Auditory and visual memory in musicians and nonmusicians

**Michael A. Cohen · Karla K. Evans ·
Todd S. Horowitz · Jeremy M. Wolfe**

Abstract Numerous studies have shown that musicians outperform nonmusicians on a variety of tasks. Here we provide the first evidence that musicians have superior auditory recognition memory for both musical and nonmusical stimuli, compared to nonmusicians. However, this advantage did not generalize to the visual domain. Previously, we showed that auditory recognition memory is inferior to visual recognition memory. Would this be true even for trained musicians? We compared auditory and visual memory in musicians and nonmusicians using familiar music, spoken English, and visual objects. For both groups, memory for the auditory stimuli was inferior to memory for the visual objects. Thus, although considerable musical training is associated with better musical and nonmusical auditory memory, it does not increase the ability to remember sounds to the levels found with visual stimuli. This suggests a fundamental capacity difference between auditory and visual recognition memory, with a persistent advantage for the visual domain.

Keywords Music · Musicians · Training · Visual memory · Auditory memory · Long term memory

Studies for the last 40 years have shown that human visual recognition memory has an astonishingly high capacity (Pezdek, Whetstone, Reynolds, Askari, & Dougherty, 1989; Shepard,

M. A. Cohen (✉)
Department of Psychology, Harvard University,
Cambridge, MA, USA
e-mail: michaelthecohen@gmail.com

K. K. Evans · T. S. Horowitz · J. M. Wolfe
Visual Attention Laboratory,
Harvard Medical School/Brigham and Women's Hospital,
Boston, MA, USA

1967; Standing, Conezio, & Haber, 1970). After viewing up to 10,000 images over the course of several hours, observers can identify which images they have previously seen with 83% accuracy (Standing, 1973). More recently, it has been shown that observers not only remember the gist of what they were shown ("I saw a beach" or "I saw an apple"), but also an incredible amount of object detail (Brady, Konkle, Alvarez, & Oliva, 2008). After viewing 2,500 images, observers were 88% correct at discriminating the target item from another object from the same category ("I saw *this* apple, not *that* one.").

Recently (Cohen, Horowitz, & Wolfe, 2009), we asked whether this same ability was present for auditory memory, and found that it was not. Auditory memory was not very massive when tested in a manner similar to the visual massive-memory experiments. Our findings ruled out the possibility that the inferiority of auditory memory was due to visual stimuli containing more information than auditory stimuli. We found that, in order to equate visual and auditory memory performance, we needed high-quality environmental sounds and severely degraded visual images. To be precise, memory was equivalent for auditory stimuli that could be correctly identified 64% of the time (e.g., a dog barking) and visual stimuli so badly blurred that they could be identified only 21% of the time.

Another possibility is that the visual advantage originates in observers' relative experience with the two modalities. If we assume that vision is the dominant perceptual modality for most humans, the auditory inferiority observed in our original data might be due to the comparative neglect of audition. To test this possibility, we sought a subject population that might be expected to pay more attention to the auditory modality than does the population at large, without being visually impaired. We chose to test trained musicians.

Of course, there is considerable debate about whether or not the effects of musical training are exclusively confined to the domain of musical abilities. Some have argued that

such training is domain specific (Peretz & Coltheart, 2003) and that there is a module for music cognition in the brain (Zatorre & Peretz, 2001). This notion is primarily supported by patients with either congenital or acquired amusia, who are unable to recognize melodies that were at one point highly familiar (in the case of acquired amusia) and are incapable of detecting wrong or out-of-tune notes (Ayotte, Peretz, & Hyde, 2002; Griffiths et al., 1997; Peretz et al., 2002, 1994). In spite of these musical deficits, these patients are normal at recognizing words and lyrics, familiar voices, and environmental sounds (i.e., dogs barking, street noises, etc.).

