Chapter 4
Memory for Timbre

Kai Siedenburg and Daniel Müllensiefen

Abstract Memory is a cognitive faculty that is of fundamental importance for human communication in speech and music. How humans retain and reproduce sequences of words and pitches has been studied extensively in the cognitive literature. However, the ability to retain timbre information in memory remains less well understood. Recent years have nonetheless witnessed an upsurge of interest in the study of timbre-related memory processes in experimental psychology and music cognition. This chapter provides the first systematic review of these developments. Following an outline of basic memory concepts, three questions are addressed. First, what are the memory processes that govern the ways in which the timbres of sound sequences are recognized? Predominantly focusing on data from short-term recognition experiments, this review addresses aspects of capacity and similarity, sequential structures, and maintenance processes. Second, is there interference of timbre with other attributes in auditory memory? In other words, how specific are memory systems for timbre and to what degree are they separate from memory systems for pitch and verbal information. Third, do vocal sounds and the sounds from familiar sources possess a special status in auditory memory and, if so, what could be the underlying mechanisms? The chapter concludes by proposing five basic principles of memory for timbre and a discussion of promising avenues for future research.

Keywords Acoustic similarity · Auditory memory · Melodic memory · Memory maintenance · Pitch interference · Sound-source familiarity · Verbal memory · Voice superiority · Working memory

K. Siedenburg (✉)
Department of Medical Physics and Acoustics, Carl von Ossietzky Universität Oldenburg, Oldenburg, Germany
e-mail: kai.siedenburg@uni-oldenburg.de

D. Müllensiefen
Department of Psychology, Goldsmiths, University of London, London, United Kingdom
e-mail: d.mullensiefen@gold.ac.uk

© Springer Nature Switzerland AG 2019

K. Siedenburg et al. (eds.), *Timbre: Acoustics, Perception, and Cognition*,
Springer Handbook of Auditory Research 69,
https://doi.org/10.1007/978-3-030-14832-4_4

4.1 Introduction

Memory, the capability to explicitly or implicitly remember past experiences, is one of the most extraordinary and mysterious abilities of the mind. Memory defines human perception, cognition, and identity. Speech and music, both fundamental to human nature and culture, are based on short- and long-term memory for acoustic patterns. Memories exist for many, but not all, experienced events: Think about which clothes you wore on an important day of your life versus which ones you wore last Wednesday (unless Wednesday was important). Not all aspects of perceptual experience are memorized equally well: Think about whether the first notes from a song you like go up or down versus the exact pitch height of the melody's first note.

While assessing memory for pitch patterns, tone sequences, and melodies has a long tradition in auditory psychology (e.g., Deutsch 1970; Müllensiefen and Halpern 2014), there are considerably fewer publications on memory for timbre. Hence, does memory for timbre exist at all? Are humans able to remember timbre information, such as the quality of an unfamiliar voice or the sonority of a particular sound sequence from a music track, over short and long time spans? Or is timbre an attribute of auditory experience that is reserved for being experienced in the moment? Only 10 years ago, research did not have proper empirical ground to answer these questions. In fact, it is important to note that timbre has not been considered critical for memory and cognition for a long time.

One reason for the lack of research on memory for timbre is the fact that speech and music have most commonly been considered within an information processing framework (e.g., Simon 1978), whereby the communicative message is conceptualized as sequences of phonemic or pitch categories that are independent from the properties of the carrier medium, which includes the sounds' vocal or instrumental timbre. Influential models of human memory (Atkinson and Shiffrin 1968) presumed that aspects of sensory information could be transformed into cognitive information and short-term or long-term memory using symbolic recoding. Any sensory information that could not be recoded was assumed to be lost from the sensory (echoic) memory store (Darwin et al. 1972). A second reason for the lack of research on timbre memory might be rooted in the fact that classic music-theoretical approaches—traditionally a driving force behind much music cognition research (Meyer 1956; Lerdahl and Jackendoff 1983)—focus on pitch and duration and their derived musical parameters harmony and rhythm but do not cover timbre as a primary musical parameter. Thirdly, the relative scarcity of empirical evidence for complex cognitive processes related to timbre, such as effects of auditory context or musical experience, may have had additional impact. Overall, this situation may have created the false impression that timbre is an auditory surface feature that is not essential to the cognitive architecture of human audition. Fortunately, this situation is beginning to change and many researchers in experimental psychology and music cognition have started to address timbre-related questions.

As summarized in this chapter, the effects of auditory context and long-term auditory experience with timbre have been demonstrated (both at a behavioral and neural level), the role of voice timbre in speech perception has become subject to experimental scrutiny, and the effects of vocal timbre on verbal memory have been known for a longer time. Clearly, timbre is becoming a rich and exciting topic for auditory cognition research, and memory obviously plays an important role in this development. Note that of the forty-five or so empirical studies on memory for timbre, more than thirty-five have been published between 2008 and 2018. This chapter provides the first systematic review of this emerging field and thereby highlights the fact that memory for timbre is a highly relevant concept in auditory cognition.

Section 4.2 describes general concepts from memory research, in particular, with regards to auditory memory systems for short-term and long-term storage, the granularity of auditory memory, and models of short-term memory. Regarding memory for timbre, four research themes stand out and provide a structure for subsequent sections. The first research theme comprises many studies that scrutinize the structure of short-term memory for timbre and have started to propose cognitive mechanisms that might be implicated. In Sect. 4.3, the presumed capacity limits of short-term memory for timbre will be discussed with a particular focus on the role of perceptual similarity and chunking.

The second theme concerns the active maintenance and imagery of timbre, which is addressed in Sect. 4.4. The tenet of this section is that memory for timbre involves elements of attentional control, which recreate facets of auditory experience.

The third theme focuses on the growing body of work that is demonstrating interference from auditory attributes on primary memory contents. For instance, variability in a task-irrelevant attribute, such as timbre, strongly impairs performance in a melodic memory task wherein the primary content (i.e., melodic structure) is conceptually independent of timbre. These findings are described in Sect. 4.5, which discusses the status of memory representations for timbre: Are they stored separately from other auditory attributes such as pitch or verbal information?

The fourth theme, discussed in Sect. 4.6, focuses on the role of sound source familiarity in memory for timbre and effects of voice superiority. Several studies have reported processing advantages for vocal timbre over timbres from nonvocal musical instruments. This finding resonates with the assumption that human listeners are specialists in voice timbre processing. To synthesize our discussion, five principles of memory for timbre are proposed that address some of the underlying cognitive processes. For a more general discussion of auditory memory, please see Demany and Semal (2007). A treatment of sound (source) recognition is included in Chap. 3 (Agus, Suied, and Pressnitzer) of this volume.

4.2 Auditory Memory Concepts

Memory is an overwhelmingly broad notion that plays a central role in almost every aspect of human cognition. At its core is the retention over time of experience-dependent internal representations and the capacity to reactivate such representations (Dudai 2007). Representations of sensory information and cognitive states thus are starting points for the formation of memories. But it is the temporal trajectory of these representations that defines memory and makes it such a rich and complex research topic.

4.2.1 Stores and Processes

An elementary conceptual distinction regarding the structure of human memory concerns the differences between the short-term and long-term memory systems. William James (1890/2004) already thought of primary (conscious, short-lived) and secondary (unconscious, long-lived) memory as independent entities. A more fine-grained distinction became the core of the classic *multistore* or *modal model*, most prominently elaborated by Atkinson and Shiffrin (1968). It posits three types of stores, namely a sensory register, a short-term memory (STM) store, and a long-term memory (LTM) store. According to Atkinson and Shiffrin (1968), sensory information is subject to modality-specific, pre-attentive storage of fast decay (within 2 s) unless there is a subject-controlled scan via selective attention, which recodes and transfers portions of the register to the short-term store. This store is thought to retain a categorical, modality-independent code where traces decay within time spans of less than 30 s. Their life spans can be lengthened by active rehearsal, which lends them more time to be transferred to the long-term store.

To refine this classic picture, Cowan (1984, 2015) proposed a taxonomy for nonverbal auditory memory that emphasized similarities with visual memory. In vision, one can find a seemingly clear structural divide between an automatic sensory storage of almost unlimited capacity and fast decay (< 200 ms)—*iconic memory*—and a more long-lived, attention-dependent, short-term memory system of constrained capacity. Cowan's *short auditory store* is hypothesized to be experienced as sensation or sensory afterimage (i.e., is distinct from the sensory type of memory required to integrate and bind perceptual features, such as loudness or amplitude modulations, over tenths of seconds). The short auditory store contains not-yet-analyzed, pre-categorical content that decays within 200–300 ms. The *long auditory store* is experienced as (short-term) memory, contains partially analyzed or categorized content, and is supposed to decay within 2–20 s. Due to the structural similarity of the long store and categorical STM (Atkinson and Shiffrin 1968) with regard to decay rates and capacity, Cowan considered the long auditory store to be a special case of STM. Contrary to the classic multistore models that assume that STM operates on verbal items, Cowan's proposal implies that STM may also operate on sensory representations.

Although Cowan's distinction between a short and automatic versus a long and consciously controlled form of auditory memory may have intuitive appeal due to its analogy to vision, recent data suggest that it is hard to find clear-cut boundaries. Several studies have highlighted difficulties in estimating the exact duration of the shorter type of auditory memory. More specifically, testing the discrimination of frequency shifts within nonharmonic tone complexes, Demany et al. (2008) observed a gradual decay in performance for increasing retention times, which is not comparable to the steep decline that is characteristic of iconic memory in vision. Importantly, there was no clear evidence for differential memory capacity (i.e., a short store of high capacity and a long store of low capacity) within the 2 s range of retention times tested. Demany et al. (2010) explicitly compared visual and auditory change detection. Whereas visual memory fidelity appeared to decay quickly and substantially within 200 ms, confirming the classical view on iconic memory, there was no such sign for auditory memory, which persisted throughout retention times of 500 ms at much lower decay rates. This finding indicates that auditory change detection may operate on much longer time scales than visual iconic memory. As a theoretical explanation, Demany et al. suggest frequency shift detectors as a cognitive mechanism that tracks spectral changes of stimuli. These detectors were shown to gradually lose their tuning specificity when inter-stimulus intervals increase (Demany et al. 2009). But the observed differences in tuning specificity were gradual rather than showing clear-cut boundaries.

