Ann. N.Y. Acad. Sci. ISSN 0077-8923

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES
Issue: *The Neurosciences and Music IV: Learning and Memory*

Working memory for speech and music

Katrin Schulze[1] and Stefan Koelsch[2]

[1]Developmental Cognitive Neuroscience Unit, UCL Institute of Child Health, London, United Kingdom. [2]Cluster "Languages of Emotion," Freie Universität Berlin, Berlin, Germany

Address for correspondence: Katrin Schulze, Developmental Cognitive Neuroscience Unit, UCL Institute of Child Health, 30 Guilford Street, London WC1N 1EH, UK. kschulze@ich.ucl.ac.uk

The present paper reviews behavioral and neuroimaging findings on similarities and differences between verbal and tonal working memory (WM), the influence of musical training, and the effect of strategy use on WM for tones. Whereas several studies demonstrate an overlap of core structures (Broca's area, premotor cortex, inferior parietal lobule), preliminary findings are discussed that imply, if confirmed, the existence of a tonal and a phonological loop in musicians. This conclusion is based on the findings of partly differing neural networks underlying verbal and tonal WM in musicians, suggesting that functional plasticity has been induced by musical training. We further propose a strong link between production and auditory WM: data indicate that both verbal and tonal auditory WM are based on the knowledge of how to produce the to-be-remembered sounds and, therefore, that sensorimotor representations are involved in the temporary maintenance of auditory information in WM.

Keywords: auditory working memory; verbal; tonal; musical expertise; strategy

Introduction

Working memory (WM) describes a brain system responsible for temporary storage and simultaneous manipulation of information,[1–3] which is critical for higher cognitive functions such as planning, problem solving, and reasoning, but also for understanding or appreciating speech and music.

The present paper is based on the WM model developed by Baddeley and Hitch,[3] which assumes an attentional control system (the "central executive") that operates in conjunction with two subsidiary systems: the visuospatial sketchpad and the phonological loop. The visuospatial sketchpad processes and stores visual and spatial information; the phonological loop represents verbal short-term memory (STM). The mutual interaction between long-term memory (LTM) and WM was recognized by the introduction of a fourth component to the model: the episodic buffer. This limited capacity system is assumed to bind information from the subsidiary systems, store information in a multimodal code, and enable the interaction between WM and LTM.[4]

We acknowledge that there are many other STM or WM models (for an overview, see Refs. 5, 6–11). Our paper, however, is theoretically embedded in the highly influential Baddeley and Hitch[3] WM model because studies exploring the question whether WM for music and language differs have been primarily based on this WM model.[3,12–18] Furthermore, although parts of this model are still discussed, no other model of verbal WM is as well investigated, developed, and accepted as the phonological loop.[19]

The terms STM and WM have not been used consistently in the literature.[20] One possibility is to use the term STM to refer to the simple temporary storage of information, and WM to refer to the maintenance and manipulation of information.[5,20] Often it is not well defined whether a task needs additional processing and/or manipulation;[a] therefore, no distinction has been made between STM and WM in the present paper, but we will refer only to WM instead. In addition, we will use the term

[a]For example, if participants in an auditory WM experiment have to decide whether one test stimulus was presented previously during a sequence consisting of several stimuli, it is not known whether or how much manipulation in addition to simply storing and rehearsing the auditory sequence is required.

doi: 10.1111/j.1749-6632.2012.06447.x

auditory WM to describe WM processes for verbal or nonverbal stimuli that were presented auditorily. Finally, it should be noted that whereas verbal WM experiments used both recall and recognition tasks, studies exploring WM for tone material were relying on recognition tasks (but see Ref. 12 for a recall task for musical stimuli).

Behavioral studies

WM for verbal information

Baddeley and Hitch[3] suggested a multicomponent WM model in which verbal information is processed by a phonological loop. This component can be further subdivided into a passive storage component (phonological storage) and an active rehearsal mechanism (articulatory rehearsal process). It is assumed that the passive storage component can store auditory or speech-based information for a few seconds.[2,8] If the information has to be maintained for longer, the articulatory rehearsal process can rehearse the verbal information, a process comparable to subvocal speech. The articulatory rehearsal can be interrupted by articulatory suppression,[1,2] which usually involves overt articulation (for example, Refs. 21 and 22–26), preventing the articulatory rehearsal mechanism to subvocally rehearse verbal material and thus reducing the verbal WM function. The word length effect, on the other hand, refers to the phenomenon that participants show a greater memory span[22] and a superior recognition accuracy[21] for short words than for long words. The effect of articulatory suppression and of word length suggested that during the articulatory rehearsal process, verbal material is maintained by using a phonological code, comparable to subvocal speech (for an overview, see Refs. 1, 2, and 5). The phonological similarity effect describes participants' inferior performance to recall[24] or recognize[27] phonologically similar verbal material compared to phonologically dissimilar verbal material.

