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

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

Acuity of mental representations of pitch

Petr Janata
Center for Mind and Brain and Department of Psychology, University of California, Davis, California

Address for correspondence: Petr Janata, Center for Mind and Brain, University of California, 267 Cousteau Pl., Davis, CA 95618. pjanata@ucdavis.edu

Singing in one's mind or forming expectations about upcoming notes both require that mental images of one or more pitches will be generated. As with other musical abilities, the acuity with which such images are formed might be expected to vary across individuals and may depend on musical training. Results from several behavioral tasks involving intonation judgments indicate that multiple memory systems contribute to the formation of accurate mental images for pitch, and that the functionality of each is affected by musical training. Electrophysiological measures indicate that the ability to form accurate mental images is associated with greater engagement of auditory areas and associated error-detection circuitry when listeners imagine ascending scales and make intonation judgments about target notes. A view of auditory mental images is espoused in which unified mental image representations are distributed across multiple brain areas. Each brain area helps shape the acuity of the unified representation based on current behavioral demands and past experience.

Keywords: imagery; music; attention; discrimination; event-related potentials

Introduction

Our internal auditory worlds are vibrant. From inner speech to anticipating the upcoming word or note in a familiar song that we are listening to, our minds maintain an "auditory" world that exists both independently and in interaction with the auditory world around us.

Before proceeding to discuss the acuity of mental auditory images, it is useful to discuss what is meant by an auditory image or mental representation of a sound. I find it most useful to start with a broad definition, such as any sound that you hear playing in your mind. A fundamental distinction can be drawn between external (physical) sources of mental representations of sound and internal (mental) sources of those representations. A slightly more specific definition can be fashioned at the neural level such that the neural activity associated with representations of external sources is contemporaneous with the external source and derives from activity ascending the auditory pathway from the ear, whereas representations deriving from internal sources need not satisfy either of those constraints. I use the term *auditory image* synonymously with mental representation of a sound.

Expectation, covert orientation of attention, and imagery

The focus of this paper is on the acuity of auditory images of pitch that are formed and maintained in a number of different task contexts. I argue that it is parsimonious to consider a unified concept of auditory image across task contexts rather than separate types of auditory images for each of the different types of tasks that in some way instantiate and cause listeners to maintain or manipulate auditory images (Fig. 1).[1] Examples of different types of tasks in which auditory images must be maintained are target detection tasks, auditory working memory tasks, and traditional imagery tasks, such as imagining a familiar melody. While the informational sources from which the auditory image derives may be different in each of those tasks, the auditory image nonetheless becomes available for comparison with mental representations of sensory auditory input so that the listener can decide whether the two mental representations are the same. One might argue that an auditory image should necessarily be associated with a single brain region or population of neurons within that brain region. In such a view, there would be a different auditory image associated with each

doi: 10.1111/j.1749-6632.2011.06441.x

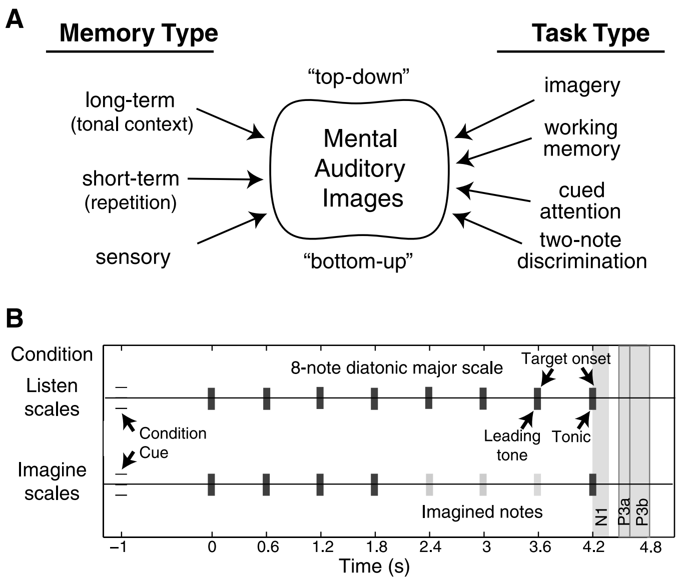

Figure 1. (A) A schematic view of the idea that mental auditory images might be regarded as distributed representations that are subject to influence by different types of memory processes associated with different brain regions as a function of the demands of any particular task. This view is distinct from one in which different types of auditory images are associated with each of the different task types. (B) Diagram of the cued-attention (listen scales) and imagery tasks (imagine scales) used to examine acuity of target tones. Adapted with permission from Ref. 1 (with permission from the Acoustic Society of America) and Ref. 14 (with permission from Elsevier).