On the other hand, there are numerous studies showing that musical training improves a variety of cognitive functions beyond the musical. For example, it is believed that musical training improves verbal memory (Chan, Ho, & Cheung, 1998), speech perception (Moreno, Marques, Santos, Castro, & Besson, 2008; Parbery-Clarke, Skoe, Lam, & Kraus, 2009), IQ scores (Schellenberg, 2004), analytic listening abilities (Oxenham, Fligor, Mason, & Kidd, 2003), and spatial abilities (Schellenberg, 2005). Furthermore, musicians also undergo numerous structural transformations in their brains, including functional differences in sensorimotor areas (Gaser & Schlaug, 2003; Schlaug, 2001), auditory areas (Lappe, Herholtz, Trainor, & Pantev, 2008; Pantev et al., 1998; Schneider et al., 2002), the brain stem (Wong, Skoe, Russo, Dees, & Kraus, 2007), and other multimodal integration areas (Zatorre, Chen, & Penhune, 2007). Musical training has also been shown to cause structural changes after only 15 months of instruction in early childhood (Hyde et al., 2009). However, the improvement in children's general and spatial cognitive development after 1 and 2 years of instructions disappeared after 3 years (Costa-Giomi, 1999). Finally, numerous longitudinal studies have shown the benefits of musical training (Hyde et al., 2009; Moreno et al., 2008; Schellenberg, 2004), suggesting that *training* per se, rather than a natural predisposition, is responsible for these differences.

Might musical training augment an individual's ability to remember not only music, but nonmusical sounds as well? Perhaps individuals who dedicate more time and resources to the auditory modality have an auditory memory capacity that is comparable to the visual memory capacity of normal individuals.

Experiments 1A and 1B

Methods

Participants We tested between 8 and 10 (depending on the condition) trained musicians (average age: 28.31 years, *SD* 8.26) on auditory and visual recognition memory tasks using a variety of stimuli, and compared their performance to that of nonmusicians (average age: 27.12 years, *SD* 6.62). Musicians were students or instructors recruited from Julliard, the New England Conservatory, Berklee College of Music, or the Harvard School of Music. Each musician had at least 15 years of music training and reported spending between 35 and 60 h per week engaged with music (e.g., playing, writing). Nonmusicians had either no musical training or limited training confined to their preteenage years. Musicians and nonmusicians were not significantly different in their ages and socioeconomic status, which was defined by parental education (measured on a 6-point scale; Norton et al., 2005), which is held to be a reliable indicator of socioeconomic status (Hollingshead & Redlich, 1958). The musicians themselves had an average educational score of 4.2 (*SD* 0.57), while the nonmusicians had an average score of 2.9 (*SD* 0.68).

Stimuli The sound stimuli comprised music, speech, and environmental sounds. Two classes of musical stimuli were used: familiar and unfamiliar music. The familiar music stimulus set comprised 258 well-known pop songs, nursery rhymes, theme songs, and so on. Clips that could not be correctly identified during a free recall classification task were excluded from analysis: These comprised 3.1% (musicians) and 4.0% (nonmusicians) of trials. Unfamiliar music came from a variety of musical styles, with the exception of jazz and classical music, since each musician was trained in one of those styles and was thus substantially less likely to be unfamiliar with the musical clip. To ensure that none of the participants were familiar with the music clips, they were instructed to identify either the title or the artist of any clip they could recognize during the experiment. Correctly identified clips were removed from analysis and resulted in the exclusion of 2.9% (musicians) and 1.3% (nonmusicians) of trials. In addition, two classes of nonmusical stimuli were used: speech clips and environmental sounds. For each of these classes, there was only one clip belonging to a particular semantic category (i.e., only one speech clip about politics, and only one environmental sound of dogs barking). In total, the set size of each stimulus set was 258 familiar music clips, 99 unfamiliar music clips, 90 environmental sound clips, and 108 speech clips.

The visual stimuli comprised 258 images of isolated objects on a white background (from the database of Brady, Konkle, Alvarez, & Oliva, 2008) and 99 abstract art pieces that were unfamiliar to all observers (from the database of Wilmer et al., 2010).

Procedure Each experiment consisted of two tasks: a recognition memory and a semantic classification task. Each recognition memory task comprised a study phase and a testing phase. The study phase consisted of 60–172 stimuli presented

sequentially for between 5 and 12 s, depending on the condition. During the test phase, participants were presented with another set of 60–172 stimuli, of which half were images they had seen in the study phase, and they had to classify each stimulus as "old" or "new." We converted recognition memory to d', the signal detection sensitivity parameter.

