

# The Mental Representation of Music Notation: Notational Audiation

Warren Brodsky and Yoav Kessler  
Ben-Gurion University of the Negev

Bat-Sheva Rubinstein  
Tel Aviv University

Jane Ginsborg  
Royal Northern College of Music

Avishai Henik  
Ben-Gurion University of the Negev

This study investigated the mental representation of music notation. Notational audiation is the ability to internally “hear” the music one is reading before physically hearing it performed on an instrument. In earlier studies, the authors claimed that this process engages music imagery contingent on subvocal silent singing. This study refines the previously developed *embedded melody task* and further explores the phonatory nature of notational audiation with throat-audio and larynx-electromyography measurement. Experiment 1 corroborates previous findings and confirms that notational audiation is a process engaging kinesthetic-like covert excitation of the vocal folds linked to phonatory resources. Experiment 2 explores whether covert rehearsal with the mind’s voice also involves actual motor processing systems and suggests that the mental representation of music notation cues manual motor imagery. Experiment 3 verifies findings of both Experiments 1 and 2 with a sample of professional drummers. The study points to the profound reliance on phonatory and manual motor processing—a dual-route stratagem—used during music reading. Further implications concern the integration of auditory and motor imagery in the brain and cross-modal encoding of a unisensory input.

*Keywords:* music reading, notational audiation, embedded melody, music expertise, music imagery

Does the reading of music notation produce aural images in trained musicians? If so, what is the nature of these images triggered during sight reading? Is the process similar to other actions involving “inner hearing,” such as subvocalization, inner voice, or inner speech? The current study was designed to investigate the mental representation of music notation. Researchers

have long been aware that when performing from notated music, highly trained musicians rely on music imagery just as much, if not more, than on the actual external sounds themselves (see Hubbard & Stoeckig, 1992). Music images possess a sensory quality that makes the experience of imagining music similar to that of perceiving music (Zatorre & Halpern, 1993; Zatorre, Halpern, Perry, Meyer, & Evans, 1996). In an extensive review, Halpern (2001) summarized the vast amount of evidence indicating that brain areas normally engaged in processing auditory information are recruited even when the auditory information is internally generated. Gordon (1975) called the internal analog of aural perception *audiation*; he further referred to *notational audiation* as the specific skill of “hearing” the music one is reading before physically hearing it performed on an instrument.

---

Warren Brodsky, Music Science Research, Department of the Arts, Ben-Gurion University of the Negev, Beer-Sheva, Israel; Yoav Kessler and Avishai Henik, Department of Psychology and Zlotowski Center for Neuroscience, Ben-Gurion University of the Negev; Bat-Sheva Rubinstein, Department of Theory and Composition, Buchmann-Mehta School of Music, Tel Aviv University, Tel Aviv, Israel; Jane Ginsborg, Center for Excellence in Teaching and Learning, Royal Northern College of Music, Manchester, England.

This work was supported by Grant 765/03-34.0 to Warren Brodsky from the Israel Science Foundation, funded by the Israel Academy of Sciences and Humanities; and by funding to Jane Ginsborg from the Royal Northern College of Music. We offer gratitude to college directorates Tomer Lev (of the Buchmann-Mehta School of Music, Tel Aviv, Israel) and Edward Gregson, Linda Merrick, and Anthony Gritten (of the Royal Northern College of Music). In addition, we offer a hearty thank you to maestro Moshe Zorman for the high-quality original music materials. Further, we deeply appreciate all the efforts of Sagi Shorer, Shimon Even-Zur, and Yehiel Braver (Klei Zemer Yamaha, Tel Aviv, Israel) as well as those of master drummer Rony Holan and Yosef Zucker (Or-Tav Publications, Kfar Saba, Israel) for their kind permission to use published notation and audio examples. We would like to thank all of the musicians who participated in the studies—they are our partners in music science research.

Correspondence concerning this article should be addressed to Warren Brodsky, Department of the Arts, Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel. E-mail: wbrodsky@bgu.ac.il

Almost 100 years ago, the music psychologist Carl Seashore (1919) proposed the idea that a *musical mind* is characterized by the ability to “think in music,” or produce music imagery, more than by any other music skill. Seventy years earlier, in the introduction to his piano method, the Romantic composer Robert Schumann (1848/1967) wrote to his students, “You must get to the point that you can hear music from the page” (p. 402). However, very little is known about the nature of the cognitive process underlying notational audiation. Is it based on auditory, phonatory, or manual motor resources? How does it develop? Is it linked to expertise of music instrument, music literacy, music theory, sight reading, or absolute perfect pitch?

Historic doctrines have traditionally put faith in *sight singing* as the musical aid to developing mental imagery of printed music (Karpinski, 2000), and several pedagogical methods (e.g., Gordon, 1993; Jacques-Dalcroze, 1921) claim to develop inner hearing.

Nonetheless, in the absence of experimental investigations, as far as cognitive science is concerned, evidence for imagery as a cued response to music notation is essentially anecdotal. It has been proposed that during auditory and music imagery, the inner voice supplies a kinesthetic stimulus that acts as raw material via some motor output or plan, which is detectable by channels of perception associated with the phonological system (Intons-Peterson, 1992; MacKay, 1992; Smith, Reisberg, & Wilson, 1992). Yet hardly any studies have targeted imagery of music triggered by music notation, and perhaps one reason for their scarcity is lack of a reliable method whereby such internal processes can be teased out and examined.

In one study, Waters, Townsend, and Underwood (1998; Experiment 6) used a *same-different* paradigm in which trained pianists silently read 30 cards each containing a single bar of piano music. They found that pianists were successful at matching silently read music notation to a subsequently presented auditory sequence. Yet the study did little to demonstrate that task performance was based on evoked music imagery rather than on structural harmonic analyses or on guesswork based on visual surface cues found in the notation. In another study, Wöllner, Halfpenny, Ho, and Kurosawa (2003) presented voice majors with two 20-note single-line C-major melodies. The first was read silently without distraction, whereas the other was read silently with concurrent auditory distraction (i.e., the Oscar Peterson Quartet was heard in the background). After reading the notation, the participants sang the melody aloud. As no differences surfaced between normal sight reading (viewed as “intact” inner hearing) and distracted sight reading (viewed as “hampered” inner hearing), the authors concluded that “inner-hearing is . . . less important in sight-reading than assumed” (Wöllner et al., 2003, p. 385). Yet this study did little to establish that sight reading was equivalent to or based on inner hearing or that a valid measure of the internal process is sight-singing accuracy.

In an earlier set of studies (Brodsky, Henik, Rubinstein, & Zorman, 1998, 1999, 2003), we developed a paradigm that exploited the compositional technique of theme and variation. This method allowed for a well-known theme to be embedded in the notation of a newly composed, stand-alone, embellished phrase (hereafter referred to as an *embedded melody* [EM]); although the original well-known theme was visually indiscernible, it was still available to the “mind’s ear.” In the experiments, after the participants silently read the notation of an EM, they heard a tune and had to decide whether this excerpt was the well-known embedded theme (i.e., target) or a different tune (i.e., melodic lure). We found that only a third of the highly skilled musicians recruited were able to perform the task reliably—although all of them were successful when the embellished phrases incorporating EMs were presented aloud. It might be inferred from these results that only one in three musicians with formal advanced training has skills that are sufficient to internally hear a well-known embedded theme when presented graphically.