successive population of neurons that participates in the representation/maintenance of the auditory image. Thus, even though behavioral responses are made as a consequence of comparing mental and sensory representations in the task types mentioned previously, implying a single locus for the auditory image, the possibility of different task-specific sources of the mental auditory images would imply that different auditory images are maintained in different areas of the brain, before being broadcast to the brain area in which the mental and sensory representations are compared. A contrasting view is one in which the neural representation of a single auditory image is delocalized, that is, distributed across multiple brain areas. In this view, each task-relevant brain area may assist in maintaining the image, and the shaping of the image by any of those areas can influence the neural representation at the other areas. I elaborate this idea for the case of pitch.

Comparisons of pitch acuity across task types

I consider three types of tasks that vary in the degree of sensory support that they provide for the auditory image that is to be compared with a probe note so that an intonation judgment can be made about the probe note. Extensive reports of experiments using these tasks have been presented elsewhere.[1,2] The tasks are as follows.

(1) A basic *two-tone discrimination task* in which two pairs of two 250 ms complex harmonic tones are presented in quick succession (600 ms stimulus-onset asynchrony) and the listener must indicate in which of the two pairs the two tones differ. This task provides the greatest degree of sensory support, as the exact information that is to be compared is maintained in sensory memory for a brief period of time. While this task would not be considered an imagery task in the traditional sense of imagining a melody, there is nonetheless a requirement to maintain a mental image of a just-heard sound in memory.

(2) A *cued-attention task* in which an expectation for a probe note is elicited by the preceding notes of an ascending major scale. Two scales are heard on each trial, and participants must indicate which of the two scales contained the mistuned terminal probe note. This task

provides an intermediate degree of sensory support in the sense that although a note corresponding to the probe note has not yet been heard—that is, there is no direct sensory representation of it—the sequence of notes and associated sensory representations leading up to it is highly predictive and can potentially support the formation of an accurate mental image to coincide with the arrival of the probe tone. While this task is not a traditional imagery task, there is nevertheless a requirement to construct a mental image of a sound that has not yet been heard.

(3) An *imagery task* that is similar to the cued-attention task, but differs in that only the initial four notes of the major scale are heard and the rest must be imagined, so that a probe note that occurs at the time it would have occurred, had the entire scale been played, can be judged. This task provides very little sensory support for the auditory image.

The cued-attention and imagery tasks can be manipulated further by varying the number of keys that the scales are presented in and by varying the terminal probe note about which the intonation judgment is made. Holding the key constant from trial to trial allows a short-term absolute memory for the target notes to build up. Placement of the probe note at the leading tone (an unstable scale position) versus the tonic (the most stable scale position) allows an estimate to be made of the influence of long-term memory for tonal knowledge on the momentary auditory image.

These brief descriptions of the task types and additional mnemonic manipulations hopefully serve to illustrate the difficulty in specifying precisely what the criteria are for labeling a process as *auditory imagery* and the elements that are being operated on as *auditory images*.

When the tasks are structured as two-alternative forced choice (2AFC) tasks, the exact amount of mistuning of the probe notes is varied from trial to trial so that each listener's intonation threshold can be determined for each task type. Not surprisingly, the smallest thresholds are obtained in the two-tone discrimination task. Interestingly, thresholds in the cued-attention task remain comparable to two-tone thresholds for many, but not all listeners (Fig. 2A). Moreover, when the key in which the scale is played is varied across trials, thresholds tend to become worse. These effects are further exacerbated in the imagery task when sensory support is further removed, even though some individuals