WM for tonal information

The Baddeley and Hitch WM model[1,3] does not specify whether the phonological loop also serves the processing of nonphonological information, or whether different subsystems (a "tonal loop"[13] or a "musical loop"[14]) exist in addition to the phonological loop. As described previously, verbal information can be maintained in verbal WM by internal articulatory rehearsal. But does internal rehearsal also work for pitch information? Studies that investigated this question yielded conflicting results, indicating either a behavioral improvement of WM performance by internal rehearsal[13,16,18] or only a small improvement or no improvement at all.[28–30] However, these studies differed with regard to the degree to which participants could imitate and repeat the experimental stimuli. Experiments that find only a small or no effect of internal rehearsal used tones, whose frequencies did not correspond to the frequencies of the Western chromatic scale;[29,30] tones with a frequency difference smaller than the smallest difference, namely one semitone, used in songs of Western tonal music;[29,30] and/or chords consisting of several simultaneously played sine-wave tones.[28] In comparison, if studies used tones whose frequencies did correspond to the frequencies of the Western chromatic scale,[13,18] or if the frequency differences between the used tones were not smaller than one semitone,[13,16,18] then the observed results support the hypothesis of a rehearsal mechanism underlying WM for tones.

Comparison between verbal and tonal WM

WM for tones is fundamental for music perception and production. However, the majority of research on auditory WM has been carried out using verbal material, namely phonemes, syllables, and words. Research on WM for pitch or the "tonal loop"[13] is rather scant and does not yet provide a consistent picture. Deutsch[31] observed that presenting intervening tones interfered more strongly with a WM task for tones than presenting phonemes, and this was interpreted as evidence for a specialized tonal WM system. Further, Salame and Baddeley[32] showed that instrumental music interfered less with verbal WM compared to vocal music, supporting the theory of two independent WM systems for verbal and tonal stimuli. On the other hand, Semal *et al.*[33] criticized that the frequency relations between the standard tones and the intervening verbal material were not controlled in Deutsch's[31] study, which might explain the missing interference between the standard tones and the intervening verbal material. Pitch similarity of the intervening stimuli (words or tones) had a greater effect on the performance rate than the modality (verbal or tonal) of the intervening stimuli,[33] indicating that pitches for both verbal and tonal stimuli are processed in the same WM system. Along these lines, Chan *et al.*[34] reported that musical training increases verbal WM performance, indicating that rather overlapping mechanisms are underlying verbal and tonal WM. Further support for similarities between auditory verbal and tonal WM comes from an experiment using a suppression paradigm in musically experienced participants.[18] Musical (singing "la") and verbal (producing the words "the") suppression decreased recognition accuracy for both digit and tone sequences, indicating that musical or verbal suppression does not selectively impair verbal or tonal WM. In a recent study, Williamson *et al.*[12] compared WM recall for tones and letters. Their results suggest that well-known characteristics of verbal WM could also be observed for the tonal modality: WM for tonal information showed limited capacity, and nonmusicians, but not musicians, showed a decreased performance if the tone sequences consisted of more proximal pitches compared to more distal pitches, an effect resembling the phonological similarity effect in the verbal WM domain.

Neuroanatomical correlates of WM

WM for verbal information

As seen for the behavioral experiments, most studies investigating the functional neuroarchitecture of auditory WM used verbal material. Neuroimaging studies indicate that mainly Broca's area and premotor areas (as well as pre-SMA and SMA) play a crucial role during the internal rehearsal of verbal material.[35–39] In addition, evidence suggests that both the insular cortex[38,40,41] and the cerebellum[39,42,43] are involved in internal rehearsal of verbal information. Whereas the involvement of Broca's area and the premotor cortex during the internal rehearsal has been supported by numerous studies, the research results regarding the phonological store have been much less conclusive. The phonological store has been suggested to rely on parietal areas, particularly the inferior parietal lobule (IPL; Refs. 35, 37, 38, and 42–46), but also on the superior parietal lobule (SPL, Ref. 35). However, the localization of the phonological store in the parietal lobe is very controversial for several reasons. First, neural activity in this area might also reflect increased engagement of attentional resources.[47,48] Second, the reported coordinates for the phonological store differ greatly between studies,[19] and finally, the IPL is not activated during passive listening,[19,49] which should be the case if this structure is involved in automatically storing incoming auditory information as suggested by the WM model.[1,3,19]

