Routledge
Taylor & Francis Group

Working memory for pitch, timbre, and words

Katrin Schulze[1,2] and Barbara Tillmann[1]

[1]Lyon Neuroscience Research Center, Auditory Cognition and Psychoacoustics Team, CNRS UMR5292, INSERM U1028, Université de Lyon, Lyon, France
[2]Developmental Cognitive Neuroscience Unit, UCL Institute of Child Health, London, UK

Aiming to further our understanding of fundamental mechanisms of auditory working memory (WM), the present study compared performance for three auditory materials (words, tones, timbres). In a forward recognition task (Experiment 1) participants indicated whether the order of the items in the second sequence was the same as in the first sequence. In a backward recognition task (Experiment 2) participants indicated whether the items of the second sequence were played in the correct backward order. In Experiment 3 participants performed an articulatory suppression task during the retention delay of the backward task. To investigate potential length effects the number of items per sequence was manipulated. Overall findings underline the benefit of a cross-material experimental approach and suggest that human auditory WM is not a unitary system. Whereas WM processes for timbres differed from those for tones and words, similarities and differences were observed for words and tones: Both types of stimuli appear to rely on rehearsal mechanisms, but might differ in the involved sensorimotor codes.

Keywords: Working memory; Auditory processing; Music cognition.

The memory system that enables humans to maintain and manipulate information for short periods of time has been referred to as working memory (WM). Our understanding of auditory WM and whether there are differences for the processing of different auditory materials (e.g., speech and music) is still elusive. Previous findings suggest that auditory WM is not a unitary system (e.g., Berz, 1995; Pechmann & Mohr, 1992; Schulze, Zysset Mueller, Friederici, & Koelsch, 2011). Our present study set out to investigate whether (i) the phonological loop, a WM component known to process verbal information (Baddeley, 2012), is also processing non-verbal auditory information (tones and timbre)

and (ii) auditory WM differs between stimuli that can be internally rehearsed (words, tones) and stimuli that are more difficult to rehearse (timbre). Therefore, to further our understanding of the underlying processes of WM for different auditory information, we compared, to our knowledge for the first time, WM performance for verbal (words) and non-verbal (pitch, timbre) materials in a forward (Experiment 1) and a backward (Experiment 2) recognition task, and in a backward recognition task using articulatory suppression (Experiment 3).

In the following we will review what is known about WM for verbal, tonal, and timbre stimuli. Whereas verbal WM studies have used both recall

and recognition to investigate participants' WM performance, experiments exploring WM for non-verbal materials (tones and timbre) have been relying on recognition tasks (but see Williamson, Baddeley, & Hitch, 2010 for a tonal WM recall task).

While various short-term memory (STM) or WM models have been proposed (for overviews see Baddeley, 2012; Cowan, 1988, 1999; Ericsson & Kintsch, 1995; Jones, 1993; Nairne, 1990), our present paper is based on the Baddeley and Hitch (1974) WM model, which has motivated numerous studies investigating WM for verbal and tonal materials (Baddeley & Hitch, 1974; Berz, 1995; Hickok, Buchsbaum, Humphries, & Muftuler, 2003; Koelsch et al., 2009; Pechmann & Mohr, 1992; Schendel & Palmer, 2007; Schulze, Zysset et al., 2011; Williamson et al., 2010). In addition, although parts of this model are still under debate, no other verbal WM component is as well investigated and accepted as the phonological loop (Buchsbaum & D'Esposito, 2008).

VERBAL INFORMATION

In the influential multi-component WM model proposed by Baddeley and Hitch (1974) verbal information is processed by a phonological loop, which is further subdivided in a passive storage component (phonological store) and an active rehearsal mechanism (articulatory rehearsal process). The passive storage component is assumed to store auditory or speech-based information for a few seconds (Baddeley, 1992).

When verbal information has to be maintained for longer time spans, it is rehearsed by the articulatory rehearsal process[1] (Baddeley, 2003). The following three effects have been interpreted to support the notion that the articulatory rehearsal is comparable to subvocal speech. First, the word length effect refers to the phenomenon that participants show a greater memory span (Baddeley, Thomson, & Buchanan, 1975) and superior recognition accuracy (Baddeley, Chincotta, Stafford, & Turk, 2002) for short words than for long words. This finding has been used as one of the major arguments for articulatory

rehearsal being comparable to subvocal speech, even though some research suggested the additional influence of other word characteristics (i.e., phonological complexity) and attentional mechanisms (Lewandowsky & Oberauer, 2008). Second, it has been shown that the articulatory rehearsal process can be interrupted by articulatory suppression (Baddeley, 1992, 2003; Hall & Gathercole, 2011; Schendel & Palmer, 2007), which prevents the internal rehearsal of verbal material and thus reduces verbal WM capacity. Third, recent neuroimaging studies have provided further evidence for subvocal rehearsal underlying verbal WM. Using a recognition paradigm these studies have reported that active rehearsal of verbal material engages motor-related areas, which are usually involved in controlling and programming of speech movements (Baddeley, 2003; Gruber & von Cramon, 2003; Hickok et al., 2003; Koelsch et al., 2009; Paulesu, Frith, & Frackowiak, 1993; Schulze, Zysset et al., 2011; Smith & Jonides, 1997). Thus participants might use their knowledge of how to produce speech in order to convert the auditorily presented verbal information into internally rehearsable motor representations or sensorimotor codes (Hickok et al., 2003; Koelsch et al., 2009; Schulze, Zysset et al., 2011).

Auditory WM has mainly been investigated using verbal stimuli. Non-verbal auditory information, however, like music, also unfolds over time, and the understanding and appreciation of music depends, as the understanding of speech, on WM.

TONES

In contrast to studies showing the improvement of verbal WM performance thanks to internal rehearsal (for a review see Baddeley, 2003), studies that have investigated whether internal rehearsal of the to-be-remembered stimuli can also improve WM for tones in recognition tasks yielded conflicting results. Some studies showed no beneficial effects of internal rehearsal for WM of tones (Demany, Montandon, & Semal, 2004; Kaernbach & Schlemmer, 2008). In these studies participants might have encountered difficulties in covertly rehearsing the experimental stimuli because the frequency of the tones used did not correspond to the frequencies of the Western chromatic scale or were ambiguous, the frequency difference between tones was smaller than the smallest

---

[1] Refreshing in order to maintain information in WM (Raye, Johnson, Mitchell, Greene, & Johnson, 2007), which is assumed to provide an alternative to the rehearsal suggested by the Baddeley and Hitch WM model, will be discussed later in the manuscript.

difference (a semitone) used in songs of Western tonal music (Kaernbach & Schlemmer, 2008) and/ or chords, consisting of several simultaneously played sine wave tones were used (Demany et al., 2004). In contrast, in studies in which the frequencies of the tones used corresponded to the frequencies of the Western chromatic scale (Pechmann & Mohr, 1992) or in which the frequency differences between tones were not smaller than one semitone (Koelsch et al., 2009; Pechmann & Mohr, 1992), it was observed that participants showed more accurate WM performance if rehearsal was possible, indicating a rehearsal mechanism underlying WM for tones.

Additional corroborating evidence comes from neuroimaging data indicating that during the internal rehearsal of tones motor-related cortical areas are activated, comparable to those activated during the internal rehearsal of verbal material (Hickok et al., 2003; Koelsch et al., 2009; Schulze, Zysset et al., 2011). In sum, these findings suggest that participants are able to translate the auditory signal of certain tones (e.g. corresponding to a known, distinguishable tonal set) into internally rehearsable (sensorimotor) representations, similar to verbal material, that can be used for the maintenance of information in WM (Koelsch et al., 2009; Schulze, Zysset et al., 2011).

COMPARISON BETWEEN VERBAL AND TONAL INFORMATION

The Baddeley and Hitch WM model (Baddeley, 2003; Baddeley & Hitch, 1974) has been designed to explain verbal WM and does not specify whether the phonological loop also processes non-verbal information, or whether different subsystems (a ''musical loop'', Berz, 1995; or a ''tonal loop'', Pechmann & Mohr, 1992) exist in addition to the phonological loop. There is also no consensus in the literature whether verbal and tonal auditory information are processed in one WM system (Chan, Ho, & Cheung, 1998; Semal, Demany, Ueda, & Halle, 1996; Williamson et al., 2010) or in two systems (Deutsch, 1970; Salame & Baddeley, 1989). Furthermore, findings differ as a function of musical expertise, which influences tonal WM: Verbal WM and tonal WM differ more strongly in musicians than they differ in non-musicians (Pechmann & Mohr, 1992; but see Schendel & Palmer, 2007; Schulze, Zysset et al., 2011; Williamson et al., 2010).

TIMBRE

To our knowledge, no study has yet investigated WM for timbre in comparison to WM for both verbal and tonal materials. According to the Acoustical Society, timbre is defined as the features that enable the distinction between two sounds of identical pitch, intensity, duration, and location. For example, an instrument or an individual's voice can be identified by its timbre. Research suggests that participants cannot imitate or repeat timbre (Crowder, 1989). Thus maintaining timbre information in WM should not benefit from internal rehearsal, in contrast to verbal material and tonal material that can be facilitated by internal rehearsal (internal speaking or singing, respectively). The question remains how timbre information can be stored, maintained, and manipulated in WM.

