Ann. N.Y. Acad. Sci. ISSN 0077-89

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Special Issue: *The Neurosciences and Music VI*
REVIEW

Musical and verbal short-term memory: insights from neurodevelopmental and neurological disorders

Anne Caclin[1,2] and Barbara Tillmann[1,2]

[1]Lyon Neuroscience Research Center (CRNL), Brain Dynamics and Cognition Team (DYCOG) and Auditory Cognition and Psychoacoustics Team, INSERM, U1028, CNRS, UMR5292, Lyon, France. [2]Université Lyon 1, Lyon, France

Address for correspondence: Anne Caclin, Lyon Neuroscience Research Center (CRNL), Brain Dynamics and Cognition Team (DYCOG) and Auditory Cognition and Psychoacoustics Team, INSERM U1028, 69675, Bron Cedex, France. anne.caclin@inserm.fr

Auditory short-term memory (STM) is a fundamental ability to make sense of auditory information as it unfolds over time. Whether separate STM systems exist for different types of auditory information (music and speech, in particular) is a matter of debate. The present paper reviews studies that have investigated both musical and verbal STM in healthy individuals and in participants with neurodevelopmental and neurological disorders. Overall, the results are in favor of only partly shared networks for musical and verbal STM. Evidence for a distinction in STM for the two materials stems from (1) behavioral studies in healthy participants, in particular from the comparison between nonmusicians and musicians; (2) behavioral studies in congenital amusia, where a selective pitch STM deficit is observed; and (3) studies in brain-damaged patients with cases of double dissociation. In this review we highlight the need for future studies comparing STM for the same perceptual dimension (e.g., pitch) in different materials (e.g., music and speech), as well as for studies aiming at a more insightful characterization of shared and distinct mechanisms for speech and music in the different components of STM, namely encoding, retention, and retrieval.

Keywords: congenital amusia; stroke; auditory short-term memory; auditory working memory; delayed matching-to-sample task; fMRI

Introduction

Auditory short-term memory (STM) allows for the retention of auditory information across short time scales (of the order of a few seconds to tens of seconds). It is closely related to working memory (WM), which entails manipulation of the memorized information.[a] As acoustic signals unfold over time, auditory STM is a critical ability to adequately process and understand these signals. Not surprisingly, numerous research has investigated auditory STM in humans (which, however, remains less often studied than its visual equivalent), with both behavioral approaches and neuroimaging. A vast majority of this research has used verbal stimuli or isolated tones, whereas STM for musical sequences has been studied less often.

Here, we focus on the research question to know whether auditory STM is a unitary phenomenon or it is based on separate subsystems for different types of information, in particular musical and verbal information. To this aim, we first review studies that have investigated this issue in healthy individuals, and then we discuss how studies of neurodevelopmental and neurological disorders shed light on the organization of auditory STM, both for its behavioral and cerebral aspects. It is worth noting that musical and verbal STM abilities have been most often studied separately, and that musical STM has been studied less often than verbal STM (for a recent review on the mechanisms of nonverbal auditory STM, focusing on tones, auditory objects, and time, see Ref. 1). To the best of our knowledge, musical STM has been investigated mainly in the

[a]Note that "working memory" is sometimes used in a larger sense encompassing situations not involving the manipulation of information. Here, we will use "working memory" only when the paradigm entails such a manipulation (e.g., backward task).

doi: 10.1111/nyas.13733

following populations: healthy participants (nonmusicians and musicians), congenital amusic individuals, and brain-damaged patients (see below for a few studies in other patient groups).

The current review integrates more generally in the expanding research field that compares perceptual and cognitive processes for music and language as well as their neural correlates. Earlier accumulated evidence, in particular from neuropsychological studies of brain-damaged patients, was in favor of separate neural networks for music and language processing,[2–4] while neuroimaging data and some behavioral data rather highlight shared networks for the processing of the two materials, with nevertheless some specialized aspects.[5–7] Here, we will analyze how this partial dichotomy applies to short-term and WM processes for both music and language materials. Music and language, as any acoustic signal, are processed in the brain including the extraction of a variety of auditory features: pitch, loudness, temporal features, timbre and phonemic features, as well as perceived location. The relative importance of these features might be different for music and language, with a particularly more prominent role of pitch in music (for melody, harmony, etc.) than in speech (for prosody), and only speech carrying phonemic information. In STM studies, it is critical to distinguish between (1) the types of sound material presented to participants: music, speech, or other; and (2) the to-be-remembered features (e.g., pitch, phoneme, timbre, etc.). It is worth pointing out that the terminology in the literature is not uniformly used: the terms of "musical STM" and "pitch STM" are used in different studies to refer to the same cognitive processes, that is, the STM of the pitches of musical sounds (as opposed to verbal sounds). In the following, we use "musical" versus "verbal" to refer to the types of sound material (melodies versus words), and, if not otherwise specified, we use "musical STM" when focusing on the pitch dimension, and "verbal STM" when focusing on phoneme (or syllables, or words) identity. In this review, we address the question of whether auditory STM is a unitary phenomenon from both points of view: the storage of distinct features (which can have different involvements in music and speech) and the storage of music and speech information (see Table 1 for a list of studies comparing musical and verbal STM). Beyond presenting a review of existing research, the present survey aims at identifying topics that warrant further investigation to build a comprehensive model of auditory STM.

Musical and verbal short-term memory in healthy individuals

STM lies in-between sensory memory (which consists of automatic storage of very recent auditory information, i.e., the past few seconds, with a rapid decline) and long-term memory (LTM), with STM mostly referring to the storage of recent auditory information over short-time scales (typically a few tens of seconds at most). STM can be studied with recall and recognition tasks. However, for nonmusicians, recall tasks are not easy to perform with musical stimuli, as they require singing or the use of musical notation. Hence, studies investigating musical STM or comparing musical and verbal STM typically use recognition paradigms, such as the classical delayed matching-to-sample task (DMST, but see Ref. 8 for a study using a recall task). In a DMST trial, a first stimulus (S1), which can be an isolated sound or a sound sequence, is presented, the participant has to memorize S1. After a delay (usually a few seconds), a second stimulus is presented (S2) and the participant's task is to report whether S1 and S2 are identical or different. Such a task entails at least the following processing steps: *encoding* (of S1) in memory, *retention* (over the delay), and *retrieval* of the memorized information (to compare S1 and S2). At least the encoding step could be shared with basic perceptual tasks (such as pitch discrimination[9]). Performance in DMSTs is dependent on numerous factors, some of them being controlled by the experimenter, such as memory load or duration of the retention delay, and others depending on the participant, notably the individual discrimination ability of the relevant feature(s) (see examples in Refs. 10 and 11).

