Dynamic Aspects of Musical Imagery

Andrea R. Halpern
Bucknell University, ahalpern@bucknell.edu

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*

Dynamic aspects of musical imagery

Andrea R. Halpern

Psychology Department, Bucknell University, Lewisberg, Pennsylvania

Address for correspondence: Andrea R. Halpern, Psychology Department, Bucknell University, Lewisburg, PA 17837. ahalpern@bucknell.edu

Auditory imagery can represent many aspects of music, such as the starting pitches of a tune or the instrument that typically plays it. In this paper, I concentrate on more dynamic, or time-sensitive aspects of musical imagery, as demonstrated in two recently published studies. The first was a behavioral study that examined the ability to make emotional judgments about both heard and imagined music in real time. The second was a neuroimaging study on the neural correlates of anticipating an upcoming tune, after hearing a cue tune. That study found activation of several sequence-learning brain areas, some of which varied with the vividness of the anticipated musical memory. Both studies speak to the ways in which musical imagery allows us to judge temporally changing aspects of the represented musical experience. These judgments can be quite precise, despite the complexity of generating the rich internal representations of imagery.

Keywords: auditory imagery; memory; emotion; anticipation; music

Introduction

The introduction to this symposium referred to some of the characteristics of mental imagery as memory representations. People report these experiences as capturing perceptual aspects of the event or object, but in addition to the phenomenology, numerous reports suggest that people can make accurate decisions about objects or experiences that are only imagined.[1,2] Although auditory imagery is studied less often than visual, people commonly claim that they can recollect music in such a way as to simulate the actual hearing of the tune. For instance, they can reproduce timing and pitch relationships as well as evoke the sound qualities of musical instruments (reviewed in Ref. 3).

The relationship between heard and imagined sounds is of interest for both theoretical and applied reasons. The theoretical interest lies in the ways in which mental representation both preserves but also changes experience. Memory is a dynamic process, and memory researchers have long since abandoned thinking of memory as an actual audio or video recorder. Not even "flashbulb" memories are veridical,[4] and many psychologists caution against the reification of eyewitness testimony in legal systems.[5] However, as noted previously, the extent to which imagery representations maintain aspects of experience that are not easily verbalizable or even obviously codable suggests dissociations between the content and labeling of memories. For instance, I tested people's abilities to reproduce the opening pitch of familiar songs, like "Happy Birthday." Although people differed in what pitch they chose, individuals were quite reliable in reproducing or choosing the same opening pitch for the tune on two occasions several days apart.[6] But none of these people had absolute pitch ability; that is, none could name the notes they were selecting, thus they apparently were accessing a perceptual code.

Understanding auditory imagery is also useful in several applied domains. Consider mental practice. Musicians can suffer from crippling dystonias or other medical conditions that can shorten a career.[7] Relieving some of the motor stress by engaging in mental practice (which involves both auditory and motor imagery) could help alleviate some of these concerns, and indeed mental practice is recommended for even healthy musicians or athletes when physical practice is not practical. Knowing how

doi: 10.1111/j.1749-6632.2011.06442.x

imagery processes work could lead to recommendations about how to implement these strategies, as well as offering guidance to teachers for students who have more and less vivid imagery.[8] In addition, the emerging field of neural prosthetics could benefit from understanding the neural underpinnings of these vivid mental representations.[9]

My focus in this symposium was on two aspects of auditory imagery that could be considered *dynamic* in the sense that the imagery representation is required to be updated continuously to serve the purpose at hand. I use this term to contrast with the more static representations required in some of my previous auditory imagery studies. For instance, in the study on mental pitches mentioned earlier,[6] the pitch selection does not require any processing past the initial memory retrieval. Likewise, comparing the sound of imagined musical instruments requires a memory retrieval of the sound of a violin or saxophone, and maintenance long enough to generate a similarity rating, but no additional operations.[10] However, to accomplish a wider variety of tasks, it is likely we use dynamic imagery processes to track time-varying information, make novel or creative judgments, generate predictions, or to represent multiple aspects of information simultaneously.