One possibility is that timbre is stored as an ''acategorical'' sensory memory trace (for an overview, see Cowan, 1984, 1988; Winkler & Cowan, 2005). Whereas WM is assumed to be the active, attentional maintenance of information (e.g., via rehearsal), sensory memory refers to a preattentive representation of information (Cowan, 1984; Kaernbach, 2004a, 2004b). Two types of sensory stores have been described: A short sensory store, which is part of perception, and a long sensory store, which is part of memory (Cowan, 1984; Kaernbach, 2004a, 2004b). Sensory memory traces have been traditionally investigated with electrophysiological methods (mismatch negativity: MMN; for an overview, see Schroger, 2007). Results suggest that the long sensory store can hold auditory information for durations of 10 to 20 seconds (Cowan, 2000; Fritz, Elhilali, David, & Shamma, 2007; Kaernbach, 2004b; Sams, Hari, Rif, & Knuutila, 1993). For example, white noise, which cannot be repeated or categorised, can be stored for up to 20 seconds (Kaernbach, 2004b). Another indication that timbre is stored using a sensory trace comes from the Timbre Memory Model (TMM; McKeown & Wellsted, 2009; Mercer & McKeown, 2010). This model assumes that distractors that share features with stimuli stored in memory are likely to overwrite these stored stimuli and thus degrade their memory trace (McKeown & Wellsted, 2009; Mercer & McKeown, 2010), but that articulatory suppression does not have the same degrading effect: McKeown, Millis, and Mercer (2011) designed an

experiment in which participants were asked to compare two complex periodic sounds that could not be verbally labelled. These two sounds were separated by a retention interval lasting 5 to 30 seconds. Even though participants were reading aloud during the retention interval, performance was robust, also for the longer interval. The authors concluded that this form of auditory memory does not depend on verbal rehearsal.

Overall, findings suggest that timbre information might be stored as acategorical information using sensory memory traces (Cowan, 1984; Kaernbach, 2004a, 2004b; McKeown et al., 2011; McKeown & Wellsted, 2009; Mercer & McKeown, 2010).

COMPARISON BETWEEN TIMBRE AND TONES/WORDS

Memory for timbre has mainly been investigated in comparison to memory for tones using interference experiments. That is, it was investigated whether the presentation of timbre stimuli interferes with WM for tones or vice versa. Semal and Demany (1991) reported that WM for periodic complex tones was only influenced by pitch, but not by timbre. These results, suggesting two independent WM systems for pitch and timbre, were supported by (1) Krumhansl and Iverson (1992) who observed that timbre variations did not influence pitch memory, even though memory for timbre was slightly influenced by pitch, and also by (2) Starr and Pitt (1997), who reported that memory for timbre was only minimally disrupted by irrelevant pitch variation. Overall it has been proposed that pitch and timbre are processed independently in WM recognition tasks (Krumhansl & Iverson, 1992; Semal & Demany, 1991; Starr & Pitt, 1997). This might be due to the fact that participants can sing or hum a melody and can therefore rehearse pitch information internally (Hickok et al., 2003; Koelsch et al., 2009; Schulze, Zysset et al., 2011), while they might have severe difficulties to sub-vocally repeat or produce timbre information (Crowder, 1989; Halpern, Zatorre, Bouffard, & Johnson, 2004).

However, it has been shown that participants are able to imagine timbre, which might be based on the creation of a sensory image. Crowder (1989) asked participants to indicate whether two tones, played by a guitar, flute, or trumpet, were the same or different in pitch. In a timbre imagery task participants were presented with a sine wave tone, and were then asked to imagine this tone being played by a guitar, a flute, or a trumpet. Then the second tone, played by a guitar, flute, or trumpet timbre, was presented, and participants had to decide whether the two tones differed in pitch. Interestingly, participants were slowed down in this judgement when the second tone was played by an instrument that differed from the instrument that they imagined during the imagery condition. Although this finding suggested that participants might be able to imagine timbre, it is worth noting that, first, participants only had to imagine one timbre stimulus, and second, they had as much time as they needed to do so as this experiment was self-paced.

Further findings support the hypothesis that WM for timbre does not rely on an internal rehearsal mechanism, but rather on a sensory imagery mechanism. Functional imaging data have shown for tones that imagery (Halpern & Zatorre, 1999) and internal rehearsal (Koelsch et al., 2009; Schulze, Zysset et al., 2011) rely on motor-related neural structures, indicating that production systems are involved during these processes. However, for timbres the involvement of motor-related structures has not been observed during an imagery task (Halpern et al., 2004).[2] Although there has been only little research on WM for timbres, results so far suggest that WM for timbres differs from that for tones and words. It has been proposed (for overviews see, Lewandowsky & Oberauer, 2008; Raye et al., 2007) that the process of refreshing, like internal rehearsal, can actively maintain verbal information in WM. Refreshing involves an attention-based augmentation and maintenance of a memory trace, and could be used to maintain timbre information in auditory WM. However, for timbre, McKeown et al. (2011) argued that refreshing is not responsible for the robust memory trace over time, notably because reading during the retention delay did not impair WM performance for auditory stimuli that could not be verbally labelled.

FORWARD AND BACKWARD WM RECOGNITION

WM enables the temporally limited storage and maintenance (e.g., by active rehearsal) as well as

---

[2] Activation of the supplementary motor area did not reach significance (Halpern et al., 2004).

the manipulation of information (Baddeley, 2003; Baddeley & Hitch, 1974). While maintenance is sufficient for forward WM tasks, manipulation (i.e., reordering of the elements of the sequence) is required for backward WM tasks (Zatorre, Halpern, & Bouffard., 2009). Previous studies have shown that participants perform better during forward recall than during backward recall (e.g., Farrand & Jones, 1996; Hulme et al., 1997). Manipulation appears to rely on a different mechanism or to require additional processes compared to maintenance only (as indicated by studies showing differences between forward and backward recall for verbal stimuli; Bireta et al., 2010; Hulme et al., 1997; Surprenant et al., 2011; Tehan & Mills, 2007), for example by involving processes of the central executive (Baddeley, 2003, 2012; Baddeley & Della Sala, 1996). In contrast to recall tasks, it is still not well known whether forward and backward WM (thus maintenance and manipulation) of verbal and nonverbal materials might differ in recognition tasks. Therefore our present study investigated WM for auditory stimuli using forward and backward WM recognition paradigms.

MEMORY FOR SERIAL ORDER

One component not included in the Baddeley and Hitch WM model (Baddeley & Hitch, 1974) is a timing component; that is, how participants re-member the order of items. The existence of a ''timing signal'' that supports memory for serial order has been postulated by different models (Brown, Preece, & Hulme, 2000; Burgess & Hitch, 1999), in which it is assumed, that memory for item and memory for order information are based on different processes. Henson, Hartley, Burgess, Hitch, and Flude (2003) investigated memory for items and serial order using visually presented letters in a WM recognition paradigm. Whereas in the task investigating memory for items partici-pants were asked to indicate whether a control item had been presented in a previous sequence, in the task investigating memory for serial order participants were asked to indicate whether a sequence of items was presented in the same order as before (thus being comparable to the forward task used in our study). Irrelevant speech, articu-latory suppression, and paced finger tapping dur-ing the presentation of the to-be-remembered visual items impaired memory for serial order more strongly than memory for items. These findings suggested that (i) memory for serial order and memory for items are based on separate processes and that (ii) memory for serial order relies to a greater degree on internal rehearsal. Memory for serial order, compared to memory for items, involved the left dorsal premotor cortex more strongly, a brain structure also involved in rhythm processing (Henson, Burgess, & Frith, 2000).

Although the timing component has not been investigated as systematically for the auditory domain as it has been investigated for the visual domain (e.g., Henson et al., 2000; Henson et al., 2003), previous research suggested that timing and ordering mechanisms also underlie verbal and non-verbal WM. In particular, WM for rhythmic patterns (tones presented auditorily) has been shown to share mechanisms with verbal WM (Hall & Gathercole, 2011; Saito, 1994, 2001; Saito & Ishio, 1998), whether auditorily presented digits or letters (Hall & Gathercole, 2011; Saito, 2001) or visually presented letters (Saito, 1994). Thus auditory WM has been suggested to involve timing control mechanisms as well (Hall & Gathercole, 2011; Saito, 2001).

However, recognition memory for serial order has, to our knowledge, only been compared to memory for items for visual material in a forward task (i.e., requiring maintenance only), thus call-ing for research investigating memory for serial order using auditory materials in forward and backward recognition tasks (as used in our study). This is especially the case for the backward task, because this task requires participants to reorder, that is to manipulate, the presented items. Mem-ory for serial order relies more strongly on internal rehearsal than memory for items (Henson et al., 2003), thus using this task allowed us to compare WM rehearsal processes for different auditory stimuli.

