

Neurophysiological Mechanisms Underlying Auditory Image Formation in Music

Petr Janata

Introduction

The formation of contextually dependent expectancies is an important feature of music cognition. Both explicit and implicit knowledge about the structure of a piece of music serve to establish highly specific expectations about the pitch, timbre, and other features of ensuing musical information. Musical expectancies represent a specific type of musical imagery. On the one hand, musical imagery might be thought of as a mental process that occurs over an extended period as a person imagines hearing or performing a piece of music. This type of imagery differs from expectancy formation in that it may transpire in the absence of sensory input. Active expectancy formation, on the other hand, generally requires that specific images for subsequent sensory input are based on preceding sensory input and established knowledge of what type of sensory input to expect.

A neuroscientific way of framing the general question is, ‘What are the brain areas and mechanisms that support the formation of such images and the interaction of these images with incoming sensory information?’ Electrophysiological measures of brain activity provide a description of how the human brain implements these types of processes, and a variety of different experimental designs can be used to address various components of these processes. Over the past 30 years, studies of auditory evoked potentials have provided support for models of how voluntarily maintained

images (expectancies) of single tones interact with sensory input consisting of multiple auditory stimuli. Stimuli based on musical considerations 1) extend the types of designs that can be used to study mechanisms of auditory image formation, 2) provide important tests of the existing models, and 3) provide a framework, rooted in the neuroethological tradition (Pflüger & Menzel, 1999), for understanding the neural underpinnings of human musical behavior.

Forms of musical imagery

Perhaps the first step in studying musical imagery is to place musical imagery in the broader context of auditory imagery and the general domain of mental imagery, if for no other reason than to borrow from definitions of mental imagery derived primarily from considerations of the form and formation of visual images. Finke (1989) defines mental imagery as, 'the mental invention or recreation of an experience that in at least some respects resembles the experience of actually perceiving an object or an event, either in conjunction with, or in the absence of, direct sensory stimulation.'

In order to pinpoint and characterize specific neural mechanisms underlying musical imagery, it is necessary to define what a musical image is and what the processes of forming such an image or series of images are. Mirroring Finke's definition, I consider two contexts in which musical imagery occurs. In the first context, musical imagery is purely a mental act: an endogenous phenomenon in which the content of the images is internally generated from long-term memory stores of musical knowledge and is uninfluenced by any concurrent sensory input. In the second context, formation of musical images depends on an interaction of memory-dependent processes (expectancies) with representations of incoming auditory input.

The relationship between perception and mental imagery has been elaborated and tested extensively with visual material by Kosslyn (1980, 1994). In Kosslyn's view (1994, p. 287), 'images are formed by the same processes that allow one to anticipate what one would see if a particular object or scene were present.' Thus, postulating two contexts for musical imagery is in keeping with other theories of mental imagery.

Figure 1 on the facing page shows a theoretical framework for thinking about how imagery processes within these two different contexts might be instantiated in a set of brain structures. Because a complete theory of musical imagery should include also imagery for musical performance and the interaction of sensory and motor information, the diagram in Figure 1 is restricted, in the interest of relative simplicity, to represent processes that may be involved in 'sensory' imagery rather than 'motor' imagery. Following a brief description of the framework, I summarize the physiological methods and experiments in support of it.

The arrows represent the flow of information across different general brain areas listed at the right through time. Those brain areas involved more immediately with sensory processing are listed at the bottom, while those involved in abstract reasoning and memory storage/retrieval are listed toward the top. The first type of auditory imagery unfolds over longer time periods (seconds or minutes) and is generally unconstrained by sensory input. I call it 'non-expectant' because we neither expect to hear anything as we are imagining, nor are we forming expectations of what we will

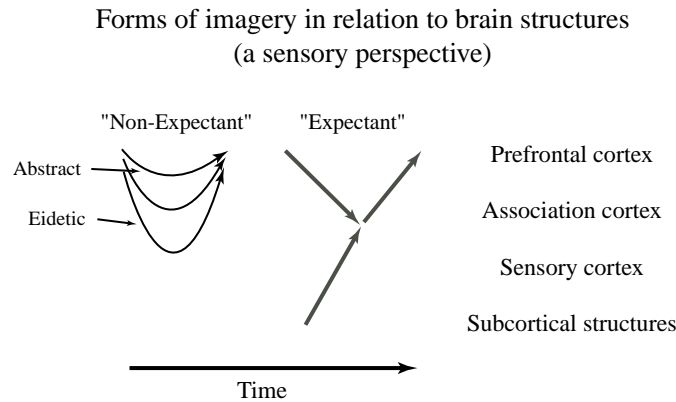


Figure 1. Schematic view of different types of auditory imagery and how these might be instantiated in the human brain (see text for details).

hear in the immediate future. This is the type of imagery we engage in when we imagine a melody in our mind. Similarly, we might mentally improvise melodies that we have never heard before but are able to compose based on knowledge, either explicit or implicit, of tonal sequences, etc. Thus, this type of imagery relies on long-term memories we have of specific musical material, or on a more abstract knowledge of musical structure, e.g. the tonal relationships in western tonal music.

Non-expectant imagery may be differentiated further into two modes of imagery that I call 'abstract' and 'eidetic'. In the abstract mode, the sequence of pitches in a melody might be imagined without the sense of 'hearing' the melody being played by any particular instrument. In the eidetic mode, a strong component of the image is the impression that an instrument or group of instruments is performing the imagined piece of music, and the image has a strong sensory quality to it.

