

Auditory Imagery: Empirical Findings

Timothy L. Hubbard
Texas Christian University

The empirical literature on auditory imagery is reviewed. Data on (a) imagery for auditory features (pitch, timbre, loudness), (b) imagery for complex nonverbal auditory stimuli (musical contour, melody, harmony, tempo, notational audiation, environmental sounds), (c) imagery for verbal stimuli (speech, text, in dreams, interior monologue), (d) auditory imagery's relationship to perception and memory (detection, encoding, recall, mnemonic properties, phonological loop), and (e) individual differences in auditory imagery (in vividness, musical ability and experience, synesthesia, musical hallucinosis, schizophrenia, amusia) are considered. It is concluded that auditory imagery (a) preserves many structural and temporal properties of auditory stimuli, (b) can facilitate auditory discrimination but interfere with auditory detection, (c) involves many of the same brain areas as auditory perception, (d) is often but not necessarily influenced by subvocalization, (e) involves semantically interpreted information and expectancies, (f) involves depictive components and descriptive components, (g) can function as a mnemonic but is distinct from rehearsal, and (h) is related to musical ability and experience (although the mechanisms of that relationship are not clear).

Keywords: imagery, audition, language, music, cognition

Despite the resurgence in imagery research beginning in the late 1960s and early 1970s, auditory forms of imagery have received relatively little interest. The bulk of the research on imagery has focused on visual imagery and on spatial imagery (e.g., Finke, 1980, 1985, 1989; Finke & Shepard, 1986; Kosslyn, 1980, 1994; Reisberg & Heuer, 2005; Shepard & Cooper, 1982; but see Reisberg, 1992; Stevenson & Case, 2005), but nonvisual forms of imagery are common and useful in everyday life (e.g., see Eardley & Pring, 2006). This review will focus on empirical findings regarding (a) imagery for auditory features (e.g., pitch, loudness) and for more complex auditory stimuli (e.g., music, language), (b) the relationship between auditory imagery and other domains (e.g., perception, memory), and (c) variables potentially related to individual differences in auditory imagery ability and occurrence (e.g., musical experience, psychopathology). The definition of *auditory imagery* to be used is that provided by Intons-Peterson (1992, p. 46), who referred to auditory imagery as “the introspective persistence of an auditory experience, including one constructed from components drawn from long-term memory, in the absence of direct sensory instigation of that experience.” This definition is analogous to those used for other modalities of imagery (e.g., Finke, 1989, and Stevenson & Case, 2005, in their considerations of visual imagery and olfactory imagery, respectively).

The experience of imagery is subjective, and images cannot be directly observed by an experimenter; the existence and properties of imagery must be inferred on the basis of more indirect measures that are hypothesized to be influenced by imagery in predictable and systematic ways (e.g., selective facilitation or interference with a concurrent or subsequent task). Studies that purport to

examine auditory imagery vary in the extent to which they offer convincing evidence that auditory imagery (rather than another form of representation) was generated, evoked, or used, and so the strength of the evidence for auditory imagery in individual studies must be considered in any assessment of this literature. Unfortunately, some studies that purport to examine auditory imagery fail to provide convincing evidence that auditory imagery was actually generated, evoked, or used, and so even though the results of those studies might be consistent with a role of auditory imagery, such results do not necessarily demonstrate that auditory imagery occurred or was responsible for the results obtained in those studies. Also, consideration of auditory imagery includes domains that are not necessarily or solely auditory (e.g., language and music can be represented in visual written form as well as in auditory form; speech imagery or music imagery can include motor information relevant to articulation or to performance) but that can be experienced in the auditory modality, and so it is important to separate the *modality* of imagery from the *content* of imagery.

The types of data to be surveyed include (a) subjective reports of experimental participants, (b) comparisons of performances on tasks in which experimental participants are instructed to generate auditory imagery to performances on tasks in which participants are not instructed to generate auditory imagery or in which an auditory stimulus is presented, (c) brain imaging studies that examine cortical areas that are activated when experimental participants are engaged in tasks hypothesized to involve auditory imagery, and (d) clinical data regarding psychopathologies or other conditions in which auditory imagery or perception is affected. In Part 1, auditory imagery for simple features (pitch, timbre, loudness) is reviewed, and in Part 2, auditory imagery for complex musical and nonverbal stimuli (musical contour, melody, harmony, tempo, notational audiation, environmental sounds) is reviewed. In Part 3, auditory imagery for verbal stimuli (speech, text, in dreams, interior monologue) is reviewed. In Part 4, the relationship of

Correspondence concerning this article should be addressed to Timothy L. Hubbard, Department of Psychology, Texas Christian University, Fort Worth, TX 76129. E-mail: timothy.leehubbard@gmail.com

auditory imagery to processes or structures in perception and memory (detection, encoding, recall, mnemonics, phonological loop) is reviewed. In Part 5, how individual differences (in vividness, musical ability and experience, synesthesia, psychopathology) might influence or be influenced by auditory imagery is reviewed. In Part 6, several questions are asked and conclusions drawn regarding general properties and characteristics of auditory imagery. Finally, in Part 7, a brief summary and conclusion are provided.

Part 1: Auditory Features

Numerous studies of auditory imagery have been concerned with structural properties of imagery involving elementary or basic features such as pitch, timbre, and loudness. The representation of such features in auditory imagery is examined in this section.

Pitch

Farah and Smith (1983) had participants perceive or image tones of 715 Hz or 1000 Hz while simultaneously listening for a target signal of 715 Hz or 1000 Hz. The dependent measure was the loudness that would be required of a target for that target to be detected. Performance was facilitated (i.e., the target could be detected at a lower loudness) when the pitch of the image matched the pitch of the target. Furthermore, when the image was generated prior to the presentation of the target, this facilitation was increased. Okada and Matsuoka (1992) had participants image a tone of 800 Hz. Participants had to detect which of five different target tones was then presented, and contrary to Farah and Smith's findings, auditory imagery interfered more with detection of the target when the image was of the same frequency. These findings initially seemed inconsistent with Farah and Smith's assertion that imagery facilitates perception. Okada and Matsuoka pointed out several differences in the methodology of their experiment and the methodology of Farah and Smith's experiment (e.g., Farah & Smith, 1983, used a staircase design and focused on discrimination; Okada & Matsuoka, 1992, used an ascending method of limits and focused on detection), and they suggested that differences between their findings and Farah and Smith's findings were consistent with Finke's (1986) notion that imagery improves discrimination but does not improve detection.

Halpern (1989) had participants hum the starting pitch of each of several well-known melodies. Melodies that descended in the first few notes (e.g., *Joy to the World*) typically had higher hummed starting pitches, and melodies that ascended in the first few notes (e.g., *Somewhere Over the Rainbow*) typically had lower hummed starting pitches. Halpern (1992) suggested that if humming was an adequate externalization of mental pitch, then such findings pointed to a form of absolute pitch. However, and as Halpern noted, the observed regularities might have been due to physical constraints imposed by the vocal apparatus of each participant. Halpern (1989) also had participants indicate the starting pitch of a well-known melody by choosing a key on a (nonvisible) keyboard. In a production condition, male participants chose lower starting pitches than did female participants, whereas in a recognition condition, there were no differences in the starting pitches chosen by male or by female participants. Also, whether a melody was given the highest or lowest starting pitch was not as strongly related to whether that melody immediately descended or ascended

in pitch. In a follow-up study, the same participants were presented with several potential starting pitches, one of which was their previously indicated preferred pitch. They rated how acceptable each of those pitches was as a starting pitch for that melody. Interestingly, pitches that were a specific musical interval (i.e., major third, perfect fifth) away from the previously preferred starting pitch were rated as more acceptable than were closer pitches. Halpern (1992) suggested that this was consistent with a form of memory for the absolute pitch of the starting tone.

Janata and Paroo (2006) examined pitch acuity and temporal acuity of auditory images. Participants were presented with ascending diatonic tone sequences and judged whether the final note was (a) in tune or out of tune or (b) on time or not on time. In the perception condition (referred to by Janata & Paroo, 2006, as *attentional cueing*), participants heard every note of the scale before the final note was presented. In the imagery condition, participants heard the first three or five notes, and they were instructed to image the remaining notes leading up to the final note before the final note was presented. Judgments of the pitch of the final note were not influenced by whether participants heard all the notes leading up to the final note or imaged some of the notes, but judgments of the timing of the final note were severely impaired when participants imaged some of the notes leading up to the final note. Janata and Paroo suggested that their data were consistent with the notion that expectations were the same as images and fit well with Neisser's (1976) notion of images as expectations (see also Janata, 2001). Janata and Paroo also suggested that images of time were more susceptible to distortion in the absence of external stimuli than were images of pitch. Perhaps not surprisingly, increases in musical training were related to more narrowly tuned images, especially along the pitch dimension.

Intons-Peterson, Russell, and Dressel (1992) collected ratings of the typical pitch height of common environmental sounds (e.g., cat purring, wind chimes tinkling). These ratings defined pitch and pitch distance in subsequent studies with other participants. In one study, participants were presented with pairs of phrases that named two objects. Participants were instructed to form an auditory image of the sound named by each member of the pair, and then they mentally adjusted the pitch of one image to match the pitch of the other image. Response times increased with increases in the distance in pitch between the initial pitches. In a second study, participants identified which of two named objects would produce a higher pitch or a lower pitch. Response times decreased as pitch distance increased. This pattern is consistent with a symbolic distance effect (i.e., time required to discriminate between two stimuli along a dimension decreases if the distance between the stimuli along that dimension increases; Moyer, 1973; Moyer & Bayer, 1976) based on the preservation of pitch distance within auditory imagery. For judgments of high pitch, response time decreased when participants judged high pitches relative to when participants judged low pitches, but an equivalent semantic congruity effect (i.e., time required to judge which of two stimuli is more extreme along a dimension decreases when the judgment is congruent with that end of the dimension; Banks, 1977) for judgments of low pitches did not occur.

Yoo, Lee, and Choi (2001) instructed participants to form an auditory image of a single pitch (described as "C major," but a specific frequency was not reported) and acquired functional magnetic resonance imaging (fMRI) with an event-related methodol-

ogy. A single pitch was used to avoid potential confounds that might result if a more complex speech stimulus or musical stimulus (that might automatically activate higher order phonological or musical knowledge distinct from auditory imagery) was imaged. Participants were familiarized with the pitch before brain imaging began, and during brain imaging, participants were prompted to generate an auditory image by a tactile cue to their right thumbs. Significant increases in activation during the time participants were to be imaging occurred in inferior frontal gyri, precuneus, superior temporal gyri, anterior cingulate gyri, and middle and inferior frontal gyri. Activation was bilaterally symmetrical except for slightly higher activation on the left side for inferior frontal gyri and slightly higher activation on the right side for precuneus and superior temporal gyri. Maximum change in activated areas was reached 5 s after auditory image generation was cued, and the amount of change was smaller in primary auditory cortex than in secondary auditory cortex. There was residual activity in primary auditory cortex 15 s after image generation was cued; such residual activity does not persist as long after perception, and it is not clear why it was present after imagery. However, Yoo et al. did not provide any behavioral evidence that participants formed images or that activation patterns were not due to another form of representation.

Timbre

Crowder (1989) presented participants with a sine wave tone of a specific pitch and visually presented the name of a specific musical instrument (e.g., guitar, piano). Participants were instructed to image the pitch of the sine wave tone in the timbre of the named instrument. After participants indicated they had an adequate image, a second tone was presented, and this tone (a) could match the pitch of the previous tone or not, and (b) could match the timbre of the named instrument or not. Participants judged whether the pitch of the second tone matched the pitch in their image. Response times were faster and participants were more accurate, when the timbre of the second tone matched the timbre of the instrument that participants had been instructed to image. Analogous priming on the basis of differences in timbre within a single instrument was reported in a similar experiment by Pitt and Crowder (1992), who manipulated timbre of a cello tone by varying whether the cello was bowed or plucked; judgments of whether the pitch of the image matched the pitch of a subsequent tone were facilitated when the timbre of the subsequent tone matched the timbre that participants had been instructed to image. Also, Pitt and Crowder suggested that the use of nonvocal timbre afforded an examination of sensory aspects of auditory imagery and ruled out potential motor components to imagery; this was based on the idea that subvocalization or other motor processes would not be activated in nonvocal imagery (but see Baddeley & Logie, 1992; Smith, Wilson, & Reisberg, 1995).

Pitt and Crowder (1992; Crowder & Pitt, 1992) examined whether spectral properties (i.e., differences in which harmonics [overtones] were present) and dynamic properties (i.e., differences in onset and offset times and rates of amplitude change of harmonics) of the timbre of an auditory perceptual stimulus were present in an auditory image of that stimulus. Differences in spectral properties were manipulated by synthesizing tones containing different harmonics (keeping dynamic information con-

stant), and differences in dynamic properties were manipulated by varying whether onsets of synthesized tones were sudden or gradual (keeping spectral information constant). Prior to imagery trials, participants were exposed to and learned labels for each of the synthesized timbres. On each trial, participants heard a sine wave tone and saw a visual presentation of the name of one of the timbres. Participants were instructed to image the pitch of the sine wave tone in the named timbre. Once participants indicated they had an adequate image, a second tone was presented, and participants judged whether the pitch of the second tone matched the pitch in their image. Response times were influenced by whether spectral properties of timbre were manipulated, but response times were not influenced by whether dynamic properties of timbre were manipulated. On the basis of this pattern, Pitt and Crowder suggested that spectral properties of timbre were included in auditory images, but dynamic properties of timbre were not included in auditory images.

Halpern, Zatorre, Bouffard, and Johnson (2004) had participants rate similarities of timbres. In a perception condition, participants were presented with all possible pairs of recordings of eight different musical instruments. In an imagery condition, participants were presented with all possible pairs of names of the same eight musical instruments and were asked to image the sounds of those instruments. Additionally, fMRI was acquired during timbre perception and during timbre imagery. Multidimensional scaling of ratings in timbre perception was similar to previous findings (e.g., Kendall, Carterette, & Hajda, 1999) and involved a two-dimensional space defined by an axis of *brilliance* and an axis of *nasality*. Multidimensional scaling of ratings in timbre imagery was highly similar, and ratings of timbre perception and ratings of timbre imagery were highly correlated ($r = .84$). Secondary auditory cortex showed significantly increased activation relative to a visual imagery control task for timbre perception and for timbre imagery, but activation was less during timbre imagery than during timbre perception. Activation for timbre perception and for timbre imagery was bilateral but was stronger on the right side, especially in the superior temporal region. Primary auditory cortex was activated during timbre perception but was not activated during timbre imagery. Supplementary motor cortex exhibited subthreshold activation during timbre imagery, and Halpern et al. suggested that this might reflect (a) subvocalization of pitch or (b) involvement of supplementary motor cortex in a more general aspect of auditory imagery (i.e., generation) distinct from any potential contribution of subvocalization.

Loudness

Intons-Peterson (1980) presented visual phrases that named common environmental sounds. Participants rated the typical loudness of sounds referred to by each phrase, and then they generated images of sounds referred to by the phrases. The time required to generate images was not related to loudness ratings, and Intons-Peterson suggested that this occurred because auditory images did not necessarily encode loudness. However, evidence for an encoding of loudness was observed in a follow-up study in which participants were visually presented with two phrases. Participants formed an image of the sound named by the phrase on the left, and then they formed an image of the sound named by the phrase on the right. Participants then adjusted the subjective loudness of the

second image to match the subjective loudness of the first image. Response times increased with increases in the difference in the typical loudness levels of the named objects. In another follow-up study, participants generated images of the same pairs, and they judged which member of the pair was louder. Response times decreased with larger differences in initial imaged loudness (cf. a symbolic distance effect). Results of the matching and judgment tasks suggested that auditory images could contain loudness information, but results of the generation task suggested that loudness information was not necessarily retrieved (or incrementally generated) when an auditory image was generated. Interestingly, results of matching and judgment tasks for imaged loudness in Intons-Peterson's study paralleled results of matching and judgment tasks for imaged pitch in Intons-Peterson et al.'s (1992) study.

Pitt and Crowder (1992) also suggested that loudness was not necessarily encoded in auditory images. They presented participants with a sine wave tone, and they visually presented the word *soft* or the word *loud* to indicate to participants whether an image based on that sine wave tone should be imaged subjectively soft or subjectively loud. A probe tone was then presented, and that probe was either soft or loud, and participants judged whether the probe was the same pitch as their image or a different pitch. Whether loudness levels of the image and probe matched or did not match did not influence response times (i.e., equivalent loudness levels did not prime judgments) or accuracy rates, and Pitt and Crowder suggested that this occurred because loudness information was not contained within the auditory image. Pitt and Crowder suggested that if loudness information was not encoded in an auditory image, then loudness information might be represented in another form (e.g., loudness might be encoded in motor imagery). Such a lack of encoding of loudness information in auditory imagery might account for the lack of evidence for dynamic information in the image of a specific auditory timbre in Pitt and Crowder's data noted earlier. However, rejection of a necessary loudness component in auditory imagery by Pitt and Crowder and by Intons-Peterson (1980) seems extreme, as they appear to assume that if loudness information is present, then it must influence judgments of other auditory dimensions (i.e., pitch) and must be generated incrementally.¹

Part 2: Music and Environmental Sounds

Although simple auditory features are of interest, many auditory stimuli (and by extension, images of those auditory stimuli) involve combinations of multiple features to form more complex stimuli. Examples of these more complex stimuli involve musical contour, melody, harmony, tempo and duration, notational audiation, and sounds produced by nonhuman animals, objects, and events in the environment. The representation of musical stimuli and environmental sounds in auditory imagery is examined in this section.

Musical Contour

Weber and Brown (1986) had participants learn eight note melodies. On half of the trials, the melodies had an accompanying lyric, and on the other half of the trials, the melodies were sung to the syllable "ba." The participants then drew a sequence of short horizontal lines (moving from left to right) to indicate the pitch

height of each note relative to the preceding note while they sang or imaged each melody. Response times were faster for melodies accompanied by lyrics than for melodies that were not accompanied by lyrics. More relevant to the current concern is that drawing times and error rates were not influenced by whether participants sang or imaged the melody. Indeed, given the importance of musical contour in melodic recognition (e.g., Dowling, 1978; Dowling & Fujitani, 1971), musical contour information would be expected to be specified within an auditory image of a melody. In a follow-up study, participants indicated the pitch height of each note relative to the preceding note by saying "higher" or "lower." Response times were longer for spoken responses than for drawn lines, but there were no other differences between conditions. Given the longer response times for verbal responses, Weber and Brown concluded that musical imagery was a form of auditory imagery that was open to interference from verbal responses. However, it is possible that judgments of contour could have been made without appeal to auditory imagery, and the possibility that a nonimagery form of representation was used cannot be ruled out.

Melody

Zatorre and Halpern (1993) visually presented participants with the title of a melody (e.g., *Jingle Bells*) followed by the first line of the lyric, and in the first line of the lyric, two words were printed in capitals (e.g., "Dashing through the SNOW, in a one horse open SLEIGH"). Participants were patients with a left temporal lobe lesion, patients with a right temporal lobe lesion, or nonpatient controls. Each participant completed an imagery condition and a perception condition, and the task of participants in each condition was to decide whether the pitch of the second capitalized word was higher or lower than the pitch of the first capitalized word (see also Aleman, Nieuwenstein, Boecker, & de Hann, 2000). For all participants, response times increased with increases in the number of intervening beats (see also Halpern, 1988a; Halpern & Zatorre, 1999; Zatorre, Halpern, Perry, Meyer, & Evans, 1996). Also, all participants were generally less accurate in the imagery condition than in the perception condition. The performance of patients with a left temporal lobe lesion did not differ from the performance of control participants, but the performance of patients with a right temporal lobe lesion was significantly worse than that of control participants or of patients with a left temporal lobe lesion. The decreased performance of patients with a right temporal lobe lesion occurred in imagery and in perception.