Rejecting the idea of clear separations between hypothetical memory stores resonates with the *proceduralist approach* to memory (Crowder 1993; Jonides et al. 2008). Instead of conceptualizing memory as a separate cognitive system, implemented by a multitude of interacting modules (e.g., sensory, STM, and LTM), the unitary or proceduralist approach understands memory as an emergent property of the ways in which mental processes operate on perceptual representations or cognitive states. As noted by Craik and Lockhart (1972), "It is perfectly possible to draw a box around early analyses and call it sensory memory and a box around intermediate analyses called short-term memory, but that procedure both oversimplifies matters and evades the more significant issues" (p. 675). A classical illustration of the idea of memory being a byproduct of perceptual processing is given by the *levels of processing effect* (Craik and Lockhart 1972): If experimental participants' attention in an encoding phase is drawn toward "deep" semantic features of words (as in a semantic categorization task), recall is better than if participants judge "shallow" perceptual features of the stimuli (as in phonemic categorization). Contemporary neuroimaging studies support unitary views of memory in the sense that, in general, the same neural ensembles are found to be responsible for perceptual processing and memory storage (D'Esposito and Postle 2015).

Note that even if one does not believe in the existence of dedicated short-term and long-term memory systems, the notions of STM and LTM may be used as referents to memory function over short or long time intervals. This agnostic usage acknowledges that there may be different time scales of memory persistence but does not presuppose any particular stores or cognitive mechanisms.

4.2.2 Granularity of Auditory Memory

Another line of research has raised the question of how fine-grained auditory memory representations can be. In other words, what is the smallest detail of a sound that can be remembered? Using noise waveforms that are completely identical according to macroscopic auditory features, such as spectral and temporal envelope, Kaernbach (2004) showed that repetitions of noise segments could be well detected up to at least 10 s of segment length; single, seamless repetitions of noise waveforms were detected with above-chance accuracy up to 2 s. Agus et al. (2010) even demonstrated that there is a form of long-term persistence for features of noise waveforms (also see Agus, Suied, and Pressnitzer, Chap. 3). When requiring listeners to detect repetitions of noise segments, recurring noise stimuli featured far superior hit rates compared to novel noise waveforms. Notably, subjects were not aware that segments reoccurred and must have implicitly picked up idiosyncratic features of the presented noise tokens. This demonstrates that there is implicit, nondeclarative long-term auditory memory even for small sensory details. This memory process appears to be fully automatic: Andrillon et al. (2017) even demonstrated that noise snippets were memorized during rapid-eye-movement sleep.

What is the relation between this detailed form of memory and the formation of general auditory categories? McDermott et al. (2013) had listeners discriminate different classes of resynthesized environmental textures (e.g., rain versus waves) and exemplars of textures (e.g., one type of rain versus another). Texture category discrimination performance gradually increased with excerpt length (40–2500 ms) but, curiously, the discrimination of exemplars within categories gradually worsened. This was interpreted as an indication that summary statistics underlie the representation of sound textures: Representations of two exemplars from the same category converge with increasing excerpt length because averaging over increased lengths removes idiosyncratic sound features. In sum, this implies that humans can possess fine-grained memories of auditory events (Agus et al. 2010), but the recognition of sound (texture) categories likely relies on robust summary statistics that are less affected by idiosyncratic details (McDermott et al. 2013).

4.2.3 Capacity Limits in Short-Term Memory

A common assumption in studies of human short-term memory is its limited capacity. The famous conjecture by Miller (1956) states that people can retain 7±2 independent chunks of information in immediate memory. This idea has been of enormous impact in cognitive (and popular) science. Miller's core idea was that the informational bottleneck of short-term memory does not strictly depend on the number of items, but that there is a general limit on the number of *independent chunks* of information in short-term memory. The concept of item and chunk are distinct because sequences of items may be recoded into fewer chunks. More technically, a chunk can be defined as a "collection of concepts that have strong associations to one another and much weaker associations to other chunks concurrently in use" (Cowan 2001, p. 89). For example, sequences of letters, such as IRSCIAFBI, are far easier to memorize when remembered as chunks IRS CIA FBI (familiar US federal agencies) than as raw item-by-item successions (Cowan 2008).

Presenting a contemporary revision of Miller's original hypothesis, Cowan (2001) reviewed empirical evidence across a wide range of domains such as verbal, visual, and auditory memory. Specifically, Cowan argued that the capacity limit of short-term memory (STM) is only about 4±1 chunks if the involvement of other factors, such as long-term memory (LTM) and active rehearsal, is limited. The above example illustrates this proposal because long-term memory enables participants to form chunks such as IRS, CIA, and FBI. The role of active rehearsal, classically considered as vocal or subvocal (i.e., silent) repetition of the stimuli in verbal memory research (Baddeley 2012), would be to actively maintain the memory trace.

Despite its considerable influence, Cowan's 4±1 proposal has received harsh criticism from the very beginning (see the peer commentaries in Cowan 2001). An alternative framework that has gained momentum in visual memory research replaces the idea of magical numbers in STM (7±2 or 4±1) by *resource-based models* (Ma et al. 2014). These models of short-term memory assume limited resources in terms of the representational space or medium shared by items but not a limit to the exact number of items that can be maintained. Stimulus representations are considered to be corrupted by noise. The level of the noise increases with more items to be held in memory because items interfere with each other in their representational space. In other words, resource models assume that short-term memory is fundamentally limited in the quality, rather than the quantity, of information. These assumptions imply an increased probability of memory lapses in situations when items are perceptually similar.

Transferring the concept of capacity limits or a capacity-similarity tradeoff to timbre entails the question of what constitutes the basic unit to be memorized, that is, the item. In the study of verbal memory, individual words naturally qualify as items because language is composed of strings of words. However, there are many other domains for which the situation is not as clear. As Ma et al. (2014) noted with regards to vision, "An 'item' is often relatively easy to define in laboratory experiments, but this is not necessarily the case in real scenes. In an image of a bike, for example, is the entire bike the item, or are its wheels or its spokes items?" Similar complications may be in place for auditory memory beyond speech. In the context of polyphonic music, there can be plenty of timbral contrast that arises in short time spans from the sounds of various instruments. But it is not intuitively clear what constitutes the unit of the item in this case: individual tones, fused auditory events, or segments of auditory streams? In analogy to the existing verbal memory research, many studies of STM for musical timbre (see Sect. 4.3) use sequences of individual tones that differ by timbre, for instance, with sounds from diff erent orchestral instruments changing on a note-by-note basis. Although this operationalization may be seen as a plausible perceptual model for an orchestration technique, such as *Klangfarbenmelodie* (i.e., timbre melodies; Siedenburg and McAdams 2018; McAdams, Chap. 8) or percussion music (Siedenburg et al. 2016), it does not seem to be an appropriate model for many other types of music, for which this type of strong timbral contrast on a note-to-note basis represents a rare exception.

4.3 Factors in Short-Term Recognition

4.3.1 Memory Capacity and Similarity

Similarity effects are a hallmark of verbal and visual STM (Baddeley 2012; Ma et al. 2014). Despite being perceptually discriminable, similar items are more frequently confused in memory compared to dissimilar ones. With regards to memory for timbre, however, research is only beginning to account for effects of perceptual similarity relations.

Starr and Pitt (1997) used an interpolated tone paradigm (cf., Deutsch 1970) that required participants to match a standard and a comparison stimulus, separated by a 5 s interval with intervening distractor tones. Their first experiment demonstrated an effect of timbre similarity: The more similar in brightness the interfering tones were to the target tone, the more detrimental was their effect on retention in memory. Visscher et al. (2007) tested auditory short-term recognition in an item recognition experiment using auditory ripple stimuli (i.e., amplitude-modulated sinusoid complexes). They observed that two independent factors caused decreases in false alarm rates on a trial-by-trial basis: (a) increases of the mean dissimilarity of the probe sound to the sequence and (b) increases of the perceptual homogeneity of the sounds in the sequence, that is, the average similarity between the sounds in the sequence.

In one of the first studies, Golubock and Janata (2013) set out to measure capacity limits of short-term memory for timbre. They used an item recognition task with synthetic sounds differing by timbre (constituting the items). They synthesized sounds that varied along the dimensions of spectral centroid, attack time, and spectral flux, the discriminability of which was ensured via separate just-noticeable-difference measurements. Sequences of 2–6 tones that differed in timbre were presented, but the tones were of constant pitch and loudness. Each sequence was followed by a silent retention interval of 1–6 s, and then a single probe tone was presented for which participants had to judge whether it was part of the sequence or not. The authors observed memory capacities at around K = 1.5 items, estimated according to the formula

$$K = (\text{hit rate} + \text{correct rejection rate} - 1)^{*}N,$$

where N denotes the number of items in the test sequence. Capacities significantly decreased with increasing sizes of the retention intervals, with K = 1.7 for 1 s and K = 1.3 for 6 s.

The large difference between the capacity estimate of an average of 1.5 timbre items from Golubock and Janata (2013) and the supposedly universal estimate of 3–5 items according to Cowan (2001) seems striking. Why should memory for timbre be so much worse? Notably, the sounds in Golubock and Janata's first experiment only varied along three timbral dimensions.

A second experiment used a more heterogeneous set of sounds from a commercial keyboard synthesizer and measured a significantly greater capacity of around 1.7 items. Figure 4.1 displays hit and correct rejection rates averaged across reten-

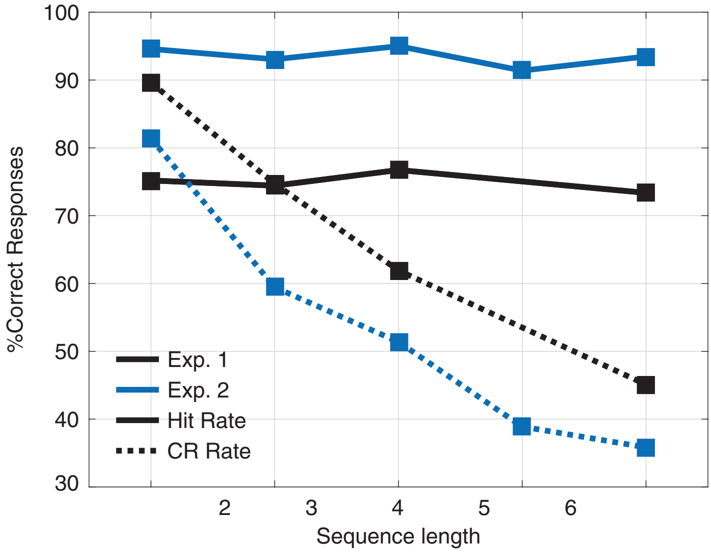

Fig. 4.1 Accuracy (percentage of correct responses) as a function of the sequence length in the two-item recognition experiments. Experiment 1 used abstract synthetic sounds; experiment 2 used sounds selected from a commercial sound sampler. Hits correspond to correct identification of match trials, correct rejections (*CR*) to correct identification of nonmatch trials. (Adapted from Table 1 in Golubock and Janata 2013; used with permission from the American Psychological Association)

tion intervals from the memory experiments in their study. Hit rates are higher in experiment 2 compared to experiment 1, but the false alarm rates of experiment 2 also exceed those of experiment 1. However, no trial-by-trial analyses of these data were conducted, and it remains unclear whether the increase in capacity in the second experiment was primarily caused by a global increase in the timbral homogeneity of sounds or by greater probe list dissimilarities.