Alternatively, area Spt (Sylvian–parietal–temporal, left posterior planum temporale) has been suggested to be involved in the temporary storage of verbal information during WM tasks.[19] This is because activation in the left Spt has been observed to be enhanced during the delay period of a WM task[17,50] and to be independent of the modality of the presented stimuli (auditory or visual; Ref. 50). On the basis of these findings and because area Spt also supports speech processing, it has been proposed that area Spt acts as an auditory–motor interface for WM.[17,19,50] This proposition fits nicely with the hypothesis of a dual-stream model of speech processing.[51–54] In this model, a ventral stream supports speech comprehension via a lexical access while a left dominant dorsal stream, which comprises also area Spt, enables sensory–motor integration, i.e., the mapping of the perceived speech signals onto articulatory representations.

WM for tonal information

In comparison to the underlying networks of the phonological loop, far fewer neuroimaging studies have investigated WM for tones. In participants who were not selected for musical expertise, Gaab *et al.*[55] showed activation of the supramarginal gyrus (SMG), the SPL, the planum temporale, premotor regions encroaching on Broca's area, and cerebellar regions during a pitch memory task. This network is surprisingly similar to the network subserving the phonological loop described above. A similar network, including the inferior frontal and insular cortex, the planum temporale, and the SMG, had previously been reported to be activated during the active retention of pitch.[56]

Comparison between verbal and tonal WM

To our knowledge, only three neuroimaging studies have directly compared the neural correlates underlying auditory WM for tonal and verbal material.[15–17] Hickok *et al.*[17] compared the neural correlates underlying verbal and tonal WM in nonmusicians using functional magnetic resonance imaging (fMRI). Melodic sequences (tonal condition) and sentences consisting of pseudowords (verbal condition) were presented auditorily, and subsequently participants rehearsed internally the verbal and tonal stimuli. Results showed that internal rehearsal of both verbal and tonal material activated the area Spt, Broca's area, and left premotor regions.[17] Very similar activations were observed in the study by Koelsch *et al.*,[16] in which similarities between the neural components underlying WM for verbal (syllables) and tonal (pitch) material were investigated using a recognition task. During the verbal rehearsal, a neural network comprising the premotor cortex, the anterior insula, the SMG/intraparietal sulcus (IPS), the planum temporale, the inferior frontal gyrus, pre-SMA, and the cerebellum was activated, mainly in the left hemisphere. Importantly, the neural network activated during the tonal rehearsal was virtually identical to that observed during verbal rehearsal. In an fMRI study by Schulze *et al.*,[15] similarities and differences of the functional networks underlying the internal rehearsal of verbal and tonal WM were investigated using a recognition task. Similar to our previous study,[16] both verbal and tonal WM-activated areas typically reported in previous experiments on either verbal[1,35,37,38] or tonal WM[17,55,56] in nonmusicians. The fact that both verbal and tonal WM activated these core structures, namely Broca's area, the left premotor cortex, (pre-)SMA, left insular cortex, and left IPL, corroborates previous results showing considerable overlap of the networks underlying verbal and tonal WM.[16,17] Importantly, only in nonmusicians, all structures involved in tonal WM were also involved in verbal WM; in contrast, verbal but not tonal WM relied on additional structures that have previously been implicated in verbal WM.[1,47,57] This difference in activation of WM resources in nonmusicians is reflected in the behavioral data that showed better performance during verbal compared to tonal WM.