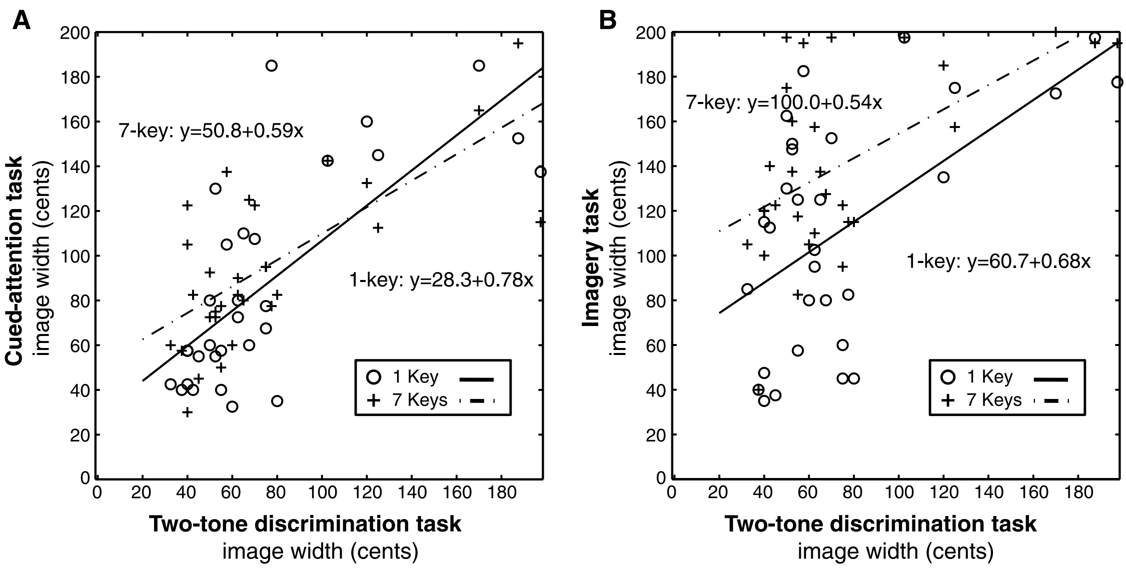

Figure 2. Correlations between auditory image widths in a two-tone discrimination task and image widths for the last note of an ascending major scale in a cued-attention task (A) and imagery task (B). The influence of short-term memory build-up was assessed by contrasting performance on blocks of trials in which the key was held constant and blocks in which the key varied randomly among seven keys. Reprinted with permission from Ref. 1.

perform as well in the imagery task as they do in the two-tone discrimination task (Fig. 2B). These observations indicate that auditory images are shaped by multiple mechanisms, that comparable quality of the image can be obtained whether it arises directly from sensory input or from internal sources, and that individuals vary considerably in the degree to which they utilize these different mechanisms when forming and maintaining auditory images.[1,2] Although considerable variance in pitch image acuity is observed in individuals with no musical training, with some untrained individuals exhibiting pitch images that are as precise as those of trained individuations, overall, musical training significantly improves acuity in both cued-attention and imagery conditions.[1,2]

Electrophysiological measures of image acuity

The decisions made by listeners in the behavioral experiments summarized earlier depend on the comparison of an internally maintained auditory image (top–down) with incoming sensory information (bottom–up). The suggestion that internal auditory images might be maintained across a number of brain areas implies that comparisons between top–down and bottom–up representations could also transpire at and across multiple brain areas. Particular attention has been paid to the auditory cortex. Mismatch negativity (MMN) experiments have demonstrated a role of the auditory cortex in the detection of acoustic deviants outside the focus of attention,[3] though auditory representations in earlier stages of the ascending auditory pathway have been shown to be modulated by attention or the preceding stimulus context.[4] Here, I also focus on the auditory cortex, or, more specifically, electrophysiological signatures of secondary auditory cortex processing with a particular emphasis on the N1 component of the auditory evoked potential.[5,6]

Both attentional cuing and auditory imagery influence activity in auditory cortical areas as demonstrated by electrophysiological, positron emission tomography (PET), and functional magnetic resonance imaging (fMRI) studies.[7–11] Imagining notes in a melody establishes a voltage distribution across the scalp with a topography that resembles that of the N1, and can be modulated focally in time to appear as an evoked potential.[12] MMN responses have also been observed in response to imagined notes in familiar melodies.[13]

