2015

Differences in Auditory Imagery Self-Report Predict Neural and Behavioral Outcomes

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

Psychomusicology: Music, Mind, and Brain
2015, Vol. 25, No. 1, 37–47

0275-3987/15/$12.00 http://dx.doi.org/10.1037/pmu0000081

Differences in Auditory Imagery Self-Report Predict Neural and Behavioral Outcomes

Andrea R. Halpern
Bucknell University

Mental imagery abilities vary among individuals, as shown both by objective measures and by self-report. Few imagery studies consider auditory imagery, however. The Bucknell Auditory Imagery Scale is a short self-report measure encompassing both Vividness and Control subscales for musical, verbal, and environmental sounds. It has high internal reliability, no relation to social desirability, and only a modest relation to musical training. High scores on Vividness predict fewer source memory errors in distinguishing heard from imagined tunes on a recognition test, and better performance on pitch imitation tasks. Furthermore, higher scores are related to hemodynamic response and gray matter volume in several brain areas that are known to be involved in auditory imagery. Even though self-report measures encompass both cognitive and metacognitive aspects, they are useful tools in accounting for individual differences in high-level cognitive skills.

Keywords: auditory imagery, self-report scale, vividness, control, neural correlates of imagery, individual differences

Mental imagery has been a topic of psychological inquiry for a number of years, with most researchers focusing on visual imagery. However, auditory imagery has recently been studied as a topic in its own right (see Hubbard, 2010 for a review). Auditory imagery has been implicated in a variety of mental processes, such as reality monitoring (Johnson & Raye, 1981), working memory and rehearsal (Rudner, Rönnberg & Hugdahl, 2005; Tinti, Cornoldi, & Marschark, 1997), and hallucinations (Vitrovic & Biller, 2013), as well as musical processing (Crowder, 1989; Cupchik, Phillips, & Hill, 2001; Halpern, 1988a, 1988b). The experience of "hearing" something in one's head is a phenomenologically strong one, particularly for music (Bailes, 2007) and is accompanied by identifiable changes in cerebral blood flow (Halpern & Zatorre, 1999; Zvyagintsev et al., 2013) as well as neural electrical signal (Schaefer, Vlek, & Desain, 2011).

One aspect common to many of these studies is that performance is usually averaged over all the participants in a study. Sometimes group differences, particularly with regard to musical training, are examined. For instance, Aleman, Nieuwenstein, Böcker, and de Haan (2000b) found that musicians were superior to nonmusicians in both musical and nonmusical auditory imagery tasks, but not in a visual imagery task. However, individual differences in auditory imagery at a finer level have not typically been addressed.

In contrast, individual differences in visual imagery have been studied extensively, most commonly by self-report measures such as the Vividness of Visual Imagery Questionnaire (VVIQ; Marks, 1973). This questionnaire is designed to elicit self-report measures of visual imagery vividness, and has been used in hundreds of studies either probing differences in reported imagery ability for different participant groups or assessing the predictive value of self-reported imagery for more objective tasks presumed to require imagery. McKelvie (1995) reviewed both reliability and validity of the VVIQ in a large meta-analysis. Reliability as measured by internal consistency was quite good. Predictive validity, on the other hand, seems to vary by type of task. As examples, VVIQ scores predict reasonably well performance on tasks such as memory for visual detail, but not performance on mental rotation tasks. Kozhevnikov, Kozhevnikov, Yu, and Blazhenkova (2013) link vividness to one aspect of visual imagery skill called object visualization; this predicts artistic creativity—in contrast to spatial visualization, which predicts scientific creativity.

Similar questions about individual variability are of potential interest to auditory imagery researchers. For instance, Aleman, Nieuwenstein, Böcker, and de Haan (2000a) compared the imagery abilities of nonpsychotic individuals who report occasional verbal hallucinations to those of nonhallucinators. Hallucinators reported more vivid visual imagery on the VVIQ than nonhallucinators. Barrett (1993) found a similar result using the visual vividness subscale of Betts' (1909) Questionnaire Upon Mental Imagery or QMI (see more below). On the other hand, both studies reported that the two groups did not differ on self-reported vividness of auditory imagery using the auditory subscale of the QMI.

Despite this interest in documenting individual differences in auditory imagery, scales for capturing this ability have been slow to develop. An early attempt to index individual differences was Betts' (1909) QMI, which assesses imagery in seven sensory modalities, including audition. The QMI presents people with

The initial scale development formed part of the Honors Thesis of David J. Lizotte, at Bucknell University. I thank Drs. Craig Colder, Jean Lamont, J. T. Ptacek, and Joel Wade for their consultations about statistical analysis, and the Psychology Department of Bloomsburg University for allowing their students to participate in the research.

Correspondence concerning this article should be addressed to Andrea R. Halpern, Department of Psychology, Bucknell University, One Dent Drive, Lewisburg, PA 17837. E-mail: ahalpern@bucknell.edu

written descriptions, using visual, auditory, or other sensory detail, and asks them to rate the vividness of their mental images of certain aspects of the descriptions. Participants use a 7-point rating scale for vividness and clarity, ranging from "perfectly clear and vivid as the actual experience" to "no image present at all."

The QMI was a significant step in measuring individual differences in auditory imagery because it did so via quantifiable self-report ratings. However, Betts' QMI is unwieldy because it consists of 150 items: 40 in the visual domain and 10 to 20 each in the other domains. The presentation of descriptions and items on the questionnaire is also relatively unsystematic, as well as complicated by the use of many imagery modalities.

A more recent strictly auditory self-report imagery questionnaire was designed by Gissurarson (1992). His Auditory Imagery Scale (AIS) is an outgrowth of the auditory portion of Betts' (1909) QMI. The scale consists of seven items: six environmental sound items, such as "Imagine the sound of water dripping," and one open-ended music item, "Imagine your favorite piece of music." Participants are asked to rate how well they can imagine the sound or noise using a 4-point rating scale ranging from "very clear sound/noise" to "no sound/noise at all." Gissurarson (1992) reported an internal consistency of $\alpha = .80$. He also performed a principal components-based factor analysis and found that all items loaded on a single common factor.