Using an item recognition task, Siedenburg and McAdams (2017) observed significant correlations between participants' response choices (i.e., whether they recognized a probe sound as match or nonmatch) and the mean perceptual dissimilarity from the probe to the tones in the sequence. However, no significant correlation between timbral homogeneity and response choices was observed.

Siedenburg and McAdams (2018) further evaluated the role of similarity in a serial recognition task. They had participants indicate whether the order of the timbres of two subsequently presented sound sequences was identical or not. In the non-identical case, two sounds were swapped. A correlation analysis showed that the timbral dissimilarity of swapped items (TDS) was a good predictor of response choice in serial recognition and predicted around 90% of the variance of response choices throughout four experiments. This study also tested for the role of sequence homogeneity but did not find a consistent effect: Homogeneity and response choice were significantly correlated in only one out of four experiments. Moreover,

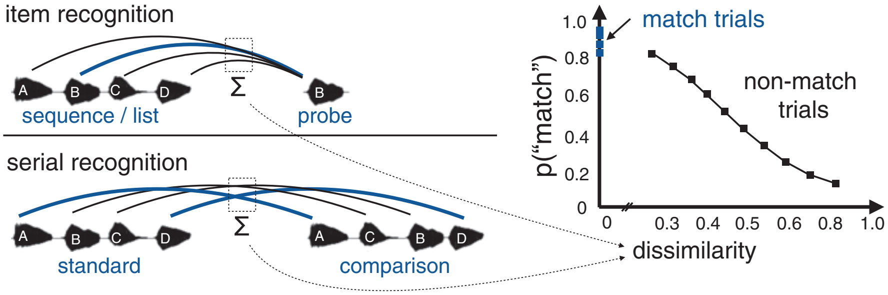

Fig. 4.2 Schematic depiction of the relationship between response choice (probability of "match" responses) and timbre dissimilarity. For item recognition tasks, the hypothetical dissimilarity measure corresponds to the sums of dissimilarities (Σ) of the probe item to all the items in the sequence (indicated by *connecting lines*). The *blue line* indicates a match and, hence, zero dissimilarity. For serial recognition tasks, the dissimilarity measure could be derived from the sum of the item-wise dissimilarities, resulting in the dissimilarity of the two items that were swapped (here: items *C* and *B*). Dissimilarity is normalized between 0 and 1

stepwise regression analysis failed to include homogeneity as a predictor of response choices in any experiment, indicating that a parsimonious account would not consider homogeneity as a crucial factor for timbre recognition. Figure 4.2 provides a schematic visualization of the described relation between response choice and both the probesequence dissimilarity in item recognition and the timbral dissimilarity of the swap in serial recognition.

Taken together, the strong effects of similarity (Siedenburg and McAdams 2018) and the wide range of estimates for timbre STM capacity (that differ clearly from STM capacity estimates for other auditory material; Golubock and Janata 2013) indicate that fixed-slot models of STM capacity may not be suitable as a model of STM for timbre. On the contrary, resource-based approaches that assume limited representational resources, and thus take into account similarity relations from the very beginning, appear to be better suited for the data from timbre experiments, although no formal model evaluation has been conducted yet. This is in line with the observed trade-off between the number of items that can be maintained in short-term memory and their timbral similarity.

4.3.2 Sequential Chunking

As already mentioned in Sect. 4.2.3, many memory studies try to avoid sequences with an explicit sequential structure. In order to measure memory proper, the rationale is that sequences should not explicitly allow for chunking made possible through grouping or repetition (Cowan 2001). At the same time, it is likely that affordances for sequential processing are important ecological factors in memory for timbre.

Siedenburg et al. (2016) considered the case of timbral sequencing as part of the tabla drumming tradition from North India. The tabla is a pair of hand drums with an extremely rich timbral repertoire and is considered the most important percussion instrument in North Indian classical music (Saxena 2008). Tabla music exhibits intricate serial patterns with hierarchical dependencies, for instance, through the nested repetition of groups of sounds. The centuries old tradition of tabla is taught as part of an oral tradition. Compositions are learned via the memorization of sequences of bols, that is, solfège-like vocalizations associated with drum strokes. In tabla solo performances, the verbal recitation of the composition oftentimes precedes the actual drumming. Furthermore, North Indian classical music is unfamiliar to most (but not all) western listeners and hence is well-suited for exploration of the effects of long-term memory on sound sequence recognition.

The experiment compared the recognition of tabla sequences between a group of tabla students and a group of western musicians unfamiliar with tabla music. As depicted in Fig. 4.3, four distinct sequencing conditions were used in the experiment: (1) idiomatic tabla sequences, (2) reversed sequences, (3) sequences of random order, and (4) sequences of random order and randomly drawn items without replacement. In the serial order recognition experiment, participants indicated whether the sounds in two consecutively played sequences were presented in the same order.

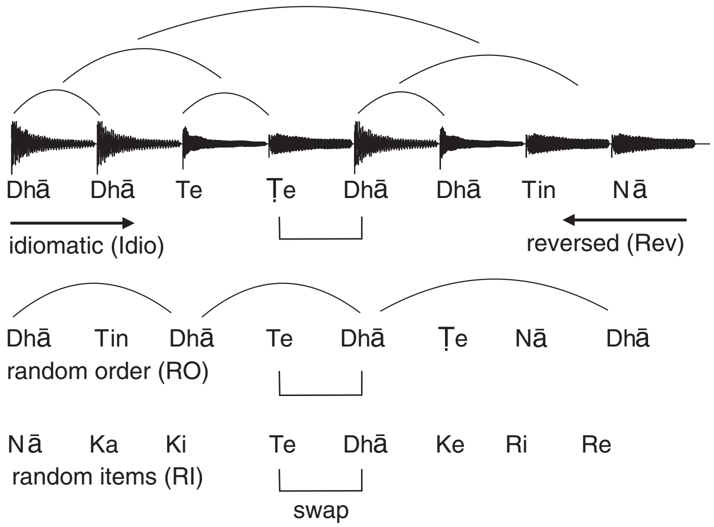

Fig. 4.3 Examples of the four sequencing conditions: an *idiomatic* sequence of bols (*Dha, Te, Tin, Na*) and the corresponding *reversed*, *random order*, and *random items* (adding *Ke, Ri, Re*) conditions (drawn without replacement). Note that in the idiomatic, reversed, and random order condition, there are items that occur multiple times in the sequence. (From Siedenburg et al. 2016; used with permission of the American Psychological Association)

The results showed a very strong effect of sequential structure: Idiomatic sequences of tabla strokes and their reversed versions were recognized best, followed by their counterparts with randomly shuffled order, followed by fully random sequences without repetitions of items. The latter effect indicated a facilitation of chunking due to the repetition of items. Because serial-order recognition was tested, it could be concluded that the advantage of redundancy primarily goes back to chunking and not a reduced load in terms of item identity. The advantage of reversed sequences over randomly shuffled ones was suspected to be related to the hierarchical structure inherent in the idiomatic sequences or their reversed versions. The reversed versions not only contained item repetitions, but repeating subsequences of items, such that sequences could be encoded hierarchically. Notably, effects of familiarity with idiomatic sequences (comparing tabla students versus naıve controls) only occurred for the vocal sounds but not for the drum sounds. This result indicates that vocal sounds are particularly well suited for chunking via long-term associations. Participants who are familiar with tabla can simply represent idiomatic sequences of bols (tabla words) via one item and hence have a significant mnemonic advantage over naıve participants. However, memory for instrumental sounds did not follow the same pattern, which may indicate that familiarity-based chunking is particularly effective for vocal sounds for which humans have a natural proficiency for combining basic building blocks in endless ways (e.g., Hagoort and Indefrey 2014).

An example of long-term recognition of timbre sequences was provided by Tillmann and McAdams (2004) who adopted the sequence-learning paradigm made famous by Saffran et al. (1999). Their results indicated that memory for timbre sequences is strongly affected by grouping cues provided by perceptual dissimilarity relations between subsequent tone pairs in the sequences (for more information, see McAdams, Chap. 8).

From a general perspective, these results indicate that auditory sequences can be stored much more efficiently if chunked in appropriate ways. Chunking could make memory for sequences more robust by structuring the memory trace along a hierarchy of time scales that is provided by grouping cues (if this sounds abstract, think about how to memorize, ABCXYZABCQ). This perspective allows us to explain effects in both short-term (Siedenburg et al. 2016) and long-term recognition (Tillmann and McAdams 2004).

4.4 Active Maintenance and Imagery of Timbre

In this section, it is argued that memory for timbre is not a fully automatic process that is solely based on persistence of passive information. Timbre representations can be consciously refreshed in working memory and recreated from long-term memory.

4.4.1 Maintenance in Working Memory

The key property that distinguishes the concept of working memory (WM) from that of short-term memory is the role of active manipulation and maintenance of the memory contents (although both terms are often used interchangeably). In contrast to the presumably passive and automatic process of auditory short-term memory, WM is usually defined as an active form of memory that, as a whole, underpins a range of important cognitive faculties such as problem solving and action control. The active nature of verbal WM becomes apparent when thinking of how phone numbers, street names, or vocabulary words in foreign language classes are commonly memorized. People tend to vocalize, openly or covertly, in order to retain verbal information in mind. This observation was captured by Baddeley's influential multicomponent model of working memory, which described verbal WM as governed by a phonological storage buffer and a rehearsal mechanism, overall giving rise to the *phonological loop* (Baddeley 2012). The memory trace in the buffer would decay gradually but could be refreshed by (sub)vocal rehearsal in order to be kept in the loop. In other words, the original auditory event undergoes a form of recoding into a sensorimotor code that allows conscious rehearsal.

Because of their success in explaining verbal working memory, the concept of the phonological loop has also influenced nonverbal auditory memory research and research into melodic memory in particular (Berz 1995; Schulze and Koelsch 2012). More specific to our concerns is the question of whether nonverbal auditory working memory and STM for timbre are subject to similar active maintenance processes. In other words, in which sense is short-term memory for timbre *working*?