I will illustrate dynamic auditory imagery processes with two studies that I was recently involved with. In the first case, Lucas, Schubert, and I[11] examined how well people could extract emotional judgments of music in real time as they listened to, and then imagined, a familiar piece of music. In this case, even the perceptual judgment is dynamic in that listeners must monitor the piece as well as make a corresponding judgment in continuous, real time. Dynamic imagery processes are even more important when the piece is being generated internally.

In the second case, I will describe a study carried out in conjunction with Leaver and Rauschecker[12] that looked at people's ability to predict an upcoming melody based on having heard a melody previously associated with it. We call this "anticipatory imagery," in that the imagery is evoked only in response to the cue, but then must be retrieved in a prospective manner, as the second melody, in this case, is not actually played. The analogy from real life would be anticipating the next track on one's favorite album or the next movement in a familiar symphony as the previous segment ends. In fact, the first experiment in that paper used favorite CDs as stimuli. However, I was only involved in the second study that used paired associate learning of tunes, so I will be describing that one. And I will be making a link between these, as it is entirely possible that the tracking displayed in the emotion judgments was facilitated by anticipation.

During talks, I sometimes ask the audience if they ever deliberately imagine music so as to regulate their moods; I usually get wide agreement with this statement. Even persistent musical memories, or "earworms," seem to consist of preferred and pleasant music.[13] So it would not be surprising if people could extract emotional judgments from imagined music, at least on a global basis. What was less evident to us was whether emotional judgments could be made in real time as the music was actually proceeding (either real or imagined). Some preliminary evidence by Schubert *et al.*[14] suggested a mixed answer. They asked a pianist to listen to a recording of himself and make continuous judgments of valence and arousal in the piece. He then did the same thing as he imagined the piece. The two response profiles were similar, but the pianist began lagging in the responses to the imagined relative to the heard piece. This is consistent with imagery being a fairly expensive cognitive process in that it requires full consciousness and responses in typical mental imagery experiments are often rather slow compared to other kinds of tasks. However, we did not know if these results would generalize beyond a case study.

We therefore asked undergraduate students to participate in two tracking studies. We tested musicians (with at least 8 years of private instrument playing, averaging about 10.5 years) partly to insure that our participants would have the fine motor control needed for continuous response, but also because of the music we used. After considerable discussion, we decided that classical music would be a good genre to test because we needed pieces that varied on the selected dimensions of valence, arousal, and emotionality over the time span of about one minute. We decided that music from genres more familiar to nonmusicians, such as pop, rock, and movie scores, tend not to vary as much over short durations. Furthermore, we needed the pieces to be highly familiar to our listeners for the imagery condition, which again pointed to students with extensive musical training.

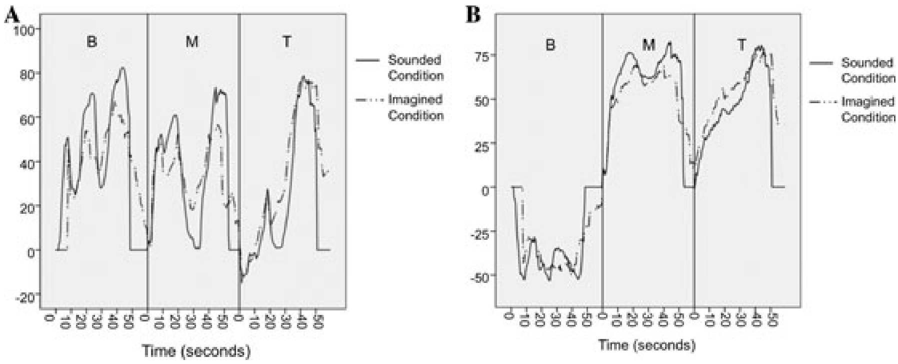

Figure 1. Mean sounded and imagined time-series responses. (A) Mean arousal response for Beethoven, B; Mozart, M; and Tchaikovsky, T. (B) Mean valence response for Beethoven, B; Mozart, M; and Tchaikovsky, T. Figure reprinted with permission from Ref. 11.