SEQUENCE LENGTH

Baddeley (2003) suggested that the amount of material that can be rehearsed and stored in WM is restricted; the longer the sequence (or the more events there are), the weaker the WM perfor-mance (Cowan, 2000). It has been shown that certain characteristics of auditory WM depend on sequence length. For example Williamson et al. (2010) reported that the pitch proximity effect (better recall performance for dissimilar com-pared to similar tone sequences) decreases as

sequence length increases. Furthermore, Schulze, Dowling, and Tillmann (2012) showed that the tonality effect (recognition performance in a forward task was better for tonal compared to atonal sequences) was modulated by the length of the sequences. The length effect seems to be further influenced by the task (recall vs recognition; forward vs backward) and/or by the auditory stimulus type: For verbal WM a length effect has been observed in a forward recall task (Baddeley et al., 1975; Bireta et al., 2010; Surprenant et al., 2011), but not in a backward recall task (Bireta et al., 2010; Surprenant et al., 2011; Tehan & Mills, 2007), while for tonal material a length effect has been observed in a forward recall task (Williamson et al., 2010) and in both forward (Croonen, 1994; Schulze et al., 2012) and backward recognition tasks (Schulze et al., 2012). Our present study investigated whether a length effect would be observed for different auditory materials in forward and backward recognition tasks by using two sequence lengths: five- and six-item sequences in the forward task, and three- and four-item sequences in the backward task.[3]

To summarise, our study investigated WM for three types of materials (words, tones, timbre) using a recognition task for serial order. This combination allowed us to compare WM for verbal material (words) and non-verbal (tone, timbre) materials, as well as WM for which internal rehearsal has been reported previously (words, tones) or not (timbres). Furthermore we manipulated sequence length and used both forward (Experiment 1) and backward (Experiment 2) recognition tasks in order to study WM with its components of maintenance and manipulation. In Experiment 3 we investigated the influence of articulatory suppression on WM for words, tones, or timbres.

We hypothesised that if WM for all auditory stimuli relies on the same mechanisms, then the same data pattern should be observed across material types, e.g., decreased performance for longer sequences compared to shorter sequences (Baddeley, 2003; Croonen, 1994; Schulze et al., 2012; Williamson et al., 2010). However, if WM for the three auditory stimuli relies on different mechanisms, then an interaction between length

and material during the forward task and/or the backward task was expected. Differences might emerge because participants are able to rehearse words and tones internally (based on their knowledge how to produce words and tones), and these motor representations could support WM performance (Halpern & Zatorre, 1999; Koelsch et al., 2009; Schulze, Zysset et al., 2011), while being subjected to length effects (Baddeley et al., 1975; Bireta et al., 2010; Croonen, 1994; Schulze et al., 2012; Surprenant et al., 2011; Williamson et al., 2010). This might not be the case for timbres, and participants might rely on sensory memory rather than on internal rehearsal (Crowder, 1989; Halpern et al., 2004; Kaernbach, 2004b; McKeown et al., 2011; McKeown & Wellsted, 2009; Mercer & McKeown, 2010).

EXPERIMENT 1 (FORWARD TASK)

Using a forward recognition task, in Experiment 1 we investigated WM for words, timbres, and tones. We expected decreased performance for longer sequences (i.e., length effect) for words (Baddeley, 2003; Baddeley et al., 1975) and tones (Croonen, 1994; Schulze et al., 2012; Williamson et al., 2010), which both might rely on rehearsal mechanisms, but probably not for timbre, which is presumably based on sensory memory traces (Kaernbach, 2004b; McKeown et al., 2011; McKeown & Wellsted, 2009; Mercer & McKeown, 2010). Previous literature suggests that participants perform better for words compared to tones based on their everyday life experience and training with verbal stimuli (Koelsch et al., 2009; Schulze, Zysset et al., 2011). Alternatively, the additional contour information and tonal-like structure might lead to better performance for tones than for words (Schulze et al., 2012; Tillmann et al., 2009). For the timbre stimuli we hypothesised that participants showed, compared to words and tones, a decreased performance level.

Method

Participants. A total of 20 participants (12 female) took part in Experiment 1. The mean age was 22.65 years ($SD = 3.01$ years; age range: 18–28). Number of years of musical training, as measured by years of instrumental instruction, ranged from 0 to 10, with a mean of 2.08

---

[3] In a pilot experiment we observed that participants showed decreased performance during the backward task compared to the forward task. To account for this and to avoid floor effects in performance, we presented participants with shorter sequences in the backward task.

($SD = 3.01$) and a median of 0. Of these participants, 12 had not received any musical instruction (0 years).

Materials. Participants were presented with auditory sequences, consisting of either five or six auditory items. Each item had a duration of 500 ms. In the sequences the items were presented with an inter-stimulus interval (ISI) of 20 ms, resulting in an stimulus-onset asynchrony of 520 ms. For the pitch task six tones were used. These tones differed in pitch height, namely C4 (262 Hz), D4 (294 Hz), E4 (330 Hz), F4 (349 Hz), G4 (392 Hz), and A4 (440 Hz). All tones were generated using a cello timbre. For the timbre task we used six timbres (guitar, cello, flute, trumpet, vibes, and piano), all played at 330 Hz (E4). For the word task six monosyllabic meaningful French words (consonant – vowel combinations, using the vowel /u/) were used (/tu/, /lu/, /bu/, /mu/, /gu/, and /pu/). All words were spoken by a female voice and recordings were adjusted to the pitch of 230 Hz with "STRAIGHT" (Kawahara & Irino, 2004). The words were selected from a pool of recorded words on the basis of subjective ratings indicating easy intelligibility (i.e., using a subjective scale from 1 (very easy to understand) to 5 (not easy to understand)) by eight native French speakers. To account for the phonological similarity effect (Conrad & Hull, 1964), phonologically similar monosyllabic words were chosen, so that timbre stimuli did not sound more similar to each other than the words.[4]

For the different pairs of the pitch task two non-adjacent tones were exchanged aiming to preserve the melodic contour (e.g., A G *F* C *D* – A G *D* C *F*). As in the backward task (see Experiment 2) the first tone of the second sequence was never changed. For the different sequence pairs of all six-item sequences the melodic contour was preserved. However, it was not possible to preserve the melodic contour in 3 of the 14 five-item sequences (e.g., G D E **_F A_** – G D E **_A F_**), because of the constraint to keep the first tone unchanged.

The pattern of how the item order was changed in the different sequence pairs of the pitch sequences (i.e., the positions of the changed

items) served as a template to generate the different sequence pairs for the timbre (cello guitar **_flute_** piano **_trumpet_** – cello guitar **_trumpet_** piano **_flute_**) and the word (/bu//gu//**_mu_**//tu//**_pu_**/ – /bu//gu//**_pu_**//tu//**_mu_**/) conditions.

Example sequences for each stimulus type are provided as Supplementary Material, which is available via the supplementary tab on the article's online page at http://dx.doi.org/10.1080/09658211.2012.731070.

Apparatus. The tone and timbre stimuli were created with the software Cubase 5.1 and Halion Sampler (Steinberg Media Technologies). The word stimuli were spoken by a female voice and recorded using the software Audacity (http://audacity.sourceforge.net/). The program Adobe Audition 3 (San Jose, CA, USA) was used to slightly adapt the duration of all auditory stimuli to 500 ms. All stimuli were saved as 44.1 kHz, 16-bit resolution mono files. Participants were presented with the auditory stimuli via closed headphones (Sennheiser HD200). The software Presentation (Neurobehavioural Systems, Albany, USA) was used to present the stimuli, collect data, and record participants' responses.

Procedure. Participants listened to sequences of auditory stimuli, either tones differing in pitch (*pitch task*), tones differing in timbre (*timbre task*), or monosyllabic words (*word task*). A first sequence (e.g., 1 2 3 4 5) was presented and after 3 seconds of silence a second sequence was presented with all auditory items being either in the same order (e.g., 1 2 3 4 5) or not (e.g., 1 2 5 4 3). Participants pressed one of two mouse buttons to indicate whether the two sequences were the same or different (with "same" being defined as all items played correctly in the same order). Subsequently participants pressed the space bar to continue with the next trial of this self-paced experiment.

The task was demonstrated with two five-item sequence trials for each material at the beginning of the first three blocks. For these practice trials (one required the answer "same" and one the answer "different"), feedback was given, and it was ascertained that participants understood the task.

The experiment consisted of 168 trials: 28 pairs for each Material (words, timbre, pitch) and Length (five- and six-item sequences), with half of them being different and half being the same. A pseudo-randomised presentation was used so that: (1) the same sequence was not presented consecutively in a same pair and a different pair

---

[4] Timbre stimuli were chosen so that they sounded as dissimilar as possible and were easy to distinguish. However, one might argue that due to the nature of the timbre material, the timbre stimuli might still sound somewhat similar to each other. Therefore, to account for this, phonologically similar monosyllabic words were chosen.

and (2) the type of pair (same/different) changed after at most three trials (i.e., no more than three consecutive ''same'' or ''different'' trials). The experiment was structured into six experimental blocks of 28 pairs each (two blocks for timbre, two blocks for pitch, and two blocks for words). The blocks always started with the shorter sequences (five-item sequences). Over participants, six different orders of blocks were used (differing in the order of presentation within the blocks and the order of the experimental blocks), which were counterbalanced over participants. The experiment had a duration of approximately 45 minutes.

Results

Recognition performance was analysed by calculating the hit rate (number of correct responses for different trials/number of all different trials) minus the false alarm rate (number of incorrect responses for same trials/number of all same trials) for each participant and each condition. Performance was significantly better than chance for each of the conditions as shown by one-sample t-tests ($ps < .001$).