Nees et al. (2017) tested whether melodic short-term recognition is supported by active rehearsal or by attention-based processes. Using a sequence matching task, participants listened to two melodies separated by an 8 s retention interval and judged the melodies as identical or non-identical. As is common in verbal WM research, a dual task paradigm was used. The basic assumption is that if a secondary task severely impairs the accuracy in the target task, the latter can be assumed to rely on similar cognitive processes and resources. Nees et al. (2017) used four secondary task conditions. An articulatory suppression (AS) condition required participants to read out loud solved math problems that were presented visually (e.g., 2 + 3 = 5). In an attentional refreshing suppression (ARS) condition, participants silently read math problems presented on a screen and needed to type the correct response on a computer keyboard. A third condition combined both articulatory and attentional refreshing suppression by having participants read aloud the math problem and provide the response orally (AS+ARS). A silent condition without suppression served as a baseline. Notably, the authors found that performance did not differ between the control and the ARS condition, but both the AS and AS+ARS conditions yielded a marked decline of sensitivity. These results indicate that melodic short-term memory is supported by subvocal rehearsal and not by attentional refreshing, suggesting strong structural similarities to verbal memory. As described in the following, the clarity of these findings for melody recognition by Nees et al. (2017) differs from the situation that we find for timbre.

Three distinct mechanisms for the maintenance of timbre in WM appear to be possible a priori. First, timbre recognition could be a passive process, which would imply that maintenance in fact does not play a strong role. The retention of timbre would instead primarily rely on the persistence of the sensory memory trace. Second, participants could attach labels to timbres (e.g., piano-violin-weird voice) and subsequently rehearse the verbal labels. This would constitute a verbal surrogate of memory for timbre. Third, listeners could allocate attention to the auditory memory trace and mentally replay timbre representations in their minds, a process that has been called *attentional refreshing* (Camos et al. 2009).

Several studies have gathered data that have implications for deciding on the plausibility of the mechanisms. McKeown et al. (2011) had three participants discriminate small changes in the spectral distribution of tones and showed that sensitivity was above chance even for extended retention intervals of 5–30 s. This effect was robust to an articulatory suppression task in which participants were required to read aloud during the retention time. These results were interpreted as evidence for a type of sensory persistence that is neither based on verbal labeling nor due to attentional refreshing. Schulze and Tillmann (2013) compared the serial recognition of timbres, pitches, and words in various experimental variants, using sampled acoustical-instrument tones and spoken pseudowords. They found that the retention of timbre, contrary to that of pitches and words, did not suffer from concurrent articulatory suppression, speaking against the involvement of labeling. In line with McKeown et al. (2011), they concluded that STM for timbre is structured differently than working memory for words or pitches and is unlikely to be facilitated by verbal labeling and (sub)vocal rehearsal. Nonetheless, their results did not rule out the possibility of attentional refreshing.

On the other hand, there are studies that have underlined the necessity of attentional refreshing for maintaining timbre information in memory. Soemer and Saito (2015) observed that short-term item recognition of timbre was only inconsistently disrupted by articulatory suppression but was more strongly impaired by a concurrent auditory imagery task. The authors interpreted these results as evidence that memory for timbre can be an active, re-enacting process that relies on the support of attentional resources. Siedenburg and McAdams (2017) more directly compared the effect of articulatory suppression with a suppression condition that captured listeners' visual attention. They used an item recognition task with familiar and unfamiliar sounds that were controlled for their timbral similarity relations. Three different suppression tasks filled the 6 s retention interval between the sound sequence and the probe sound. Participants either waited in silence, counted out loud (articulatory suppression), or detected identical exemplars in sequences of black and white grids (visual suppression). Results showed a clear advantage for familiar sounds that persisted throughout all experimental conditions. Surprisingly, there was no difference between articulatory and visual suppression, neither for familiar nor for unfamiliar sounds. However, both types of suppression affected timbre memory negatively compared to the silence condition.

Considering these empirical results from Siedenburg and McAdams (2017), multiple reasons speak for attentional refreshing as an important maintenance strategy for timbre. Firstly, verbal labeling was unlikely to act as a dominant maintenance strategy for timbre: Performance on unfamiliar sounds that were difficult to label was impaired under both articulatory and visual suppression. It seems much more plausible that the detrimental effect of articulatory suppression was due to interference with the auditory trace. Secondly, the plausibility of passive sensory storage without any active maintenance was ruled out by the detrimental effect of visual suppression, which should not interfere if auditory and visual WM are fully separated. Finally, it could be assumed that attentional refreshing was moderately disrupted by both types of suppression because the visual distractor task reduced attentional resources that refreshing relies on, and articulatory suppression interfered with the auditory trace that is subject to refreshing (beyond rather minor attentional requirements). Overall, these reasons indicated that attentional refreshing was the most likely candidate for active maintenance of timbre in the experiments by Siedenburg and McAdams (2017). The results by McKeown et al. (2011), to the contrary, indicated that neither verbal labeling and rehearsal nor attentional refreshing was necessary for successful timbre recognition. Taken together, this finding suggests that attentional refreshing is likely a sufficient, but not a necessary, condition of WM for timbre.

4.4.2 Mental Imagery of Timbre

Beyond the realm of maintaining information in short-term memory, research has also provided evidence for the feasibility of a closely related mental faculty: imagery for timbre. Whereas attentional refreshing was understood as a sort of attention-driven

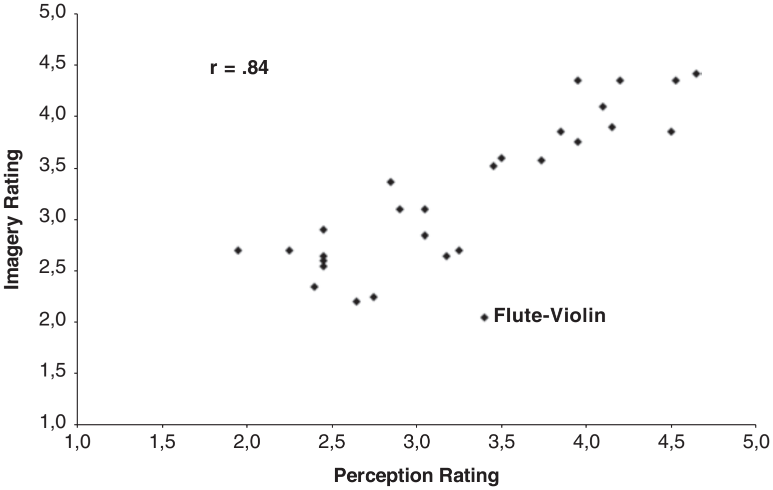

Fig. 4.4 Scatterplot of mean similarity ratings for each instrument pair in the perception and imagery conditions. Correlation coefficient, r = 0.84. (From Halpern et al. 2004; used with permission from Elsevier)

"replay function" of an initial sensory trace, imagery supposedly activates sensory representations without prior stimulation. This means that imagery solely makes use of long-term memory contents and constitutes a form of memory recollection in the perceptual domain. Studying the similarity of timbre imagery and perception, Halpern et al. (2004) had musicians rate perceived dissimilarity of subsequently presented pairs of timbres while recording brain activity with functional magnetic resonance imaging. The same procedure (including the dissimilarity ratings) was repeated in a condition in which the auditory stimuli were to be actively imagined. Figure 4.4 depicts the significant correlation between the behavioral dissimilarity data in the perception and imagery conditions. When compared to a visual imagery control condition, both auditory perception and imagery conditions featured activity in the primary and secondary auditory cortices with a right-sided asymmetry. Results such as these speak for the accuracy of auditory imagery for timbre: Sensory representations activated by imagery can resemble those activated by sensory stimulation.

These empirical findings have a bearing on the conceptualization of the active facets of timbre cognition. Working memory for timbre seems to be characterized as relying on concrete sensory refreshing or re-enactment and differs from the motor-based articulation processes found for pitch and verbal memory. Auditory imagery based on LTM representations of timbre appears to accurately resemble actual sensory stimulation. Both processes, refreshing and imagery, are related to the notion of *active perceptual simulation*, which is defined as a re-creation of facets of perceptual experience. Theories of perceptual symbol systems advocate that cognition is grounded in perceptual simulation (Barsalou 1999). This view stands in direct contrast to classic theories of cognition, which presume that perceptual processing leads to a transduction of sensory states into configurations of amodal symbols (Atkinson and Shiffrin 1968). Perceptual symbol systems assume that sensory schemata are abstracted from sensory states via perceptual learning, and cognition consists of simulating these schematic representations in concrete sensory form. That framework would be able to account for this phenomenon: When listeners actively maintain timbre in WM, they "hear" the original sound. Similarly, when a conductor reads a score, they will not perceive the music through the abstract application of a set of music-theoretical rules but through the mental restaging of the notated musical scene (cf., Zatorre and Halpern 2005).

4.5 Interference Effects in Memory for Timbre

An important part of the characterization of auditory memory concerns the question of whether timbre is encoded and stored independently from other auditory attributes. In this section, three specific scenarios will be described that address aspects of interference in short-term memory for timbre, effects of musical timbre on long-term melodic memory, and effects of voice timbre on verbal memory.

4.5.1 Interference in Short-Term Memory

Short-term storage of timbre is closely related to the perceptual encoding stage. In basic perceptual experiments that have tested the independence of pitch and timbre, results indicate that pitch and timbral brightness information are integral attributes (Melara and Marks 1990; Allen and Oxenham 2014; and for a more detailed discussion, see McAdams, Chap. 2). There is evidence to suggest that interactions between pitch and timbre extend to memory.

Siedenburg and McAdams (2018) studied the short-term recognition of timbre by using a serial matching task wherein participants judged whether the timbres of two subsequent (standard and comparison) sequences of tones were of the same order. When the tone sequences comprised concurrent variation in pitch, the performance of nonmusicians was impaired more strongly than was the performance of musicians. When pitch patterns differed across standard and comparison sequences, however, musicians showed impaired performances as well. This means that musicians may require higher degrees of complexity of pitch patterns in order to exhibit impaired timbre recognition. More generally speaking, these results indicate that pitch and timbre are not encoded independently in short-term memory—these features are part of an integrated memory trace.