For a number of reasons, Gissurarson's (1992) scale is not an ideal candidate for an auditory imagery assessment instrument of interest to music researchers. Only seven items appear on the questionnaire, most of which are environmental. Interestingly, Gissurarson (1992) found that interitem correlations were generally high, accounting for the high reliability, but correlations of the musical item with each other item were much lower than the other ones. The only type of validation that Gissurarson (1992) provided was a construct validation between his measure and the VVIQ, which were significantly correlated, $r(158) = 0.48$. This result provides some evidence that there may be a general self-report imagery capacity across modalities, at least vision and audition. However, Gissurarson did not present his scale with any auditory tasks to see if it actually predicted anything.

A few other scales have recently been offered to the auditory imagery research community. Beaty et al. (2013) devised a five-item scale that captured how often, important, and enjoyable people found their musical imagery experience, and related those reports to musical preferences and personality. One interesting correlation was that people reporting more frequent musical imagery preferred intense and reflective music. Andrade, May, Deeprose, Baugh, and Ganis (2013) included five environmental auditory items on the Plymouth Sensory Imagery Questionnaire which probes imagery vividness, and found the different sensory items each formed a separate subscale on a factor analysis. Willander and Baraldi (2010) take issue with how well vividness may capture the imagery experience and prefer to assess clarity. They devised a Clarity of Auditory Imagery Scale, which includes 16 mostly environmental items and found high internal reliability and a single factor explaining about 31% of the variance in answers. However, no correlations with actual tasks or neural activity were studied with respect to any of these scales.

The Bucknell Auditory Imagery Scale (BAIS) is a short questionnaire that covers the three main domains of auditory experience: musical, verbal, and environmental. The instrument has two subscales to capture vividness as well as control of auditory images. I will first describe its development and initial psychometric road testing, followed by a summary of ways in which this simple scale has predicted some interesting behavior and also neural concomitants.

Development and Psychometrics of the BAIS

In designing the BAIS, previous work in the visual domain, especially with the widely employed VVIQ, was used as a guideline. Self-report measures for visual imagery have typically focused on both vividness and control of one's images; hence, both of these aspects are included in the BAIS. Vividness of imagery was already described for the VVIQ and the QMI. Gordon (1949) defined control of imagery as the ability to call up new images, dismiss old images, and manipulate images while they are present. Thus, control of imagery taps the more dynamic aspects of mental imagery.

Gordon (1949) developed the first measure of control of visual imagery, the Test of Visual Imagery Control (TVIC). The TVIC consists of 12 questions or statements regarding an image of a car and its surroundings. Each successive question or statement asks whether or not the participant can change the image he or she had just generated to a new image. For example, the first question is "Can you see a car standing in front of a garden gate?" which is followed by "What is its color? Try and see it in a different color." Gordon classified people as being able to control their imagery only if they could successfully change their image at each point, thus answering "yes" to all questions. A slightly modified version of the TVIC (Richardson, 1969) is typically used in visual imagery research. On the other hand, no auditory control-of-imagery measure has heretofore been published.

The BAIS captures self-report of auditory imagery vividness and control using two subscales with similar items and formats. The scale incorporates a number of suggestions offered by McKelvie (1995) to improve the VVIQ. The BAIS was also developed to capture potential differences between the auditory and visual domains, as well as the desire for parallel subscales for vividness and control. The scale includes three types of sound: music, environmental sounds, and spoken voice. This section describes the construction of the BAIS, and examines its internal consistency, factor structure, and construct validity via administration of a modified form of the VVIQ.

Method

The main goals of the development phase were to test its reliability, factor structure, and construct validity. Also of interest was the relationship between the control and vividness subscales. To assess its construct validity, the same participants filled out a new version of the VVIQ: the VVIQ-Modified (VVIQ-M). Modifications included expanding the rating scale from 5 to 7 points, reversing the rating scale so that higher numbers corresponded to more vivid imagery, and some wording changes: As per McKelvie's (1995) suggestions, the rating scale labels were changed from "clear and vivid" to just "vivid," and also instead of instructing respondents to "visualize," we asked them to "try to visualize." If auditory and visual imagery processes share common cognitive structures, then people's auditory imagery abilities should predict their visual imagery abilities at least to some extent.

To assess the possibility that the BAIS is subject to social desirability biases, the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1964) was included in a pilot study. Finally, sex differences in BAIS ratings were analyzed. Sex differences are not always reported for the VVIQ, but when they are, women report more vivid visual images than men (a significant effect in McKelvie's (1995) meta-analysis). A similar difference on the BAIS was a reasonable prediction.

Participants. Seventy-six college students (54 women, 22 men) participated in the main study, either as uncompensated volunteers or for a small amount of course credit. The only requirement for participation was having corrected or normal hearing. Participants were selected without regard to age (M = 22.6 years, SD = 8.1), number of years of formal education (M = 14.0 years, SD = 1.8), or musical background (15 of the 74 participants indicated that they had had 5 or more years of formal musical training; because of this skew, this correlations with years of training was not considered further).

Materials. For each item on the BAIS, a situation is described followed by a description of a sound relevant to that situation (see Appendices A and B). Respondents are instructed to try to construct an auditory image of that sound. For the Vividness subscale (BAIS-V), people are asked to rate the vividness of their image of the sound using a 1 to 7 scale, with 1 meaning *no image is present at all*, 4 meaning *fairly vivid,* and 7 meaning *the image is as vivid as the actual sound.* No other scale points are labeled. For the Control subscale (BAIS-C), each situation and sound is again described but the task is to rate the ease of changing one's image of the original sound to a new sound. Respondents use a different 7-point scale for control ratings, with 1 meaning *no image present at all,* 4 meaning *could change the image, but with effort,* and 7 meaning *extremely easy to change the image.* For instance, one item is "Consider the beginning of the song 'Happy Birthday.'" The BAIS-V asks the respondent to rate the vividness of a trumpet playing the opening. The BAIS-C asks the respondent to rate the ease of imagining a change from a trumpet to a violin.

The 14 items of the BAIS cover the domains of music, the environment, and voice. Some items straddle two categories, such as choir singing, which includes both music and voice. The same items are probed on each subscale (with the necessary further detail for the BAIS-C), but in a different order. All items were selected from a large pool of items that were piloted tested with a group of 36 undergraduates so as to ensure familiarity to most undergraduates in this population. Participants were asked to use all levels of the rating scales as appropriate when selecting their ratings. In addition to the larger pool of BAIS items, these participants received a musical background questionnaire, the VVIQ-M, and the Marlowe-Crowne Social Desirability Scale (Crowne & Marlowe, 1964). The pilot study showed that social desirability ratings did not correlate with BAIS-V ratings, $r(34) = -0.24$, *ns*, or Control ratings, $r(34) = -0.27$, *ns*, so the Marlowe-Crowne Scale was not presented in main phase. The final set of BAIS items were chosen as to be familiar, cover a range of domains, and which elicited a wide distribution of ratings.