Hits – false alarms measures are depicted in Figure 1 and were analysed using a 2×3 ANOVA with Length (five and six) and Material (pitch, timbre, words) as within-participant factors.

The main effect of Material was not significant, $F(2, 18) = 1.94$, $p = .16$, $MSE = .07$. The main effect of Length, $F(1, 19) = 7.50$, $p = .01$, $MSE = .03$, was significant and interacted with Material, $F(2, 18) = 5.20$, $p = .01$, $MSE = .03$. Performance was better for the five-item sequences than for the

six-item sequences for words, $t(19) = 2.40$, $p = .03$, and pitch, $t(19) = 3.81$, $p < .01$, while for timbres no difference between sequence lengths was observed, $t(19) = .68$, $p = .50$. Furthermore, for the five-item sequences, performance for pitch was higher than for timbre, $t(19) = 2.67$, $p = .02$) and words, $t(19) = 2.38$, $p = .03$, while these latter two did not differ ($p = .59$). For the six-item sequences no significant differences between materials were observed ($ps > .16$).

Recognition performance was further investigated by analysing hit rates and false alarm rates separately (see Supplementary Material).

No significant correlation (Pearson) was observed between the number of years of musical training and the performance for pitch, $r(18) = .027$; $p = .911$, timbre, $r(18) = .310$; $p = .183$, and words, $r(18) = -.007$; $p = .976$.

An additional analysis was performed only with the contour-preserved sequences in the pitch conditions (i.e., excluding the three contour violated five-item sequences, see Method). This analysis confirmed the main effect of Length, $F(1, 19) = 4.91$, $p = .039$, $MSE = .03$, and its interaction with Material, $F(2, 18) = 3.83$, $p = .03$, $MSE = .03$, as described for the main analysis.

DISCUSSION

In Experiment 1 we investigated WM maintenance for words, timbre, and tones using five- and six-item sequences. Participants' performance for the different lengths depended on the to-be-remembered material. For the pitch and word conditions an influence of length was observed,
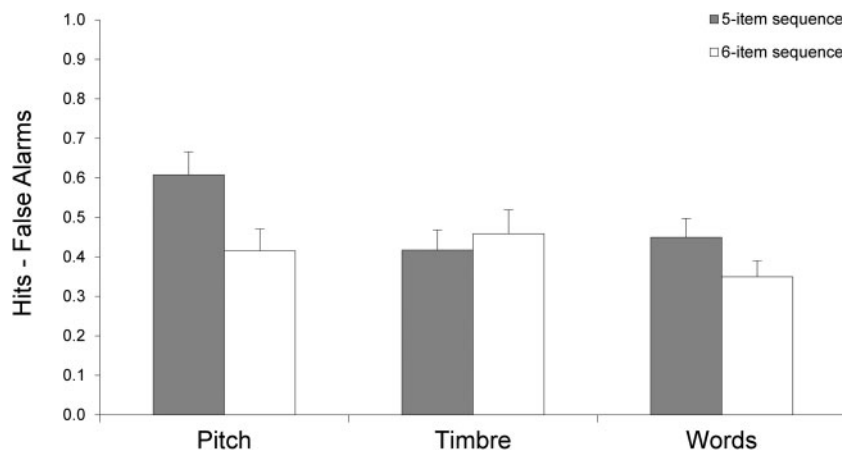
Figure 1. WM performance during the forward WM recognition task for pitch, timbre, and word information (Experiment 1). Error bars indicate the standard error of mean (SEM).

notably with better performance for shorter than for longer sequences. This finding is in agreement with previously reported length effects for verbal material (Baddeley, 2003; Baddeley et al., 1975; but see Lewandowsky & Oberauer, 2008) and tonal material (Croonen, 1994; Schulze et al., 2012; Williamson et al., 2010), indicating that verbal and tonal WM are subject to similar limiting factors, hence similar processes, notably linked to rehearsal mechanisms (Baddeley, 2003). More specifically, it has been suggested that WM for words and tones can be based on subvocalisation or internal singing (Baddeley, 2003; Baddeley et al., 1975; Gruber & von Cramon, 2003; Hickok et al., 2003; Koelsch et al., 2009; Paulesu et al., 1993; Schulze, Zysset et al., 2011) and sensorimotor codes might be used to actively rehearse the information in WM (Koelsch et al., 2009; Schulze, Zysset et al., 2011).

Beyond this similarity in the benefits of rehearsal, the verbal and tonal materials differ on the following features that might have increased WM performance for tones (in particular for the five-item sequences). First, the tones of a sequence define a melodic contour (i.e., patterns of up-and-downs) that can be used as an additional information to increase WM performance of the tones (Dowling & Fujitani, 1971).[5] Second, the tone set used is part of a tonal set of Western music, the C Major key, a stimulus-inherent structure that might create associations between elements (based on tonal relations) and allow participants to chunk the to-be-remembered information (Gobet et al., 2001) and thus decrease WM load. The improvement of memory thanks to stimulus-inherent structure has been shown previously for tone sequences (Schulze et al., 2012; Schulze, Mueller et al., 2011), but also for word lists (Savage et al., 2001; Tulving, 1962) and for spatial patterns (Bor, Duncan, Wiseman, & Owen, 2003). Stimulus-inherent structure, which could improve WM performance, was not included in our word material. In contrast, phonologically similar words were used (see Method), with phonological similarity being known to decrease WM performance (Baddeley, 2003; Conrad & Hull, 1964; Surprenant, Neath, & LeCompte, 1999; Williamson et al., 2010).

By comparing verbal and tonal memory to timbre memory in the same participants, our

study went beyond previous research that was restricted to two material types. In contrast to performance for words and tones, performance for timbres did not differ between shorter and longer sequences. This finding might suggest that the mechanisms underlying WM for timbre differ from those underlying WM for tones and words. In our study the duration of one trial (sequence 1 – delay – sequence 2) was ∼8 seconds for the five-item sequences and ∼9 seconds for the six-item sequences. Studies investigating the auditory long sensory store have indicated that participants can hold auditory traces for up to 20 seconds (see Cowan, 1984), e.g., for non-imitable white noise segments (Kaernbach, 2004b). Therefore maintenance of timbre information in our experiment might have been mainly dependent on a sensory memory trace (Crowder, 1989; Fritz et al., 2007; Halpern et al., 2004; Kaernbach, 2004b; Pitt & Crowder, 1992; Sams et al., 1993), in particular as the sequences were short enough to be compared in full length or being reactivated by refreshing (Lewandowsky & Oberauer, 2008; Raye et al., 2007), thus not revealing a length effect. Our interpretation, that WM for timbre is independent of WM for verbal stimuli, is corroborated by a recent study (McKeown et al., 2011) reporting that WM for timbre is not facilitated by verbal rehearsal processes.

EXPERIMENT 2 (BACKWARD TASK)

To summarise Experiment 1, the observed length effect for words and tones, together with the literature about the limited capacity of WM for verbal material (Baddeley, 2003; Baddeley et al., 1975) and tone material (Croonen, 1994; Williamson et al., 2010) supports the hypothesis that performance for these two stimulus types were based on similar WM mechanisms related to rehearsal. For timbre participants might have relied more strongly on a sensory memory trace, notably because timbre is assumed to be difficult to imitate and thus to rehearse subvocally (Crowder, 1989; Fritz et al., 2007; Halpern et al., 2004; Kaernbach, 2004b). In Experiment 2 we investigated whether the same pattern of results can be observed for WM for words, tones, and timbre during a backward recognition task that required the reordering of the items. In contrast to maintenance (as in the forward task, Experiment 1), only a few studies have investigated the manipulation of auditory events in WM. While

---

[5] WM performance is improved when to-be-detected differences change the melodic contour in comparison to differences that preserve the contour.

both verbal and tonal WM rely on internal rehearsal, participants might be more trained on the manipulation of verbal information in WM (Koelsch et al., 2009; Schulze, Zysset et al., 2011). Better performance was expected for words compared to tones, and both materials should be subjected to length effects (with weaker performance for longer sequences; Baddeley, 2003; Baddeley et al., 1975; Croonen, 1994; Schulze et al., 2012; Williamson et al., 2010). In addition, manipulating tone sequences (as requested in the backward task) has been shown to be a difficult task for participants, presumably due to the difficulty of reversing the contour of the perceived melody (Schulze et al., 2012).

Because timbre does not seem to be supported by internal rehearsal in WM (Crowder, 1989; Halpern et al., 2004; Kaernbach, 2004b; McKeown et al., 2011; McKeown & Wellsted, 2009; Mercer & McKeown, 2010), we hypothesised that WM for timbre would be inferior to WM for the other two stimuli types (words, tones). To our knowledge no study so far has investigated manipulation of timbre information in WM, as requested here in the backward recognition task. Our study thus tested for the first time whether the additional manipulation of the timbre information stored in the sensory memory might give rise to a length effect also for timbres comparable to tones (Croonen, 1994; Schulze et al., 2012; Williamson et al., 2010) and words (Baddeley et al., 1975; Bireta et al., 2010; Surprenant et al., 2011), as for example due to the intervention of the central executive (Baddeley, 2003, 2012; Baddeley & Della Sala, 1996).