There were two types of classification tasks: free recall and multialternative forced choice. For the free recall task, which was used for the music memory tasks, participants would hear each clip one by one and provide the name of either the song or the performing artist. For the multi-alternative forced choice task, participants would hear or see each stimulus sequentially and choose the name that described that stimulus from a list of options. For each stimulus set, more options were provided than there were stimuli, and no label could be applied to multiple stimuli.

Results

Experiment 1A examined whether musicians have superior auditory memory abilities relative to nonmusicians for both musical and nonmusical stimuli (see Fig. 1a). Musicians were significantly better than nonmusicians at remembering familiar music [$t(7) = 2.54$, $p < .05$], unfamiliar music [$t(9) = 5.39$, $p < .001$], speech [$t(7) = 2.39$, $p < .05$], and environmental [$t(9) = 6.61$, $p < .001$] sound clips.

Perhaps musicians were better at remembering sounds because their initial percept was richer, and therefore they had more information available to encode in the first place. As in our previous study (Cohen et al., 2009), we tested this hypothesis by looking at the classification data. We reasoned that better perceptual information would lead to better identification. However, using the multialternative forced choice classification task, musicians and non-musicians performed equally well at classifying both the speech [musicians, 94.2%; nonmusicians, 94.7%; $t(7) = 0.67$, $p = .52$] and environmental [musicians, 89.7%; nonmusicians, 92.7%; $t(7) = 1.17$, $p = .28$] clips .

The most striking aspect of these data is that musicians' superior memory was clearly not confined to the musical domain and extended to nonmusical stimuli such as spoken words and environmental sounds.

Experiment 1B tested whether this increased memory capacity was limited to the auditory modality or whether it extended to visual memory as well. Previous studies have not definitively shown whether musical training extends into the domain of vision (Chan et al., 1998; Ho, Cheung, & Chan, 2003; Jakobson, Lewycky, Kilgour, & Stoesz, 2008). We tested our participants' recognition memory for visual objects, as well as abstract art. There was no difference between musicians' and nonmusicians' visual memory for either the objects [$t(9) = 1.09$, $p = .31$] or the abstract art [$t(9) = 0.49$, $p = .647$].

Discussion

Experiments 1A and 1B demonstrate that while musicians have increased abilities that extend across the auditory modality, this is not because they have better memory in general. Musicians' and nonmusicians' visual memory performance did not differ.

Experiments 2A and 2B

After having demonstrated musicians' superior auditory memory abilities, we asked whether these abilities could overcome the inferiority of auditory memory observed in our previous study that compared auditory and visual recognition memory (Cohen et al., 2009). In Experiments 2A and 2B, we sought to directly compare auditory and visual memory in musicians and nonmusicians using the same number of visual objects, familiar music clips, and speech clips. Experiment 2A had one group of musicians and nonmusicians complete recognition memory tasks for visual objects and familiar music clips, and Experiment 2B

Fig. 1 (a) Performance on auditory recognition memory tests in both musicians and nonmusicians. The various conditions are labeled on the x-axis, and performance is measured in terms of d' (the signal detection metric) on the y-axis. (b) Performance for the two groups on the visual recognition memory task. Again, conditions are on the x-axis and performance on the y-axis
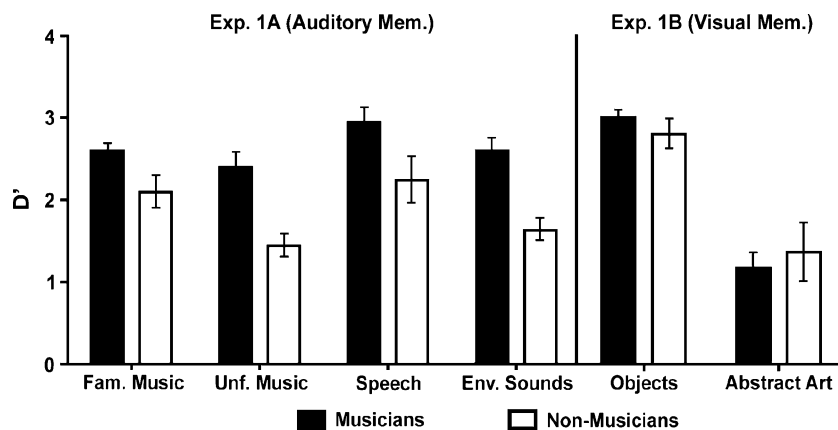
had another group of participants complete recognition memory tasks for visual objects and speech clips.