The N1

Utilizing the intonation judgment tasks described earlier, we demonstrated that N1 amplitude can serve as a marker of auditory image acuity.[14] The amplitude of the N1 depends on the rate at which successive tones are presented, and decreases rapidly when multiple tones are presented.[15] If the tones cease, the amplitude of the N1 in response to an eventual tone is large again. What might happen to the amplitude of the N1 to a probe tone given the task of imagining an ascending scale? If the notes in an ascending scale are imagined successfully and each of those imagery events activates the auditory cortex, the N1 to the probe note should be small, as if each of the intervening imagined notes had actually been heard. By contrast, if no imagery task is performed, or if a participant who is supposed to engage in imagining the sequence of tones does not do so successfully, the N1 amplitude in response to a probe tone about which a judgment must be made should be large because it is effectively the first input to the auditory cortex in several seconds. Figures 3A and 3B illustrate this effect. Those individuals who succeeded in forming accurate auditory images showed smaller N1 amplitudes than those who did not, suggesting that by virtue of imagining the tones leading up to the probe tone they instantiated representations of those tones in the auditory cortex.

These observations parallel the findings of studies in which the conceptual framework has centered on the orienting of attention. Visual and auditory tasks in which expectations are manipulated in time and another feature dimension, for example, space or pitch, result in faster reaction times and reduced N1 amplitudes when expectations are met.[16] Utilizing a very similar paradigm, in which all the notes of an ascending scale across two octaves were heard except for the two notes preceding the final tonic, Lange[16] observed reductions in N1 amplitude as temporal and pitch expectations were met. Although no explicit instruction was provided to imagine the missing notes, the results illustrate that regularity in the stimulus structure implicitly facilitates the instantiation of a mental pitch representation at a specific moment in time that can be compared with the incoming sensory representation.[17,18]

