Auditory and visual short-term memory: Influence of material type, contour, and musical expertise

Francesca Talamini^{1,3,4}, Salomé Blain¹, Jérémie Ginzburg¹, Olivier Houix², Patrick Bouchet¹, Massimo Grassi³, Barbara Tillmann^{1*}, and Anne Caclin^{1*}

¹ Lyon Neuroscience Research Center (CRNL), INSERM U1028, CNRS UMR 5292, Université Claude Bernard Lyon 1, Université de Lyon, Lyon, France

² Ircam, CNRS, Sorbonne Université, Ministère de la Culture, Sciences et Technologies de la Musique et du son, STMS, F-75004 Paris, France

³ Università degli Studi di Padova - Dipartimento di Psicologia Generale. Italy

⁴ Institute für Psychologie, Universität Innsbruck. Austria

*The authors share the last authorship

Corresponding author: Francesca Talamini, e-mail: Francesca.Talamini@uibk.ac.at

Abstract

Short-term memory has mostly been investigated with verbal or visuospatial stimuli and less so with other categories of stimuli. Moreover, the influence of sensory modality has been explored almost solely in the verbal domain. The present study used the same experimental paradigm to investigate auditory and visual short-term memory for different types of stimuli. In each trial, participants were presented with two sequences of events, separated by a silent delay, and had to indicate whether the two sequences were identical or different. Performance in this recognition (delayed-matching-to-sample) paradigm was compared for materials that were either verbal (i.e., chained syllables without meaning) or nonverbal (i.e., not easily described by verbal labels). For the latter ones, the event sequence could either entail a contour, which is a pattern of up and down changes (based on non-pitch features), or not. All materials were implemented in both auditory and visual modalities. As previous research has reported better auditory memory (and to some extent, visual memory), and better auditory contour recognition for musicians than non-musicians, the recognition tasks were performed by a group of musicians and a group of non-musicians. Results revealed a selective advantage of musicians for the auditory no-contour stimuli and for the contour stimuli (both visual and auditory), suggesting that musical expertise is associated with specific short-term memory advantages in domains close to the trained domain, even extending cross-modally. These findings offer new insights into the role of encoding strategies and their effect on short-term memory performance across modalities.

Keywords: short-term memory; working memory; auditory and visual modalities; cross-modal comparisons; memory encoding; music expertise.

1. Introduction

For memory, different functional systems have been distinguished depending on the duration of the memory traces (e.g., short or long, see Cowan, 2008). Here, we focus on short-term memory¹, that is the active maintenance of information for a short period of time, without manipulation of this information (maintenance with manipulation is classically defined as working memory; Baddeley & Hitch, 1974). Short-term memory has been usually investigated with tasks that first present a set of items, and then participants have to either recall or recognize the previously presented information. How does short-term memory work? Are there different processes and “storages” for different types of the to-be-remembered information and depending on the sensory modality (e.g., visual, auditory)?

Short-term memory has been mostly investigated for verbal stimuli (e.g., words, numbers, syllables) and visual stimuli (e.g., figures, spatial positions). The influential memory model of Baddeley and Hitch (1974) proposed three different components, with two temporary storages depending on the material type, and a central executive. The two storages are the phonological loop, which stores verbal information, and the visuospatial sketchpad, which stores visuospatial information. The central executive is responsible for controlling and distributing attentional resources. Later on, Baddeley added a further component, the episodic buffer, which integrates multidimensional information and links working memory to long-term memory (Baddeley, 2000). The distinction between verbal storage and visuospatial storage made by the model is supported by numerous behavioral and neuroimaging data (Logie, Zucco, & Baddeley, 1990; Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Smith, Jonides, & Koeppel, 1996; Rottschy et al., 2012; Baldo & Dronkers, 2006; Harel et al., 2017).

Beyond verbal material, auditory short-term memory has been investigated more recently with tonal material (e.g., chords, melodies), and two separate subsystems have been proposed: a phonological loop for verbal information and a tonal loop for music information (Berz, 1995; Schulze, Zysset, Mueller, Friederici, & Koelsch, 2011; see also Caclin & Tillmann, 2018 for a discussion). Evidence for two separate systems has also been provided by studies investigating congenital amusia, which show impaired short-term memory for tone and timbre sequences, but intact short-term memory for spoken syllable sequences (Tillmann, Lvque, Fornoni, Albouy, & Caclin, 2016; Tillmann, Schulze, & Foxton, 2009, Albouy, Schulze, Caclin, & Tillmann, 2013; Williamson & Stewart, 2010). Interestingly, for the non-amusic control participants, better performance was observed for tone sequences than for verbal

¹From a theoretical perspective, here we will discuss also working memory models, because the separation of short-term memory and working memory for some authors is not clear and working memory models can also apply to the functioning of short-term memory.

sequences (see also Schulze & Tillmann, 2013). This observation might be due to a salient aspect of tone sequences, the pitch contour, which is the pattern of up and down pitch variations regardless of the exact pitch-value of each individual tone and the exact size of the defined intervals. These patterns can include, for example, rises (a pitch is higher than the preceding one), falls (a pitch is lower than the preceding one), or no change, that is a repeated tone. In short-term memory recognition tasks, different trials can be defined with changes that either keep constant the contour or not. Some studies have reported that it is easier to distinguish melodies when there is a contour change than when contour is preserved (e.g., Dowling & Fushitani, 1971; Dowling & Hollombe, 1977; Idson & Massaro, 1978; Peretz & Babai, 1992, Tillmann et al., 2009). Recent studies have reported that contour is a privileged feature for processing and storing auditory information, not only with pitch variations, but also with loudness and brightness variations (McDermott, Lehr, & Oxenham, 2008; Graves et al., 2019).

Beyond verbal, visuospatial, and tonal materials, only few studies have investigated memory for other categories of stimuli, through recognition tasks. In the visual modality, some studies have used 1) abstract images or images that are not associated with a semantic representation (Hirel et al., 2017; Tallon-Baudry, Bertrand, Peronnet, & Pernier, 1998; Tallon-Baudry, Bertrand, & Fisher, 2001; Cohen, Evans, Horowitz, & Wolfe, 2011) and 2) visual luminance variations (Aizenman, Gold, & Sekuler, 2017), and in the auditory modality 1) nonmusical sounds (such as environmental sounds, Thompson & Paivio, 1994; Cohen et al., 2011), 2) loudness variations, and 3) brightness variations (McDermott et al., 2008; Graves et al., in press). Short-term memory has been partly investigated also for other sensory modalities, such as olfaction or touch (e.g., Gallace, Tan, Haggard, & Spence, 2008; Herz & Engen, 1996).

Previous research comparing short-term memory for different types of stimuli has mainly focused on content (e.g., verbal vs. visuospatial, verbal vs. musical) rather than modality (e.g., visual vs. auditory). Sensory modality has been taken into account in some theoretical models, such as in the memory model of Atkinson and Shiffrin (1968). This model proposes a sensory memory buffer, where modality-specific information of the input is held active for a few seconds, and then this information is transferred to a short-term memory storage, where it can lose its modality specificity (e.g., auditory information can be retained as an image). A recent review by Christophel and colleagues (2017) presents evidence from neuroimaging studies suggesting that working memory is more likely to be based on a distributed network (and this can apply to short-term memory too). According to the authors, there is not a unique working memory storage in a single brain region or specific brain regions, but sensory materials are first processed and stored by specific sensory brain areas (e.g., auditory cortex for auditory stimuli,

Gottlieb, Yaadia, & Abeles, 1989), and then, less specialized brain areas are involved for more abstract representations (e.g., prefrontal cortex, Spitzer, Goltz, Wacker, Auksztulewicz, & Blankenburg, 2014).

Despite the attention some models gave to sensory modality, most of studies that carried on systematic modality comparisons (i.e., visual vs. auditory) focused mainly on verbal stimuli only (e.g., Murdock, 1966; Corballis, 1966; Murdock, 1967; Conrad & Hull, 1968; Drewnowski, & Murdock, 1980; Maylor, Vousden, & Brown, 1999; Frankish, 2008; Macken, Taylor, Kozlov, Hughes, & Jones, 2016) making it difficult to generalize their results. In serial recall tasks, there is evidence that auditory presentation of verbal stimuli leads to greater recency effects (i.e., better memory for last items of a list) than visual presentation, whereas the latter leads to greater mid-list effect (i.e., better memory for central items of a list) than auditory presentation (see Macken et al., 2016). This would suggest that indeed different processes and/or stores come into play depending on the presentation modality. Some authors argue that these auditory recency effects depend on a pre-categorical acoustical storage (PAS, Crowder & Morton, 1969), a temporary store where the properties of an auditory object are stored for few seconds, and replaced by new-coming auditory stimuli. The durability of the mnemonic trace would be stronger in this store in comparison to the pre-categorical visual store. Other authors claim that representations of auditory stimuli are richer than visual ones, and that visual stimuli are dominated more strongly by modality-independent features (e.g., semantic representations) than are auditory stimuli (Nairne, 1990). More recently, other authors claimed that the visual and auditory differences depend on the way the object representation is formed: auditory item-lists are more likely to be combined into a single object (especially when the items have similar auditory features), than visual item-lists (Macken et al., 2016). Modality differences in verbal memory were also observed at a cerebral level (e.g., different brain regions in the left hemisphere were activated for visual vs. auditory digits during an n-back task, Crottaz-Herbette, Anagnoson, & Menon, 2004).

Beyond these studies suggesting modality effects for verbal memory, only few studies investigated modality effects with non-verbal stimuli. Lehnert and Zimmer (2006) compared spatial short-term memory across modalities. The authors hypothesized that if there were separate systems depending on the modality, the combined presentation of visual and auditory spatial stimuli would enhance memory performance. As this was not observed in their study, the authors concluded that memory for location is stored in a unique system, regardless of sensory modality. Tremblay and colleagues (2006) also investigated spatial immediate serial recall for visual and auditory stimuli, found a greater recency effect for auditory spatial stimuli than visual ones, in line with the modality effects found with verbal stimuli (Tremblay, Parmentier, Guérard, Nicholls, & Jones, 2006). However, other studies with musical stimuli

(i.e., sequences of chords and tones, either played or displayed on a staff) did not find a modality-specific recency effect (Roberts, 1986), or for notes, they reported a recency effect similar to the auditory verbal one, but without a comparable visual control task (Greene & Samuel, 1986), so there is no evidence for modality-specificity yet. To sum up, sensory modality effects in short-term memory remain to be investigated for other-than-verbal categories. Moreover, the effect of modality has been mostly investigated with recall tasks, but less is known about these effects in recognition tasks.