The topic of pitch-timbre interference implies an answer to the question of whether the defining units of working memory are constituted by integrated auditory events (called *sound objects*) or by individual features. Joseph et al. (2015) investigated the recognition of narrowband noise segments. Two features of these

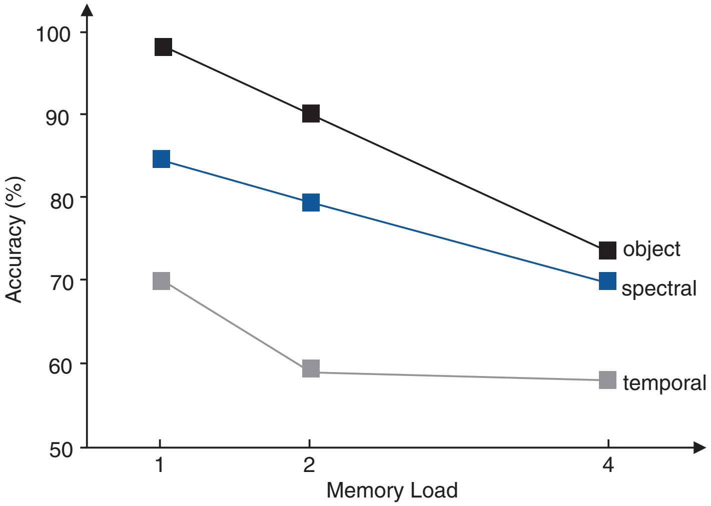

Fig. 4.5 Accuracy by memory load and condition in the item recognition task. Participants were required to match a probe sound to sounds from a previously presented sequence of length 1, 2, or 4 (*Memory Load*) according to *spectral*, *temporal*, or both spectral and temporal features (*object*). (Recreated from Joseph et al. 2015; used with permission from Frontiers/Creative Commons)

sounds were manipulated in the experiment: the spectral passband (i.e., yielding differences in the spectral centroid) and the amplitude modulation (AM) rate imposed on the waveform. Listeners were presented with a sequence of three sounds (each of 1 s duration with 1 s inter-stimulus intervals). They were instructed to judge whether there was a match between the third probe sound and the first or second sound presented. In two feature conditions, a match was defined as having one identical feature: passband or AM rate. In the object condition, a match was defined as both features being identical. As depicted in Fig. 4.5, accuracy in the object condition exceeded that in the feature condition by far (although accuracy for the spectral feature alone was better compared to the AM feature alone). This means that even if the task required participants only to memorize individual component features, there was a significant extraction cost when features had to be encoded and recollected individually.

Whether concerning the interference of pitch and timbre (Siedenburg and McAdams 2018) or spectral and temporal features of noise realizations (Joseph et al. 2015), the empirical evidence indicates that the content of short-term storage appears to be integrated auditory events (or "objects" as termed by Joseph et al. 2015) rather than individual features. The same position will be corroborated in the following review of effects of timbre on memory for melodies.

4.5.2 Timbre and Long-Term Melodic Memory

This section summarizes studies that have investigated the effects of timbre on melodic memory at time spans in the range of at least several minutes, which generally would be considered as LTM rather than STM processes. Although timbre does not affect a melody's pitch and rhythm structure, many studies have highlighted the role of timbre as a salient auditory feature for memorizing melodies. In experiments by Radvansky et al. (1995), participants identified which of two test melodies, a target and a distractor, was heard in the experiment's exposure phase. The accuracy of recognition judgements by both musicians and nonmusicians was higher when the timbre of the test melody equaled the timbre of the exposure melody, that is, a change in instrumentation clearly impaired melody recognition. This result was replicated with a sample of 6-month-old infants (Trainor et al. 2004).

Using richer musical stimuli, Poulin-Charronnat et al. (2004) studied recognition memory for tonal music (Liszt) and atonal contemporary music (Reynolds). A change of instrumentation from piano to orchestra or vice versa impaired recognition of tonal excerpts in both musicians and nonmusicians compared to conditions in which the instrumentation was held constant. For contemporary music, recognition performance by musicians was strongly impaired for instrumentation changes, whereas there was no effect for nonmusicians who performed poorly regardless of instrumentation. Halpern and Müllensiefen (2008) observed that the detrimental effect of timbre change is unaffected by whether the participant's attention at the exposure stage was directed toward timbral features (through an instrument categorization task) or to the melodic structure (through a judgement of melody familiarity).

Most recently, Schellenberg and Habashi (2015) explored the temporal dynamics of musical memory by testing melody recognition with delays between the exposure and the test that spanned 10 min, 1 day, and 1 week. Surprisingly, recognition accuracies were similar for all three retention intervals, and there even seemed to be a trend for consolidation as reflected by a small but significant increase in accuracy for a delay of 1 week compared to 10 min. Pitch transpositions of six semitones or a tempo shift of sixty-four beats per minute impaired recognition after 10 min and 1 day but not after 1 week. Notably, a change of instrument from piano to saxophone impaired melody recognition as strongly as the aforementioned changes in pitch or tempo but, unlike these parameters, the effect of timbre change did not reduce over time. This means that in contrast to key or tempo shifts, timbre information was not abstracted over time but stayed integral to the identity of the melody.

Schutz et al. (2017) considered melodic memory and object-to-melody association with a specific focus on the role of the amplitude envelopes of tones, which are closely related to the ways in which a sounding object is set into vibration. The excitations of a resonator by an impact usually generate rapid increases and exponentially decaying amplitude envelopes, whereas continuous excitations generate amplitude envelopes that tend to be rather flat. Schutz et al. (2017) let participants listen to melodies consisting of four pure tones with a flat or an exponentially decaying envelope. Each melody was presented three times and listeners were asked to associate the melody with a household object (e.g., digital clock, keys, calculator, etc.) that was physically presented by the experimenter during the presentation of the melodies. After a delay of more than 6 min, participants were presented with a recognition and recollection task; if melodies were identified as old, listeners also were asked to recall the associated object. Although their results only exhibited insignificant trends toward better melody recognition for percussive envelopes, melody-to-object association was significantly better for tones with percussively decaying envelopes. In two additional experiments, the authors observed that melodies of tones with reverse-ramped (i.e., increasing) envelopes were poorly associated with objects (performance was even worse than with flat envelopes). The results indicated that associative memory was better for decaying envelopes compared to flat or reversed envelopes, potentially due to their higher ecological familiarity. Although it may not be clear a priori why this stimulus manipulation only had an effect on associative memory but not on recognition memory, differences between associative and recognition memory are frequently observed in the literature (Kahana 2012).

Taken together, these studies strongly suggest that memory for melodies does not solely draw from an abstract lexicon of melodies represented by pitch interval information. Instead, melody recognition appears to rely on a rich auditory representation that integrates various features including timbre. Similar results have been found for verbal memory as described in the next section.

4.5.3 Timbre and Verbal Memory

The classic study on the role of voice timbre in spoken word recognition was conducted by Goldinger (1996) (for a discussion of more recent studies, see Goh 2005). Goldinger (1996) let participants listen to sequences of words recorded by 2, 6, or 10 different speakers. After three different delay periods, participants were required to distinguish old from new words in a recognition task. The results indicated that listeners better recognized words spoken by the voices of the exposure phase: the same-voice advantage was 7.5% after 5 min, 4.1% after 1 day, and an unreliable 1.6% after 1 week. Beyond the coarse same/different distinction, however, there was also a more fine-grained correlation of voice similarity with the percentage of correct rejections. In a second experiment, the delay interval was held constant at 5 min, but there were three different encoding conditions. Using a speeded classification task, participants either categorized voice gender, the initial phoneme from a list of alternatives, or the word's syntactic category (e.g., verb versus adjective). From a levels-of-processing perspective (Craik and Lockhart 1972), these tasks enforce shallow (gender), intermediate (phoneme), or deep (syntax) encoding of the words, respectively. The word recognition scores were as expected in that hit rates increased with the depth of encoding (i.e., gender < phoneme < syntax). The strength of the voice effect was reversed across encoding conditions. Whereas old voices had an advantage of around 12% for the gender condition, this advantage shrank to around 5% in the syntax condition. Because the effects were robust to a variety of encoding conditions, Goldinger (1996) concluded that the results "support an episodic view of the lexicon, in which words are recognized against a background of countless, detailed traces. Speech is not a noisy vehicle of linguistic content; the medium may be an integral dimension of later representation" (p. 1180). These findings suggest that the long-standing idea of the *mental lexicon* (Oldfield 1966), supposedly based on an amodal representation of words, is not enough to account for human recognition of spoken words.

Van Berkum et al. (2008) specifically investigated the time course of the integration of speaker and message information. In their experiment, participants passively listened to sentences while electroencephalography (EEG) signals were recorded. In two anomalous conditions, sentences could either feature *semantic anomalies* (e.g., Dutch trains are *sour* and blue; target word in italic) or *speaker inconsistencies* (e.g., I have a large *tattoo* on my back, spoken with an upper-class accent). They found that semantic anomalies elicited a standard N400 response for deviant trials, that is, an inflection of the EEG signal with a negative peak around 400 ms after the target word. Interestingly, the same time course was observed for the speaker inconsistency condition, where a similar N400 response was observed (albeit of much smaller magnitude). The clear onset of the deviant EEG response at around 200–300 ms after the acoustic onset of the deviant word indicated the rapid extraction and processing of timbre-specific information. These results suggest that voice-specific information is integrated into linguistic processing around the same point in time when language interpretation mechanisms construct meaning based on the lexical content of the words.

In sum, the studies presented in this section have shown strong associations between individual features in auditory memory. Because of their shared underlying tonotopic dimension, pitch and timbral brightness may be particularly intertwined. However, some of the evidence suggests that even amplitude envelope features affect aspects of melodic memory (Schutz et al. 2017). If features are to be accessed, recollected, or recognized individually, an extraction cost can be assumed (Joseph et al. 2015). This cost may be reduced by enhanced auditory attention and listening experience to some extent, but it is unlikely to ever vanish completely (Allen and Oxenham 2014; Siedenburg and McAdams 2018). The notion of integrated memory representations appears to contradict the seemingly abstract nature of auditory cognition (e.g., Obleser and Eisner 2009; Patel 2008). Sensory information related to timbre is not simply "left behind" in the process of information transduction from sensory to more symbolic forms of representations. On the contrary, timbre stays integral to both word and melody recognition over long retention spans—the medium and the message are intertwined.

4.6 Familiarity and Voice Superiority

The last theme in this review of memory for timbre concerns the roles of long-term familiarity with sound sources. A sound source of particular relevance and familiarity for humans is the voice. For that reason, the role of the voice in timbre processing has been studied with particular scrutiny (for an overview, see Mathias and Kriegstein, Chap. 7). This section discusses studies that have investigated the role of long-term familiarity with musical-instrument sounds in timbre processing (Sect. 4.6.1) and the special status of voice timbre in melodic memory (Sect. 4.6.2).

4.6.1 Familiarity in Short-Term Recognition

A factor that significantly increases the complexity of STM research and modeling relates to the presumption that STM is not completely distinct from LTM as suggested by procedural memory approaches. In fact, there is further evidence to assume a strong link between the two systems (e.g., Jonides et al. 2008). The experimental cornerstone regarding this link in verbal memory research is the *lexicality effect*: short-term memory for the identity of words or syllables (i.e., verbal items) is generally better for words than for *pseudowords* or nonsense syllables (Thorn et al. 2008). Pseudowords are defined as meaningless strings of letters that respect a language's phonotactic constraints but are not part of the dictionary (e.g., bech, chaf, tog, wesh, etc.).

