from each of the 12 pools, therefore balancing equally for starting item. The amount of overlap in structure across sequences was minimized by inspection. Finally, 3 sequences, 1 of each starting with the same number, were selected from the remaining pool of items to serve as practice sequences.

Within each list of 36 sequences, 12 began on the same number. Within these 12 sequences, we balanced for number of manipulations (order and interval) and place of manipulation (start, middle, or end of sequence). Distribution of manipulations (i.e., interval vs. order) was made randomly, but according to the constraints of sequence (i.e., an order manipulation is not always suitable as sometimes it creates a repeated number). Half the practice items were the same, and half had a different manipulation.

**Procedure.** The 12-trial pitch training was modified to reflect the recognition format of the present experiment, as opposed to the recall format used in Experiment 1. Participants first saw a cross on screen, which was followed by the C-major chord. After a 2-sec silence, the first tone was played. After a further 2-sec silence, the second tone was played. Participants were asked to say whether the second tone was the same or different by pressing one of two buttons on the keyboard. Immediate feedback on performance was provided.

Each participant completed two memory tasks, with a 5-min break in between. Each condition involved 3 practice trials followed by 36 experimental trials. Trials were broken up into three blocks of 12 with a 1-min pause between blocks. During each trial, a cross appeared on the screen for 2 sec, followed by the musical chord. After a 2-sec silence, the first sequence was played. After a further 2-sec pause, the second sequence was played. Participants responded im-

mediately by pressing one of two keys: "S" for same, and "D" for different. Their response triggered the next trial.

## Results

Two participants failed to meet criterion for the pitch-training task in the proximal condition and were replaced. We converted each participant's data into a  $d'$  score using the same-different independent observations model (Macmillan & Creelman, 2005). There were no significant main effects or interactions associated with the interval versus order manipulation, so the data were collapsed. Figure 3 shows the results.

A two-way ANOVA revealed a significant effect of similarity [ $F(1,22) = 15.40$ ,  $MS_e = 0.52$ ,  $p = .001$ ,  $\eta_g^2 = .16$ ], indicating better performance in the dissimilar conditions ( $M = 1.07$  vs. 1.89). The effect of stimulus type (verbal vs. musical) was not significant [ $F(1,22) = 0.65$ ,  $MS_e = 1.34$ , n.s.], and neither was the interaction between the two variables [ $F(1,22) = 1.17$ ,  $MS_e = 0.52$ , n.s.].

## Discussion

The results of this serial-recognition experiment suggest that, as predicted, the detrimental impact of acoustic similarity was present in both verbal and musical conditions. This pattern of results remained when the participants were divided into low performers and high performers, indicating that neither effect was driven by performance level. The finding of a phonological similarity effect in the verbal data supports previous literature that has reported similar findings in verbal serial-recognition paradigms (Henson et al., 2003; Nimmo & Roodenrys, 2005). It also extends these findings to conditions in which sequences are created from a pool of only three verbal items.

One surprising finding was the lack of a difference between the interval and contour manipulation in the tone condition. It was expected that the contour change would be easier to detect, given the importance of contour to basic melody memory compared with information about intervals (Dowling, 1991, 1994). This finding may have resulted from the use of a limited pool of tones. In these circum-

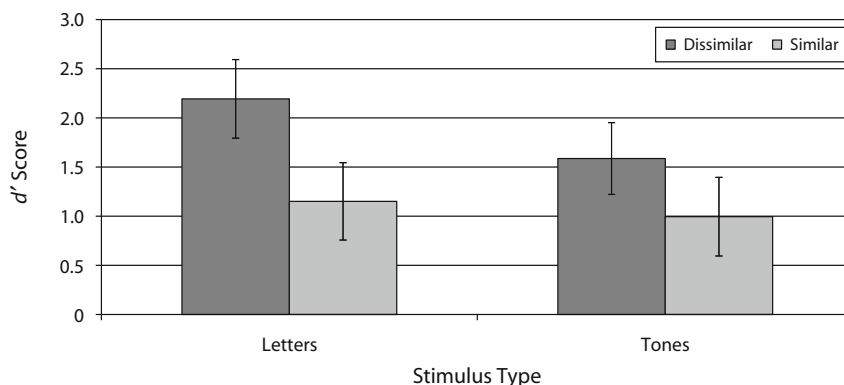


Figure 3. Experiment 2:  $d'$  scores for both letter and tone serial recognition in acoustically dissimilar and similar conditions. Chance performance is  $d' = 0$ . Error bars represent SEMs.

stances, participants could have encoded details about the number of times that each tone appears and thereby noticed when an “extra” tone was present. This finding does not affect the overall conclusion. The results suggest that pitch proximity is not an artifact of the new serial-recall method used in the pilot and in Experiment 1, but is an effect, like phonological similarity, that can be replicated in an alternative, established short-term memory paradigm.

### EXPERIMENT 3

In the final experiment, we attempted to again replicate the pitch-proximity effect in serial recall, but this time we included the methodological improvements from Experiment 2 (e.g., use of artificial tones, rotation of sequence structure). In order to make a direct comparison between the effects of phonological similarity and pitch proximity, we included an equivalent verbal recall test (i.e., using a grid response and a set pool of only three stimuli).

We also looked at the issue of expertise. There is an inherent confound present in the comparison of verbal and musical memory. All literate participants are arguably experts in dealing with verbal materials, but, when it comes to musical stimuli, there is a continuum of expertise. For this reason, we sought to determine how musical expertise affected short-term memory performance for our tone and letter recall.

Musicians’ tone recall is likely to be better than that of nonmusicians. Musicians are well practiced in the skills required to encode, retain, and recall large pieces of music. Musicians create and use hierarchical retrieval structures (Clarke, 1988; Williamon & Egner, 2004; Williamon & Valentine, 2002) akin to the type proposed by skilled memory theory and long-term working-memory theory (Chase & Ericsson, 1981; Ericsson & Kintsch, 1995). They may also have improved pitch-perception abilities (Besson, Schön, Moreno, Santos, & Magne, 2007; Micheyl, Delhommeau, Perrot, & Oxenham, 2006), increased activation in areas of the brain associated with auditory short-term memory storage (Gaab & Schlaug, 2003), and improved working-memory operations with regard to musical materials (Pechmann & Mohr, 1992). This experiment was designed, not to distinguish among the potential explanations for musicians’ superior memory, but to establish whether improved performance could be found by using the new method of testing serial recall.