Procedure. Participants were given the packet of questionnaires and told that the study concerned auditory imagery. The directions for each questionnaire were printed on the questionnaires themselves, but directions were also orally presented. The experimenter emphasized to the participants that for each item on the BAIS-C they were to physically cover up the second sentence describing the changed sound while constructing their original image. This was done to encourage them to process the two parts of each item serially. They were instructed to complete the questionnaires in the order given, without returning to any one questionnaire after moving on to the next. Participants were tested in groups and the experimenter was available at all times to answer any questions.

The BAIS-V, the BAIS-C, and the VVIQ-M were given in one of four different orders (thus sometimes the two subscales were presented in adjacent order but sometimes not). The musical background questionnaire was always given last. At the end of the 20-min session, participants were debriefed.

Results

Results primarily concern the BAIS, but a few relevant results from the VVIQ-M are also reported.

Scale ratings. The mean rating for the BAIS-V and BAIS-C each was based on 14 items, and based on 16 items for the VVIQ-M. Participants' mean ratings varied widely: Mean ratings on the BAIS-V ranged from 2.9 to 6.9, with a grand mean of 5.1, SD = 0.9. Mean ratings on the BAIS-C ranged from 3.2 to 6.7 (M = 5.1, SD = 0.9). Mean ratings for the VVIQ-M ranged from 3.0 to 7.0 (M = 5.4, SD = 1.0). Mean ratings on the BAIS-V and the BAIS-C did not differ based on the order of presentation.

People also distributed their ratings fairly normally over items. On the BAIS-V, 68 of 76 participants used at least four scale points in their answers, as did 69 of 76 people on the BAIS-C. Only two individuals used fewer than four points on both scales, meaning that 74 of 76 people used at least four scale points on at least one of the scales. On average, people showed a SD of 1.5 in their ratings over the 14 items on both the BAIS-V and BAIS-C, and the equivalent SD for the VVIQ-M was 1.2.

Reliability. The internal consistency of each of the three self-report measures was assessed using Cronbach's alpha. The BAIS-V had a reliability of $\alpha = .83$. Interitem correlations for the BAIS-V ranged from −0.07 to 0.57, with a mean correlation of 0.27. The BAIS-C had a reliability of $\alpha = .81$. Interitem correlations for the BAIS-C ranged from −0.10 to 0.56 (M = 0.23). Combining the subscales, $\alpha = .91$. The VVIQ-M had a reliability of $\alpha = .91$. Interitem correlations for the VVIQ-M ranged from 0.05 to 0.75 (M = 0.40).

Factor analysis. Exploratory factor analyses of each subscale were performed in SPSS, using the method of principal components.[1] Initial analyses run without constraining the number of components suggested that a solution constrained to three components would provide the most interpretable solution; this also took into account the modest sample size. The solutions using varimax rotation and three components are reported here. The analyses were run on raw and standardized scores; the results were very similar and the raw score results are presented. See Tables 1 and 2.

In interpreting the rotated solutions, loadings whose absolute value exceeded 0.35 were considered significant. For the BAIS-V, the solution converged in five iterations and accounted for 58% of

[1] Communalities for each subscale averaged .63, and were above .50 for all but three items of the total of 28 items.

Table 1
Component Loadings of the Items of the BAIS-V

Item	Type	Com1	Com2	Com3
3	E	**.72**	.24	.15
4	E	**.48**	**.38**	.21
6	E	**.55**	.30	−.04
9	E/V	**.81**	.21	.07
10	E/V	**.55**	.05	.28
13	E/V	**.61**	.04	**.42**
12	E/M	**.62**	**.48**	.07
14	E/M/V	**.75**	−.08	.02
5	M	.26	**.72**	.20
8	M	−.03	**.89**	.03
1	M/V	.06	.03	**.77**
7	M/V	.06	**.49**	**.65**
11	M/V	.25	**.66**	.25
2	V	.21	.14	**.77**

Note. BAIS-V = Bucknell Auditory Imagery Scale–Vividness; COM = Component. The type of item is either music (M), environmental sound (E), voice (V), or a combination. Loadings (> |.35|) are in bold.

the variance. Component 1 included all of the environmental items and several other items that had environmental contexts: cheering at a baseball game, a classroom, and listening to the radio in the car; loadings ranged from .48 to .81. Component 2 encompassed all of the items involving music, including tap dancing (which had a higher loading on Component 1) except the car radio item. One anomalous item is imagining the sound of the dentist drill, although that loading was low at .38; the other loadings ranged from .48 to .87. Component 3 largely comprised items with voices, either in a musical context or not (imagining a choir and the sound of a clerk), with loadings ranging from .42 to .77. One exception here was that imagining a teacher's voice loaded only .28 on this component but did load moderately highly (.55) on the environmental component.

For the BAIS-C, the solution converged in nine iterations and accounted for 59% of the variance. The solution overall is not as interpretable as for BAIS-V: The first factor shows high loadings (.52 to .82) for the last six items on the scale, which comprise verbal and environmental items, except for one musical item (Beethoven). Component 2 comprises all the musical items, except for Happy Birthday (which in both scales clusters with verbal rather than music items), and includes the rainstorm item. Component 3 comprises mostly items with a verbal component, with loadings from .44 to .73, 73 but omits the two classroom items.

Construct validity. We found a significant correlation between the VVIQ-M and BAIS-V, as well as BAIS-C, both $r(74) = 0.62$. We also found the BAIS-V and BAIS-C to be significantly correlated, $r(74) = 0.74$; all $p < .01$.

Sex differences. The mean vividness rating for the 54 women on the BAIS-V was 5.2, $SD = 0.8$, whereas that for the 22 men was 4.9, $SD = 1.1$. The mean rating for women on the BAIS-C was 5.2, $SD = 0.7$, and for men was 4.9, $SD = 0.8$. These means did not differ significantly by a *t* test. However, the mean vividness rating for women on the VVIQ-M was 5.6, $SD = 0.8$, whereas that for men was 4.8, $SD = 1.3$. These mean ratings did differ significantly, $t(74) = 2.76$, $p < .01$.