In summary, consistent across studies,[15–17] data obtained from nonmusicians indicate a considerable overlap of neural resources underlying WM for both verbal and tonal information. This common network includes a mainly left-lateralized fronto-parietal network (premotor cortex, Broca's area, and in two of the three studies, the IPL,[15,16] the cerebellum,[15,16] and the planum temporale/area Spt[16,17]).[b]

Comparison between nonmusicians and musicians

Because speech is a fundamental human skill typically acquired during early childhood, nonmusicians can be considered to be trained in processing and producing speech, but they possess less expertise in the music domain. Thus, for a more balanced comparison of verbal and tonal WM, Schulze *et al.*[15] investigated in addition highly trained musicians. Interestingly, many of the structures, namely Broca's area, left premotor cortex, left insular cortex, (pre-)SMA, cingulate gyrus, and left IPL, which were activated more strongly in nonmusicians during verbal compared to tonal WM, were activated more strongly in musicians compared to nonmusicians during tonal WM. That is, the functional network on which nonmusicians relied for verbal WM was also used by musicians for tonal WM. In contrast to nonmusicians, musicians recruited a number of structures exclusively for either verbal or tonal WM. For tonal information these areas were the left cuneus, the right globus pallidus, and the right caudate nucleus, as well as the left cerebellum, and for verbal information the right insular cortex.

In addition, activation differences between verbal and tonal WM were observed in a number of structures in musicians, providing a first indication of the existence of two WM systems, namely a phonological loop maintaining phonological information and a tonal loop dedicated to the maintenance of tonal information. Both systems activated the same core structures of WM and therefore showed considerable overlap, but both systems also differed in that they relied on different neural subcomponents. Importantly, the structural differences between the verbal and tonal loop in musicians could not be explained simply by performance differences between the tonal and the verbal tasks, because several brain structures were recruited selectively for verbal or tonal WM (see discussion in Refs. 15 and 58). One hypothesis, based on the assumption of functional plasticity induced by music, is that musical expertise leads to a network comprising more structures underlying tonal WM, therefore showing a considerable overlap with the functional network

[b]The fMRI studies that detected activation in Spt used continuous scanning,[16,17,50] whereas Schulze *et al.*[15] used a sparse temporal sampling scanning technique that might have not been sensitive enough to capture Spt activation.

subserving verbal WM, but also exhibiting substantial differences.

Sensorimotor codes

To account for the similarities between verbal WM with speech production (and speech perception), the underlying representations of verbal WM have been termed sensorimotor codes.[59] The following results indicate that, indeed, internal verbal rehearsal shares some characteristics with speech production. The word-length effect and the articulatory-suppression effect suggest that verbal WM is comparable to subvocal speech (for an overview see Ref. 1). Furthermore, the phonological loop is conceived as a memory system involving internal articulatory speech actions implemented by motor-related areas such as Broca's area, premotor and insular cortex,[40] (pre-)SMA,[37–39] and the cerebellum.[39,42]

For tonal WM, findings indicate that internal rehearsal mainly improves WM performance for tones if participants are able to imitate and repeat the auditory stimuli,[13,16,18] in contrast to studies in which this might have been more difficult or impossible.[28–30]

Remarkably, the superior performance of nonmusicians during verbal compared to tonal WM, and the better performance of musicians compared to nonmusicians during tonal WM, were primarily associated with activation differences in structures known to be involved in the control, programming, and planning, in addition to execution of actions, such as Broca's area, premotor cortex, (pre-)SMA, left insular cortex, IPS, IPL, and the cerebellum.[15]

The behavioral and neurophysiological differences between WM for verbal and tonal information in nonmusicians were interpreted as a consequence of a more extensive production and rehearsal of verbal information in everyday life. Musicians, on the other hand, might have more elaborate sensorimotor codes underlying the internal rehearsal of tones compared to nonmusicians. This indicates functional plasticity induced by musical training and, more specifically, might be a consequence of musicians' long-term learning of associations between pitch information and motor actions.[60–64]

Previous research has established that Broca's area and the premotor cortex are involved in the planning and controlling of vocal and hand actions[65,66] and in auditory-motor mapping;[67] movement representations for both speech and music are supported by the anterior insula;[51,68,69] and these structures serve voluntary motor control and contribute to the programming, initiation, and execution of movements.[70–73]

Therefore, sensorimotor processes may assist with the representation and manipulation of information, and sensorimotor coding could play an important role for WM processes. This points toward a basic mechanism of auditory WM: to translate the sensory auditory event into a rehearseable sensorimotor code. Action-related sensorimotor codes are assumed to be based on motor knowledge—how to produce the auditory stimulus (e.g., syllable, tone)—and are thought to be involved in the rehearsal and representation of information in auditory, verbal, and tonal WM.[15–17,74]

The dual-stream model of speech perception[51–53] assumes that one of the functions of the dorsal path of the auditory system is sensory–motor integration, i.e., mapping the perceived speech signals onto articulatory representations. The left-dominant dorsal stream for sensory–motor integration involves structures at the parietal–temporal junction and projects to the premotor cortex and Broca's area,[52] structures that were also observed for verbal and tonal WM.[15–17] Speech production requires motor speech representations but also representations of sensory speech targets that are important for comparing between predicted and actual consequences of motor speech acts.[54] Furthermore, in a recent paper it has been suggested that sensorimotor integration also plays a role during singing.[75] In conclusion, we propose that internal rehearsal associated with auditory WM relies on sensorimotor representations, which might also be crucial for singing and speaking.