We selected approximately the first minute of Beethoven's Symphony no. 5 in C minor," op. 67, Tchaikovsky's "Waltz of the Flowers" from the *Nutcracker Suite*, op. 71, and the "Allegro" from Mozart's *Serenade in G*, K. 525 (*Eine Kleine Nachtmusik*). These were very familiar to our participant pool, had the requisite changes in emotional characteristics in their opening sections, and were rather different from each other. For instance, the Beethoven piece was in a minor key and the others in major. [The figures refer to these pieces by the first letter of the composers: B, T, and M.]

We trained the participants to use a continuous response recorder. As the piece played (or as they imagined it), they moved a mouse on a display; mouse positions were recorded twice per second. In the first study, the task was to track valence on the x-axis simultaneously with arousal on the y-axis. Valence was explained as the range of negative to positive emotion and arousal as the range of sleepy to excited emotion. Experiment 2 used the single dimension of emotionality (in the range of high to low). Several trials of practice preceded the real trials. In the sounded condition, the piece was played over speakers. In the imagined condition, the participants heard the first few notes of the to-be-imagined piece and also could refer to a musical score. We analyzed data from 17 participants in experiment 1, and 11 participants (4 from the previous study) in experiment 2, but I will discuss the results together. We also gave participants a tapping synchronization task, wherein they heard two measures of an isochronous beat at 160 bpm, then had to keep the beat steady without the cue for 40 measures.

Figures 1 and 2 show the averaged response profiles for the three dimensions: the solid line indicates the mouse position over time in the sounded condition and the dashed line in the imagined condition. As is evident, the response profiles on average were very similar in both conditions on all three dimensions, both in pattern (the same peaks and valleys) and timing. Quantitative analysis showed that the two profiles were highly correlated in most participants. A cross-correlation function analysis allowed us to determine the lead or lag at which the highest profile correlations were obtained. We found that depending on the dimension, 85–100% of the analysis showed significant correlation at some lead or lag; and the modal lead–lag was 0 samples. That is, most people most of the time tracked the emotional profile similarly, in both pattern and timing, in the sounded and imagined conditions. The small leads and lags shown by some people in some conditions were actually within the small error range on the tapping synchronization task. We also related many of the peaks and valleys in the profiles to predesignated "turning points" in the music. For instance, we predicted an increase in arousal when loudness increased in the music. Over the eight identified turning points across the three pieces, 84% of the participants responded in the predicted way in their tracking responses.

Thus, we concluded that musicians, at least, could indeed make temporally fine-grained judgments. In addition, the emotional message of music can be

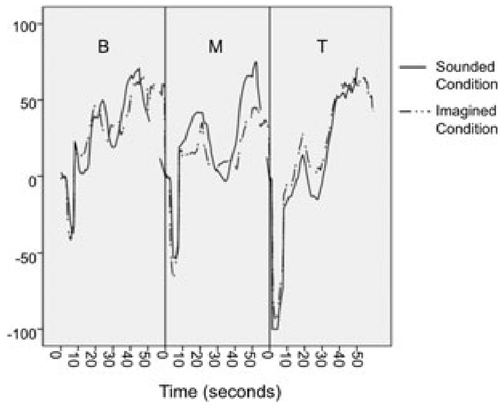

Figure 2. Mean emotionality response for Beethoven, B; Mozart, M; and Tchaikovsky, T. Figure reprinted with permission from Ref. 11.

decoded, and presumably enjoyed in real time, even in a mental representation. In that context, it is interesting to hear the words of Romel Joseph, a violinist buried in the rubble of his collapsed music school for 18 hours before being rescued from the 2010 Haiti earthquake. He later related how running through familiar pieces in his mind helped him cope: "For example, if I perform the Franck sonata, which is [sic] 35 minutes long in my honors recital at Juilliard, then I would bring myself to that time. That allows me ... to mentally take myself out of the space where I was. . ."[15]

In this study, we were somewhat surprised that the tracking in the imagined condition did not slow down. Although carried out before the one just described, the other study I am using to illustrate dynamic auditory imagery could help explain the outcome. Perhaps the participants were anticipating the notes in the imagined music, which counteracted what would otherwise have been a slowdown in response because of the exigencies of generating the auditory image.