By comparison, the literature on the performance of musicians in verbal tasks is limited. Some evidence suggests that musicians outperform nonmusicians on verbal memory tasks (Besson et al., 2007; Brandler & Rammsayer, 2003; Chan, Ho, & Cheung, 1998; Franklin et al., 2008; Ho, Cheung, & Chan, 2003; Jakobson, Lewycky, Kilgour, & Stoesz, 2008). However, the tests reported typically were carried out on small populations and used nonstandard measures adapted for logographic languages (Chan et al., 1998), were part of larger intelligence testing sessions (Brandler & Rammsayer, 2003), used long-term memory measures (Franklin et al., 2008; Jakobson et al., 2008), or tested children only (Besson et al., 2007; Ho et al., 2003). At this point, there is not

enough evidence to support a prediction that adult musicians would show significantly better verbal serial recall than nonmusicians would.

The first hypothesis was that both musicians and nonmusicians would show phonological similarity effects, but that only nonmusicians would show effects of pitch proximity (Surprenant et al., 1993). The second hypothesis was that musicians would show significantly better tone recall, but not necessarily better verbal recall, than would nonmusicians.

### Method

**Participants.** Sixty-four University of York students completed the study for either course credit or £4. Thirty-two participants (16 male, 16 female) were self-reported nonmusicians, according to the criteria adopted by previous experiments. Their mean age was 21.97 ( $SD = 7.30$ ). None had taken part in any previous experiments from the present article. The other 32 participants (16 males and 16 females) were musicians (mean age = 21.78,  $SD = 4.09$ ). They had at least 8 years of musical training or regular practice, were able to read standard music notation, and reported that they did not possess absolute pitch. They reported a mean of 13.41 years training (range = 8–31 years).

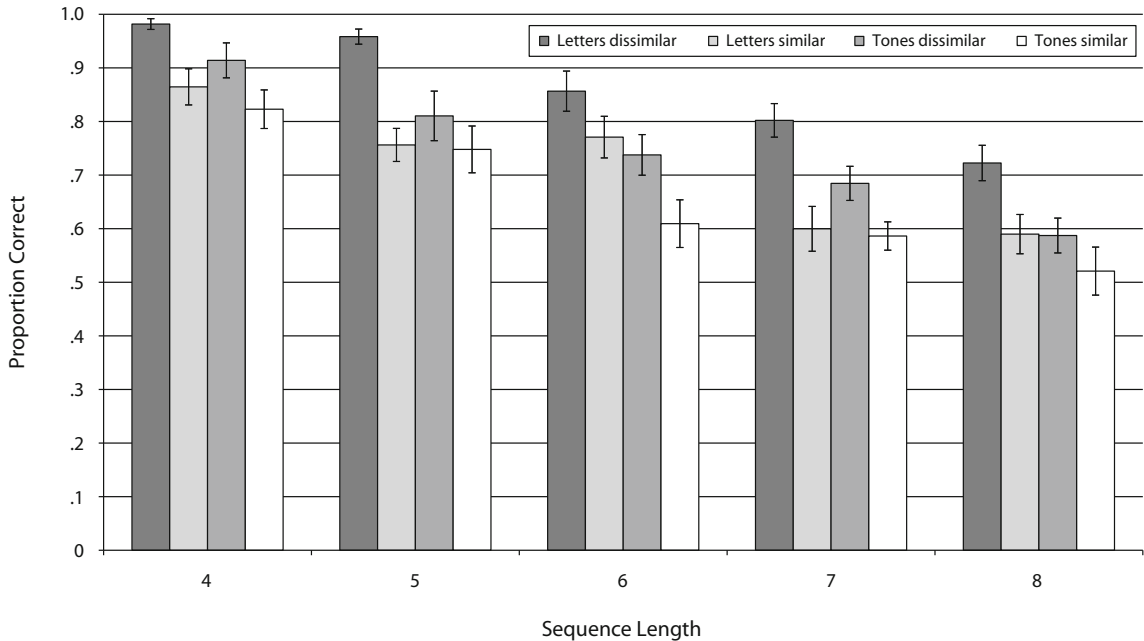
**Design.** We used a  $2 \times 2 \times 2 \times 5$  split-plot design. The first between-subjects variable was group (musicians vs. nonmusicians). The second between-subjects variable was stimulus type (verbal vs. musical). The verbal conditions used two sets of letters from the English alphabet, whereas the musical conditions used sets of tones from two different musical keys. The third within-subjects variable was similarity. The final within-subjects variable was sequence length (four to eight items). The dependent variable was the proportion of items correct at each sequence length for each condition.

**Materials.** In order to ensure generality of the findings to different stimuli (Clark, 1973), we used two separate stimulus sets for each condition. The first musical condition used tones from previous experiments (C4, D4, and E4 vs. C4, G4, and B4). The second used tones with the same intervals from B $\flat$ -major (B $\flat$ 3, C4, and D4 vs. B $\flat$ 3, F4, and A4). Tones and chords were generated by using the same procedure as was used in Experiment 2.

In the first verbal condition, the phonologically similar letters were B, D, and G and the dissimilar letters were M, Q, and R. The .wav files of the original letters from Experiment 2 were extended to exactly 800 msec using a time-stretch (high-precision) setting in Audition. A 200-msec gap of silence was added to the end of each file. In the second verbal condition, the phonologically similar letters were F, S, and X and the dissimilar letters were C, L, and Y. The female participant who spoke the letters in Experiment 2 recorded the additional files using the same procedure and equipment. These letters were then altered to match the length and amplitude of the other files. The average fundamental frequencies of the letters are presented in the Appendix.