WM and strategy

The amount of information that can be maintained by the WM system is limited.[1,22,76] However, the use of a strategy, for example chunking the to-be-remembered information,[77,78] can improve WM performance. Chunking refers to a process in which elements of information are organized into one unit or chunk,[76] with stronger associations between items within one chunk than between chunks.[79] This process is assumed to be supported by the episodic buffer enabling features from different sources to be bound into chunks and new information to be integrated into an existing context stored in LTM.[5] Previously, the neural correlates underlying such strategy-based memorization were explored using visual–spatial or verbal material,[80–83] but it was mainly unknown whether a similar network is also involved during strategy-based WM for tones.

By using structured (all tones belonged to one tonality) and unstructured (atonal) five-tone sequences, Schulze *et al.*[84] investigated whether musical structure influences encoding and rehearsal in a nonverbal auditory WM task and how this is reflected in the brain of nonmusicians and musicians. Musicians, but not nonmusicians, showed better performance for structured than for unstructured sequences, indicating that musicians' knowledge about musical regularities[85–88] helped them to keep the structured sequences in WM. The data[84] in musicians showed that a lateral (pre-)frontal–parietal network, including the right inferior precentral sulcus, the premotor cortex, and the left IPS, was more strongly involved during WM for structured compared to unstructured auditory sequences. Previous research reported the involvement of a similar network during strategy-based WM processing for visual and auditory–verbal stimuli,[81–83] therefore pointing toward a modality-independent (pre-)frontal–parietal network subserving strategy-based WM.

In a behavioral study by Schulze *et al.*,[89] participants had to indicate whether two sequences were the same or different, the facilitating effect of tonality (structure) on WM performance for tones could be confirmed, and was also observed for nonmusicians. Tonality, however, only improved WM performance for tones during maintenance (forward task), but not during manipulation (backward task).

Summary and conclusion

This paper reviewed research results indicating differences and similarities between verbal and tonal WM related to the underlying mechanisms and neural correlates. Whereas the core structures, namely Broca's area, premotor cortex, and IPL, show a considerable overlap, these preliminary findings in musicians suggest that there are also different subcomponents activated either during verbal or tonal WM. These results indicate, if confirmed, the existence of both a tonal and a phonological loop in musicians. We further propose a strong link between production and auditory WM. Both verbal and tonal auditory WM appear to be based on the knowledge of how to produce the to-be-remembered sounds, and we therefore suggest that sensorimotor representations are involved in the temporary maintenance of auditory information in WM.

Conflicts of interest

The authors declare no conflicts of interest.

References

1. Baddeley, A.D. 2003. Working memory: looking back and looking forward. *Nat. Rev. Neurosci.* **4:** 829–839.
2. Baddeley, A.D. 1992. Working memory. *Science* **255:** 556–559.
3. Baddeley, A.D. & G.J. Hitch. 1974. Working memory. In *Recent Advances in Learning and Motivation.* Vol VIII. G.A. Bower, Ed.: 47–89. Academic Press. New York.
4. Baddeley, A.D. 2000. The episodic buffer: a new component of working memory? *Trends. Cogn. Sci.* **4:** 417–423.
5. Baddeley, A.D. 2011. Working memory: theories, models, and controversies. *Annu. Rev. Psychol.* doi:10.1146/annurev-psych-120710-100422.
6. Cowan, N. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychol. Bull.* **104:** 163–191.
7. Cowan, N. (1999) An embedded-processes model of working memory. In *Models of Working Memory*. A. Miyake & P. Shah, Eds.: 62–101. University Press. Cambridge.
8. Baddeley, A.D. 2010. Working memory. *Curr. Biol.* **20:** R136–R140.
9. Ericsson, K.A. & W. Kintsch. 1995. Long-term working-memory. *Psychol. Rev.* **102:** 211–245.
10. Jones, D.M. 1993. Objects, streams and threads of auditory attention. In *Attention: Selection, Awareness and Control.* A.D. Baddeley & L. Weiskrantz, Eds.: 87–104. Clarendon. Oxford, U.K.
11. Nairne, J.S. 1990. A feature model of immediate memory. *Mem. Cogn.* **18:** 251–269.
12. 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. Cogn.* **38:** 163–175.
13. Pechmann, T. & G. Mohr. 1992. Interference in memory for tonal pitch: implications for a working-memory model. *Mem. Cogn.* **20:** 314–320.
14. Berz, W.L. 1995. Working memory in music: a theoretical model. *Music Percept.* **12:** 353–364.
15. 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.
16. Koelsch, S. *et al.* 2009. Functional architecture of verbal and tonal working memory: an FMRI study. *Hum. Brain Mapp.* **30:** 859–873.
17. Hickok, G., B. Buchsbaum, C. Humphries & T. Muftuler. 2003. Auditory-motor interaction revealed by fMRI: speech,