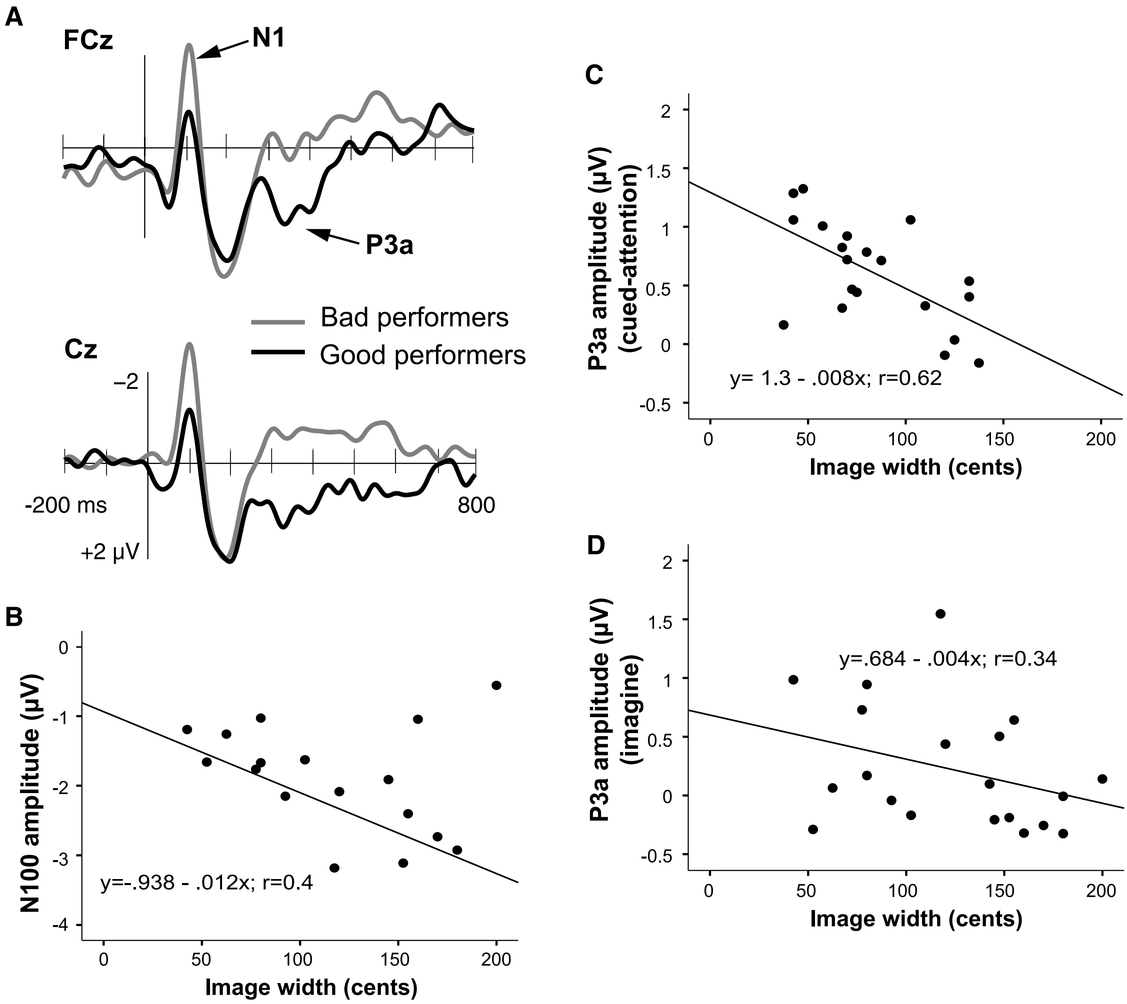

Figure 3. Event-related potential (ERP) indexes of auditory image acuity. (A) Listeners with poor pitch image acuity (bad performers) exhibited a larger N1 component of the auditory-evoked potential, indicative of a failure to imagine the notes leading up to the final note properly. (B) N1 amplitudes increased (became more negative) as image acuity decreased. The P3a component, a measure of deviance detection, was larger in those listeners who formed more accurate images in the cued-attention (C) and imagery conditions (D). Adapted from Ref. 14 with permission.

Auditory images within a perception–action cycle framework

I suggested earlier that the role of establishing and maintaining auditory images might be shared across multiple brain areas depending on the particular task context. Why, however, should such sharing take place and what would the other areas be that participate in the shaping of the representation? Paralleling the work of others, I believe it is useful to think of auditory images in terms of a perception–action cycle.[19] In simplest terms, the perception side of the perception–action cycle might be associated with representing bottom–up sensory information, while the action side mediates the projection of top–down information, not only to motor areas that drive effectors, but also to sensory areas in the form of "efference-copy" or expectations more generally.[20–22] The concept of cyclicality becomes pertinent whenever a sequence of events/actions comes into play, in that an image/action sets the stage for an upcoming image/action, often shaped by the interaction of the image with sensory input. Musical examples include both covert and overt singing of a melody with or without actual accompaniment, or adjusting intonation across time while singing a single note.

P3a

The P3a is an event-related potential (ERP) component that appears as a frontocentral positivity along the midline in response to novel auditory stimuli.[23] The P3a can even appear in response to musical events, for example, chords, that are incongruent with a highly expected and imagined chord.[24] I believe the P3a response observed in the cued-attention and imagery tasks may reflect activity of an incongruity detector within the perception–action cycle. As with the N1, its amplitude was related to image acuity in both the cued-attention and imagery conditions: listeners with more precise mental images had a larger P3a amplitude (Fig. 3C and D). Note, that in contrast to the reduction of N1 amplitude for the probe note, presumably due to activation of the auditory cortex by the preceding imagined notes, the P3a increased in amplitude, consistent with its interpretation as a marker of deviance detection. Attributing to the P3a, a relationship to the perception–action cycle stems from its focal frontocentral midline topography with a presumed source in the anterior cingulate cortex,[25] the ACCs known role in conflict and action monitoring[26,27] and auditory target detection.[23,25,28] However, given the location of the probes at the end of a note sequence (chain of perception–action cycle events), it cannot be ruled out that the P3a simply reflects the conscious marking of a probe note as a deviant note for a subsequent response, rather than an error signal that would influence the production of the next note in a sequence.

Is the concept of a delocalized auditory image useful?