We created two lists of sequences to be rotated across conditions. Seven sequences (six experimental and one practice) were generated at each sequence length by using the computer program and procedure designed for Experiment 2. For the verbal conditions, the response grid was adapted by placing the three letters in the left axis in alphabetical order. So in the case of condition B-D-G, “B” replaced “low,” “D” replaced “medium,” and “G” replaced “high.”

**Procedure.** Participants completed either two musical memory tests (pitch proximal and distal) or two verbal memory tests (phonologically similar and different). B-D-G was always paired with M-Q-R, and F-S-X was always paired with C-L-Y. In the musical condition, C4-D4-E4 was paired with C4-G4-B4 and B $\flat$ 3-C4-D4 was paired with B $\flat$ 3-F4-A4. The order of presentation of similarity condition was counterbalanced. The two lists of sequences were rotated across similarity condition. This meant that 16 participants were required in order to complete one design rotation.<sup>2</sup>



**Figure 4. Experiment 3: Nonmusicians' proportion correct performance for both sound types (verbal and musical) in acoustically dissimilar and similar conditions. Error bars represent SEMs.**

Pitch training followed the same procedure as that in Experiment 1. On each experimental trial, participants saw a fixation cross for 2 sec, followed by a musical chord. A C-major chord was played in the B-D-G and C-major conditions. A B-major chord was played in the F-S-X and B-major conditions. After a 2-sec silence, the to-be-remembered sequence was played; immediately afterward, participants completed the response grid. An increase in sequence length was preceded by a warning on screen. Finally, participants were asked about strategy use. Each session lasted 50 min.

## Results

Three nonmusicians failed to pass the proximal pitch training, so their data were replaced. Figure 4 (nonmusicians) and Figure 5 (musicians) illustrate the performance scores obtained.

Preliminary analysis indicated no significant differences between performance across the two levels of verbal conditions and musical conditions [ $F(3,56) = 0.57$ ,  $MS_e = 730.17$ , n.s.], so scores were collapsed. A 2 (group)  $\times$  2 (stimulus type)  $\times$  2 (similarity)  $\times$  5 (sequence length) ANOVA revealed a significant effect of similarity in the predicted direction [ $F(1,60) = 113.84$ ,  $MS_e = 105.29$ ,  $p < .001$ ,  $\eta_g^2 = .13$ ;  $M = 61.23$  vs.  $70.28$ ]. There was also a significant effect of sequence length [ $F(4,240) = 178.67$ ,  $MS_e = 65.71$ ,  $p < .001$ ,  $\eta_g^2 = .29$ ], indicating decreased performances at increased sequence lengths. Finally, there was a significant effect of group [ $F(1,60) = 10.28$ ,  $MS_e = 709.36$ ,  $p = .002$ ,  $\eta_g^2 = .08$ ], indicating better performances for musicians than for nonmusicians ( $M = 69.32$  vs.  $M = 62.60$ ). The main effect of stimulus type did not reach significance ( $F < 0.5$ ).

The interaction between group and stimulus type was significant [ $F(1,60) = 6.79$ ,  $MS_e = 703.36$ ,  $p = .01$ ,  $\eta_g^2 = .06$ ],

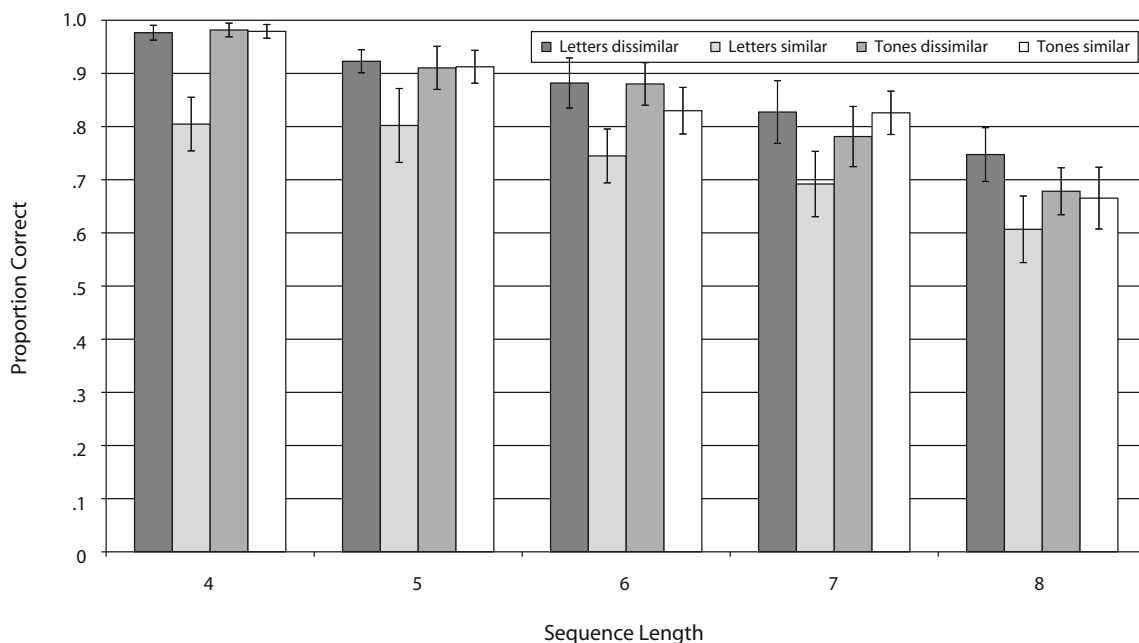
indicating that musicians performed better in tone recall but not letter recall. The group  $\times$  similarity  $\times$  stimulus type interaction approached significance [ $F(1,60) = 3.49$ ,  $MS_e = 105.59$ ,  $p = .06$ ,  $\eta_g^2 = .03$ ], suggesting that, although both groups were vulnerable to phonological similarity, only nonmusicians showed an effect of pitch proximity. Further analysis confirmed a significant phonological similarity effect for musicians [ $F(1,30) = 63.08$ ,  $MS_e = 1,392.60$ ,  $p < .001$ ,  $\eta_g^2 = .15$ ], but no significant effect of pitch proximity [ $F(1,30) = 0.39$ ,  $MS_e = 8.55$ , n.s.]. There was no interaction in the nonmusicians' data ( $F < 0.05$ ).