Replications and Extensions

In a recent paper on poor-pitch singing, Pfordresher and Halpern (2013) presented the BAIS-V to 120 undergraduates (114 completed both scales): Most had fewer than 5 years of musical training. Cronbach's alpha for the BAIS-V was $\alpha = .83$, for BAIS-C $\alpha = .91$, and the same for total BAIS, $\alpha = .91$. The two subscales correlated $r(112) = 0.50$, $p < .01$. Also, a study currently under review (Lima et al., 2015) administered the BAIS-V and the original VVIQ to 46 adults spanning ranging in age from 20 to 81. Those two scales correlated significantly, $r(44) = .51$, $p < .001$. Finally the BAIS has recently been reconfigured for online presentation, which controls the timing of each item, particularly useful in the BAIS-C subscale where two scenarios comprise each item. A recent sample in my lab of 52 respondents yielded nearly identical means and distributions of ratings as the original sample; in particular, only 3 of the 52 used fewer than four scale points across both scales.

Discussion

Overall, responses to the BAIS indicated that it is a well-behaved psychometric instrument. One important observation is that people did seem to report varying degrees of auditory imagery vividness and control ratings. The range of mean ratings for participants on both subscales was large and the standard deviation of the grand mean rating for each measure was moderately high. Furthermore, ratings did not hover at the ceiling: Grand mean ratings were closer to the middle rating of 4 than the highest rating of 7. Individuals by and large also used a variety of scale values in their responses. Thus, we have some confidence that people are willing to differentiate imagery experiences in their ratings, and that they differ from one another in the average degree of control and vividness they report.

The fact that a questionnaire picks up individual differences says nothing about the quality of the scale itself. For example, a scale might be composed of a number of unrelated items that do not pertain to any one mental ability, and still be measuring individual differences. However, the internal consistency (reliabil-

Table 2
Loadings of the Items of the BAIS-C

Item	Type	Com1	Com2	Com3
3	E	.26	**.84**	−.11
14	E	**.82**	.02	−.02
7	E/V	.14	.25	**.73**
11	E/V	.52	−.01	**.64**
4	M/V to E	.17	**.65**	.22
10	E/M	**.55**	.05	.28
9	M	**.76**	**.37**	.08
2	M	−.05	**.56**	**.44**
6	M/V	.29	**.67**	−.03
8	M/V	.24	−.07	**.73**
1	M/V	−.19	**.45**	**.38**
13	V	**.71**	.11	.25
12	V	**.74**	.19	.32
5	V	.19	**.38**	**.61**

Note. BAIS-C = Bucknell Auditory Imagery Scale—Control; COM = Component. The type of item is either music (M), environmental sound (E), voice (V), or a combination. Loadings (> |.35|) are in bold.

ity) of the BAIS-V and the BAIS-C in two samples were acceptably high (greater than .80). The internal consistencies of the BAIS-V and BAIS-C were of equivalent magnitude to those reported by McKelvie (1995) for the VVIQ, indicating that the reliabilities of the BAIS conform fairly well to the reliability of the most widely used imagery self-report measure.

The rotated solutions for the BAIS-V suggested that the components corresponded more or less to the three types of probed items, with some exceptions, similar to the four factors corresponding to the four vignettes of the VVIQ scale (Kihlstrom, Glisky, Peterson, Harvey, & Rose, 1991).

The component structure for BAIS-C is less clear, perhaps because many of the items can be classified under several domains. For instance, Item 7 refers to the sound of a dentist's drill that stops and is replaced by the sound of the receptionist: this involves a shift from environmental to verbal. As noted above, in both scales, people were asked to imagine Happy Birthday as an instrumental piece, but that item clustered with other verbal items, suggesting that activation of the lyrics is obligatory for that oft-sung song. Future studies with larger samples sizes would be worthwhile to confirm these findings. It is worth noting, however, that no studies reported herein separate out the different item types either to present as subscales, nor in analysis, so the results relating to component structure remain of secondary interest.

The current data do not allow us to settle whether control and vividness are separable aspects of self-reported imagery. The high correlation of the two subscales, and their identical correlation with the VVIQ-M, may reflect the fact that each control item started with the need to construct a vivid image. In other words, controlling or changing one's images requires generating vivid images in the first place, at least for this instrument. On the other hand, the tasks requested of respondents for each subscale are on the face of it quite different. This prima facie consideration suggests keeping the individual subscales and not combine them in one 28-item scale; predictive validity studies described hereafter reinforce that decision.

One additional result meriting attention is the pattern of sex differences in ratings on the VVIQ-M, BAIS-V, and BAIS-C. The VVIQ-M showed the commonly reported result of women giving higher ratings; however, neither the BAIS-V nor the BAIS-C showed a sex difference in ratings for the same sample. It may be the case that women have particularly vivid visual imagery, but that this does not extend to all other modalities. Finding sex differences on the VVIQ-M but on neither BAIS subscale underscores the possibility that visual and auditory imagery are different constructs.

Behavioral Correlates of BAIS

The section above presented evidence that the BAIS-V and BAIS-C have good reliabilities and construct validity. People differ from one another in the ratings they provide, they use a wide range of the scale, and they seem to respond consistently. Here, we ask whether these individual differences in ratings correspond to individual differences on tasks requiring the use of auditory imagery.

One example comes from a source memory study (Herholz, Halpern, & Zatorre, 2012). In this fMRI study, 10 volunteers were selected so as to have a wide range of scores on the BAIS-V (we did not administer the BAIS-C) and also had a wide range of musical experience, which did not correlate with BAIS-V score. They first saw a karaoke-type presentation of lyrics to familiar songs. On each trial, the lyrics were either accompanied by a sound file playing the melody synchronized to the lyrics, or the participant had to imagine the tune being played.

After about 30 min during which time anatomical scans were collected, a surprise recognition test ensued. Titles of familiar songs appeared one at a time, and participants had to make an old/new judgment; if they said "old," they were asked whether the melody had been heard or imagined in the first phase. We considered two kinds of source memory errors: HI errors were mistaking a previously heard song for an imagined one, and IH errors were the reverse mistake. We found that people with high self-reported auditory imagery vividness made fewer HI errors, $r(8) = -.69$, $p = .027$: that is, when a melody had been heard, they rarely said it had been imagined; BAIS-V was not related to the reverse error (which has been reported for visual imagery (Finke, Johnson, & Shyi, 1988)). Thus, the more vivid imagers seemed reinstate strong memory representations for actually heard melodies, which they accurately distinguished from the weaker traces of imagined melodies.