A way of investigating whether there are common and/or separate processes in memory for various types of information (e.g., auditory vs. visual; verbal vs. nonverbal), can be via the comparison of different populations with and without specific expertise (as expertise can be associated with specific enhanced memory processes). Musicians represent potential candidates for this goal, as they are a class of experts that undergo a long and intense training, which requires mostly auditory processing and, to a lesser extent, visual processing, as well as the integration of crossmodal information and related motor responses. Likely because of the peculiarity of their training, an advantage of musicians over non-musicians has been observed for music-related tasks (e.g., frequency discrimination, Spiegel & Watson, 1984), and other tasks that are not directly linked to the trained domain (i.e., far transfer). For example, musicians have a memory advantage over non-musicians not only for musical stimuli (e.g., Pallesen et al., 2010; Williamson, Baddeley, & Hitch, 2010; Schulze, Dowling, & Tillmann, 2012), but also for verbal stimuli, mostly when delivered auditorily (e.g., Hansen, Wallentin, & Vuust, 2013; Suárez, Elangovan, & Au, 2016; Talamini, Carretti & Grassi, 2016) and to some extent, for nonverbal visual stimuli (e.g., George & Coch, 2011; Bidelman, Hutka, & Moreno, 2013; Yang, Lu, Gong, & Yao, 2016). A recent meta-analysis has highlighted that the difference in memory performance between musicians and non-musicians is consistent for tonal stimuli, where effect-sizes are large, and for verbal stimuli (whether auditory or visual), where effect sizes are moderate, both in short-term memory and working memory (Talamini, Altoè, Carretti, & Grassi, 2017). However, it is not clear whether the advantage of musicians over non-musicians in verbal memory is linked to a general superior auditory memory capacity in musicians. In fact, most of verbal memory tasks administered to musicians and non-musicians have been presented auditorily, but musicians might be able to create a stronger mnemonic trace thanks to their enhanced auditory perception skills (e.g., Spiegel & Watson, 1984, Tervaniemi, Just, Koelsch, Widmann, & Schröger, 2005). For memory of visuospatial stimuli, the meta-analysis showed that the effect size ranges from small (i.e., small difference between musicians and nonmusicians, with musicians performing better) to null (i.e., no difference between groups) (Talamini et al., 2017).

Our present study aimed to compare visual and auditory short-term memory for different categories of stimuli, namely nonverbal stimuli with and without contour information, and verbal stimuli, in both musicians and non-musicians. For the auditory modality, nonverbal stimuli with contour information were created with loudness variations, and not with pitch variation (to overcome the superiority of musicians in pitch discrimination abilities). For the visual modality, nonverbal stimuli with contour information were created with luminance variations (i.e., darker vs brighter), based on the crossmodal correspondence with loudness stimuli (Marks, 1987). Note that luminance variations have been previously used in short-term memory studies (Aizenman et al., 2017; Gold, Aizenman, Bond, & Sekuler, 2014). In our study, we included also contour conditions because it has been reported that musicians process pitch contour more automatically and more easily than non-musicians (Fujioka, Trainor, Ross, Kakigi, & Pantev, 2004), but less is known on how musicians perform in non-pitch contour recognition, both in the auditory and visual modality (in comparison to nonmusicians). For the no-contour stimuli, we designed nonverbal stimuli (i.e., to which it is difficult to attach a verbal label), where contour cues were as minimal as possible.

To optimize the comparison between sensory modalities and among different categories of stimuli, we used a short-term memory recognition task rather than a recall task, which is not easily adaptable to nonverbal material. In the recognition task, two sequences of items were presented, and participants had to judge whether the second sequence was the same or different from the first one. For different trials, two versions were implemented: changes in the second sequence were elicited by either replacing one of the items with a new one, or by exchanging the serial position of two (or more) items (referred to as item change vs. order change, respectively). These two versions of the recognition paradigm might tap into slightly different mechanisms, in particular regarding serial order processing. In fact, some studies (either using recall and recognition tasks) have suggested different processes for item and serial order memory (Gorin, Kowialiewski, & Majerus, 2016; Majerus, Poncelet, Greffe, & Van der Linden, 2006; Martinez-Perez, Majerus, Mahot, & Poncelet, 2012). For contour stimuli, the order of the items could take a greater importance as contour is defined by serial relationships between the items. We thus compared here the two versions of the recognition paradigm (item vs. order changes).

In sum, we presented short-term memory recognition tasks to a group of highly-trained musicians and a group of non-musicians, with the aims of comparing (1) visual and auditory short-term memory for different types of stimuli, both verbal and nonverbal, with and without contour information; (2) performance between musicians and non-musicians, as musicians have previously shown a selective advantage over non-musicians in some short-term memory tasks (e.g., verbal spans, music recognition);

and (3) two versions of the recognition task, with possibly different reliance on serial order and item memory. To the best of our knowledge, the present study incorporates the most complete design tested so far to compare types of stimuli and modality of presentation in a short-term memory investigation.

2. Method

2.1 Participants

Forty-eight French speakers participated in the study. Demographic details are reported in Table 1. They were between 18 and 30 years old, had normal or corrected-to-normal vision, had normal peripheral hearing (thresholds below 20 dB HL for 250, 500, 1000, 2000, 4000, and 6000 Hz, assessed with standard audiometry) and did not report any kind of neurological or psychiatric disorders (assessed with self-report questionnaires). Participants were 24 non-musicians and 24 musicians. They were all right-handed except for three non-musicians who were left-handed. Musicians were professional players or students at the music conservatory of Lyon and Villeurbanne (France), with a minimum training of 10 years. Seven musicians self-reported to have perfect pitch for every instrument.

Table 1

Age, education, music education, and weekly musical practice for the two participant groups.

	Musicians <i>N</i> =24 (18 women)	Non-musicians <i>N</i> =24 (19 women)	p-value (group comparison)
Age (yrs.)	23.3 (3.62)	23.7 (2.87)	$p = .693$
Education (yrs.)	15.2 (2.19)	15.3 (1.02)	$p = .235$
Musical training (yrs.)	16.2 (4.3)	2.2 (0.96)	$p < .001$
Music Weekly Practice (hrs)	23.9 (14.16)	/	

Note. Mean (SD). The last column represents the p-value of the two-sided independent t-test used to compare the two groups.

2.2 Neuropsychological testing

2.2.1 WAIS-IV (Wechsler Adult Intelligence Scale, Wechsler, 2008): Working Memory tests

To test for working memory capacity, we administered the two working memory tests of the WAIS-IV: the digit span of the WAIS-IV with its three subtests, and the arithmetic test.

Digit span subtests: 1) In the forward digit span, participants were asked to repeat a list of digits that was read by the experimenter at a pace of 1 digit per second, in the same order of presentation (e.g., the sequence “3 – 1 – 5”, had to be repeated as “3 – 1 – 5”). There were two trials of each length, starting with two digits, and then the list increased by one digit every two trials. When participants made two mistakes in two trials of the same length, the task ended. 2) The backward digit span followed the same procedure as the forward span, but participants were required to repeat the digits in the reverse order with respect to their presentation (e.g., the sequence “3 – 1 – 5”, had to be repeated as “5 – 1 – 3”). 3) In the digit span sequencing, participants were required to repeat the presented digits starting from the smallest number and proceeding in ascending order (e.g., the sequence “3 – 1 – 5”, had to be repeated as “1 – 3 – 5”). For each subtest, each correct response scored one point, and the total score was composed by the sum of correct responses.

Arithmetic subtest: Participants listened to a series of arithmetic problems that were read by the experimenter, one at a time, and they had to solve mental operations starting from easy ones and proceeding with more difficult ones (e.g., one simple problem could be the following: “John has six balls and he loses three balls. How many balls has John left?”). When participants made three mistakes in a row, the task ended. Each correct response scored one point, and the total score was composed by the sum of correct responses.

2.2.2 WAIS-IV: Speed of processing tests

Processing speed tests of the WAIS-IV were included as the ability of quickly processing new information is usually related to how efficient a person is in learning quickly and accurately a new task (that in our case would be the short-term memory recognition task) (Lichtenberger, & Kaufman, 2012; Fry & Hale, 1996).

Symbol Search: Participants had to compare whether two symbols depicted on the left side of a sheet of paper were included in a list of symbols on the right side of the sheet. If one of the two symbols was present, participants had to mark it, otherwise, they had to mark “no”. There were 10 trials per page, and six pages in total, and participants had to perform the task as fast as possible, within two minutes. The score was calculated by adding up all correct responses and by subtracting all incorrect responses.

Coding: In this timed task, participants saw a reference grid that contained a series of numbers, each one associated to a nonsense symbol. There were nine numbers and nine corresponding symbols. Participants had to fill up a series of empty boxes below a series of numbers, with the respective symbol, as indicated in the reference grid, which was always present. This had to be done as fast as possible, within two minutes. The score was calculated by counting all the symbols correctly reported in the boxes.

2.3 Short-term memory recognition tasks

We designed 12 short-term memory recognition tasks, by manipulating the factors Modality (auditory or visual), Type of Stimulus (verbal, non-verbal with contour, or non-verbal without contour) and Change (item or order changes). Examples of stimuli are depicted in Fig. 1.

2.3.1 Auditory stimuli. All auditory items were 500 ms long, and were presented in sequences with silent 100 ms ISI (Inter-Stimulus Interval). A given item was never repeated within a sequence.

AUDITORY VERBAL: Five-item sequences were created from a pool of six syllables: /pu/, /mu/, /bu/, /lu/, /gu/, /tu/. These syllables were adapted (i.e., shortening some of the vowels) from Tillmann et al. (2009) to sound more isochronous when presented in the sequence. All syllables had a fundamental frequency of 232 Hz, and were presented at ~67 dBA.

AUDITORY CONTOUR: Five-item sequences were created from a pool of six instances of one tone (with a fundamental frequency of 232 Hz) played with a piano timbre and presented with different loudness levels and edited with the Cool Edit Pro software. The six piano tones were presented from 55dBA (the softest tone) to 80 dBA (the loudest tone) in 5 dB steps.

AUDITORY NO-CONTOUR: Four-item sequences were created from a pool of five noises, which were created from a pink noise (generated with Cool Edit Pro software) modified with the “particle” module of the Cecilia software (Ajax Sound Studio, 2017). This manipulation allowed changing the texture of the noise with a granulator synthesis method (Roads, 1988), and to create noise stimuli that can be discriminable, but without clear contour variations. Noises were presented at ~67 dBA.