¹ There is precedent for incremental changes of qualities in imagery (e.g., changes in scanning across an image [Kosslyn, 1980] or in mentally rotating an object [Shepard & Cooper, 1982]). Kosslyn (1980) distinguished between *shift* and *blink* transformations in a visual image: A shift transformation refers to a continuous process of small incremental changes (in which the image appears to go through or experience intermediate values), and a blink transformation refers to a discontinuous sudden change (i.e., in which the image appears to change "in the blink of an eye" without going through or experiencing intermediate values). It is possible that initial generation of loudness information might generally involve a blink transformation (in which the final loudness level is generated directly rather than beginning with a low loudness level and then incrementing the loudness level) rather than a shift transformation, whereas subsequent adjustment of loudness information might generally involve a shift transformation.

Zatorre et al. (1996) used a similar pitch comparison task in a positron emission tomography (PET) imaging study. Rather than presenting the full first line of the lyric, only the two words that indicated the pitches to be compared were presented. In the imagery and perception conditions of interest, these words were drawn from the same melody. In the subtraction control condition, these words were drawn from different melodies, and participants judged which word was longer. As would be expected given previous behavioral and lesion data, participants made fewer errors in the perception condition than in the imagery condition; furthermore, in the perception condition and in the imagery condition, participants took longer to make pitch comparisons when there were more intervening beats. The superior temporal gyrus, as well as areas of the frontal lobe and parietal lobe, were activated in imagery and in perception. The supplementary motor area was activated in imagery and in perception, and this activation was stronger in imagery than in perception. Areas adjacent to the superior temporal gyrus were activated in imagery and in perception, and the thalamus and inferior frontopolar areas were activated only during imagery. Halpern (2003) suggested that these lesion (Zatorre & Halpern, 1993) and imaging (Zatorre et al., 1996) data demonstrated that the right superior temporal gyrus is involved in comparisons of pitch in imagery (cf. Yoo et al., 2001) and in perception.

Activation of the supplementary motor area and cortical areas in the left hemisphere in Zatorre et al. (1996) was initially surprising (see also Halpern et al., 2004), but this might have resulted from effects of verbal lyrics associated with the melody rather than from effects of melody per se; therefore, in a follow-up study, Halpern and Zatorre (1999) presented melodies that did not have associated lyrics. The first few notes of a well-known musical theme (e.g., movie or television themes, classical excerpts) were presented, and participants were instructed to image the remainder of the theme to the end of the first phrase. Upon reaching the end of the first phrase, participants pressed a button. The length of time that elapsed before participants pressed the button increased with increases in the number of intervening beats between the final presented note and the end of the first phrase. There was significant increased activation in the right superior temporal gyrus but no activation in the left superior temporal gyrus. There was also increased activation in the right frontal lobe and supplementary motor area; activation in the supplementary motor area even when no words (lyrics) were presented is consistent with some form of subvocal rehearsal (e.g., humming the appropriate pitches) or a role other than subvocalization (cf. Halpern et al., 2004). Halpern (2003) suggested that because the task in Halpern and Zatorre's study did not involve comparison of pitches, the areas of activation in that study likely reflected musical semantic memory rather than working memory.

Kraemer, Macrae, Green, and Kelley (2005) acquired fMRI from participants who listened to excerpts of music that contained lyrics or to excerpts of instrumental music. The pieces of music were familiar or unfamiliar (names of excerpted pieces were pre-rated by participants so that an appropriate playlist could be created for each participant). Short sections of each musical stimulus (between 2 and 5 s in duration) were deleted and replaced with silent gaps. Listening to unfamiliar musical stimuli resulted in greater activation in auditory association areas than did listening to familiar musical stimuli, but activation in primary auditory cortex

was not influenced by familiarity. Of greater interest, gaps in familiar musical pieces resulted in greater activation in auditory association areas than did gaps in unfamiliar musical pieces. Interestingly, participants reported hearing a continuation of the music in imagery during gaps in familiar pieces but not hearing a continuation of the music in imagery during gaps in unfamiliar pieces. Kraemer et al. argued that even short gaps were enough to evoke auditory imagery in familiar pieces and, furthermore, that this indicated the obligatory nature of auditory imagery. However, there is some question as to whether activation patterns during gaps in Kraemer et al.'s study reflect imagery, as information regarding the length of the continuation of the melody in imagery during a gap was not reported or correlated with fMRI (see Zatorre & Halpern, 2005).

The potential obligatory nature of auditory imagery was further examined by Janata (2001), who measured emitted potentials related to auditory imagery. In two experiments, Janata presented participants with ascending or descending melodic phrases consisting of eight notes. In a perception condition, the entire melodic phrase was presented. In different imagery conditions, the first three or five notes were presented, and participants were instructed to image the remaining notes. Additionally, in some imagery conditions, participants were cued to expect to hear three notes (and so they would image five notes) or cued to expect to hear five notes (and so they would image three notes), and in a small percentage of the trials, the cue was invalid (e.g., participants expected to hear five notes but only heard three notes). In all conditions, participants pressed a lever when they thought the final note of the sequence would occur. Imaging a continuation, as well as expectation of a note that was not presented, resulted in emitted potentials highly similar to evoked potentials for perceived notes. Janata suggested that similarities of topologies of electroencephalogram (EEG) during the N100 time window when a continuation was imaged, when a note was expected but not presented, and when a note was perceived are consistent with the hypothesis that auditory imagery, auditory expectation, and auditory perception activate similar brain mechanisms. Similarities in the topologies between subsequent imaged notes and perception were weaker, and this might reflect decreased expectations that those subsequent notes would be perceived.

Leaver, van Lare, Zielinski, Halpern, and Rauschecker (2009) examined anticipatory auditory imagery of melodies. In one study, participants listened to the final portion of a music track from a familiar or unfamiliar compact disc (CD). When the track was from a familiar CD, participants reported auditory imagery of the upcoming track occurred during the subsequent silence, but when the track was from an unfamiliar CD, no such imagery was reported; such a pattern is consistent with effects of familiar or unfamiliar music on auditory imagery during a silent gap in Kraemer et al. (2005). Additionally, fMRI was acquired, and reports of auditory imagery were associated with increased activation in right hemisphere superior frontal gyrus, presupplementary motor cortex, dorsal premotor cortex, and inferior frontal gyrus. In a second study, participants learned pairs of novel melodies. Participants were then instructed to form an image of the second member of a pair when they heard the first member of that pair. Increases in rated vividness of auditory imagery correlated with increased activation of right globus pallidus/putamen and left inferior ventral premotor cortex, and Leaver et al. argued that this

was consistent with reports that participants generated auditory imagery. Also, more activation occurred in parietal and subcortical (e.g., basal ganglia and cerebellum) structures for newly learned melodies in the second study than for well-learned melodies in the first study, and Leaver et al. suggested that this reflected a progression in the learning of melodies consistent with motor sequence learning.

Zatorre, Halpern, and Bouffard (2009) examined manipulation of auditory images of melodies. Participants were presented with an auditory excerpt of the opening notes of a well-known melody (e.g., the *Pink Panther* theme) or with the visual written title of such a melody. A probe string of auditory notes was presented, and in the conditions of interest, participants judged whether the probe was an exact reversal (i.e., the ordering of notes in the probe was reversed) or a false reversal (i.e., an exact reversal in which one or two of the notes were changed) of the beginning of the previously indicated melody. Additionally, fMRI was acquired. Participants reported using auditory imagery in the task, and activation of posterior parietal cortex during judgment correlated with ratings of auditory imagery vividness. Zatorre et al. suggested that parietal activation reflected an amodal mechanism for manipulation of auditory imagery rather than a mechanism for generation of auditory imagery, and this is consistent with findings that posterior parietal cortex exhibits significant activation during manipulation of visual imagery (e.g., Zacks, 2008). Ventrolateral and dorsolateral frontal cortex and right auditory cortex were also activated during reported auditory imagery. Participants were faster to respond to false reversals than to exact reversals, and Zatorre et al. suggested that participants stopped scanning the auditory image as soon as they reached an incorrect note in a false reversal but scanned the entire auditory image in an exact reversal.

Harmony

Hubbard and Stoeckig (1988) presented participants with a cue composed of a single tone or a major chord, and participants were instructed to form an image of what that cue would sound like if it were raised in pitch one semitone (participants were given examples of and practice with feedback in imaging an increase in pitch of one semitone). After participants had an image, they pressed a key, and a probe tone or probe major chord was presented; the participants' task was to compare the pitch(es) in their image with the pitch(es) in the probe. The patterns of response times and accuracy rates as a function of harmonic relatedness of the image and the probe matched patterns for perceived cues and probes reported in the harmonic priming literature (e.g., Bharucha & Stoeckig, 1986). Additionally, images of major chords required more time to generate than did images of single tones, and this was consistent with findings in the visual imagery literature that images of complex objects required more time to generate than did images of simple objects (e.g., Kosslyn, Reiser, Farah, & Fliegel, 1983). The pattern of priming suggested that processing of musical images involved the same mechanisms as did processing of musical percepts; given that speeded response measures were used, the alternative possibility of demand characteristics (in which participants within a few hundred milliseconds identified the harmonic relationship between the image and the probe and adjusted their response times and accuracy rates to reflect that relationship) was rejected.

Meyer, Elmer, Baumann, and Jancke (2007) examined EEG responses when participants perceived or were instructed to image (a) major or minor chords or (b) verbal syllables (data regarding verbal syllables are discussed in Part 3). Auditory stimulus presentation or image formation was time-locked to presentation of a visual stimulus, and participants pressed a button concurrent with the offset of the auditory percept or image. Instructions to generate auditory images of chords resulted in a N1 component in the window of 109–143 ms (cf. Janata, 2001) and a late positive component (LPC) in the window of 400–460 ms (cf. Wu, Mai, Chan, Zheng, & Luo, 2006). Topology of the EEG response was consistent with the hypothesis that the N1 in auditory imagery was associated with anterior temporal regions and that the LPC in auditory imagery was associated with activity in the cingulate, cuneus, medial frontal regions, and right auditory association cortex (cf. Zatorre & Halpern, 1993). A significant difference in button press latencies for perception and for imagery was reported; however, in the absence of concurrent behavioral evidence of imagery, such a difference does not (a) rule out the possibility other types of representation were involved or (b) support the hypothesis that observed EEG reflects activity unique to auditory imagery. Furthermore, it is not clear why image offset would be measured in the absence of specific instructions regarding how long an image should be maintained or any task in which the image could be used.

Tempo and Duration

As noted earlier, Janata and Paroo (2006) reported that temporal acuity in auditory imagery was reduced relative to temporal acuity in auditory perception. Halpern (1988a) visually presented participants with two lyrics from well-known melodies (e.g., *Do Re Mi*). In the imagery condition, participants were instructed to focus on the first lyric and then “mentally play” through the melody until they arrived at the second lyric. In a nonimagery control condition, participants judged whether the second lyric was in the same melody as was the first lyric. For the imagery condition, response times increased with increases in the distance of the second lyric from the first lyric (see also Zatorre & Halpern, 1993; Zatorre et al., 1996). This suggests the auditory image preserved the temporal structure of the melody. Response times also increased the further into the melody the first lyric was located, and this suggests participants began scanning at the beginning of the melody regardless of the location of the first lyric. However, the use of lyrics allows for the possibility that participants could have used a form of representation other than auditory imagery (e.g., abstract verbal or propositional representation). To address this possibility, Halpern presented pairs of lyrics, and participants judged whether the pitch of the second lyric was higher or lower than the pitch of the first lyric. As before, response times increased with increases in the distance between the lyrics.

The findings of Halpern (1988a) suggested that auditory images were extended in time, and so Halpern (1988b) examined whether tempo was represented in a consistent way. The ability of musical ensembles to perform without a conductor suggests tempo can be represented, and Halpern (1988b) noted anecdotal reports that auditory images of melodies seemed to specify the tempi of those melodies. In a perception condition, participants could adjust the tempo on a computer recording. In an imagery condition, partici-

pants were given the title of a familiar melody, instructed to image that melody, and then adjusted a metronome to correspond to the tempo in their image. Tempo settings differed across melodies, suggesting that participants differentiated the melodies. More importantly, the correlation between perceived tempo and imaged tempo for each melody was high ($r = .63$), suggesting that auditory imagery preserved tempo information. Halpern (1992) had participants image familiar melodies and tap a finger along with the imaged beat. The tapping rate for each melody was measured four times, and there was little within-melody variability in the measurements. When participants adjusted a metronome to reflect the acceptable maximum or minimum tempo in their image for a given melody, participants with more musical experience chose more extreme (faster or slower) tempi than did participants with less musical experience and so were apparently able to image melodies at more extreme tempi. This suggests increases in musical experience might increase flexibility in representing tempo within auditory imagery.

Notational Audiation

Using auditory imagery to “hear” music that is notated in a visually perceived musical score is referred to as *notational audiation* (e.g., Gordon, 1975). Waters, Townsend, and Underwood (1998) had trained pianists silently read cards that contained notation of a measure of piano music. The pianists were successful in judging whether subsequently presented auditory sequences matched the notations on the cards. Wöllner, Halfpenny, Ho, and Kurosawa (2003) presented undergraduate voice majors with notations of single line melodies. The notation was read in silence or in the presence of an auditory distractor. Participants then sang the melody aloud, and whether the initial reading (i.e., the notational audiation) had been accompanied by an auditory distractor did not influence subsequent performance. However, Brodsky, Kessler, Rubenstein, Ginsborg, and Henik (2008) suggested that performance in Waters et al.’s and Wöllner et al.’s studies did not conclusively address the potential role of auditory imagery in notational audiation but might reflect nonimagery information or processes (e.g., structural harmonic analyses, cues based on visual contour). More specifically, Brodsky et al. suggested that if silent reading of visually notated music evokes auditory imagery, then processing of notated music should be influenced by a concurrent stimulus or task that engages mechanisms used in audiation (i.e., used in auditory imagery) but should not be influenced by a concurrent stimulus or task that does not engage mechanisms used in audiation.

Brodsky, Henik, Rubenstein, and Zorman (2003) developed the embedded melody task as a way to investigate notational audiation. In this task, a well-known theme is incorporated into visual notation of a larger phrase. This theme is visually indiscernible in the phrase within which it was embedded, but Brodsky et al. suggested that it might be available to the “mind’s ear” (i.e., to notational audiation) when reading musical notation. Participants silently read notation containing an embedded melody in one of four conditions: (a) nondistracted reading, (b) while tapping a steady rhythm and listening to a task-irrelevant rhythmic pattern (rhythmic interference), (c) while wordlessly singing or humming (phonatory interference), or (d) while listening to a recording of themselves wordlessly singing or humming (auditory interfer-

ence). Participants then heard an auditory presentation of a melody and judged whether that melody was the same as the melody embedded in the previously read notation. A perceptual condition in which the larger phrase within the embedded melody was presented auditorily was also examined. On the basis of similarities of speech and vocal music, Brodsky et al. predicted that recognition of embedded melodies in imagery would be disrupted more by phonatory interference than by rhythmic or auditory interference, and indeed, recognition of embedded melodies in imagery was lowest when phonatory interference was present. Also, participants were more successful recognizing embedded melodies in perception than in imagery. Brodsky et al. suggested that notational audiation involves kinesthetic-like covert phonatory processes.

Brodsky et al. (2008) examined potential motor contributions to notational audiation. Participants completed the same embedded melody task with many of the same interference conditions as in Brodsky et al. (2003). In one study, activity level of muscles near the vocal folds was monitored during the embedded melody task and was compared with activity level of those same muscles during control tasks (e.g., reading printed text, silent mathematical reasoning). Recognition of the embedded melody was lowest when phonatory interference was present, and interestingly, the pattern of subvocal muscle activity was much more dynamic during silent reading of visual notation than during control tasks. In a second study, participants also made motor movements on their instrument appropriate to the notated music (e.g., pressing keys on a silent keyboard). Recognition of the embedded melody was lowest when phonatory interference was present, but addition of motor movements improved recognition of embedded melodies when rhythmic interference was present. In a third study carried out with professional drummers as participants, recognition of the embedded melody was lowest when phonatory interference was present, and this occurred even though the visual (drum-kit) notation did not involve pitch or tonality. Brodsky et al. (2008) suggested that both phonatory and motor processing were involved in notational audiation and that phonatory resources in notational audiation are not influenced by instrument or by notational system.

Highben and Palmer (2004) assessed a possible contribution of notational audiation to the learning and performance of novel music. The participants were pianists and were given a musical score to learn. In the conditions of interest, participants received auditory feedback (i.e., heard themselves play) or did not receive auditory feedback (i.e., pressing the piano keys did not produce sound) as they practiced. In the latter condition, participants were instructed to imagine how the notated music would sound as they practiced the movements necessary to perform that music (i.e., to engage in notational audiation). The notation was then removed, and all participants performed the piece from memory and with normal auditory feedback. Performance was better for participants who received auditory feedback during practice than for participants who did not receive auditory feedback (i.e., who engaged in notational audiation). All participants then completed a test of notational audiation in which they listened to a short melody while viewing a musical score, and they indicated whether the melody in the auditory presentation was the same as the melody in the visual score. Performance on the prior learning task was influenced less by absence of auditory feedback in participants with higher notational audiation scores; Highben and Palmer suggested that audi-

tory imagery aids memorization of music and is important for successful musical performance from memory.

Kalakoski (2007) presented musicians and nonmusicians with serial sequences of visual notes on a musical staff. Musicians recalled notes better than did nonmusicians, and musicians' recall was influenced by the musical well-formedness of the note patterns (similar to effects of expertise in recall of meaningful patterns in other domains, e.g., Chase & Simon, 1973). When notes were presented as lists of tone chroma (e.g., "A," "F-sharp," "D-flat"), musicians' recall did not differ from when notes were presented on a musical staff, and so better performance of musicians than of nonmusicians when notes were presented on a musical staff was not due to nonmusical elements of presentation (e.g., perceptual visual chunking). Kalakoski suggested that musicians formed auditory images of the indicated pitches (i.e., engaged in notational audiation) that aided memory. Schurmann, Raji, Fujiki, and Hari (2002) presented musicians with visual notes and instructed them to image the corresponding sounds (i.e., engage in notational audiation). Brain activity was measured using magnetoencephalography, and an initial activation of left and right occipital areas spread to the midline parietal cortex (precuneus) and then to the left temporal auditory association areas and the left and right premotor areas. However, neither Kalakoski nor Schurmann et al. reported convergent behavioral data that would support the claim their findings were due to properties of auditory imagery rather than to any other form of representation.

Environmental Sounds

As noted earlier, participants can adjust the pitch (Intons-Peterson et al., 1992) or loudness (Intons-Peterson, 1980) in an auditory image of the sound made by a common environmental stimulus to match the pitch or loudness in an auditory image of the sound made by a different common environmental stimulus. Stuart and Jones (1996) instructed participants to form an auditory image of the sound made by an object named by a visually presented word, and they found that imaged environmental sounds could prime subsequently perceived environmental sounds from the same category (e.g., sounds of transport, nature, common household objects). Similarly, Schneider, Engel, and Debener (2008) presented participants with visual pictures of objects and recordings of different object sounds. In the condition most relevant to the current concern, a picture of an object was presented, and then a sound was presented that was appropriate or not appropriate as an example of a sound an object of the type in the picture would make. Participants judged whether the presented sound was appropriate to the object, and responses were faster when the sound was appropriate to the pictured object than when the sound was not appropriate to the pictured object. To the extent that a visual picture can evoke an auditory image of that pictured object, the results of Schneider et al. are consistent with the hypothesis that an auditory image evoked by a visual picture can prime a subsequent auditory percept. However, the extent to which any representation of sound evoked by the picture would necessarily involve auditory imagery is not clear.