However, from the outset, we felt that the ability to recognize original well-known embedded themes does not necessarily provide conclusive evidence in itself that notational audiation exists (or is being used). Empirical caution is warranted here as there may be several other explanations for how musicians can perform the EM task, including informed guesswork and harmonic structural analyses. Hence, on the basis of the conception that overt

measurement can provide evidence of covert mental processes, whereby one infers the existence of a process by observing some effect caused by that process, we used a distraction paradigm (Brodsky et al., 1998, 1999, 2003). Specifically, we argued that if theme recognition can be hampered or blocked by conditions that may be assumed to engage audiation processes, then we could be justified in concluding that notational audiation was in operation during nondistracted score reading. Accordingly, four music-reading conditions were used: normal nondistracted sight reading, sight reading with concurrent auditory distraction, sight reading with concurrent rhythmic distraction, and sight reading with concurrent phonatory interference. The study found that phonatory interference impaired recognition of themes more than did the other conditions, and consequently, we surmised that notational audiation occurs when the silent reading of music notation triggers auditory imagery resulting in measurable auditory perception. Therefore, we suggested that notational audiation elicits kinesthetic-like covert phonatory processes such as silent singing.

Sight reading music notation is an extremely complicated task: Elliott (1982) delineated seven predictor variables of ability; Waters, Underwood, and Findlay (1997) showed there to be at least three different types of processing ability required; and Lee (2004) identified 20 component skills. In a more recent study, Kopiez, Weihs, Ligges, and Lee (2006) formulated three composite groupings of requisite skills engaged during sight reading, each with associate subskills; a total of 27 subskills were documented. The grouping concerned with practice-related skills is of particular relevance here because *auditory imagery* is listed among its associated subskills. On the basis of previous studies (e.g., Lehmann & Ericsson, 1993, 1996), Kopiez et al. assumed that sight reading relies to some extent on the ability to generate an aural image (i.e., inner hearing) of the printed score. To test this assumption, they used our EM task (Brodsky et al., 1998, 1999, 2003). In Kopiez et al.’s study, participants read notated variations of five well-known classical piano pieces, and then after they had heard a tune (i.e., the original target theme or a lure melody), they had to decide whether the excerpt was the theme embedded in the variation previously read. The dependent variable used was  $d'$ . On the surface, it might appear that Kopiez et al. validated our experimental task. However, the procedures they used were only faintly similar to our protocol, and unfortunately, the  $d'$  calculations were not published.

The nature of notational audiation is elusive, and one has only to look at the various descriptive labels used in the literature to understand this bafflement. The skill has been proposed to be a process of inner hearing or *auralization* (Karpinski, 2000; Larson, 1993; Martin, 1952) as well as a form of silent singing (Walters, 1989). The resulting internal phenomenon has been perceived as an “acoustic picture” or “mental score” (Raffman, 1993), as supplied by the “hearing eye” (Benward & Carr, 1999) or the “seeing ear” (Benward & Kolosic, 1996). Yet these phenomena should not necessarily be equated with each other, nor do they fundamentally represent the same processes. Our previous findings (Brodsky et al., 1998, 1999, 2003) suggest that notational audiation is a process engaging kinesthetic-like covert excitation of the vocal folds, and hence we have theorized that the mind’s representation of music notation might not have anything at all to do with hearing per se.

Kalakoski (2001) highlighted the fact that the literature is unequivocal in pointing to the underlying cognitive systems that maintain and process auditory representations as expressed in

music images. Gerhardstein (2002) cited Seashore's (1938) landmark writings, suggesting that kinesthetic sense is related to the ability to generate music imagery, and Reisberg (1992) proposed the crisscrossing between aural and oral channels in relation to the generation of music imagery. However, in the last decade, the triggering of music images has been linked to motor memory (Mikumo, 1994; Petsche, von Stein, & Filz, 1996), whereas neuromusical studies using positron-emission tomography and functional MRI technologies (Halpern, 2001; Halpern & Zatorre, 1999) have found that the supplementary motor area (SMA) is activated in the course of music imagery—especially during covert mental rehearsal (Langheim, Callicott, Mattay, Duyn, & Weinberger, 2002). Accordingly, the SMA may mediate rehearsal that involves motor processes such as humming. Further, the role of the SMA during imagery of familiar melodies has been found to include both auditory components of hearing the actual song and carrier components, such as an image of subvocalizing, of moving fingers on a keyboard, or of someone else performing (Schneider & Godoy, 2001; Zatorre et al., 1996).

However, it must be pointed out that although all of the above studies purport to explore musical imagery and imagined musical performance—that is, they attempt to tease out the physiological underpinnings of musical cognition in music performance without sensorimotor and auditory confounds of overt performance—some findings may be more than questionable. For example, Langheim et al. (2002) themselves concluded the following:

While cognitive psychologists studying mental imagery have demonstrated creative ways in which to ascertain that the imagined task is in fact being imagined, our study had no such control. We would argue, however, that to the musically experienced, imagining performance of a musical instrument can more closely be compared to imagining the production of language in the form of subvocal speech. (p. 907)

Nonetheless, no neuromusical study has explored imagery triggered by music notation. The one exception is a study conducted by Schurmann, Raji, Fujiki, and Hari (2002), who had 11 trained musicians read music notation while undergoing magnetoencephalogram scanning. During the procedure, Schurmann et al. presented the participants with a four-item test set, each item consisting of only one notated pitch ( $G_1$ ,  $A_1$ ,  $B^p_1$ , and  $C_2$ ). Although it may be necessary to use a minimalist methodology to shed light on the time course of brain activation in particular sites while magnetoencephalogram is used to explore auditory imagery, it should be pointed out that such stimuli bear no resemblance to real-world music reading, which never involves just one isolated note. It could be argued, therefore, that this study conveys little insight into the cognitive processes that underlie notational audiation.

Neuroscience may purport to finally have found a solution to the problem of measuring internal phenomena, and that solution involves functional imaging techniques. Certainly this stance relies on the fact that the underlying neural activity can be measured directly rather than by inferring its presence. However, establishing what is being measured remains a major issue. Zatorre and Halpern (2005) articulate criticism about such a predicament with their claim that “merely placing subjects in a scanner and asking them to imagine some music, for instance, simply will not do, because one will have no evidence that the desired mental activity is taking place” (p. 9). This issue can be addressed by the development of behavioral paradigms measuring overt responses that

either depend on or correlate with the internal activity under investigation. Only then, by recruiting these responses in combination with state-of-the-art neuroimaging techniques, can cognitive neuroscientists make strides in uncovering the particular processes underlying music imagery in their effort to understand the musical mind. To this end, we carried out the current study. Our goal was to refine our previously developed EM task by developing new stimuli that are more exacting in their level of difficulty, highly flexible in their functionality, and able to transcend the boundaries of music literacy. We developed this task as a means to demonstrate music imagery triggered by music notation, with which we could then explore the phonatory nature of notational audiation in conjunction with physiological measurements of throat-audio and larynx-electromyography (EMG) recordings.

### Experiment 1

The purpose of Experiment 1 was to replicate and validate our previous findings with a new set of stimuli. This new set (described below) was designed specifically to counteract discrepancies that might arise between items allocated as targets or as lures, as well as to rule out the possibility that task performance could be biased by familiarity with the Western classical music repertoire. We achieved the former goal by creating pairs of stimuli that can be presented in a counterbalanced fashion (once as a target and once as a lure); we achieved the latter goal by creating three types of stimuli (that were either well-known, newly composed, or a hybrid of both). With these in hand, we implemented the EM task within a distraction paradigm while logging audio and EMG physiological activity. We expected that the audio and EMG measures would reveal levels of subvocalization and covert activity of the vocal folds. Further, we expected there to be no differences in performance between the three stimulus types.