2.3.2 Visual stimuli. All visual items (except for the visual contour items) were 500 ms long, and were presented in sequences with 100 ms blank ISI. In visual contour sequences, each item was presented for 600 ms and there was no ISI, to avoid a flickering effect that could disturb luminance perception. A given item was never repeated within a sequence.

VISUAL VERBAL: Five-item sequences were created from the same pool of six syllables, as presented in the auditory modality, and spelled as follows: “pou”, “mou”, “bou”, “lou”, “gou”, “tou”. The syllables were presented in black colour on a white background, in five different positions on the horizontal axis of the computer screen, from left to right, and always centrally in the vertical axis. Note that here the choice of presenting the syllables in different positions on the screen was made because

presenting them at the centre might have invited subjects to look only at the first letter (as the vowels were always the same).

VISUAL CONTOUR: Five-item sequences were created from a pool of six squares with different grades of luminance in steps of 5 cd/m². The dimmest level had a luminance of 2.26 cd/m². The square size was 490 by 490 pixels, always presented at the centre of the screen, on a black background.

VISUAL NO-CONTOUR: Four-item sequences were created from a pool of five different Japanese ideograms (i.e., kanji). We chose to have items that were belonging to the same category (i.e., ideograms) to avoid as much as possible semantic and/or shape-related labelling. The number of elements (i.e., horizontal, vertical, and oblique lines) in each item was also equalized (e.g., by cutting or adding some elements with the software Inkscape) to minimize differences among stimuli. Stimuli were presented in four different positions on the horizontal axis of the computer screen, from left to right, and always centrally in the vertical axis, to avoid focusing on a single feature instead that on the whole item (e.g., focusing on the upper line would have produced a different “height” feeling that could have reminded contour, see stimuli in Fig. 1).

2.3.3 Task design. On each trial, a sequence of items (from now on, S1) was followed by a second sequence (from now on, S2) after a 2 sec. delay. S2 was either the same (50% of trials) or different from S1 (50%). S2 was different in two ways (presented in separate blocks): the order of the items of S1 were changed (order change) or a new element replaced one of the items of S1 (item change). In the order change, adjacent items of S1 were inverted to create S2, and this occurred equally likely in every position of the sequence (e.g., 25% of the changes were made by switching position 1 and 2, 25% by switching position 2 and 3, and so on with 3 and 4, and 4 and 5). In the item change, the position in which the new item replaced one of the items of S1 was equally distributed (e.g., 1/5 of the changes were made in position 1, 1/5 in position 2, and so on).

For verbal and contour stimuli, we created five-item sequences by using six items. All sequence structures were unique, and they could be used for verbal and contour stimuli (in different participants, e.g., the sequences’ structures were used for the verbal stimuli for a participant, and for the contour stimuli for another participant). For contour stimuli, each S1 contained at least one ascending interval and one descending one, and when S2 was different from S1 there was always a change in contour. For no-contour stimuli, we created the structures (all unique) of four-item sequences because pilot testing revealed high difficulty in performing the task with 5-item sequences for these stimuli. All prepared sequences’ structures were distributed into eight sets (separately for five-item and four-item sequences), that were counterbalanced across participants for visual and auditory modalities, for item and order

changes and for same and different trials. The counterbalancing was done separately within each group (i.e., in musicians and nonmusicians), and it was the same for both groups.

Fig. 1. Examples of sequences for the different stimuli. Indicated ISI lasted 100 ms and was set to 0 for visual contour stimuli. Left: Representation of the waveforms of the auditory stimuli. Right: Representation of the visual stimuli; in the verbal and no-contour condition, the syllables and the kanji ideograms were appearing from left to right, one at a time, and in the contour condition, the square was always presented at the centre of the screen.

2.4 Apparatus

The visual stimuli were presented on a 17" monitor (Compaq 9500) with a resolution of 1280*1024 pixels (Graphic Card NVIDIA Geforce GTX 750) and the distance between the participant and the screen was ~75cm. All auditory stimuli had a 44,100-Hz sampling rate with a 32-bit resolution. The audio card was a SPDIF out Creative SB X Ti. The used headphones were a pair of Sennheiser HD 280 pro. Presentation software (Neurobehavioral systems, Albany, CA, USA) was used to present the stimuli and record participants' responses.

2.5 Procedure

Participants first completed a set of questionnaires about 1) demographic details, such as age, sex, education, profession, handedness, presence of hearing or vision problems, 2) potential musical training (e.g., if they played, sang or danced, for how long, starting age, practice hours, whether they underwent specific training, they had perfect pitch and subjective reports about their musical ability and about music in general). Then, participants underwent an audiogram screening in a sound-attenuated small room, where they also performed the computerized short-term memory tasks. Before beginning the task, instructions for the short-term memory task were given, and each single visual or auditory item that could be presented in the task was shown or played to participants.

The six types of stimuli were presented in separate blocks. Each experimental block was composed of 40 trials (for the verbal and contour stimuli) and of 48 trials (for the no-contour stimuli). The number of trials was larger for the no-contour stimuli to complete the counterbalancing with the four-item sequences. In each block, there were 50% of same trials (i.e., $S1 = S2$) and 50% of different trials (i.e., $S1 \neq S2$). On each trial, participants saw/heard $S1$ and $S2$, and after the presentation of $S2$, they had to judge by clicking on the mouse whether $S2$ was the same as $S1$ or different. Participants completed a short training block (4 trials) before each block. Participants were first tested with six blocks for one of the two change types (item or order), followed by six blocks for the other change type (with the same blocks' order as for the first part). Participants were informed on the type of change in the task before starting the blocks. Each part lasted for about 45 minutes. In the middle, participants were taking a break, and the WAIS-IV subtests were administered (duration of about 30 minutes). Half of the participants were starting with the order change tasks, and the other half with the item change tasks. The order of the 6 blocks of each task was counterbalanced across participants using a Latin square controlling for first-order carry-over effects. In each block, there were no more than three trials in a row with the same expected answer (i.e., three consecutive trials with "same" $S2$ or with "different" $S2$). After the last block, a questionnaire was administered to collect participants' opinions about the difficulty of the tasks (if it varied depending on the various type of stimuli and conditions) and potentially used strategies. The duration of the entire experimental session, including breaks, was about three hours. The current study was approved by the regional ethical committee, and informed consent was obtained prior to participation.

2.6. Analysis

1) Demographic details and control measures (i.e., WAIS-IV subtests) of the two groups of participants were compared with two-sided, independent Student's *t*-tests. Effect sizes are reported using Cohen's *d*.

2) Memory task data were analyzed in terms of Signal Detection Theory, notably by computing d' , for accuracy, and the c criterion, for response bias. The computation of these values is based on hit (i.e., number of correct responses for different trials/number of different trials) and false alarm (i.e., number of incorrect responses for same trials/number of same trials) rates of each participant in a given condition. In the case of no false alarms or maximum number of hits, we applied a correction of .01 and .99 respectively. For the response bias c , positive values indicate the tendency of answering "same" and negative values reflect the tendency of answering "different".

3) d' and c values were tested against zero, separately for each group, for each of the six types of stimuli, and for item and order tasks (i.e., the 12 blocks presented to participants).

4) d' and c criterion were analyzed with two Bayesian repeated measure ANOVAs, with Group (i.e., musicians and nonmusicians) as between-subject factor, and Modality (i.e., visual and auditory), Type (i.e., verbal, contour, and no-contour), and Change (i.e., item and order), as within-subject factors. The Bayesian approach allows comparing different models (that represent the data of interest) in order to select the best one (Wagenmakers et al., 2017). There is no rejection of null or alternative hypothesis, but only how much evidence there is towards a specific hypothesis. According to Lee and Wagenmakers (2014), Bayes Factor (BF_{10}) value can be interpreted as how many times a specific model is more likely to represent the data with respect to the null model; specifically, a value under three reflects weak evidence, a value between three and 10 medium evidence, between 10 and 100 strong evidence, and over 100 decisive evidence. We also report the $BF_{inclusion}$, which indicates the strength of the evidence of a single effect with respect to all possible models.

5) The factors retained based on the Bayesian ANOVA were further explored with post-hoc tests (i.e., Student's *t*-test, with FDR correction for multiple comparisons).

6) Six further Bayesian repeated measure ANOVAs were computed to explore any possible Modality or Group effects depending on the Position of change in different trials (e.g., first items, last items). These analyses were run separately for Change and Type, as the sequence length and possible position of changes differed depending on these conditions. The analyses included the percentage of correct responses in different trials as dependent variable, and Modality, Position of change (within-subjects), and Group (between-subjects) as independent variables. For the sake of brevity, post-hoc tests (*t*-tests with FDR correction) were run only to explore effects or interactions involving the Position of change factor included in the models selected.

7) Pearson correlations were computed to investigate the relationships between the different measures.

8) An exploratory analysis (ANOVAs and

Wilcoxon tests) was run to investigate the role of strategies (reported by the participants) used during the short-term memory task.

The data were analyzed with R (R Development Core Team, 2008) and JASP (JASP team, 2018) packages.

3. Results

3.1 Demographic data. The participants of the two groups were comparable in terms of age, sex distribution, education (Table 1) and in all WAIS-IV subtests, except for the Coding subtest, in which musicians outperformed non-musicians (Table 2).

Table 2.

Performance on the WAIS-IV subtests, separately for musicians and nonmusicians.

WAIS-IV SUBTESTS	Musicians	Nonmusicians	Group comparison	
	<i>N</i> = 24 (18 females)	<i>N</i> = 24 (19 females)	p-value	Cohen's <i>d</i>
Digit Span Forward (WM)	11 (1.89)	11 (1.35)	$p = 1$	$d = 0$
Digit Span Backward (WM)	9.79 (2.17)	9.08 (1.56)	$p = .53$	$d = .37$
Digit Span Sequencing (WM)	9.87 (2.72)	9.79 (1.15)	$p = 1$	$d = .04$
Arithmetic (WM)	15.83 (2.93)	14.92 (3.52)	$p = .53$	$d = .28$
WM index (composite score)	106.7 (14.87)	102.3 (12.97)	$p = .53$	$d = .32$
Coding (PS)	91.58 (15.68)	78.83 (13.6)	$p = .032$	$d = .87$
Symbol Search (PS)	41.29 (8.35)	41.08 (6.6)	$p = 1$	$d = .03$
PS index (composite score)	117.8 (15.73)	108.8 (14.18)	$p = .164$	$d = .61$

Note. p-values refer to the comparison between groups through t-tests (FDR corrected). The last column represents the effect size expressed with the *d* of Cohen. WM: Working Memory; PS: Processing Speed. The WM and PS indexes (composite scores) are computed based on the raw score of each subtest, which was converted into a scaled score taking into account the age of participants (based on normative data), and then transformed in the final standardized index with $M = 100$ and $SD = 15$.