Similar enhancements of STM performance have also been demonstrated for related linguistic variables, including word frequency and imaginability (Thorn et al. 2008). The analogous question for timbre, and of particular concern for the current purpose, is whether STM is facilitated by long-term familiarity with sounds produced by well-known musical instruments. If this were the case, it would constitute a timbral analogy to the verbal lexicality effect. More importantly, it would suggest that STM for timbre cannot be properly placed in a one-size-fits-all principle of sensory persistence—one would need to consider existing auditory categories as well.

To study the role of familiarity in STM for timbre, Siedenburg and McAdams (2017) compared the recognition of recorded tones from familiar acoustical instruments with that of unfamiliar synthesized tones that do not readily evoke sound-source categories. Steps were taken in order to manipulate familiarity while controlling for dissimilarity relations within the stimulus set. First, the spectrotemporal signal envelopes and temporal fine structures of recorded sounds were mismatched to generate novel and unfamiliar sounds. Second, familiarity ratings by musicians were collected for the transformed sounds, ensuring that the transformed sounds used in the main experiment were rated as significantly less familiar compared to the original recordings. Third, the main experiment used an item recognition task with sequences of three sounds. The mean timbral dissimilarity between the sounds in the sequence and those in the probe was equalized across recordings and transformations, using previously obtained pairwise dissimilarity ratings. Two experiments revealed greater recognition accuracy for timbres of familiar recorded sounds compared to unfamiliar transformations, as well as better performance at shorter delays (2 s versus 6 s), but no interaction between the factors of delay and stimulus material. These results point toward a generally more robust form of encoding of timbral properties coming from familiar acoustical instruments. The superior memory performance for familiar instruments proved to be independent of effects of perceptual similarity.

Prior knowledge of instrument categories for familiar acoustical-instrument sounds helps to associate sounds with auditory knowledge categories or schemas. In other words, familiar instrument sounds activate not only auditory sensory representations but, possibly to some extent, also activate semantic, visual, and even sensorimotor networks. These sounds are not necessarily rehearsed in STM, but could act as representational anchors for the associated auditory sensory traces. Saitis and Weinzierl (Chap. 5) further describe the nuanced cross-modal associations that timbre can elicit.

The special role of sound source familiarity has gained support from neurophysiological studies on timbre processing. Pantev et al. (2001) observed that professional trumpet players and violinists exhibited stronger event-related potentials to sounds from their own instrument at around 100 ms after sound onset (the N1 component), indexing stronger pre-attentive processes related to stimulus detection. In addition, there is evidence that learning not only affects cortical activity but can even modulate low-level processing in the brainstem. Strait et al. (2012) demonstrated that recordings of electrical brainstem activity taken from pianists more closely correlated with the amplitude envelopes of the original piano sounds when compared to recordings taken from musicians who did not play the piano as their primary instrument. However, brainstem activity did not differ between pianists and other musicians for sounds from the tuba and the bassoon. This result indicates that there may be instrument-specific neural adaptations that affect the perceptual processing of certain classes of instrumental sounds. Apparently, musical training can affect the fine-tuning of subcortical structures to more efficiently process sounds that are of particular relevance to the listener. These findings refute the idea that timbre could be a less important auditory surface feature. On the contrary, elementary aspects of auditory processing appear to be shaped by experience with sound source categories.

Unfortunately, none of the studies discussed here have been able to completely control low-level factors and the individual experience of the participants. Therefore, the exact origins of the effects may remain contentious. Future experiments that familiarize listeners with certain classes of novel timbres in the lab may help to more precisely characterize the underlying mechanisms of familiarity in timbre processing.

4.6.2 Voice Superiority

A sound source that all humans should be particularly familiar with, from both an evolutionary and ontogenetic point of view, is the human voice. Recent studies have suggested that sounds of vocal origin are faster and more robustly categorized compared to instrumental musical sounds. Many of these studies are also discussed in greater depth by Agus, Suied, and Pressnitzer (Chap. 3); hence, they will only be summarized here to set the stage for the consideration of additional memory effects.

Employing a go/no-go task, Agus et al. (2012) asked listeners to indicate as quickly as possible whether sounds were part of a target category (voice, percussion, or strings). Results showed faster reaction times for voices. Importantly, the effect did not arise for auditory chimeras that retained either spectral or temporal envelope shapes of vocal sounds. Suied et al. (2014) further observed that voices were more robustly recognized compared to other instrumental sounds even for very short snippets (durations from 2 ms to 128 ms). The exact acoustic features responsible for this advantage must be of spectrotemporal nature because neither solely spectral nor solely temporal cues sufficed to yield a processing advantage. Furthermore, Agus et al. (2017) only observed an increase of activity in areas of the human temporal lobe that have documented sensitivity to vocal stimuli (see Mathias and Kriegstein, Chap. 7) for nonchimaeric stimuli. This means that there are brain areas that selectively react to the full set of spectrotemporal cues of voices but not to isolated spectral or temporal cues.

Across several recent studies, Weiss and colleagues (see Weiss et al. 2017, and references therein) accumulated evidence for a memory advantage of vocal melodies compared to melodies played by nonvocal musical instruments (specifically piano, banjo, and marimba). In all of these studies, the basic experimental approach

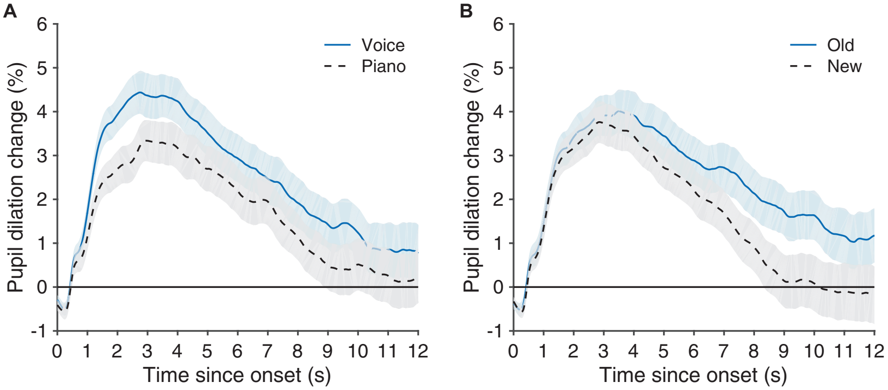

Fig. 4.6 Pupil dilation response as a function of the time since melody onset: (**A**) vocal versus piano melodies; (**B**) old versus new melodies. (From Weiss et al. 2016; used with permission of the American Psychological Association)

was to have participants listen to a set of melodies presented with a vocal or instrumental timbre. After a 5–10 min break, participants heard the exposure melodies intermixed with a set of new melodies and rated their confidence in having heard a melody previously on a seven-point scale. Analyses of the recognition ratings for old and new melodies revealed that adults more confidently and correctly recognized vocal compared to instrumental melodies (Weiss et al. 2012). The effect generalized to musicians with and without absolute pitch, and even pianists recognized more vocal melodies correctly with higher confidence in their correct ratings than for piano melodies (Weiss et al. 2015). This finding suggests that sensorimotor representations and perceptual familiarity with certain classes of sounds are an unlikely locus of the observed effect. Otherwise, pianists should have shown a reduced voice advantage due to their ability to recruit motor representations for piano melodies and to their high familiarity with piano sounds.

It was further shown that the presentation of vocal melodies, as well as previously encountered melodies, was accompanied by an increase in pupil dilation (Weiss et al., 2016). Increases in pupil dilation are generally interpreted as an indicator of heightened engagement and potentially a greater recruitment of attentional resources (Kang et al. 2014). The results by Weiss et al. (2016) are depicted in Fig. 4.6. Note that the difference in pupil dilation between piano and vocal melodies is most pronounced around 3 s after the onset of melodies. To the contrary, the difference between old and new melodies appears to accumulate across the full length of the melodies, indexing the distinct time courses of melody recognition and vocal superiority.

Although the memory advantage for melodies with a vocal timbre has turned out to be stable across several studies, there remain several open questions to explore within this paradigm (e.g., the role of signal amplitude normalizations, see Bigand et al. 2011). Most importantly, the psychophysical origin of any of the reported vocal superiority effects (Agus et al. 2012; Weiss et al. 2012) is not clear. Could vocal superiority be a result of the involvement of motor processes (Liberman and

Mattingly 1985)? Is there a particular spectrotemporal feature in the acoustics of voices that boosts the processing of these sounds? Or is it the case that all auditory stimuli that indicate a vocal sound source happen to be preferentially processed once a voice has been implicitly recognized? Differentiating these hypotheses would require disentangling top-down and bottom-up effects. As discussed in greater depth by Mathias and Kriegstein (Chap. 7), there are voice-selective areas in the auditory cortex that only react to vocal input sounds, even if low-level cues, such as temporal or spectral envelopes, are matched with other sounds (Agus et al. 2017). But what exactly is the representational content of these voice-selective areas? Is this cortical selectivity the origin or the result of vocal superiority? Future research may be able to shed light on these intriguing questions.

4.7 Summary and Future Perspectives

This chapter provides a review of important research threads in memory for timbre. These threads concern the role of perceptual similarity relations and chunking in short-term memory for timbre, active imagery of timbre, the role of interference of auditory attributes in memory, and questions regarding the privileged processing of familiar and vocal timbres. Only 10 years ago these topics had not been covered to any serious degree within auditory cognition research. Since then, many studies have been published that provide valuable insights into the processing of timbre in memory, but they also open up new perspectives for future research. Today, we think we have sufficient empirical grounds to formulate a few principles of how memory for timbre works. In the following, five such principles will be outlined, followed by a brief discussion of what we consider to be relevant questions for future research.

4.7.1 Principles of Memory for Timbre

In contrast to other sets of memory principles that have been proposed to hold for all types memory (Surprenant and Neath 2009), the current principles are specifically derived from empirical studies on timbre and they serve two purposes. First, these principles will act as concise summaries of the empirical data collected up to date. Second, they will be considered as intermediate explanations of empirical effects. From this perspective, a principle should be more abstract than an effect. At the same time, a principle can be less specific than a model because it does not need to provide a comprehensive list of components and their functional interrelations for the overall system. In this sense, the following principles highlight what we currently understand about memory for timbre but also expose how incomplete the current state of knowledge is. Figure 4.7 provides a schematic of how these processes could function for the example of an item recognition task.