Method

Participants We tested 8 trained musicians from the New England Conservatory, Berklee College of Music, the Harvard School of Music, and the Oberlin College Conservatory. The criterion used to define musicians and nonmusicians (i.e., amount of musical training) was the same that had been used in Experiments 1A and 1B. Musicians (average age: 26.88 years, *SD* 3.72) and non-musicians (average age: 30.5 years, *SD* 5.73) were once again matched as closely as possible in age and parental education socioeconomic status. The musicians themselves had an average educational score of 4.1 (*SD* 0.64), and the nonmusicians had an average score of 4.4 (*SD* 0.52).

Stimuli For Experiment 2A, 258 visual images and 258 music clips were used, which were the same stimuli used in Experiments 1A and 1B. For Experiment 2B, 111 visual images were selected from the Brady et al. (2008) database, and speech clips were obtained online and were recorded and edited in the laboratory. Each image/clip was presented sequentially for exactly 5 s.

Procedure Once again, for the memory task, participants simply had to classify stimuli as being "old" or "new" during the testing phase. To determine whether the stimuli were understood, participants performed the free recall classification task for the music and visual stimuli. They performed the multialternative forced choice classification task for the speech clips.

Results

The results for Experiments 2A and 2B are presented in Fig. 2. For Experiment 2A, we found that both groups of participants remembered visual objects better than familiar music [musi-

cians, $t(7) = 3.75$, $p < .01$; nonmusicians, $t(7) = 3.1$, $p < .05$]. Experiment 2B showed that visual images were also remembered better than speech clips [musicians, $t(7) = 2.74$, $p < .05$; nonmusicians, $t(7) = 3.44$, $p < .01$]. Furthermore, Experiment 2A replicated our results from Experiment 1A, showing that musicians were better than nonmusicians at remembering familiar music [$t(7) = 2.53$, $p < .05$] but were no better than nonmusicians at remembering visual images [$t(7) = 0.69$, $p = .51$]. The same trend was found in Experiment 2B for the visual images, in that there was no difference between the two groups [$t(7) = 0.52$, $p = .61$], whereas for the speech clips, the effect between the groups was trending but not significant [$t(7) = 1.94$, $p = .09$].

As before, one might suspect that the differences between auditory and visual stimuli might lie in the initial perceptual representation. If so, the superiority for visual stimuli should also show up in a nonmemory task. We addressed this question by asking each participant to complete a categorization task following the recognition memory experiment. We found that visual images and music clips were categorized equally well, and this held true for both musicians [images, 92.6%; music, 94.7%; $t(7) = 1.48$, $p = .19$] and nonmusicians [images, 95.4%; music, 94.4%; $t(7) = 1.02$, $p = .34$]. Similarly, there was no difference in categorizing the visual images and speech clips for the musicians [images, 94.8%; speech, 95.2%; $t(7) = 0.89$, $p = .4$] or the nonmusicians [images, 93.0%; speech, 94.2%; $t(7) = 0.42$, $p = .67$]. Thus, we found no difference in the ability to provide semantic labels for visual and auditory stimuli that could be used to explain the differences between auditory and visual memory performance.

Discussion

Taken together, Experiments 2A and 2B show that even musicians, who have superior memory abilities across the auditory modality, cannot remember sounds as well as visual images. It should be noted that we specifically used familiar music and speech clips because these stimuli yielded the best memory performance in Experiment 1.

Fig. 2 Performance on the visual, music, and speech memory conditions in both musicians and nonmusicians (*x*-axis). The left panel shows compared performance for the picture versus music conditions. The right panel shows compared performance for the picture versus speech conditions. Performance was measured in terms of *d′* (*y*-axis)
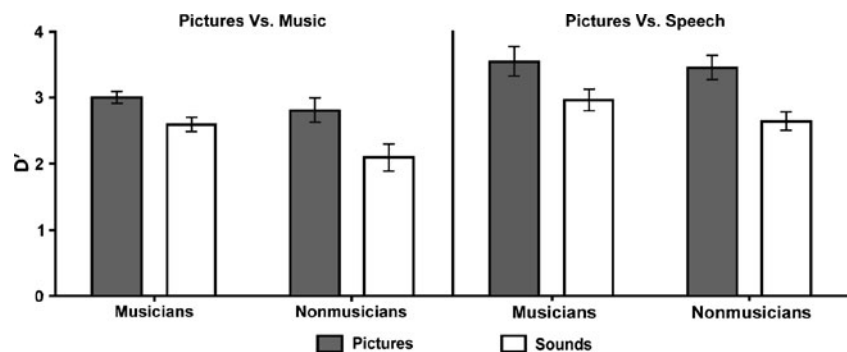
Thus, even when we chose our two best classes of auditory stimuli, the musicians *still* could not recall those sounds as well as a random selection of visual objects.