Behavioral studies of musical and verbal short-term memory

In a seminal study, Deutsch reported that performance in a pitch memory task is affected by the presentation of irrelevant tones varying in pitch during the retention delay between the two to-be-compared stimuli, but not by the presentation of spoken numbers during the same retention delay.[12] This finding suggested an autonomous store for pitch in STM, but raised numerous questions as to how auditory

Table 1. Summary of the studies comparing musical and verbal short-term memory in adult populations

Study	Feature(s) for musical STM	Feature(s) for verbal STM	Task	Measures	Participants	Main relevant finding(s)
Semal *et al.*[13]	Pitch	Pitch	DMST with isolated sounds and interfering material during the delay	Behavior	Healthy participants[a]	– Pitch memory (in tones or words) is affected by irrelevant tones or words varying in pitch
						– The nature of the irrelevant sounds (tones or words) has weak influence, if any
Schulze and Tillmann[22]	Pitch and timbre	Syllables	DMST (forward and backward tasks)	Behavior	Healthy participants (nonmusicians)	– Pitch memory more sensitive to backward task than timbre or word memory
						– Only memory for words sensitive to verbal interference
Benassi-Werke *et al.*[26]	Pitch and pitch contour	Digits and pseu-dowords	Recall tasks: forward and backward spans	Behavior	Healthy participants (amateur singers and professional singers)	– Musical memory decreases more strongly in the backward task than did verbal memory
Gorin *et al.*[29]	Pitch	Dissyllabic nonwords	DMST	Behavior	Healthy participants (nonmusicians)	– Similar time-based serial-order mechanisms in musical and verbal STM
Williamson *et al.*[8]	Pitch	Letters	Recall task	Behavior	Healthy participants (musicians and nonmusicians)	– Similar effects of pitch proximity and phonological proximity
Talamini *et al.*[34]	Various materials across studies		Meta-analysis	Behavior	Healthy participants (musicians and nonmusicians)	– Large musicianship effect for tonal STM
						– Moderate musicianship effect for verbal STM
Gorin *et al.*[36]	Pitch	Digits	Serial-order reconstruction	Behavior	Healthy participants (musicians and nonmusicians)	– Similar serial-order mechanisms in musical and verbal STM
						– Musicians use contour and metric cues for musical STM better than do nonmusicians
Hickok *et al.*[49]	Pitch	Nonsense speech	Rehearsal	fMRI	Healthy participants[a]	– Overlapping brain networks for musical and verbal STM
Koelsch *et al.*[50]	Pitch	Syllables	DMST	fMRI	Healthy participants (nonmusicians)	– Overlapping brain networks for musical and verbal STM
Schulze *et al.*[51]	Pitch	Syllables	DMST	fMRI	Healthy participants (musicians and nonmusicians)	– Overlapping brain networks for musical and verbal STM in both nonmusicians and musicians
						– Specific areas of activation for musical STM in musicians
Bittrich *et al.*[53]	Pitch	Syllables	DMST	EEG	Healthy participants (nonmusicians)	– EEG responses to old and new items are better differentiated for verbal stimuli than for musical stimuli
Jeong and Ryu[54]	Pitch and timbre	Words and syllables	Target sound recognition	fNIRS	Healthy participants (nonmusicians)	– Enhanced prefrontal interhemispheric asymmetries for pitch and timbre STM in comparison to verbal STM
Weiss *et al.*[57]	Pitch and rhythm	Syllables	Span for verbal material, DMST for musical material	Behavior	Musicians with and without dyslexia and nonmusician dyslexics	– Dyslexic musicians have impaired musical and verbal STM compared to nondyslexic musicians.
						– Dyslexic musicians have verbal and musical STM performance similar to that of dyslexic nonmusicians, but better performance in a pitch discrimination task
Tillmann *et al.*[63]	Pitch and timbre	Syllables	DMST	Behavior	Congenital amusics and nonmusician controls	– Selective deficit of musical STM in congenital amusia
						– Larger impairment of pitch (versus timbre) STM in congenital amusia
Tillmann *et al.*[67]	Pitch	Pitch	DMST	Behavior	Congenital amusics and nonmusician controls	– Pitch STM deficit for both musical and verbal material in congenital amusia
						– Pitch STM deficit is more pronounced for musical (versus verbal) material in congenital amusic individuals with large pitch discrimination thresholds
Hirel *et al.*[77]	Pitch	Syllables	DMST	Behavior and MRI	Stroke patients and matched controls	– Double dissociations with cases of impairment in verbal STM only or

Continued

Table 1. *Continued*

Study	Feature(s) for musical STM	Feature(s) for verbal STM	Task	Measures	Participants	Main relevant finding(s)
						musical STM only, and cases with impairments in both
						– No evidence for interhemispheric differences
Ménard and Belleville[79]	Pitch	Syllables	DMST	Behavior	Alzheimer's patients and matched controls	– Impairment in both musical and verbal STM (and LTM) in patients relative to controls, with no impairment in discrimination tasks
Unpublished, see Tillmann *et al.*[80]	Pitch	Syllables	DMST	Behavior	Landau–Kleffner syndrome patients and matched controls	– Impairment in both musical and verbal STM relative to controls

Note: See the main text for details. These studies were systematically searched for in the PubMed database using the keywords "musical," "verbal," and "short-term memory" (or "working memory"). We have added other relevant references to the best of our knowledge. Studies are presented with the same order as in the main text. Except the Semal *et al.* paper,[13] all these studies required participants to memorize sound sequences (or at least pitch variations in Tillmann *et al.*[67]). When not otherwise specified, DMSTs are forward tasks only.
[a]Musical training of participants not reported.