Finally the analysis indicated nonspecific variations in the size of the similarity effects at some sequence lengths for both groups [ $F(4,240) = 2.30$ ,  $MS_e = 66.76$ ,  $p = .05$ ,  $\eta_g^2 = .01$ ]. Wilcoxon tests corrected for multiple comparisons conducted on the nonmusicians' data indicated significant effects of phonological similarity at all sequence lengths except for six-item sequences [ $Z(16) = -1.91$ ,  $p = .06$ ] and eight-item sequences [ $Z(16) = -2.25$ ,  $p = .03$ ] and significant effects of pitch proximity at all lengths apart from five-item sequences [ $Z(16) = -2.23$ ,  $p = .03$ ] and eight-item sequences [ $Z(16) = -1.70$ ,  $p = .09$ ]. For the musicians there were significant effects of phonological similarity at all sequence lengths apart from five-item sequences [ $Z(16) = -2.07$ ,  $p = .04$ ] and eight-item sequences [ $Z(16) = -2.27$ ,  $p = .02$ ].

## Discussion

The results of this experiment suggest the effect of pitch proximity in nonmusicians' serial recall, which was found in the pilot and Experiment 1, was replicated using different sequence patterns and tones. These data





**Figure 5. Experiment 3: Musicians' proportion correct performance for both sound types (verbal and musical) in acoustically dissimilar and similar conditions. Error bars represent SEMs.**

suggest that acoustic similarity impacts upon nonmusicians' immediate serial recall of both verbal and tonal pitch materials. This correspondence suggests support for a degree of shared storage in short-term memory (Salamé & Baddeley, 1989; Semal et al., 1996). It is still possible to postulate the existence of separate stores for verbal and pitch sequences (Berz, 1995; Deutsch, 1970; Pechmann & Mohr, 1992) but with the caveat that they operate using similar principles (i.e., storage of pitch sounds).

The second finding was a significant phonological similarity effect for the verbal materials in all participants. There are a number of aspects to the new methodology that differ from standard serial-recall tasks, including the use of sequences containing repeating items from a limited pool and the visual grid response. Replication of the phonological similarity effect despite these alterations reflects the robustness of the effect. However, it remains to be established whether the reliability of the methodology is as strong when using other manipulations known to affect serial recall from short-term memory.<sup>3</sup>

An additional aim for this experiment was to assess the impact of musical expertise on memory performance. The first hypothesis was that musicians would perform better in tone recall. Direct support for the hypothesis was seen in the significant group by sound interaction. The nature of the experiment means that there is no way to distinguish between the different theories of musicians' superior performance (Chase & Ericsson, 1981; Ericsson & Kintsch, 1995; Williamon & Egner, 2004; Williamon & Valentine, 2002). However, now that the effect has been replicated, future work can analyze how perceptual abilities, encod-

ing techniques or long-term knowledge combine to influence performance.

The second hypothesis, no difference in group performances across verbal conditions, was also in line with the findings. However, there is a discrepancy with the results of some studies that have suggested musical training is associated with superior performance on verbal memory tasks (Besson et al., 2007; Brandler & Rammsayer, 2003; Chan et al., 1998; Ho et al., 2003; Jakobson et al., 2008). This discrepancy may be explained by the type of task administered. Alternatively, the association could be stronger in children than in adults (Besson et al., 2007; Ho et al., 2003) or could be present only in adult musicians who speak a logographic language (Chan et al., 1998). Another possibility is that the present experiment is lacking in the statistical power necessary to detect a difference, since only 16 members of each group completed a verbal task. More wide-ranging tests of musician's short-term memory skills using standardized tests are necessary before conclusions can be drawn about the association between musical ability and verbal short-term memory.

Finally, we found that musicians, unlike nonmusicians, showed no pitch-proximity effect. This finding supports the results of Surprenant et al. (1993). We went a step further in attempting to determine the reason for the lack of a pitch-proximity effect in musicians. All participants were asked about their strategy use post hoc. The relevant means are shown in Table 1. Musicians relied more on a combination of encoding techniques (either dual or multiple), including auditory (tone sound), verbal (labeling tones or contour patterns), and tactile (playing an imaginary instrument) encoding, whereas nonmusicians relied

**Table 1**  
**Experiment 3: Total Number of Post Hoc Reported Strategies From**  
**Both Groups for the Letter- and Tone-Recall Conditions**

Strategy	Letter Recall		Tone Recall	
	Musicians	Nonmusicians	Musicians	Nonmusicians
Single strategy	7 (7 verbal)	8 (7 verbal)	5 (2 musical)	10 (6 musical)
Two strategies	6	7	8	5
Multiple strategies	2	0	3	1

more heavily on a single strategy, usually of maintaining the sound of the tone sequences. Musicians' lower dependence on an auditory representation of music may explain their decreased vulnerability to the pitch-proximity effect. The implications for memory processing are presented in the General Discussion.

Another possible explanation for musicians' lack of pitch-proximity effect is based on tonality. The proximal and distant tones were selected according to their position in tonal hierarchy, in order to minimize the influence of tonality on recall. However, musicians show increased sensitivity to tonal hierarchy (Halpern, Kwak, Bartlett, & Dowling, 1996). Therefore, they may have found the two stimulus sets to be more equated for the purpose of recall. The nonmusicians conversely may have been more influenced by the proximity manipulation than tonality.<sup>4</sup> There is no way to avoid incurring tonality in the present experiment, because sequences created using three tones will always trigger a tonal center. A way to address this concern would be to create sequences in which pitch distance is held constant and tonal strength is manipulated. If musicians showed an effect of such a manipulation, it could be assumed that, in the present paradigm, musicians' serial recall is affected more by tonality than by pitch proximity.