Method

Participants. A total of 20 participants (13 female) took part in Experiment 2. The mean age was 21.65 years ($SD = 4.01$ years; age range: 18–34). Number of years of musical training, as measured by years of instrumental instruction, ranged from 0 to 17, with a mean of 1.90 ($SD = 4.2$) and a median of 0. Of these participants, 13 had not received any musical instruction (0 years). None of the participants took part in Experiment 1. The forward and backward tasks were run in separate experiments, with different participants taking part, to prevent potential carry-over effects of strategies between the instructions for both tasks.

Years of musical training (years of musical instruction) of participants in Experiments 1 and 2 did not differ ($p > .88$).

Material and apparatus. Sequences consisted of three or four items. For "different" trials the exchanged tones (see Method for Experiment 1) introduced a contour change in all pitch sequences (e.g., E F C and C E F). The contour of the sequences in the different trials could not be preserved because of the short sequences (i.e., the three-item sequences) and the constraint to keep the first tone unchanged. (As in Experiment 1, the pattern of how the item order was changed in the different trials of the pitch sequences served as a template to generate the different trials for the timbre and the word condition.) Besides this, materials and apparatus were as described for Experiment 1. Example sequences for each stimulus type are provided as Supplementary Material.

Procedure. Participants listened to the auditory stimuli as in Experiment 1, either tones differing in pitch (*pitch task*), tones differing in timbre (*timbre task*) or monosyllabic words (*word task*). A first sequence was presented (e.g., 1 2 3) and after a period of silence (3 seconds) participants heard a second sequence that was either the same or had a changed order (two items were exchanged). Participants had to indicate whether the second sequence had all items in a backward order (e.g., same: 3 2 1) or whether this order had been altered (e.g., different: 3 1 2). To prevent participants from solving the task by comparing only the last stimuli of the first sequence with the first stimuli of the second sequence to decide whether the sequence had been played backwards correctly, the first stimuli of the second sequence was never changed (as in Experiment 1). Besides presenting the second sequence backwards and using different sequence lengths, the procedure was as described for Experiment 1.

As with Experiment 1, Experiment 2 consisted of 168 trials overall: 28 pairs for each Material (words, timbre, pitch) and Length (three- and four-item sequences), with half of them being different and half being the same. The pseudo-randomised presentation and the presentation of the experimental items in four blocks were as described in Experiment 1 (adapted to the sequence lengths used here).

Results

As with Experiment 1, recognition performance was analysed by calculating the hit rate (number of correct responses for different trials/number of all different trials) minus the false alarm rate (number of incorrect responses for same trials/number of all same trials) for each participant and each condition. Performance was significantly better than chance for each of the conditions as shown by one-sample t-tests (all $ps < .001$, except for the four-item pitch sequences, $p = .06$).

Hits – false alarms measures are depicted in Figure 2 and were analysed using a 2×3 ANOVA with Length (three- and four-items) and Material (pitch, timbre, words) as within-participant factors. Main effects of Length, $F(1, 19) = 38.99$, $p < .001$, MSE $= .03$, and Material, $F(2, 18) = 7.82$, $p = .001$, $MSE = .08$, were observed, but no significant interaction between Length and Material ($p = .19$). Performance was better for timbre, $t(19) = 2.56$, $p = .02$, and for words, $t(19) = 3.49$, $p = .002$, compared to pitch, while no performance difference was observed between timbre and words ($p = .11$). Performance was better for the three-item sequences than for the four-item sequences, $t(19) = 6.24$, $p < .001$.

Recognition performance was further investigated by analysing separately hit rates and false alarm rates (see Supplementary Material).

We observed a significant positive Pearson correlation between musical training and the performance for the pitch task, $r(18) = .653$; $p = .002$ (indicating better performance for participants with more years of musical training), a tendency for the timbre task, $r(18) = .438$; $p = .053$, but no correlation for the word task, $r(18) = .349$; $p = .132$.

Discussion

In Experiment 2 we investigated manipulation of auditory information in WM for words, timbres, and tones for three- and four-item sequences. In this backward recognition task performance was better for words compared to tones, while in the forward recognition task of Experiment 1 performance was better for tones compared to words. Two possible hypotheses will be discussed aiming to explain these different patterns in WM performance.

In contrast to the forward task, data of the backward task suggest that the stimulus-inherent features of the tone sequences (contour information and tonal-like structure) did not support WM for tones when reordering was required. This has also been observed in a previous study where tonality improved WM performance in the forward task, but not in the backward task (Schulze et al., 2012). Musical information unfolds over time in a structured, directional way. Backward recognition requires the reversal of this structure (Zatorre et al., 2009), thus the breaking up of a melody (Schulze et al., 2012), so that contour information and the tonal-like structure might not have been supporting the WM recognition process. This could have led to a decreased performance for tones compared to words in Experiment 2.
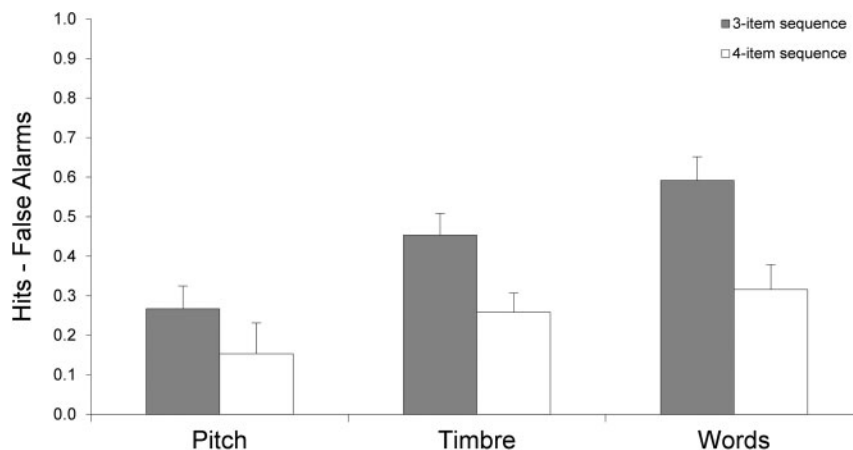
Figure 2. WM performance during the backward recognition task for pitch, timbre, and word information (Experiment 2). Error bars indicate the standard error of mean (*SEM*).

A second hypothesis is based on neuroimaging data: Previous studies have indicated that in musicians and non-musicians active rehearsal of verbal and tonal materials relies on motor-related areas, which are usually involved in controlling and programming of movements and therefore on internally rehearsable sensorimotor codes (Koelsch et al., 2009; Schulze, Zysset et al., 2011). It has also been reported that these motor-related areas were less activated during tonal WM than during verbal WM in non-musicians (Schulze, Zysset et al., 2011), whereas motor-related areas were involved during verbal and tonal WM in musicians (Schulze, Zysset et al., 2011) and during the mental reversal of imagined melodies in musicians (Zatorre et al., 2009). These previous findings have suggested that non-musicians have more elaborate sensorimotor representations for the rehearsal of verbal information than for the rehearsal of tonal information. Participants might have been able to benefit from these motor representations for verbal material even during manipulation (required by the backward task), leading to better WM performance for words compared to tones in Experiment 2. This assumption is corroborated by the finding that musical training was only correlated with WM performance in the backward task for pitch.

Despite this reported different pattern (tonal > verbal during forward and verbal > tonal during backward), both pitch and word materials showed a length effect in the backward task, i.e., performance was better for the shorter compared to the longer sequences. In contrast to the forward task, such a length effect was also observed for the timbre material in the backward task. Although the speed of presentation made it unlikely, and recent data have shown that WM for timbre is independent of verbal rehearsal (McKeown et al., 2011), it could be argued that participants actually manipulated timbre information in the backward task by verbally labelling the timbre stimuli. Experiment 3 was conducted to investigate this hypothesis.

EXPERIMENT 3 (BACKWARD SUPPRESSION TASK)

To investigate whether participants used verbal labels to manipulate timbre information in WM Experiment 3 used the backward task of Experi-

ment 2, but introduced an articulatory suppression task during the delay: Participants were required to count aloud from 1 to 5 during the 3-second period of silence between the to-be-compared auditory sequences (as in Baddeley et al., 1975; Henson et al., 2003; Surprenant et al., 1999).

Articulatory suppression has been shown to lead to a reduction of the verbal WM capacity. This reduction has been interpreted as being caused by the additional competing processes that are disturbing the articulatory rehearsal process of the phonological loop (Baddeley, 1992, 2003). Based on previous interference and articulatory suppression experiments (Deutsch, 1970; Gruber & von Cramon, 2001; Koelsch et al., 2009; Pechmann & Mohr, 1992; Schendel & Palmer, 2007), we expected the verbal suppression to interfere with WM for words (verbal rehearsal), but not for tones. In particular, while both materials are based on internal rehearsal (Hickok et al., 2003; Koelsch et al., 2009; Schulze, Zysset et al., 2011), they seem to rely on different sensorimotor codes (Koelsch et al., 2009; Schulze, Zysset et al., 2011). If participants used verbal labelling for timbres in the backward task (Experiment 2), then articulatory suppression should also interfere with timbre. If timbre is not decoded verbally, but rather maintained and manipulated relying on a sensory trace and the central executive (Baddeley, 2003, 2012; Baddeley & Della Sala, 1996), we should not observe interference by articulatory suppression (Crowder, 1989; Halpern et al., 2004; Kaernbach, 2004b; McKeown et al., 2011; McKeown & Wellsted, 2009; Mercer & McKeown, 2010). The result for timbres will thus indicate whether the observed sequence length effect in Experiment 2 might be due to shared verbalising processes or to shared processes for instance on the level of the central executive, which contributes to manipulation and applies to various material types.