STM is organized. Are verbal and musical information stored separately? Are different auditory qualities (here pitch and timbre) stored separately? How pitch memory traces are organized and in particular, could this initial finding be an effect of the pitch distance between the relevant tones and the irrelevant speech sounds?[13] Since 1970, numerous behavioral studies have investigated these issues, revealing that auditory STM does not seem to be a unitary system, but might include some autonomous storage of distinct auditory features.

Short-term memory of distinct auditory features. A first set of studies used the interference paradigm developed by Deutsch.[12] This paradigm requires the comparison of two sounds separated by a delay filled with irrelevant sounds. By using various combinations of (1) the to-be-memorized auditory features and (2) the auditory features of irrelevant sounds presented during the retention delay, a fairly autonomous storage of the different features was observed with, in particular, separate memory traces for pitch and timbre,[14,15] pitch and loudness,[16,17] or pitch and location[18] (see Ref. 19 for converging evidence with 1- and 2-back tasks). Proximity (perceptual similarity) effects were observed for all auditory features. For example, pitch memory traces are degraded when the interfering tones have fundamental frequencies close to that of the to-be-remembered tone, and timbre memory traces are degraded by interfering tones of similar timbres.[15] Perceptual properties (e.g., perceived sound location) and not acoustic features of the sounds (e.g., interaural time or level differences) appear to be stored in STM.[20] It is worth noting that this relative independence of the memory traces for various auditory features does not preclude a storage of auditory objects as an entity in memory.[21]

Short-term memory of pitch in speech and music. The studies using the interference paradigm reviewed above have mostly used nonverbal stimuli. In one study using either speech sounds or musical tones as the to-be-memorized material and either of the two stimuli as the interfering sounds, only very limited evidence for a separate storage of pitch information in verbal and musical material was observed (Table 1), in particular when controlling for the pitch distance between the different types of sounds.[13] Studies using the interference paradigm required participants to memorize single sounds. More recent studies have used sound sequences rather than isolated events to compare verbal and musical STM (see Table 1 for a summary presentation).

Short-term memory of verbal and musical sequences. One study systematically compared STM and WM for musical sequences (pitch and timbre sequences) and verbal sequences (word sequences) with forward and backward recognition tasks (backward tasks require manipulating (reversing) the information held in STM).[22] The findings suggest that memory mechanisms are different for the three types of sequences. Only WM for word sequences (but not WM for timbre and pitch sequences) was sensitive to a verbal interference task during the retention delay of a backward recognition task. Length effects were more pronounced for pitch and word sequences than for timbre sequences. Overall result patterns suggest that pitch and word sequences might rely on similar rehearsing mechanisms (subvocal rehearsal), while timbre might rather be stored as acategorical information using sensory memory traces that are not sensitive to verbal interference (in agreement with Refs. 23 and 24, but see Ref. 25 for contrasting results). Memory for pitch sequences decreased more strongly in the backward task compared to the forward task than memory for word and timbre sequences, suggesting either a specific role of contour-information and tonal structure in pitch memory, facilitating performance in the forward task, or an effect of expertise with the auditory material (in favor of words) when its manipulation is requested, facilitating performance in the backward task. A more deleterious influence of backward tasks, relative to forward tasks, for tones as compared to words was also observed in recall tasks administered to musician participants.[26] Evidence for a distinction between the functioning of STM for verbal and nonverbal information was also obtained in studies using recall paradigms with speech and environmental sounds.[27,28]

Despite distinct mechanisms for the building and/or maintenance of representations in memory, musical and verbal STM might share domain-general (or amodal) time-based serial-order processes. This has been proposed in a recent study showing similar sensitivity to rhythmic interference for the two materials in a serial-order hybrid recognition/recall task without interference in the item task (see experiment 2 in Ref. 29). Other common principles in musical and verbal STM have been reported, with similar effects of pitch proximity and phonological proximity in particular.[8]

Summary. Behavioral data have demonstrated that auditory STM is not a unitary system, even though common structuring principles (e.g., time-based mechanisms and proximity effects) have been observed in different subsystems. Studies with nonverbal stimuli have provided ample evidence for a separate storage of different auditory features, in particular pitch and timbre.[14,15] Studies comparing tone sequences (musical memory, in particular pitch or timbre memory) and word sequences (verbal memory) support the view that these materials are encoded and maintained with somewhat different mechanisms. However, very few data allow for a link between findings on STM of auditory features (pitch, timbre, rhythm, and location) and findings on STM of more complex sound materials (speech, music, and environmental sounds). Are the differences between musical and verbal memory only a consequence of low-level acoustic differences between the two types of sounds? Or are there two, at least partly, separate systems for STM of speech and music sounds? Neuroimaging studies have endeavored to shed light on these questions, as reviewed below. We will now discuss studies that have assessed effects of expertise on auditory STM, as they are also informative with regard to the existence or not of separate subsystems within auditory STM.

Effects of expertise on musical and verbal short-term memory. Studies with expert participants allow for uncovering the domain-specificity of trained skills. Not surprisingly, musicians have been shown to outperform nonmusicians in a variety of auditory tasks with musical stimuli, but interestingly, these improvements have been shown to extend to other domains, in particular language and reading skills (for a review, see Ref. 30). Improvements due to musical training can be observed for a variety of perceptual and cognitive tasks, including speech-in-noise perception,[31] auditory selective attention,[32] and executive functions.[33]

A meta-analysis of studies comparing STM and WM (as well as LTM) in musicians and nonmusicians was recently performed.[34] It confirmed that musicians perform better than nonmusicians in memory tasks. For STM and WM, this expertise effect was large for tonal stimuli, moderate for verbal stimuli, and small or null for visuospatial stimuli. The advantage for verbal stimuli in musicians has been observed to be larger when stimuli are presented auditorily, compared to visually.[35] In a recent study comparing serial-order effects in verbal and musical STM in musicians and nonmusician controls, similar effects, such as sequence length, primacy and recency effects, were observed for musical and verbal materials in both participant groups.[36] Better performance for the musical task was observed in musicians relative to controls, possibly due to an efficient use of configuration (contour) and metrical cues. Overall, studies comparing musicians and nonmusicians are in keeping with the hypothesis of only partly separate subsystems for verbal and musical memory.