General discussion

We tested the hypothesis that the superiority of visual recognition memory over auditory recognition memory is a by-product of the dominance of the visual modality in humans. This hypothesis would predict that greater experience with a different sensory modality would improve recognition memory in that modality. We selected trained musicians as our study population on the basis of the wide body of research showing that their auditory abilities are fundamentally different from nonmusicians.

The results from Experiment 1 clearly demonstrated that musicians have superior auditory memory relative to non-musicians for a range of both musical and nonmusical sounds, supplementing the large body of work showing that musical training increases a variety of auditory abilities (e.g., Hannon & Trainor, 2007; Kraus & Chandrasekaran, 2010; Patel & Iversen, 2007). Our results agree with studies that have found that musicians have improved memory across the auditory domain but not in the visual domain (Chan et al., 1998; Ho et al., 2003). They are not in agreement with studies that have suggested that the effects of musical training are confined to the realm of musical ability (Peretz & Coltheart, 2003), nor are they in agreement with studies that hold that musical training improves visual memory (Jakobson et al., 2008).

Experiment 2 compared auditory and visual memory in musicians and nonmusicians. Although musicians once again performed better than nonmusicians at auditory memory, reducing the difference between their memory for auditory and visual stimuli, they still could not remember these sounds as well as visual images. This is not to say that the auditory memories of musicians or nonmusicians are impoverished in general. After all, both groups of participants were able to classify a variety of songs and sounds, which requires that those stimuli be stored in memory. Furthermore, many musicians have an enormous amount of music committed to memory that they are able to perform at will. It was specifically the *episodic* memory for a recent set of stimuli that was better for all observers in the visual than in the auditory domain and better for musicians than for nonmusicians in the auditory domain.

It is worth stressing that, in attempting to close the gap between auditory and visual memory, we stacked the deck somewhat in favor of auditory stimuli. Not only did we find a population of participants with overall superior auditory memory abilities (musicians), we also specifically chose the auditory stimuli they were able to remember the most easily (familiar music and speech clips). The visual objects, however, were not specifically chosen for any particular reason and were randomly selected from a large object database. Thus, even when numerous steps were taken in an attempt to increase auditory memory performance, we *still* could not find a group of participants and a stimulus set that allowed performance to equal that for visual memory. This is consistent with the idea that some fundamental difference exists between visual and auditory stimuli, or visual and auditory processing, when it comes to recognition memory capacities, with the advantage persistently going to vision.

References

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain, 125*, 238–251.

Brady, T. F., Konkle, T., Alvarez, G. A., & Oliva, A. (2008). Visual long-term memory has a massive storage capacity for object details. *Proceedings of the National Academy of Sciences, 105*, 14325–14329.

Chan, A. S., Ho, Y.-C., & Cheung, M.-C. (1998). Music training improves verbal memory. *Nature, 396*, 128. doi:10.1038/24075

Cohen, M. A., Horowitz, T. S., & Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences, 106*, 6008–6010.

Costa-Giomi, E. (1999). The effects of three years of piano instruction on children's cognitive development. *Journal of Research in Music Education, 47*, 198–212.

Gaser, C., & Schlaug, G. (2003). Brain structures differ between musicians and non-musicians. *The Journal of Neuroscience, 5*, 688–694.

Griffiths, T. D., Rees, A., Witton, C., Cross, P. M., Shakir, R. A., & Green, G. G. (1997). Spatial and temporal auditory processing deficits following right hemisphere infarction: A psychophysical study. *Brain, 120*, 785–794.