To the extent that this 'non-expectant' imagery depends on retrieving information from long term memory stores, it may rely heavily on areas of prefrontal cortex which have been implicated in general memory functions (Goldman-Rakic, 1996). One prediction of the 'abstract/eidetic' distinction is that the eidetic qualities of images engage brain regions more immediately involved in processing sensory information. Increased eidetic qualities of the images are represented by increases in the depths of the arcs of the solid arrows in Figure 1. For example, a strong impression of hearing an orchestra in our minds might be indicative of auditory cortex involvement in the imagery process. These relationship between the vividness of the mental image and the brain areas that are activated by the image remains to be tested. A modicum of support for the notion that more 'true-to-life' images activate those brain areas more immediately involved in the formation of sensory representations comes from neuroimaging studies of schizophrenic patients experiencing auditory hallucinations which show that auditory cortical areas, including primary auditory cortex are activated during hallucinations (Dierks et al., 1999; Griffiths, Jackson, Spillane, Friston, & Frackowiak, 1997).

'Expectant' imagery refers to the process of forming mental images when listening, attentively, to music or sounds in general. In addition to relying on long-term memory for musical structure or a specific piece of music, the mental images are additionally constrained by the interactions of contemporaneous sensory information with the memorized information. In other words, as we listen to the notes of an ascending major scale we can form a very specific image/expectation of the next note in the scale. The specificity arises in part from our knowledge of the intervalic relationships between successive notes in a major scale, as well as the exact frequencies of the notes being used to create this particular instance of the scale. If the playing of the notes in the scale were to stop, we could continue forming images of the remaining notes. Similarly, in listening to a chamber ensemble playing a familiar piece of music, our expectancies are formed from familiarity with the piece of music as well as the sensory information we are receiving about the tone of these particular instruments, or the expressive tendencies of the particular ensemble. In Figure 1, the merging of the straight arrows represents the interaction of 'top-down' expectancies with 'bottom-up' sensory input, and a subsequent influence of this interaction on the expectancy/image forming processes in a continuous, iterative, process. The extensive literature on the formation of auditory representations (reviewed in Näätänen & Winkler, 1999) and the interaction of these representations with top-down influences such as selective attention (Näätänen, 1992) implicate areas such as the secondary auditory cortex as a neural substrate for these interactions.

Methods for probing human brain activity

Inferences about brain function are made by measuring changes in one or several dependent variable(s) as a subject performs specific cognitive tasks. The dependent variables can range from behavioral measures such as reaction time and accuracy to physiological measures of regional cerebral blood flow or field potentials generated by neurons. I will focus on the physiological measures, first providing a brief description of the signals being measured along with the measurement and data analysis techniques and then summarizing results of studies that are of particular relevance to musical imagery.

Studies of the brain's physiological responses typically strive to 1) identify brain areas that are responsible for performing specific tasks or computations, and 2) describe the neural mechanisms by which stimuli are represented and cognitive tasks are performed. Functional imaging techniques such as positron emission tomography (PET) and functional magnetic resonance imaging (fMRI) are well-suited to address the first goal. Both of these methods monitor blood flow changes in the brain. As large populations of neurons in those brain areas that perform a cognitive task become active and increase their metabolism, the blood supply to those areas increases in order to meet the increased demand for oxygen (Villringer, 1999). Note that neuronal activity is modulated more quickly than is the associated blood flow, resulting in a temporal resolution limit (peak response is 2-4 seconds from onset of event) of PET and fMRI. While PET and fMRI provide only an indirect measure of neural activity, they do so with much better spatial resolution than do direct, non-invasive measurements

of neural activity. Thus, these methods are invaluable tools for localizing cognitive functions, and their application to issues of auditory imagery is described below.

The temporal properties of neural responses in cognitive tasks are best captured by direct measures of the neural activity. The electrical fields generated by large populations of neurons comprise the electroencephalogram (EEG), and the magnetic counterpart to the electrical fields forms the magnetoencephalogram (MEG). Given the distance of EEG recording electrodes positioned at the scalp from the cortical surface, large populations of similarly oriented neurons must be active synchronously in order for them to create a sufficiently strong electrical field that can be detected at the scalp surface (Nunez, 1981). The superposition of many electrical fields from many neuronal populations, along with the low-pass filtering characteristics of the skull, makes the problem of unambiguously localizing the neural sources based on EEG data a difficult one.

The nature of the experimental situation generally dictates how the EEG is analyzed. When the experiment consists of short, discrete, and clearly defined stimuli, such as single tones embedded in longer sequences of tones, the stimuli are presented many (30-1000) times while the EEG is recorded. The EEG responses to each presentation are then averaged in order to extract the mean waveform. This waveform is interpreted as the unique response to the particular type of stimulus, and is generally referred to as an event-related potential (ERP) waveform. The ERP waveform is analyzed in terms of the amplitudes, latencies, and areas of the peaks and troughs. The nomenclature reflects the polarity and typical latency of the deflection in the ERP waveform. For example, the auditory N100 is a negative peak (as measured from an electrode at the vertex of the head relative to a mastoid, ear, noise or non-cephalic reference electrode) which occurs approximately 100 ms following the onset of the stimulus. Features of the ERP waveform, such as the N100 or P300, are commonly called 'components' to indicate their dependence on the perceptual/cognitive factors that modulate their presence, size, and latency. Individual ERP components, particularly those occurring in the hundreds of milliseconds do not necessarily reflect unitary cognitive phenomena. For example, the amplitude of the N100 is modulated by both physical features of the stimulus as well as the attentional state of the subject. The N100 represents a conglomerate of brain processes associated with processing an auditory stimulus (Näätänen & Picton, 1987), and it may overlap with other ERP components such as the mismatch negativity (MMN) (Näätänen, 1992).