Neuroimaging studies of musical and verbal short-term memory

Neuroimaging studies of auditory STM (e.g., Refs. 37–40) have revealed the involvement of distributed brain networks encompassing superior temporal areas (in the auditory cortex), frontal areas (dorso-lateral prefrontal cortex and ventrolateral prefrontal cortex), parietal areas (in particular in the supra-marginal gyrus), and other brain regions (e.g., cerebellum, basal ganglia, and premotor areas). As for the behavioral studies reviewed above, we review neuroimaging studies that have addressed (1) whether STM for different auditory features (manipulated in nonverbal sounds) relies on separate networks (or not), (2) whether STM for music and speech relies on separate networks (or not), and (3) how the comparison between musicians and nonmusicians shed light on these issues.

Neuroimaging studies of STM for distinct auditory features. Several studies using nonverbal material have revealed at least partly distinct brain networks involved in the STM of various auditory features, for example, EEG/MEG studies using pitch and location STM tasks. During the encoding of the sounds, differences are observed in the generators of early evoked responses such as N1 or N2, depending on the to-be-encoded features.[41,42] This segregation seems to be maintained by associative areas activated in the P3 latency range.[43] Late slow waves observed during the retention delay (i.e., later than 500 ms poststimulus onset), however, are affected by memory load, but not by the to-be-encoded features.[42,44] Regarding other auditory features, a positron emission tomography (PET) study comparing pitch and rhythm memory has revealed different brain networks for the two dimensions, a fronto-parieto-temporal network for pitch STM, and a network encompassing cerebellar and cortical areas for rhythm STM.[45] The dissociation between pitch and rhythm STM was confirmed in a tDCS study revealing that the involvement of left and right parietal cortex depends on the feature to memorize.[46] These results are in agreement with mismatch negativity studies that have revealed distinct storage of various features in auditory sensory memory (see Refs. 47 and 48 for examples).

Neuroimaging studies of STM for musical and verbal sequences. Neuroimaging studies directly comparing verbal and musical STM remain rather scarce. A first functional magnetic resonance imaging (fMRI) study reported overlapping brain networks in superior temporal, inferior parietal, and frontal areas for the STM of speech (nonsense sentences) and music (novel melodies).[49] In agreement with this observation, a second fMRI study with sung syllables revealed very similar networks during the rehearsal of syllables (verbal information) or pitches (tonal information).[50] In another fMRI study comparing musicians with nonmusicians, considerable overlap in the cerebral networks engaged in verbal and musical STM was observed. However, it also uncovered specific areas of activation for musical STM in musicians (in frontal and subcortical structures in particular), suggesting that at least in this population, neural correlates related to STM of the two materials are partly segregated.[51]

Overall, fMRI data support the view that some brain areas involved in STM are shared for verbal and musical materials: frontal areas (including Broca's area and premotor areas), inferior parietal structures, and auditory areas (review in Ref. 52). Only limited evidence for specialized brain networks for speech and music STM arose in fMRI data, and comes from the comparison of musician data and nonmusician data. However, it remains possible that specialized networks coexist within the same brain regions for musical and verbal STM, but that the spatial resolution of classic fMRI paradigms and analyses does not allow for disentangling them.[7] Some evidence for distinct mechanisms for verbal and musical memory was observed with EEG[53] and fNIRS.[54] We will now tackle the question of the existence of separate networks for musical and verbal STM from the perspective of neurodevelopmental and neurological disorders.

Selective impairment of musical short-term memory in congenital amusia

Auditory STM has been studied in dyslexia (and other language-related disorders) and in congenital amusia. Whereas a verbal STM deficit has been heavily documented in dyslexia and is considered a hallmark of this condition,[55,56] musical STM in dyslexia has received little attention. One study systematically compared musical and verbal STM in musicians with and without dyslexia, revealing a general auditory STM impairment in dyslexic musicians.[57] For the purpose of the present review on auditory STM, the most informative studies in neurodevelopmental disorders concern congenital amusia.

Congenital amusia, sometimes also referred to as "tone deafness," is a lifelong disorder of music perception and production (for a review, see Ref. 58). Individuals with congenital amusia have difficulty detecting when someone sings out of tune (including themselves), recognizing familiar tunes without lyrics, detecting a wrong or out-of-tune note, and memorizing even short melodies.[59,60] Studying STM in congenital amusia provides a unique opportunity to focus on the investigation of nonverbal auditory memory, with its associated deficits. Various studies have provided evidence for amusic individuals' impaired musical (pitch) STM, notably for all steps required for memory, such as encoding, maintenance, and retrieval (for a review, see Ref. 61). Deficits are also observed in pitch discrimination tasks in congenital amusia (e.g., Refs. 60 and 62), but importantly, STM deficits are still observed when the used pitch changes exceeded participants' individual pitch discrimination thresholds.[63] Amusics' deficit is central for the STM of tones and tone sequences, but it also extends to timbral material, even though to a lesser extent.[63,64] However, individuals with congenital amusia have unimpaired verbal STM. They do not differ from their matched controls when tested in the same way for the tone and timbre sequences with word sequences presented in a DMST,[63] and when tested with classical verbal memory spans.[65,66] These findings show that STM deficits in congenital amusia are specific to pitch (and to a lesser extent timbre), suggesting a pitch-memory system that is, at least partly, separated from verbal memory.