### Method

*Participants.* Initially, 74 musicians were referred and tested. Referrals were made by music theory and music performance department heads at music colleges and universities, by ear-training instructors at music academies, and by professional music performers. The criteria for referral were demonstrable high-level abilities in general music skills relating to performance, literature, theory, and analysis, as well as specific music skills relating to dictation and sight singing. Out of the 74 musicians tested, 26 (35%) passed a prerequisite threshold inclusion criterion demonstrating notational audiation ability; the criterion adopted for inclusion in the study (from Brodsky et al., 1998, 1999, 2003) represents a significant task performance ( $p < .05$  using a sign test) during nondistracted sight reading, reflecting a 75% success rate in a block of 12 items. Hence, this subset ( $N = 26$ ) represents the full sample participating in Experiment 1. The participants were a multicultural group of musicians comprising a wide range of nationalities, ethnicities, and religions. Participants were recruited and tested at either the Buchmann-Mehta School of Music (formerly the Israel Academy of Music) in Tel Aviv, Israel, or at the Royal Northern College of Music in Manchester, England. As there were no meaningful differences between the two groups in terms of demographic or biographical characteristics, nor in terms of their task performance, we have pooled them into a combined

sample group. In total, there were slightly more women (65%) than men, with the majority (73%) having completed a bachelor's of music as their final degree; 7 (27%) had completed postgraduate music degrees. The participants' mean age was 26 years ( $SD = 8.11$ , range = 19–50); they had an average of 14 years ( $SD = 3.16$ , range = 5–20) of formal instrument lessons beginning at an average age of 7 years ( $SD = 3.17$ , range = 4–17) and had an average of 5 years ( $SD = 3.99$ , range = 1–16) of formal ear-training lessons beginning at an average age of 11 ( $SD = 6.41$ , range = 1–25). In general, they were right-handed (85%), pianists (80%), and music performers (76%). Less than half (38%) of the sample claimed to possess absolute perfect pitch. Eighty-eight percent described themselves as avid listeners of classical music. Using a 4-point Likert scale (1 = *not at all*, 4 = *highly proficient*), the participants reported an overall high level of confidence in their abilities to read new unseen pieces of music ( $M = 3.27$ ,  $SD = 0.667$ ), to “hear” the printed page ( $M = 3.31$ ,  $SD = 0.617$ ), and to remember music after a one-time exposure ( $M = 3.19$ ,  $SD = 0.567$ ). However, they rated their skill of concurrent music analysis while reading or listening to a piece as average ( $M = 2.88$ ,  $SD = 0.567$ ). Finally, more than half (54%) of the musicians reported that the initial learning strategy they use when approaching a new piece is to read silently through the piece; the others reported playing through the piece (35%) or listening to an audio recording (11%).

**Stimuli.** Twenty well-known operatic and symphonic themes were selected from Barlow and Morgenstern (1975). Examples are seen in Figure 1 (see target themes). Each of these original well-known themes was then embedded into a newly composed embellished phrase (i.e., an EM) by a professional composer–arranger using compositional techniques such as quasi-contrapuntal treatment, displacement of registers, melodic ornamentation, and rhythmic augmentation or diminution (see Figure 1, embedded melodies). Each pair was then matched to another well-known theme serving as a melodic lure; the process was primarily one of locating a tune that could mislead the musician reader into assuming that the subsequently heard audio exemplar was the target melody embedded in the notation even though it was not. Hence, the process of matching lures to EMs often involved sophisticated deception. The decisive factor used in choosing the melodies for lures was that there be thematic or visual similarities of at least seven criteria, such as contour, texture, opening interval, rhythmic pattern, phrasing, meter, tonality, key signature, harmonic structure, and music style (see Figure 1, melodic lures). The lures were then treated in a similar fashion as the targets—that is, they were used as themes to be embedded in an embellished phrase. This first set of stimuli was labeled Type I.

A second set of 20 well-known themes, also selected from Barlow and Morgenstern (1975), was treated in a similar fashion except that the lures were composed for the purposes of the experiment; this set of stimuli was labeled Type II. Finally, a third set of 20 themes was composed for the purposes of the experiment, together with newly composed melodic lures; this set of stimuli was labeled Type III. All three stimulus types were evaluated by B.-S.R. (head of a Music Theory, Composition, and Conducting Department) with respect to the level of difficulty (i.e., the recognizability of the original well-known EM in the embellished phrase) and the structural and harmonic fit of *target–lure functional reversibility* (whereby a theme chosen as a lure can become

a target while the original target can newly function in a mirrored fashion as the appropriate melodic lure). It should be pointed out that although functional reversibility is commonplace for visual or textual stimuli, such an approach is clearly more of a challenge when developing music stimuli. Further, the functional reversibility of target melodies and melodic lures has not as yet been reported in the music cognition literature. Subsequently, of the 10 foursomes in each stimulus type, roughly 40% were deemed inappropriate and dropped from the test pool. The remaining 18 reversible item pairs (36 items) were randomly assigned to three blocks (one block per experimental condition), each containing an equal number of targets, lures, and types. Each block was also stratified for an equal number of items on the basis of tonality (major or minor) and meter (2/4, 4/4, 3/4, or 6/8). All items were recorded live (performed by the composer–arranger) with a Behringer B-2 (Behringer) dual-diaphragm studio condenser microphone suspended over a Yamaha upright piano (with lid open), to a portable Korg 1600 (Korg) 16-track digital recording-studio desk. The recordings were cropped with Soundforge XP4.5 (RealNetworks) audio-editing package and were standardized for volume (i.e., reduced or enhanced where necessary). On average, the audio files (i.e., target and lure tunes) were approximately 24 s ( $SD = 6.09$  s) in exposure length. The notation was produced with the Finale V.2005 (Coda Music Technologies, MakeMusic) music-editing package and formatted as 24-bit picture files. The notation was presented as a G-clef single-line melody, with all stems pointing upward, placed in standardized measure widths, of an average of 16 bars in length ( $SD = 4.98$ , range = 8–24 bars).

As reading music notation clearly involves several actions not necessarily linked to the music skill itself, such as fixed seating position, visual scanning and line reading, and analytic processes, a three-task pretest baseline control (BC) task was developed. Two 650-word texts were chosen for silent reading: the Hebrew text was translated from Brodsky (2002), whereas the English text was taken from Brodsky (2003). In addition, there were five mathematical number line completion exercises (e.g., 22, 27, 25, 30, ?) selected from the 10th-grade math portion of the Israel High School National Curriculum.

**Apparatus.** We used two laptop computers, an integrated biomonitor, and experiment delivery software. The experiment ran on a ThinkPad T40 (IBM) with Intel Pentium M 1.4-GHz processor, 14-in. TFT SXGA plus LCD screen, and an onboard SoundMAX audio chipset driving a palm-sized TravelSound (Creative) 4-W digital amplifier with two titanium-driver stereo speakers. The biomonitor ran on a ThinkPad X31 (IBM) with Intel Centrino 1.4-GHz processor and 12-in. TFT XGA LCD screen. The three-channel integrated biomonitor was a NovaCorder AR-500 Custom (Atlas Researches, Israel), with a stretch-band, Velcro-fastened strap to mount a throat-contact microphone (to record subvocalizations, with full-scale integrated output) and two reusable gold-dry EMG electrodes (to monitor covert phonatory muscle activity, each composed of a high-gain low-noise Atlas Bioamplifier [100–250 Hz bandwidth, full scale 0.0–25.5  $\mu$ V]), an optic-fiber serial PC link, and an integrated PC software package (for recording, display, and editing of data files). The experiment was designed and operated with E-Prime (Version 1.1; Psychology Software Tools).