In tasks employing continuous stimulus situations in which individual discrete events cannot be identified, such as listening to a recorded piece of music, the EEG is typically analyzed in the frequency domain. Here, the assumption is that the performance of any given cognitive task will be associated with a sustained and stable pattern of neural activity involving some number of brain areas. The field potential arising from the neural activity pattern is then described by its frequency spectrum. Typically, successive 2s EEG epochs are converted into the frequency domain and the average power spectrum is computed. The magnitude of the power spectrum indicates the strength with which oscillations at a particular frequency are represented in the EEG during a cognitive state. Brain activity associated with any given cognitive task can be isolated by subtracting the average power spectrum during the task from the average power spectrum during rest, and the synchronization of different brain regions

can be assessed through the frequency-specific coherence between pairs of electrodes (Rappelsberger & Petsche, 1988; Srinivasan, Nunez, & Silberstein, 1998). Overall, the spectral analysis of the EEG tends to provide information about global brain dynamics associated with a particular cognitive task, whereas ERPs are used to elucidate the sequence of processing steps associated with discrete stimulus events.

Physiological measures of mental images

‘Non-expectant imagery’

What is the manifestation of different types of auditory/musical imagery in the brain, and what is the evidence supporting the functional architecture described above? To the extent that ‘non-expectant’ imagery establishes a set of stationary processes in the brain, i.e. a stable pattern of activity within specific neural circuits over the duration that subjects perform an imagery task, it should be possible to capture signatures of these processes in EEG recordings. Petsche and colleagues have found wide-spread coordination, as manifested in the coherence of the EEG, of brain areas as subjects imagine hearing or composing musical material (Petsche, Richter, von Stein, Etlinger, & Filz, 1993) or mentally play an instrument (Petsche, von Stein, & Filz, 1996). Imagining a piece of music leads to an increase in the number of observed coherence changes compared to listening to the same piece of music. The exact patterns of changes differ appreciably among subjects, however. For example, in one subject imagery is associated with theta and alpha band decreases and beta band increases, whereas in another subject there are coherence increases across all frequency bands.

So far, the best evidence for those brain areas involved in auditory imagery in the absence of acoustic input come from PET studies by Zatorre and colleagues (Halpern & Zatorre, 1999; Zatorre, Evans, & Meyer, 1994; Zatorre, Halpern, Perry, Meyer, & Evans, 1996). In their tasks, mentally scanning through a melody results in superior temporal gyrus (STG) and right frontal lobe activations. The auditory cortex lies along the STG. The frontal lobes are widely implicated in memory retrieval processes (Goldman-Rakic 1996). The observation that the auditory cortex is activated during these tasks is extremely important because it indicates that those structures responsible for the processing of auditory stimuli are also activated under certain conditions of musical/auditory imagery when no sensory stimuli are present. Indirectly, these results suggest that discrete, mentally generated, auditory images might be compared against incoming sensory information in the auditory cortex.

Formation of specific musical expectancies

Although spectral analysis of the EEG is typically applied to sustained tasks rather than to event-related tasks, it has been applied to analyzing the build up and resolution of harmonic expectancies (Janata & Petsche, 1993), implicating right frontal and temporal areas in the processing of cadences and their resolutions. ERPs analyzed as time-domain averages have been used by several researchers to probe musical expectancy. Although the results of the studies differ slightly, the general finding is that unexpected notes and chords elicit larger positive potentials from 300 to 600 ms following the onset of the stimulus than do highly expected notes and chords (Besson

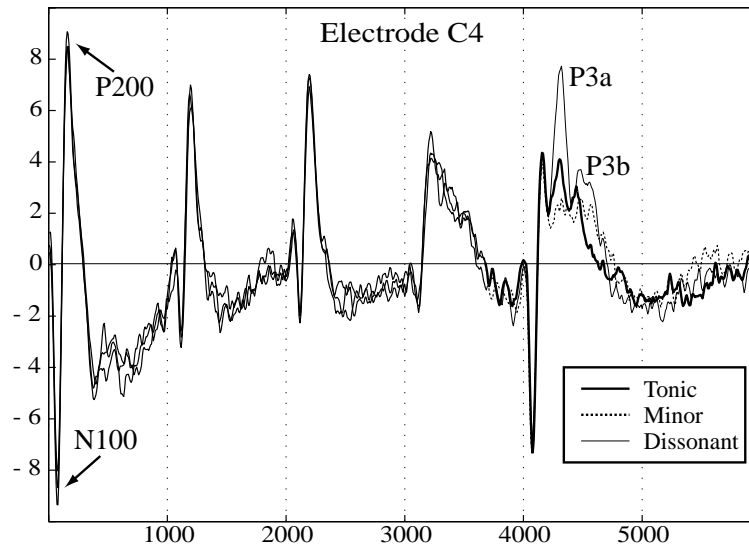


Figure 2. Averaged event related potentials (ERPs) recorded during a harmonic priming task in which subjects heard a I, IV, V cadence (at 0 ms, 1000 ms, 2000 ms, respectively), imagined the best possible resolution (3000-4000 ms) and heard one of three possible resolutions at 4000 ms, whereupon they had to decide if it was the resolution they had imagined. The waveform components labeled P3a and P3b typically vary in amplitude as a function of expectancy. Adapted from Janata, 1995.

& Faïta, 1995; Besson & Macar, 1987; Hantz, Kreilick, Kananen, & Swartz, 1997; Janata, 1995; see also Janata & Petsche, 1993 for a frequency domain analysis of event-related EEG data; Patel, Gibson, Ratner, Besson, & Holcomb, 1998). The general names in the ERP literature for the large, late positive waves are 'P300' and late positive complex (LPC), and their amplitude is inversely proportional to the subjective probability of the eliciting event (for a review, see Donchin & Coles, 1988; Verleger, 1988). More recently, researchers have focussed on contributions of frontal brain areas to the processing of harmonic expectancies, showing negative shifts in the waveforms for harmonically deviant chords, compared to contextually consonant chords (Koelsch, Gunter, Friederici, & Schröger, 2000; Patel et al., 1998).