A second example comes from the study of poor-pitch singing referred to above (Pfordresher & Halpern, 2013). We were particularly interested in the ability to imitate pitch, in people who do not necessarily show impairments in pitch discrimination. By the reasoning of an inverse model, we thought that poor imitators might have deficient auditory imagery representations of the to-be-imitated pitch. We predicted that self-report of auditory imagery vividness would predict pitch error in imitation. We did not expect control of imagery to be related to pitch error, as the target needs to be represented but not changed into another sound. We asked a sample of undergraduates to sing back single pitches presented in their comfortable singing range. They also carried out a pitch discrimination task. The percentage of trials sung in tune (within a musical semitone of the correct pitch) was not correlated with pitch discrimination, but was predicted by BAIS-V, $r(118) = .28$, $p < .01$; it was not predicted by BAIS-C, $r(112) = .11$, $p > .10$.

In an extension of that study, Greenspon, Pfordresher, and Halpern (2013) presented actual auditory imagery tasks to good and poor vocal imitators. We asked people to sing back or recognize a target that was mentally transposed, reversed, or serially shifted. Once again production error rates correlated with BAIS-V scores. Figure 1 shows the mean absolute note error for all the production conditions, plotted against BAIS-V score: People with higher self-reported vividness made fewer errors on these auditory imagery tasks, $r(38) = -.35$, $p < .05$. In these auditory imagery tasks, BAIS-C also predicted overall error, $r(36) = -.31$ (two people did not fill out the BAIS-C).

Self-reported control also predicted performance in another active imagery task. Gelding, Thompson, and Johnson (2014) devised a novel auditory imagery task in which participants were given a starting note and then subsequent notes were cues to whether the next note would move up or down in scale steps. The final note was cued only by the arrow, not played, and participants had to judge whether the played note was the correct ending; the trials varied in difficulty. Better performers (success at higher difficulty level) on the task indicated that

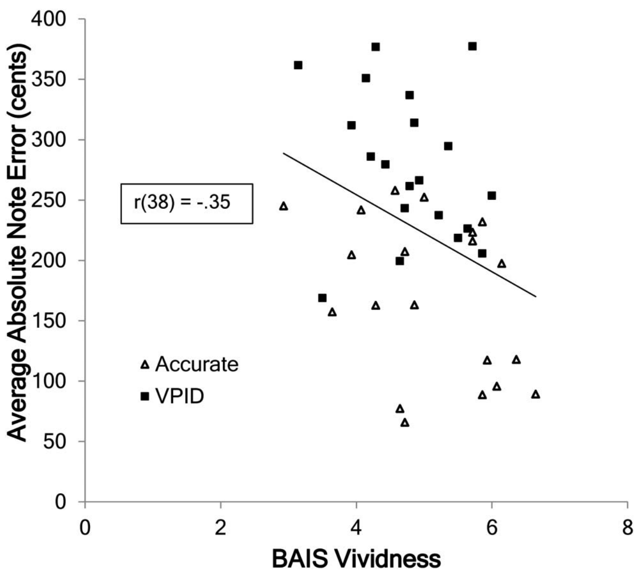

Figure 1. Mean absolute note error on three auditory transformation tasks, plotted against self-reported auditory imagery vividness on the Bucknell Auditory Imagery Scale (BAIS).

they used a pitch imagery strategy, and better performance correlated with BAIS-C, $r(38) = 0.56$, $p < .001$, as did faster reaction time (RT) on correct trials, $r(38) = -0.51$, $p < .01$. BAIS-V did not predict these outcomes.

Another skill that may relate to reported ability to manipulate sound could be generating predictions about what sounds might come next in complex sequence. Focusing on the timing aspects of music, a project in progress in my lab with Ian Colley and Peter Keller, looks at whether people with higher versus lower self-reported BAIS-C differ in being able to learn and then predict an expressive (variable) timing pattern in piano music. Preliminary results show high BAIS-C (but not BAIS-V) predicts less asynchrony (difference in timing between the musical beat, and the participant's tapping response), or better synchronization, $r(34) = -.40$, $p = .01$. A prediction/tracking (P/T) ratio was calculated as a measure of the extent to which participants were anticipating tempo changes by tapping early (prediction) or reacting by tapping late (tracking; Pecenka & Keller, 2011). Anticipation connotes a more accurate internal model of expressive timing. In a hierarchical regression model, BAIS-C predicted anticipatory synchronization above and beyond an objective test of imagery (pitch arrow task of Gelding et al., 2014), and an operations span test of working memory (R^2 change $= .20$, $p = .003$).

Finally, although the BAIS was developed to capture voluntary, controlled auditory imagery, some aspects of this process could relate to involuntary musical imagery (INMI, or "earworms"). Recent research reported by Floridou, Williamson, Stewart, and Müllensiefen (2014) tested a large online sample ($N = 2{,}286$) and found that score on the BAIS-V was positively correlated with frequency of INMI ($r = .23$) and also the extent to which people reported a tendency to move in rhythm with those imagined tunes ($r = .25$).

Neural Correlates of BAIS

Cognitive neuroscientists have begun to employ brain imaging technologies to better understand brain mechanisms mediating several modalities of mental imagery, such as vision (Kosslyn et al., 1999), movement (Parsons et al., 1995), and of course audition (Zatorre & Halpern, 2005). Correlating amount or spatial extent of activation with self-reported imagery experiences can lead us to a new way of externalizing what remains, at the core, the intensely internal experience of mental imagery.

A few recent studies in musical imagery have shown some interesting neural correlations with the BAIS.[2] In the study by Herholz, Halpern, and Zatorre (2012) mentioned above, fMRI results revealed that a network consisting of right anterior superior temporal gyrus and right dorsolateral prefrontal cortex was active during encoding of imagined melodies. It makes sense that imagery would be mediated by a connection between a secondary auditory and a working memory area. Interestingly, cerebral blood flow in both those areas was higher among people with higher scores on the BAIS-V. And consistent with the better ability of more vivid imagers to distinguish previously heard from previously imagined melodies during title recognition, one area of left temporal pole was more active in more vivid imagers during recognition of previously heard melodies.