music, and working memory in area Spt. *J. Cogn. Neurosci.* **15:** 673–682.
18. Schendel, Z.A. & C. Palmer. 2007. Suppression effects on musical and verbal memory. *Mem. Cogn.* **35:** 640–650.
19. Buchsbaum, B.R. & M. D'Esposito. 2008. The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* **20:** 762–778.
20. Cowan, N. 2008. What are the differences between long-term, short-term, and working memory? *Prog. Brain. Res.* **169:** 323–338.
21. Baddeley, A.D., D. Chincotta, L. Stafford & D. Turk. 2002. Is the word length effect in STM entirely attributable to output delay? Evidence from serial recognition. *Q. J. Exp. Psychol. A* **55:** 353–369.
22. Baddeley, A.D., N. Thomson & L. Buchanan. 1975. Word length and the structure of short-term memory. *J. Verbal Learn. Verbal Behav.* **14:** 575–589.
23. Neath, I., A.M. Surprenant & D.C. LeCompte. 1998. Irrelevant speech eliminates the word length effect. *Mem. Cogn.* **26:** 343–354.
24. Surprenant, A.M., I. Neath & D.C. LeCompte. 1999. Irrelevant speech, phonological similarity, and presentation modality. *Memory* **7:** 405–420.
25. Henson, R.N.A., T. Hartley, N. Burgess, *et al*. 2003. Selective interference with verbal short-term memory for serial order information: a new paradigm and tests of a timing-signal hypothesis. *Q. J. Exp. Psychol. A* **56:** 1307–1334.
26. Larsen, J.D. & A.D. Baddeley. 2003. Disruption of verbal STM by irrelevant speech, articulatory suppression, and manual tapping: do they have a common source? *Q. J. Exp. Psychol. A* **56:** 1249–1268.
27. Nimmo, L.M. & S. Roodenrys. 2005. The phonological similarity effect in serial recognition. *Memory* **13:** 773–784.
28. Demany, L., G. Montandon & C. Semal. 2004. Pitch perception and retention: two cumulative benefits of selective attention. *Percept. Psychophys.* **66:** 609–617.
29. Keller, T.A., N. Cowan & J.S. Saults. 1995. Can auditory memory for tone pitch be rehearsed? *J. Exp. Psychol. Learn. Mem. Cogn.* **21:** 635–645.
30. Kaernbach, C. & K. Schlemmer. 2008. The decay of pitch memory during rehearsal. *J. Acoust. Soc. Am.* **123:** 1846–1849.
31. Deutsch, D. 1970. Tones and numbers: specificity of interference in immediate memory. *Science* **168:** 1604–1605.
32. Salame, P. & A.D. Baddeley. 1989. Effects of background music on phonological short-term memory. *Q. J. Exp. Psychol.* **41:** 107–122.
33. Semal, C., L. Demany, K. Ueda & P.A. Halle. 1996. Speech versus nonspeech in pitch memory. *J. Acoust. Soc. Am.* **100:** 1132–1140.
34. Chan, A.S., Y.C. Ho & M.C. Cheung. 1998. Music training improves verbal memory. *Nature* **396:** 128.
35. Awh, E. *et al*. 1996. Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. *Psychol. Sci.* **7:** 25–31.
36. Fiez, J.A. *et al*. 1996. A positron emission tomography study of the short-term maintenance of verbal information. *J. Neurosci.* **16:** 808–822.