Figure 2 shows ERP waveforms in response to chords in a priming (I, IV, V) cadence and three resolutions of the cadence. In this experiment (Janata, 1995), chords in the priming cadences were presented in numerous inversions and several different keys, using a sampled grand piano sound. Each chord was presented for 1 s. For 1s between the offset of the final chord and the onset of the resolution (3000-4000s), subjects imagined the best possible resolution. At 4000 ms, one of three possible resolutions was heard: the expected resolution to the I (thick solid line), a harmonically plausible resolution to the tonic of the relative minor (dashed line), or a harmonically implausible resolution to a triad based on the tritone (thin solid line). The large negative (N100) and positive peaks (P200) characterize the auditory evoked potential. The

response to the harmonically incongruous ending elicited the largest amplitudes in the two P300 components. The brain circuitry that gives rise to these late potentials is not well understood. Intracerebral recordings indicate multiple sites that exhibit P300-like activations (Halgren, Marinkovic, & Chauvel, 1998). Thus, the details of how activity in auditory cortical regions is coordinated with activity in other brain areas necessarily remain murky.

Interestingly, Janata (1995) observed a large waveform in the period when subjects were asked to imagine the best possible resolution, suggesting that this was a measurable brain response to the act of imagining the resolution to the tonic. Unfortunately, the evoked potential may have been caused by the offset of the previous chord, rather than voluntary image formation. Further studies have been performed to investigate evoked potentials elicited by imagined events, and some preliminary results are presented below.

Measures of expectancies in the absence of sensory input

In ERP studies of auditory expectancy, brain potentials resulting from the expectancy forming process and potentials arising in response to sensory stimulation are combined. Although it is possible to measure the outcome of the interaction of the expectancy (mental image) with the sensory information, determining the electrophysiological signature of each information stream poses a greater challenge. One way of studying the expectation is to simply omit an expected stimulus, thereby removing all sensory components from the ERP response (Besson & Faïta, 1995; Besson, Faïta, Czternasty, & Kutas, 1997; Ruchkin, Sutton, & Tueting, 1975; Simson, Vaughan, & Ritter, 1976; Sutton, Tueting, Zubin, & John, 1967; Weinberg, Walter, & Crow, 1970). Such omissions generate a large P300, typical of unexpected stimuli. In some cases, earlier components, reminiscent of the auditory evoked potential are also present (Besson et al., 1997). The relationship between potentials generated in response to unexpected stimulus omissions and voluntarily generated images has not been explored in more detail, however.

These earlier studies raise several questions about the neural processes involved in forming auditory expectancies and the interactions of these expectancies with sensory input. Specifically, can the mental processes involved in mental image formation be dissociated further from the process of expectation? In other words, is it possible to measure emitted potentials associated with the formation of a mental image in the complete absence of an expectation that a stimulus will occur? If the answer is yes, is there any evidence that such emitted potentials arise from the auditory cortex where expectancies and sensory input are believed to interact?

To begin investigating these questions, I performed a preliminary study in which musically trained subjects were asked to first listen to and then imagine completions of simple eight-note melodic phrases. Examples of the melodic fragments and the various experimental conditions are schematized in Figure 3 on the facing page. On each 'imagery' trial (Figure 3B), subjects first heard all eight notes of the melody, and made a key-press synchronously with the last note ('All Heard' condition). Next, they heard the first five notes of the same melody and continued imagining the remaining three, pressing a key at the time they thought the last note would have occurred ('3 Imagined' condition). They then heard the first three notes of the melody, imag-

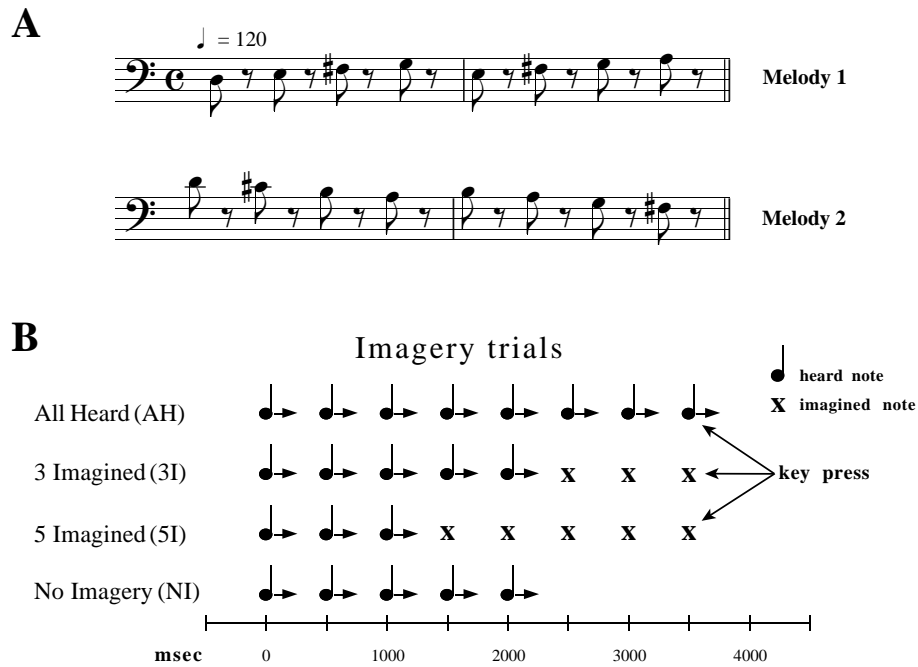


Figure 3. Melodies and experimental conditions used in the auditory imagery experiments. A) The two simple melodies in musical notation. B) Schematic diagram of imagery trials. In the ‘All Heard’ condition subjects heard the entire melodic fragment. In the ‘3 Imagined’ condition, the initial five notes were heard and the remaining three imagined. In the ‘5 Imagined’ condition, subjects heard the initial three notes and imagined the remaining five. On each trial, these three conditions appeared in immediate succession. In separate blocks of ‘No Imagery’ trials, subjects heard the initial five notes but did not imagine the remaining three.