Considering more difficult auditory imagery tasks, Zatorre, Halpern, and Bouffard (2010) asked people to mentally reverse a melody. They did this by first presenting the first notes of the first phrase of a familiar song. After a pause, a comparison was presented that was either that phrase in reverse order, or the reversed phrase had a note error in it. The task was to reverse the reversal and say whether it was a valid or invalid reversal. Fairly trained musicians were used in this study, as the task was difficult, but BAIS scores still varied (on average, BAIS total score correlates about .30 with years of musical training). This study used the BAIS total score as a predictor. During the pause after the first phrase, BAIS score was positively correlated with activity in right secondary auditory cortex, as presumably the model was being represented in imagery. When the reversal phase was in process, we saw activity in intraparietal sulcus, consistent with the attentional demands of the task. This activity was bilateral in the overall contrast with a control condition. However, high BAIS scorers showed higher IPS activity during reversal on the right side. Note the right-sided asymmetry in all these correlations.

The final example comes from a structural study (Lima et al., 2015) in which a large sample of 74 individuals participated in a passive listening task of different kinds of human vocal sounds during fMRI scanning. The main results of interest here were the structural scan results, measuring gray matter volume via voxel-based morphometry. Higher BAIS-V correlated with positively with gray matter volume in several areas, including left inferior parietal lobule and left supplementary motor area (SMA). Both areas have been implicated in functional studies of musical imagery (Halpern & Zatorre, 1999; Zatorre, Halpern, & Bouffard, 2010; Foster, Halpern, & Zatorre, 2013). No results were modulated by age, and increased functional discriminability of sounds was as-

[2] Exact correlations are not given in this section, as some studies presented results over a region, rather than a single voxel.

sociated with larger regional gray matter volume, $r = .28$, $p = .03$ in the SMA. Interestingly, SMA volume was also higher in those reporting more vivid visual imagery; the auditory and visual scales correlated at $r = .51$ ($p < .001$), as they did in the initial scale development study.

Cross-Language and Culture Versions

Although the BAIS has been unpublished up to now, researchers have been free to contact me to get a version, and I have worked with teams in other countries to develop slightly modified versions for other countries. The item referring to baseball has been changed to different sports even for other English-speaking countries: Cricket is used in the United Kingdom, for instance, and football (soccer) in other European countries (the crowd noise is a common element). A few other small modifications to items are necessary in different cultures. Versions have been created in German, French, and Hebrew: In all cases, a native bilingual speaker has translated the scale, and then a second native bilingual has backtranslated. I have then worked with the researchers to make sure the translation matches the original as much as possible. A version has been created and piloted for children (in German), using modified items and scales.

Conclusions and Limitations

The BAIS scale has good reliability and an interpretable factor structure at least for the BAIS-V, and predicts a number of behavioral and neural (sometimes voxel-specific) results over a wide range of tasks. It is easy to be skeptical of self-report scales in cognition: They are subject to biases, interpretation, and the fallibility of retrospective report. Nevertheless, results from this scale are highly suggestive of people being able to query their own auditory imagery in a reliable way. Furthermore, this "trait" measure has predicted "state" performance in all the ways described. It is remarkable that a simple self-report scale has been so widely applicable.

Nevertheless, limitations and cautions are warranted. Some tasks that would seem to be sensitive to vividness were not predicted by the scale: For instance, in unpublished data from my lab, the scale failed to predict performance on a mental pitch comparison and mental loudness comparison task in a sample of 30 participants, despite a wide range of performance accuracy on both tasks. It may be the case that different aspects of auditory imagery, analogous to object versus spatial imagery (Kozhevnikov et al., 2013), comprise the overall ability, which is not detected by this scale. And in most of the correlations with behavioral results, the proportion of variance accounted for, while significant, is modest. Clearly factors other than self-reported vividness or control contribute to these complex memory tasks. In the expressive timing project mentioned previously, we added a working memory task to see to what extent differences in self-reported imagery is mediated by, or independent of, working memory differences.

Within the neural domain, the structural and functional results have so far not implicated all the same areas: Functional correlations pointed to the perceptual-memory network and attentional control mediated by the IPS, whereas the structural results were primarily found in the SMA, which may be related to the sequencing or covert motor planning activated during auditory imagery. It was particularly surprising that right superior temporal gyrus volume was not predicted by the BAIS-V.

It is not yet entirely clear the extent to which the two subscales capture shared versus unique aspects of imagery. They do correlate at a fairly robust level, but they also dissociate in some of the studies described above. It is also evident that visual and auditory imagery share components. In addition to positive correlations in self-report scales for visual and auditory imagery we found in our initial scale development, in the Lima et al. study (2015), left SMA correlated with visual imagery vividness, as well as auditory imagery vividness. Zvyagintsev et al. (2013) identified neural networks common to visual and auditory imagery, as well as unique areas. Cross-modality individual differences in imagery may include ability to generate and maintain resource-demanding perceptual representations.

Finally, BAIS-V or total BAIS have been more successful predictors than BAIS-C in tasks reported to date. The more complex self analysis of "control" may be related to outcomes in a less straightforward way, although recent reports of BAIS-C predicting the dynamic aspect of mental imagery and temporal prediction are quite promising. With the publication of this scale, it will be interesting to see what other tasks might be predicted by BAIS-C, and I invite researchers to keep me informed about their results.

References

Aleman, A., Nieuwenstein, M. R., Böcker, K. B. E., & de Haan, E. H. F. (2000a). Mental imagery and perception in hallucination-prone individuals. *Journal of Nervous and Mental Disease, 188,* 830–836. http://dx.doi.org/10.1097/00005053-200012000-00007

Aleman, A., Nieuwenstein, M. R., Böcker, K. B., & de Haan, E. H. (2000b). Music training and mental imagery ability. *Neuropsychologia, 38,* 1664–1668. http://dx.doi.org/10.1016/S0028-3932(00)00079-8

Andrade, J., May, J., Deeprose, C., Baugh, S. J., & Ganis, G. (2014). Assessing vividness of mental imagery: The Plymouth Sensory Imagery Questionnaire. *British Journal of Psychology, 105,* 547–563.

Bailes, F. (2007). The prevalence and nature of imagined music in the everyday lives of music students. *Psychology of Music, 35,* 555–570. http://dx.doi.org/10.1177/0305735607077834

Barrett, T. R. (1993). Verbal hallucinations in normals—II. Self-reported imagery vividness. *Personality and Individual Differences, 15,* 61–67. http://dx.doi.org/10.1016/0191-8869(93)90041-Z

Beaty, R. E., Burgin, C. J., Nusbaum, E. C., Kwapil, T. R., Hodges, D. A., & Silvia, P. J. (2013). Music to the inner ears: Exploring individual differences in musical imagery. *Consciousness and Cognition: An International Journal, 22,* 1163–1173. http://dx.doi.org/10.1016/j.concog.2013.07.006

Betts, G. H. (1909). *The distribution and functions of mental imagery*. New York, NY: Columbia University Press.