Bunzeck, Wuestenberg, Lutz, Heinze, and Jancke (2005) acquired fMRI data from participants who (a) viewed familiar visual scenes and heard sounds appropriate to those scenes, (b) viewed familiar visual scenes and were instructed to image sounds appro-

priate to those scenes, or (c) viewed scrambled versions of visual scenes without any accompanying sound and without any instructions regarding auditory imagery. When participants heard appropriate sounds, there was bilateral activation in primary auditory cortex (Heschl's gyrus) and secondary auditory cortex (planum temporale), but when participants were instructed to generate appropriate sounds in auditory imagery, there was activation in secondary auditory cortex but not in primary auditory cortex. This pattern is consistent with activation of secondary auditory cortex and relative lack of activation of primary auditory cortex in Yoo et al. (2001) and with greater activation of auditory association areas during silent gaps embedded in familiar music (and in which auditory imagery was reported) in Kraemer et al. (2005). Bunzeck et al. concluded that auditory imagery and auditory perception involve overlapping neural structures in secondary auditory cortex but that auditory imagery does not involve structures in primary auditory cortex that are involved in auditory perception (cf. Halpern et al., 2004).

Wu et al. (2006) had participants (a) view a picture of an animal or (b) view a picture of an animal and simultaneously generate an auditory image of the sound typically made by that type of animal. Differences in event-related potential waveforms between these two conditions were assumed to reflect neural correlates of auditory imagery. A perception control condition was also included in which an animal sound was paired with an animal picture. To ensure attention to the stimuli, a question mark would appear after some trials, and participants judged whether the stimulus presented after the question mark was the same as the stimulus presented before the question mark. There was no difference in N1 and P1 between imagery and control conditions, but P2 in the imagery condition was larger than P2 in the control condition, and this was interpreted as reflecting the additional attentional component in imagery. There was also a larger LPC in auditory imagery during the window of 350–600 ms after stimulus presentation (cf. Meyer et al., 2007), and this was interpreted as reflecting retrieval of auditory imagery information from memory and activation of the phonological loop for rehearsal. However, neither Wu et al., Bunzeck et al. (2005), nor Yoo et al. (2001) presented behavioral evidence that participants in imagery conditions were actually generating images (and that the auditory image involved a sound appropriate to the type of stimulus being viewed) and that the patterns of brain activation that they reported reflected activity specific to auditory imagery.

Part 3: Language

Much of introspective experience involves verbal information and an "interior monologue" (i.e., using auditory imagery of speech to "talk to one's self"), and so it could be argued that auditory imagery for language is a primary component of conscious experience and cognition. Accordingly, auditory imagery for speech, the extent to which speech images can be reversed or reinterpreted, speech imagery evoked by reading visually presented text, and speech imagery in dreaming and in waking interior monologue are examined in this section.

Speech

Weber and Bach (1969) instructed participants to form visual images or speech images of letters of the alphabet. Participants

were instructed to image each letter in sequence until they had completed two passes through the alphabet. Response times were collected and compared with responses times in a control condition in which participants recited each letter in sequence until they had completed two passes through the alphabet. Response times for speech and for speech imagery were identical, and both were faster than response times for visual imagery. Additionally, when asked to localize the experience within their heads, participants indicated a more frontal location in the visual imagery condition than in the other conditions. Weber and Castleman (1970) instructed participants to form visual images or speech images of letters of the alphabet in sequence until they completed one pass through the alphabet. Weber and Castleman replicated the faster response times for speech imagery than for visual imagery found by Weber and Bach and also reported that visual imagery was rated as more fatiguing than was speech imagery. Although Weber and Castleman included measures designed to help ensure visual images were being generated in the visual imagery condition (e.g., having participants judge whether lower case letters were vertically large [e.g., “d,” “f,” “g,” “h,” “j”] or small [e.g., “a,” “c,” “e,” “i,” “n”]), there were no equivalent measures to help ensure auditory images were being generated in the speech imagery conditions in Weber and Castleman’s or in Weber and Bach’s studies.

Anderson (1982) had participants rehearse (a) sequences of letters, days of the week, and months of the year or (b) spatial arrays corresponding to the living room, kitchen, and bedroom in the participant’s home. Rehearsal involved generating speech imagery or visual imagery of the stimuli. The faster speech imagery than visual imagery reported by Weber and colleagues (Weber & Bach, 1969; Weber & Castleman, 1970) for letters was replicated, but rehearsal (i.e., image generation) times were influenced by the type of object being imaged; for sequential/serial stimuli (letters, days, month), rehearsal with speech imagery was faster than was rehearsal with visual imagery, whereas for spatial/pictorial stimuli (living room, kitchen, bedroom), rehearsal with speech imagery was slower than or equal to rehearsal with visual imagery. Anderson concluded that rehearsal is faster and easier when processing mode (i.e., speech imagery, visual imagery) is compatible with characteristics of the material (e.g., sequential/serial, spatial/pictorial) than when processing mode and materials are incompatible. A similar effect of compatibility was found in Stuart and Jones (1996), who visually presented a word and instructed participants to form an auditory image of (a) what that word would sound like when pronounced or (b) the sound typically made by the object named by that word. Auditory images of spoken words primed subsequent perceptual recognition of those words, and as noted earlier, auditory images of environmental sounds primed subsequent perceptual recognition of those environmental sounds; however, priming did not occur when the auditory image did not match (i.e., was incompatible with) the subsequent perceived stimulus to be recognized.

Meyer et al. (2007) examined EEG responses when participants perceived or were instructed to image verbal syllables (i.e., “ka,” “ta,” “pa”). As noted earlier, auditory stimulus presentation or image formation in Meyer et al.’s study was time-locked to presentation of a visual stimulus, and participants pressed a button concurrent with offset of the auditory percept or image. Instructions to generate auditory images of syllables resulted in a N1 component in the window of 109–143 ms (the same as Meyer et

al., 2007, reported for images of chords; cf. Janata, 2001) and a LPC in the window of 290–330 ms (cf. Wu et al., 2006). Topology of the EEG response was consistent with the hypothesis that the N1 in auditory imagery was associated with activity in anterior temporal regions and that the LPC in auditory imagery was associated with activity in the cingulate, cuneus, medial frontal regions, and bilateral anterior temporal lobes. However, and as noted earlier, Meyer et al. did not offer independent behavioral evidence that participants were generating imagery or that the EEG patterns in the imagery condition reflected processes unique to auditory imagery. Aleman and van’t Wout (2004) instructed participants to form speech images of bisyllabic words and indicate which syllable carried the stress, and performance was compared with distractor conditions in which participants engaged in concurrent articulation of an irrelevant sound or in finger tapping. Articulation and tapping affected auditory imagery equally, but articulation had less effect on a visual imagery control task than did tapping. Aleman and van’t Wout suggested that the greater effect of articulation on auditory imagery than on visual imagery indicated a role of subvocalization in auditory imagery.

Oppenheim and Dell (2008) created four-word sequences in which the last two words of each sequence had been previously shown to lead to an increased probability of speech errors. Participants repeated each sequence four times, there was a brief pause, and then participants were cued whether the next set of repetitions should involve overt speech (i.e., be spoken) or inner speech (i.e., be imagined). After another brief pause, participants began speaking or imaging the sequence at a faster tempo. Errors in overt speech and errors in inner speech were self-reported. For overt speech and for inner speech, a lexical bias (i.e., errors tended to result in words rather than in nonwords; Costa, Roelstraete, & Hartsuiker, 2006) was reported. However, a phonemic similarity effect (i.e., errors tended to result in exchanging similar phonemes; Shattuck-Hufnagel & Klatt, 1979) was reported for errors in overt speech but not for errors in inner speech. Oppenheim and Dell suggested that inner speech contains lexical information but does not contain phonological information that would support phonemic similarity. However, this study relies on veridicality of self-reports, and it is not clear whether participants reported all errors or whether there was a bias such that lexical errors were noticed and reported in imagery more than were phonemic errors. Without independent measures of the accuracy of self-reporting of different types of speech errors in imagery and in perception, no conclusions can be drawn.

Geiselman and Glenny (1977) had participants listen to recordings of a male voice or a female voice. Participants were then visually presented with word pairs and instructed to image the words spoken by the male voice, female voice, or in their own voice. Participants were more likely to later recognize words spoken by the same voice that they had imaged during learning. Such a finding probably reflects the general cognitive principle of encoding specificity (i.e., probability of recall is maximized when the cues present at retrieval match the cues present at encoding; Tulving, 1983) rather than a unique property of auditory imagery. Johnson, Foley, and Leach (1988) presented lists of words that were spoken or imaged in a specific speaker’s voice. Participants later had difficulty distinguishing between words that were spoken and words that they had imaged being spoken. McGuire, Silbersweig, Murray, et al. (1996) presented visual words, and partici-

pants generated sentences in response to those words. In an inner speech condition, sentences were silently articulated, and in an auditory imagery condition, participants were instructed to image sentences spoken in another person's voice. Brain imaging using PET was acquired; both inner speech and auditory imagery were associated with increased activity in left inferior frontal gyrus, and auditory imagery was additionally associated with increased activity in the left premotor cortex, supplementary motor area, and left temporal cortex.

Regions of the left hemisphere corresponding to Broca's area exhibit increased activity during the generation of covert speech (i.e., during speech imagery; e.g., Bookheimer, 2002; Hinke et al., 1993), and application of transcranial magnetic stimulation (TMS) to the left hemisphere can disrupt overt speech production (e.g., Stewart, Walsh, Frith, & Rothwell, 2001). Building on these two findings, Aziz-Zadeh, Cattaneo, Rochat, and Rizzolatti (2005) examined whether application of TMS to areas of the left hemisphere might disrupt inner (covert) speech production (i.e., if TMS might disrupt verbal auditory imagery). Participants counted the number of syllables in visually displayed letter strings, and the latency for responding increased with increases in the number of syllables. Performance was generally faster when participants counted covertly than when participants counted overtly (i.e., performance was faster when speech imagery was generated than when actual speech was generated; cf. Weber & Bach, 1969). Application of TMS over anterior or posterior areas (over Broca's area and motor areas, respectively) of the left hemisphere resulted in (a) distortion of speech when participants counted overtly, and (b) significant increases in response time when participants counted overtly or covertly. Application of TMS over posterior/motor areas of the right hemisphere interfered with overt speech but did not appear to interfere with covert speech.

Shergill et al. (2001) acquired fMRI when participants were instructed to engage in auditory imagery that involved inner speech (silent articulation about the participant), first-person (about the participant and in the participant's own voice), second-person (about the participant but spoken by another voice), or third-person (about another person and spoken by another voice) content. Inner speech resulted in increased activation in left inferior frontal/insula cortex, left temporo-parietal cortex, right cerebellum, superior temporal gyri, and supplementary motor area. Relative to inner speech, first-person imagery resulted in greater activation in left insula, precentral gyrus, and lingua gyrus; bilateral activation in middle temporal gyri and posterior cerebellar cortex; and changes in right middle frontal gyrus, inferior parietal lobule, hippocampus, and thalamus. Relative to first-person imagery, second- and third-person imagery resulted in greater activation of the supplementary motor area, left precentral and middle temporal gyri and inferior parietal lobule, and right superior temporal gyrus and posterior cerebellar cortex. Relative to a perception control in which participants listened to speech, second- and third-person imagery resulted in greater activation in medial parietal lobule and posterior cingulate gyrus. However, the extent to which participants imaged different voices was not independently assessed, and it is not clear that articulatory properties of the voice in inner speech would necessarily be different from articulatory properties of the voice in first-, second-, or third-person imagery.

Reversibility and Reinterpretation

Research on visual imagery suggests that experimental participants cannot reverse or reinterpret a visual image of a visually ambiguous figure, and this notion is based on findings that participants who generated an image of a visually ambiguous figure according to one interpretation were not able to "see" an alternative interpretation in a visual image of that figure (e.g., Chambers & Reisberg, 1985; Reisberg & Chambers, 1991; but see Finke, Pinker, & Farah, 1989; Peterson, Kihlstrom, Rose, & Glisky, 1992). This finding has been interpreted as suggesting visual images are inherently meaningful (i.e., interpreted). In an extension of this idea to auditory imagery, Reisberg, Smith, Baxter, and Sonenshine (1989) exploited the verbal transformation effect, in which some speech sounds, if repeated, offer multiple interpretations (e.g., a string of repetitions of the word "life" could also be parsed as a string of repetitions of the word "fly"; Warren, 1968). If auditory images are inherently meaningful, then participants who image such a stimulus under one specific interpretation (e.g., repetitions of the word "life") should not be able to subsequently "hear" an alternative interpretation (e.g., repetitions of the word "fly") of that image. In Reisberg et al.'s study, participants who imaged repetitions of the word "stress" were able to "hear" the alternative interpretation "dress" only when auditory imagery was accompanied by subvocalization. When participants were prevented from subvocalizing during auditory imagery (e.g., by chewing candy, keeping their jaws closed), those participants were unable to report an alternative interpretation (see also Smith et al., 1995).

Text

Auditory imagery can influence reading of text. Abramson and Goldinger (1997) found that participants required more time to respond to visual presentations of words containing long vowel sounds than to respond to visual presentations of words containing short vowel sounds, even when words were equated for orthographic length. Even when reading silently, participants appeared to access phonological representations that preserved temporal properties of spoken words. Further evidence that auditory temporal information might be accessed during reading was reported by Kosslyn and Matt (1977), who had participants read silently or read aloud passages they were told had been written by an author with a fast speaking rate or a slow speaking rate. When participants read aloud, reading times were significantly slower for passages attributed to an author with a slow speaking rate than for passages attributed to an author with a fast speaking rate. However, when participants read silently, reading times were not influenced by the speaking rate of the author. A similar result was reported by Alexander and Nygaard (2008) when participants read silently or read aloud a passage attributed to a previously heard speaker who had spoken at a slow rate or a fast rate. The findings of Kosslyn and Matt and of Alexander and Nygaard are different from the findings of Abramson and Goldinger, but this might be because the task of the former involved articulatory and auditory information, whereas the task of the latter involved primarily auditory information.

The effect of speaking rate when participants read aloud but not when participants read silently in Kosslyn and Matt (1977) and in

Alexander and Nygaard (2008) is consistent with the hypothesis that phonological aspects of auditory imagery were activated when participants read aloud but not when participants read silently. Alternatively, in reading aloud, differences in reading times might have occurred if participants mimicked vocal characteristics of the original speaker and not because of any intrinsic or necessary property of auditory imagery. Alexander and Nygaard also collected self-ratings of imagery, assessed imagery ability of participants (with Aleman et al.'s, 2000, task of asking which one of three visually named items would produce the most different sound), and varied text difficulty. More difficult text slowed reading speed and increased the likelihood participants would report using speaker-specific representation (i.e., imagery). For difficult text, engaging in auditory imagery for the speaker's voice was not related to imagery ability, as all participants were more likely to access speaker-specific (rather than abstract) representations; however, for easy text, participants high in auditory imagery were more likely to engage in auditory imagery for the speaker's voice than were participants low in auditory imagery. Self-report measures revealed that a significant portion of participants imagined hearing the voices of the speakers in their images; also, participants who reported hearing the voices of the speakers in their images were more likely to score high on auditory imagery ability.

Dreaming

One domain in which auditory imagery is potentially widespread, but that has not received sufficient laboratory investigation, is dreaming. Snyder (1970) collected 635 reports of dream content from adult participants awakened from rapid eye movement (REM) sleep. A small percentage of reports (13%) mentioned nonverbal auditory imagery, but a majority of reports (94%) mentioned speech or conversation imagery. These data suggest auditory imagery is frequent during REM sleep, and when auditory imagery occurs during REM sleep, that auditory imagery usually involves speech. Heynick (1983) had participants report, upon awakening, the last line of speech they remembered from a dream. This sentence was usually one uttered by the self-character in the dream and was judged by dreamers to be a sentence they might say or hear during waking life. Examination of children's dreams (e.g., Foulkes, 1982) suggests significant auditory imagery does not appear in dreams prior to the transition between preoperational thinking and concrete operational thinking. Although there are numerous cases in the clinical literature regarding abnormalities in visual imagery in dreams following neurological damage that resulted in abnormalities in visual perception, there are few cases in the literature regarding abnormalities of auditory imagery in dreams following neurological damage (e.g., a single report of aphasic imagery in REM dreams described by M. Harrison, 1981; see discussion in Solms, 1997). Although suggestive, findings regarding auditory imagery in dreaming must be viewed tentatively until more controlled studies can be carried out.

Waking Interior Monologue

Speech imagery directed toward one's self (i.e., *interior monologue*) is a major component of general waking cognitive functioning (e.g., see Farthing, 1992; Klinger, 1999), but like auditory imagery in dreaming, it has not received sufficient investigation.

Hogenraad and Orianne (1983) had participants "think aloud" for a 3-hr period, and they found 60–90-min periodicities in the strength of imagery and interior monologue; interestingly, this periodicity is similar to that of REM sleep in which the majority of vivid nocturnal dreaming occurs (and as noted earlier, in which significant speech and language imagery occurs), and this similarity suggests that a consistent rhythm in imagery strength occurs across the circadian cycle (peaking in daydreams and REM dreams). Klinger and Cox (1987–1988) had participants carry beepers as those participants went about their daily lives. The beeper signaled at an unpredictable time, and when the beeper signaled, participants answered questions concerning their most recent thoughts. Nearly three quarters (73%) of the sampled thoughts contained some degree of interior monologue. Perhaps the most famous discussion of auditory imagery in interior monologue is Jaynes (1976), in which he posited that prior to the time of the Trojan War, verbal imagery generated by the right hemisphere of the brain was interpreted as voices of the gods by the left hemisphere. Jaynes's theory has been rejected by most scholars (e.g., Cavanna, Trimble, Cinti, & Monaco, 2007), but many issues involving auditory imagery and the role of interior monologue in everyday cognition and consciousness still await rigorous empirical investigation.

Part 4: Perception and Memory

Many of the studies and findings reviewed in Parts 1, 2, and 3 address relationships between auditory imagery and perception or between auditory imagery and memory. However, those studies and findings were narrowly focused on relatively specific elements of auditory imagery. In this section, relationships of auditory imagery to perception and to memory are more broadly considered, and the extent to which an auditory image can influence detection, encoding, and recall of a stimulus is examined. The mnemonic properties of auditory imagery, and the possibility that the phonological loop component of working memory (involving subvocalization and the phonological store) is involved in auditory imagery, are also examined.

Detection

There is a large quantity of literature that examines whether visual imagery can influence perception (e.g., Craver-Lemley & Reeves, 1992), but there is a much smaller quantity of literature that examines whether auditory imagery can influence perception. In a classic study, Segal and Fusella (1970) asked participants to detect a faint visual stimulus or a faint auditory stimulus while those participants simultaneously generated a visual image or an auditory image. Detection of visual targets decreased more when participants generated a visual image than when participants generated an auditory image, whereas detection of auditory targets decreased more when participants generated an auditory image than when participants generated a visual image. More generally, imagery in the same modality interfered with detection of a stimulus more than did imagery in a different modality (for complementary evidence that detection interferes with same-modality imagery more than with different-modality imagery, see Tinti, Cornoldi, & Marschark, 1997). Such an effect is consistent with raised thresholds for perceptual detection during daydreams con-

taining imagery (Antrobus, 1968; Antrobus, Singer, Goldstein, & Fortgang, 1970). As noted earlier, imagery of a specific auditory frequency can interfere with detection of an auditory stimulus of that frequency (Okada & Matsuoka, 1992). However, and as noted earlier, imagery can facilitate perception of a same-modality stimulus in vision (e.g., letters; Farah, 1985) and audition (e.g., major chords; Hubbard & Stoeckig, 1988) when participants judge whether a percept matches their image, and so it is not the case that auditory imagery always interferes with perception of auditory stimuli.

Encoding and Recall

Sullivan, Urakawa, and Cossey (1996) attempted in a pitch comparison task to separate effects of general alertness and effects of auditory imagery of a specific pitch. Participants heard two tones and indicated whether the tones were the same or different in pitch. The tones were each either 500 Hz or 1000 Hz, and the inter-stimulus interval (ISI) between the tones varied. Participants received one of two types of cue immediately prior to the first tone: a nonspecific cue (a vertical line) that did not indicate the pitch of the first tone, or a specific cue (the printed word “low” or “high”) that indicated the pitch of the first tone. Presentation of nonspecific cues allowed assessment of nonspecific effects of alertness, whereas presentation of specific cues allowed assessment of effects of expectation of a specific pitch (and presumably activation of imagery of that pitch) in addition to nonspecific effects of alertness. Response times were highest at the short ISI, but accuracy rates were not influenced by ISI. Response times were faster with *same* judgments than with *different* judgments and faster when a specific cue rather than a nonspecific cue preceded the first tone. However, Sullivan et al. noted that specific information did not change the general shape of the encoding function (i.e., the Response Time \times ISI function; see Posner, 1978), and they concluded that facilitating effects of auditory imagery were due to general increases in alertness rather than to specific pitches that were imaged. Even so, Sullivan et al. acknowledged the lack of specific pitch effects in imagery might reflect the restricted stimulus range.