However, this conclusion is only valid when focusing on the memory of the content of the verbal material, such as the phonemes, syllables, or words carried by the signal, but not when focusing on the memory of the pitch information carried by the verbal material. When Mandarin or Thai words, which have the same phonemic content but differ in the pronounced pitch trajectory (e.g., /ma/ rising or falling in pitch), are used in STM paradigms, amusic individuals (native French speakers) were impaired in comparison to their matched controls.[67] Interestingly, for amusics with the worst pitch discrimination thresholds (as measured with tones), the pitch memory deficit with the verbal material was less pronounced than the memory deficit with nonverbal analogue materials imitating the pitch trajectory with a musical timbre. Thus, speech may enhance performance in pitch discrimination[68,69] and STM[67] tasks in amusia, even though it does not necessarily restore normal processing.

Musical and verbal short-term memory in brain-damaged patients

In addition to the neurodevelopmental disorder of congenital amusia, auditory STM has also been examined in brain-damaged patients, with some of them also showing (acquired) amusia. Studies of verbal STM with either visual or auditory stimuli in patients with cortical lesions revealed the involvement of left temporal auditory areas and left inferior parietal lobe in phonological memory, as well as of the left inferior frontal lobe, which seems to be involved in particular in the articulatory rehearsal component.[70–73] Musical (pitch) STM was assessed in patients who underwent partial temporal and/or frontal resection (for the relief of pharmacoresistant epilepsy), revealing musical STM deficits in patients with right-sided resections.[74] A recent study reported that nonfluent forms of primary progressive aphasia (related to a selective degeneration of language networks with a left-sided prominence of the anomalies)[75] include a WM deficit for nonverbal sounds, in particular in the time (rhythm) domain.[76] The patients under study exhibited stronger deficits in tasks requiring the processing of sequences of tones (pitch STM and rhythm sequence tasks) as compared to psychoacoustic tasks using isolated sounds or tone pairs.

To the best of our knowledge, only three studies have directly compared musical and verbal STM in patients with neurological disorders. These studies also allow us to make links between auditory STM and visual STM, and between STM and more general cognitive abilities. STM for pitch sequences and word sequences was systematically assessed in patients with unilateral middle cerebral artery infarct.[77] Individual data of the stroke patients revealed double dissociations with cases of impairment in verbal STM only or musical STM only, but also patient cases with impairments in both tasks.

However, in contrast to what could be expected from previous neuropsychological studies having tested verbal and musical STM separately (see above), there was no evidence for interhemispheric differences in this study. Exploratory analyses using voxel-based lesion-symptom mapping were in keeping with the hypothesis of the involvement of the dorsal auditory pathway for both musical and verbal STM (see Ref. 78 for a review on language pathways), and a potential additional involvement of the ventral auditory pathway in musical STM. In this study with stroke patients,[77] we also investigated visual STM. At the individual level, we observed double dissociations between auditory and visual STM deficits; but across all patients, we also observed a correlation between auditory verbal and visual (nonverbal) STM performance, which points to at least partially shared mechanisms in auditory and visual STM.

Musical and verbal STM have been assessed jointly in patients with Alzheimer's disease (AD) with impairments observed for both materials (i.e., syllable and tone sequences).[79] Performance in musical and verbal STM was correlated, suggesting mechanisms shared between auditory domains (possibly related to sequential-order processing). Memory deficits are central in AD and in addition to STM deficits, AD patients exhibited LTM deficits for musical (melodies) and verbal (pseudoword sequences) stimuli, but no deficits in discrimination tasks. However, the deficits between STM and LTM were not correlated, raising questions on the relationships between short- and LTM.

Finally, musical and verbal STM are being assessed jointly in a currently ongoing study, investigating patients with the Landau–Kleffner syndrome (see Ref. 80 for a preliminary report). In this rare epileptic encephalopathy, a severe verbal agnosia is observed in the active phase of the disease in childhood. By testing musical and verbal STM in adults who suffered from this disorder in childhood, we are observing that the deficits affect not only verbal STM, but also nonverbal musical STM. The results thus far reveal an impairment that is more general than just related to the verbal deficits that are the hallmark of this pathology.

Overall, studies in patients with neurological disorders have revealed that deficits of either musical STM only, verbal STM only, or both types of STM can be observed in various conditions. The patterns of associations and dissociations of the deficits suggest both shared and distinct mechanisms for musical and verbal STM, as well as some shared mechanisms between auditory and visual STM. The current (limited) data call for a more detailed assessment of the roles of the involved cerebral structures and the respective contribution of the two cerebral hemispheres in auditory STM as a function of the material. Auditory STM deficits are observed in a variety of neurological conditions, where the prominent symptoms are observed in different realms: speech and language (such as in primary progressive aphasia or Landau–Kleffner syndrome) or memory (in AD). Future studies in various patient groups are needed to refine our understanding of the relationships between STM and other cognitive abilities, and to further characterize shared and distinct mechanisms between STM for auditory (music and speech) and visual materials.

Synthesis and discussion

The behavioral, neuroimaging, and neuropsychological findings reviewed here highlight that human auditory STM is not a unitary system. At least partly separate storage for different auditory features (pitch, timbre, loudness, localization, and rhythm) has been demonstrated with behavioral and neuroimaging methods in typical (healthy) individuals (including control participants). This was confirmed with the observed selective deficits (for nonverbal material) in some conditions: for example, a deficit of pitch STM in congenital amusia,[61] and a deficit of nonverbal STM in a form of primary progressive aphasia that was most prominent in the rhythm domain (versus pitch or timbre).[76]

Whether or not musical and verbal STM are separate systems remains a matter of debate. Musical STM is often reduced to pitch STM in the literature, and verbal STM to the memory of linguistic content (phonemes, syllables, and words). Musical/pitch STM and verbal/content STM are at least partly separate systems, as evidenced from (1) behavioral studies in healthy participants (e.g., Ref. 22), (2) behavioral and fMRI studies comparing nonmusicians and musicians,[34,51] and (3) studies in congenital amusia[63] and in stroke patients.[77] However, as highlighted in Table 1, only very few studies to date have compared the storage of the same perceptual dimension (e.g., pitch) in musical versus verbal material, an aspect that thus warrants further investigation to fully decipher the relationships between music and language processing in the brain. A study with healthy individuals was in favor of only very few differences between the storage of pitch in verbal and nonverbal stimuli,[13] whereas a study in congenital amusia revealed more pronounced pitch STM deficits with musical stimuli than with verbal stimuli in this population.[67]