3.2 Accuracy in the short-term memory tasks. In the memory task, all *d'* values were significantly different from zero, separately for group and for each of the conditions with $p < .001$. The Bayesian repeated measure ANOVA indicated several complex models to best represent the data: Table 3 summarizes the three best models (in comparison to the null model) and the data are illustrated Fig. 2.

Table 3.

The three best models among all possible ones in the Bayesian repeated measures ANOVA on d'

Model	BF ₁₀
Modality + Type + Group + Change + Modality * Type + Modality * Group + Type * Group + Change * Modality + Change * Type + Modality * Type * Group + Change * Modality * Type	4.244e+33
Modality + Type + Group + Change + Modality * Type + Modality * Group + Type * Group + Change * Modality + Change * Type + Change * Group + Modality * Type * Group + Change * Modality * Type	6.445e+32
Modality + Type + Group + Change + Modality * Type + Type * Group + Change * Modality + Change * Type + Change * Modality * Type	5.042e+32

Note. Bold characters indicate the best model.

The best model includes all our four single factors (i.e., Modality, Type, Group, and Change), and multiple interactions: The factor Modality interacted in three two-way interactions with Type, Group, or Change, respectively. There was also a three-way interaction between Modality, Type, and Group, and between Change, Modality and Type. In addition, Type interacted with Group and with Change.

Regarding the factors present in the best model, the subsequent Bayesian analysis of effects did not show any evidence for the effect of Change, for the interaction between Modality and Group, and for the interaction between Change and Modality (i.e., $BF_{inclusion} < 3$). The other factors and interactions of the best model had all a $BF_{inclusion} > 3$, and they will be described and further explored with post-hoc tests. The full results of the analysis of effects are presented in appendix in Table A.1.

The factor Group was associated with strong evidence, $BF_{inclusion} = 51.62$. The interaction between Type and Group was associated with a decisive evidence, $BF_{inclusion} = 200743$. The interaction between Type, Modality, and Group was associated with strong evidence, $BF_{inclusion} = 48.63$. Post-hoc tests revealed that in the visual modality (Fig. 2), musicians outperformed non-musicians with contour stimuli, $t(46) = 4.54, p < .001, d = 1.31$, where the mean of musicians ($M = 2.41, SD = 1.14$) was more than the double of the one of non-musicians ($M = 1.18, SD = .92$), but there was no statistical group difference for no-contour stimuli, $t(46) = .80, p = .476, d = .23$ and for verbal stimuli, $t(46) = 1.28, p = .309, d = .37$. In the auditory modality, musicians outperformed non-musicians with the contour stimuli, $t(46) = 4.11, p < .001, d = 1.19$ (musicians $M = 1.86, SD = .87$; non-musicians $M = .99, SD = .75$); and with the no-contour stimuli, $t(46) = 4.39, p < .001, d = 1.27$ (musicians $M = 2.25, SD = .90$; non-musicians $M = 1.41, SD = .78$), but did not differ significantly from non-musicians with the auditory verbal stimuli, $t(46) = .72, p = .476, d = .21$.

The factor Modality (i.e., visual, auditory) was associated with a decisive evidence as a main effect, $BF_{inclusion} = 2.334e+6$, and in interaction with Type, $BF_{inclusion} = 1900$. Post-hoc tests revealed that visual material led to better performance than auditory material for verbal stimuli ($M = 2.57$, $SD = .90$ and $M = 1.94$, $SD = .81$, respectively, $t(47) = -6.46$, $p < .001$, $d = 1.32$), for contour stimuli ($M = 1.79$, $SD = 1.2$ and $M = 1.43$, $SD = .92$, respectively, $t(47) = -3.12$, $p = .005$, $d = .64$), but not for no-contour stimuli (visual: $M = 1.84$, $SD = .81$; auditory: $M = 1.83$, $SD = .94$, $p = .945$, $d = .01$).

The type of stimuli (i.e., verbal, contour, no-contour) was associated with a decisive evidence, $BF_{inclusion} = 5.181e+16$. The post-hoc analysis revealed that performance was higher for verbal stimuli ($M = 2.26$, $SD = .91$) than contour stimuli ($M = 1.48$, $SD = 1.06$), $t(47) = -6.74$, $p < .001$, $d = 1.38$, and no-contour stimuli ($M = 1.74$, $SD = .85$), $t(47) = 6.41$, $p < .001$, $d = 1.31$. Performance with contour stimuli was also significantly different from the one with no-contour stimuli, but here the difference was smaller than the two previous ones, $t(47) = 2.26$, $p = .029$, $d = .46$.

The factor Change (i.e., order, item) was not associated with any positive evidence as a main effect, $BF_{inclusion} = .14$, but interacted with Type of stimuli, associated with a small evidence, $BF_{inclusion} = 3.79$. The Change and the Type of stimuli interacted with Modality in a three-way interaction, which was associated with a decisive evidence, $BF_{inclusion} = 139$. Post-hoc tests revealed that order change led to higher performance than item change only with auditory no-contour stimuli, $t(47) = -7.17$, $p < .001$, $d = 1.46$, (order change: $M = 2.20$, $SD = .959$; item change: $M = 1.45$, $SD = .756$), but performance did not differ for the other conditions (all p-values $> .063$).

Fig. 2. Accuracy (d') for the memory task, separately for Modality, Type of stimuli and Group. Inside each box, the horizontal line indicates the median. The red diamond represents the mean. The edges of the box represent the first quartile (Q1) and the third quartile (Q3) (i.e., the 25th and the 75th percentiles). The interquartile (IQR) range is the height of the box (i.e., $Q3 - Q1$). The upper whisker extends from Q3 to the largest value, no further than $1.5 * IQR$ from Q3. The lower whisker extends from Q1 to the smallest value no further than $1.5 * IQR$ from Q1. The black dots represent the outliers inferior or superior to these cut-off values. The red asterisks represent the statistically significant differences between groups (post-hoc tests).

3.3 Response bias. c values (Table 4) were tested against zero separately for groups. A positive bias (i.e., $c > 0$) describes the tendency of answering “same”, whereas a negative bias (i.e., $c < 0$) describes the tendency of answering “different”. Here, some conditions were associated with a positive bias, but differently between groups. See table 4 for details.

Table 4

Mean (SD) of the c criterion, separately for musicians and nonmusicians

	Musicians		Nonmusicians	
	Item	Order	Item	Order
Verbal (A)	-.12 (.47)	-.05 (.48)	.06 (.34)	.002 (.43)
Contour (A)	.3 (.42)*	.32 (.27)*	.26 (.33)*	.47 (.32)*
No-contour (A)	.49 (.36)*	.25 (.39)*	.38 (.34)*	.27 (.26)*
Verbal (V)	-.1 (.39)	-.23 (.42)*	-.16 (.36)(*)	.03 (.38)
Contour (V)	.07 (.36)	.21 (.41)*	.26 (.27)*	.24 (.29)*
No-contour (V)	.37 (.34)*	.35 (.25)*	.41 (.26)*	.29 (.35)*

Note. The asterisk * indicates $p < .05$ (FDR corrected) for the test against zero. (*) indicates p-values that were significant without FDR correction.

The Bayesian repeated measures ANOVA on the c criterion revealed that the best model to represent the data included only two factors, Modality and Type. Table 5 represents the three best models in comparison to the null model.

Table 5

The best three models among all the possible models in the Bayesian repeated measures ANOVA on c criterion.

Model	BF ₁₀
Modality + Type	3.376e +32
Modality + Type + Modality * Type	1.715e+32
Modality + Type + Change + Change * Type	1.147e+32

Note. In bold characters the best model.

The best model included two main factors (Modality and Type), without further factors and/or interactions. The Bayesian analysis of effects confirmed the evidence for Modality and Type, and further showed evidence for the interaction between Change and Type, and between Change, Modality, Type and Group ($BF_{inclusion} > 3$). The full analysis of effects is represented in Table A.2. Specifically, Modality was found to be associated with a weak evidence, $BF_{inclusion} = 5.93$, with the mean c value slightly higher in the auditory modality ($M = .22$, $SD = .41$) than in the visual modality ($M = .14$, $SD = .4$), $d = .2$. Type was associated with a decisive evidence, $BF_{inclusion} = 1.394e +32$. Post-hoc tests showed that specifically the c value was different for each one of the possible comparisons, namely, between verbal ($M = -.08$,

$SD = .41$) and contour stimuli ($M = .26$, $SD = .35$), $t(47) = -7.70$, $p < .001$, $d = 1.57$, between verbal and no-contour stimuli ($M = .35$, $SD = .33$), $t(47) = -9.68$, $p < .001$, $d = 1.98$ and between contour and no-contour stimuli, $t(47) = -2.25$, $p = .029$, $d = .46$. A weak evidence was also found to be associated with the interaction between Change and Type, $BF_{inclusion} = 3.98$, and to a four-way interaction, that is Change by Modality by Type by Group, $BF_{inclusion} = 3.81$. To further investigate this interaction, post-hoc t-tests were run separately for musicians and nonmusicians, and the only significant differences between the item and the order change were found to be in the auditory contour condition in nonmusicians, $t(23) = -3$, $p = .038$, $d = .87$, and in the auditory no-contour condition in musicians, $t(23) = 3.27$, $p = .038$, $d = .94$ (see Table 4 for the means and SD).

3.4 Position of change analysis. For the item change conditions, there were five possible positions of change, for verbal and contour stimuli, and four for no-contour stimuli. With verbal stimuli the best model representing our data was the following: Modality + Position + Modality * Position, $BF_{10} = 1.109e + 8$. The analysis of effects also confirmed the evidence for the interaction between Modality and Position, $BF_{inclusion} > 3$. Post-hoc tests for the interaction between Modality and Position revealed that all the positions of change differed between auditory and visual modality ($p \leq .041$), except of the fourth position, $p = .122$ (Fig. 3). However, it is interesting to observe that, in contrast to all other positions, in the last position of change auditory presentation had higher accuracy than the visual one. When comparing the positions of change within each modality, in the auditory modality a difference between position 1 and 2, 1 and 3, 2 and 5, 3 and 4, and 3 and 5 emerged (all p-values $\leq .043$). The other comparisons were not statistically significant (all p-values $\geq .083$). In the visual modality, no difference between positions of change emerged (all p-values $\geq .09$).