Several studies presented participants with a pitch comparison task in which a standard pitch was presented, an interval of time that was silent or filled with a distractor stimulus elapsed, and then a comparison pitch was presented (e.g., Berti, Munzer, Schröger, & Pechmann, 2006; Deutsch, 1970, 1972, 1974; Pechmann & Mohr, 1992). In such studies, participants judged whether the comparison was the same pitch as the standard, and performance was usually best when the interval between the standard and comparison was silent. One possible explanation is that participants used auditory imagery to rehearse the pitch of the standard and that such imagery is easier to generate or maintain in the absence of distracting or competing perceptual stimuli. T. A. Keller, Cowan, and Sauls (1995) examined this hypothesis, and in a pitch comparison task, they (a) instructed participants to rehearse the pitch of the standard, (b) presented a verbal distractor (i.e., sequence of digits) designed to prevent rehearsal, or (c) presented a nonverbal distractor (i.e., sequence of tones) designed to prevent rehearsal. Performance decreased when a distractor was presented, but there was no difference in performance with verbal or nonverbal distractors. Furthermore, the decrease in performance in-

creased with longer intervals between the standard and the comparison. T. A. Keller et al. suggested that auditory imagery could slow decay of memory for pitch. However, an independent assessment of the extent to which auditory imagery was generated or used by participants was not presented, and it is not clear that rehearsal necessarily involved auditory imagery.

Mnemonic Properties

The mnemonic properties of visual imagery are well established. Techniques such as the Method of Loci (e.g., Bower, 1970b) and the use of pegwords (e.g., Bugelski, Kidd, & Segmen, 1968) increase the probability of recall of verbal material, and Paivio (1969, 1971, 1986) demonstrated that more visually concrete (i.e., visualizable) stimuli were recalled better. However, there has been less investigation of potential mnemonic properties of auditory imagery. One of the first examinations of auditory imagery as a mnemonic for free recall was by Sharps and Price (1992), who presented printed verbal labels, auditory recordings, and visual pictures of stimuli. Recall was better when participants were presented with auditory recordings or with visual pictures than when participants were presented with verbal labels, and recall did not differ as a function of whether participants were presented with auditory recordings or visual pictures. In a follow-up study, participants were presented with either auditory recordings or visual pictures or with both auditory recordings and visual pictures; recall in the combined condition was not significantly greater than when only visual pictures or only auditory recordings were presented. Sharps and Price suggested that auditory imagery possessed a mnemonic value similar to that of visual imagery but that the two modalities of imagery were not processed in different memory systems (as different systems would presumably have combined effects, and use of combined auditory and visual imagery would have resulted in greater recall than would use of a single modality of imagery).

In a further investigation aimed at examining the mnemonic advantage of more concrete auditory stimuli, Sharps and Pollitt (1998) presented sets of printed verbal labels, auditory recordings, and visual pictures similar to those of Sharps and Price (1992). In one experiment, stimuli were blocked by category (musical instruments, vehicles, animals, tools) or presented in a random order. Recall for auditory recordings and for visual pictures did not differ, and recall for both were higher than was recall for verbal labels. For verbal labels and for auditory recordings, recall was better when stimuli had been blocked (i.e., a category superiority effect occurred; cf. Gollin & Sharps, 1988), but blocking did not influence recall for visual pictures. Sharps and Pollitt suggested that nonverbal and nonmusical auditory imagery “has a somewhat intermediate character between visual images and verbal labels” (p. 114) because the pattern of recall for auditory recordings was similar to the pattern of recall for verbal labels in being influenced by the blocking manipulation, but the overall level of recall for auditory recordings was more similar to that obtained for visual pictures. Sharps and Pollitt further suggested that auditory images might involve processing resources from the visuospatial component and from the phonological loop component of working memory. However, neither Sharps and Price nor Sharps and Pollitt produced independent evidence that auditory imagery was necessarily used in encoding auditory stimuli.

Mnemonic effectiveness of visual images is enhanced if images of separate visual stimuli portray those stimuli as interacting (e.g., Bower, 1970a; Kroll, Schepeler, & Angin, 1986), and Tinti et al. (1997) examined whether a similar enhancement would occur if auditory images of separate auditory stimuli portray those stimuli as interacting. In one study, participants formed a visual image of a speaker and either simultaneous or successive auditory images of two words coming from that speaker. More words were recalled in the simultaneous condition than in the sequential condition. In a second study, participants were visually shown pairs of words and formed a visual interactive image or an auditory interactive image. After presentation of a word pair, participants had to detect a faint visual target or a faint auditory target. In general, more words were recalled with visual interactive imagery than with auditory interactive imagery, and recall decreased when the target was in the same modality as the image. In a third study, participants verbally elaborated word pairs (i.e., formed, spoke, and rehearsed a sentence containing the stimuli) or generated auditory interactive images. Auditory interactive imagery resulted in better recall than did verbal elaboration. Better recall with interaction of auditory images is consistent with previous findings of better recall with interaction of visual images, but better recall with visual interactive imagery than with auditory interactive imagery is not consistent with Sharp and Price's (1992) suggestion that auditory imagery had the same mnemonic value as visual imagery. More intriguingly, better recall with interactive auditory imagery than with verbal elaboration suggests that processes involved in auditory imagery do not depend directly or solely on the phonological loop.

The Phonological Loop

Baddeley and Logie (1992) suggested that the phonological loop subsystem of working memory provides a basis for auditory imagery and, furthermore, that such a role for the phonological loop is consistent with the articulatory suppression effect (i.e., memory for verbal materials is disrupted when an experimental participant suppresses rehearsal of those verbal materials by articulating an irrelevant sound; Baddeley, Lewis, & Vallar, 1984; Murray, 1965) and the irrelevant speech effect (i.e., memory for visually presented stimuli is disrupted by simultaneous presentation of spoken material that the participant is instructed to ignore; Salame & Baddeley, 1982). An objection to a phonological or articulatory (i.e., subvocalization) mechanism as the basis for auditory imagery initially raised by Crowder (1989; Pitt & Crowder, 1992) is that a significant amount of auditory imagery does not involve speech or speech-like stimuli (e.g., instrumental music, environmental sounds); however, Baddeley and Logie suggested that covert rehearsal of nonspeech sounds could be carried out by phonological mechanisms even if overt generation of nonspeech sounds by phonological mechanisms was not possible. Along these lines, Baddeley and Logie suggested that evidence is relatively stronger for involvement of the phonological loop in temporary storage of auditory imagery than for involvement of the phonological loop in generation of auditory imagery retrieved from long-term memory (i.e., evoked in the absence of an appropriate auditory stimulus).

Evidence consistent with a contribution of the phonological loop in auditory imagery is found in studies of the role of subvocalization and the inner voice in auditory imagery. Reisberg et al.

(1989); Smith, Reisberg, and Wilson (1992); and Smith et al. (1995) discussed differences between the *inner voice* and the *inner ear*. The inner voice is linked to subvocalization and involves articulatory information, whereas the inner ear is linked to the phonological store and involves more purely auditory information. Smith et al. (1995) replicated Reisberg et al.'s finding of the importance of subvocalization in reinterpreting an auditory image of a perceptually ambiguous stimulus; additionally, Smith et al. (1995) found that preventing subvocalization decreased performance when participants judged whether the pitch of an imaged melody increased or decreased between the second and third notes of that melody. Such a decrease in performance when subvocalization was blocked is consistent with findings of increased activity in the supplementary motor area during auditory imagery of a melody (Halpern & Zatorre, 1999; Zatorre et al., 1996) or timbre (Halpern et al., 2004). Smith et al. (1995) suggested that subvocalization and the phonological store worked in partnership in many auditory imagery tasks (e.g., reinterpreting a perceptually ambiguous auditory stimulus, scanning a familiar melody), but subvocalization could be used without the phonological store in other tasks (e.g., distinguishing voiced from unvoiced consonants), and neither subvocalization nor the phonological store was required in other tasks (e.g., judging homophones).

Several additional findings are consistent with involvement of the phonological loop in auditory imagery. Sharps and Pollitt (1998) suggested that a category superiority effect in memory for auditory recordings (i.e., auditory imagery) and for verbal labels supported a role for the phonological loop in auditory imagery. Aleman and van't Wout (2004) reported that auditory imagery for verbal materials was impaired when subvocalization was blocked. Weber and Brown (1986) reported that verbal responses interfered with musical imagery, and Brodsky et al. (2003, 2008) reported that phonatory processes interfered with recognition of an embedded melody in imagery. Aleman et al. (2005) reported that brain areas implicated in the phonological loop are activated by auditory imagery. As noted earlier, Smith et al. (1992, 1995) suggested that a role for subvocalization in several auditory imagery tasks, and linkage of the inner voice and the inner ear with subvocalization and the phonological store, respectively, is consistent with phonological loop architecture. However, evidence inconsistent with a role for the phonological loop in auditory imagery was presented by Tinti et al. (1997). An interpolated articulatory judgment task decreased recall of words that had been verbally elaborated, whereas when imagery of the word was emphasized, recall of words was decreased by an interpolated task involving recognition of sounds. Tinti et al. suggested that differential effects of an interpolated articulatory task demonstrated that auditory imagery does not necessarily depend on the phonological loop (consistent with Smith et al., 1995). Therefore, the phonological loop might contribute in more or in different ways to some types of auditory imagery than to other types of auditory imagery.

Part 5: Individual Differences

A consideration of responses to questionnaires and of records in diaries shows that many respondents report having experienced some form of auditory imagery (e.g., Klinger & Cox, 1987–1988; Kosslyn, Seger, Pani, & Hillger, 1990; McKellar, 1965). Despite this, there have been relatively few examinations of potential

individual differences in auditory imagery. In this section, potential individual differences in auditory imagery involving vividness, musical ability and experience, synesthesia, and the presence of some types of psychopathology are examined.

Vividness

There has been surprisingly little research on the vividness of auditory imagery in nonclinical populations. On the Auditory Imagery Scale (Gissurason, 1992), participants rate vividness of auditory images of sounds of common objects (e.g., telephone). Ratings of vividness from the Auditory Imagery Scale correlate positively ($r = .48$) with ratings of vividness from the revised Vividness of Visual Imagery Questionnaire (D. F. Marks, 1995). However, unlike ratings of vividness of visual imagery, which correlate with social desirability (e.g., see McKelvie, 1995), ratings of vividness of auditory imagery do not correlate with social desirability (Allbutt, Ling, Heffernan, & Shafiullah, 2008). Kosslyn et al. (1990) and Tinti et al. (1997) reported that auditory images are rated as more vivid than are visual images; additionally, Kosslyn et al. reported that vividness and typicality of auditory imagery are inversely related, and Tinti et al. reported that interactive auditory images are more vivid than noninteracting auditory images. Baddeley and Andrade (2000) suggested that vividness of auditory imagery is related to the strength of (i.e., the lack of interference with) the representation in the phonological loop. More generally, Baddeley and Andrade suggested that an experience of vividness of imagery requires abundant sensory information be available from long-term memory and maintained in an appropriate slave system in working memory. Attempts to relate vividness to activation patterns in brain imaging studies have not yielded consistent results across studies (cf. Leaver et al., 2009; Zatorre et al., 2009). Whether vividness of auditory imagery is related to musical ability and experience or to psychopathology is addressed below.

Musical Ability and Experience

A consideration of auditory imagery is a component of many tests of musical aptitude. Also, the extent to which auditory imagery is potentially involved in, related to, or influenced by ability, training, or experience in musical perception, performance, and composition is considered.²

Tests of musical aptitude. Many of the tests developed to assess musical talent or aptitude include measures of auditory imagery. Indeed, Seashore (1938/1967) suggested that a rating scale for mental imagery should be included in every set of measures of musical talent, and he suggested that the most outstanding mark of the musical mind is a high capacity for auditory imagery. The Seashore Measures of Musical Talent (originally developed in 1919 and then later revised; Seashore, Lewis, & Saetveit, 1960) was one of the first attempts to produce a standardized test of musical ability and included a scale for self-reported imagery. Imagery (in the form of remembered tones) is a component of the Standardized Tests of Musical Intelligence (Wing, 1961). The more commonly used Musical Aptitude Profile (Gordon, 1965) contains an extensive battery of tests that includes assessment of tonal imagery (involving melody and harmony) and of rhythm imagery (involving tempo and meter). Such measures

were developed to assess musical aptitude and predict future musical performance, and so far have contributed little to an understanding of auditory imagery or musical imagery per se.

Musical training. Aleman et al. (2000) visually presented participants with two lyrics from a well-known melody and asked which of the two lyrics would normally be sung at a higher pitch (see also Halpern, 1988a; Zatorre & Halpern, 1993; Zatorre et al., 1996). Participants with musical training exhibited more correct responses than did participants with little or no musical training (although response times were not influenced by musical training). In a control task intended to evoke nonmusical auditory imagery, participants were visually presented with names of three objects, and they reported which object would have produced a sound least like the sounds produced by the other objects. Participants with musical training performed better than did participants with little or no musical training. There was no difference between musically trained participants and musically untrained participants on a perceptual version of the task or on a visual imagery task. Aleman et al. suggested that the lack of differences between musically trained participants and musically untrained participants on the perception task or on the visual imagery task ruled out a better processing of auditory information or an enhancement of imagery in general and instead indicated an enhanced ability to organize and manipulate musical information within working memory. Seashore (1938/1967) reported that a group of musicians possessed stronger auditory imagery than did a group of psychologists, and as noted earlier, Janata and Paroo (2006) reported that participants with musical training exhibited better pitch acuity and better temporal acuity in auditory imagery than did participants with little or no musical training.

P. E. Keller and Koch (2008) had participants with varying levels of musical training respond to each of several different visually presented colored patches by producing a unique sequence of taps on three vertically aligned keys. Each tap triggered an auditory tone, and the key-tone pairing was compatible (taps on the top, middle, and bottom key produced a high, medium, and low tone, respectively) or incompatible (key-tone pairing was either scrambled, reversed, or the same tone was associated with each key). For participants with musical training, response times were shorter with compatible pairings than with incompatible pairings; however, for participants with little or no musical training, response times were not related to compatibility of the key-tone pairings. P. E. Keller and Koch suggested that increases in auditory imagery accompanied increases in musical training, and this increased auditory imagery primed compatible responses in participants with more musical training but not in participants with little or no musical training. Given such priming, incompatible pairings would result in slower responding by participants with more musical training. The notion that imagery could prime subsequent perception or action is consistent with the view that imagery involves expectation (cf. Janata, 2001), but an independent mea-

² Although imagery in music involves nonauditory components in addition to auditory components (e.g., motor representations appropriate to musical performance; e.g., Kristeva, Chakarov, Schulte-Monting, & Spreer, 2003; Meister et al., 2004; Zatorre & Halpern, 2005), consideration of nonauditory components of musical imagery is beyond the scope of this review.

sure of auditory imagery was not obtained, and the extent to which auditory imagery was actually involved in P. E. Keller and Koch's task is not clear.

Bailes (2007) assessed the prevalence and nature of musical auditory imagery in the everyday lives of music students. Participants received six telephone calls between 10 a.m. and 10 p.m. on each of 7 consecutive days, and upon receiving a telephone call, participants filled out a detailed questionnaire regarding their activities and whether they were hearing music or imaging music at the time of the call. Episodes of musical imagery were most common during working or during interaction with others and were also common during "filler" activities (e.g., waiting, lying in bed). Melody and lyrics were rated as the most vivid elements of musical auditory imagery, and expressive elements (such as dynamics and harmony) were rated as less vivid. The majority of reports that mentioned musical auditory imagery described the image as a repeated musical fragment rather than as a full or extended performance. Imaged music could generally be identified by name, and this suggested a high level of familiarity with the original musical piece; indeed, many participants reported having recently perceived the music that they subsequently experienced in imagery.

Sight-reading. The initial (unrehearsed) performance of a given piece of music is referred to as *sight-reading*, and the ability to sight-read is considered a basic skill for musicians (for reviews, see Gabrielsson, 1999; Sloboda, 1984). Curiously, among musicians there is only a weak relationship between general performance ability and sight-reading ability (Wolf, 1976). Increased auditory imagery was suggested by Kornicke (1995) to contribute to improved sight-reading ability. Waters et al. (1998) reported that better sight-readers exhibited greater priming in a harmonic priming task, and to the extent that auditory imagery involves expectations, this finding is consistent with the notion that better sight-readers have increased or more effective auditory imagery. Also, when briefly presented with visual music notation immediately followed by an auditory excerpt of music, better sight-readers were more accurate in judging whether the auditory excerpt matched the visual notation. Given that such a task appears to involve notational audiation, this also suggests that increased auditory imagery can contribute to improved sight-reading (but see Fourie, 2004). Kopiez, Weihs, Ligges, and Lee (2006) collected measures of several cognitive variables that potentially contribute to skill in sight-reading; for their measure of auditory imagery, they adapted the embedded melody task of Brodsky et al. (2003). Performance on the embedded melody task (and presumably auditory imagery) was not significant in a two-class or three-class discriminant analysis; however, as Kopiez et al. noted, the embedded melody task can be very difficult, and this might have restricted the range of its predictive power.

Composition. As Sloboda (1985, p. 103) noted, "composition is the least studied and least well understood of musical processes, and . . . there is no substantial psychological literature to review." The state of the psychological literature on musical composition has not changed significantly since Sloboda's assessment was made, and the empirical literature on the use of auditory imagery in musical composition is almost nonexistent (see also Mountain, 2001). Indeed, in a recent detailed longitudinal case study of composition carried out on a single composer (Collins, 2005), the words "image" or "imagery" do not even appear.³ Seashore (1938/

1967) presented several anecdotal examples from the writings of Robert Schumann, Wolfgang Amadeus Mozart, Hector Berlioz, and Richard Wagner that suggest that these composers utilized vivid auditory and musical imagery during the process of composition. Perhaps the most impressive example of the use of imagery in composition is Ludwig Beethoven's composition of the Ninth Symphony; this work was composed after Beethoven became deaf, and so he could not have perceived the music (Lockwood, 2002). Even if Beethoven relied primarily on semantic or syntactic knowledge of chords, intervals, and voice leading during the composition of the Ninth Symphony, he might have also relied on notational audiation and other aspects of auditory imagery to recreate or simulate auditory qualities of the Ninth Symphony (see also P. Harrison, 1988).

Synesthesia

In *synesthesia* (also referred to as *synaesthesia*), a stimulus in one modality or dimension induces sensory experiences in a second modality or dimension (for reviews, see Baron-Cohen & Harrison, 1997; Cytowic, 2002; L. E. Marks, 1975, 1978; Robertson & Sagiv, 2005). These experiences are automatic and reliable over time (Ward & Mattingly, 2006). Given that a stimulus in the second modality or dimension is not actually present, and given that imagery is considered to reflect sensory experience in the absence of an appropriate stimulus (cf. Finke, 1989; Intons-Peterson, 1992), synesthetic experiences clearly involve imagery. Given this, the types of auditory images induced by nonauditory stimuli in synesthesia should be of interest in any consideration of auditory imagery. However, consideration of auditory imagery in synesthesia should be careful to exclude those cases that do not meet the criteria for synesthesia (e.g., that are not automatic, vivid, consistent across time), and so more voluntary types of auditory images in response to a nonauditory stimulus (e.g., notational audiation upon reading a musical score) should not be considered examples of synesthesia. Also, synesthesia is generally not bidirectional (in the sense that either sensation can induce the other; e.g., Ward, Simner, & Auyeung, 2005; although a few exceptions have been reported; e.g., Kadosh, Kadosh, & Henik, 2007), and so cases in which an auditory stimulus induces nonauditory imagery do not imply that an equivalent nonauditory stimulus would induce auditory imagery.