ined the remaining five, once again making a key-press synchronously with the last imagined note (‘5 Imagined’ condition). In a separate control block of ‘no-imagery’ trials, subjects heard the first five notes of the melody but did not continue imagining the remaining notes and made no key presses (‘No Imagery’ condition). They were explicitly instructed to not imagine a continuation of the melody and to try to hear the five notes as a complete phrase. In order to achieve an appropriate signal to noise ratio in the ERP waveform, subjects performed 100 imagery trials and 100 no-imagery trials.

Subjects’ brain electrical activity was recorded throughout each trial using a geodesic array of 129 electrodes (Electrical Geodesics Inc., Eugene, OR) distributed across the scalp. Such dense sampling of the electrical potential at the scalp surface allows one to construct an accurate topographical map of the voltage at each sampled time point.

Figure 4A (see page 10) illustrates a typical auditory N100 topography in which

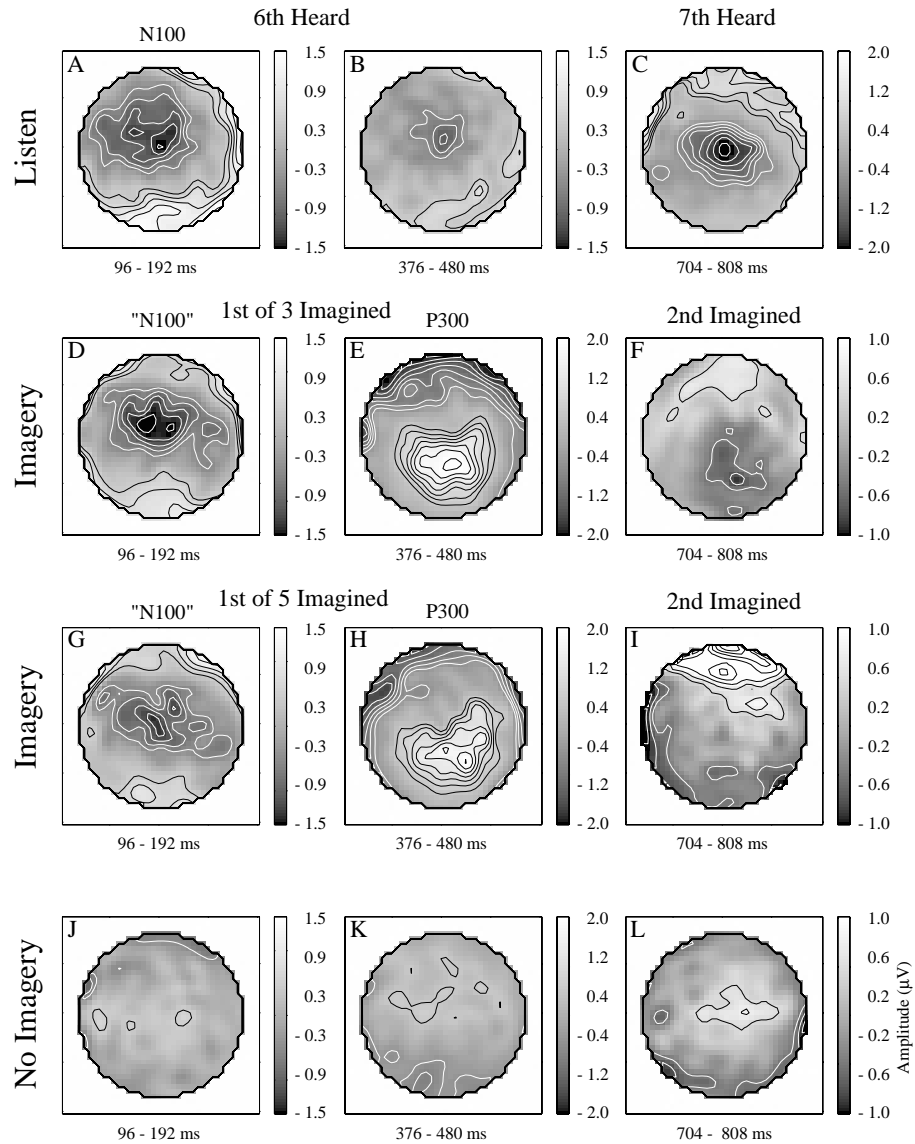


Figure 4. Summary topographical maps of the activation elicited by different tasks (each of the four rows) in the auditory imagery experiment. Each circle represents a view down onto the top of the head. The nose would be at the top of the circle. Plotted are the average voltage values in time windows that encompass the N100 (left column), the P300 (center column), and the N100/P200 for the next note. Maps with negative values at centro-frontal sites and positive values around the perimeter (A, D, G) are typical of the auditory N100 response. The large parietal positivities in the imagery conditions (E, H) are characteristic of a P300 response. Each map is the activation averaged across seven subjects and 60-90 trials/subject.

there is a large negative focus across centro-frontal electrode sites on top of the head, and a ring of positive voltage at electrodes at more inferior positions of the scalp around the perimeter. Within the same time window, imagining the first of a sequence of notes elicited a topographical pattern in the brain electrical activity that resembled the topographical pattern elicited by the corresponding heard note in the original phrase (Figure 4D, G). No such pattern was elicited in the condition in which subjects were asked to abstain from imagining the continuation of the phrase (Figure 4J). In the imagery conditions, another stable topographical pattern was assumed from 375-480 ms after the time at which the first note was to be imagined (Figure 4E, H). The positive peak above centro-parietal electrodes is characteristic of the P300 component mentioned earlier. Because subjects were expecting the sounds to cease, the presence of the P300 does not indicate an expectancy violation response to an unexpected cessation of input. This interpretation is further supported by the absence of a P300 response in the no-imagery condition (Figure 4K). The presence of the P300 is associated specifically with the imagery task, though it is difficult to assign a further functional role at this time.