Fig. 3. Accuracy (mean percentage of correct responses in different trials) for different positions of change, separately for type of stimuli (i.e., verbal, contour, no-contour) and type of change (i.e., item, order). In each plot, group (i.e., musicians, non-musicians) and modality (i.e., auditory, visual) are represented separately.

With contour stimuli, the best model representing our data was the following: Modality + Group, $BF_{10} = 201.92$. This is in line with the main analysis showing again an effect of Modality and Group with contour stimuli, but differently from verbal stimuli, here no clear effect of Position of change emerged (see Fig. 3).

With no-contour stimuli, the best model representing our data included only the factor Modality, $BF_{10} = 66.21$. As Fig. 3 shows, visual stimuli were associated to better recognition than auditory stimuli.

In the order change conditions, the possible positions of change were four for the verbal and contour stimuli, and three for the no-contour stimuli. It is to note that here each change corresponds to two items

that are inverted: for example, position 1-2 will indicate the inversion of item 1 and item 2, position 2-3 of item 2 and item 3, and so on. With verbal stimuli the best model representing our data was the following: Modality + Position, $BF_{10} = 2.877e+7$. Here the model did not include any interactions, and the graph shows greater accuracy for visual items (as emerged in the main analysis). Concerning the position of change, post-hoc tests revealed that there was a statistically significant difference between position 1-2 and position 2-3, and between position 2-3 and position 4-5 (both p -values = .025). No other statistically significant difference emerged (all other p -values $\geq .192$) (see Fig. 3).

With contour stimuli, the best model representing our data was the following: Modality + Position + Group, $BF_{10} = 6.704e+7$, again revealing no interactions. The effect of Group and Modality is in line with the main analysis but concerning the position of change, post-hoc tests revealed that there was no statistically significant difference between the different positions of change (all p -values $\geq .071$).

Finally, with no-contour stimuli, the best model representing our data was the following: Modality + Position + Group + Modality * Group, $BF_{10} = 35061.91$. The effects of modality and group were again in line with the main analysis, and concerning the positions of change, post-hoc tests revealed that position 1-2 differed from position 2-3, $p = .008$, and that position 2-3 differed from position 3-4, $p = .024$. No difference between position 1-2 and position 3-4 was found, $p = .495$. (Fig. 3)

3.5 Correlation analysis. We investigated the possible relation between memory and control tasks with correlation analyses. Tables 6 (musicians) and 7 (non-musicians) present the correlations between the performance in recognition memory in the different conditions (collapsed across change type because the main analysis showed no important differences between the two types of change). Tables 8 (musicians) and 9 (non-musicians) present the correlations between the performance in recognition memory task for the different conditions (collapsed across change type) and participants' performance in the different WAIS-IV subtests.

Finally, we ran additional correlations within the musician group only, aiming to investigate potential relationships between degree of musical expertise (i.e., starting age, years of musical training, and hour of weekly practice) and recognition memory performance, but no significant correlations were observed (all p -values $> .511$).

Table 6. Pearson correlations in musicians among the recognition memory tasks, separately for modality and type.

	Verbal (A)	Contour (A)	No-contour (A)	Verbal (V)	Contour (V)	No-contour (V)
Verbal (A)	—					
Contour (A)	.59**	—				
No-contour (A)	.61**	.63**	—			
Verbal (V)	.48*	.28	.48*	—		
Contour (V)	.46*	.68**	.71**	.30	—	
No-contour (V)	.39	.12	.26	.18	.39	—

* $p \leq .05$, ** $p < .01$, *** $p < .001$

Note. $N = 24$. The p-values reported were corrected for multiple comparisons (FDR).

Table 7. Pearson correlations in non-musicians among the recognition memory tasks, separately for modality and type.

	Verbal (A)	Contour (A)	No-contour (A)	Verbal (V)	Contour (V)	No-contour (V)
Verbal (A)	—					
Contour (A)	.50*	—				
No-contour (A)	.48*	.67**	—			
Verbal (V)	.45*	.45*	.52*	—		
Contour (V)	.47*	.39	.31	.51*	—	
No-contour (V)	.17	.24	.33	.40(*)	-.01	—

* $p \leq .05$, ** $p < .01$, *** $p < .001$

Note. $N = 24$. The p-values reported were corrected for multiple comparisons (FDR). (*) indicates p-values that were significant before the FDR correction. (A): auditory; (V): visual.

Table 8. Pearson correlations in musicians among the WM and PS indexes of the WAIS-IV and the recognition memory tasks, separately for modality and type.

	Verbal (A)	Contour (A)	No-contour (A)	Verbal (V)	Contour (V)	No-contour (V)
WM index	.33	.12	.11	.27	.40 (*)	.34
PS index	.30	.02	-.12	.03	.27	.01

* $p \leq .05$, ** $p < .01$, *** $p < .001$

Note. $N = 24$. The p-values were corrected for multiple comparisons (FDR). (*) indicates p-values that were significant before the FDR correction. WM: Working Memory; PS: Processing Speed. (A): auditory; (V): visual.

Table 9. Pearson correlations in non-musicians among the WM and PS indexes of the WAIS-IV and the recognition memory tasks, separately for modality and type.

	Verbal (A)	Contour (A)	No-contour (A)	Verbal (V)	Contour (V)	No-contour (V)
WM index	.08	.1	.43(*)	.7***	.41(*)	.43(*)
PS index	.08	-.3	-.08	.07	.15	.19

* $p \leq .05$, ** $p < .01$, *** $p < .001$

Note. $N = 24$. The p-values were corrected for multiple comparisons (FDR). WM: Working Memory; PS: Processing Speed. (A): auditory; (V): visual.

3.6 Exploratory analysis based on participants' reported strategies. Based on the strategies reported by the participants at the end of the experiment, participants were sorted into different categories to investigate whether performance depended to some extent on the strategies used. In particular, we were interested in strategies used in the contour conditions, because the encoding of contour is likely crucial to perform well in these conditions. Moreover, the large differences between groups we observed could be linked to different ways of encoding the sequences (e.g., relying on the contour or not). Strategies were categorized as follows: “contour”, when participants reported to rely on contour by either picturing a mental line with curves, or imagining a melody; “possibly involving contour”, when participants reported to use strategies that might involve contour, such as imagining spatial positions; “none”, when they reported not to be able to use any strategy, and “other”, when the strategies reported were not belonging to any of the previous categories (e.g., verbal strategies, such as attaching a label to each element). Categories with two or less participants were removed from the analysis. Performance in the contour conditions depending on the used strategy are presented in Fig. 4.

Fig. 4. Discrimination performance (d') presented as a function of the type of strategies used in the contour conditions (visual and auditory), and separately for musicians and non-musicians. Above each box, the number of participants that reported using that strategy. Inside each box, the horizontal line indicates the median. The red diamond represents the mean. The edges of the box represent the first quartile (Q1) and the third quartile (Q3) (i.e., the 25th and the 75th percentiles). The interquartile (IQR) range is the height of the box (i.e., $Q3 - Q1$). The upper whisker extends from Q3 to the largest value, no further than $1.5 * IQR$ from Q3. The lower whisker extends from Q1 to the smallest value no further than $1.5 * IQR$ from Q1. The black dots represent the outliers inferior or superior to these cut-off values.

We performed two two-way ANOVAs with d' as dependent variable, and the strategies used and group as independent variables, for auditory contour and visual contour conditions respectively. In the auditory contour condition, there was a significant effect of the group (as in the main analysis), $F(1,39) = 8.88$, $p = .005$ (i.e., musicians performed better), and also a significant effect of strategies, $F(2,39) = 11.71$, $p < .001$, but no significant interaction between the two factors, $F(2,39) = .22$, $p = .802$.

To investigate the effect of strategies, we performed Wilcoxon tests, as we had large sample size differences across the various categories. There was a significant difference between “contour” and “none” strategies, $W = 240$, $p = .002$, where contour strategies ($M = 2.06$, $SD = .8$) were associated with better performances than no strategies ($M = .99$, $SD = .69$). Another significant difference emerged

between “possibly involving contour” ($M = 1.70$, $SD = .61$) and “none” categories, $W = 48$, $p = .019$. No difference between the “contour” and “possibly involving contour” categories emerged, $p = .257$.

In the visual contour condition, there was again a significant effect of the group (as in the main analysis) $F(1,41) = 7.93$, $p = .007$ (i.e., musicians performed better), and also a significant effect of strategies, $F(3,41) = 8.09$, $p < .001$, but no significant interaction between the two factors, $F(1,41) = 1.22$, $p = .276$. The effect of strategies was also investigated with Wilcoxon tests. There was a significant difference between “contour” and “none” categories, $W = 235.5$, $p = .001$, where contour strategies ($M = 2.4$, $SD = 1.08$) were associated with better performances than no strategies ($M = .99$, $SD = .69$). Another significant difference emerged between “possibly involving contour” ($M = 1.91$, $SD = .49$) and “none” categories, $W = 5$, $p = .04$. Again, no difference between the “contour” and “possibly involving contour” categories emerged, $p = .499$. Furthermore, there was a significant difference between the “other” ($M = 1.63$, $SD = .81$) and “none” categories, $W = 19$, $p = .035$, with the latter associated again with the lowest performance.

4. Discussion

The present study aimed to compare visual and auditory short-term memory for different types of materials, and to investigate whether musical expertise could influence short-term memory for some materials in a specific way. A recognition memory task was implemented with verbal stimuli and nonverbal stimuli with and without contour, implemented both in the visual and auditory modality, and with item and order changes for the different trials. One of the most salient findings of the study was that musicians showed a specific advantage over non-musicians with auditory contour and no-contour (nonverbal) stimuli, and with visual contour stimuli. Furthermore, results revealed that short-term memory performance differed depending on the modality, with different recency effects, but not with all types of stimuli. Concerning the general accuracy across materials, verbal stimuli were associated to higher accuracy than nonverbal stimuli. However, the difference between verbal and some of the nonverbal stimuli (i.e., the contour stimuli) decreased when participants reported to be able to use specific strategies (e.g., using contour information improved performance). We discuss below these findings with respect to the possible theoretical implications for short-term memory and their contribution to further understand the potential benefits of intensive musical training.