Arguing for a view of auditory images as neural representations that span several brain areas involved to varying degrees in perception and action is somewhat perilous because it risks making the concept of an auditory mental representation unspecific and difficult to identify in the brain. Many might argue that it is easier to regard the representations of auditory objects in terms of discrete processing stages that are accessible to top–down influences and conscious awareness to varying degrees.[29] For example, in the context of attention-orienting paradigms, one can ask in what part of the brain is activity modulated when an expectation is established for a particular auditory feature. However, the processes and focal representations in the attention/expectation case are traditionally viewed as different than the distributed representations maintained during an auditory working memory task across the brain areas that constitute the phonological loop. In attention/expectation paradigms, the emphasis is on a single locus in feature space at a certain point in time, with an emphasis on perception, whereas in the working memory case there is an emphasis on the maintenance of multiple objects over time across both sensory and motor areas. Moreover, in the context of natural behaviors, such as controlling intonation of a pitch while singing, or singing along with a piece of familiar music in one's mind, it is very difficult to specify where the sensory representation at one moment in time stops and the motor representation of the next moment in time begins. In this regard it seems more parsimonious to consider a delocalized sensorimotor representation that is then subject to influences from and accessible to multiple other brain regions as a function of the broader behavioral context in which the sensorimotor representation is being created. A delocalized representation should not imply a lack of constraints, however.

Anatomical connections impose constraints on the activity in the connected areas insofar as the pattern of activity in one area will be correlated with the pattern of activity in the other. In the case of auditory images, the arcuate fasciculus is a connection that binds sensory (lateral temporal) and motor (lateral prefrontal) brain areas and supports verbal and presumably tonal working memory.[30,31] It is also associated with the ability to perceive pitch sequences, such that individuals with poor ability to discriminate pitch contours (amusics) have a thinner arcuate fasciculus in the right hemisphere.[32] Interestingly, amusics show a dissociation in their representations of pitch contours, with more accurate, albeit not fully tuned, representations evident if they are probed with a production task instead of a perception task.[33] This observation serves as a cautionary note that the inferred quality of a mental representation may depend on the behavioral context in which it is probed.

If a task involves a decision, the accuracy of which depends on pitch image quality, the behavioral and neural acuity measures simply reflect the multiple influences that have combined to shape the top–down and bottom–up representations up to the point that they interact. However, if the behavioral or neural measure probes the quality of the auditory image at some other location in the brain, the inferred quality may be different. For example, take two individuals who are asked to imagine one of the songs of their favorite band that they heard the previous evening at a concert they attended together. They are both performing the same imagery task. Both of them might recall the same song equally well in terms of the melodic contour and the words if probed with a production task, whereas only one of them might be able to determine accurately whether sounded probe tones are "in tune" with the imagined notes. Rather than considering that one of these individuals deploys more mental auditory images than the other during the imagery task, it may make sense to consider mental auditory images as unified mental constructs. However, the exact appearance of the image at any point in the brain nonetheless depends on the functional perspective of the brain area engaging the construct and the individual's success in engaging the functions of that brain area. Whether such a view is useful from an experimentalist's perspective remains to be seen. One advantage of it is that it diminishes the drive to consider every different type of task that in some way manipulates an auditory representation to be giving rise to a different type of represented entity.

Conflicts of interest

The author declares no conflicts of interest.