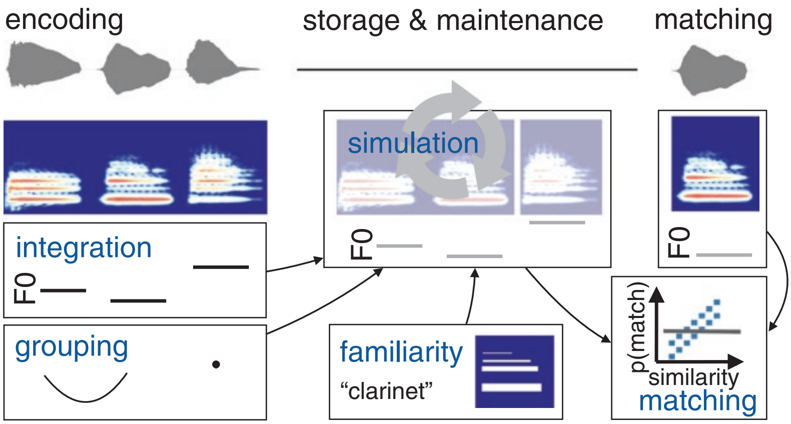

Fig. 4.7 Schematic of the five proposed principles of memory for timbre. The example shows how the five principles might relate to each other in a timbre memory recognition task. The auditory spectrogram indicates that the timbres of a sequence of sounds are represented in terms of their spectrotemporal properties. The structure of the memory trace is shaped by the process of *integration* (Principle I) as concurrently varying features, such as pitch, are integrated with the timbre memory trace. Sequential *grouping* (Principle II) provides additional temporal structure to the memory trace (in this example by separating the last sound from the first two). Timbre *familiarity* (Principle III) provides representational anchor points and cross-modal associations, for instance, by readily yielding semantic labels for certain sounds (here, the clarinet). Attention-based refreshing, a form of perceptual *simulation* (Principle IV), may be a maintenance strategy specifically suited for timbre. Here, perceptual simulation is graphically represented by a circle, denoting the cyclical process of refreshing the memory trace by means of attention. Finally, the *matching* (Principle V) stage takes the collection of features of a probe sound and compares them to stored memory traces. If the similarity measure exceeds the listener's internal threshold, the probe is considered a match

4.7.1.1 Integration: Timbre Information as Integrated Representations in Memory

Several experiments have shown that the perceptual discrimination of pitch and timbre (and more specifically, timbral brightness) is subject to symmetric interference effects (e.g., Allen and Oxenham 2014). As reviewed in Sect. 4.5.1, recent experiments on short-term recognition found detrimental effects of concurrent variations of irrelevant features (Joseph et al. 2015; Siedenburg and McAdams 2018) and hence suggested that integrated representations (or events/auditory objects) are stored in STM. The elaborations in Sect. 4.5.2 have illustrated that experiments on long-term melodic memory corroborated these findings. Whenever there is a shift of timbre, it is harder to discriminate new from old melodies (Schellenberg and Habashi 2015). The analogous effect even constitutes a classic effect in verbal memory: Spoken words are harder to recognize whenever they stem from a different speaker in the test phase (Goldinger 1996), and the time courses of semantic and speaker information processing are very similar (Van Berkum et al. 2008).

4.7.1.2 Grouping: Memory for Timbre Sequences is Affected by Grouping Cues

The item-to-item structure of auditory sequences strongly affects their mnemonic affordances. As reviewed in Sect. 4.3.2, hierarchically structured sequences are easier to chunk and encode compared to random sequences (Siedenburg et al. 2016). Furthermore, acoustic cues such as strong acoustic dissimilarity between statistically distinct groups of sounds may enhance the separated encoding of such groups (Tillmann and McAdams 2004). Whether based on chunking or acoustic dissimilarity, grouping cues powerfully enrich memory traces by structuring them along a hierarchy of time scales.

4.7.1.3 Familiarity: Better Memory Performance and Processing Accuracy

As discussed in Sect. 4.6, familiar sounds from well-known musical instruments are easier to recognize compared to unfamiliar transformed sounds (Siedenburg and McAdams 2017). Familiar musical-instrument sounds not only activate auditory sensory representations but to some extent also elicit semantic, visual, and even sensorimotor representations, which may act as anchors for the associated auditory sensory traces. Human voices may be considered as sound sources that are particularly familiar both ontogenetically and evolutionarily, and corresponding vocal advantage effects have been demonstrated (Agus et al. 2012; Weiss et al. 2012).

4.7.1.4 Perceptual Simulation: Active Memory Rehearsal and Timbre Imagery

As described in Sect. 4.4, short-term recognition of timbre can be impaired by attention-demanding tasks such as visual change detection (Siedenburg and McAdams 2017) or auditory imagery (Soemer and Saito 2015). Furthermore, precise timbre representations can be obtained through auditory imagery (Halpern et al. 2004), that is, through a simulation of sensory schemata from long-term memory. This means that timbre is part of an active form of auditory cognition that operates at the level of sensory representations.

4.7.1.5 Matching: Timbre Recognition via Similarity-Based Matching

The similarity effects observed by Siedenburg and McAdams (2018), as discussed in Sect. 4.3.1, suggest that a similarity-based matching mechanism could be at the basis of timbre recognition. This mechanism could be conceived as an ongoing computation of similarity of the current auditory input with past representations that are stored in memory. For item recognition tasks, the matching process could effectively be modeled as a similarity computation (Kahana 2012), indicating a match if the summed perceptual similarities of the probe item to the items in the memory sequence exceeds a certain threshold. Serial recognition tasks could be based on a matching process that computes item-wise dissimilarities between two sequences and hence corresponds to the dissimilarity of the swap criterion (Siedenburg and McAdams 2018).

4.7.2 Future Perspectives

We wish to close by discussing four potentially productive avenues for future research. Obtaining a more substantiated understanding of these questions appears to be of central importance for the topic of memory for timbre itself and might even have important implications for practical applications, such as music composition and production, sonification for human-computer interactions, and speech communication technology.

A first apparent gap in the literature concerns our knowledge about the basic memory persistence of different timbre features. For example, is a set of sounds varying along temporal features (e.g., the attack time) as easily retained in memory as sounds varying along spectral timbre features (e.g., brightness)? So far, most research has either considered minute details of spectral composition (e.g., McKeown and Wellsted 2009) or has not touched at all on the question of individual perceptual features, even if global similarity relations were considered (Golubock and Janata 2013; Siedenburg and McAdams 2017). An exception might be the experiments by Schutz et al. (2017), which indicated that flat amplitude envelopes are less well-suited for soundobject associations compared to percussive (i.e., exponentially decaying) envelopes.

Closely related to this question, and even more specific than the last point, is the need to specify the origin of vocal superiority effects. Two studies have already addressed this aspect in detail (Agus et al. 2012; Suied et al. 2014) but were not able to identify acoustic features that are specific to the vocal superiority effect. It is also not clear whether the recognition advantage observed by Weiss et al. (2012) has an acoustic or a cognitive origin. In other words, we still do not know what the basic acoustic or cognitive ingredients are that make memory for voices special.

Second, despite a plethora of memory models in other domains (e.g., Kahana 2012), there is no formal model of memory for timbre that predicts listeners' responses in memory tasks on the basis of the presented audio signals. The existence of such a model would mean a significant contribution, because it would help to make explicit the set of underlying assumptions of this research field. Perhaps the greatest hurdle for constructing a timbre memory model is the difficulty of agreeing on a signal-based representation for approximating the timbre features that are most relevant perceptually. Nonetheless, significant progress has been achieved over recent years regarding the latter (see McAdams, Chap. 2; Caetano, Saitis, and Siedenburg, Chap. 11; and Elhilali, Chap. 12).

Third, interindividual differences in memory for timbre and the role of formal musical training, as well as informal music learning, in memory for timbre have not been fully addressed yet. Whereas in timbre dissimilarity perception, musical training does not appear to affect perceptual space (McAdams et al. 1995), recognition memory for timbre may be more accurate in musicians compared to nonmusicians (Siedenburg and McAdams 2017). However, no rigorous attempt has been undertaken so far to control other individual differences that might act as confounding factors (e.g., verbal working memory, general cognitive ability). In addition, it is unclear whether the differences due to musical training observed for musicians versus nonmusicians also extend to varying levels of musical training found in the general population. It is unclear (a) how large individual differences in memory for timbre are, and (b) to what other cognitive abilities are these potential differences related. Insights regarding the latter questions might give an indication of the origin of individual differences in timbre memory. The development of a standardized test of timbre memory would represent a significant step forward in this respect. Such a test could build on existing experimental paradigms (Golubock and Janata 2013) for which factors that contribute to task difficulty have been studied already.

Finally, Agus et al. (2010) demonstrated a rapid and detailed form of implicit auditory memory for noise clips, and similar processes might be at play for the timbres of unfamiliar sound sources. Nonetheless, no study has yet addressed the time course of familiarization (i.e., learning trajectory) with sound sources. Lately, Siedenburg (2018) showed that the perception of brightness can be affected strongly by context effects. This implies that listeners not only memorize timbral associations within sequences of sounds, but the percept of a sound itself can be altered by the timbral properties of the auditory context. Hence, there exists an implicit form of memory for auditory properties, including timbre, that subconsciously affects present perceptual processing and that is in urgent need of further scientific exploration.

Acknowledgements KS was supported by a Carl von Ossietzky Young Researcher's Fellowship from the University of Oldenburg.

Compliance with Ethics Requirements Kai Siedenburg declares that he has no conflict of interest.
Daniel Müllensiefen declares that he has no conflict of interest.