In conclusion, Experiment 3 has provided evidence that musical expertise leads to improved performance on a tone-based, but not an equivalent verbal, immediate serial-recall task. Both musicians and nonmusicians showed a phonological similarity effect, indicating that storage of verbal items in memory is not influenced by music expertise. Nonmusicians showed a pitch-proximity effect, supporting the existence of shared processing or overlap in verbal and musical short-term memory. Finally, musicians were not vulnerable to the manipulation of pitch proximity. One explanation for this finding is group differences in memorization strategies.

## GENERAL DISCUSSION

Despite an increasing interest in the comparison of short-term memory for language and music, there is a lack of behavioral paradigms for directly comparing serial recall. The present study had a practical aim and a theoretical aim. The practical aim was to develop a method of serial recall that musicians and nonmusicians could use in order that we might compare verbal and musical memory. The theoretical aim was to test a musical equivalent of phonological similarity, a manipulation well known to impact on auditory-verbal immediate serial recall. Similarities across verbal and musical performance would provide an argument for similarity in memory processes.

We developed a new response method for testing immediate serial recall of letters and tones. The visual-recall grid makes no demands on either a musical response (singing or musical terminology) or verbal recoding.<sup>5</sup> The method for generating the tones was refined (prerecorded tones vs. artificially created .wav files), and a number of new constraints important to the construction of tone sequences were identified (balancing for tonal hierarchy, control of start and end item, and rotation of contour pattern).

We compared the effects of acoustic similarity upon serial recall of novel letter and tone sequences in nonmusicians. The phonological similarity effect was found in Experiment 3. The pitch-proximity effect was found in serial recall (Experiments 1 and 3) and recognition tasks (Experiment 2). These findings suggest a degree of overlap in the processing of musical and verbal sounds in short-term memory. Theoretical models, such as the multicomponent working-memory model (Baddeley & Hitch, 1974), may need to be adapted to include the processing of pitch sounds. There are alternative models that have the potential to explain some of the present findings, including those that make no clear distinction between the processing of speech and musical sounds in memory (e.g., feature model [Nairne, 1990; Neath, 2000], SIMPLE [Neath & Brown, 2006], and O-OER [D. M. Jones, 1993]). It is beyond the remit of the present article to derive and compare the predictions of these models. However, such comparisons could add further valuable debate to the nature of language-music processing in memory.

The effect of musical expertise was tested in Experiment 3. The effect of phonological similarity was consistent across musicians and nonmusicians, but the pitch-proximity effect was not found in musicians. These findings support and extend those of Surprenant et al. (1993). One possible explanation is that musical training results in changes to short-term memory for music, meaning that musicians are no longer vulnerable to a manipulation of acoustic similarity. This might include the development of a specialized storage system for musical materials (Pechmann & Mohr, 1992).

Another possibility mentioned in discussion is that musicians rely more on multidimensional codes to generate and maintain music (i.e., visual, auditory, and tactile). The episodic buffer provides a system wherein unitary multidimensional codes can be stored and accessed (Baddeley, 2007), as does the feature model (Neath, 2000). If musicians store multidimensional codes then impairing only one of their encoded representations (auditory), as happens in the pitch proximity paradigm, would have minimal effect on performance. Future work could seek to determine if musicians are generating and retaining multiple codes by

determining whether, in the present task, there is specific neural activation in areas of the brain associated with such codes (possible areas include the right prefrontal cortex, the temporoparietal junction and the posterior parietal cortex; Rudner, Fransson, Ingvar, Nyberg, & Rönnerberg, 2007).

Taking a wider view, how do the present findings contribute to the debate regarding language–music cognitive processing overlap? There is a great deal of evidence to support the existence of distinct networks in the brain devoted to language and music (Peretz & Zatorre, 2005). However, in the shared syntactic integration resource hypothesis (SSIRH), Patel (2003, 2008) draws a distinction between domain-specific knowledge for language and music, and shared cognitive operations. Patel argues that similar operations are required when analogous task demands arise, such as maintenance or integration of incoming, evolving sound patterns.

SSIRH has not yet been applied to short-term memory specifically, but it has many theoretical parallels to the working-memory model and specifically the operations of the phonological loop. Both theories postulate a dual-component framework that draws a theoretical distinction between the processes involved in storage of sound and operations that are carried out on that sound for the purposes of higher cognition (rehearsal in the case of the phonological loop and syntactic integration in SSIRH).

If Patel's concept of resource sharing is applicable to short-term memory, it is possible that there is a tonal store that does not overlap with the phonological store (Deutsch, 1970), but that there is an articulatory rehearsal process that is common to both. This idea is consistent with recent neuroimaging (Koelsch et al., 2009; Mandell, Schulze, & Schlaug, 2007) and behavioral evidence (Williamson, 2008). How might a theory of distinct stores and common rehearsal be reconciled with the findings of the present experiment, which might equally suggest overlap in storage in nonmusicians? Because of the similarities in the structure of language and music, there may be features that the two stores have in common (such as coding according to pitch), but storage itself could still be fundamentally separable. In summary, cognitive-resource-sharing theories may provide the key to understanding the relationship between language and music processing, both within short-term memory and as part of wider cognition. It is a premise that explains a good deal of the current data.

The present research provides both a methodological development and a set of results that support the value in comparing the processing of speech and tonal pitch materials in short-term memory. There is potential for a great deal to be gained by continuing to adapt established models of verbal memory to testing musical memory skills.

#### AUTHOR NOTE

V.J.W. was supported by a University of York PhD studentship and an ESRC postdoctoral fellowship. We thank Paul McLaughlin for his invaluable assistance in creating computer programs and Dave Moore for his Disklavier recordings. In addition, we are grateful to three anonymous reviewers and to Peter Bailey for their helpful comments on earlier drafts. Address correspondence to V.J. Williamson, Psychology Department, Goldsmiths, University of London, New Cross, London, SE14 6NW, England (e-mail: v.williamson@gold.ac.uk).