In addition to the evidence summarized above in favor of (partly) separate systems for verbal and musical STM, a number of results point to shared mechanisms between STM in the two domains, namely, (1) behavioral findings in healthy participants, suggesting common mechanisms for some aspects of STM (e.g., subvocal rehearsal[22] and time-based serial-order mechanisms[29]); (2) fMRI data in healthy participants showing similar brain networks involved in STM for the two materials;[50] and (3) the joint alteration of verbal and musical STM in various patient groups, including some stroke patients,[77] patients with AD,[79] and patients suffering from the Landau–Kleffner syndrome.[80] We can very tentatively speculate that shared mechanisms between domains are found in particular in the active maintenance phase of STM, while encoding and the formation of memory traces might entail more specific mechanisms. A more detailed investigation of the alterations of the different processing steps involved in STM (encoding, maintenance, and retrieval) in pathological brains, on the one hand, and of the refinements of these various processing steps in expert participants (e.g., musicians), on the other hand, should provide highly relevant data to construct a comprehensive model of auditory (and more general) STM. Neuroimaging studies (in particular EEG/MEG studies with their high temporal resolution) could shed light on these issues, notably to reveal the underlying neural dynamics (which could differ according to sound materials) with connectivity analyses and/or analyses of oscillatory activities. Note that a similar proposal has been made regarding spatial versus nonspatial STM in nonverbal stimuli.[41–43]

Taken together, the currently available research does not yet allow for drawing a full picture of shared versus specialized mechanisms of STM for musical and verbal materials, even though providing some insights. The literature survey provided here highlights the need of more systematic research testing verbal and musical STM in the same participants, both for healthy (including experts versus novices) and patient populations. For that goal, it is crucial to aim for the same experimental implementation of the two material types with comparable levels of task difficulty. Because of the critical role of STM processes in audition and the relation between STM and more general cognitive abilities, the investigation of musical and verbal STM involves the potential to develop perspectives for training and rehabilitation for the impaired brain as it is currently the case for perception (e.g., rhythm processing[81]).

Acknowledgments

This work was supported by grants from "Agence Nationale de la Recherche" (ANR) of the French Ministry of Research ANR-11-BSH2-001-01 to B.T. and A.C., ANR-14-CE30-0001-01 to A.C., and ANR-16-CE28-0012-02 to B.T. This work was conducted in the framework of the LabEx CeLyA ("Centre Lyonnais d'Acoustique," ANR-10-LABX-0060) and of the LabEx Cortex ("Construction, Function and Cognitive Function and Rehabilitation of the Cortex," ANR-11-LABX-0042) of Université de Lyon, within the program "Investissements d'avenir" (ANR-11-IDEX-0007) operated by the French National Research Agency (ANR).

Competing interests

The authors declare no competing interests.

References

1. Joseph, S., S. Teki, S. Kumar, *et al.* 2016. Resource allocation models of auditory working memory. *Brain Res.* **1640:** 183–192.
2. Peretz, I., R. Kolinsky, M. Tramo, *et al.* 1994. Functional dissociations following bilateral lesions of auditory cortex. *Brain J. Neurol.* **117**(Pt 6): 1283–1301.
3. Griffiths, T.D., A. Flees & G.G.R. Green. 1999. Disorders of human complex sound processing. *Neurocase* **5:** 365–378.
4. Peretz, I. & M. Coltheart. 2003. Modularity of music processing. *Nat. Neurosci.* **6:** 688–691.
5. Patel, A.D. 2008. *Music, Language, and the Brain.* New York: Oxford University Press.
6. Schön, D., R. Gordon, A. Campagne, *et al.* 2010. Similar cerebral networks in language, music and song perception. *Neuroimage* **51:** 450–461.
7. Peretz, I., D. Vuvan, M.-É. Lagrois, *et al.* 2015. Neural overlap in processing music and speech. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370:** 20140090.
8. Williamson, V.J., A.D. Baddeley & G.J. Hitch. 2010. Musicians' and nonmusicians' short-term memory for verbal and musical sequences: comparing phonological similarity and pitch proximity. *Mem. Cognit.* **38:** 163–175.