References

Agus TR, Thorpe SJ, Pressnitzer D (2010) Rapid formation of robust auditory memories: insights from noise. Neuron 66:610–618

Agus TR, Suied C, Thorpe SJ, Pressnitzer D (2012) Fast recognition of musical sounds based on timbre. J Acou Soc Am 131(5):4124–4133

Agus TR, Paquette S, Suied C et al (2017) Voice selectivity in the temporal voice area despite matched low-level acoustic cues. Sci Rep 7(1):11526

Allen EJ, Oxenham AJ (2014) Symmetric interactions and interference between pitch and timbre. J Acous Soc Am 135(3):1371–1379

Andrillon T, Pressnitzer D, Léger D, Kouider S (2017) Formation and suppression of acoustic memories during human sleep. Nat Commun 8(1):179

Atkinson RC, Shiffrin RM (1968) Human memory: a proposed system and its control processes. In: Spence KW, Spence JT (eds) The psychology of learning and motivation: advances in research and theory (vol 2). Academic Press, New York, pp 89–195

Baddeley AD (2012) Working memory: theories models and controversies. Ann Rev Psy 63:1–29

Barsalou LW (1999) Perceptual symbol systems. Beh Brain Sci 22:577–660

Berz WL (1995) Working memory in music: a theoretical model. Music Percept 12(3):353–364

Bigand E, Delbé C, Gérard Y, Tillmann B (2011) Categorization of extremely brief auditory stimuli: domain-specific or domain-general processes? PLoS One 6(10):e27024

Camos V, Lagner P, Barrouillet P (2009) Two maintenance mechanisms of verbal information in working memory. J Mem Lang 61(3):457–469

Cowan N (1984) On short and long auditory stores. Psy Bull 96(2):341–370

Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Beh Brain Sci 24(1):87–114

Cowan N (2008) What are the differences between long-term short-term and working memory? Prog Brain Res 169:323–338

Cowan N (2015) Sensational memorability: working memory for things we see hear feel or somehow sense. In: Jolicoeur P, Levebre C, Martinez-Trujillo J (eds) Mechanisms of sensory working memory/ Attention and perfomance XXV. Academic Press, London, pp 5–22

Craik FI, Lockhart RS (1972) Levels of processing: a framework for memory research. J Verb Learn Verb Beh 11(6):671–684

Crowder RG (1993) Auditory memory. In: McAdams S, Bigand E (eds) Thinking in sound: the cognitive psychology of human audition. Oxford University Press, Oxford, pp 113–143

Darwin CJ, Turvey MT, Crowder RG (1972) An auditory analogue of the sperling partial report procedure: evidence for brief auditory storage. Cog Psy 3(2):255–267

D'Esposito M, Postle BR (2015) The cognitive neuroscience of working memory. Ann Rev Psych 66:1–28

Demany L, Semal C (2007) The role of memory in auditory perception. In: Yost WA, Fay RR (eds) Auditory perc of sound sources. Springer, New York, pp 77–113

Demany L, Trost W, Serman M, Semal C (2008) Auditory change detection: simple sounds are not memorized better than complex sounds. Psy Sci 19(1):85–91

Demany L, Pressnitzer D, Semal C (2009) Tuning properties of the auditory frequency-shift detectors. J Acou Soc Am 126(3):1342–1348

Demany L, Semal C, Cazalets J-R, Pressnitzer D (2010) Fundamental differences in change detection between vision and audition. Exp Brain Res 203(2):261–270

Deutsch D (1970) Tones and numbers: specificity of interference in immediate memory. Sci 168(3939):1604–1605

Dudai Y (2007) Memory: it's all about representations. In: Roediger HL III, Dudai Y, Fitzpatrick SM (eds) Science of memory: concepts. Oxford University Press, Oxford, pp 13–16

Goh WD (2005) Talker variability and recognition memory: instance-specific and voice-specific effects. J Exp Psy:LMC 31(1):40–53

Goldinger SD (1996) Words and voices: episodic traces in spoken word identification and recognition memory. J Exp Psy: LMC 22(5):1166–1183

Golubock JL, Janata P (2013) Keeping timbre in mind: working memory for complex sounds that can't be verbalized. J Exp Psy: HPP 39(2):399–412

Hagoort P, Indefrey P (2014) The neurobiology of language beyond single words. Ann Rev Neuosci 37:347–362

Halpern AR, Müllensiefen D (2008) Effects of timbre and tempo change on memory for music. Q J Exp Psy 61(9):1371–1384

Halpern AR, Zatorre RJ, Bouffard M, Johnson JA (2004) Behavioral and neural correlates of perceived and imagined musical timbre. Neuropsy 42(9):1281–1292

James W (1890/2004) The principles of psychology. http://www.psychclassicsyorkuca/James/Principles. Accessed 9 Nov 2015

Jonides J, Lewis RL, Nee DE et al (2008) The mind and brain of short-term memory. Ann Rev Psy 59:193–224

Joseph S, Kumar S, Husain M, Griffiths T (2015) Auditory working memory for objects vs features. Front Neurosci 9(13). https://doi.org/10.3389/fnins201500013

Kaernbach C (2004) The memory of noise. Exp Psy 51(4):240–248

Kahana MJ (2012) Foundations of human memory. Oxford University Press, New York

Kang OE, Huffer KE, Wheatley TP (2014) Pupil dilation dynamics track attention to high-level information. PLoS One 9(8):e102463

Lerdahl F, Jackendoff R (1983) A generative theory of tonal music. MIT Pr, Cambridge

Liberman AM, Mattingly IG (1985) The motor theory of speech perception revised. Cognition 21:1–36

Ma WJ, Husain M, Bays PM (2014) Changing concepts of working memory. Nat Neurosci 17(3):347–356

McAdams S, Winsberg S, Donnadieu S et al (1995) Perceptual scaling of synthesized musical timbres: common dimensions specificities and latent subject classes. Psy Res 58(3):177–192

McDermott JH, Schemitsch M, Simoncelli EP (2013) Summary statistics in auditory perception. Nat Neurosci 16(4):493–498

McKeown D, Wellsted D (2009) Auditory memory for timbre. J Exp Psy: HPP 35(3):855–875

McKeown D, Mills R, Mercer T (2011) Comparisons of complex sounds across extended retention intervals survives reading aloud. Perception 40(10):1193–1205

Melara RD, Marks LE (1990) Interaction among auditory dimensions: timbre pitch and loudness. Perc Psyphys 48(2):169–178

Meyer LB (1956) Emotion and meaning in music. Chicago U Pr, Chicago

Miller GA (1956) The magical number seven plus or minus two: some limits on our capacity for processing information. Psy Rev 63(2):81–97

Müllensiefen D, Halpern AR (2014) The role of features and context in recognition of novel melodies. Music Percept 31(5):418–435

Nees MA, Corrini E, Leong P, Harris J (2017) Maintenance of memory for melodies: articulation or attentional refreshing? Psy Bull Rev 24(6):1964–1970

Obleser J, Eisner F (2009) Pre-lexical abstraction of speech in the auditory cortex. Tr Cog Sci 13(1):14–19

Oldfield RC (1966) Things words and the brain. Q J Exp Psy 18(4):340–353

Pantev C, Roberts LE, Schulz M et al (2001) Timbre-specific enhancement of auditory cortical representations in musicians. Neur Rep 12(1):169–174

Patel AD (2008) Music language and the brain. Oxford University Press, Oxford

Poulin-Charronnat B, Bigand E, Lalitte P et al (2004) Effects of a change in instrumentation on the recognition of musical materials. Music Percept 22(2):239–263

Radvansky GA, Fleming KJ, Simmons JA (1995) Timbre reliance in non-musicians' and musicians' memory for melodies. Music Percept 13(2):127–140

Saffran JR, Johnson EK, Aslin RN, Newport EL (1999) Statistical learning of tone sequences by human infants and adults. Cogn 70:27–52

Saxena SK (2008) The art of Tabla rhythm: essentials tradition and creativity. In: New vistas in Indian performing arts. DK Printworld Ltd, New Dehli

Schellenberg EG, Habashi P (2015) Remembering the melody and timbre forgetting the key and tempo. Mem Cog 43(7):1021–1031

Schulze K, Koelsch S (2012) Working memory for speech and music. A NY Ac Sci 1252(1):229–236

Schulze K, Tillmann B (2013) Working memory for pitch timbre and words. Mem 21(3):377–395

Schutz M, Stefanucci JK, Baum SH, Roth A (2017) Name that percussive tune: Associative memory and amplitude envelope. Q J Exp Psy 70(7):1323–1343

Siedenburg K (2018) Timbral Shepard-illusion reveals perceptual ambiguity and context sensitivity of brightness perception. J Acou Soc Am 143(2):EL-00691

Siedenburg K, McAdams S (2017) The role of long-term familiarity and attentional maintenance in auditory short-term memory for timbre. Mem 25(4):550–564

Siedenburg K, McAdams S (2018) Short-term recognition of timbre sequences: music training pitch variability and timbral similarity. Music Percept 36(1):24–39

Siedenburg K, Mativetsky S, McAdams S (2016) Auditory and Verbal Memory in North Indian Tabla Drumming. Psychomusicology 26(4):327–336

Simon HA (1978) Information-processing theory of human problem solving. In: Estes WK (ed) Handbook of learning and cognitive processes, vol 5, pp 271–295

Soemer A, Saito S (2015) Maintenance of auditory-nonverbal information in working memory. Psy Bull Rev 22(6):1777–1783

Starr GE, Pitt MA (1997) Interference effects in short-term memory for timbre. J Acou Soc Am 102(1):486–494

Strait DL, Chan K, Ashley R, Kraus N (2012) Specialization among the specialized: auditory brainstem function is tuned in to timbre. Cortex 48(3):360–362

Suied C, Agus TR, Thorpe SJ et al (2014) Auditory gist: recognition of very short sounds from timbre cues. J Acou Soc Am 135(3):1380–1391

Surprenant A, Neath I (2009) Principles of memory. Psy Pr, New York

Thorn AS, Frankish CR, Gathercole SE (2008) The influence of long-term knowledge on short-term memory: evidence for multiple mechanisms. In: Thorn AS, Page M (eds) Interactions between short-term and long-term memory in the verbal domain. Psy Pr, New York, pp 198–219

Tillmann B, McAdams S (2004) Implicit learning of musical timbre sequences: statistical regularities confronted with acoustical (dis)similarities. J Exp Psy: LMC 30(5):1131–1142

Trainor LJ, Wu L, Tsang CD (2004) Long-term memory for music: infants remember tempo and timbre. Dev Sci 7(3):289–296

van Berkum JJ, van den Brink D, Tesink CM et al (2008) The neural integration of speaker and message. J Cog Neurosci 20(4):580–591

Visscher KM, Kaplan E, Kahana MJ, Sekuler R (2007) Auditory short-term memory behaves like visual short-term memory. PLoS Bio 5(3):e56. https://doi.org/10.1371/journal.pbio.0050056

Weiss MW, Trehub SE, Schellenberg EG (2012) Something in the way she sings enhanced memory for vocal melodies. Psy Sci 23(10):1074–1078

Weiss MW, Vanzella P, Schellenberg EG, Trehub SE (2015) Pianists exhibit enhanced memory for vocal melodies but not piano melodies. Q J Exp Psy 68(5):866–877

Weiss MW, Trehub SE, Schellenberg EG, Habashi P (2016) Pupils dilate for vocal or familiar music. J Exp Psy: HPP 42(8):1061–1065

Weiss MW, Schellenberg EG, Trehub SE (2017) Generality of the memory advantage for vocal melodies. Music Percept 34(3):313–318

Zatorre RJ, Halpern AR (2005) Mental concerts: musical imagery and auditory cortex. Neur 47(1):9–12