**Design and test presentation.** The overall plan included a pretest control task and an experiment set. It should be pointed out



Embedded Melody



**Target Theme: Boccherini, Minuet (1st theme in A major)**  
3rd Movement from “Quintet For Strings in E Major.

**Melodic Lure: Beethoven, Minuet For Piano in G Major**  
(transcribed to A major)



Embedded Melody



**Target Theme: Beethoven, Minuet For Piano in G Major**  
(transcribed to A major)

**Melodic Lure: Boccherini, Minuet (1st theme in A major)**  
3rd Movement from “Quintet For Strings in E Major.



Figure 1. Embedded melodies: Type I. The embedded melodies were created by Moshe Zorman, Copyright 2003. Used with kind permission of Moshe Zorman.

that although each participant underwent the control task (serving as a baseline measurement) prior to the experiment set, the two tasks are of an entirely different character, and hence we assumed that there would be no danger of carryover effects. The BC task comprised three subtasks: sitting quietly (90 s), silent text reading (90 s), and completing five mathematical number lines. The order of these subtasks was counterbalanced across participants. The experimental task required the participants to silently read and recognize themes embedded in the music notation of embellished phrases and then to correctly match or reject tunes heard aloud

after the notation had disappeared from the computer screen. Sight reading was conducted under three conditions: (a) normal, non-distracted, silent music reading (NR); (b) silent music reading with concurrent rhythmic distraction (RD), created by having the participant tap a steady pulse (knee patch) while hearing an irrelevant random cross-rhythm beaten out (with a pen on tabletop) by the experimenter; and (c) music reading with concurrent phonatory interference (PI), created by the participant him- or herself by singing a traditional folk song but replacing the words of the song with the sound *la*. The two folk songs used in the PI condition were

“David, King of Israel” (sung in Hebrew in those experiments conducted in Israel) and “Twinkle, Twinkle, Little Star” (sung in English in those experiments conducted in England); both songs were chosen for their multicultural familiarity, and both were in a major key tonality in a 4/4 meter. In total, there were 36 trials, each consisting of an EM paired with either the original target theme or a melodic lure. Each block of 12 pairs represented one of the three music-reading conditions. We presented condition (NR, RD, PI), block (item set 1–12, 13–24, 25–36), and pairs (EM–target, EM–lure) in a counterbalanced form to offset biases linked to presentation order.

**Procedure.** The experiment ran for approximately 90 min and consisted of three segments: (a) fitting of biomonitor and pretest BC; (b) oral or written instructions, including four-item demonstration–practice trial; and (c) 36-trial EM task under three reading conditions. In a typical session, each participant was exposed to the following sequence of events: The study was introduced to the participant, who signed an informed consent form and completed a one-page questionnaire containing demographic information and self-ranking of music skills. Then, participants were fitted with a throat-contact microphone and two reusable gold-dry EMG electrodes mounted on a stretch-band, Velcro-fastened choker strap; the throat-contact microphone was placed over the thyroid cartilage (known as the laryngeal prominence or, more commonly, the Adam’s apple) with each electrode positioned roughly 5 cm posterior to the right and left larynges (i.e., the voice box housing the vocal cords). While seated alongside the experimenter, the participants completed the BC task. Thereafter, instructions were read orally, and each participant was exposed to a demonstration and four practice trials for clarification of the procedure and experimental task. The participants were instructed to silently read the music notation in full (i.e., not to focus on just the first measure), to respond as soon as possible when hearing a tune, and to try not to make errors. Then, the notation of the first EM appeared on the screen and stayed in view for up to 60 s (or until a key was pressed in a self-paced manner). After 60 s (or the key press), the EM disappeared and a tune was heard immediately. The participants were required to indicate as quickly as possible whether the tune heard was the original theme embedded in the embellished phrase. They indicated their response by depressing a color-coded key on either side of the keyboard space bar; green stickers on the *Ctrl* keys indicated that the tune heard was the *original* theme, and red stickers on the *Alt* keys indicated that it was *not original*. A prompt indicating the two response categories with associated color codes appeared in the center of the screen. Response times were measured in milliseconds from the onset of

the audio file to the key press. Subsequently, the second trial began, and so forth. The procedure in all three conditions was similar. It should be noted, however, that the rhythmic distracter in the RD condition varied from trial to trial because the pulse tapping was self-generated and controlled by the participant, and the cross-rhythms beaten by the experimenter were improvised (sometimes on rhythmic patterns of previous tunes). Similarly, the tonality of the folk song sung in the PI condition varied from trial to trial because singing was self-generated and controlled by the participants; depending on whether participants possessed absolute perfect pitch, they may have changed keys for each item to sing in accordance with the key signature as written in the notation.

**Biomonitor analyses.** The Atlas Biomonitor transmits serial data to a PC in the form of a continuous analog wave form. The wave form is then decoded during the digital conversion into millisecond intervals. By synchronizing the T40 laptop running the experiment with the X31 laptop recording audio–EMG output, we were able to data link specific segments of each file representing the sections in which the participants were reading the music notation; because music reading was limited to a 60-s exposure, the maximum sampling per segment was 60,000 data points. The median output of each electrode per item was calculated and averaged across the two channels (left–right sides) to create a mean output per item. Then, the means of all correct items in each block of 12 items were averaged into a grand mean EMG output per condition. The same procedure was used for the audio data.

## Results

For each participant in each music-reading condition (i.e., NR, RD, PI), the responses were analyzed for percentage correct (PC; or overall success rate represented by the sum of the percentages of correctly identified targets and correctly rejected lures), hits (i.e., correct targets), false alarms (FAs),  $d'$  (i.e., an index of detection sensitivity), and response times (RTs). As can be seen in Table 1, across the conditions there was a decreasing level of PC, a decrease in percentage of hits with an increase in percentage of FAs, and hence an overall decrease in  $d'$ . Taken together, these results seem to indicate an escalating level of impairment in the sequenced order NR–RD–PI. The significance level was .05 for all the analyses.

Each dependent variable (PC, hits, FAs,  $d'$ , and RTs) was entered into a separate repeated measures analysis of variance (ANOVA) with reading condition as a within-subjects variable. There were significant effects of condition for PC,  $F(2, 50) = 10.58$ ,  $MSE = 153.53$ ,  $\eta_p^2 = .30$ ,  $p < .001$ ; hits,  $F(2, 50) = 10.55$ ,

Table 1  
Experiment 1: Descriptive Statistics of Behavioral Measures by Sight Reading Condition

Condition	PC		Hits		FAs		$d'$		Median RTs (s)	
	%	SD	%	SD	%	SD	<i>M</i>	SD	<i>M</i>	SD
NR	80.1	5.81	89.1	13.29	28.9	12.07	2.94	1.36	12.65	5.98
RD	67.3	14.51	73.1	22.60	38.5	21.50	1.68	1.68	13.65	6.98
PI	65.7	17.70	68.5	25.10	37.2	26.40	1.43	2.03	13.56	5.61

*Note.* PC = percentage correct; FA = false alarm; RT = response time; NR = nondistracted reading; RD = rhythmic distraction; PI = phonatory interference.

$MSE = 286.47$ ,  $\eta_p^2 = .30$ ,  $p < .001$ ; and  $d'$ ,  $F(2, 50) = 6.88$ ,  $MSE = 2.489$ ,  $\eta_p^2 = .22$ ,  $p < .01$ ; as well as a near significant effect for FAs,  $F(2, 50) = 2.44$ ,  $MSE = 290.17$ ,  $\eta_p^2 = .09$ ,  $p = .09$ . There was no effect for RTs. Comparison analysis showed that as there were no significant differences between the distraction conditions themselves, these effects were generally due to significant differences between the nondistracted and distraction conditions (see Table 2).

Subsequently, for each participant in each stimulus type (I, II, III), the frequency of correct responses was analyzed for PC, hits, FAs,  $d'$ , and RTs. As can be seen in Table 3, there were few significant differences between the different stimulus types. In general, this finding suggests that musicians were just as good readers of notation regardless of whether the embedded theme in the embellished phrase had been previously known or was in fact newly composed music.