The most common type of synesthesia involves experiences of color in response to a noncolor stimulus (e.g., graphemes: Baron-Cohen, Harrison, Goldstein, & Wyke, 1993; musical notation: Ward, Tsakanikos, & Bray, 2006), and examples of synesthesia in which a nonauditory stimulus induces auditory imagery appear relatively rare. In one of the few examples in the literature of synesthesia in which auditory imagery was induced by nonauditory stimulation, composer Jean Sibelius experienced different

³ Contemporary composers often have access to computer playback and Musical Instrument Digital Interface (MIDI) systems that can immediately reveal how newly composed music would sound. This type of technologically based *externalized imagery* might be filling the role traditionally occupied by more *internalized imagery* in previous decades and centuries, thus making auditory imagery less central or critical in composition.

musical chords when viewing different colors (Pearce, 2007).⁴ It is not clear whether the paucity of reports of auditory imagery in synesthesia reflects (a) a bias in the reporting or in the literature or (b) an important characteristic or limitation of auditory imagery in synesthesia. One admittedly highly speculative possibility is that synesthesia is constrained by a visual dominance similar to that in normative perception. In visual dominance, there is a bias in favor of visual information such that visual information takes priority over nonvisual information (e.g., see Colavita, 1974; Posner, Nissen, & Klein, 1976; Welch & Warren, 1986). Thus, if a nonvisual stimulus is presented and a visual image initially induced, given visual dominance, that visual image is experienced. However, if a visual stimulus is presented and an auditory image initially induced, the auditory image is not experienced because the general bias in favor of visual information would result in the auditory image being unattended or even suppressed.

Psychopathology

With the exception of synesthesia, examples of auditory imagery considered thus far involved auditory imagery that was under some degree of voluntary control. However, amongst individuals with mental illness or other psychopathology, the presence of auditory imagery is often involuntary, as well as unwanted, intrusive, and distressing (Shergill, Murray, & McGuire, 1998). Excesses, deficits, or other abnormalities in auditory imagery are a significant component of several psychopathologies, including musical hallucinosis, schizophrenia, and (potentially) amusia.

Musical hallucinosis. A clinical condition in which musical imagery is not voluntary, but impinges upon an individual involuntarily, is referred to as *musical hallucinosis* (e.g., Griffiths, 2000).⁵ Much of the literature on musical hallucinosis involves a focus on case histories (e.g., Fischer, Marchie, & Norris, 2004; Sacks, 2007), and estimates of the prevalence of musical hallucinosis in psychiatric patients range from 0.16% (Fukunishi, Horikawa, & Onai, 1998) to 27% (Hermesh et al., 2004). Findings from brain imaging studies are consistent with the hypothesis that musical hallucinosis might result from abnormal spontaneous activity in neural substrates underlying musical perception and imagery (e.g., Griffiths, 2000; Shinosaki et al., 2003). Additionally, there is evidence that bilateral hearing loss (Tanriverdi, Sayilgan, & Ozcurumez, 2001), dysfunction of the right auditory cortex (Kasai, Asada, Yumoto, Takeya, & Matsuda, 1999), desynchronization of the right auditory cortex (Shinosaki et al., 2003), right occipital meningioma (Nagaratam, Virk, & Brdarevic, 1996), brainstem lesions (Murata, Naritomi, & Sawada, 1994; Schielke, Reuter, Hoffman, & Weber, 2000), and obsession (Gomibuchi, Gomibuchi, Akiyama, Tsuda, & Hayakawa, 2000) are linked to musical hallucinosis. Although musical hallucinosis typically involves neurologically compromised individuals, a similar case (involving a “perpetual music track”) in a noncompromised individual has been reported (Brown, 2006). In less pathological cases, a persistent musical image (i.e., “I can’t get that song out of my head!”) is colloquially referred to as an *earworm* (e.g., Levitin, 2007).

Schizophrenia. Auditory hallucinations are a diagnostic criterion (American Psychiatric Association, 2000) and an important cognitive characteristic (George & Neufeld, 1985; Nayani & David, 1996) of schizophrenia. Although musical hallucinations

have been reported in patients with schizophrenia, auditory hallucinations in schizophrenia usually involve verbal or vocal content (Baba & Hamada, 1999; Saba & Keshavan, 1997). One possible account is that auditory hallucinations result from abnormal vividness of auditory imagery (for discussion, see Seal, Aleman, & McGuire, 2004; Smith, 1992). Although some studies suggested that auditory hallucinations in schizophrenia resulted from increased vividness of auditory imagery (e.g., Mintz & Alpert, 1972), other studies suggested that patients more prone to hallucinations had less vivid auditory imagery than did patients less prone to hallucinations (e.g., Seitz & Molholm, 1947; Starker & Jolin, 1982), and still other studies suggested that there were no differences in vividness of auditory imagery between patients more prone to hallucinations and patients less prone to hallucinations (e.g., Slade, 1976). In a review of this literature, Seal et al. (2004) concluded there was no evidence that exceptionally vivid or exceptionally weak auditory imagery was generally related to the presence of auditory hallucinations in schizophrenia. Similarly, Bentall and Slade (1985) concluded that the hypothesis that auditory hallucinations in schizophrenia resulted from abnormalities in auditory imagery was not supported.

Another possible account is that auditory hallucinations in schizophrenia occur because patients are unable to distinguish between imaged speech and external speech, and so they attribute their own imaged speech to an external source (e.g., Bick & Kinsbourne, 1987). Lennox, Park, Medley, Morris, and Jones (2000) compared fMRI of auditorily hallucinating schizophrenia patients with fMRI of those same patients at rest. Auditory hallucinations were related to increased activity in right and left superior temporal gyrus, left inferior parietal cortex, and left middle frontal gyrus. Lennox et al. suggested that this pattern supported the hypothesis that auditory hallucinations reflect abnormal activation of normal auditory pathways. Evans, McGuire, and David (2000) compared patients with schizophrenia who were more prone to auditory hallucinations with patients with schizophrenia who were less prone to auditory hallucinations on tasks suggested by Smith et al. (1995) to involve both subvocalization and the phonological store (parsing letter strings, judgment of pitch, reinterpreting an ambiguous auditory stimulus), subvocalization but not the phonological store (judgment of phonemes), or neither subvocalization nor the phonological store (judgment of homo-

⁴ Interestingly, other composers—such as Franz Liszt and Nikolai Rimsky-Korsakov (Pearce, 2007), Olivier Messiaen (Bernard, 1986), and Alexander Scriabin (Peacock, 1985)—reported experiences of color in response to auditory stimulation, and individuals who experience visual images in response to music are more likely to play a musical instrument than are individuals with other types of synesthesia (Ward, Thompson-Lake, Ely, & Kaminski, 2008).

⁵ Although the term *musical hallucinosis* suggests that auditory imagery in musical hallucinosis is hallucinatory (i.e., thought to reflect an actually present external stimulus), many patients with chronic musical hallucinosis report that their imagery is not hallucinatory, and those patients “attributed the experience to a problem with the brain or the ears” (Griffiths, 2000, p. 2066). Even so, the involuntary nature of imagery in musical hallucinosis is more typical of hallucination than of normative imagery, as hallucinations are generally considered to be under less voluntary control, and normative imagery is generally considered to be under more voluntary control.

phones). Performance of patients who were more prone to auditory hallucinations did not differ from performance of patients who were less prone to auditory hallucinations on any of the tasks, and Evans et al. suggested that it was unlikely that inner speech and auditory hallucinations were connected in a direct or simple way.

A related possible account is that auditory hallucinations in schizophrenia result from decreased efficiency of the inner voice or the inner ear. McGuire, Silbersweig, Wright, et al. (1996) compared PET of patients with schizophrenia who were more prone to auditory hallucinations, patients with schizophrenia who were less prone to auditory hallucinations, and control participants. When participants imaged their own voice, there were no differences between groups, but when participants imaged sentences spoken in another person's voice, patients who were more prone to auditory hallucinations exhibited reduced activity in the left middle temporal gyrus and rostral supplementary motor area (see also McGuire et al., 1995). Shergill, Bullmore, Simmons, Murray, and McGuire (2000) compared fMRI of patients with schizophrenia who were prone to verbal hallucinations with control participants. When participants imaged their own voice, there were no differences between groups, but when participants imaged verbal stimuli in another person's voice, patients exhibited reduced activity in the posterior cerebellar cortex, hippocampi, bilateral lenticular nuclei, right thalamus, middle and superior temporal cortex, and left nucleus accumbens. Findings of McGuire, Silbersweig, Wright, et al.; McGuire et al. (1995); and Shergill et al. (2000) are consistent with the hypothesis that verbal hallucinations in schizophrenia are linked with failures to activate brain areas associated with monitoring of inner speech. On the basis of fMRI acquired from nonpatient participants, Shergill et al. (2001) suggested that such areas might involve supplementary motor cortex and the cerebellum.

A potential issue is that voluntary auditory images experienced by patients with schizophrenia (e.g., as in McGuire et al., 1995; McGuire, Silbersweig, Wright, et al., 1996; Shergill et al., 2000) might not necessarily involve the same mechanisms as do involuntary auditory hallucinations experienced by such patients; to the extent that voluntary auditory imagery and involuntary auditory hallucinations involve similar characteristics (e.g., pitch, timbre), overlap in mechanisms might be expected, but to the extent that voluntary auditory imagery and involuntary auditory hallucinations do not involve similar characteristics (e.g., extent of voluntary control over presence and content of the experience), overlap in mechanisms might not be expected. This issue might be addressed by comparing brain imaging of spontaneous auditory hallucinations in patients with schizophrenia with brain imaging of voluntary auditory imagery of the same content in the same patients and with brain imaging of voluntary auditory imagery of the same content in nonpatient participants. Brain areas exhibiting similar and significant activation for all three conditions would address general mechanisms of auditory imagery. Brain areas exhibiting similar and significant activation for auditory hallucinations and for voluntary auditory imagery in patients with schizophrenia but not for voluntary auditory imagery in nonpatient participants would address characteristics of schizophrenic imagery. Finally, comparison of areas of significant brain activation in voluntary auditory imagery and in involuntary auditory hallucinations in patients with schizophrenia would address hallucinatory aspects of auditory imagery in schizophrenia.

Amusia. An acquired disorder of music perception, performance, reading, or writing that is not attributable to a disruption of basic perceptual, motoric, or cognitive functions is referred to as an *amusia* (for review, see Brust, 2003; Marin & Perry, 1999). It is well documented in the clinical literature that disorders of visual perception are often mirrored by parallel disorders of visual imagery (e.g., Bisiach & Luzzatti, 1978; DeVreese, 1991; for review, see Farah, 1988; Kosslyn, 1994), and so it could be predicted that disorders of auditory perception should be mirrored by parallel disorders of auditory imagery. However, within the amusia literature, there are few reports of disorders of auditory imagery or of comparisons of auditory imagery and auditory perception. The standard screening battery for amusia, The Montreal Battery of Evaluation of Amusia (Peretz, Champod, & Hyde, 2003), does not explicitly test for auditory imagery per se, although it does test memory for several musical qualities that could be represented or encoded by auditory imagery (e.g., interval size, rhythm). A relationship between amusia and defects in the spatial component of visual imagery has been reported (patients with amusia perform more poorly on a mental rotation task than do nonpatient controls; Douglas & Bilkey, 2007). Evidence has also emerged for a *congenital amusia* (colloquially referred to as *tone deafness*), which has been suggested to result from a heritable defect in fine-grained pitch processing (Peretz & Hyde, 2003); studies of the effects of congenital amusia on auditory imagery have not yet been reported.

In a study discussed previously, Zatorre and Halpern (1993) visually presented pairs of lyrics of familiar melodies, and participants judged whether the pitch of the second lyric was higher or lower than was the pitch of the first lyric. Patients with a right temporal lobe lesion performed more poorly in imagery and in perception than did patients with a left temporal lobe lesion or control participants (see also Halpern & Zatorre, 1999). In a complementary investigation to this study (discussed in Halpern, 2003), temporary functional lesions (i.e., temporary amusias) were induced by application of TMS. Participants were asked whether the second note of a verbal melody (i.e., with lyrics) or a nonverbal melody (i.e., without lyrics) was higher or lower in pitch than was the first note of that melody. TMS was applied to supplementary motor cortex, left auditory cortex, right auditory cortex, or visual cortex (visual cortex served as a control condition). The application of TMS had no effect on performance when TMS was applied to visual cortex, supplementary motor cortex, or left auditory cortex, but application of TMS to the right auditory cortex decreased performance. This pattern suggests that the right auditory cortex is involved in auditory imagery, and this is consistent with poorer performance of patients with a right temporal lobe lesion in Zatorre and Halpern's study. Also, application of TMS slowed response times in the imagery task for nonverbal melodies but did not influence response times for verbal melodies.

Part 6: Integration and Implication

In the preceding discussion, a wide range of findings and claims regarding properties and characteristics of auditory imagery were considered. In addition to the specific questions asked and conclusions drawn in the individual studies discussed in Parts 1, 2, 3, 4, and 5, a number of more general questions regarding properties and characteristics of auditory imagery that draw on results across

multiple studies can be asked. These more general questions are discussed in Part 6.

Does Auditory Imagery Preserve Structural Properties of Auditory Stimuli?

Several studies are consistent with the hypothesis that auditory imagery preserves structural properties of auditory stimuli, including pitch distance (Intons-Peterson et al., 1992), loudness distance (Intons-Peterson, 1980), absolute pitch of the starting tone of a melody (Halpern, 1989, 1992), timbre (Halpern et al., 2004), musical contour (Weber & Brown, 1986), melody (Zatorre et al., 2009), intervening beats in a musical stimulus (Halpern, 1988a; Halpern & Zatorre, 1999), tempo of music (Halpern, 1988b), and tempo of speech (Abramson & Goldinger, 1997). Additionally, auditory imagery can prime a subsequent percept on the basis of harmonic relationships (Hubbard & Stoeckig, 1988), timbre (Crowder, 1989), and categories of words and environmental sounds (Stuart & Jones, 1996), and this is consistent with preservation within auditory imagery of the structures and relationships in those domains. However, other studies are not consistent with the hypothesis that auditory imagery preserves structural properties of auditory stimuli. Participants have exhibited greater difficulty in detecting embedded melodies (Brodsky et al., 2003) or alternative interpretations of an ambiguous stimulus (Reisberg et al., 1989) in auditory imagery than in auditory perception, and participants have exhibited decreased accuracy in pitch comparison tasks in auditory imagery relative to auditory perception (Zatorre & Halpern, 1993; Zatorre et al., 1996). Imaged loudness does not prime a subsequent percept (Pitt & Crowder, 1992), and coupled with the apparent absence of loudness as a necessary part of auditory imagery (e.g., Intons-Peterson, 1980), this suggests structural properties regarding loudness are not part of the basic architecture of auditory imagery.

Consideration of which structural properties appear to be preserved in auditory imagery and which structural properties do not appear to be preserved in auditory imagery does not reveal a clear principle for determining which properties are preserved and which properties are not preserved. One possible account is that structure regarding relatively simple information (e.g., pitch) is preserved but that structure regarding relatively complex information (e.g., an embedded melody) is not preserved. A second possible account is that structure regarding isolable or separable information (e.g., pitch) is preserved but that structure regarding integrated information (e.g., an embedded melody) is not preserved separately from the larger structure within which that information is integrated. However, failure of auditory imagery to necessarily incorporate loudness does not appear consistent with these two accounts. A third possible account is that complete structural information is preserved within the image, but that such information is weaker or more susceptible to interference for some stimulus qualities, types, or dimensions than for other stimulus qualities, types, or dimensions. Overall, auditory imagery appears to preserve most, but not all, of the structural properties of auditory stimuli. Rather than considering whether auditory imagery preserves structural properties of auditory stimuli in an all-or-none fashion, future research should examine (a) whether additional structural properties are preserved and (b) how the experimental context or task influences whether a given structural property is

present or not within the image; consideration of a larger sample of properties, contexts, or tasks might clarify principles regarding which structural properties are preserved in auditory imagery.

Does Auditory Imagery Preserve Temporal Properties of Auditory Stimuli?

Evidence from several studies is consistent with the hypothesis that auditory imagery preserves temporal properties of auditory stimuli. More time is required to transform an imaged pitch a greater (pitch) distance (Intons-Peterson et al., 1992), and more time is required to transform the subjective loudness level of one image to match the subjective loudness level of a second image with increases in the difference between the initial subjective loudness levels (Intons-Peterson, 1980). Similarly, more time is required to scan across more beats in an imaged melody (Halpern, 1988a; Halpern & Zatorre, 1999) or count more syllables in an imaged letter string (Aziz-Zadeh et al., 2005). An image of a particular melody appears to specify a consistent tempo similar to the tempo at which that melody is usually perceived (Halpern, 1988b). More time is required to respond to visually presented words containing long vowel sounds than to respond to visually presented words containing short vowel sounds (Abramson & Goldinger, 1977). Even so, relatively weaker temporal acuity in auditory imagery than in auditory perception (Janata & Paroo, 2006) suggests that not all temporal properties (or dimensions along which temporal properties are expressed) in auditory imagery are equally preserved or equally easy to transform. The time to generate an auditory image does not appear related to subjective loudness (Intons-Peterson, 1980), but this objection to the claim that auditory imagery preserves temporal properties of auditory stimuli would only apply if loudness is assumed to be generated incrementally. Overall, auditory imagery appears to generally preserve temporal properties of auditory stimuli.

Does Auditory Imagery Interfere With or Facilitate Auditory Perception?

Detection of a faint auditory signal is decreased if participants are instructed to be simultaneously generating auditory imagery (Okada & Matsuoka, 1992; Segal & Fusella, 1970), and this is consistent with the hypothesis that auditory imagery can interfere with auditory perception. Such interference might reflect limited capacity or limited processing resources (as suggested by Antrobus et al., 1970). However, judgments of whether pitch (Farah & Smith, 1983; Hubbard & Stoeckig, 1988) or timbre (Crowder, 1992) of an imaged tone match pitch or timbre of a subsequently perceived tone are faster and more accurate if pitches or timbres of the image and the percept match, and this is consistent with the hypothesis that auditory imagery can facilitate auditory perception. Such facilitation might reflect priming of a representation common to imagery and perception (as suggested by Hubbard & Stoeckig, 1988) or a general increase in alertness separate from effects of imaging a specific pitch (as suggested by Sullivan et al., 1996) or a specific timbre. The notion that images involve expectation (e.g., Janata, 2001; Janata & Paroo, 2006) suggests that imagery should facilitate perception when the stimulus to be perceived matches the expectation (e.g., priming in Crowder, 1989; Farah & Smith, 1983) and should interfere with percep-

tion when the stimulus to be perceived does not match the expectation (e.g., as hypothesized for incompatible pairings in P. E. Keller & Koch, 2008); however, even when a stimulus to be perceived matches the content of the image exactly (and the match between expectation and percept strongest), interference can occur (e.g., Okada & Matsuoka, 1992).

There are several possible accounts of why auditory imagery appears to sometimes interfere with auditory perception and sometimes facilitate auditory perception. One possible account is that facilitation occurs up to that point at which processing capacity or resources are exceeded, and then interference occurs. However, either interference (e.g., Okada & Matsuoka, 1992) or facilitation (Farah & Smith, 1983) can be found when participants image just a single stimulus that presumably would not exceed processing capacity or resources. A second possible account is that facilitation occurs when generation of the auditory image precedes the auditory stimulus, and interference occurs when generation of the auditory image is concurrent with the auditory stimulus. However, either interference (e.g., Okada & Matsuoka, 1992) or facilitation (Farah & Smith, 1983) can be found when generation of the auditory image precedes the auditory stimulus. A third possible account is that facilitation occurs when the content of the auditory image and the auditory percept match, and interference occurs when the content of the auditory image and the auditory percept do not match, but interference can occur when the auditory stimulus exactly matches the auditory percept (Okada & Matsuoka, 1992). A fourth possible account (consistent with suggestions of Finke, 1986, regarding the interaction of visual imagery and visual perception) is that auditory imagery interferes with auditory detection but facilitates auditory discrimination or identification.