#### REFERENCES

- ANVARI, S. H., TRAINOR, L. J., WOODSIDE, J., & LEVY, B. A. (2002). Relations among musical skills, phonological processing, and early reading ability in preschool children. *Journal of Experimental Child Psychology*, **83**, 111-130. doi:10.1016/S0022-0965(02)00124-8
- AYOTTE, J., PERETZ, I., & HYDE, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, **125**, 238-251. doi:10.1093/brain/awf028
- BADDELEY, A. D. (1966). The influence of acoustic and semantic similarity on long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, **18**, 302-309.
- BADDELEY, A. [D.] (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, **4**, 417-423. doi:10.1016/S1364-6613(00)01538-2
- BADDELEY, A. [D.] (2007). *Working memory, thought, and action*. Oxford: Oxford University Press.
- BADDELEY, A. [D.], CHINCOTTA, D., & ADLAM, A. (2001). Working memory and the control of action: Evidence from task switching. *Journal of Experimental Psychology: General*, **130**, 641-657. doi:10.1037/0096-3445.130.4.641
- BADDELEY, A. [D.], GATHERCOLE, S. [E.], & PAPAGNO, C. (1998). The phonological loop as a language learning device. *Psychological Review*, **105**, 158-173.
- BADDELEY, A. D., & HITCH, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47-89). New York: Academic Press.
- BADDELEY, A. [D.], & LARSEN, J. D. (2003). The disruption of STM: A response to our commentators. *Quarterly Journal of Experimental Psychology*, **56A**, 1301-1306.
- BADDELEY, A. D., LEWIS, V., & VALLAR, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, **36A**, 233-252.
- BADDELEY, A. [D.], PAPAGNO, C., & VALLAR, G. (1988). When long-term learning depends on short-term storage. *Journal of Memory & Language*, **27**, 586-595. doi:10.1016/0749-596X(88)90028-9
- BERZ, W. L. (1995). Working memory in music: A theoretical model. *Music Perception*, **12**, 353-364.
- BESSON, M., SCHÖN, D., MORENO, S., SANTOS, A., & MAGNE, C. (2007). Influence of musical expertise and musical training on pitch processing in music and language. *Restorative Neurology & Neuroscience*, **25**, 399-410.
- BRANDLER, S., & RAMMSAYER, T. H. (2003). Differences in mental abilities between musicians and non-musicians. *Psychology of Music*, **31**, 123-138. doi:10.1177/0305735603031002290
- BROWN, S., & MARTINEZ, M. J. (2007). Activation of premotor vocal areas during musical discrimination. *Brain & Cognition*, **63**, 59-69. doi:10.1016/j.bandc.2006.08.006
- BROWN, S., MARTINEZ, M. J., HODGES, D. A., FOX, P. T., & PARSONS, L. M. (2004). The song system of the human brain. *Cognitive Brain Research*, **20**, 363-375. doi:10.1016/j.cogbrainres.2004.03.016
- BROWN, S., MARTINEZ, M. J., & PARSONS, L. M. (2006). Music and language side by side in the brain: A PET study of the generation of melodies and sentences. *European Journal of Neuroscience*, **23**, 2791-2803.
- CALLAN, D. E., TSYTSAREV, V., HANAKAWA, T., CALLAN, A. M., KATSUHARA, M., FUKUYAMA, H., & TURNER, R. (2006). Song and speech: Brain regions involved with perception and covert production. *NeuroImage*, **31**, 1327-1342. doi:10.1016/j.neuroimage.2006.01.036
- CHAN, A. S., HO, Y.-C., & CHEUNG, M.-C. (1998). Music training improves verbal memory. *Nature*, **396**, 128.
- CHASE, W. G., & ERICSSON, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141-189). Hillsdale, NJ: Erlbaum.
- CLARK, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, **12**, 335-359.
- CLARKE, E. F. (1988). Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music: The psychology of performance, improvisation, and composition* (pp. 1-26). Oxford: Oxford University Press, Clarendon Press.
- COLLE, H. A., & WELSH, A. (1976). Acoustic making in primary memory. *Journal of Verbal Learning & Verbal Behavior*, **15**, 17-31. doi:10.1016/S0022-5371(76)90003-7