Given the relatively recent advent of dense-EEG methods, statistical techniques for quantitative assessment of the similarity or dissimilarity of different topographies have not been well established. Nonetheless, one way to compare topographical maps is to calculate the correlation between them. The similarity of the topographical states across the different conditions was assessed by correlating the topographies in the 3 Imagined condition with corresponding topographies in the other conditions. The temporal evolution of correlations among voltage topographies elicited in the four experimental conditions is depicted in Figure 5 on the next page. Average topographies were computed for successive 100 ms epochs. A 100ms window size reduces the amount of data as much as possible while preserving the most prominent and stable topographical distributions. In order to compare the 3 Imagined and 5 Imagined conditions, the topographies from the two conditions had to be aligned with respect to the onsets of the first imagined events. Figure 5A shows the correlations of topographies among the conditions during the first two notes of the melodies. During these epochs, the acoustical parameters were identical across the conditions. Although the degree of correlation between the 3 Imagined condition and the other conditions varied from time-window to time-window, there were no differences among any of the conditions, as was expected. Figure 5B shows the correlations when the tasks and acoustic stimulation diverged. The two imagery conditions were most highly correlated, whereas the topographies in the 3 Imagined and no-imagery conditions were uncorrelated. Note that the acoustic input in these latter conditions was identical. The correlation between the 3 Imagined and All heard conditions assumed intermediate values, particularly during the first portion of the first imagined note epoch. As a point of reference, the topographies shown in the leftmost column of Figure 4 correspond to the second 100 ms window in Figure 5B.

While the instruction to imagine the continuation of the melody resulted in a clear emitted potential in response to the first imagined note, and the topographical pattern was similar to the N100 elicited by the corresponding heard note, the same pattern was not observed for the subsequent imagined notes. Rather, the topographical activation pattern shifted to a frontal-positive peak around the time that the second note was to

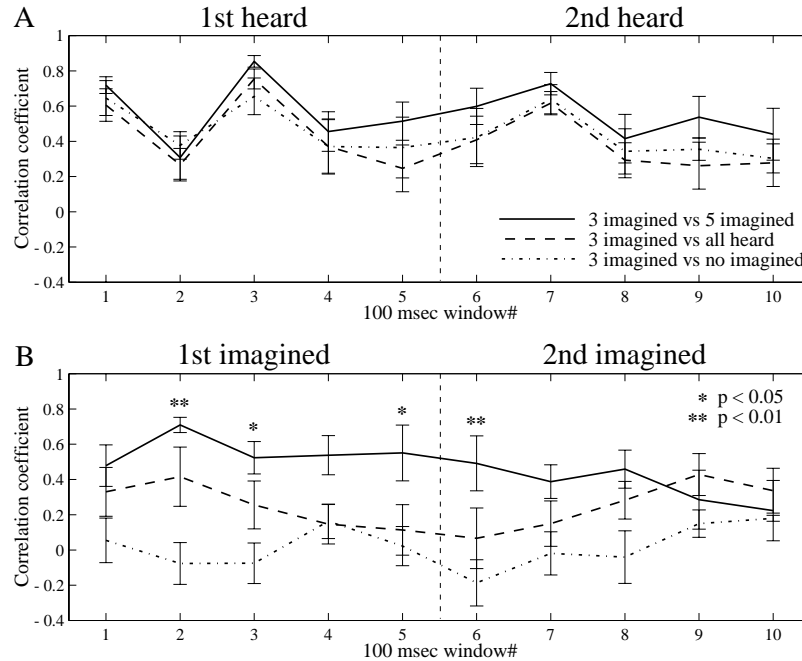


Figure 5. Time-series of the correlations among the experimental conditions. For each subject, the average scalp-topographies in successive 100 ms windows from the 3 Imagined (3I) condition were correlated with corresponding topographies in the other task conditions. The correlations between conditions (averaged across subjects) are plotted in A & B. Error bars indicate the standard error of the mean. A) The solid line shows the correlation of evoked responses during the first and second heard notes in the 3I and 5 Imagined (5I) task conditions. The dashed line shows the correlation between the 3I and All Heard (AH) conditions during the first and second heard notes. The dashed-dotted line shows the correlation between the 3I and No-Imagery (NI) conditions during the first and second heard notes. The acoustic stimulation was identical across task conditions during this 1 second epoch. B) The solid line shows the correlation between the 3I and 5I conditions during the first and second imagined notes in each condition. The dashed line shows the comparison between the first and second imagined notes in the 3I condition and the sixth and seventh heard notes in the AH condition. The dashed-dotted line shows the comparison of the first and second imagined notes in the 3I condition and corresponding silence in the NI condition. Asterisks indicate windows in which there were significant differences in the correlation coefficients between conditions.

be imagined (Figure 4F, I). When no imagery task was performed, no topographical transformations were observed (Figure 4J-L). The presence of a sequence of distinct topographies that was unique to the imagery conditions indicates that the instructions to imagine successive notes result in a measurable series of brain states. The similarity of the initial component during imagery and the auditory evoked potential N100 component suggest that auditory areas may be activated as subjects begin to imagine the continuation of the melody.