An account based on the notion that auditory imagery interferes with detection of an auditory stimulus but facilitates discrimination or identification of an auditory stimulus appears consistent with the widest range of data, as studies in which participants detect an auditory stimulus generally provide evidence of interference (e.g., Okada & Matsuoka, 1992; Segal & Fusella, 1970), whereas studies in which participants discriminate between auditory stimuli or identify an auditory stimulus generally provide evidence of facilitation (e.g., Farah & Smith, 1983; Hubbard & Stoeckig, 1988). Such a difference might occur if detection involves different cognitive resources or levels of processing than does discrimination or identification. Detection of a stimulus involves deployment of limited attentional resources, and if some of those resources are already occupied in generating or maintaining an image, then fewer attentional resources would be available for detecting a stimulus. Effects of a decrease in attentional resources would be especially noticeable if the stimulus to be detected was faint. Also, if at least some attentional resources are modality-specific, this could account for differences in effects of same-modality and different-modality imagery on detection. Discrimination or identification of an already detected stimulus would involve comparing or matching stimulus information against large amounts of data in memory. Imagery of an appropriate stimulus could lower the threshold of the relevant representation in memory so that the representation could be more easily triggered by the detected information.

Does Auditory Imagery Involve the Same Brain Areas as Auditory Perception?

Behavioral evidence involving priming (Crowder, 1989; Farah & Smith, 1983; Hubbard & Stoeckig, 1988), similarity ratings of timbre (Halpern et al., 2004), and detection of embedded melodies in the presence of auditory distractors (Brodsky et al., 2008), as well as clinical studies of musical hallucinosis (Kasai et al., 1999; Shinosaki et al., 2003) and schizophrenia (Lennox et al., 2000), is consistent with the hypothesis that auditory imagery involves brain areas involved in auditory perception. Patients with a right temporal lobe lesion perform more poorly on a pitch comparison task in imagery and in perception than do patients with a left temporal lobe lesion or control participants (Zatorre & Halpern, 1993). Similar decreases in performance following application of TMS to the right temporal lobe but not to the left temporal lobe or occipital lobe occur in nonpatient participants (Halpern, 2003). The superior temporal gyrus, frontal and parietal lobes, and supplementary motor cortex are activated in pitch (Halpern & Zatorre, 1999; Zatorre et al., 1996) and timbre (Halpern et al., 2004) comparisons in imagery and in perception. The planum temporale is activated by instructions to form an auditory image or by auditory perception of environmental sounds (Bunzeck et al., 2005), and activation level of this area during reported auditory imagery correlated with ratings of imagery vividness (Zatorre et al., 2009). Auditory and premotor cortical areas might be activated in notational audiation (Schurmann et al., 2002). Findings that application of TMS to the left hemisphere disrupts covert speech and overt speech (Aziz-Zadeh et al., 2005), silent articulation and speech imagery activate left inferior frontal gyrus (McGuire, Silbersweig, Murray, et al., 1996), and speech imagery activates Broca's area (Bookheimer, 2002; Hinke et al., 1993) are also consistent with the hypothesis that auditory imagery involves brain areas involved in auditory perception.

An issue with many studies that report brain imaging and that make claims regarding auditory imagery is that behavioral evidence that auditory imagery was generated is often not reported (e.g., as in Bunzeck et al., 2005; Kraemer et al., 2005; McGuire, Silbersweig, Murray, et al., 1996; Meyer et al., 2007; Schurmann et al., 2002; Shergill et al., 2000, 2001; Wu et al., 2006; Yoo et al., 2001). Imagery is assumed or suggested to have occurred because (a) participants were instructed to generate images or (b) imagery offers a plausible explanation for patterns in the data. However, without behavioral evidence that imagery was generated or used, conclusions regarding brain areas involved in imagery cannot be drawn. As Zatorre and Halpern (2005, p. 9) noted, "merely placing subjects in a scanner and asking them to image . . . simply will not do, because one will have no evidence that the desired mental activity was actually taking place." Claims regarding brain areas involved in imagery should present behavioral evidence that imagery was actually generated or used as well as presenting brain imaging or other physiological measures (e.g., as in Halpern et al., 2004; Zatorre et al., 1996). Also, it should be noted that a lack of behavioral evidence regarding auditory imagery is not limited to brain imaging studies, as some behavioral studies make claims regarding auditory imagery without presenting evidence that auditory imagery occurred or was used in the experimental task (e.g., P. E. Keller & Koch, 2008; Sharps & Pollitt, 1998). Examples of behavioral measures that could be used to support claims that

imagery was generated include priming (e.g., Crowder, 1989; Hubbard & Stoeckig, 1988), interference on behavioral tasks involving specific processes or types of stimuli (e.g., Aleman & van't Wout, 2004; Brodsky et al., 2008), and measurement of imaged qualities (e.g., Halpern, 1992; Intons-Peterson et al., 1992).

Answers to earlier questions regarding the extent to which auditory imagery preserves structural or temporal properties of an auditory stimulus, and whether auditory imagery interferes with or facilitates perception of an auditory stimulus, might be directly related to the extent to which auditory imagery involves the same brain areas involved in auditory perception. Even so, the full extent of brain areas involved in auditory imagery cannot be identical to the full extent of brain areas involved in auditory perception, as that would make it difficult for persons to distinguish between their auditory imagery and their auditory perception (i.e., all auditory images would be hallucinatory). Consistent with this, differences between brain areas activated in auditory perception and brain areas activated when participants are instructed or assumed to be generating auditory imagery have been found (e.g., primary auditory cortex is activated by auditory perception but not activated [or activated as strongly] by instructions to generate or use auditory imagery; Bunzeck et al., 2005; Halpern et al., 2004; Yoo et al., 2001; Zatorre & Halpern, 2005). Observations that auditory imagery can usually be distinguished from auditory perception despite occasional auditory hallucinations (e.g., Griffiths, 2000) or failures of reality/source monitoring (e.g., Johnson et al., 1988), coupled with substantial but not complete overlap in brain areas suggested by behavioral, clinical, and brain imaging data, suggest that auditory imagery involves many, but not all, of the brain areas involved in auditory perception. Interestingly, such a conclusion is consistent with conclusions that visual imagery involves many, but not all, of the brain areas involved in visual perception (e.g., Kosslyn & Thompson, 2000, 2003).

Does Auditory Imagery Involve Subvocalization?

Crowder and Pitt (1992) manipulated timbre in studies of auditory imagery in an attempt to minimize the motor component of auditory imagery, and they suggested that subvocalization was not involved in imagery of timbres that could not be produced by a human vocal tract. Baddeley and Logie (1992; Baddeley & Andrade, 2000) challenged this suggestion and proposed that the phonological loop could rehearse and maintain nonvocal timbres. The articulatory suppression effect and irrelevant speech effect (Baddeley & Logie, 1992), decreases in performance in auditory imagery tasks when subvocalization is blocked (Aleman & van't Wout, 2004; Brodsky et al., 2003, 2008; Reisberg et al., 1989; Smith et al., 1995), dynamic activity of muscles surrounding the vocal folds during reading of music notation (Brodsky et al., 2008), activation in supplementary motor area during comparisons of pitch (Halpern & Zatorre, 1999; Zatorre et al., 1996) or timbre (Halpern et al., 2004), increases in activation of Broca's area during speech imagery (Bookheimer, 2002; Hinke et al., 1993), disruption of covert speech by application of TMS to the left hemisphere (Aziz-Zadeh et al., 2005), and activation of brain areas implicated in the phonological loop by auditory imagery (Aleman et al., 2005) are consistent with the hypothesis that auditory imagery can involve subvocalization. However, differential effects of an interpolated task on recall of words or of sounds (Tinti et al.,

1997) and lack of an effect of subvocalization when judging homophones (Smith et al., 1995) suggest that auditory imagery does not necessarily involve subvocalization. Also, it is possible that apparent subvocalization of nonvocal timbres might involve vocal approximations of nonvocal stimuli (e.g., humming pitches of a melody) rather than representation of nonvocal timbres per se.

Is There a Separate Inner Voice and Inner Ear?

Findings from several behavioral studies of auditory imagery are consistent with a distinction between the inner voice and the inner ear. Silent reading time is longer for words containing long vowel sounds than for words containing short vowel sounds (Abramson & Goldinger, 1977), but silent reading time is not influenced by whether the passage being read is attributed to an author with a slow speaking rate or a fast speaking rate (Alexander & Nygaard, 2008; Kosslyn & Matt, 1977). The former difference might reflect characteristics of the inner voice (i.e., auditory and articulatory information), and the latter lack of difference might reflect characteristics of the inner ear (i.e., auditory information). Subvocalization influences some types of judgments involving auditory imagery (e.g., reinterpreting ambiguous images) but not other types of judgments involving auditory imagery (e.g., whether letter strings when pronounced would sound like actual words), and this might reflect different relative contributions of the inner voice and the inner ear (Smith et al., 1992, 1995). However, although these differences are consistent with a separation of the inner ear and the inner voice, the data do not demonstrate that such a difference must necessarily exist. It is possible that differences in reading times might reflect differences in the material being processed (e.g., difficulty of the text) or strategy (e.g., intentional or unintentional mimicking of speaking rate) and not reflect differences in the structures or mechanisms processing that material.

Several claims involving brain areas involved in auditory imagery are consistent with a distinction between the inner voice and the inner ear. Auditory imagery of one's own voice (involving the inner voice) and auditory imagery of another person's voice (involving the inner ear) activate inferior frontal gyrus, and auditory imagery of another person's voice also activates left premotor cortex, supplementary motor cortex, and left temporal cortex (McGuire, Silbersweig, Murray, et al., 1996). Auditory imagery of another person's voice results in greater activation of supplementary motor cortex, left precentral and middle temporal gyri, inferior parietal cortex, and right superior temporal gyri and cerebellar cortex (Shergill et al., 2001). Patients with schizophrenia who are more prone to auditory hallucinations exhibit less activity in the left middle temporal gyrus or rostral supplementary motor cortex (McGuire, Silbersweig, Wright, et al., 1996) and posterior cerebellar cortex, bilateral lenticular nuclei, middle and superior temporal cortex, and left nucleus accumbens (Shergill et al., 2000) when imaging another person's voice than when imaging their own voice. However, behavioral evidence that imagery occurred was not reported in any of these studies. Also, even if appropriate imagery was being generated in these studies, there is a confound involving greater articulatory information in imagery of one's own voice (but see below) and a second confound in that the timbre of one's own voice is usually different from the timbre of someone else's voice. The extent to which differences in brain activity in imaging one's own voice or in imaging another person's voice

were due to differences in source (referent) of the voice rather than to differences in amount of articulatory information present or to differences in timbre is not clear.

Although behavioral and brain imaging data are consistent with a distinction between the inner voice and the inner ear, it is not clear whether such a distinction necessarily reflects differences in mechanisms or structures. Perception of another person's actions has been proposed to involve activation of one's own motor or action plans (e.g., Knoblich & Sebanz, 2006; Liberman & Mattingly, 1985; Wilson & Knoblich, 2005). To the extent that auditory imagery involves the same mechanisms as auditory perception, such a motor view of perception suggests articulatory information should be activated for the inner voice and for the inner ear. Thus, even though an auditory image of another person's voice is considered to involve the inner ear rather than the inner voice (e.g., McGuire, Silbersweig, Murray, et al., 1996; Shergill et al., 2000), it is possible that an auditory image of another person's voice (involving the inner ear) might include articulatory information (cf. activation of supplementary motor cortex when imaging another person's voice; McGuire, Silbersweig, Murray, et al., 1996; McGuire et al., 1995; Shergill et al., 2001). Similarly, if subvocalization can be used to rehearse nonvocal timbres (cf. Baddeley & Logie, 1992; Halpern et al., 2004), then articulatory information might not be limited to the inner voice. If auditory imagery previously attributed to the inner ear can involve articulatory information, and if auditory imagery attributed to the inner voice does not necessarily involve articulatory information, then the primary difference between the inner voice and inner ear is reduced or even eliminated. Although useful as a heuristic, it is not clear that the distinction between the inner voice and the inner ear is necessarily valid; perhaps data consistent with this distinction reflect differences in content rather than differences in mechanisms or structures.

Are Auditory Images Interpreted?

The idea of "hearing in the mind's ear" suggests that auditory imagery involves a relatively uninterpreted sensory copy of an auditory stimulus. Such a sensory copy might be interpreted in different ways at different times, just as a perceptually ambiguous stimulus might be interpreted in different ways at different times. However, findings that participants have greater difficulty detecting an embedded melody (Brodsky et al., 2003) or an alternative interpretation of a string of speech sounds (Reisberg et al., 1989) in auditory imagery than in auditory perception suggest that an auditory image cannot be easily reinterpreted once that image has been generated. This further suggests that an auditory image has undergone significant interpretation by the time that image is generated. One possible type of interpretation involves creation of a reference frame; indeed, Reisberg and Heuer (2005) suggested that the inability to reinterpret visual images of visually perceptually ambiguous figures might result from the presence of such reference frames in visual imagery. In the case of visual imagery, a reference frame would not specify what the image is an image of (i.e., would not specify the referent), but it would specify how the content of the image is to be understood (e.g., figure-ground relationships, how the form is parsed, the top and bottom). By analogy, a reference frame for an auditory image might specify how the content of the auditory image is to be parsed, thus

rendering it more difficult to detect an embedded melody or to reinterpret a string of speech sounds in that image. Regardless, auditory images do not appear to be uninterpreted sensory copies played back before the "mind's ear" but rather appear to be interpreted experiences that reflect considerable prior processing.⁶

Are Auditory Images Depictive?

To the extent that auditory images are based on interpreted information, such images might be considered descriptive rather than depictive (in the senses suggested by Kosslyn, 1980). However, Reisberg and Heuer (2005) argued that visual imagery is at least partly depictive because the content of a visual image "looks like" what is being represented, and an analogous argument can be made that auditory imagery is at least partly depictive because the content of an auditory image "sounds like" what is being represented. Also, just as preservation of spatial relationships in visual imagery has been used to argue that visual images are depictive (e.g., more time is required to scan a longer imaged distance; Kosslyn, 1980), preservation of temporal relationships in auditory imagery (e.g., Aziz-Zadeh et al., 2005; Halpern, 1988b; Intons-Peterson, 1980; Intons-Peterson et al., 1992) can be used to argue that auditory images are depictive. However, if auditory imagery was completely depictive, then basic features such as pitch, loudness, and timbre would necessarily be specified in an auditory image (e.g., an image of a given pitch would have to specify loudness and timbre of that pitch; Hubbard & Stoeckig, 1992), but some basic features are not necessarily specified in auditory imagery (e.g., loudness; Intons-Peterson, 1980; Pitt & Crowder, 1992). Also, and as noted above, participants have difficulties detecting embedded melodies (Brodsky et al., 2003) or alternative interpretations of a string of speech sounds (Reisberg et al., 1989) in auditory imagery, and this suggests that auditory images are at least partly descriptive. Thus, auditory imagery appears to involve both depictive components and descriptive components.

Does Auditory Imagery Involve Expectancies?

The findings that participants report hearing a continuation of music in auditory imagery during a silent gap in a familiar piece of music (Kraemer et al., 2005), report auditory imagery of an anticipated familiar piece of music during a period of silence prior to presentation of that music (Leaver et al., 2009), and exhibit emitted potentials when instructed to form an auditory image or when an expected auditory stimulus is not presented (Janata, 2001; Meyer et al., 2007) are consistent with the hypothesis that imagery involves expectation. Such expectations are consistent with Neisser's (1976) notion that imagery arises (at least in part) from schemata: Processing of perceptual input would activate a schema, and when an expected stimulus was not presented, the schema would "fill in" the silent gap. If the gap is prolonged or the image not attended, activation of the schema would decay and filling in cease (cf. the decrease in emitted potentials for subsequent imaged notes in Janata, 2001). Such filling in would be more likely or

⁶ It would not be surprising that considerable processing and interpretation preceded creation of an image, as most theorists believe that considerable processing and interpretation are involved in perception (from which elements of imagery are drawn).

more complete with familiar pieces than with unfamiliar pieces, as schemata for familiar pieces are more detailed or elaborated (see Deliege, Melen, Stammers, & Cross, 1996; Dowling & Harwood, 1986; Krumhansl, 1990; Narmour, 1990). The hypothesis that auditory images prime compatible responses in action planning (P. E. Keller & Koch, 2008), and the finding that auditory images prime judgments of chords as a function of harmonic relatedness (Hubbard & Stoeckig, 1988), are consistent with a view of imagery as involving expectation. However, auditory imagery can interfere with perceptual detection even when content of the image matches content of the percept (and expectation should be strongest; Okada & Matsuoka, 1992), and this suggests that effects of expectation are task-specific rather than generally facilitating (cf. Sullivan et al., 1996).

Is Auditory Imagery Obligatory?

The finding that participants report hearing a continuation of music in auditory imagery during a silent gap in a familiar piece of music has been hypothesized to demonstrate that auditory imagery is obligatory (Kraemer et al., 2005; but see Zatorre & Halpern, 2005). The strongest sense of “obligatory” suggests an involuntary experience of auditory imagery regardless of other cognitive activity, and although this can occur in musical hallucinosis and in schizophrenia, it is not typical of normative auditory imagery. A weaker sense of “obligatory” involves specific perceptual stimuli or schemata automatically evoking a specific auditory image, but in the absence of such specific stimuli or schemata, that auditory image would not be evoked. However, findings that loudness information was not necessarily present in auditory imagery (Intons-Peterson, 1980; Pitt & Crowder, 1992), phonological components of speech imagery were not necessarily activated during silent reading (Alexander & Nygaard, 2008; Kosslyn & Matt, 1977), and auditory imagery of the sound an animal would make was not reported following visual presentation of a picture of that animal unless participants had been instructed to form such an auditory image (Wu et al., 2006) are consistent with the hypothesis that auditory imagery might not be obligatory in even this weaker sense. Even if an auditory image is evoked, certain features might not be automatically generated (e.g., loudness), and in this case, at least some aspects of an auditory image would not be obligatory. Rather than considering whether auditory images are obligatory in an all-or-none fashion, future research should examine which aspects of auditory imagery are automatically evoked by which stimuli or under which conditions.

Is Auditory Imagery Related to Rehearsal?

Increased memory for an auditory recording of a stimulus relative to a verbal label of that stimulus (Sharps & Pollitt, 1998; Sharps & Price, 1992) and for interactive auditory images relative to verbal elaboration (Tinti et al., 1997) is consistent with the notion that auditory imagery might provide a concrete sensory code in addition to any abstract verbal code that might be available. In this regard, auditory imagery might be useful as a mnemonic. Indeed, for auditory stimuli for which there is not an adequate verbal description, a verbal code might not be available, and an imagery code might be the most useful or even the only encoding format available (cf. Hubbard, 1996). The effectiveness

of melody as an aid in initial learning (e.g., Dickson & Grant, 2003; VanVoorhis, 2002) and in relearning (e.g., Rainey & Larsen, 2002) suggests that auditory imagery has mnemonic properties and can function as a mnemonic. Use of auditory imagery as a mnemonic is consistent with dual-coding accounts of memory (Paivio, 1971, 1986); indeed, auditory imagery might function as a third code in addition to the visual codes and verbal codes that have been previously and more extensively studied (e.g., the melodic code suggested by Samson & Zatorre, 1991). Also, to the extent that auditory imagery involves or evokes schemata or other knowledge, that imagery might direct subsequent processing (encoding) and thus influence the likelihood that specific elements of the environment might be selected for further or deeper processing. To the extent that auditory imagery can improve performance in encoding, storage, or retrieval, auditory imagery would function as a mnemonic.