9. Albouy, P., M. Cousineau, A. Caclin, *et al.* 2016. Impaired encoding of rapid pitch information underlies perception and memory deficits in congenital amusia. *Sci. Rep.* **6:** 18861.
10. Jiang, C., V.K. Lim, H. Wang, *et al.* 2013. Difficulties with pitch discrimination influences pitch memory performance: evidence from congenital amusia. *PLoS One* **8:** e79216.
11. Albouy, P., J. Mattout, R. Bouet, *et al.* 2013. Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain J. Neurol.* **136:** 1639–1661.
12. Deutsch, D. 1970. Tones and numbers: specificity of interference in immediate memory. *Science* **168:** 1604–1605.
13. Semal, C., L. Demany, K. Ueda, *et al.* 1996. Speech versus nonspeech in pitch memory. *J. Acoust. Soc. Am.* **100:** 1132–1140.
14. Semal, C. & L. Demany. 1991. Dissociation of pitch from timbre in auditory short-term memory. *J. Acoust. Soc. Am.* **89:** 2404–2410.
15. Starr, G.E. & M.A. Pitt. 1997. Interference effects in short-term memory for timbre. *J. Acoust. Soc. Am.* **102:** 486–494.
16. Semal, C. & L. Demany. 1993. Further evidence for an autonomous processing of pitch in auditory short-term memory. *J. Acoust. Soc. Am.* **94:** 1315–1322.
17. Clément, S., L. Demany & C. Semal. 1999. Memory for pitch versus memory for loudness. *J. Acoust. Soc. Am.* **106:** 2805–2811.
18. Clarke, S., M. Adriani & A. Bellmann. 1998. Distinct short-term memory systems for sound content and sound localization. *Neuroreport* **9:** 3433–3437.
19. Anourova, I., P. Rämä, K. Alho, *et al.* 1999. Selective interference reveals dissociation between auditory memory for location and pitch. *Neuroreport* **10:** 3543–3547.
20. Mathias, S.R. & K. von Kriegstein. 2014. Percepts, not acoustic properties, are the units of auditory short-term memory. *J. Exp. Psychol. Hum. Percept. Perform.* **40:** 445–450.
21. Joseph, S., S. Kumar, M. Husain, *et al.* 2015. Auditory working memory for objects vs. features. *Front. Neurosci.* **9:** 13.
22. Schulze, K. & B. Tillmann. 2013. Working memory for pitch, timbre, and words. *Memory* **21:** 377–395.
23. McKeown, D. & D. Wellsted. 2009. Auditory memory for timbre. *J. Exp. Psychol. Hum. Percept. Perform.* **35:** 855–875.
24. McKeown, D., R. Mills & T. Mercer. 2011. Comparisons of complex sounds across extended retention intervals survives reading aloud. *Perception* **40:** 1193–1205.
25. Siedenburg, K. & S. McAdams. 2017. The role of long-term familiarity and attentional maintenance in short-term memory for timbre. *Memory* **25:** 550–564.
26. Benassi-Werke, M.E., M. Queiroz, R.S. Araújo, *et al.* 2012. Musicians' working memory for tones, words, and pseudowords. *Q. J. Exp. Psychol.* **65:** 1161–1171.
27. Rowe, E.J. & W.G. Rowe. 1976. Stimulus suffix effects with speech and nonspeech sounds. *Mem. Cognit.* **4:** 128–131.
28. de Gelder, B. & J. Vroomen. 1997. Modality effects in immediate recall of verbal and non-verbal information. *Eur. J. Cogn. Psychol.* **9:** 97–110.
29. Gorin, S., B. Kowialiewski & S. Majerus. 2016. Domain-generality of timing-based serial order processes in short-term memory: new insights from musical and verbal domains. *PLoS One* **11:** e0168699.
30. Tierney, A. & N. Kraus. 2013. Music training for the development of reading skills. *Prog. Brain Res.* **207:** 209–241.
31. Slater, J., E. Skoe, D.L. Strait, *et al.* 2015. Music training improves speech-in-noise perception: longitudinal evidence from a community-based music program. *Behav. Brain Res.* **291:** 244–252.
32. Strait, D.L., J. Slater, S. O'Connell, *et al.* 2015. Music training relates to the development of neural mechanisms of selective auditory attention. *Dev. Cogn. Neurosci.* **12:** 94–104.
33. Habibi, A., A. Damasio, B. Ilari, *et al.* 2018. Music training and child development: a review of recent findings from a longitudinal study. *Ann. N.Y. Acad. Sci.* https://doi.org/10.1111/nyas.13606.
34. Talamini, F., G. Altoè, B. Carretti, *et al.* 2017. Musicians have better memory than nonmusicians: a meta-analysis. *PLoS One* **12:** e0186773.
35. Talamini, F., B. Carretti & M. Grassi. 2016. The working memory of musicians and nonmusicians. *Music Percept. Interdiscip. J.* **34:** 183–191.
36. Gorin, S., P. Mengal & S. Majerus. 2017. A comparison of serial order short-term memory effects across verbal and musical domains. *Mem. Cognit.* **46:** 464–481.
37. Zatorre, R.J., A.C. Evans & E. Meyer. 1994. Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* **14:** 1908–1919.
38. Schumacher, E.H., E. Lauber, E. Awh, *et al.* 1996. PET evidence for an amodal verbal working memory system. *Neuroimage* **3:** 79–88.
39. Gaab, N., C. Gaser, T. Zaehle, *et al.* 2003. Functional anatomy of pitch memory—an fMRI study with sparse temporal sampling. *Neuroimage* **19:** 1417–1426.
40. Crottaz-Herbette, S., R.T. Anagnoson & V. Menon. 2004. Modality effects in verbal working memory: differential prefrontal and parietal responses to auditory and visual stimuli. *Neuroimage* **21:** 340–351.
41. Anourova, I., V.V. Nikouline, R.J. Ilmoniemi, *et al.* 2001. Evidence for dissociation of spatial and nonspatial auditory information processing. *Neuroimage* **14:** 1268–1277.
42. Anurova, I., D. Artchakov, A. Korvenoja, *et al.* 2003. Differences between auditory evoked responses recorded during spatial and nonspatial working memory tasks. *Neuroimage* **20:** 1181–1192.
43. Anurova, I., D. Artchakov, A. Korvenoja, *et al.* 2005. Cortical generators of slow evoked responses elicited by spatial and nonspatial auditory working memory tasks. *Clin. Neurophysiol.* **116:** 1644–1654.
44. Rämä, P., L. Paavilainen, I. Anourova, *et al.* 2000. Modulation of slow brain potentials by working memory load in spatial and nonspatial auditory tasks. *Neuropsychologia* **38:** 913–922.
45. Jerde, T.A., S.K. Childs, S.T. Handy, *et al.* 2011. Dissociable systems of working memory for rhythm and melody. *Neuroimage* **57:** 1572–1579.
46. Schaal, N.K., B. Pollok & M.J. Banissy. 2017. Hemispheric differences between left and right supramarginal gyrus for pitch and rhythm memory. *Sci. Rep.* **7:** 42456.
47. Giard, M.H., J. Lavikahen, K. Reinikainen, *et al.* 1995. Separate representation of stimulus frequency, intensity, and