As mentioned earlier, Leaver, Rauschecker, and I built on a prior study in Rauschecker's lab that examined neural correlates of anticipation of the next track in familiar CDs.[12] To reduce individual differences and to gain more experimental control, we taught participants seven otherwise unrelated pairs of unfamiliar but melodious tunes. The training session involved repeated study of a pair, with the goal that when presented with the first member of the pair, everyone could conjure up the image of the second member of the pair. The scanning session involved some additional training before entering the fMRI scanner, and after the first run, we asked them to overtrain by hearing the pairs for about 10 additional minutes. We tested 10 people who had a minimum of two years of musical training (averaging about 6.5 years).

Although in the scanner, trials were of three types: silent baseline, anticipatory imagery trials in which

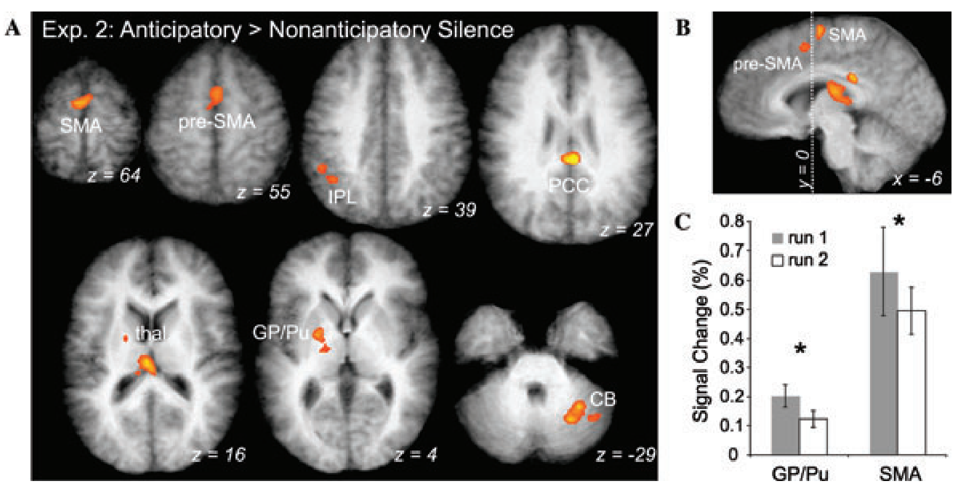

Figure 3. (A) Significant activation associated with anticipatory imagery (AS > NS). Areas include cerebellum, globus pallidus/putamen (GP/Pu), thalamus, posterior cingulate cortex (PCC), presupplementary motor area (pre-SMA), and SMA proper. (B) Sagittal view of medial frontal activation. Dotted line indicates Talairach coordinate axis, $y = 0$, separates pre-SMA and SMA proper. (C) ROI analysis reveals percentage signal change differences in the AS conditions between run 1 (shaded) and run 2 (white) ($^{*}P < 0.05$). ROIs were defined by analysis shown in A. Error bars indicate SEM. Figure reprinted with permission from Ref. 12.