Each dependent variable was entered into a separate repeated measures ANOVA with stimulus type as a within-subjects variable. There were significant effects of stimulus type for RTs,  $F(2, 50) = 4.19$ ,  $MSE = 749,314$ ,  $\eta_p^2 = .14$ ,  $p < .05$ ; no effects for PC, hits, FAs, or  $d'$  surfaced. Comparison analysis for RTs showed that effects were due to significant differences between well-known (shortest RTs) and hybrid stimuli (longest RTs; see Table 3).

Finally, the audio and EMG data relating to correct responses for each participant in each condition were calculated for each music-reading segment; measurements taken during the pretest control task (BC) representing a baseline level data were also calculated (see Table 4).

The data for each participant with condition (baseline and three music-reading conditions) as a within-subjects variable were entered into a repeated measures ANOVA. There were significant effects of condition for audio output,  $F(3, 54) = 31.61$ ,  $MSE = 461.56$ ,  $\eta_p^2 = .64$ ,  $p < .0001$ ; as well as for EMG output,  $F(3, 60) = 14.11$ ,  $MSE = 0.6048$ ,  $\eta_p^2 = .41$ ,  $p < .0001$ . As can be seen in Table 5, comparison analysis for audio data revealed no significant differences between the silent-reading conditions, but there were significant differences between both silent-reading conditions and the vocal condition. Further, in an analysis of audio data, we found no significant differences between the BC tasks and either silent-reading condition, but significant differences did surface between the control tasks and the vocal condition. In addition, comparison analysis of EMG data showed no significant differ-

ences between the silent-reading conditions, but there were significant differences between both silent-reading conditions and the vocal condition. Finally, in our analysis of EMG data, we also found significant differences between the BC tasks and all three music-reading conditions, thus suggesting covert activity of the vocal folds in the NR and RD conditions.

Last, we explored the biographical data supplied by the participants (i.e., age, gender, handedness, possession of absolute perfect pitch, onset age of instrument learning, accumulated years of instrument lessons, onset age of ear training, and number of years of ear-training lessons) for correlations and interactions with performance outcome variables (PC, hits, FAs,  $d'$ , and RTs) for each reading condition and stimulus type. The analyses found a negative relationship between onset age of instrument learning and the accumulated years of instrument lessons ( $R = -.71$ ,  $p < .05$ ), as well as onset age of ear training and the accumulated years of instrument lessons ( $R = -.41$ ,  $p < .05$ ). Further, a negative correlation surfaced between stimuli Type III–hits and age ( $R = -.39$ ,  $p < .05$ ); there was also a positive correlation between stimuli Type I– $d'$  and age ( $R = .43$ ,  $p < .05$ ). These correlations indicate relationships of age with dedication (the younger a person is at onset of instrument or theory training, the longer he or she takes lessons) and with training regimes (those who attended academies between 1960 and 1980 seem to be more sensitive to well-known music, whereas those who attended academies from 1980 onward seem to be more sensitive to newly composed music). Further, the analyses found only one main effect of descriptive variables for performance outcome variables; there was a main effect of absolute perfect pitch for RTs,  $F(1, 24) = 7.88$ ,  $MSE = 77,382,504.00$ ,  $\eta_p^2 = .25$ ,  $p < .01$ , indicating significantly decreased RTs for possessors of absolute perfect pitch ( $M = 9.74$  s,  $SD = 4.99$  s) compared with nonpossessors ( $M = 15.48$  s,  $SD = 5.14$  s). No interaction effects surfaced.

Discussion

The results of Experiment 1 constitute a successful replication and validation of our previous work. The current study offers tighter empirical control with the current set of stimuli: The items used were newly composed and tested and then selected on the basis of functional reversibility (i.e., items functioning as both targets and lures). We assumed that such a rigorous improvement of music items used (in comparison to our previously used set of stimuli) would offer assurances that any findings yielded by the study would not be contaminated by contextual differences between targets and lures—a possibility we previously raised. Further, we enhanced these materials by creating three stimulus types: item pairs (i.e., targets–lures) that were either well-known, newly composed, or a hybrid of both. We controlled each block to facilitate analyses by both function (target, lure) and type (I, II, III).

The results show that highly trained musicians who were proficient in task performance during nondistracted music reading were significantly less capable of matching targets or rejecting lures while reading EMs with concurrent RD or PI. This result, similar to the findings of Brodsky et al. (1998, 1999, 2003), suggests that our previous results did not occur because of qualitative differences between targets and lures. Moreover, the results show no significant differences between the stimulus types for

Table 2  
Experiment 1: Contrasts Between Sight Reading Conditions for Behavioral Measures

Dependent variable	Conditions	$F(1, 25)$	$MSE$	$\eta_p^2$	$p$
PC	NR vs. RD	16.64	128.41	.40	<.001
	NR vs. PI	15.32	176.55	.38	<.001
Hits	NR vs. RD	14.67	227.56	.37	<.001
	NR vs. PI	17.39	314.53	.41	<.001
FAs	NR vs. RD	4.88	246.36	.16	<.05
$d'$	NR vs. RD	10.15	2.03	.29	<.01
	NR vs. PI	8.37	3.532	.25	<.01

Note. PC = percentage correct; NR = nondistracted reading; RD = rhythmic distraction; PI = phonatory interference; FA = false alarm.

Table 3

Experiment 1: Descriptive Statistics of Behavioral Measures by Stimuli Type and Contrasts Between Stimuli Type for Behavioral Measures

Condition	% PC		% hits		% FAs		$d'$		Median RTs (s)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Type I	72.8	13.20	77.6	19.90	32.1	19.40	2.36	1.99	12.78	5.68
Type II	70.4	16.00	78.8	18.60	38.4	26.50	1.95	1.67	14.96	6.42
Type III	70.1	11.10	74.6	19.60	33.9	19.70	1.72	1.32	13.63	6.75

Note. For the dependent variable RT, Type I versus Type II,  $F(1, 25) = 8.39$ ,  $MSE = 7,355,731$ ,  $\eta_p^2 = .25$ ,  $p < .01$ . PC = percentage correct; FA = false alarm; RT = response time.

overall PC. Hence, we might also rule out the possibility that music literacy affects success rate in our experimental task. That is, one might have thought that previous exposure to the music literature would bias task performance as it is the foundation of a more intimate knowledge of well-known themes. Had we found this, it could have been argued that our experimental task was not explicitly exposing the mental representation of music notation but rather emulating a highly sophisticated cognitive game of music literacy. Yet, there were no significant differences between the stimulus types, and the results show not only an equal level of PC but also equal percentages of hits and FAs regardless of whether the stimuli presented were well-known targets combined with well-known lures, well-known targets combined with newly composed lures, or newly composed targets with newly composed lures. It would seem, then, that a proficient music reader of familiar music is just as proficient when reading newly composed, previously unheard music.

Further, the results show no significant difference between RTs in the different reading conditions; in general, RTs were faster with well-known themes (Type I) than with newly composed music (Type II or Type III). Overall, this finding conflicts with findings from our earlier studies (Brodsky et al., 1998, 1999, 2002). In this regard, several explanations come to mind. For example, one possibility is that musicians are accustomed to listening to music in such a way that they are used to paying attention until the final cadence. Another possibility is that musicians are genuinely interested in listening to unfamiliar music materials and therefore follow the melodic and harmonic structure even when instructed to identify it as the original or that they otherwise identify it as quickly as possible and then stop listening. In any event, in the current study we found that RTs were insensitive to differences

Table 4

Experiment 1: Descriptive Statistics of Audio and EMG Output by Sight Reading Condition

Condition	Audio ( $\mu V$ )		EMG ( $\mu V$ )	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
BC	4.40	0.33	1.70	0.37
NR	4.54	0.92	2.00	0.64
RD	4.39	0.44	2.00	0.59
PI	59.87	42.93	3.15	1.68

Note. EMG = electromyography; BC = baseline control; NR = nondistracted reading; RD = rhythmic distraction; PI = phonatory interference.

between conditions and were hence ineffective as a behavioral measure for musician participants in a task using EM music stimuli.