duration in auditory sensory memory: an event-related potential and dipole-model analysis. *J. Cogn. Neurosci.* **7:** 133–143.
48. Caclin, A., E. Brattico, M. Tervaniemi, *et al.* 2006. Separate neural processing of timbre dimensions in auditory sensory memory. *J. Cogn. Neurosci.* **18:** 1959–1972.
49. Hickok, G., B. Buchsbaum, C. Humphries, *et al.* 2003. Auditory–motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *J. Cogn. Neurosci.* **15:** 673–682.
50. Koelsch, S., K. Schulze, D. Sammler, *et al.* 2009. Functional architecture of verbal and tonal working memory: an FMRI study. *Hum. Brain Mapp.* **30:** 859–873.
51. Schulze, K., S. Zysset, K. Mueller, *et al.* 2011. Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Hum. Brain Mapp.* **32:** 771–783.
52. Schulze, K. & S. Koelsch. 2012. Working memory for speech and music. *Ann. N.Y. Acad. Sci.* **1252:** 229–236.
53. Bittrich, K., K. Schulze & S. Koelsch. 2012. Electrophysiological correlates of verbal and tonal working memory. *Brain Res.* **1432:** 84–94.
54. Jeong, E. & H. Ryu. 2016. Nonverbal auditory working memory: can music indicate the capacity? *Brain Cogn.* **105:** 9–21.
55. Liebel, S.W. & J.M. Nelson. 2017. Auditory and visual working memory functioning in college students with attention-deficit/hyperactivity disorder and/or learning disabilities. *Arch. Clin. Neuropsychol.* **32:** 980–991.
56. Männel, C., L. Meyer, A. Wilcke, *et al.* 2015. Working-memory endophenotype and dyslexia-associated genetic variant predict dyslexia phenotype. *Cortex* **71:** 291–305.
57. Weiss, A.H., R.Y. Granot & M. Ahissar. 2014. The enigma of dyslexic musicians. *Neuropsychologia* **54:** 28–40.
58. Tillmann, B., P. Albouy & A. Caclin. 2015. Congenital amusias. *Handb. Clin. Neurol.* **129:** 589–605.
59. Peretz, I., J. Ayotte, R.J. Zatorre, *et al.* 2002. Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* **33:** 185–191.
60. Ayotte, J., I. Peretz & K. Hyde. 2002. Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain J. Neurol.* **125:** 238–251.
61. Tillmann, B., Y. Lévêque, L. Fornoni, *et al.* 2016. Impaired short-term memory for pitch in congenital amusia. *Brain Res.* **1640:** 251–263.
62. Hyde, K.L. & I. Peretz. 2004. Brains that are out of tune but in time. *Psychol. Sci.* **15:** 356–360.
63. Tillmann, B., K. Schulze & J.M. Foxton. 2009. Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. *Brain Cogn.* **71:** 259–264.
64. Marin, M.M., B. Gingras & L. Stewart. 2012. Perception of musical timbre in congenital amusia: categorization, discrimination and short-term memory. *Neuropsychologia* **50:** 367–378.
65. Williamson, V.J. & L. Stewart. 2010. Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory* **18:** 657–669.
66. Albouy, P., K. Schulze, A. Caclin, *et al.* 2013. Does tonality boost short-term memory in congenital amusia? *Brain Res.* **1537:** 224–232.
67. Tillmann, B., D. Burnham, S. Nguyen, *et al.* 2011. Congenital amusia (or tone-deafness) interferes with pitch processing in tone languages. *Front. Psychol.* **2:** 120.
68. Patel, A.D., J.M. Foxton & T.D. Griffiths. 2005. Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain Cogn.* **59:** 310–313.
69. Tillmann, B., E. Rusconi, C. Traube, *et al.* 2011. Fine-grained pitch processing of music and speech in congenital amusia. *J. Acoust. Soc. Am.* **130:** 4089–4096.
70. Baldo, J.V. & N.F. Dronkers. 2006. The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology* **20:** 529–538.
71. Leff, A.P., T.M. Schofield, J.T. Crinion, *et al.* 2009. The left superior temporal gyrus is a shared substrate for auditory short-term memory and speech comprehension: evidence from 210 patients with stroke. *Brain J. Neurol.* **132:** 3401–3410.
72. Buchsbaum, B.R., J. Baldo, K. Okada, *et al.* 2011. Conduction aphasia, sensory–motor integration, and phonological short-term memory—an aggregate analysis of lesion and fMRI data. *Brain Lang.* **119:** 119–128.
73. Baldo, J.V., S. Katseff & N.F. Dronkers. 2012. Brain regions underlying repetition and auditory–verbal short-term memory deficits in aphasia: evidence from voxel-based lesion symptom mapping. *Aphasiology* **26:** 338–354.
74. Zatorre, R.J. & S. Samson. 1991. Role of the right temporal neocortex in retention of pitch in auditory short-term memory. *Brain J. Neurol.* **114**(Pt 6): 2403–2417.
75. Mesulam, M.-M., S. Weintraub, E.J. Rogalski, *et al.* 2014. Asymmetry and heterogeneity of Alzheimer's and frontotemporal pathology in primary progressive aphasia. *Brain J. Neurol.* **137:** 1176–1192.
76. Grube, M., R. Bruffaerts, J. Schaeverbeke, *et al.* 2016. Core auditory processing deficits in primary progressive aphasia. *Brain J. Neurol.* **139:** 1817–1829.
77. Hirel, C., N. Nighoghossian, Y. Lévêque, *et al.* 2017. Verbal and musical short-term memory: variety of auditory disorders after stroke. *Brain Cogn.* **113:** 10–22.
78. Catani, M. & M. Mesulam. 2008. The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *Cortex* **44:** 953–961.
79. Ménard, M.-C. & S. Belleville. 2009. Musical and verbal memory in Alzheimer's disease: a study of long-term and short-term memory. *Brain Cogn.* **71:** 38–45.
80. Tillmann, B., C. Hirel, Y. Lévêque, *et al.* 2017. Musical connections: music perception and neurological deficits. In *The Routledge Companion to Music Cognition.* R. Ashley & R. Timmers, Eds.: 225–235. New York, NY: Routledge.
81. Fujii, S. & C.Y. Wan. 2014. The role of rhythm in speech and language rehabilitation: the SEP hypothesis. *Front. Hum. Neurosci.* **8:** 777.