The possibility that auditory imagery can function as a mnemonic suggests that auditory imagery is related to rehearsal (cf. T. A. Keller et al., 1995). Findings that subvocalization is involved in some auditory imagery tasks (e.g., Smith et al., 1995), coupled with the notion that there is a close relationship between working memory and articulatory coding (see Smith et al., 1992), suggest that subvocalization is common to both auditory imagery and rehearsal. Indeed, Smith et al. (1992, p. 108) suggested that “memory rehearsal requires both the inner ear and the inner voice,” and MacKay (1992, p. 142) suggested that “rehearsal and its effects on memory may represent one of the main functions of internal speech.” Such notions imply a close correspondence between auditory imagery and rehearsal. However, although evidence is relatively strong for involvement of the phonological loop in temporary storage of at least some types of auditory imagery, evidence is much less strong for involvement of the phonological loop in generation of auditory imagery (Baddeley & Logie, 1992). Also, learning (and presumably rehearsal) can occur without involvement of subvocalization or the phonological store (e.g., learning in nonhuman animals) and in nonauditory domains (e.g., visual or olfactory distinctions that cannot be verbally labeled by the learner), and as noted earlier, some auditory imagery tasks do not involve subvocalization or the phonological store (e.g., judgment of homophones). Rehearsal can involve auditory imagery but also involves additional mechanisms, and auditory imagery can involve mechanisms used for rehearsal but also involves additional mechanisms; thus, auditory imagery and rehearsal involve overlapping but distinct processes.

Is Auditory Imagery Related to Musical Ability and Experience?

Despite Seashore’s (1938/1967) provocative claim that the most outstanding mark of a musical mind is a high capacity for auditory imagery, there has been relatively little empirical research on the relationship of auditory imagery to musical ability and experience. Participants with more musical training perform better when visually presented with two lyrics of a familiar melody and asked which lyric would normally be sung at a higher pitch (Aleman et al., 2000), exhibit better pitch acuity and temporal acuity in auditory imagery (Janata & Paroo, 2006), exhibit increased flexibility in representing tempo in auditory imagery (Halpern, 1992), perform better when asked to identify the most different sound of a

group of sounds (Aleman et al., 2000), and have been hypothesized to possess greater vividness of auditory imagery (P. E. Keller & Koch, 2008) than do participants with little or no musical training. Auditory imagery has been suggested to contribute to increased ability in sight-reading (Kornicke, 1995; although see Kopiez et al., 2006), musical performance (Highben & Palmer, 2004), and musical note identification (Kalakoski, 2007). Auditory imagery is evoked during notational audiation by trained musicians (Brodsky et al., 2008; although see Wöllner et al., 2003), and there are anecdotal reports that auditory imagery was used in composition by several noted composers. However, it is not clear whether differences in musical ability and experience result from differences in auditory imagery or whether differences in auditory imagery result from differences in musical ability and experience. More research is needed before general conclusions regarding the relationship of auditory imagery with musical ability and experience can be drawn.

Part 7: Summary and Conclusions

Auditory imagery involves a perceptual-like experience of an auditory stimulus in the absence of that stimulus. Auditory imagery preserves many structural and temporal properties of auditory stimuli, and generation of auditory imagery appears to involve activation of many brain areas involved in perception of auditory stimuli. Auditory imagery appears to interfere with auditory perceptual detection but facilitate auditory perceptual discrimination or identification, and auditory imagery can function as a mnemonic. Some tasks involving auditory imagery appear to involve subvocalization, and this suggests that the phonological loop of working memory might provide a basis for at least some auditory imagery. An inability to hear an alternative interpretation of a perceptually ambiguous stimulus within an auditory image, or to detect an embedded melody within an auditory image, suggests auditory images reflect considerable interpretation and are not uninterpreted sensory experiences; rather, auditory images contain both depictive information and descriptive information. Priming involving imaged pitch, timbre, words, and common environmental sounds, as well as existence of emitted potentials during auditory imagery, suggests that auditory imagery involves expectancies. Auditory imagery is related to musical ability or experience, but whether differences in auditory imagery lead to differences in musical ability and experience or whether differences in musical ability and experience lead to differences in auditory imagery is not clear.

As demonstrated by the wide range of data surveyed here, auditory imagery is an ubiquitous phenomenon related to many different aspects of cognitive functioning. Auditory imagery is an important element of cognitive processing related to language, music, and other environmental stimuli, and abnormalities in cognitive processing related to auditory imagery are exhibited in or contribute to psychopathology (e.g., musical hallucinosis, schizophrenia). A variety of methodologies involving introspective reports, comparisons of performances and of brain activation patterns during behavioral tasks in which participants were instructed to generate an auditory image or perceived an auditory stimulus, and clinical data regarding various psychopathologies converge on a set of general conclusions regarding properties that characterize auditory imagery. The representation of features of auditory stim-

uli, such as pitch, timbre, and loudness, and the way those features are represented in images of more complex structures, such as language and music, appear highly systematic. There is as yet no general theory of auditory imagery (although suggestions regarding subvocalization and the phonological loop point in one possible direction), but empirical findings and conclusions discussed here provide an important step toward such a theory. In conjunction with data and theories addressing imagery in other modalities, perhaps such a theory of auditory imagery could contribute to more general theories of imagery and mental representation.

References

- Abramson, M., & Goldinger, S. D. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics*, *59*, 1059–1068.
- Aleman, A., Formisano, E., Koppenhagen, H., Hagoort, P., de Hann, E. H. F., & Kahn, R. S. (2005). The functional neuroanatomy of metrical stress evaluation of perceived and imaged spoken words. *Cerebral Cortex*, *15*, 221–228.
- Aleman, A., Nieuwenstein, M. R., Boecker, K. B. E., & de Hann, E. H. F. (2000). Music training and mental imagery ability. *Neuropsychologia*, *38*, 1664–1668.
- Aleman, A., & van't Wout, M. (2004). Subvocalization in auditory-verbal imagery: Just a form of motor imagery? *Cognitive Processing*, *5*, 228–231.
- Alexander, J. D., & Nygaard, L. C. (2008). Reading voices and hearing text: Talker-specific auditory imagery in reading. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 446–459.
- Allbutt, J., Ling, J., Heffernan, T. M., & Shafiuallah, M. (2008). Self-report imagery questionnaire scores and subtypes of social-desirable responding. *Journal of Individual Differences*, *29*, 181–188.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Anderson, R. E. (1982). Speech imagery is not always faster than visual imagery. *Memory & Cognition*, *10*, 371–380.
- Antrobus, J. S. (1968). Information theory and stimulus-independent thought. *British Journal of Psychology*, *59*, 423–430.
- Antrobus, J. S., Singer, J. L., Goldstein, S., & Fortgang, M. (1970). Mindwandering and cognitive structure. *Transactions of the New York Academy of Sciences*, *32*, 242–252.
- Aziz-Zadeh, L., Cattaneo, L., Rochat, M., & Rizzolatti, G. (2005). Covert speech arrest induced by rTMS over both motor and nonmotor left hemisphere frontal sites. *Journal of Cognitive Neuroscience*, *17*, 928–938.
- Baba, A., & Hamada, H. (1999). Musical hallucinations in schizophrenia. *Psychopathology*, *32*, 242–251.
- Baddeley, A. D., & Andrade, J. (2000). Working memory and the vividness of imagery. *Journal of Experimental Psychology: General*, *129*, 126–145.
- Baddeley, A. D., Lewis, V. J., & Vallar, G. (1984). Exploring the articulatory loop. *Quarterly Journal of Experimental Psychology*, *36*, 233–252.
- Baddeley, A. D., & Logie, R. H. (1992). Auditory imagery and working memory. In D. Reisberg (Ed.), *Auditory imagery* (pp. 179–197). Hillsdale, NJ: Erlbaum.
- Bailes, F. (2007). The prevalence and nature of imagined music in the everyday lives of music students. *Psychology of Music*, *35*, 555–570.
- Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 11, pp. 101–159). New York, NY: Academic Press.
- Baron-Cohen, S., & Harrison, J. (Eds.). (1997). *Synaesthesia: Classic and contemporary readings*. Oxford, England: Blackwell.

- Baron-Cohen, S., Harrison, J., Goldstein, L. H., & Wyke, M. (1993). Coloured speech perception: Is synaesthesia what happens when modularity breaks down? *Perception*, *22*, 419–426.
- Bentall, R. P., & Slade, P. D. (1985). Reality testing and auditory hallucinations: A signal detection analysis. *British Journal of Clinical Psychology*, *24*, 159–169.
- Bernard, J. W. (1986). Messiaen's synaesthesia: The correspondence between color and sound structure in his music. *Music Perception*, *4*, 41–68.
- Berti, S., Munzer, S., Schröger, E., & Pechmann, T. (2006). Different interference effects in musicians and a control group. *Experimental Psychology*, *53*, 111–116.
- Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 403–410.
- Bick, P. A., & Kinsbourne, M. (1987). Auditory hallucinations and subvocal speech in schizophrenic patients. *American Journal of Psychiatry*, *144*, 222–225.
- Bisiach, E., & Luzzatti, C. (1978). Unilateral neglect of representational space. *Cortex*, *14*, 129–133.
- Bookheimer, S. (2002). Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience*, *25*, 151–188.
- Bower, G. H. (1970a). Analysis of a mnemonic device. *American Scientist*, *58*, 496–510.
- Bower, G. H. (1970b). Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, *9*, 529–533.
- Brodsky, W., Henik, A., Rubenstein, B. S., & Zorman, M. (2003). Auditory imagery from musical notation in expert musicians. *Perception & Psychophysics*, *65*, 602–612.
- Brodsky, W., Kessler, Y., Rubenstein, B. S., Ginsborg, J., & Henik, A. (2008). The mental representation of music notation: Notational audiation. *Journal of Experimental Psychology: Human Perception and Performance*, *34*, 427–445.
- Brown, S. (2006). The perpetual music track: The phenomenon of constant musical imagery. *Journal of Consciousness Studies*, *13*, 25–44.
- Brust, J. C. M. (2003). Music and the neurologist: A historical perspective. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 181–191). New York, NY: Academic Press.
- Bugelski, B. R., Kidd, E., & Segman, J. (1968). Image as mediator in one-trial paired associate learning. *Journal of Experimental Psychology*, *76*, 69–73.
- Bunzeck, N., Wuestenberg, T., Lutz, K., Heinze, H. J., & Jancke, L. (2005). Scanning silence: Mental imagery of complex sounds. *NeuroImage*, *26*, 1119–1127.
- Cavanna, A. E., Trimble, M., Cinti, F., & Monaco, F. (2007). The “bicameral mind” 30 years on: A critical reappraisal of Julian Jaynes' hypothesis. *Functional Neurology*, *22*, 11–15.
- Chambers, D., & Reisberg, D. (1985). Can mental images be ambiguous? *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 317–328.
- Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55–81.
- Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*, 409–412.
- Collins, D. (2005). A synthesis process model of creative thinking in music composition. *Psychology of Music*, *33*, 193–216.
- Costa, A., Roelstraete, B., & Hartsuiker, R. J. (2006). The lexical bias effect in speech production: Evidence for feedback between lexical and sublexical levels across languages. *Psychonomic Bulletin & Review*, *13*, 972–977.
- Craver-Lemley, C., & Reeves, A. (1992). How visual imagery interferes with vision. *Psychological Review*, *99*, 633–649.
- Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 472–478.
- Crowder, R. G., & Pitt, M. A. (1992). Research on memory/imagery for musical timbre. In D. Reisberg (Ed.), *Auditory imagery* (pp. 29–44). Hillsdale, NJ: Erlbaum.
- Cytowic, R. E. (2002). *Synesthesia: A union of the senses* (2nd ed.). Cambridge, MA: MIT Press.
- Deliege, I., Melen, M., Stammers, D., & Cross, I. (1996). Musical schemata in real-time listening to a piece of music. *Music Perception*, *14*, 117–160.
- Deutsch, D. (1970, June 26). Tones and numbers: Specificity of interference in short-term memory. *Science*, *168*, 1604–1605.
- Deutsch, D. (1972, March 3). Mapping of interactions in the pitch memory store. *Science*, *175*, 1020–1022.
- Deutsch, D. (1974). Generality of interference by tonal stimuli in recognition memory for pitch. *Quarterly Journal of Experimental Psychology*, *26*, 229–234.
- DeVreese, L. P. (1991). Two systems for colour-naming defects: Verbal disconnection vs. colour imagery disorder. *Neuropsychologia*, *29*, 1–18.
- Dickson, D., & Grant, L. (2003). Physics karaoke: Why not? *Physics Education*, *38*, 320–323.
- Douglas, K. M., & Bilkey, D. K. (2007). Amusia is associated with deficits in spatial processing. *Nature Neuroscience*, *10*, 915–921.
- Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*, 341–354.
- Dowling, W. J., & Fujitani, D. S. (1971). Contour, interval, and pitch recognition in memory for melodies. *The Journal of the Acoustical Society of America*, *49*, 524–531.
- Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. New York, NY: Academic Press.
- Eardley, A. F., & Pring, L. (2006). Remembering the past and imaging the future: A role for nonvisual imagery in the everyday cognition of blind and sighted people. *Memory*, *14*, 925–936.
- Evans, C. L., McGuire, P. K., & David, A. S. (2000). Is auditory imagery defective in patients with auditory hallucinations? *Psychological Medicine*, *30*, 137–148.
- Farah, M. J. (1985). Psychophysical evidence for a shared representational memory for mental images and percepts. *Journal of Experimental Psychology: General*, *114*, 91–103.
- Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review*, *95*, 307–317.
- Farah, M. J., & Smith, A. F. (1983). Perceptual interference and facilitation with auditory imagery. *Perception & Psychophysics*, *33*, 475–478.
- Farthing, G. W. (1992). *The psychology of consciousness*. Englewood Cliffs, NJ: Prentice Hall.
- Finke, R. A. (1980). Levels of equivalence in imagery and perception. *Psychological Review*, *87*, 113–132.
- Finke, R. A. (1985). Theories relating mental imagery to perception. *Psychological Bulletin*, *98*, 236–259.
- Finke, R. A. (1986). Some consequences of visualization in pattern identification and detection. *American Journal of Psychology*, *99*, 257–274.
- Finke, R. A. (1989). *Principles of mental imagery*. Cambridge, MA: MIT Press.
- Finke, R. A., Pinker, S., & Farah, M. J. (1989). Reinterpreting visual patterns in mental imagery. *Cognitive Science*, *13*, 51–78.
- Finke, R. A., & Shepard, R. N. (1986). Visual functions of mental imagery. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 37.1–37.55). New York, NY: Wiley.
- Fischer, C. E., Marchie, A., & Norris, M. (2004). Musical and auditory hallucinations: A spectrum. *Psychiatry and Clinical Neurosciences*, *58*, 96–98.
- Foulkes, D. (1982). *Children's dreams: Longitudinal studies*. New York, NY: Wiley.

- Fourie, E. (2004). The processing of music notation: Some implications for piano sight-reading. *Journal of the Musical Arts in Africa*, 1, 1–23.
- Fukunishi, I., Horikawa, N., & Onai, H. (1998). Prevalence rate of musical hallucinations in a general hospital setting. *Psychosomatics*, 39, 175.
- Gabrielsson, A. (1999). The performance of music. In D. Deutsch (Ed.), *The psychology of music* (pp. 501–602). New York, NY: Academic Press.
- Geiselman, R. E., & Glenny, J. (1977). Effects of imaging speakers' voices on the retention of words presented visually. *Memory & Cognition*, 5, 499–504.
- George, L., & Neufeld, R. W. J. (1985). Cognition and symptomatology in schizophrenia. *Schizophrenia Bulletin*, 11, 264–284.
- Gissurarson, L. R. (1992). Reported auditory imagery and its relationship with visual imagery. *Journal of Mental Imagery*, 16, 117–122.
- Gollin, E. S., & Sharps, M. J. (1988). Facilitation of free recall by categorical blocking depends upon stimulus type. *Memory & Cognition*, 16, 539–544.
- Gomibuchi, T., Gomibuchi, K., Akiyama, T., Tsuda, H., & Hayakawa, T. (2000). Obsession of hearing music: From the viewpoint of Morita theory. *Psychiatry and Clinical Neurosciences*, 54, 203–206.
- Gordon, E. E. (1965). *Musical aptitude profile*. Boston, MA: Houghton-Mifflin.
- Gordon, E. E. (1975). *Learning theory, patterns, and music*. Buffalo, NY: Tometic Associates.
- Griffiths, T. D. (2000). Musical hallucinosis in acquired deafness: Phenomenology and brain substrate. *Brain*, 123, 2065–2076.
- Halpern, A. R. (1988a). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 434–443.
- Halpern, A. R. (1988b). Perceived and imaged tempos of familiar songs. *Music Perception*, 6, 193–202.
- Halpern, A. R. (1989). Memory for the absolute pitch of familiar songs. *Memory & Cognition*, 17, 572–581.
- Halpern, A. R. (1992). Musical aspects of auditory imagery. In D. Reisberg (Ed.), *Auditory imagery* (pp. 1–27). Hillsdale, NJ: Erlbaum.
- Halpern, A. R. (2003). Cerebral substrates of musical imagery. In I. Peretz & R. Zatorre (Eds.), *The cognitive neuroscience of music* (pp. 217–230). New York, NY: Oxford University Press.
- Halpern, A. R., & Zatorre, R. J. (1999). When the tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, 9, 697–704.
- Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, 42, 1281–1292.
- Harrison, M. (1981). Dysphasia during sleep due to an unusual vascular lesion. *Journal of Neurology, Neurosurgery, and Psychiatry*, 44, 739.
- Harrison, P. (1988). The effects of deafness on musical composition. *Journal of the Royal Society of Medicine*, 81, 598–601.
- Hermesh, H., Konas, S., Shiloh, R., Dar, R., Marom, S., Weizman, A., & Gross-Isseroff, R. (2004). Musical hallucinations: Prevalence in psychotic and non-psychotic outpatients. *Journal of Clinical Psychiatry*, 65, 191–197.
- Heynick, F. (1983). *Theoretical and empirical investigation into verbal aspects of the Freudian model of dream generation*. Unpublished doctoral dissertation, University of Groningen, the Netherlands.
- Highben, Z., & Palmer, C. (2004). Effects of auditory and motor mental practice in memorized piano performance. *Bulletin of the Council for Research in Music Education*, 159, 58–65.
- Hinke, R. M., Hu, X., Stillman, A. E., Kim, S. G., Merkle, H., Salmi, R., & Ugurbil, K. (1993). Functional magnetic resonance imaging of Broca's area during internal speech. *NeuroReport*, 4, 675–678.
- Hogenraad, R., & Orianne, E. (1983). "A penny for your thoughts": Imagery value and the periodicity of interior monologue. *Journal of Mental Imagery*, 7, 51–61.
- Hubbard, T. L. (1996). The importance of a consideration of qualia to imagery and cognition. *Consciousness and Cognition*, 5, 327–358.
- Hubbard, T. L., & Stoeckig, K. (1988). Musical imagery: Generation of tones and chords. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 14, 656–667.
- Hubbard, T. L., & Stoeckig, K. (1992). The representation of pitch in musical images. In D. Reisberg (Ed.), *Auditory imagery* (pp. 199–235). Hillsdale, NJ: Erlbaum.
- Intons-Peterson, M. J. (1980). The role of loudness in auditory imagery. *Memory & Cognition*, 8, 385–393.
- Intons-Peterson, M. J. (1992). Components of auditory imagery. In D. Reisberg (Ed.), *Auditory imagery* (pp. 45–71). Hillsdale, NJ: Erlbaum.
- Intons-Peterson, M. J., Russell, W., & Dressel, S. (1992). The role of pitch in auditory imagery. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 233–240.
- Janata, P. (2001). Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain Topography*, 13, 169–193.
- Janata, P., & Paroo, K. (2006). Acuity of auditory images in pitch and time. *Perception & Psychophysics*, 68, 829–844.
- Jaynes, J. (1976). *The origin of consciousness in the breakdown of the bicameral mind*. Boston, MA: Houghton Mifflin.
- Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition*, 16, 337–342.
- Kadosh, R. C., Kadosh, K. C., & Henik, A. (2007). The neuronal correlate of directional synesthesia: A combined event-related potential and functional magnetic resonance imaging study. *Journal of Cognitive Neuroscience*, 19, 2050–2059.
- Kalamoski, V. (2007). Effect of skill level on recall of visually presented patterns of musical notes. *Scandinavian Journal of Psychology*, 48, 87–96.
- Kasai, K., Asada, T., Yumoto, M., Takeya, J., & Matsuda, H. (1999, November 13). Evidence of functional abnormality in the right auditory cortex during musical hallucinations. *Lancet*, 354, 1703–1704.
- Keller, P. E., & Koch, I. (2008). Action planning in sequential skills: Relations to music performance. *Quarterly Journal of Experimental Psychology*, 61, 275–291.
- Keller, T. A., Cowan, N., & Saults, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 21, 635–645.
- Kendall, R. A., Carterette, E. C., & Hajda, J. M. (1999). Natural vs. synthetic instrument tones. *Music Perception*, 16, 327–364.
- Klinger, E. (1999). Thought flow: Properties and mechanisms underlying shifts in content. In J. A. Singer & P. Salovey (Eds.), *At play in the fields of consciousness: Essays in honor of Jerome L. Singer* (pp. 29–50). Mahwah, NJ: Erlbaum.
- Klinger, E., & Cox, W. M. (1987–1988). Dimensions of thought flow in everyday life. *Imagination, Cognition, and Personality*, 7, 105–128.
- Knoblich, G., & Sebanz, N. (2006). The social nature of perception and action. *Current Directions in Psychological Science*, 15, 99–104.
- Kopiez, R., Weihs, C., Ligges, U., & Lee, J. I. (2006). Classification of high and low achievers in a music sight-reading task. *Psychology of Music*, 34, 5–26.
- Kornicke, L. E. (1995). An exploratory study of individual difference variables in piano sight-reading achievement. *Quarterly Journal of Music Teaching and Learning*, 6, 56–79.
- Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press.
- Kosslyn, S. M., & Matt, A. M. C. (1977). If you speak slowly, do people read your prose slowly? Person-particular speech recoding during reading. *Bulletin of the Psychonomic Society*, 9, 250–252.
- Kosslyn, S. M., Reiser, B. J., Farah, M. J., & Fliegel, S. L. (1983).