The results suggest that age, gender, and handedness had no influence on the participants' task performance (reflected by their overall success rate). However, we did find that age positively correlated with  $d'$  scores of Type I (i.e., the facility to differentiate between original themes and melodic lures) and that age negatively correlated with the percentage of hits of Type III (i.e., the aptitude to detect newly composed themes embedded in embellished variations). These relationships may be explained as resulting from pedagogical trends and training regimes of music theory and ear training classes, which typically focus on the development of finely honed skills serving to retrieve and store melodic-rhythmic fragments (or gestalts) that are considered to be the building blocks of 20th-century modern music. Hence, our results seem to point to a delicate advantage for older musicians in terms of experience and for younger musicians in terms of cognitive dexterity. These differences between older musicians and younger musicians also seem to map onto changes in crystallized intelligence versus fluid intelligence, which are seen to occur with aging.

Table 5

Experiment 1: Contrasts Between Sight Reading Conditions for Audio and EMG Output

Conditions	<i>F</i>	<i>MSE</i>	$\eta_p^2$	<i>p</i>
Audio output <sup>a</sup>				
NR vs. RD	1.19	0.1873	.06	.29
NR vs. PI	31.49	923.33	.64	<.0001
RD vs. PI	31.53	927.36	.64	<.0001
BC vs. NR	0.30	0.6294	.02	.59
BC vs. RD	0.01	0.243	.00	.94
BC vs. PI	31.85	917.58	.64	<.0001
EMG output <sup>b</sup>				
NR vs. RD	0.01	0.0233	.00	.92
NR vs. PI	14.13	0.9702	.41	<.01
RD vs. PI	15.67	0.8813	.44	<.001
BC vs. NR	4.71	0.1993	.19	<.05
BC vs. RD	5.64	0.1620	.22	<.05
BC vs. PI	15.67	201.3928	.44	<.001

Note. EMG = electromyography; NR = nondistracted reading; RD = rhythmic distraction; PI = phonatory interference; BC = baseline control. <sup>a</sup>  $df = 1, 18$ . <sup>b</sup>  $df = 1, 20$ .



However, we found little benefit for those musicians who possess absolute perfect pitch. That is, although they were significantly faster in their responses (i.e., shorter RTs), they were just as accurate as nonpossessors (i.e., there were no differences of PC, hits, FAs, or  $d'$ ), which suggests the likelihood of a speed–accuracy trade-off. This latter finding is especially interesting because most music researchers and music educators tend to believe that possession of such an ability is favorable for reading music, composing, doing melodic or harmonic dictation, and sight singing.

However, the main interest of Experiment 1 lies in the findings that we obtained from the physiological data. First, there were few differences between audio-output levels in all the tasks that were performed silently. That is, the measured output was roughly the same when participants sat silently, silently read a language text, and silently completed a mathematical sequence (BC) as it was when participants silently read music notation (NR, RD). Further, there were significant differences of audio-output level between all of these silent conditions compared with when participants sang a traditional folk song aloud (PI). Both of these outcomes were to be expected. However, it is extremely interesting to note that the associated EMG-output levels were of a very different character. That is, when monitoring the muscle activation of the vocal folds, we found that not only were there significant differences between subvocal activity occurring during silent reading and the vocal activity of singing aloud, but that significant differences also surfaced within the subvocal conditions themselves. Indeed, silent reading of language texts and silent mathematical reasoning have long been associated with “internal mutter” (Sokolov, 1972; Vygotsky, 1986). Further, there is also considerable evidence indicating that printed stimuli are not retained in working memory in their visual form but that they are instead recoded in a phonological format (Wilson, 2001). Therefore, observable subvocal activity during silent reading and reasoning tasks was to be expected as were output levels that were considerably lower than during overt vocal activity. Yet the current experiment clearly demonstrates that covert vocal fold activity is significantly more dynamic when the same participants silently read music notation than when they read printed text or work out mathematical sequences (the BC task). We feel that this finding is *prima facie* evidence corroborating our previous proposal that notational audiation is a process engaging kinesthetic-like covert excitation of the vocal folds linked to phonatory resources.

Nevertheless, the results of Experiment 1 are not conclusive in differentiating between the effects of RD and PI (as was seen previously in Brodsky et al., 2003). Although there is an indication of impairment in both the RD and PI conditions—and although on the basis of a decrement of PC and an increment of FAs, the PI condition seems to result in greater interference—differences between the two remain statistically nonsignificant. In an effort to interpret this picture, it may be advantageous to look at the RD condition as supplying a rhythmic distractor that disrupts temporal processing and at the PI condition as supplying a pitch distractor that disrupts spatial (tonal) processing. For example, Waters and Underwood (1999) viewed a comparable set of conditions in this manner and reported each one as disrupting particular codes or strategies necessary to generate imagery prompted by the visual surface cues provided by the music notation—each different but equally potent. Or perhaps another explanation could be found if

we were to use Baddeley’s (1986) proposal to distinguish between the “inner ear” (i.e., subvocal rehearsal) and the “inner voice” (i.e., phonological storage). This is especially appropriate as not all forms of auditory imagery rely on the same components. For example, using various interference conditions to highlight each of these components, Aleman and Wout (2004) demonstrated that articulatory suppression interferes with the inner voice while concurrent irrelevant auditory stimuli interfere with the inner ear. A third possibility, akin to that reported by Smith, Wilson, and Reisberg (1995), is that both components are essential and thus that no significant difference is to be expected at all. In fact, Wilson (2001) not only argues in support of the findings by Smith et al., but points out that by interfering with either component, one should not be able to reduce task performance to chance level, because if both components are essential, then when one element is blocked the other resource is still available to maintain information.

In light of these three explanations, a further exploration of differences between the RD and PI conditions seems warranted. Given the fact that our EMG data suggest that there is vocal fold activity (VFA) in all three music-reading conditions, it would seem appropriate to relabel the conditions accordingly: NR becomes VFA alone; RD becomes VFA plus manual motor activity (finger tapping) plus irrelevant temporal auditory stimuli (heard counter-rhythms); and PI becomes VFA plus phonatory muscle activity (singing aloud) plus irrelevant spatial–tonal auditory stimuli (heard melody). Bearing this reclassification in mind, it is interesting to look again at Aleman and Wout (2004), who found effects of auditory suppression (akin to our PI condition) on auditory–verbal visual tasks, whereas their tapping task (akin to our RD condition) affected the visual-alone tasks. They concluded that the two conditions do not by definition interfere with the same processing system and that tapping clearly interfered with visuospatial processing. Therefore, considering the results of Experiment 1, and taking into account Aleman and Wout’s results, we now ask whether covert rehearsal with the mind’s voice does in fact involve actual manual motor processing systems. That is, because the distraction from RD is as large as the interference from PI, then we might assume there to be an important reliance on kinesthetic phonatory and manual motor processing during subvocalization of music notation. Ecological evidence would support such a stance: Unlike text reading, the reading of music notation is seldom learned in isolation from learning to play an instrument (involving the corresponding manual motor sequences). However, hard empirical evidence to support the above hypothesis might be obtained if a nonauditory manual motor action could facilitate task performance in the RD condition—while not improving task performance in the PI conditions. This led to Experiment 2.