If the set of brain areas involved in forming successive mental images is engaged in an iterative manner, one might expect to detect repetitions of voltage topographies at the scalp. The failure to record the same emitted potential to each note in a sequence of imagined notes may stem from technical limitations of the event-related potential (ERP) methodology, e.g. the need to average together many responses in which the phase-jitter of the components one is interested in observing is minimized. Alternatively, the switch from listening to a sequence of notes to imagining a sequence of notes may reorganize the functional relationship of brain areas involved in the task. The pattern of brain activity evoked at the onset of imagery may differ from that generated during more sustained imagery. For example, the response to the first imagined note may represent the interaction of the image with a decaying sensory memory trace of the preceding note. Cowan (1995) reviews evidence for two types of auditory sensory memory. The shorter form lasting several hundred (250) ms from the stimulus may be perceived as a continuation of the stimulus, whereas the other lasts 10-20s and takes the form of a vivid recollection of the sound. The process of image generation during the first note might interact with either form of the auditory sensory memory. Näätänen and Winkler (Näätänen & Winkler, 1999) have recently argued that representations of auditory stimuli that are accessible to top-down processes are formed in the auditory cortex and indexed by the MMN. A correspondence has been proposed between the MMN and the longer lasting form of auditory sensory memory (Cowan, 1995).

If, in the auditory imagery tasks described above, the N100-like topography is due to the interaction of the image generation process with the short-duration sensory component, the same topography should be absent for subsequent imagined notes, as the sensory memory has decayed. If the N100-like topography to the imagined note is the result of an interaction with the longer-duration sensory memory store, one might expect its presence for the remaining imagined notes also. It is possible, however, that either sensory memory store is perturbed by forming an image of the next, different, note in the melody, in which case the N100-like topography may no longer be possible. A slightly different explanation for the absence of an N100-like topography for each imagined note is that the processes at the onset of auditory imagery and those in continued imagery differ and are therefore associated with different voltage topographies. In the case of visual mental imagery, Kosslyn (1994) makes a strong case for dissociating between processes of 'image generation' and 'image maintenance', arguing that the two processes have different brain activation signatures, the latter being closely associated with working memory processes. Clearly, additional experiments are needed to disentangle the pattern of activations observed in the auditory imagery experiment described above.

The neural circuitry of musical imagery

What are the brain circuits that facilitate the formation of musical images, either in the form of expectancies in musical contexts or sequences of imagined notes of a melody? Recent functional neuroimaging studies using PET by Zatorre and colleagues found that regions of the STG, home to the auditory cortex, and the right frontal lobes are activated when subjects mentally scan a familiar melody in order to compare the pitch height of two notes (Zatorre et al., 1996), or continue imagining melody (Halpern & Zatorre, 1999). While directly implicating the auditory cortex and frontal areas in musical imagery, the PET data do not have the temporal resolution to specify how these areas interact in time.

The temporal resolution of EEG and MEG methods can potentially address the issue of temporal interactions of different brain areas, but these methods are hampered by the need to average many trials due to signal-to-noise ratio considerations and the difficulty of inferring intracranial sources of the potentials measured at the scalp. Inferences about the loci of brain activation that give rise to the electrical field measured at the scalp are based on analyses of the topographical distribution of the voltage across the scalp. For instance, the hallmark topography of the N100 is a polarity inversion between negative values at the vertex of the head and positive values at sites around the lower perimeter of head. Models, based on MEG recordings, of equivalent current dipoles that account for such a polarity inversion imply activation of auditory cortical regions along the superior temporal gyrus (Pantev et al., 1995; Verkindt, Bertrand, Perrin, Echallier, & Pernier, 1995). MEG and EEG data recorded during scanning of auditory memory (Kaufman, Curtis, Wang, & Williamson, 1992), and auditory selective attention tasks (reviewed in Näätänen, 1992) all implicate the STG as an area where stored images are compared with sensory input. Thus, the similarity in the topography during the first imagined note and the N100 component of the auditory evoked potential to the corresponding heard note is consistent with the notion that the process of generating an image of the next note in a melody activates the auditory cortex.

Combined neurophysiological and functional neuroimaging work may provide a way of determining whether focal expectancies and sustained imagery of auditory sequences are simply different facets of the same process, i.e. dependent on the same neural architecture. Currently, the available physiological data are too sparse to describe in detail the mechanisms of auditory and musical imagery in the brain. The spatiotemporal activity pattern in the averaged ERPs described above is consistent with suggestions from the functional neuroimaging literature that a circuit subserving musical imagery may consist of auditory cortex (N100 activation in the ERPs), frontal cortex (positive anterior focus, e.g. Figure 4G, 3I), and posterior parietal cortex (P300 activation). However, the process of auditory imagery in a musical context requires a more precise theoretical specification. This specification must account for ways in which domain general functions, such as the numerous forms of memory (long-term, working, echoic), attention, expectancy, and perception interact to imbue mental images with their content.

Acknowledgements

The research was supported by NIH grants GM07257, NS10395, and P50 NS17778-18, and the McDonnell/Pew Center for the Cognitive Neuroscience of Attention at the University of Oregon. I thank several anonymous reviewers for their helpful critiques of an earlier version of this manuscript.