Hannon, E. E., & Trainor, L. J. (2007). Music acquisition: Effects of enculturation and formal training on development. *Trends in Cognitive Sciences, 11*, 466–472.

Hollingshead, A., & Redlich, F. (1958) Social Class and mental illness. New York: Wiley

Ho, Y.-C., Cheung, M.-C., & Chan, A. S. (2003). Music training improves verbal but not visual memory: Cross-sectional and longitudinal exploration in children. *Neuropsychology, 17*, 439–445.

Hyde, K. L., Lerch, J., Norton, A., Forgeard, M., Winner, E., Evans, A. C., et al. (2009). Musical training shapes structural brain development. *The Journal of Neuroscience, 29*, 3019–3025.

Jakobson, L. S., Lewycky, S. T., Kilgour, A. R., & Stoesz, B. M. (2008). Memory for verbal and visual material in highly trained musicians. *Music Perception, 26*, 41–55.

Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews. Neuroscience, 11*, 599–605.

Lappe, C., Herholtz, S. C., Trainor, L. J., & Pantev, C. (2008). Cortical plasticity induced by short-term unimodal and multimodal musical training. *The Journal of Neuroscience, 28*, 9632–9639.

Moreno, S., Marques, C., Santos, A., Castro, S. L., & Besson, M. (2008). Musical training influences linguistic abilities in 8-year-old children: More evidence for brain plasticity. *Cerebral Cortex, 19*, 712–723.

Norton, A., Winner, E., Cronin, K., Overy, K., Lee, D. J., & Schlaug, G. (2005). Are there pre-existing neural, cognitive, or motoric markers for musical ability? *Brain and Cognition, 59*, 124–134.

Oxenham, A. J., Fligor, B. J., Mason, C. R., & Kidd, G., Jr. (2003). Informational masking and musical training. *The Journal of the Acoustical Society of America, 114*, 1543–1549.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., & Hoke, M. (1998). Increased auditory cortical representation in musicians. *Nature, 392*, 811–814.

Parbery-Clark, A., Skoe, E., Lam, C., & Kraus, N. (2009). Musicians enhancement for speech-in-noise. *Ear and Hearing, 30*, 653–661.

Patel, A. D., & Iversen, J. R. (2007). The linguistic benefits of musical abilities. *Trends in Cognitive Sciences, 11*, 369–372.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., et al. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron, 33*, 185–191.

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience, 6*, 688–691.

Peretz, I., Kolinsky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., et al. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117*, 1283–1301.

Pezdek, K., Whetstone, T., Reynolds, K., Askari, N., & Dougherty, T. (1989). Memory for real-world scenes: The role of consistency with schema expectation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 15*, 587–595.

Schellenberg, E. G. (2004). Music lessons enhance IQ. *Psychological Science, 15*, 511–514.

Schellenberg, E. G. (2005). Music and cognitive abilities. *Current Directions in Psychological Science, 14*, 317–320.

Schlaug, G. (2001). The brain of musicians: A model for functional and structural adaptation. *Annals of the New York Academy of Sciences, 930*, 281–299.

Schneider, P., Scherg, M., Dosch, H. G., Specht, H. J., Gutschalk, A., & Rupp, A. (2002). Morphology of Heschl's gyrus reflects enhanced activation in the auditory cortex of musicians. *Nature Neuroscience, 5*, 688–694.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior, 6*, 156–163.

Standing, L. (1973). Learning 10, 000 pictures. *The Quarterly Journal of Experimental Psychology, 25*, 207–222.

Standing, L., Conezio, J., & Haber, R. N. (1970). Perception and memory for pictures: Single-trial learning of 2, 500 visual stimuli. *Psychonomic Science, 19*, 73–74.

Wilmer, J. B., Germine, L., Chabris, C. F., Chatterjee, G., Williams, M., Loken, E., et al. (2010). Human face recognition ability is highly heritable. *Proceedings of the National Academy of Sciences, 107*, 5238–5241.

Wong, P. C., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*, 420–422.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory–motor interactions in music perception and production. *Nature Reviews. Neuroscience, 8*, 547–558.

Zatorre, R. J., & Peretz, I. (2001). *The biological foundations of music (Annals of the New York Academy Sciences* (Vol. 930)). New York: New York Academy of Sciences.