## Experiment 2

The purpose of this experiment was to shed further light on the two distraction conditions used in Experiment 1. We asked whether covert rehearsal with the mind’s voice does in fact involve actual motor processing systems beyond the larynx and hence a reliance on both articulatory and manual motor activity during the reading of music notation. To explore this issue, in Experiment 2 we added the element of finger movements emulating a music performance during the same music-reading conditions with the

same musicians who had participated in Experiment 1. We expected that there would be improvements in task performance resulting from the mock performance, but if such actions resulted in facilitated RD performances (while not improving performances in the NR or PI conditions), then the findings might provide ample empirical evidence to resolve the above query.

### Method

Approximately 8 months after Experiment 1 (hereafter referred to as T1), all 14 Israeli musician participants were contacted to participate in the current Experiment 2 (hereafter referred to as T2). In total, 4 were not available: 1 declared lack of interest, 2 had since moved to Europe for advanced training, and 1 was on army reserves duty. The 10 available participants (71% of the original Israeli sample) were retested at the Buchmann-Mehta School of Music in Tel Aviv, in the same room and daytime-hour conditions as in Experiment 1. In general, this subsample included slightly more men (60%) than women. Their mean age was 23 years ( $SD = 2.98$ , range = 19–30), and they had an average of 13 years ( $SD = 3.27$ , range = 5–16) of formal instrumental lessons beginning at an average age of 8 ( $SD = 3.49$ , range = 5–17) and an average of 6 years ( $SD = 1.83$ , range = 3–9) of formal ear-training lessons beginning at an average age of 13 ( $SD = 5.21$ , range = 6–25).

The stimuli, apparatus, design and test presentation, and procedure were the same as in Experiment 1, but with two adjustments. First, all music-reading conditions (NR, RD, and PI) were augmented with overt finger movements replicating a music performance of the presented EM notation—albeit without auditory feedback (i.e., the music instruments remained silent). One might, then, consider such activity to be a virtual performance. Given that the notation represents a single-line melody, only one hand was actively involved in the virtual performance, freeing the other to implement the tapping task required during the RD condition. For example, using only one hand, pianists pressed the keys of a MIDI keyboard without electric power, string players placed fingers on the fingerboard of their instrument muted by a cloth, and wind players pressed on the keypads of their wind instrument without its mouthpiece. The second difference between Experiments 1 and 2 was that in Experiment 2, the biomonitor was dropped from the procedure.

### Results

For each participant at each time point (T1: Experiment 1; T2: Experiment 2) in each music-reading condition (NR, RD, PI), the responses were analyzed for PC, hits, FAs,  $d'$ , and RTs (see Table 6).

Each dependent variable was entered separately into a 2 (time: T1, T2)  $\times$  3 (condition: NR, RD, PI) repeated measures ANOVA. No main effects of time for PC, hits,  $d'$ , or RTs surfaced. However, there were significant main effects of time for FAs,  $F(1, 9) = 5.61$ ,  $MSE = 475.31$ ,  $\eta_p^2 = .38$ ,  $p < .05$ . These effects indicated an overall decreased percentage of FAs at T2 ( $M = 26\%$ ,  $SD = 12.51$ ) as compared with T1 ( $M = 40\%$ ,  $SD = 8.21$ ). Further, there were significant main effects of condition for PC,  $F(2, 18) = 7.16$ ,  $MSE = 151.88$ ,  $\eta_p^2 = .44$ ,  $p < .01$ ; hits,  $F(2, 18) = 6.76$ ,  $MSE = 193.93$ ,  $\eta_p^2 = .43$ ,  $p < .01$ ; and  $d'$ ,  $F(2, 18) = 7.98$ ,  $MSE = 2.45$ ,  $\eta_p^2 = .47$ ,  $p < .01$ ; there were near significant effects for FAs,  $F(2,$

Table 6  
*Experiment 2: Descriptive Statistics of Behavioral Measures by Sight Reading Condition and Time (Experiment Session)*

Dependent variable	NR		RD		PI	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
PCs (%)						
T1	82.5	6.15	66.7	18.84	64.2	16.22
T2	81.7	8.61	86.7	15.32	70.8	20.13
Hits (%)						
T1	95.0	8.05	76.7	30.63	73.2	23.83
T2	90.0	8.61	86.7	17.21	80.0	20.49
FAs (%)						
T1	30.0	13.15	43.3	21.08	45.0	29.45
T2	26.7	16.10	13.3	17.21	38.3	28.38
$d'$						
T1	3.45	1.22	2.09	2.07	1.04	1.47
T2	2.96	1.56	4.60	2.87	2.09	2.96
RTs (s)						
T1	11.88	5.87	13.86	8.43	13.14	5.73
T2	11.38	6.78	11.92	6.56	12.66	6.82

Note. NR = nondistracted reading; RD = rhythmic distraction; PI = phonatory interference; PC = percentage correct; T1 = Experiment 1; T2 = Experiment 2; FA = false alarm; RT = response time.

18) = 3.14,  $MSE = 377.57$ ,  $\eta_p^2 = .26$ ,  $p = .07$ ; and RTs,  $F(2, 18) = 2.64$ ,  $MSE = 9,699,899$ ,  $\eta_p^2 = .23$ ,  $p = .10$ . In general, comparison analysis between the conditions demonstrated that effects were due to significant differences between nondistracted music reading and reading under distraction or interference conditions (see Table 7, Contrasts of condition). Further, the ANOVA found significant interactions of Time  $\times$  Condition for FAs,  $F(2, 18) = 3.93$ ,  $MSE = 268.52$ ,  $\eta_p^2 = .30$ ,  $p < .05$ ; and  $d'$ ,  $F(2, 18) = 4.50$ ,  $MSE = 2.5121$ ,  $\eta_p^2 = .33$ ,  $p < .05$ ; as well as a near significant interaction for PC,  $F(2, 18) = 3.23$ ,  $MSE = 172.20$ ,  $\eta_p^2 = .26$ ,  $p = .06$ . The interaction was nonsignificant for hits or RTs. As can be seen in Table 7 (Time  $\times$  Condition interaction), comparison analyses found that interaction effects were solely due to improvements in the RD condition for T2; this would seem to indicate the efficiency of overt motor finger movements in overcoming the effects of RD.

Last, for each participant at each time point (T1, T2) in each music stimulus type (I, II, III) the frequency of correct responses was analyzed for PC, hits, FAs,  $d'$ , and RTs (see Table 8).

Each variable was entered separately into a 2 (time: T1, T2)  $\times$  3 (stimulus type: I, II, III) repeated measures ANOVA. There were main effects only for time for FAs,  $F(1, 9) = 5.61$ ,  $MSE = 475.31$ ,  $\eta_p^2 = .38$ ,  $p < .05$ ; no main effects for PC, hits,  $d'$ , or RTs surfaced. The effects indicate an overall decreased percentage of FAs at T2 ( $M = 26\%$ ,  $SD = 3.52$ ) compared with T1 ( $M = 39\%$ ,  $SD = 5.87$ ). Further, no main effects of stimulus type were found. Finally, the ANOVA found no significant interaction effects.