Contour plays an important role in short-term memory performance across modalities. Musicians’ performance was better than non-musicians’ performance in the contour conditions (regardless of the

sensory modality). For both auditory and visual modalities, most of musicians reported to use contour as a cue for remembering the sequences. These findings might suggest that short-term memory for contour stimuli might be based on mechanisms shared between the auditory modality and the visual modality, and that these mechanisms are more efficient in individuals with musical training. The self-reported use of contour cues led to higher performance in the contour conditions, as revealed by the analysis of strategies. In fact, also non-musicians who reported to use contour strategies performed better than non-musicians who did not report the use of any strategy, supporting the hypothesis that the way material is encoded contributes to short-term memory performance. Moreover, for the visual modality, among participants who reported the use of contour strategies, 14 musicians and four non-musicians reported to have automatic mental associations between a pitch and a specific level of luminance and that they retained an imagined melody. Note that this association between the luminance sequences and auditory imagery has also been reported by some participants in the study of Gold et al. (2014), where participants had to spot whether the second half of an eight-item sequence was a repetition of its first half. However, the authors did not control for musical expertise and did not report any musical training information, so it might be possible that the association luminance-music has been reported by some of their participants that had musical training. Concerning our results, interestingly, in the auditory contour condition, fewer participants (than in the visual contour condition) reported to retain the sequence as an imagined melody (i.e., three musicians and one nonmusicians) even though previous studies observed that loudness changes can also induce a feeling of contour changing in pitch. This was observed, for example, by comparing systematically contour representations with different variations of sounds' proprieties, such as pitch, brightness, and loudness (McDermott et al., 2008).

Overall, in both modalities our data suggests that the ability to extract contour was stronger in musicians than in non-musicians, because the non-musicians who reported to use contour strategies did not reach the same level of accuracy as musicians. Moreover, a correlation between auditory and visual performance for contour stimuli was observed in musicians only. This reinforces the idea that when using the same encoding strategy across materials of different sensory modalities, a common memory process could be recruited instead of separate processes for each type of material and/or sensory modality (the latter having been observed in previous studies: Crottaz-Herbette et al., 2006; Fougny & Marois, 2011; Macken et al., 2016). This view could be integrated into the conception that short-term memory (but also working memory) has a distributed nature, as proposed by Christophel and colleagues (2017), with sensory modality representations linked to the sensory-specific brain areas, and more abstract

representations linked to sensory-independent brain areas. Contour representation might thus be memorized in an abstract way and become independent from its sensory modality (even though at early stages the sensory modality likely plays a role in encoding the information, e.g., here, visual information was generally better retained than auditory information). These findings and their interpretation can also be integrated into the working memory model of Cowan (1999, 2010). In this model, the stimuli are stored initially into a (modality-specific) sensory store for hundreds of milliseconds, and they later activate either a portion of long-term memory (if the stimuli are already known and the person has already some representations of them), or (if the stimuli are new) they will stay temporarily active thanks to the attentional focus. Cowan's model postulates the existence of a unique temporary storage, and that eventual differences across materials are linked to different long-term memory representations that are activated. Based on this model, one might hypothesize that the improved performance of musicians in the contour conditions might be due to their stronger contour representation in long-term memory.

Sensory modality is associated to different memory performance. Concerning the difference between auditory and visual modalities, in the main analysis, performance (as measured by d') was higher with visual verbal stimuli than with auditory verbal stimuli, here composed of exactly the same syllables. A difference in performance as a function of the modality of presentation was observed also for the contour conditions (non-verbal), where performance to visual stimuli (i.e., luminance) was higher than performance to auditory stimuli (i.e., loudness) in both participant groups. Interestingly, when looking at the position of change analysis, in the verbal task our data confirmed previous studies that showed an auditory recency effect (but not a visual one) (Frankish, 1989; Maylor et al., 1999; Macken et al. 2016). In the item condition, moreover, in the last position of change there was an auditory advantage over the visual modality, which it is interesting because the main analysis showed a general advantage of visual presentation over the auditory one. This data would support the idea that there are different memory processes for auditory and visual verbal memory.

Concerning contour stimuli, interestingly, no auditory recency effect was observed: our results seems to be in agreement with the study by Roberts (1986), where no modality-specific effect was found with melodic (i.e., sequences of tones, played or displayed on a staff) and harmonic stimuli (sequences of chords, played or displayed on a staff). It might thus be possible that contour representations are processed differently than representations of verbal stimuli. If so, this could be an important information for revising some of the ideas of how auditory objects are formed. For some authors, auditory verbal sequences are represented as unique object, and this would occur more strongly than with visual verbal

sequences (see Macken et al., 2016). However, contour sequences are also likely to be represented as a single object, rather than as single items, but we did not observe a similar accuracy pattern to auditory verbal stimuli. Moreover, with contour stimuli, we neither observed a modality-specific effect (i.e., interaction between position of change and modality), nor a recency effect, suggesting that the functioning of short-term memory is not always influenced by the sensory modality, and that other processes should be taken into account (e.g., material-specific encoding strategies).

Finally, with no-contour stimuli, even though we did not find any interaction between modality and position, our data seems to suggest that an auditory recency effect exists, as with verbal stimuli, particularly in the item change condition (see Fig.3).

To conclude, the present data confirms recency effects in the verbal domain with auditory but not visual material; but when another type of encoding is used (i.e., contour-based), modality effects are reduced, revealing a potential different memory process that could be used with musical and contour stimuli.

No overall performance differences between item and order change conditions. There was only one condition in which there was a significant difference between the two types of change, that is the auditory no-contour condition. Notably, performance was higher with order changes than with item changes, and this could be related to the fact that the items were new to participants and likely more difficult to memorize. With all other types of stimuli, the performance did not differ between the item and the order change conditions. This differs from previous studies (that used both recall and recognition tasks), which observed a different performance in terms of accuracy between the item memory and serial order memory (Gorin et al., 2016; Hachmann et al., 2013; Majerus et al., 2006). Here, we did not find an overall difference in terms of accuracy between item and order change: it could be possible that our recognition paradigm is not the best task to investigate the two processes. In fact, in previous studies (e.g., Majerus et al., 2006), the authors tested serial order memory with a recall task, by presenting some cards (picturing different animals) that were given to participants, who had to use the cards to reconstruct the order of a sequence previously presented. Gorin and colleagues (2016) observed, differently from us a difference in accuracy between item and order memory (i.e., order memory had higher accuracy than item memory). However, their manipulation in the item condition was probably more difficult than ours, as they were changing only one phoneme of a non-word (in a sequence of non-words) instead of the whole item (as, in our case, the syllable). It might be possible that order and item memory processes are sensitive to the type of task and material. Future research should further investigate the conditions under

which the difference between item and order memory emerges, to uncover its role in memory mechanisms.

Musicians showed better short-term memory performance than non-musicians for auditory and visual contour stimuli and auditory no-contour stimuli, but not with verbal stimuli. Previous studies have shown better memory skills for musicians (in comparison to non-musicians), leading to discussions about the possible benefits of musical training (Talamini et al., 2017). Here we investigated the extent of the advantage of musicians in short-term memory tasks with different types of materials and modalities. In particular, we expected a stronger advantage of musicians over non-musicians for auditory memory than for visual memory (see Talamini et al., 2016, Tierney, Bergeson-Dana, & Pisoni, 2008). This was partially confirmed by the present results. We observed a main effect of group, along with a significant interaction between group, modality and type factors, indicating that musicians performed better than non-musicians for both visual and auditory contour stimuli, and for auditory no-contour stimuli. The fact that musicians outperformed non-musicians for contour stimuli, regardless of the modality of presentation, suggests that musicians are able to efficiently extract a structure (i.e., the contour), as discussed above. Musical expertise might train and facilitate contour extraction (even though we cannot infer causality with the present study). However, with auditory contour stimuli we cannot exclude that musicians and non-musicians have different loudness discriminations thresholds, as musicians are trained to play with different intensities. If so, this would point out the importance of considering discrimination abilities and their influence on short-term memory performance, for example, not by simply assessing hearing with an audiogram test, but also by assessing individual intensity discrimination thresholds between pair of tones. Note, however, that the stimuli differed by 5 dB (minimum) and this is above the discrimination threshold of loudness (e.g., Kidd, Watson, Gygy 2007).

For the no-contour stimuli, musicians performed better than non-musicians in the auditory modality (i.e., the pink noise sequences). This result might also have been influenced by musicians' enhanced auditory perception skills (Kraus & Chandrasekaran, 2010; Strait & Kraus, 2014), which could have helped them perceiving subtle pitch differences of these stimuli, allowing them to rely on a contour also in this condition. Even though the pink noises selected for the present study did not have a specific fundamental frequency, it is possible that the different texture of the noises could elicit different pitch percept or spectral pitch. Some musicians reported, in fact, that they could hear a contour with the pink noise stimuli, even if we aimed to have an auditory condition that could not elicit verbal labels or contour feeling. This could have helped them in remembering the sequences as having also a contour. This reflects the difficulty of finding different sounds that when presented in a sequence do not elicit any feeling of

contour (especially when testing musician participants). In any case, the possibility that musicians might have an advantage in auditory discrimination skills is probably not the only explanation contributing to our results: in fact, we found that musicians performed better than non-musicians also in the luminance task, and a superior auditory perception ability is unlikely linked to also a superior visual perception ability.

Previous research suggests that musicians perform better than non-musicians in verbal memory tasks, and the advantage seems to be stronger when the task is presented auditorily (e.g., Franklin et al., 2008; Tierney et al., 2008, Talamini et al., 2016). Moreover, in a recent meta-analysis the advantage of musicians over non-musicians in verbal short-term memory tasks was found to be quite consistent, as reflected by a moderate effect size and a small variability (Talamini et al., 2017). Based on the literature, we were thus expecting an advantage of musicians over non-musicians with the verbal sequences, especially in the auditory presentation. However, the present findings did not show any performance differences between groups in the verbal memory tasks. A possible explanation for this discrepancy is that we presented a recognition task, whereas most of the studies that investigated verbal short-term memory in musicians and non-musicians presented recall tasks (see Talamini et al., 2017 for a review). Another possibility for the lack of difference in the auditory verbal condition, is that here we used synthesized syllables that sounded “artificial” whereas most of auditory verbal tasks are usually using natural speech stimuli (e.g., the experimenter reads aloud the stimuli). This in any case would be an incomplete explanation, because in the digit span task of the WAIS-IV, which was presented auditorily with the experimenter reading the number sequences, there was again no difference between groups. It might be possible that the effect observed in the past is not as strong as it was suggested by a previous meta-analysis (Talamini et al., 2017), or that other variables contribute to performance (note that here we equated education levels between groups, and checked peripheral hearing in all participants).