- CONRAD, R. (1964). Acoustic confusions in immediate memory. *British Journal of Psychology*, **55**, 75-84.
- CONRAD, R., & HULL, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, **55**, 429-432.
- CROONEN, W. L. M. (1994). Effects of length, tonal structure, and contour in the recognition of tone series. *Perception & Psychophysics*, **55**, 623-632.
- DARWIN, C. (1871). *The descent of man and selection in relation to sex*. New York: Appleton.
- DEUTSCH, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, **168**, 1604-1605.
- DOWLING, W. J. (1991). Tonal strength and melody recognition after long and short delays. *Perception & Psychophysics*, **50**, 305-313.
- DOWLING, W. J. (1994). Melodic contour in hearing and remembering melodies. In R. Aiello & J. A. Sloboda (Eds.), *Musical perceptions* (pp. 173-190). New York: Oxford University Press.
- ERICSSON, K. A., & KINTSCH, W. (1995). Long-term working memory. *Psychological Review*, **102**, 211-245.
- FRANKLIN, M. S., MOORE, K. S., YIP, C.-Y., JONIDES, J., RATTRAY, K., & MOHER, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, **36**, 353-365. doi:10.1177/0305735607086044
- GAAB, N., & SCHLAUG, G. (2003). The effect of musicianship on pitch memory in performance matched groups. *NeuroReport*, **14**, 2291-2295.
- GAAB, N., TALLAL, P., KIM, H., LAKSHMINARAYANAN, K., ARCHIE, J. J., GLOVER, G. H., & GABRIELI, J. D. E. (2005). Neural correlates of rapid spectrotemporal processing in musicians and nonmusicians. In G. Avanzini, L. Lopez, & S. Koelsch (Eds.), *The neurosciences and music II: From perception to performance* (Annals of the New York Academy of Sciences, Vol. 1060, pp. 82-88). New York: New York Academy of Sciences.
- GATHERCOLE, S. E., & BADDELEY, A. D. (1990). Phonological memory deficits in language disordered children: Is there a causal connection? *Journal of Memory & Language*, **29**, 336-360. doi:10.1016/0749-596X(90)90004-J
- GREENE, R. L., & SAMUEL, A. G. (1986). Recency and suffix effects in serial recall of musical stimuli. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **12**, 517-524.
- HADLINGTON, L. [J.], BRIDGES, A. M., & DARBY, R. J. (2004). Auditory location in the irrelevant sound effect: The effects of presenting auditory stimuli to either the left ear, right ear or both ears. *Brain & Cognition*, **55**, 545-557. doi:10.1016/j.bandc.2004.04.001
- HALPERN, A. R., KWAK, S., BARTLETT, J. C., & DOWLING, W. J. (1996). Effects of aging and musical experience on the representation of tonal hierarchies. *Psychology & Aging*, **11**, 235-246.
- HANLEY, J. R., & BAKOPOULOU, E. (2003). Irrelevant speech, articulatory suppression, and phonological similarity: A test of the phonological loop model and the feature model. *Psychonomic Bulletin & Review*, **10**, 435-444.
- HENSON, R., HARTLEY, T., BURGESS, N., HITCH, G., & FLUDE, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *Quarterly Journal of Experimental Psychology*, **56A**, 1307-1334. doi:10.1080/02724980244000747
- HICKOK, G., BUCHSBAUM, B., HUMPHRIES, C., & MUFTULER, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in Area Spt. *Journal of Cognitive Neuroscience*, **15**, 673-682. doi:10.1162/jocn.2003.15.5.673
- HO, Y.-C., CHEUNG, M.-C., & CHAN, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal explorations in children. *Neuropsychology*, **17**, 439-450.
- JAKOBSON, L. S., LEWYCKY, S. T., KILGOUR, A. R., & STOESZ, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception*, **26**, 41-55. doi:10.1525/mp.2008.26.1.41
- JONES, D. [M.] (1993). Objects, streams, and threads of auditory attention. In A. [D.] Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness, and control: A tribute to Donald Broadbent* (pp. 87-104). Oxford: Oxford University Press.
- JONES, D. M., & MACKEN, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 369-381.
- JONES, D. M., MACKEN, W. J., & HARRIES, C. (1997). Disruption of short-term recognition memory for tones: Streaming or interference? *Quarterly Journal of Experimental Psychology*, **50A**, 337-357.
- JONES, D. M., MACKEN, W. J., & MURRAY, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: The role of segmentation. *Memory & Cognition*, **21**, 318-328.
- JONES, J. L., LUCKER, J., ZALEWSKI, C., BREWER, C., & DRAYNA, D. (2009). Phonological processing in adults with deficits in musical pitch recognition. *Journal of Communication Disorders*, **42**, 226-234. doi:10.1016/j.jcomdis.2009.01.001
- KELLER, T. A., COWAN, N., & SAULTS, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 635-645.
- KISHON-RABIN, L., AMIR, O., VEXLER, Y., & ZALTZ, Y. (2001). Pitch discrimination: Are professional musicians better than non-musicians? *Journal of Basic & Clinical Physiology & Pharmacology*, **12**, 125-143.
- KOELSCH, S., SCHULZE, K., SAMMLER, D., FRITZ, T., MÜLLER, K., & GRUBER, O. (2009). Functional architecture of verbal and tonal working memory: An fMRI study. *Human Brain Mapping*, **30**, 859-873. doi:10.1002/hbm.20550
- KRAUS, N., & BANAI, K. (2007). Auditory-processing malleability: Focus on language and music. *Current Directions in Psychological Science*, **16**, 105-110. doi:10.1111/j.1467-8721.2007.00485.x
- KRUMHANSL, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.
- LEVITIN, D. J. (2006). *This is your brain on music: The science of a human obsession*. New York: Dutton.
- LIEFOOGHE, B., BAROUILLET, P., VANDIERENDONCK, A., & CAMOS, V. (2008). Working memory costs of task switching. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 478-494. doi:10.1037/0278-7393.34.3.478
- LURIA, A. R., TSVETKOVA, L. S., & FUTER, D. S. (1965). Aphasia in a composer. *Journal of the Neurological Sciences*, **2**, 288-292.
- MACKEN, W. J., TREMBLAY, S., HOUGHTON, R. J., NICHOLLS, A. P., & JONES, D. M. (2003). Does auditory streaming require attention? Evidence from attentional selectivity in short-term memory. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 43-51.
- MACMILLAN, N. A., & CREELMAN, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah, NJ: Erlbaum.
- MANDELL, J., SCHULZE, K., & SCHLAUG, G. (2007). Congenital amusia: An auditory-motor feedback disorder? *Restorative Neurology & Neuroscience*, **25**, 323-334.
- MARCUS, G. F., VOULOUMANOS, A., & SAG, I. A. (2003). Does Broca's play by the rules? *Nature Neuroscience*, **6**, 651-652.
- MARIN, O. S. M., & PERRY, D. W. (1999). Neurological aspects of music perception and performance. In D. Deutsch (Ed.), *The psychology of music* (2nd ed.), (pp. 653-724). San Diego: Academic Press.
- MICHEYL, C., DELHOMMEAU, K., PERRON, X., & OXENHAM, A. J. (2006). Influence of musical and psychoacoustical training on pitch discrimination. *Hearing Research*, **219**, 36-47. doi:10.1016/j.heares.2006.05.004
- MIYAKE, A., FRIEDMAN, N. P., EMERSON, M. J., WITZKI, A. H., HOWERTER, A., & WAGER, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, **41**, 49-100. doi:10.1006/cogp.1999.0734
- MOORE, B. C. J. (2003). *An introduction to the psychology of hearing* (5th ed.). San Diego: Academic Press.
- NAIRNE, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, **18**, 251-269.
- NEATH, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, **7**, 403-423.
- NEATH, I., & BROWN, G. D. A. (2006). SIMPLE: Further applications of a local distinctiveness model of memory. In B. H. Ross (Ed.), *The psychology of learning and motivation*. (Vol. 46, pp. 201-243). San Diego: Academic Press.
- NIMMO, L. M., & ROODENRYS, S. (2005). The phonological similarity effect in serial recognition. *Memory*, **13**, 773-784. doi:10.1080/09658210444000386
- OVERY, K. (2003). Dyslexia and music: From timing deficits to musical intervention. In G. Avanzini, C. Faienza, L. Lopez, M. Majno,