Method

Participants. A total of 25 participants (13 female) took part in Experiment 3. The mean age was 22.12 years ($SD = 1.45$ years; age range: 20–26). Number of years of musical training, as measured by years of instrumental instruction, ranged from 0 to 10, with a mean of 3.40 ($SD = 3.76$) and a median of 3. Of these participants, 11 had not received any musical instruction (0 years).

None of the participants took part in Experiments 1 or 2. Years of musical training (years of musical instruction) did not differ between participants of Experiment 3 and 2 ($p > .2$).

Materials, apparatus, and procedure. Because performance for the four-item sequences was relatively low in Experiment 2, only three-item sequences were used in this articulatory suppression task to prevent floor effects in performance. Besides this, stimuli and experimental task were identical to Experiment 2. The experiment had a duration of approximately 25 minutes. In the 3-second silence period between the standard sequence and the comparison sequence, participants were asked to count out loud from 1 to 5, which was recorded using a microphone. After the presentation of the first sequence the fixation cross in the middle of the computer screen started flashing (at the same rhythm as the stimuli were presented before) to indicate beginning and end of counting, as well as to provide a temporal pace for counting.

Results and discussion

As in Experiments 1 and 2, recognition performance was analysed by calculating the hit rate (number of correct responses for different trials/ number of all different trials) minus the false alarm rate (number of incorrect responses for same trials/number of all same trials) for each participant and each condition (Figure 3). Performance was significantly better than chance for each of the conditions as shown by one-sample t-tests (all $ps < .001$).

Hits – false alarms measures were analysed using an ANOVA with the within-participant factor Material (pitch, timbre, words). The main effect of Material was significant, $F(2, 23) = 6.20$, $p < .01$, $MSE = .07$. As in Experiment 2, performance was better for timbres (mean $= .40$, $SEM = .06$), $t(24) = 2.83$, $p = .01$, and words (mean $= .43$, $SEM = .05$), $t(24) = 3.18$, $p < .01$, than for pitch (mean $= .19$, $SEM = .06$), while the performance for words and timbre did not differ, $t(24) = .48$, $p = .63$. We observed a significant positive Pearson correlation between musical training and the performance for the pitch task, $r(23) = .530$; $p = .006$ (indicating better performance for participants with more years of musical training), and no correlation for the timbre task, $r(23) = .090$; $p = .670$, or the word task, $r(23) = -.244$; $p = .239$.

Recognition performance was further investigated by analysing hit rates and false alarm rates separately (see Supplementary Material). In comparison to the three-item sequence in Experiment 2 (Figure 2), the suppression task created a significant decrease of performance only for the words, $t(43) = 2.11$, $p = .04$, but not for timbre, $t(43) = .74$, $p = .46$, or pitch, $t(43) = .88$, $p = .37$. In summary, only performance for words was affected by the verbal interference task. We did not observe a similar decrease in WM for timbre stimuli, suggesting that participants did not maintain and manipulate timbre stimuli by assigning verbal labels to them.

GENERAL DISCUSSION

In the present study we compared recognition WM performance for words, pitch, and timbre using forward and backward recognition tasks. Both forward and backward tasks require the
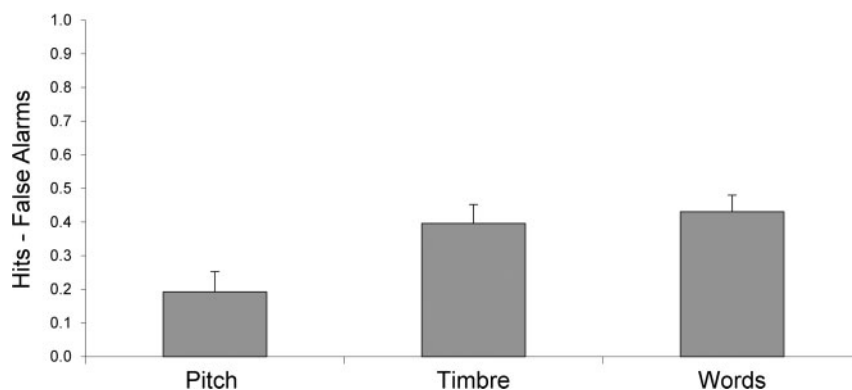
Figure 3. WM performance during the backward suppression recognition task for pitch, timbre, and word information with articulatory suppression (Experiment 3). Error bars indicate the standard error of mean (*SEM*).

memorisation of the serial order of the auditory items (e.g., Henson et al., 2000, 2003). This is enhanced for the backward task that requires participants to reorder and manipulate the presented items. The following main results were observed: (1) in the forward task performance was better for tones than for words (in particular for the five-item sequences), but no difference was observed between words and timbres, (2) in the backward task performance was better for words than for tones, and again no difference was observed between words and timbres, (3) in the backward task articulatory suppression led to impaired performance for words, but not for timbres and tones, (4) better performance for short compared to long sequences was observed for words and tones (but not for timbres) during the forward task, and for all stimuli during the backward task, and (5) a positive correlation between musical training and WM performance for pitch in the backward task, but not in the forward task and not with the other stimuli in either task.

Comparison between tones and words

Our data indicate differences and similarities for WM for tones and words. First we will discuss the similarities, and subsequently the differences.

In Experiments 1 and 2 we observed a length effect for words and tones, suggesting (at least partly) similar processes for both stimuli types. Words and tones might be remembered using an internal rehearsal (internal speaking/singing) mechanism, based on motor knowledge (sensorimotor codes) to produce these stimuli, as suggested by behavioural and neuroimaging studies (Baddeley, 2003; Conrad & Hull, 1964; Halpern & Zatorre, 1999; Hickok et al., 2003; Koelsch et al., 2009; Paulesu et al., 1993; Schulze, Zysset et al., 2011; Williamson et al., 2010).[6] Only a limited number of words (Baddeley, 2003; Conrad & Hull, 1964) or tones (Croonen, 1994; Williamson et al., 2010) can be maintained by internal rehearsal. The observed length effect together with the literature

thus suggest that participants used internal rehearsal to maintain this information active in WM, presumably using sensorimotor codes (Baddeley, 2003; Koelsch et al., 2009; Schulze, Zysset et al., 2011). It is worth noting that whereas Experiment 2 and Schulze et al. (2012) reported a length effect for a backward recognition task using tones, this was not the case for previous studies using a backward recall task with verbal material (Bireta et al., 2010; Surprenant et al., 2011). Future studies thus need to investigate more systematically differences and similarities between recall and recognition paradigms for different auditory materials. However, our results also indicate differences between verbal and tonal WM. First, as hypothesised, articulatory suppression (Experiment 3) decreased WM performance for words, but not for tones. Our data together with previous research thus suggest that both words and tones are internally rehearsed (subvocal speech/singing based on sensorimotor codes; Baddeley, 2003; Koelsch et al., 2009; Schulze, Zysset et al., 2011), but that the subvocal rehearsal, or possibly the internal representation/sensorimotor codes used to maintain and manipulate the stimuli, might differ between words (subvocal speech) and tones (internal singing; Deutsch, 1970; Schulze, Zysset et al., 2011).

Second, we observed a selective positive correlation between the years of musical training and the WM performance for the pitch sequences during the backward task (Experiment 2 and 3), but not during the forward task. In addition, no correlation between the years of musical training and WM for words or timbre was observed. This finding suggests that musical training improved WM for pitch information during the backward task, presumably mediated by more elaborate knowledge how to analyse and produce, and presumably also how to manipulate tonal information. Previous studies indicated that musical training improves tonal WM also during a forward recognition task (Schulze et al., 2012; Schulze, Zysset et al., 2011) and a forward recall task (Williamson et al., 2010).

Third, the relative performance pattern for words and tones differed between both tasks, with better performance for tones during the forward task and worse performance for tones during the backward task compared to words. As discussed above, this might be a consequence of stimulus-inherent properties. WM performance for words might have been decreased, compared

---

[6] Although premotor areas are involved during the memory for serial order (Henson et al., 2003), activation in the premotor cortex was also observed in item memory, notably auditory memory tasks investigating the memory for verbal and/or tonal items (Gruber & von Cramon, 2003; Paulesu et al., 1993; Schulze, Zysset et al., 2011). Therefore the observed involvement of motor-related structures during WM for words and also tones is not solely due to memory for serial order.

to tones, due to the phonological similarity in the forward task (Baddeley, 2003; Conrad & Hull, 1964), but to a lesser degree during the backward task (Bireta et al., 2010). For tonal WM it has been shown that WM performance was increased when the to-be-detected changes altered the melodic contour (Dowling & Fujitani, 1971) and by tonal structures (Schulze, Mueller et al., 2011) in a forward task, but not in a backward task (Schulze et al., 2012). Therefore, in the present study, participants' tonal WM performance might have benefited from melodic contour and tonal-like structures in the forward task, but not in the backward task, which required reversing the melody. Although it was not possible to compare the forward and backward experiments directly (due to the different sequence length used), it looks as if the costs for performing the backward task compared to the forward task are greater for tones than for words. Therefore the present findings suggest that forward and backward WM recognition differs, as has been previously shown also for verbal recall (Bireta et al., 2010; Hulme et al., 1997) and tone recognition (Schulze et al., 2012), which might be due to the intervention of the central executive (Baddeley, 2003, 2012; Baddeley & Della Sala, 1996).