CONCLUSIONS and PERSPECTIVES

To conclude, the findings of the present study point to an important role of contour cues in short-term memory performance, with possibly shared mechanisms for contour memory across modalities. When the same type of encoding strategy is used, sensory modality likely becomes less important during short-term memory maintenance. In support of this, the present study showed that the classical modality effects (e.g., recency effects with auditory, but not visual, presentation of verbal stimuli) was not observed with this contour stimuli, and the same pattern of accuracy in the different positions of

change was observed between visual and auditory modalities. Moreover, this is further supported by the fact that the encoding of contour was more efficient in musician participants: there was an advantage of musicians over non-musicians both with auditory and with visual stimuli. The selective advantage of musicians might be related to musicians' ability to encode the contour of the sequence, or more generally, it might be possible that they have general better maintenance in short-term memory for certain types of stimuli. Musicians' auditory perception skills (at least in the auditory tasks) might contribute in creating a stronger mnemonic trace (as they performed better also in the auditory no-contour task). Future work would be needed to investigate whether there is a causal relationship between musical training and enhanced memory skills for certain types of stimuli, and whether this advantage is linked to a general better memory, to a better use of strategies, and/or to better perception skills.

To conclude, the paradigm devised here aims to provide a comprehensive description of auditory and visual short-term memory performance. Future work will need to further investigate modality effects with nonverbal material, as it seems that short-term memory modality effects vary depending on the type of stimuli. Future work using this paradigm in patient groups or other expert groups, as well as its use in combination with brain imaging methods, should help to further characterize shared and distinct mechanisms in auditory and visual short-term memory during encoding and maintenance of various types of information.

DECLARATIONS

Funding. This work was supported by a fellowship from the LABEX CeLyA – Centre Lyonnais d'Acoustique (ANR-10-LABX-0060) to FT. This work was performed within the framework of the LABEX CeLyA (ANR-10-LABX-0060) and the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon, within the program “Investissements d'Avenir” (ANR-16-IDEX-0005) operated by the French ANR.

Conflicts of interest/Competing interests. The authors declare no conflicts of interest/competing interests.

Ethics approval: The current study was approved by the regional ethical committee.

Consent to participate: all participants signed the informed consent prior to participation.

References

- Aizenman, A. M., Gold, J. M., & Sekuler, R. (2017). Multisensory integration in short-term memory: Musicians do rock. *Neuroscience*, 1–11. <https://doi.org/10.1016/j.neuroscience.2017.04.031>
- Albouy, P., Schulze, K., Caclin, A., & Tillmann, B. (2013). Does tonality boost short-term memory in congenital amusia? *Brain Research*, 1537, 224–232. <https://doi.org/10.1016/j.brainres.2013.09.003>
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes1. In *Psychology of learning and motivation* (Vol. 2, pp. 89-195). Academic Press. [https://doi.org/10.1016/S0079-7421\(08\)60422-3](https://doi.org/10.1016/S0079-7421(08)60422-3)
- Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11), 417–423. [https://doi.org/10.1016/S1364-6613\(00\)01538-2](https://doi.org/10.1016/S1364-6613(00)01538-2)
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *The Psychology of Learning and Motivation Advances in Research and Theory*. [https://doi.org/10.1016/S0079-7421\(08\)60452-1](https://doi.org/10.1016/S0079-7421(08)60452-1)
- Balch, W. R., & Muscatelli, D. L. (1986). The interaction of modality condition and presentation rate in short-term contour recognition. *Perception & psychophysics*, 40(5), 351-358. DOI: 10.3758/BF03203027
- Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology*, 20(5), 529. <http://dx.doi.org/10.1037/0894-4105.20.5.529>
- Berz, W. L. (1995). Working memory in music: A theoretical model. *Music Perception: An Interdisciplinary Journal*, 12(3), 353-364. DOI: 10.2307/40286188
- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone Language Speakers and Musicians Share Enhanced Perceptual and Cognitive Abilities for Musical Pitch: Evidence for Bidirectionality between the Domains of Language and Music. *PLoS ONE*, 8(4). <https://doi.org/10.1371/journal.pone.0060676>
- Caclin, A., & Tillmann, B. (2018). Musical and verbal short-term memory: Insights from neurodevelopmental and neurological disorders. *Annals of the New York Academy of Sciences*. <https://doi.org/10.1111/nyas.13733>
- Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The distributed nature of working memory. *Trends in cognitive sciences*, 21(2), 111-124. <https://doi.org/10.1016/j.tics.2016.12.007>

- Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, 30(7), 1086-1095. <https://doi.org/10.3758/BF03194326>
- Cohen, M. A., Evans, K. K., Horowitz, T. S., & Wolfe, J. M. (2011). Auditory and visual memory in musicians and nonmusicians. *Psychonomic bulletin & review*, 18(3), 586-591. DOI: 10.3758/s13423-011-0074-0
- Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science*, 10, 135–136. DOI: 10.3758/BF03331446
- Corballis, M. C. (1966). Rehearsal and decay in immediate recall of visual recall of visually and aurally presented items. *Canadian Journal of Psychology*, 20, 43–50. <https://doi.org/10.1037/h0082923>
- Cowan, N. (1999). An embedded-processes model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control*, 20, 506.
- Cowan, N. (2000). Processing limits of selective attention and working memory: Potential implications for interpreting. *Interpreting*, 5(2), 117-146. <https://doi.org/10.1075/intp.5.2.05cow>
- Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in brain research*, 169, 323-338. [https://doi.org/10.1016/S0079-6123\(07\)00020-9](https://doi.org/10.1016/S0079-6123(07)00020-9)
- Crottaz-Herbette, S., Anagnoson, R. T., & Menon, V. (2004). Modality effects in verbal working memory: Differential prefrontal and parietal responses to auditory and visual stimuli. *NeuroImage*, 21(1), 340–351. <https://doi.org/10.1016/j.neuroimage.2003.09.019>
- Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 5(6), 365-373. DOI: 10.3758/BF03210660
- Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, 49(2B), 524-531. <https://doi.org/10.1121/1.1912382>
- Dowling, W. J., & Hollombe, A. W. (1977). The perception of melodies distorted by splitting into several octaves: Effects of increasing proximity and melodic contour. *Perception and Psychophysics*, 21(1), 60-64. <https://doi.org/10.3758/BF03199469>
- Drewnowski, A., & Murdock, B. A. (1980). The role of auditory features in memory span for words. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 319–322. <https://doi.org/10.1037/0278-7393.6.3.319>

- Fougnie, D., & Marois, R. (2011). What limits working memory capacity? Evidence for modality-specific sources to the simultaneous storage of visual and auditory arrays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37(6), 1329. <http://dx.doi.org/10.1037/a0024834>
- Frankish, C. (2008). Precategorical acoustic storage and the perception of speech. *Journal of Memory and Language*, 58, 815–836. <https://doi.org/10.1016/j.jml.2007.06.003>
- Franklin, M. S., Sledge Moore, K., Yip, C. Y., Jonides, J., Rattray, K., & Moher, J. (2008). The effects of musical training on verbal memory. *Psychology of Music*, 36(3), 353-365. <https://doi.org/10.1177/0305735607086044>
- Fry, A. F., & Hale, S. (1996). Processing speed, working memory, and fluid intelligence: Evidence for a developmental cascade. *Psychological science*, 7(4), 237-241. <https://doi.org/10.1111/j.1467-9280.1996.tb00366.x>
- Gallace, A., Tan, H. Z., Haggard, P., & Spence, C. (2008). Short term memory for tactile stimuli. *Brain research*, 1190, 132-142. <https://doi.org/10.1016/j.brainres.2007.11.014>
- George, E. M., & Coch, D. (2011). Music training and working memory: An ERP study. *Neuropsychologia*, 49(5), 1083–1094. <https://doi.org/10.1016/j.neuropsychologia.2011.02.001>
- Gold, J. M., Aizenman, A., Bond, S. M., & Sekuler, R. (2014). Memory and incidental learning for visual frozen noise sequences. *Vision research*, 99, 19-36. <https://doi.org/10.1016/j.visres.2013.09.005>
- Gorin, S., Kowialiewski, B., & Majerus, S. (2016). Domain-generalty of timing-based serial order processes in short-term memory: New insights from musical and verbal domains. *PloS one*, 11(12), e0168699. <https://doi.org/10.1371/journal.pone.0168699>
- Gottlieb, Y., Vaadia, E., & Abeles, M. (1989). Single unit activity in the auditory cortex of a monkey performing a short term memory task. *Experimental Brain Research*, 74(1), 139-148. <https://doi.org/10.1007/BF00248287>
- Graves J. E., Pralus, A., Fornoni L., Oxenham A. J., Caclin A., & Tillmann B. (2019) Short- and long-term memory for pitch and non-pitch contours: Insights from congenital amusia. *Brain and Cognition*. <https://doi.org/10.1016/j.bandc.2019.103614>
- Greene, R. L., & Samuel, A. G. (1986). Recency and suffix effects in serial recall of musical stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(4), 517.
- Hansen, M., Wallentin, M., & Vuust, P. (2013). Working memory and musical competence of musicians and non-musicians. *Psychology of Music*, 41(6), 779–793. <https://doi.org/10.1177/0305735612452186>

- Herz, R. S., & Engen, T. (1996). Odor memory: Review and analysis. *Psychonomic Bulletin & Review*, 3(3), 300-313. <https://doi.org/10.3758/BF03210754>
- Hirel, C., Nighoghossian, N., Lvque, Y., Hannoun, S., Fornoni, L., Daligault, S., ... & Caclin, A. (2017). Verbal and musical short-term memory: Variety of auditory disorders after stroke. *Brain and cognition*, 113, 10-22. <https://doi.org/10.1016/j.bandc.2017.01.003>
- Idson, W. L., & Massaro, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception and Psychophysics*, 24(6), 551-565. <https://doi.org/10.3758/BF03198783>
- JASP Team. (2018). JASP (Version 0.8. 6)[Computer software].
- Kidd, G. R., Watson, C. S., & Gygi, B. (2007). Individual differences in auditory abilities. *The Journal of the Acoustical Society of America*, 122(1), 418-435. <https://doi.org/10.1121/1.2743154>
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature reviews neuroscience*, 11(8), 599. <https://doi.org/10.1038/nrn2882>
- Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.
- Lehnert, G., & Zimmer, H. D. (2006). Auditory and visual spatial working memory. *Memory & cognition*, 34(5), 1080-1090. <https://doi.org/10.3758/BF03193254>
- Lichtenberger, E. O., & Kaufman, A. S. (2012). *Essentials of WAIS-IV assessment* (Vol. 96). John Wiley & Sons.
- Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica*, 75(1), 55-74. [https://doi.org/10.1016/0001-6918\(90\)90066-O](https://doi.org/10.1016/0001-6918(90)90066-O)
- Macken, B., Taylor, J. C., Kozlov, M. D., Hughes, R. W., & Jones, D. M. (2016). Memory as embodiment: The case of modality and serial short-term memory. *Cognition*, 155, 113-124. <https://doi.org/10.1016/j.cognition.2016.06.013>
- Majerus, S., Poncelet, M., Greffe, C., & Van der Linden, M. (2006). Relations between vocabulary development and verbal short-term memory: The relative importance of short-term memory for serial order and item information. *Journal of experimental child psychology*, 93(2), 95-119. <https://doi.org/10.1016/j.jecp.2005.07.005>
- Marks, L. E. (1987). On cross-modal similarity: Auditory–visual interactions in speeded discrimination. *Journal of Experimental Psychology: Human Perception and Performance*, 13(3), 384. <http://dx.doi.org/10.1037/0096-1523.13.3.384>
- Maylor, E. A., Vousden, J. I., & Brown, G. D. A. (1999). Adult age differences in short-term

memory for serial order: Data and a model. *Psychology and Aging*, 14, 572–594.
<https://doi.org/10.1037/0882-7974.14.4.572>