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 472–478. http://dx.doi.org/10.1037/0096-1523.15.3.472

Crowne, D. P., & Marlowe, D. (1964). *The approval motive: Studies in evaluative dependence*. New York, NY: Wiley.

Cupchik, G. C., Phillips, K., & Hill, D. S. (2001). Shared processes in spatial rotation and musical permutation. *Brain and Cognition, 46,* 373–382. http://dx.doi.org/10.1006/brcg.2001.1295

Finke, R. A., Johnson, M. K., & Shyi, G. C. W. (1988). Memory confusions for real and imagined completions of symmetrical visual patterns. *Memory and Cognition, 16,* 133–137. http://dx.doi.org/10.3758/BF03213481

Floridou, G., Williamson, V. J., Stewart, L., & Müllenseifen, D. (2014). *The Involuntary Musical Imager Scale (IMIS)*. Poster presented at the International Conference on Music Perception and Cognition, Seoul, South Korea.

Foster, N. E. V., Halpern, A. R., & Zatorre, R. J. (2013). Common parietal activation in musical mental transformations across pitch and time. *NeuroImage, 75,* 27–35. http://dx.doi.org/10.1016/j.neuroimage.2013.02.044

Gelding, R. W., Thompson, W. F., & Johnson, B. W. (2015). The pitch imagery arrow task: Effects of musical training, vividness, and mental control. *PloS One, 10,* e0121809. http://dx.doi.org/10.1371/journal.pone.0121809

Gissurarson, L. R. (1992). Reported auditory imagery and its relationship with visual imagery. *Journal of Mental Imagery, 16,* 117–122.

Gordon, R. (1949). An investigation into some of the factors that favour the formation of stereotyped images. *British Journal of Psychology, 39,* 156–167. http://dx.doi.org/10.1111/j.2044-8295.1949.tb00215.x

Greenspon, E., Pfordresher, P. Q., & Halpern, A. R. (2013, November). *Mental transformations of auditory imagery during vocal imitation*. Paper presented at the Auditory Perception, Cognition, and Action meeting, Toronto, Canada.

Halpern, A. R. (1988a). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 434–443. http://dx.doi.org/10.1037/0278-7393.14.3.434

Halpern, A. R. (1988b). Perceived and imagined tempos of familiar songs. *Music Perception, 6,* 193–202. http://dx.doi.org/10.2307/40285425

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex, 9,* 697–704. http://dx.doi.org/10.1093/cercor/9.7.697

Herholz, S. C., Halpern, A. R., & Zatorre, R. J. (2012). Neuronal correlates of perception, imagery, and memory for familiar tunes. *Journal of Cognitive Neuroscience, 24,* 1382–1397. http://dx.doi.org/10.1162/jocn_a_00216

Hubbard, T. L. (2010). Auditory imagery: Empirical findings. *Psychological Bulletin, 136,* 302–329. http://dx.doi.org/10.1037/a0018436

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85. http://dx.doi.org/10.1037/0033-295X.88.1.67

Kihlstrom, J. F., Glisky, M. L., Peterson, M. A., Harvey, E. M., & Rose, P. M. (1991). Vividness and control of mental imagery: A psychometric analysis. *Journal of Mental Imagery, 15,* 133–142.

Kosslyn, S. M., Pascual-Leone, A., Felician, O., Camposano, S., Keenan, J. P., Thompson, W. L., . . . Alpert, N. M. (1999). The role of area 17 in visual imagery: Convergent evidence from PET and rTMS. *Science, 284,* 167–170. http://dx.doi.org/10.1126/science.284.5411.167

Kozhevnikov, M., Kozhevnikov, M., Yu, C. J., & Blazhenkova, O. (2013). Creativity, visualization abilities, and visual cognitive style. *British Journal of Educational Psychology, 83,* 196–209. http://dx.doi.org/10.1111/bjep.12013

Lima, C. F., Lavan, N., Evans, S., Agnew, Z., Halpern, A. R., Meekings, S., . . . Scott, S. K. (2015). *Relating individual differences in auditory imagery to the structure and function of sensorimotor systems*. Manuscript submitted for publication.

Marks, D. F. (1973). Visual imagery differences in the recall of pictures. *British Journal of Psychology, 64,* 17–24. http://dx.doi.org/10.1111/j.2044-8295.1973.tb01322.x

McKelvie, S. J. (1995). The VVIQ as a psychometric test of individual differences in visual imagery vividness: A critical quantitative review and plea for direction. *Journal of Mental Imagery, 19,* 1–106.

Parsons, L. M., Fox, P. T., Downs, J. H., Glass, T., Hirsch, T. B., Martin, C. C., . . . Lancaster, J. L. (1995). Use of implicit motor imagery for visual shape discrimination as revealed by PET. *Nature, 375,* 54–58. http://dx.doi.org/10.1038/375054a0

Pecenka, N., & Keller, P. E. (2011). The role of temporal prediction abilities in interpersonal sensorimotor synchronization. *Experimental Brain Research, 211,* 505–515. http://dx.doi.org/10.1007/s00221-011-2616-0

Pfordresher, P. Q., & Halpern, A. R. (2013). Auditory imagery and the poor-pitch singer. *Psychonomic Bulletin and Review, 20,* 747–753. http://dx.doi.org/10.3758/s13423-013-0401-8

Richardson, A. (1969). *Mental imagery*. London: Routledge & Kegan Paul. http://dx.doi.org/10.1007/978-3-662-37817-5

Rudner, M., Rönnberg, J., & Hugdahl, K. (2005). Reversing spoken items—Mind twisting not tongue twisting. *Brain and Language, 92,* 78–90. http://dx.doi.org/10.1016/j.bandl.2004.05.010

Schaefer, R. S., Vlek, R. J., & Desain, P. (2011). Music perception and imagery in EEG: Alpha band effects of task and stimulus. *International Journal of Psychophysiology, 82,* 254–259.

Tinti, C., Cornoldi, C., & Marschark, M. (1997). Modality-specific auditory imaging and the interactive imagery effect. *European Journal of Cognitive Psychology, 9,* 417–436. http://dx.doi.org/10.1080/713752569

Vitrovic, D., & Biller, J. (2013). Musical hallucinations and forgotten tunes: A case report and brief literature review. *Frontiers in Neurology, 4,* 1–3.