In addition, previous findings have suggested that non-musicians have more elaborate sensorimotor representations for the rehearsal of verbal information than for the rehearsal of tonal information (Schulze, Zysset et al., 2011). Thus participants might have been able to benefit from these elaborate, and possibly more stable, sensorimotor representations for verbal material during manipulation, leading to a better WM performance for words compared to tones in Experiment 2.

However, if verbal rehearsal and tonal rehearsal rely on similar WM mechanisms, i.e., internal rehearsal, why would we observe a suppression effect only for words, but not for tones? Numerous previous research findings have suggested that verbal and tonal information can be internally rehearsed in WM (Hickok et al., 2003; Koelsch et al., 2009; Pechmann & Mohr, 1992; Schulze, Zysset et al., 2011; Williamson et al., 2010), while the rehearsal of both types of materials might rely on different (sensorimotor) codes. Schulze, Zysset et al. (2011) observed that neural networks underlying verbal and tonal WM show considerable overlap in cerebral core structures for both non-musicians and musicians. This was interpreted as a shared rehearsal process that

is using different codes, notably sensorimotor codes involved in the rehearsal of verbal and tonal material. A similar hypothesis was proposed earlier for storage of information in memory: Based on Patel's (2003) proposition of shared processes but distinct representations between music and language, Williamson et al. (2010) suggested that the WM rehearsal might process verbal and tonal information, but storage could be separable. Therefore the internal rehearsal process might underlie both verbal and tonal WM, whereas both types of WM operate with distinct sensorimotor representations (Hickok et al., 2003; Koelsch et al., 2009; Schulze, Zysset et al., 2011). This interpretation is in line with the finding that participants with congenital amusia showed WM impairment for tones, but not for words (Tillmann, Schulze, & Foxton, 2009) and could explain the differences and similarities we observed in the present study between WM performance for verbal and tonal material during the forward and backward recognition tasks.

Timbre in comparison to tones and words

WM performance for timbre differed from tones in all experiments. This finding is in agreement with previous data suggesting that pitch and timbre are processed independently in WM (Krumhansl & Iverson, 1992; Semal & Demany, 1991; Starr & Pitt, 1997).

In contrast to words and tones, for which a length effect was observed in the present study and in previous studies (Baddeley, 2003; Croonen, 1994; Schulze et al., 2012; Williamson et al., 2010), no length effect was observed for timbres in Experiment 1 (forward task). This missing length effect of timbres together with results from previous studies investigating sensory memory (Crowder, 1989; Fritz et al., 2007; Kaernbach, 2004b; Winkler & Cowan, 2005) could suggest that WM for timbre relies on a sensory trace during WM maintenance. While participants might be unable to repeat or imitate timbre information (i.e., processes that support internal rehearsal), previous studies suggested that timbre can be imagined (Crowder, 1989; Halpern et al., 2004). This imagery might engage sensory rather than motor (production) related processes (Crowder, 1989; Halpern et al., 2004; Pitt & Crowder, 1992). McKeown (2011) suggested that

the processes underlying WM for timbre are not dependent on attention or rehearsal, and that this might be supported by oscillatory processes in the theta and gamma band (Duzel, Penny, & Burgess, 2010). WM for timbre thus seems to rely on different mechanisms than WM for tones (Halpern & Zatorre, 1999; Koelsch et al., 2009; Schulze, Zysset et al., 2011), and verbal materials (Baddeley, 2003; Koelsch et al., 2009; Paulesu et al., 1993; Schulze, Zysset et al., 2011), which both have been shown to rely on internal rehearsal and motor-related brain structures. Although future studies are necessary to further investigate WM for timbre, for example by directly comparing the neural correlates underlying WM for timbre with those underlying WM for words and tones, our findings suggest different mechanisms for timbre than for words and tones.

In the backward task (Experiment 2) a length effect was observed for timbres, notably with three-item sequences being better remembered than the four-item sequences. The possibility that participants have verbally labelled the timbre stimuli was ruled out by Experiment 3. If, as discussed above, memory for timbre relies on a sensory memory trace, the observed length effect in Experiment 2 suggests that this coding is less efficient when manipulation is required for an increasing number of items (as in the backward task). One possible explanation is that during the backward task, participants needed to compare the passively stored sensory trace of the timbre events of the first sequence with the reversed timbre events of the second sequence to detect deviations from the presented order. This might be more time consuming than in the forward task, and this strategy might also be more inefficient for the longer (four-item) sequences than for the shorter (three-item) sequences.

Furthermore the additional processes necessary for manipulation (beyond maintenance), such as the involvement of the central executive (Baddeley, 2003, 2012; Baddeley & Della Sala, 1996), seem to be subjected to limitations of elements (which are manipulated) and therefore leading to a length effect, resulting in decreased performance with increased number of to-be-manipulated elements. Future studies need to further investigate manipulation of timbral information in WM, as started in Experiments 2 and 3 with the present backward task. WM performance for timbre did not differ from WM performance for words, neither for the forward task (with the short sequences in Experiment 1) nor for the backward task (Experiment 2). However, when

articulatory suppression was added, WM performance was impaired for words, but not for timbres. If words and timbres had been processed by the same WM mechanisms and systems, WM performance for timbre should have been impaired as well.

The results from the present study thus suggest that WM differs for timbre and words. The assumption of separate mechanisms underlying WM for timbres and words is in agreement with previous studies (i) observing that verbal rehearsal did not interfere with WM for timbre (McKeown et al., 2011) and (ii) reporting that participants with congenital amusia showed WM impairment for timbres, but not for words (Tillmann et al., 2009).

In addition our data support the observation by Kaernbach (2001, 2004a, 2004b) who proposed that categorical information can be sustained with rehearsal, while this is not the case for sensory storage with acategorical information.

Our findings suggest that different mechanisms underlie WM for timbre compared to words and tones. More specifically, our results together with previous observations suggest that verbal and tonal material can be rehearsed in WM (categorical information), whereas WM for timbre might be stored as a sensory memory trace in the long sensory store. This has to be further investigated, for example, by increasing the duration of the items, the retention delay, the number of items to remember, by using different auditory interference material, and by comparing directly the neural correlates underlying WM for timbre with those underlying WM for words and tones.

Limitations of the present study and suggestions for future studies

First, it is challenging to design three different kinds of auditory stimuli that are comparable in terms of absolute levels of WM performance, mainly because humans are more familiar with some auditory stimuli, namely words, but less so with others (timbre). In particular we discussed that acoustic similarity and stimulus-inherent structure could have had an influence on WM performance (see results by Williamson et al., 2010), even though this cannot explain the differences between the two tasks (forward, backward).

Second, some aspects of auditory WM are still elusive, which is reflected in the literature as there is no consensus for the interpretation. For

example, in the present paper the articulatory suppression and the length effect have been interpreted as an indication for internal rehearsal processes (Baddeley, 2003, 2012; Surprenant et al., 1999; Williamson et al., 2010). As indicated throughout the paper, the length effect has been controversially discussed in the literature and alternative explanations have been suggested: The word length effect is confounded with other word characteristics like phonological complexity, and evidence suggests an attentional mechanism that could operate in addition to articulatory rehearsal (Lewandowsky & Oberauer, 2008). However, although being aware of these potential alternative interpretations, we propose—based on recent neuroimaging results (Koelsch et al., 2009; Schulze, Zysset et al., 2011)—that our results indicate that timbre relies mainly on a sensory store (long sensory store) while tones and words rely on categorical WM.

Summary and conclusion

Together with previous research (e.g., Pechmann & Mohr, 1992; Schulze, Zysset et al., 2011; Williamson et al., 2010), our findings suggest that human auditory WM is not a unitary system. Our data indicate that: First, auditory WM differs for the here tested auditory stimuli presumably because some can be internally rehearsed (e.g., words) whereas others cannot (e.g., timbre) and rely on a sensory trace. In addition we suggest that rehearsal mehanisms might be underlying verbal and tonal WM, which, however, are operating with different (sensorimotor) representations (Patel, 2003; Schulze, Zysset et al., 2011; Williamson, et al., 2010). Second, WM recognition differs for forward and backward tasks. Third, musical training can be selectively associated with better WM performance for tones during the backward task, involving manipulation.

REFERENCES

Baddeley, A. D. (1992). Working memory. *Science*, *255*, 556–559.

Baddeley, A. D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829–839.

Baddeley, A. D. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology*, *63*, 1–29.

Baddeley, A. D., Chincotta, D., Stafford, L., & Turk, D. (2002). Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology*, *55*, 353–369.

Baddeley, A. D., & Della Sala, S. (1996). Working memory and executive control. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *351*, 1397–1403.

Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. A. Bower (Ed.) *Recent advances in learning and motivation* (Vol. VIII, pp. 47–89). New York: Academic Press.

Baddeley, A. D., Thomson, N., & Buchanan, L. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589.