- Generating visual images: Units and relations. *Journal of Experimental Psychology: General*, 112, 278–303.
- Kosslyn, S. M., Seger, C., Pani, J. R., & Hillger, L. A. (1990). When is imagery used in everyday life? A diary study. *Journal of Mental Imagery*, 14, 131–152.
- Kosslyn, S. M., & Thompson, W. L. (2000). Shared mechanisms in visual imagery and visual perception: Insights from cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (2nd ed., pp. 975–985). Cambridge, MA: MIT Press.
- Kosslyn, S. M., & Thompson, W. L. (2003). When is early visual cortex activated during visual mental imagery? *Psychological Bulletin*, 129, 723–746.
- Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelly, W. M. (2005, March 10). Sound of silence activates auditory cortex. *Nature*, 434, 158.
- Kristeva, R., Chakarov, V., Schulte-Monting, J., & Spreer, J. (2003). Activation of cortical areas in music execution and imaging: A high-resolution EEG study. *Neuroimage*, 20, 1872–1883.
- Kroll, N. E. A., Schepeler, E. M., & Angin, K. T. (1986). Bizarre imagery: The misremembered mnemonic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12, 42–53.
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York, NY: Oxford University Press.
- Leaver, A. M., van Lare, J., Zielinski, B., Halpern, A. R., & Rauschecker, J. P. (2009). Brain activation during anticipation of sound sequences. *Journal of Neuroscience*, 29, 2477–2485.
- Lennox, B. R., Park, S. B. G., Medley, I., Morris, P. G., & Jones, P. B. (2000). The functional anatomy of auditory hallucinations in schizophrenia. *Psychiatry Research: Neuroimaging*, 100, 13–20.
- Levitin, D. J. (2007). *This is your brain on music: The science of a human obsession*. New York, NY: Plume Books.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Lockwood, L. (2002). *Beethoven: The music and the life*. New York, NY: Norton.
- MacKay, D. G. (1992). Constraints on theories of inner speech. In D. Reisberg (Ed.), *Auditory imagery* (pp. 121–149). Hillsdale, NJ: Erlbaum.
- Marin, O. S. M., & Perry, D. W. (1999). Neurological aspects of music perception and performance. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 653–724). New York, NY: Academic Press.
- Marks, D. F. (1995). New directions for mental imagery research. *Journal of Mental Imagery*, 19, 153–166.
- Marks, L. E. (1975). On colored-hearing synesthesia: Cross-modal translations of sensory dimensions. *Psychological Bulletin*, 82, 303–331.
- Marks, L. E. (1978). *The unity of the senses*. New York, NY: Academic Press.
- McGuire, P. K., Silbersweig, D. A., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1996). Functional neuroanatomy of inner speech and auditory verbal imagery. *Psychological Medicine*, 26, 29–38.
- McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., David, A. S., Frackowiak, R. S. J., & Frith, C. D. (1995, September 2). Abnormal monitoring of inner speech: A physiological basis for auditory hallucinations. *Lancet*, 346, 596–600.
- McGuire, P. K., Silbersweig, D. A., Wright, I., Murray, R. M., Frackowiak, R. S., & Frith, C. D. (1996). The neural correlates of inner speech and auditory verbal imagery in schizophrenia: Relationship to auditory verbal hallucinations. *British Journal of Psychiatry*, 169, 148–159.
- McKellar, P. (1965). The investigation of mental images. In S. A. Barnett & A. McClaren (Eds.), *Penguin Science Survey* (pp. 79–94). Harmondsworth, England: Penguin Books.
- McKelvie, S. J. (1995). Responses to commentaries: The VVIQ and beyond: Vividness and its measurement. *Journal of Mental Imagery*, 19, 197–252.
- Meister, I. G., Krings, T., Foltys, H., Borojerd, B., Müller, M., Töpper, R., & Thron, A. (2004). Playing piano in the mind—An fMRI study on music imagery and performance in pianists. *Cognitive Brain Research*, 19, 219–228.
- Meyer, M., Elmer, S., Baumann, S., & Jancke, L. (2007). Short-term plasticity in the auditory system: Differential neural responses to perception and imagery of speech and music. *Restorative Neurology and Neuroscience*, 25, 411–431.
- Mintz, A., & Alpert, M. (1972). Imagery vividness, reality testing, and schizophrenic hallucinations. *Journal of Abnormal Psychology*, 79, 310–316.
- Mountain, R. (2001). Composers and imagery: Myths and realities. In R. I. Godoy & H. Jorgensen (Eds.), *Musical imagery* (pp. 271–288). New York, NY: Taylor & Francis.
- Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics*, 13, 180–184.
- Moyer, R. S., & Bayer, R. H. (1976). Mental comparison and the symbolic distance effect. *Cognitive Psychology*, 8, 228–246.
- Murata, A., Naritomi, H., & Sawada, T. (1994). Musical auditory hallucinations caused by a brainstem lesion. *Neurology*, 44, 156–158.
- Murray, D. J. (1965). Vocalization-at-presentation and immediate recall, with varying presentation rates. *Quarterly Journal of Experimental Psychology*, 17, 47–56.
- Nagaratam, N., Virk, S., & Brdarevic, O. (1996). Musical hallucinations associated with recurrence of a right occipital meningioma. *British Journal of Clinical Practice*, 50, 56–57.
- Narmour, E. (1990). *The analysis and cognition of basic melodic structures*. Chicago, IL: University of Chicago Press.
- Nayani, T. H., & David, A. S. (1996). The auditory hallucination: A phenomenological survey. *Psychological Medicine*, 26, 179–192.
- Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. New York, NY: Freeman.
- Okada, H., & Matsuoka, K. (1992). Effects of auditory imagery on the detection of a pure tone in white noise: Experimental evidence of the auditory Perky effect. *Perceptual & Motor Skills*, 74, 443–448.
- Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, 106, 528–537.
- Paivio, A. (1969). Mental imagery in associative learning and memory. *Psychological Review*, 76, 241–263.
- Paivio, A. (1971). *Imagery and verbal processes*. New York, NY: Holt, Rinehart, & Winston.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.
- Peacock, K. (1985). Synaesthetic perception: Alexander Scriabin's color hearing. *Music Perception*, 2, 483–505.
- Pearce, J. M. S. (2007). Synaesthesia. *European Neurology*, 57, 120–124.
- Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition*, 20, 314–320.
- Peretz, I., Champod, A. S., & Hyde, K. (2003). Varieties of musical disorders: The Montreal battery of evaluation of amusia. *Annals of the New York Academy of Sciences*, 999, 58–75.
- Peretz, I., & Hyde, K. (2003). What is specific to music processing? Insights from congenital amusia. *Trends in Cognitive Sciences*, 7, 362–367.
- Peterson, M., Kihlstrom, J., Rose, P., & Glisky, M. (1992). Mental images can be ambiguous: Reconstructions and reference frame reversals. *Memory & Cognition*, 20, 107–123.
- Pitt, M. A., & Crowder, R. G. (1992). The role of spectral and dynamic cues in imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 728–738.
- Posner, M. I. (1978). *Chronometric explorations of mind*. Englewood Heights, NJ: Erlbaum.
- Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An

- information-processing account of its origins and significance. *Psychological Review*, 83, 157–171.
- Rainey, D. W., & Larsen, J. D. (2002). The effects of familiar melodies on initial learning and long-term memory for unconnected text. *Music Perception*, 20, 173–186.
- Reisberg, D. (1992). *Auditory imagery*. Hillsdale, NJ: Erlbaum.
- Reisberg, D., & Chambers, D. (1991). Neither pictures nor propositions: What can we learn from a mental image? *Canadian Journal of Psychology*, 45, 336–352.
- Reisberg, D., & Heuer, F. (2005). Visuospatial images. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 35–80). New York, NY: Cambridge University Press.
- Reisberg, D., Smith, J. D., Baxter, D. A., & Sonenshine, M. (1989). “Enacted” auditory images are ambiguous; “Pure” auditory images are not. *Quarterly Journal of Experimental Psychology*, 41(A), 619–641.
- Robertson, L., & Sagiv, N. (Eds.). (2005). *Synesthesia: Perspectives from cognitive neuroscience*. Oxford, England: Oxford University Press.
- Saba, P. R., & Keshavan, M. S. (1997). Musical hallucinations and musical imagery: Prevalence and phenomenology in schizophrenic patients. *Psychopathology*, 30, 185–190.
- Sacks, O. (2007). *Musophilia: Tales of music and the brain*. New York, NY: Knopf.
- Salame, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unintended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21, 150–164.
- Samson, S., & Zatorre, R. J. (1991). Recognition memory for text and melody of songs after unilateral temporal lobe lesion: Evidence for dual encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 17, 793–804.
- Schielke, E., Reuter, U., Hoffman, O., & Weber, J. R. (2000). Musical hallucinations with dorsal pontine lesions. *Neurology*, 55, 454–455.
- Schneider, T. R., Engel, A. K., & Debener, S. (2008). Multisensory identification of natural objects in a two-way crossmodal priming paradigm. *Experimental Psychology*, 55, 121–132.
- Schurmann, M., Raji, T., Fujiki, N., & Hari, R. (2002). Mind’s ear in a musician: Where and when in the brain. *Neuroimage*, 16, 434–440.
- Seal, M. L., Aleman, A., & McGuire, P. K. (2004). Compelling imagery, unanticipated speech, and deceptive memory: Neurocognitive models of auditory verbal hallucinations in schizophrenia. *Cognitive Neuropsychiatry*, 9, 43–72.
- Seashore, C. E. (1967). *Psychology of music*. New York, NY: Dover. (Original work published 1938)
- Seashore, C. E., Lewis, L., & Saetveit, J. G. (1960). *Seashore measures of musical talents*. New York, NY: The Psychological Corporation.
- Segal, S. J., & Fusella, V. (1970). Influence of imaged pictures and sounds in detection of visual and auditory signals. *Journal of Experimental Psychology*, 83, 458–474.
- Seitz, P., & Molholm, H. B. (1947). Relation of mental imagery to hallucinations. *Archives of Neurology and Psychiatry*, 57, 469–480.
- Sharps, M. J., & Pollitt, B. K. (1998). Category superiority effects and the processing of auditory images. *Journal of General Psychology*, 125, 109–116.
- Sharps, M. J., & Price, J. L. (1992). Auditory imagery and free recall. *Journal of General Psychology*, 119, 81–87.
- Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior*, 18, 41–55.
- Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- Shergill, S. S., Bullmore, E. T., Brammer, M. J., Williams, S. C. R., Murray, R. M., & McGuire, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine*, 31, 241–253.
- Shergill, S. S., Bullmore, E., Simmons, A., Murray, R., & McGuire, P. (2000). Functional anatomy of auditory verbal imagery in schizophrenic patients with auditory hallucinations. *American Journal of Psychiatry*, 157, 1691–1693.
- Shergill, S. S., Murray, R. M., & McGuire, P. K. (1998). Auditory hallucinations: A review of psychological treatments. *Schizophrenia Research*, 32, 137–150.
- Shinosaki, K., Yamamoto, M., Ukai, S., Kawaguchi, S., Ogawa, A., Ishii, R., . . . Takeda, M. (2003). Desynchronization in the right auditory cortex during musical hallucinations: A MEG study. *Psychogeriatrics*, 3, 88–92.
- Slade, P. D. (1976). An investigation of psychological factors involved in the predisposition to auditory hallucinations. *Psychological Medicine*, 6, 123–132.
- Sloboda, J. A. (1984). Experimental studies of music reading: A review. *Music Perception*, 2, 222–236.
- Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. New York, NY: Oxford University Press.
- Smith, J. D. (1992). The auditory hallucinations of schizophrenia. In D. Reisberg (Ed.), *Auditory imagery* (pp. 151–178). Hillsdale, NJ: Erlbaum.
- Smith, J. D., Reisberg, D., & Wilson, M. (1992). Subvocalization and auditory imagery: Interactions between the inner ear and inner voice. In D. Reisberg (Ed.), *Auditory imagery* (pp. 95–119). Hillsdale, NJ: Erlbaum.
- Smith, J. D., Wilson, M., & Reisberg, D. (1995). The role of subvocalization in auditory imagery. *Neuropsychologia*, 33, 1433–1454.
- Snyder, F. (1970). The phenomenology of dreaming. In L. Madlow & L. H. Snow (Eds.), *The psychodynamic implications of the physiological studies on dreams* (pp. 124–151). Springfield, IL: Thomas.
- Solms, M. (1997). *The neuropsychology of dreams: A clinico-anatomical study*. Mahwah, NJ: Erlbaum.
- Starker, S., & Jolin, A. (1982). Imagery and hallucination in schizophrenic patients. *Journal of Nervous and Mental Disease*, 170, 448–451.
- Stevenson, R. J., & Case, T. I. (2005). Olfactory imagery: A review. *Psychonomic Bulletin & Review*, 12, 244–264.
- Stewart, L., Walsh, V., Frith, U., & Rothwell, J. C. (2001). TMS produces two dissociable types of speech disruption. *Neuroimage*, 13, 472–478.
- Stuart, G. P., & Jones, D. M. (1996). From auditory image to auditory percept: Facilitation through common processes? *Memory & Cognition*, 24, 296–304.
- Sullivan, C., Urakawa, K. S., & Cossey, V. L. (1996). Separating the effects of alertness from the effects of encoding in a pitch-imagery task. *Journal of General Psychology*, 123, 105–114.
- Tanriverdi, N., Sayilgan, M. A., & Ozcurumez, G. (2001). Musical hallucinations associated with abruptly developed bilateral loss of hearing. *Acta Psychiatrica Scandinavica*, 103, 153–155.
- Tinti, C., Cornoldi, C., & Marschark, M. (1997). Modality-specific auditory imaging and the interactive imagery effect. *European Journal of Cognitive Psychology*, 9, 417–436.
- Tulving, E. (1983). *Elements of episodic memory*. New York, NY: Oxford University Press.
- VanVoorhis, C. R. W. (2002). Stat jingles: To sing or not to sing. *Teaching of Psychology*, 29, 249–250.
- Ward, J., & Mattingley, J. B. (2006). Synaesthesia: An overview of contemporary findings and controversies. *Cortex*, 42, 129–136.
- Ward, J., Simner, J., & Auyeung, V. (2005). A comparison of lexical-gustatory and grapheme-colour synesthesia. *Cognitive Neuropsychology*, 22, 28–41.
- Ward, J., Thompson-Lake, D., Ely, R., & Kaminski, F. (2008). Synaesthesia, creativity, and art: What is the link? *British Journal of Psychology*, 99, 127–141.
- Ward, J., Tsakanikos, E., & Bray, A. (2006). Synaesthesia for reading and playing musical notes. *Neurocase*, 12, 27–34.
- Warren, R. M. (1968). Verbal transformation effect and auditory perceptual mechanisms. *Psychological Bulletin*, 70, 261–270.
- Waters, A. J., Townsend, E., & Underwood, G. (1998). Expertise in

- musical sight-reading: A study of pianists. *British Journal of Psychology*, 89, 123–149.
- Weber, R. J., & Bach, M. (1969). Visual and speech imagery. *British Journal of Psychology*, 60, 199–202.
- Weber, R. J., & Brown, S. (1986). Musical imagery. *Music Perception*, 3, 411–426.
- Weber, R. J., & Castleman, J. (1970). The time it takes to imagine. *Perception & Psychophysics*, 8, 165–168.
- Welch, R. B., & Warren, D. H. (1986). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, 88, 638–667.
- Wilson, M., & Knoblich, G. (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, 131, 460–473.
- Wing, H. D. (1961). *Standardized tests of musical intelligence*. The Mere, England: National Foundation for Educational Research.
- Wolf, T. (1976). A cognitive model of musical sight-reading. *Journal of Psycholinguistic Research*, 5, 143–171.
- Wöllner, C., Halfpenny, E., Ho, S., & Kurosawa, K. (2003). The effects of distracted inner hearing on sight-reading. *Psychology of Music*, 31, 377–389.
- Wu, J., Mai, X., Chan, C. C. H., Zheng, Y., & Luo, Y. (2006). Event-related potentials during mental imagery of animal sounds. *Psychophysiology*, 43, 592–597.
- Yoo, S. S., Lee, C. U., & Choi, B. G. (2001). Human brain mapping of auditory imagery: Event-related functional MRI study. *NeuroReport*, 12, 3045–3049.
- Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience*, 20, 1–19.
- Zatorre, R. J., & Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia*, 31, 221–232.
- Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Musical imagery and the auditory cortex. *Neuron*, 47, 9–12.
- Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2009). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience*. Advance online publication. doi:10.1162/jocn.2009.21239
- Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, 8, 29–46.

Received January 1, 2009

Revision received October 12, 2009

Accepted October 30, 2009 ■

ORDER FORM

Start my 2010 subscription to ***Psychological Bulletin***
ISSN: 0033-2909

___ \$102.00	APA MEMBER/AFFILIATE	_____
___ \$261.00	INDIVIDUAL NONMEMBER	_____
___ \$699.00	INSTITUTION	_____
	<i>In DC and MD add 6% sales tax</i>	_____
	TOTAL AMOUNT DUE	\$ _____

Subscription orders must be prepaid. Subscriptions are on a calendar year basis only. Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.



AMERICAN
PSYCHOLOGICAL
ASSOCIATION

SEND THIS ORDER FORM TO

American Psychological Association
Subscriptions
750 First Street, NE
Washington, DC 20002-4242

Call **800-374-2721** or 202-336-5600
Fax **202-336-5568** : TDD/TTY **202-336-6123**
For subscription information,
e-mail: subscriptions@apa.org

Check enclosed (make payable to APA)

Charge my: Visa MasterCard American Express

Cardholder Name _____

Card No. _____ Exp. Date _____

Signature (Required for Charge)

Billing Address

Street _____

City _____ State _____ Zip _____

Daytime Phone _____

E-mail _____

Mail To

Name _____

Address _____

City _____ State _____ Zip _____

APA Member # _____

BULA10