McDermott, J. H., Lehr, A. J., & Oxenham, A. J. (2008). Is relative pitch specific to pitch? *Psychological Science*. <https://doi.org/10.1111/j.1467-9280.2008.02235.x>

McDermott, J. H., Keebler, M. V., Micheyl, C., & Oxenham, A. J. (2010). Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special. *The Journal of the Acoustical Society of America*, 128(4), 1943-1951. <https://doi.org/10.1121/1.3478785>

Murdock Jr, B. B. (1966). Visual and auditory stores in short-term memory. *The Quarterly Journal of Experimental Psychology*, 18(3), 206-211. <https://doi.org/10.1080/14640746608400031>

Murdock Jr, B. B. (1967). Recent developments in short-term memory. *British Journal of Psychology*, 58(3-4), 421-433. <https://doi.org/10.1111/j.2044-8295.1967.tb01099.x>

Pallesen, K. J., Brattico, E., Bailey, C. J., Korvenoja, A., Koivisto, J., Gjedde, A., & Carlson, S. (2010). Cognitive control in auditory working memory is enhanced in musicians. *PLoS ONE*, 5(6). <https://doi.org/10.1371/journal.pone.0011120>

Peretz, I., & Babai, M. (1992). The role of contour and intervals in the recognition of melody parts: Evidence from cerebral asymmetries in musicians. *Neuropsychologia*, 30(3), 277-292. [https://doi.org/10.1016/0028-3932\(92\)90005-7](https://doi.org/10.1016/0028-3932(92)90005-7)

Martinez-Perez, T. M., Majerus, S., Mahot, A., & Poncelet, M. (2012). Evidence for a specific impairment of serial order short-term memory in dyslexic children. *Dyslexia*, 18(2), 94-109. <https://doi.org/10.1002/dys.1438>

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, 18(3), 251-269. DOI: 10.3758/BF03213879

Prince, J. B., Schmuckler, M. A., & Thompson, W. F. (2009). Cross-modal melodic contour similarity. *Canadian Acoustics*. Retrieved from: <https://researchrepository.murdoch.edu.au/id/eprint/3014/>

Roads, C. (1988). Introduction to granular synthesis. *Computer Music Journal*, 12(2), 11-13. DOI: 10.2307/3679937

Roberts, L. A. (1986). Modality and suffix effects in memory for melodic and harmonic musical materials. *Cognitive Psychology*, 18(2), 123-157. [https://doi.org/10.1016/0010-0285\(86\)90010-1](https://doi.org/10.1016/0010-0285(86)90010-1)

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, A. R., Schulz, J. B., ... & Eickhoff, S. B. (2012). Modelling neural correlates of working memory: a coordinate-based meta-analysis. *Neuroimage*, 60(1), 830-846. <https://doi.org/10.1016/j.neuroimage.2011.11.050>

- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, 32(5), 771–783. <https://doi.org/10.1002/hbm.21060>
- Schulze, K., Dowling, W. J., & Tillmann, B. (2012). Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Perception: An Interdisciplinary Journal*, 29(3), 255-267. DOI: 10.1525/mp.2012.29.3.255
- Schulze, K., & Tillmann, B. (2013). Working memory for pitch, timbre, and words. *Memory*, 21(3), 377-395. <https://doi.org/10.1080/09658211.2012.731070>
- Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral cortex*, 6(1), 11-20. <https://doi.org/10.1093/cercor/6.1.11>
- Spiegel, M. F., & Watson, C. S. (1984). Performance on frequency-discrimination tasks by musicians and nonmusicians. *The Journal of the Acoustical Society of America*. <https://doi.org/10.1121/1.391605>
- Spitzer, B., Goltz, D., Wacker, E., Auksztulewicz, R., & Blankenburg, F. (2014). Maintenance and manipulation of somatosensory information in ventrolateral prefrontal cortex. *Human brain mapping*, 35(5), 2412-2423. <https://doi.org/10.1002/hbm.22337>
- Strait, D. L., & Kraus, N. (2014). Biological impact of auditory expertise across the life span: musicians as a model of auditory learning. *Hearing research*, 308, 109-121. <https://doi.org/10.1016/j.heares.2013.08.004>
- Suárez, L., Elangovan, S., & Au, A. (2016). Cross-sectional study on the relationship between music training and working memory in adults. *Australian Journal of Psychology*, 68(1), 38–46. <https://doi.org/10.1111/ajpy.12087>
- Talamini, F., Altoè, G., Carretti, B., & Grassi, M. (2017). Musicians have better memory than nonmusicians: A meta-analysis. *PloS one*, 12(10), e0186773. • <https://doi.org/10.1371/journal.pone.0186773>
- Talamini, F., Carretti, B., & Grassi, M. (2016). The working memory of musicians and nonmusicians. *Music Perception*, 34(2). <https://doi.org/10.1525/MP.2016.34.2.183>
- Tallon-Baudry, C., Bertrand, O., & Fischer, C. (2001). Oscillatory synchrony between human extrastriate areas during visual short-term memory maintenance. *Journal of Neuroscience*, 21(20), RC177-RC177. Retrieved from: <https://www.jneurosci.org/content/jneuro/21/20/RC177.full.pdf>

Tallon-Baudry, C., Bertrand, O., Peronnet, F., & Pernier, J. (1998). Induced γ -band activity during the delay of a visual short-term memory task in humans. *Journal of Neuroscience*, 18(11), 4244-4254. DOI: <https://doi.org/10.1523/JNEUROSCI.18-11-04244.1998>

Team, R. C. (2013). R: A language and environment for statistical computing.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schröger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*. <https://doi.org/10.1007/s00221-004-2044-5>

Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 48(3), 380. <http://dx.doi.org/10.1037/1196-1961.48.3.380>

Tierney, A. T., Bergeson-Dana, T. R., & Pisoni, D. B. (2008). Effects of early musical experience on auditory sequence memory. *Empirical musicology review: EMR*, 3(4), 178. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3050509/>

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain and Cognition*, 71(3), 259–264. <https://doi.org/10.1016/j.bandc.2009.08.003>

Tillmann, B., Leveque, Y., Fornoni, L., Albouy, P., & Caclin, A. (2016). Impaired short-term memory for pitch in congenital amusia. *Brain Research*, 1640, 251-263. <https://doi.org/10.1016/j.brainres.2015.10.035>

Tremblay, S., Parmentier, F. B. R., Guérard, K., Nicholls, A. P., & Jones, D. M. (2006). A spatial modality effect in serial memory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 32, 1208–1215. <https://doi.org/10.1037/0278-7393.32.5.1208>

Wagenmakers, E. J., Verhagen, J., Ly, A., Matzke, D., Steingroever, H., Rouder, J. N., & Morey, R. D. (2017). The need for Bayesian hypothesis testing in psychological science. *Psychological science under scrutiny: Recent challenges and proposed solutions*, 123-138.

Wechsler, D. (2008). Wechsler adult intelligence scale—Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*, 22, 498. DOI: [10.2298/psi1710010011](https://doi.org/10.2298/psi1710010011)

Williamson, V. J., & Stewart, L. (2010). Memory for pitch in congenital amusia: Beyond a fine-grained pitch discrimination problem. *Memory*, 18(6), 657-669. <https://doi.org/10.1080/09658211.2010.501339>

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory and Cognition*, 38(2), 163–175. <https://doi.org/10.3758/MC.38.2.163>

Yang, H., Lu, J., Gong, D., & Yao, D. (2016). How do musical tonality and experience affect visual working memory? *NeuroReport*, 27(2), 94–98. <https://doi.org/10.1097/WNR.0000000000000503>

APPENDIX

Table A1.

Bayesian analysis of effects for the analysis of d'

Effects	P(incl)	P(incl data)	BF Inclusion
Group	0.114	4.634e -7	51.619
Modality	0.114	2.955e -6	2.334e +6
Type	0.114	4.345e -11	5.181e +16
Change	0.114	0.002	0.137
Group * Modality	0.299	0.017	0.152
Group * Type	0.299	0.128	200742.676
Group* Change	0.299	0.137	0.173
Modality * Type	0.299	0.008	1899.931
Change * Modality	0.299	0.011	0.378
Change * Type	0.299	0.024	3.797
Modality * Type * Group	0.114	0.866	48.630
Change * Modality * Group	0.114	0.025	0.175
Change * Type * Group	0.114	0.033	0.210
Change * Modality * Type	0.114	0.943	138.925
Change * Modality * Type * Group	0.006	0.003	0.796

Table A2.

Bayesian analysis of effects for the analysis of c criterion

Effects	P(incl)	P(incl data)	BF Inclusion
Group	0.114	0.217	0.309
Modality	0.114	0.569	5.928
Type	0.114	0.473	1.394e +32
Change	0.114	0.062	0.097
Modality * Type	0.299	0.294	0.483
Modality * Group	0.299	0.028	0.116
Type * Group	0.299	0.049	0.195
Change * Modality	0.299	0.035	0.118
Change * Type	0.299	0.286	3.932
Change * Group	0.299	0.012	0.105
Modality * Type *	0.114	2.238e -4	0.118
Group			
Change * Modality *	0.114	0.001	0.128
Type			
Change * Modality *	0.114	3.322e -5	0.160
Group			
Change * Type *	0.114	1.419e -4	0.076
Group			
Change * Modality *	0.006	6.139e -9	3.808
Type * Group			