Willander, J., & Baraldi, S. (2010). Development of a new Clarity of Auditory Imagery Scale. *Behavior Research Methods, 42,* 785–790. http://dx.doi.org/10.3758/BRM.42.3.785

Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Musical imagery and auditory cortex. *Neuron, 47,* 9–12. http://dx.doi.org/10.1016/j.neuron.2005.06.013

Zatorre, R. J., Halpern, A. R., & Bouffard, M. (2010). Mental reversal of imagined melodies: A role for the posterior parietal cortex. *Journal of Cognitive Neuroscience, 22,* 775–789. http://dx.doi.org/10.1162/jocn.2009.21239

Zvyagintsev, M., Clemens, B., Chechko, N., Mathiak, K. A., Sack, A. T., & Mathiak, K. (2013). Brain networks underlying mental imagery of auditory and visual information. *European Journal of Neuroscience, 37,* 1421–1434. http://dx.doi.org/10.1111/ejn.12140

Appendix A

The Bucknell Auditory Imagery Scale—Vividness (BAIS-V)

The following scale is designed to measure auditory imagery, or the way in which you "think about sounds in your head." For the following items you are asked to do the following: Read the item and consider whether you think of an image of the described sound in your head. Then rate the vividness of your image using the following "Vividness Rating Scale." If no image is generated, give a rating of 1.

Please feel free to use all of the levels in the scale when selecting your ratings.

Vividness Rating Scale

1	2	3	4	5	6	7
No Image Present at All			Fairly Vivid			As Vivid As The Actual Sound

Vividness Rating

1. For the first item, consider the beginning of the song "Happy Birthday."
   The sound of a trumpet beginning the piece._____
2. For the next item, consider ordering something over the phone.
   The voice of an elderly clerk assisting you._____
3. For the next item, consider being at the beach.
   The sound of the waves crashing against nearby rocks. _____
4. For the next item, consider going to a dentist appointment.
   The loud sound of the dentist's drill.______
5. For the next item, consider being present at a jazz club.
   The sound of a saxophone solo.______
6. For the next item, consider being at a live baseball game.
   The cheer of the crowd as a player hits the ball._____
7. For the next item, consider attending a choir rehearsal.
   The sound of an all-children's choir singing the first verse of a song. ______
8. For the next item, consider attending an orchestral performance of Beethoven's Fifth.
   The sound of the ensemble playing. ______
9. For the next item, consider listening to a rain storm.
   The sound of gentle rain. ______
10. For the next item, consider attending classes.
    The slow-paced voice of your English teacher.______

(*Appendices continue*)

11. For the next item, consider seeing a live opera performance.
    The voice of an opera singer in the middle of a verse.______

12. For the next item, consider attending a new tap-dance performance.
    The sound of tap-shoes on the stage.______

13. For the next item, consider a kindergarten class.
    The voice of the teacher reading a story to the children.______

14. For the next item, consider driving in a car.
    The sound of an upbeat rock song on the radio.______

Appendix B

The Bucknell Auditory Imagery Scale—Control (BAIS-C)

The following scale is designed to measure auditory imagery, or the way in which you "think about sounds in your head." For the following pairs of items you are asked to do the following: Read the first item (marked "a") and consider whether you think of an image of the described sound in your head. Then read the second item (marked "b") and consider how easily you could change your image of the first sound to that of the second sound and hold this image. Rate how easily you could make this change using the "Ease of Change Rating Scale." If no images are generated, give a rating of 1. Please read "a" first and "b" second for each pair. It may be necessary to cover up "b" so that you focus first on "a" for each pair.

Please feel free to use all of the levels in the scale when selecting your ratings.

Ease of Change Rating Scale

1	2	3	4	5	6	7
No Image Present at All			Could Change the Image but With Effort			Extremely Easy to Change the Image

Change Rating

1. For the first pair, consider attending a choir rehearsal.
   a. The sound of an all-children's choir singing the first verse of a song.
   b. An all-adults' choir now sings the second verse of the song. ______

2. For the next pair, consider being present at a jazz club.
   a. The sound of a saxophone solo.
   b. The saxophone is now accompanied by a piano.______

3. For the next pair, consider listening to a rain storm.
   a. The sound of gentle rain.
   b. The gentle rain turns into a violent thunderstorm.______

4. For the next pair, consider driving in a car.
   a. The sound of an upbeat rock song on the radio.
   b. The song is now masked by the sound of the car coming to a screeching halt. ______

(*Appendices continue*)

5. For the next pair, consider ordering something over the phone.
   a. The voice of an elderly clerk assisting you.
   b. The elderly clerk leaves and the voice of a younger clerk is now on the line.______

6. For the next pair, consider seeing a live opera performance.
   a. The voice of an opera singer in the middle of a verse.
   b. The opera singer now reaches the end of the piece and holds the final note. ______

7. For the next pair, consider going to a dentist appointment.
   a. The loud sound of the dentist's drill.
   b. The drill stops and you can now hear the soothing voice of the receptionist.______

8. For the next pair, consider the beginning of the song "Happy Birthday."
   a. The sound of a trumpet beginning the piece.
   b. The trumpet stops and a violin continues the piece.______

9. For the next pair, consider attending an orchestral performance of Beethoven's Fifth.
   a. The sound of the ensemble playing.
   b. The ensemble stops but the sound of a piano solo is present.______

10. For the next pair, consider attending a new tap-dance performance.
    a. The sound of tap-shoes on the stage.
    b. The sound of the shoes speeds up and gets louder.______

11. For the next pair, consider being at a live baseball game.
    a. The cheer of the crowd as a player hits the ball.
    b. Now the crowd boos as the fielder catches the ball._____

12. For the next pair, consider a kindergarten class.
    a. The voice of the teacher reading a story to the children.
    b. The teacher stops reading for a minute to talk to another teacher. ______

13. For the next pair, consider attending classes.
    a. The slow-paced voice of your English teacher.
    b. The pace of the teacher's voice gets faster at the end of class. ______

14. For the next pair, consider being at the beach.
    a. The sound of the waves crashing against nearby rocks.
    b. The waves are now drowned out by the loud sound of a boat's horn out at sea. _____

Received May 8, 2014
Revision received March 28, 2015
Accepted April 15, 2015 ■