37. Gruber, O. & D.Y. von Cramon. 2003. The functional neuroanatomy of human working memory revisited. Evidence from 3-T fMRI studies using classical domain-specific interference tasks. *NeuroImage* **19:** 797–809.
38. Paulesu, E., C.D. Frith & R.S. Frackowiak. 1993. The neural correlates of the verbal component of working memory. *Nature* **362:** 342–345.
39. Ravizza, S.M., M.R. Delgado, J.M. Chein, *et al*. 2004. Functional dissociations within the inferior parietal cortex in verbal working memory. *NeuroImage* **22:** 562–573.
40. Bamiou, D.E., F.E. Musiek & L.M. Luxon. 2003. The insula (Island of Reil) and its role in auditory processing. Literature review. *Brain. Res. Rev.* **42:** 143–154.
41. Chein, J.M. & J.A. Fiez. 2001. Dissociation of verbal working memory system components using a delayed serial recall task. *Cereb. Cortex* **11:** 1003–1014.
42. Chen, S.H. & J.E. Desmond. 2005. Cerebrocerebellar networks during articulatory rehearsal and verbal working memory tasks. *NeuroImage* **24:** 332–338.
43. Kirschen, M.P., S.H. Chen, P. Schraedley-Desmond & J.E. Desmond. 2005. Load- and practice-dependent increases in cerebro-cerebellar activation in verbal working memory: an fMRI study. *NeuroImage* **24:** 462–472.
44. 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.
45. Henson, R.N.A., N. Burgess & C.D. Frith. 2000. Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia* **38:** 426–440.
46. Jonides, J. *et al*. 1998. The role of parietal cortex in verbal working memory. *J. Neurosci.* **18:** 5026–5034.
47. Cabeza, R. & L. Nyberg. 2000. Imaging cognition II: an empirical review of 275 PET and fMRI studies. *J. Cogn. Neurosci.* **12:** 1–47.
48. Corbetta, M. & G.L. Shulman. 2002. Control of goal-directed and stimulus-driven attention in the brain. *Nat. Rev. Neurosci.* **3:** 201–215.
49. Becker, J.T., D.K. MacAndrew & J.A. Fiez. 1999. A comment on the functional localization of the phonological storage subsystem of working memory. *Brain Cogn.* **41:** 27–38.
50. Buchsbaum, B.R., R.K. Olsen, P. Koch & K.F. Berman. 2005. Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* **48:** 687–697.
51. Hickok, G. & D. Poeppel. 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* **8:** 393–402.
52. Hickok, G. 2009. The functional neuroanatomy of language. *Phys. Life Rev.* **6:** 121–143.
53. Rauschecker, J.P. & S.K. Scott. 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat. Neurosci.* **12:** 718–724.
54. Hickok, G., J. Houde & F. Rong. 2011. Sensorimotor integration in speech processing: computational basis and neural organization. *Neuron* **69:** 407–422.
55. 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.
56. Zatorre, R.J., A.C. Evans & E. Meyer. 1994. Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* **14:** 1908–1919.