Berz, W. L. (1995). Working memory in music: A theoretical model. *Music Perception*, *12*, 353–364.

Bireta, T. J., Fry, S. E., Jalbert, A., Neath, I., Surprenant, A. M., Tehan, G., et al. (2010). Backward recall and benchmark effects of working memory. *Memory & Cognition*, *38*, 279–291.

Bor, D., Duncan, J., Wiseman, R. J., & Owen, A. M. (2003). Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron*, *37*, 361–367.

Brown, G. D., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological review*, *107*, 127–181.

Buchsbaum, B. R., & D'Esposito, M. (2008). The search for the phonological store: from loop to convolution. *Journal of Cognitive Neuroscience*, *20*, 762–778.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*, 551–581.

Chan, A. S., Ho, Y. C., & Cheung, M. C. (1998). Music training improves verbal memory. *Nature*, *396*, 128.

Conrad, R., & Hull, A. J. (1964). Information, acoustic confusion and memory span. *British Journal of Psychology*, *55*, 429–432.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin*, *96*, 341–370.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*, 163–191.

Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 62–101). Cambridge, UK: Cambridge University Press.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioural and Brain Sciences*, *24*, 87–185.

Croonen, W. L. (1994). Effects of length, tonal structure, and contour in the recognition of tone series. *Perceptual Psychophysics*, *55*, 623–632.

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology*, *15*, 472–478.

Demany, L., Montandon, G., & Semal, C. (2004). Pitch perception and retention: Two cumulative benefits of selective attention. *Perceptual Psychophysics, 66,* 609–617.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science, 168,* 1604–1605.

Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *Journal of the Acoustical Society of America, 49,* Suppl 2:524+.

Duzel, E., Penny, W. D., & Burgess, N. (2010). Brain oscillations and memory. *Current Opinion in Neurobiology, 20,* 143–149.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working-memory. *Psychological Review, 102,* 211–245.

Farrand, P., & Jones, D. (1996). Direction of report in spatial and verbal serial short-term memory. *Quarterly Journal of Experimental Psychology A, 49,* 140–158.

Fritz, J. B., Elhilali, M., David, S. V., & Shamma, S. A. (2007). Auditory attention–focusing the searchlight on sound. *Current Opinion in Neurobiology, 17,* 437–455.

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C., Jones, G., Oliver, I., et al. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Science, 5,* 236–243.

Gruber, O., & von Cramon, D. Y. (2001). Domain-specific distribution of working memory processes along human prefrontal and parietal cortices. *Neuroimage, 13,* S679–S679.

Gruber, O., & von Cramon, D. Y. (2003). The functional neuroanatomy of human working memory revisited. Evidence from 3-T fMRI studies using classical domain-specific interference tasks. *Neuroimage, 19,* 797–809.

Hall, D., & Gathercole, S. E. (2011). Serial recall of rhythms and verbal sequences: Impacts of concurrent tasks and irrelevant sound. *Quarterly journal of experimental psychology, 64,* 1580–1592.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9,* 697–704.

Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia, 42,* 1281–1292.

Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: An fMRI study. *Neuropsychologia, 38,* 426–440.

Henson, R. N. A., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: A new paradigm and tests of a timing-signal hypothesis. *Quarterly Journal of Experimental Psychology Section a-Human Experimental Psychology, 56,* 1307–1334.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience, 15,* 673–682.

Hulme, C., Roodenrys, S., Schweickert, R., Brown, G. D. A., Martin, S., & Stuart, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology-Learning Memory and Cognition, 23,* 1217–1232.

Jones, D. M. (1993). Objects, streams and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention: Selection, awareness and control* (pp. 87–104). Oxford, UK: Clarendon.

Kaernbach, C. (2001). *Parameters of echoic memory.* Paper presented at the seventeenth annual meeting of the International Society for Psychophysics, Lengerich.

Kaernbach, C. (2004a). Auditory sensory memory and short-term memory. In C. Kaernbach, E. Schröger, & H. Müller (Eds.), *Psychophysics beyond sensation: Laws and invariants of human cognition.* Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Kaernbach, C. (2004b). The memory of noise. *Experimental Psychology, 51,* 240–248.

Kaernbach, C., & Schlemmer, K. (2008). The decay of pitch memory during rehearsal. *Journal of the Acoustical Society of America, 123,* 1846–1849.

Kawahara, H., & Irino, T. (2004). Underlying principles of a high-quality speech manipulation system STRAIGHT and its application to speech segregation. In P. L. Divenyi (Ed.), *Speech separation by humans and machines* (pp. 167–180). Amsterdam: Kluwer Academic.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Muller, K., & Gruber, O. (2009). Functional architecture of verbal and tonal working memory: An FMRI study. *Human Brain Mapping, 30,* 859–873.

Krumhansl, C. L., & Iverson, P. (1992). Perceptual interactions between musical pitch and timbre. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 739–751.

Lewandowsky, S., & Oberauer, K. (2008). The word-length effect provides no evidence for decay in short-term memory. *Psychonomic Bulletin & Review, 15,* 875–888.

McKeown, D., Mills, R., & Mercer, T. (2011). Comparisons of complex sounds across extended retention intervals survives reading aloud. *Perception, 40,* 1193–1205.

McKeown, D., & Wellsted, D. (2009). Auditory memory for timbre. *Journal of experimental psychology. Human perception and performance, 35,* 855–875.

Mercer, T., & McKeown, D. (2010). Updating and feature overwriting in short-term memory for timbre. *Attention, Perception & Psychophysics, 72,* 2289–2303.

Nairne, J. S. (1990). A feature model of immediate memory. *Memory & Cognition, 18,* 251–269.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6,* 674–681.

Paulesu, E., Frith, C. D., & Frackowiak, R. S. (1993). The neural correlates of the verbal component of working memory. *Nature, 362,* 342–345.

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: implications for a working-memory model. *Memory & Cognition*, *20*, 314–320.

Pitt, M. A., & Crowder, R. G. (1992). The role of spectral and dynamic cues in imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 728–738.

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing: a minimal executive function. *Cortex*, *43*, 135–145.

Saito, S. (1994). What effect can rhythmic finger tapping have on the phonological similarity effect. *Memory & Cognition*, *22*, 181–187.

Saito, S. (2001). The phonological loop and memory for rhythms: An individual differences approach. *Memory*, *9*, 313–322.

Saito, S., & Ishio, A. (1998). Rhythmic information in working memory: Effects of concurrent articulation on reproduction of rhythms. *Japanese Psychological Research*, *40*, 10–18.

Salame, P., & Baddeley, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology*, *41*, 107–122.

Sams, M., Hari, R., Rif, J., & Knuutila, J. (1993). The human auditory sensory memory trace persists about 10 sec – neuromagnetic evidence. *Journal of Cognitive Neuroscience*, *5*, 363–370.

Savage, C. R., Deckersbach, T., Heckers, S., Wagner, A. D., Schacter, D. L., Alpert, N. M., et al. (2001). Prefrontal regions supporting spontaneous and directed application of verbal learning strategies: Evidence from PET. *Brain*, *124*, 219–231.

Schendel, Z. A., & Palmer, C. (2007). Suppression effects on musical and verbal memory. *Memory & Cognition*, *35*, 640–650.

Schroger, E. (2007). Mismatch negativity – A microphone into auditory memory. *Journal of Psychophysiology*, *21*, 138–146.

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*, *29*, 255–267.

Schulze, K., Mueller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience*, *33*, 189–196.

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*, 771–783.

Semal, C., & Demany, L. (1991). Dissociation of pitch from timbre in auditory short-term memory. *Journal of the Acoustical Society of America*, *89*, 2404–2410.

Semal, C., Demany, L., Ueda, K., & Halle, P. A. (1996). Speech versus nonspeech in pitch memory. *Journal of the Acoustical Society of America*, *100*, 1132–1140.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology*, *33*, 5–42.

Starr, G. E., & Pitt, M. A. (1997). Interference effects in short-term memory for timbre. *Journal of the Acoustical Society of America*, *102*, 486–494.

Surprenant, A. M., Brown, M. A., Jalbert, A., Neath, I., Bireta, T. J., & Tehan, G. (2011). Backward recall and the word length effect. *American Journal of Psychology*, *124*, 75–86.

Surprenant, A. M., Neath, I., & LeCompte, D. C. (1999). Irrelevant speech, phonological similarity, and presentation modality. *Memory*, *7*, 405–420.

Tehan, G., & Mills, K. (2007). Working memory and short-term memory storage: What does backward recall tell us? In N. Osaka, R. Logie, & M. D'Esposito (Eds.), *The cognitive neuroscience of working memory*. Oxford, UK: Oxford University Press.

Tillmann, B., Schulze, K., & Foxton, J. M. (2009). Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds. *Brain & Cognition*, *71*, 259–264.

Tulving, E. (1962). Subjective organization in free recall of ''unrelated'' words. *Psychological Review*, *69*, 344–354.

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 & Cognition*, *38*, 163–175.

Winkler, I., & Cowan, N. (2005). From sensory to long-term memory – Evidence from auditory memory reactivation studies. *Experimental Psychology*, *52*, 3–20.

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2009). Mental reversal of imagined melodies: a role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*, *22*, 775–789.