References

- Besson, M., & Faita, F. (1995). An event-related potential (ERP) study of musical expectancy: Comparison of musicians with nonmusicians. *JEP: Human Perception & Performance*, *21*, 1278-1296.
- Besson, M., Faita, F., Czternasty, C., & Kutas, M. (1997). What's in a pause: Event-related potential analysis of temporal disruptions in written and spoken sentences. *Biological Psychology*, *46*(1), 3-23.
- Besson, M., & Macar, F. (1987). An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology*, *24*(1), 14-25.
- Cowan, N. (1995). *Attention and Memory*. New York: Oxford University Press.
- Dierks, T., Linden, D. E., Jandl, M., Formisano, E., Goebel, R., Lanfermann, H., & Singer, W. (1999). Activation of Heschl's gyrus during auditory hallucinations* [see comments]. *Neuron*, *22*(3), 615-621.
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 Component a Manifestation of Context Updating. *Behavioral and Brain Sciences*, *11*(3), 357-374.
- Finke, R. A. (1989). *Principles of Mental Imagery*. Cambridge: The MIT Press.
- Goldman-Rakic, P. S. (1996). The prefrontal landscape: implications of functional architecture for understanding human mentation and the central executive. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences*, *351*(1346), 1445-1453.
- Griffiths, T. D., Jackson, M. C., Spillane, J. A., Friston, K. J., & Frackowiak, R. S. J. (1997). A neural substrate for musical hallucinosis. *Neurocase*, *3*(3), 167-172.
- Halgren, E., Marinkovic, K., & Chauvel, P. (1998). Generators of the late cognitive potentials in auditory and visual oddball tasks. *Electroencephalography and Clinical Neurophysiology*, *106*(2), 156-164.
- Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cereb Cortex*, *9*(7), 697-704.
- Hantz, E. C., Kreilick, K. G., Kananen, W., & Swartz, K. P. (1997). Neural responses to melodic and harmonic closure: An event-related potential study. *Music Perception*, *15*(1), 69-98.
- Janata, P. (1995). ERP measures assay the degree of expectancy violation of harmonic contexts in music. *Journal of Cognitive Neuroscience*, *7*, 153-164.
- Janata, P., & Petsche, H. (1993). Spectral analysis of the EEG as a tool for evaluating expectancy violations of musical contexts. *Music Perception*, *10*, 281-304.
- Kaufman, L., Curtis, S., Wang, J. Z., & Williamson, S. J. (1992). Changes in cortical activity when subjects scan memory for tones. *Electroencephalography and Clinical Neurophysiology*, *82*(4), 266-284.
- Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain Indices of Music Processing: 'Nonmusicians' are Musical. *Journal of Cognitive Neuroscience*, *12*(3), 520-541.
- Kosslyn, S. M. (1980). *Image and Mind*. Cambridge: Harvard University Press.
- Kosslyn, S. M. (1994). *Image and Brain*. Cambridge: The MIT Press.
- Näätänen, R. (1992). *Attention and Brain Function*. Hillsdale: Lawrence Erlbaum Associates.
- Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*(4), 375-425.
- Näätänen, R., & Winkler, I. (1999). The concept of auditory stimulus representation in cognitive neuroscience. *Psychological Bulletin*, *125*(6), 826-859.
- Nunez, P. L. (1981). *Electric Fields of the Brain*. New York: Oxford University Press.
- Pantev, C., Bertrand, O., Eulitz, C., Verkindt, C., Hampson, S., Schuierer, G., & Elbert, T. (1995). Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalography and Clinical Neurophysiology*, *94*(1), 26-40.
- Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations

- in language and music: An event-related potential study. *Journal of Cognitive Neuroscience*, 10(6), 717-733.
- Petsche, H., Richter, P., von Stein, A., Etlinger, S. C., & Filz, O. (1993). EEG Coherence and Musical Thinking. *Music Perception*, 11(2), 117-151.
- Petsche, H., von Stein, A., & Filz, O. (1996). EEG aspects of mentally playing an instrument. *Cognitive Brain Research*, 3(2), 115-123.
- Pflüger, H. J., & Menzel, R. (1999). Neuroethology, its roots and future. *Journal of Comparative Physiology A – Sensory Neural and Behavioral Physiology*, 185(4), 389-392.
- Rappelsberger, P., & Petsche, H. (1988). Probability mapping: power and coherence analyses of cognitive processes. *Brain Topography*, 1(1), 46-54.
- Ruchkin, D. S., Sutton, S., & Tueting, P. (1975). Emitted and evoked P300 potentials and variation in stimulus probability. *Psychophysiology*, 12(5), 591-595.
- Simson, R., Vaughan, H. G., & Ritter, W. (1976). The scalp topography of potentials associated with missing visual or auditory stimuli. *Electroencephalography and Clinical Neurophysiology*, 40(1), 33-42.
- Srinivasan, R., Nunez, P. L., & Silberstein, R. B. (1998). Spatial filtering and neocortical dynamics: estimates of EEG coherence. *IEEE Transactions on Biomedical Engineering*, 45(7), 814-826.
- Sutton, S., Tueting, P., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potential. *Science*, 155(768), 1436-1439.
- Verkindt, C., Bertrand, O., Perrin, F., Echallier, J. F., & Pernier, J. (1995). Tonotopic organization of the human auditory cortex: N100 topography and multiple dipole model analysis. *Electroencephalography and Clinical Neurophysiology*, 96(2), 143-156.
- Verleger, R. (1988). Event-Related Potentials and Memory – a Critique of the Context Updating Hypothesis and an Alternative Interpretation of P3. *Behavioral and Brain Sciences*, 11(3), 343-356.
- Villringer, A. (1999). Physiological changes during brain activation. In C. T. W. Moonen & P. A. Bandettini (Eds.), *Functional MRI* (pp. 3-13). Berlin: Springer-Verlag.
- Weinberg, H., Walter, W. G., & Crow, H. J. (1970). Intracerebral events in humans related to real and imaginary stimuli. *Electroencephalography and Clinical Neurophysiology*, 29(1), 1-9.
- Zatorre, R. J., Evans, A. C., & Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *Journal of Neuroscience*, 14(4), 1908-1919.
- 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.