References

1. Navarro Cebrian, A. & P. Janata. 2010. Influences of multiple memory systems on auditory mental image acuity. *J. Acoust. Soc. Am.* **127:** 3189–3202.
2. Janata, P. & K. Paroo. 2006. Acuity of auditory images in pitch and time. *Percept. Psychophys.* **68:** 829–844.
3. Näätänen, R. 1992. *Attention and Brain Function*. Lawrence Erlbaum Associates. Hillsdale, NJ.
4. Suga, N. *et al*. 2000. The corticofugal system for hearing: recent progress. *Proc. Natl. Acad. Sci. USA* **97:** 11807–11814.
5. Näätänen, R. & T. Picton. 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* **24:** 375–425.
6. Pantev, C. *et al*. 1995. Specific tonotopic organizations of different areas of the human auditory cortex revealed by simultaneous magnetic and electric recordings. *Electroencephalogr. Clin. Neurophysiol*. **94:** 26–40.
7. Halpern, A.R. & R.J. Zatorre. 1999. When that tune runs through your head: a PET investigation of auditory imagery for familiar melodies. *Cereb. Cortex.* **9:** 697–704.
8. Jäncke, L., S. Mirzazade & N.J. Shah. 1999. Attention modulates activity in the primary and the secondary auditory cortex: a functional magnetic resonance imaging study in human subjects. *Neurosci. Lett.* **266:** 125–128.
9. Woldorff, M.G. & S.A. Hillyard. 1991. Modulation of early auditory processing during selective listening to rapidly presented tones. *Electroencephalogr. Clin. Neurophysiol.* **79:** 170–191.
10. Woldorff, M.G. *et al*. 1993. Modulation of early sensory processing in human auditory cortex during auditory selective attention. *Proc. Natl. Acad. Sci. USA* **90:** 8722–8726.
11. Zatorre, R.J. *et al*. 1996. Hearing in the mind's ear: a PET investigation of musical imagery and perception. *J. Cogn. Neurosci.* **8:** 29–46.
12. Janata, P. 2001. Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain Topogr.* **13:** 169–193.
13. Herholz, S.C. *et al*. 2008. Neural basis of music imagery and the effect of musical expertise. *Eur. J. Neurosci.* **28:** 2352–2360.
14. Navarro Cebrian, A & P. Janata. 2010. Electrophysiological correlates of accurate mental image formation in auditory perception and imagery tasks. *Brain Res.* **1342:** 39–54.
15. Woods, D.L. & R. Elmasian. 1986. The habituation of event-related potentials to speech sounds and tones. *Electroencephalogr. Clin. Neurophysiol*. **65:** 447–459.
16. Lange, K. 2009. Brain correlates of early auditory processing are attenuated by expectations for time and pitch. *Brain Cogn.* **69:** 127–137.
17. Barnes, R. & M.R. Jones. 2000. Expectancy, attention, and time. *Cogn. Psychol.* **41:** 254–311.
18. Jones, M.R. *et al*. 2002. Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychol. Sci.* **13:** 313–319.
19. Fuster, J.M. 2001. The prefrontal cortex: an update—time is of the essence. *Neuron.* **30:** 319–333.
20. Haggard, P. & B. Whitford. 2004. Supplementary motor area provides an efferent signal for sensory suppression. *Cognit. Brain Res.* **19:** 52–58.
21. Baess, P., T. Jacobsen & E. Schroeger. 2008. Suppression of the auditory N1 event-related potential component with unpredictable self-initiated tones: evidence for internal forward models with dynamic stimulation. *Int. J. Psychophysiol.* **70:** 137–143.
22. Lange, K. 2011. The reduced N1 to self-generated tones: an effect of temporal predictability? *Psychophysiology* **48:** 1088–1095.
23. Polich, J. 2007. Updating P300: an integrative theory of P3a and P3b. *Clin. Neurophysiol.* **118:** 2128–2148.
24. Janata, P. 1995. ERP measures assay the degree of expectancy violation of harmonic contexts in music. *J. Cogn. Neurosci.* **7:** 153–164.
25. Crottaz-Herbette, S. & V. Menon. 2006. Where and when the anterior cingulate cortex modulates attentional response: combined fMRI and ERP evidence. *J. Cogn. Neurosci.* **18:** 766–780.
26. Gehring, W.J. & R.T. Knight. 2000. Prefrontal-cingulate interactions in action monitoring. *Nat. Neurosci.* **3:** 516–520.
27. Carter, C.S. & V. van Veen. 2007. Anterior cingulate cortex and conflict detection: an update of theory and data. *Cognit. Affect. Behav. Neurosci.* **7:** 367–379.

28. Kiehl, K.A. *et al*. 2001. Neural sources involved in auditory target detection and novelty processing: an event-related fMRI study. *Psychophysiology.* **38:** 133–142.
29. Näätänen, R. & I. Winkler. 1999. The concept of auditory stimulus representation in cognitive neuroscience. *Psychol. Bull.* **125:** 826–859.
30. Catani, M. & M. Mesulam. 2008. The arcuate fasciculus and the disconnection theme in language and aphasia: history and current state. *Cortex* **44:** 953–961.
31. Koelsch, S. *et al*. 2009. Functional architecture of verbal and tonal working memory: an fMRI study. *Hum. Brain Mapp.* **30:** 859–873.
32. Loui, P., D. Alsop & G. Schlaug. 2009. Tone deafness: a new disconnection syndrome? *J. Neurosci.* **29:** 10215–10220.
33. Loui, P. *et al*. 2008. Action-perception mismatch in tone-deafness. *Curr. Biol.* **18:** R331–R332.