- & D. Minciacchi (Eds.), *The neurosciences and music III: Disorders and plasticity* (Annals of the New York Academy of Sciences, Vol. 999, pp. 497-505). New York: New York Academy of Sciences. doi:10.1196/annals.1284.060
- PATEL, A. D. (2003). Language, music, syntax, and the brain. *Nature Neuroscience*, **6**, 674-681. doi:10.1038/nn1082
- PATEL, A. D. (2008). *Music, language, and the brain*. New York: Oxford University Press.
- PATEL, A. D. (2009). Language, music, and the brain: A resource-sharing framework. In P. Rebuschat, M. Rohrmeier, J. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems*. Oxford: Oxford University Press.
- PATEL, A. D., PERETZ, I., TRAMO, M., & LABREQUE, R. (1998). Processing prosodic and musical patterns: A neuropsychological investigation. *Brain & Language*, **61**, 123-144. doi:10.1006/brln.1997.1862
- PECHMANN, T., & MOHR, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition*, **20**, 314-320.
- PERETZ, I., & ZATORRE, R. J. (2005). Brain organization for music processing. *Annual Review of Psychology*, **56**, 89-114. doi:10.1146/annurev.psych.56.091103.070225
- ROBERTS, L. A. (1986). Modality and suffix effects in memory for melodic and harmonic musical materials. *Cognitive Psychology*, **18**, 123-157.
- RUDNER, M., FRANSSON, P., INGVAR, M., NYBERG, L., & RÖNNBERG, J. (2007). Neural representation of binding lexical signs and words in the episodic buffer of working memory. *Neuropsychologia*, **45**, 2258-2276. doi:10.1016/j.neuropsychologia.2007.02.017
- SALAMÉ, P., & BADDELEY, A. [D.] (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*, **24**, 263-265.
- SALAMÉ, P., & BADDELEY, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology*, **41A**, 107-122.
- SCHENDEL, Z. A., & PALMER, C. (2007). Suppression effects on musical and verbal memory. *Memory & Cognition*, **35**, 640-650.
- SCHLITMEIER, S. J., HELLBRÜCK, J., & KLATTE, M. (2008). Does irrelevant music cause an irrelevant sound effect for auditory items? *European Journal of Cognitive Psychology*, **20**, 252-271.
- SEMAL, C., & DEMANY, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America*, **89**, 2404-2410. doi:10.1121/1.400928
- SEMAL, C., & DEMANY, L. (1993). Further evidence for an autonomous processing of pitch in auditory short-term memory. *Journal of the Acoustical Society of America*, **94**, 1315-1322. doi:10.1121/1.408159
- SEMAL, C., DEMANY, L., UEDA, K., & HALLÉ, P.-A. (1996). Speech versus nonspeech in pitch memory. *Journal of the Acoustical Society of America*, **100**, 1132-1140. doi:10.1121/1.416298
- SLOBODA, J. A., & PARKER, D. H. H. (1985). Immediate recall of melodies. In P. Howell, I. Cross & R. West (Eds.), *Musical structure and cognition* (pp. 143-168). London: Academic Press.
- STEWART, L., VON KRIEGSTEIN, K., WARREN, J. D., & GRIFFITHS, T. D. (2006). Music and the brain: Disorders of musical listening. *Brain*, **129**, 2533-2553. doi:10.1093/brain/aw1171
- SURPRENANT, A. M., NEATH, I., & LECOMPTE, D. C. (1999). Irrelevant speech, phonological similarity, and presentation modality. *Memory*, **7**, 405-420.
- SURPRENANT, A. M., PITT, M. A., & CROWDER, R. G. (1993). Auditory recency in immediate memory. *Quarterly Journal of Experimental Psychology*, **46A**, 193-223.
- WILLIAMSON, A., & EGNER, T. (2004). Memory structures for encoding and retrieving a piece of music: An ERP investigation. *Cognitive Brain Research*, **22**, 36-44. doi:10.1016/j.cogbrainres.2004.05.012
- WILLIAMSON, A., & VALENTINE, E. (2002). The role of retrieval structures in memorizing music. *Cognitive Psychology*, **44**, 1-32. doi:10.1006/cogp.2001.0759
- WILLIAMSON, V. J. (2008). *Comparing short-term memory for sequences of verbal and tonal materials*. Unpublished PhD thesis, University of York.
- WONG, P. C. M., SKOE, E., RUSSO, N. M., DEES, T., & KRAUS, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience*, **10**, 420-422. doi:10.1038/nn1872

## NOTES

1. For example, the sequence C, E, G, E, G, E, C (1, 2, 3, 2, 3, 2, 1) could be altered to C, E, G, C, G, E, C (1, 2, 3, 1, 3, 2, 1). Despite the alteration of one tone, the overall patterns of ups and downs (contour) remain the same.
2. Each of the four stimulus type tests was experienced by 4 participants, 2 with the similar item first and 2 with the dissimilar item first. One person heard Pattern 1 first, and 1 person heard Pattern 2 first.
3. Although main effects were replicated, there was no consistent evidence from the present paradigm to suggest that the phonological similarity or pitch-proximity effects were absent at longer sequence lengths.
4. We are grateful to one reviewer for highlighting this alternative interpretation of our findings.
5. Although the strategy data from Experiment 3 indicated that some participants did choose to use a verbal code, at least in part, this coding strategy is not necessary to use the grid response.

## APPENDIX

Average Fundamental Frequencies ( $f_0$ ) for the Letter Stimuli Used in Experiments 2 and 3

Stimulus Condition	Acoustic Similarity	Letter	Mean $f_0$ (Hz)
Language 1	Similar	B	171.89
		D	173.70
		G	169.49
Language 2	Similar	F	178.07
		S	174.54
		X	180.73
Language 1	Dissimilar	M	169.54
		Q	173.56
		R	160.05
Language 2	Dissimilar	C	183.13
		L	177.19
		Y	175.26

Note—The  $f_0$  figures were obtained using Praat software.