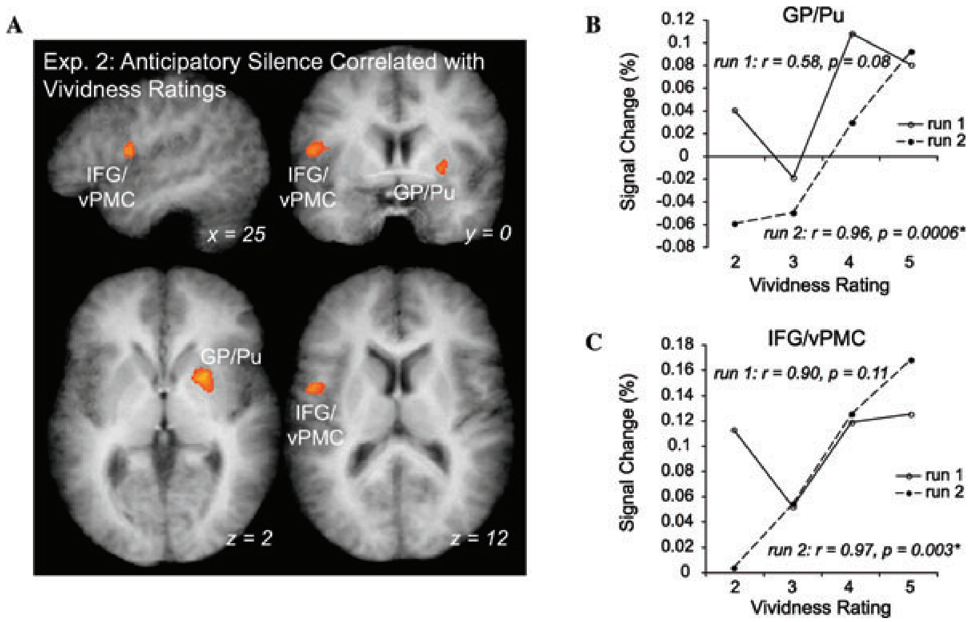

Figure 4. (A) Areas for which BOLD signal is correlated with vividness rating in the familiar silence condition (GP/Pu, inferior frontal gyrus, ventral premotor cortex (IFG/vPMC)). (B, C) Graphs show correlations between percentage signal change and vividness ratings for ROIs resulting from the parametrically weighted GLM analysis in A. Figure reprinted with permission from Ref. 12.

the first member of a learned melody pair was presented and people were asked to imagine the associated tune, and nonanticipatory trials. On these trials, a novel tune was presented so that participants could not imagine any associated tune. The acquisition of the fMRI data occurred in the silence after the tune had been presented. The silence after a learned tune ought to have engendered anticipation of the next tune, hence we called these anticipatory silences (AS). In these trials, we asked people to rate the vividness of their imagery for the anticipated tune, on a 1 (no image) to 5 (very vivid) scale. The silence after a novel tune should not have engendered any particular auditory imagery, hence nonanticipatory silence (NS). For details of the scanning parameters and analyses, see Ref. 12.

Figure 3 shows the contrast between AS and NS trials. A number of brain areas were more active in the anticipatory response, and of particular interest were several areas normally associated with motor sequence learning, such as the supplementary motor areas (SMA and pre-SMA) and the globus pallidus and putamen in the basal ganglia. These two areas showed decreased signal after overtraining, as shown on the right side of the figure. Figure 4 shows two brain areas where activity in AS trials was directly correlated with the vividness ratings: the basal ganglia areas just mentioned and a ventral premotor area. As seen in the graphs, correlations were strongly positive and even linear in the second run (correlations in the first run were positive but not significantly so). That is, for every increase in vividness ratings, a proportional increase only in these two areas occurred in the hemodynamic response of AS trials.

This experiment showed how the brain response changes as anticipatory imagery becomes stronger. When people know fairly well what tune will follow a cue tune (i.e., during the first run), sequence learning areas are fairly active. They also rate the vividness of the upcoming target tune as being moderately vivid (a mean of 3.62 of 5). But when the associations are recent and overlearned, less activity is shown in several areas, presumably because the associations have already been formed. Vividness ratings go up (a mean of 4.01), and the relationship between the behavioral rating and signal strength in two areas associated with the task becomes impressively fine tuned. Both types of findings speak to the dynamic aspects of auditory imagery: the neural locus reflects what happens as the anticipatory imagery becomes a more accurate predictor of the upcoming information. Anticipation is likely one way that musicians calibrate their timing when playing in an ensemble;[16] here we see that even small amounts of training can yield some dramatic neural changes associated with that mechanism.

From both sets of studies, we can draw some conclusions about both the construct of auditory imagery and the methods for studying it. First, it is possible to devise tasks that externalize the essentially private experience of auditory imagery. Furthermore, we can capture time-locked aspects of auditory imagery, which is important for studying imagery experiences that unfold over time. The more obvious example of this was the ability of the participants in the emotion study to track, with exquisite temporal resolution, the moment-to-moment changes in rather complex imagined music. But even the simpler task of anticipating one melody after hearing another one requires temporal tracking.