57. Petrides, M., B. Alivisatos, E. Meyer & A.C. Evans. 1993. Functional activation of the human frontal cortex during the performance of verbal working memory tasks. *Proc. Natl. Acad. Sci. USA* **90:** 878–882.
58. Wager, T.D. & E.E. Smith. 2003. Neuroimaging studies of working memory: a meta-analysis. *Cogn. Affect. Behav. Neurosci.* **3:** 255–274.
59. Wilson, M. 2001. The case for sensorimotor coding in working memory. *Psychon. Bull. Rev.* **8:** 44–57.
60. Drost, U.C., M. Rieger, M. Brass, *et al*. 2005. When hearing turns into playing: movement induction by auditory stimuli in pianists. *Q. J. Exp. Psychol. A* **58:** 1376–1389.
61. Drost, U.C., M. Rieger, M. Brass, *et al*. 2005. Action-effect coupling in pianists. *Psychol. Res.* **69:** 233–241.
62. Bangert, M. *et al*. 2006. Shared networks for auditory and motor processing in professional pianists: evidence from fMRI conjunction. *NeuroImage* **30:** 917–926.
63. D'Ausilio, A., E. Altenmüller, M. Olivetti Belardinelli & M. Lotze. 2006. Cross-modal plasticity of the motor cortex while listening to a rehearsed musical piece. *Eur. J. Neurosci.* **24:** 955–958.
64. Haslinger, B. *et al*. 2005. Transmodal sensorimotor networks during action observation in professional pianists. *J. Cogn. Neurosci.* **17:** 282–293.
65. Petrides, M., G. Cadoret & S. Mackey. 2005. Orofacial somatomotor responses in the macaque monkey homologue of Broca's area. *Nature* **435:** 1235–1238.
66. Watkins, K. & T. Paus. 2004. Modulation of motor excitability during speech perception: the role of Broca's area. *J. Cogn. Neurosci.* **16:** 978–987.
67. Lahav, A., E. Saltzman & G. Schlaug. 2007. Action representation of sound: audiomotor recognition network while listening to newly acquired actions. *J. Neurosci.* **27:** 308–314.
68. Mutschler, I. *et al*. 2007. A rapid sound-action association effect in human insular cortex. *PLoS One* **2:** e259.
69. Indefrey, P. & W.J. Levelt. 2004. The spatial and temporal signatures of word production components. *Cognition* **92:** 101–144.
70. Hoover, J.E. & P.L. Strick. 1999. The organization of cerebellar and basal ganglia outputs to primary motor cortex as revealed by retrograde transneuronal transport of herpes simplex virus type 1. *J. Neurosci.* **19:** 1446–1463.
71. Leblois, A., T. Boraud, W. Meissner, *et al*. 2006. Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia. *J. Neurosci.* **26:** 3567–3583.
72. Middleton, F.A. & P.L. Strick. 2000. Basal ganglia output and cognition: evidence from anatomical, behavioral, and clinical studies. *Brain Cogn.* **42:** 183–200.
73. Parent, A. & L.N. Hazrati. 1995. Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. *Brain. Res. Rev.* **20:** 91–127.
74. Jacquemot, C. & S.K. Scott. 2006. What is the relationship between phonological short-term memory and speech processing? *Trends Cogn. Sci.* **10:** 480–486.
75. Dalla Bella, S.D., M. Berkowska & J. Sowinski. 2011. Disorders of pitch production in tone deafness. *Front. Psychol.* **2:** 164.
76. Miller, G.A. 1956. The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.* **63:** 81–97.
77. Ericsson, K.A., W.G. Chase & S. Faloon. 1980. Acquisition of a memory skill. *Science* **208:** 1181–1182.
78. Gobet, F. 2005. Chunking models of expertise: implications for education. *Appl. Cogn. Psychol.* **19:** 183–204.
79. Gobet, F. *et al*. 2001. Chunking mechanisms in human learning. *Trends. Cogn. Sci.* **5:** 236–243.
80. Savage, C.R. *et al*. 2001. Prefrontal regions supporting spontaneous and directed application of verbal learning strategies: evidence from PET. *Brain* **124:** 219–231.
81. Bor, D., N. Cumming, C.E. Scott & A.M. Owen. 2004. Prefrontal cortical involvement in verbal encoding strategies. *Eur. J. Neurosci.* **19:** 3365–3370.
82. Bor, D., J. Duncan, R.J. Wiseman & A.M. Owen. 2003. Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron* **37:** 361–367.
83. Bor, D. & A.M. Owen. 2007. A common prefrontal-parietal network for mnemonic and mathematical recoding strategies within working memory. *Cereb. Cortex* **17:** 778–786.
84. Schulze, K., K. Mueller & S. Koelsch. 2011. Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *Eur. J. Neurosci.* **33:** 189–196.
85. Krumhansl, C.L. 1979. The psychological representation of musical pitch in a tonal context. *Cogn. Psychol.* **11**: 346–374.
86. Krumhansl, C.L. & R.N. Shepard. 1979. Quantification of the hierarchy of tonal functions within a diatonic context. (Translated from Eng) *J. Exp. Psychol. Hum. Percept. Perform.* **5:** 579–594.
87. Koelsch, S., E. Schroger & M. Tervaniemi. 1999. Superior pre-attentive auditory processing in musicians. *Neuroreport* **10:** 1309–1313.
88. Koelsch, S., E. Schroger & T.C. Gunter. 2002. Music matters: preattentive musicality of the human brain. *Psychophysiology* **39:** 38–48.
89. Schulze, K., W.J. Dowling & B. Tillmann. 2012. Working memory for tonal and atonal sequences during a forward and a backward recognition task. *Music Percept* **29:** 225–268.