We also saw that the dynamic aspect of imagery can in some respects be reflected in short-term neural changes. The additional training in the Leaver *et al.* study[12] comprised a mere 10 minutes. Nevertheless, this was enough to engender clear differences in neural response—in this case the reduction of response as the anticipation became more entrenched.

Finally, we saw remarkable correspondences in self-report measures and neural patterns. As Zatorre reports in another paper in this Symposium, self-report on an auditory imagery vividness scale predicts response in auditory cortex and intraparietal sulcus during mental musical transformation tasks. We might think of this as a relationship between a *trait* measure of auditory imagery and neural response, where overall stronger ability to imagine sounds predicts more activity. The trial-by-trial vividness measure in the anticipatory imagery study reflects more a *state* measure as participants judged each particular instance of anticipation. We cannot disentangle causality; we do not know if the ratings engender the neural changes or vice versa. However, we can say that the essentially linear relationship between self-reported vividness and neural activity suggests that these reports are psychologically robust. These results should encourage cognitive neuroscientists to include both objective and subjective behavioral measures as they probe complex mental phenomena.

Acknowledgments

The research reported in Ref. 12 was supported by the National Science Foundation.

Conflicts of interest

The author declares no conflicts of interest.

References

1. Kosslyn, S.M. 1980. *Image and Mind*. Harvard University Press. Cambridge, MA.
2. Finke, R.A. 1985. Theories relating mental imagery to perception. *Psychol. Bull.* **98:** 236–259.
3. Hubbard, T.L. 2010. Auditory imagery: empirical findings. *Psychol. Bull.* **136:** 302–329.
4. Talarico, J.M. & D.C. Rubin. 2003. Confidence, not consistency, characterizes flashbulb memories. *Psychol. Sci.* **14:** 455–461.
5. Wells, G.L., A. Memon & S.D. Penrod. 2006. Eyewitness evidence: improving its probative value. *Psychol. Sci. Pub. Int.* **7:** 45–75.
6. Halpern, A.R. 1989. Memory for the absolute pitch of familiar songs. *Mem. Cognit.* **17:** 572–581.
7. Altenmüller, E. & H.-C. Jabusch. 2010. Focal dystonia in musicians: phenomenology, pathophysiology and triggering factors. *Eur. J. Neurol.* **17**: 31–36.
8. Highben, Z. & C. Palmer. 2004. Effects of auditory and motor mental practice in memorized piano performance. *B. Coun. Res. Music Ed.* **159:** 58–65.
9. Nicolelis, M.A.L. 2001. Actions from thoughts. *Nature* **409:** 403–407.
10. Halpern, A.R., R.J. Zatorre, M. Bouffard & J. Johnson. 2004. Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia* **42:** 1281–1292.
11. Lucas, B.L., E. Schubert & A.R. Halpern. 2010. Perception of emotion in sounded and imagined music. *Music Percept*. **27:** 399–412.
12. Leaver, A.M., J. Van Lare, B. Zielinski, *et al*. 2009. Brain activation during anticipation of sound sequences. *J. Neurosci.* **29:** 2477–2485.
13. Halpern, A.R. & J.C. Bartlett. 2011. The persistence of musical memories: a descriptive study of earworms. *Music Percept.* **28:** 425–431.
14. Schubert, E., P. Evans & J. Rink. 2006. Emotion in real and imagined music: same or different? In *Proceedings of the Ninth International Conference on Music Perception and Cognition*. M. Baroni, A.R. Addessi, R. Caterina & M. Costa, Eds.: 810—814. Bologna, Italy.
15. Turpin, C. (Producer). 2010. Wife, school lost in quake, violinist vows to rebuild. *All Things Considered* [audio podcast]. Available at: http://www.npr.org/2010/01/23/122900781/wife-school-lost-in-quake-violinist-vows-to-rebuild. Accessed January 23, 2010.
16. Keller, P.E. & M. Appel. 2010. Individual differences, auditory imagery, and the coordination of body movements and sounds in musical ensembles. *Music Percept.* **28:** 27–46.