### Discussion

A comparison of the results of Experiments 1 and 2 shows that previous exposure to the task did not significantly aid task performance—except that participants were slightly less likely to mistake a lure for a target in Experiment 2 relative to that in Experiment 1. Even the level of phonatory interference remained

Table 7  
*Experiment 2: Contrasts Between Behavioral Measures for Sight Reading Conditions and Interaction Effects Between Sight Reading Condition and Time (Experiment Session)*

Dependent variable	Conditions	<i>F</i> (1, 9)	<i>MSE</i>	$\eta_p^2$	<i>p</i>
Contrasts of condition					
PC	NR ( <i>M</i> = 82%, <i>SD</i> = 0.56) vs. PI ( <i>M</i> = 68%, <i>SD</i> = 4.66)	21.0	101.27	.70	<.01
	RD ( <i>M</i> = 77%, <i>SD</i> = 14.14) vs. PI ( <i>M</i> = 68%, <i>SD</i> = 4.66)	4.46	188.27	.33	.06
Hits	NR ( <i>M</i> = 93%, <i>SD</i> = 3.54) vs. RD ( <i>M</i> = 82%, <i>SD</i> = 7.07)	5.41	216.82	.38	<.05
	NR ( <i>M</i> = 93%, <i>SD</i> = 3.54) vs. PI ( <i>M</i> = 77%, <i>SD</i> = 4.81)	17.19	145.83	.66	<.01
<i>d'</i>	NR ( <i>M</i> = 3.20, <i>SD</i> = 0.35) vs. PI ( <i>M</i> = 1.57, <i>SD</i> = 0.74)	15.71	1.7124	.64	<.01
	RD ( <i>M</i> = 3.08, <i>SD</i> = 1.39) vs. PI ( <i>M</i> = 1.57, <i>SD</i> = 0.74)	8.03	3.6595	.47	<.05
FAs	NR ( <i>M</i> = 28%, <i>SD</i> = 2.33) vs. PI ( <i>M</i> = 42%, <i>SD</i> = 4.74)	23.81	466.05	.73	<.001
RTs	NR ( <i>M</i> = 11.60 s, <i>SD</i> = 0.35) vs. PI ( <i>M</i> = 12.90 s, <i>SD</i> = 0.34)	13.16	1,216,939	.59	<.01
	NR ( <i>M</i> = 11.63 s, <i>SD</i> = 0.35) vs. RD ( <i>M</i> = 13.89 s, <i>SD</i> = 0.04)	3.52	14,454,741	.28	.09
Time × Condition interaction					
FAs	RD	10.57	425.93	.54	<.05
<i>d'</i>	RD	5.74	5.4901	.39	<.05
PC	RD	7.16	2.79	.44	<.05

Note. PC = percentage correct; NR = nondistracted reading; PI = phonatory interference; RD = rhythmic distraction; FA = false alarm; RT = response time.

significantly unchanged. Nevertheless, there were significant interactions vis-à-vis improvement of task performance for the RD condition in the second session. That is, even though all music-reading conditions were equally augmented with overt motor finger movements replicating a music performance, an overall improvement of task performance was seen only for the RD condition. Performance in the RD condition was actually better than in the NR condition (including PC, FAs, and *d'*). Furthermore, when looking at Table 6, it can be seen that at T1 the RD condition emulated a distraction condition (similar to PI), whereas

at T2 with the finger movements the RD condition emulated a nondistracted music-reading condition (similar to NR).

It is interesting to note that Aleman and Wout (2004) questioned the extent to which actual motor processing systems are active during covert rehearsal with the mind's voice. In fact, they proposed that if concurrent tapping interference conditions did not interfere to the same extent as concurrent articulatory suppression conditions, then that would be a strong indication of language processing without sensory-motor processing. However, they also proposed that, in contrast, interference from finger tapping that was as great as the interference from articulatory suppression would indicate reliance on kinesthetic phonatory and motor processing during subvocalization. We view the results of Experiment 2 as empirical support for the latter proposition. Not only in the first instance (T1) were RD and PI conditions equal in their distraction–interference effects, but through the provision of motor enhancement (T2), participants were able to combat distraction effects and attain performance levels as high or even higher than in nondistracted conditions. This demonstration confirms that the mental representation of music notation also cues manual motor imagery. This stance is similar to other conceptions, such as that regarding differences between speech and print stimuli. For example, Gathercole and Martin (1996) proposed that motoric or articulatory processes may not be the key components in verbal working memory per se, but that the representations involved in rehearsal are representations of vocal gestures—intended for speech perception but not speech production. Accordingly, such representations are long-term memory representations, and working memory is believed to consist of their temporary activation. Hence, when considering the mental representation of music notation, perhaps more than anything else a reliance on manual motor imagery is inevitable because of the closely knit cognitive relationship between reading music and the associated manual ges-

Table 8  
*Experiment 2: Descriptive Statistics of Behavioral Measures by Stimuli Type and Time (Experiment Session)*

Dependent variable	Type I		Type II		Type III	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
PCs (%)						
T1	74.2	10.72	69.2	17.15	70.0	10.54
T2	80.8	14.95	79.2	18.94	79.2	11.28
Hits (%)						
T1	81.7	19.95	83.3	15.71	80.0	21.94
T2	85.0	14.59	83.3	17.57	83.3	11.25
FAs (%)						
T1	33.3	13.61	45.0	27.30	40.0	17.90
T2	23.2	23.83	25.0	23.90	30.0	20.49
<i>d'</i>						
T1	2.28	1.62	1.84	1.80	1.79	1.29
T2	2.72	2.03	3.02	2.54	2.78	1.90
RTs (s)						
T1	11.90	6.26	14.06	5.98	13.01	6.50
T2	13.08	6.35	14.11	7.25	14.53	7.23

Note. PC = percentage correct; T1 = Experiment 1; T2 = Experiment 2; FA = false alarm; RT = response time.

tures imprinted in the minds of music readers by having a music instrument in hand throughout a lifetime of music development.

Yet, thus far all our efforts, as well as those of others reported in the literature, have been directed at classically trained musician performers on tonal instruments. Specifically, our participants were pianists or players of orchestral instruments, all of which produce tones having a definitive pitch (i.e., measurable frequency) and are associated with specific letter names (C, D, E, etc.), solfège syllables (do, re, mi, etc.), and notes (i.e., the specific graphic placement of a symbol on a music staff). One might ask, then, whether the mental representation of music notation (which seems to involve the engagement of kinesthetic-like covert excitation of the vocal folds and cued manual motor imagery) is biased by higher order tonological resources. In other words, we ask to what extent the effects found in Experiments 1 and 2 are exclusive to musicians who rehearse and rely on music notation in a tonal vernacular, or rather whether the findings reported above reflect broader perceptual cognitive mechanisms that are recruited when reading music notation—regardless of instrument or notational system. This question led to Experiment 3.

### Experiment 3

In the final experiment, we focused on professional drummers who read drum-kit notation. The standard graphic representation of music (i.e., music notation) that has been in place for over 400 years is known as the orthochronic system (OS; Sloboda, 1981). OS is generic enough to accommodate a wide range of music instruments of assorted pitch ranges and performance methods. On the one hand, OS implements a universal set of symbols to indicate performance commands (such as loudness and phrasing); on the other hand, OS allows for an alternative set of instrument-specific symbols necessary for performance (such as fingerings, pedaling, blowing, and plucking). Nevertheless, there is one distinctive and peculiar variation of OS used regularly among music performers—that is, music notation for the modern drum kit. The drum kit is made up of a set of 4–7 drums and 4–6 cymbals, variegated by size (diameter, depth, and weight) to produce a range of sonorities

and relative pitches. The kit is performed by one player in a seated position, using both upper and lower limbs; hands play with drumsticks (also beaters, mallets, and wire brushes), while the feet employ pedals. The drum kit is part of most ensembles performing popular music styles, including country and western, blues and jazz, pop and rock, ballads, polkas and marches, and Broadway theatre shows. Drum-kit notation uses a music staff, employs similar rhythmic relations between notes and groups of notes, and shares many conventions of OS, such as meter values and dynamic markings. However, there are two major differences that distinguish drum-kit notation from OS: (a) drum-kit notation uses various note heads to indicate performance sonority; and (b) the five-horizontal line grid is not indicative of pitch values (inasmuch as to reference placement of fixed-pitch notes such as C, D, and E) but rather designates location of attack (the explicit drum or cymbal to be played), performance timbre (a head or rim shot, open or closed hi-hat), and body-limb part involvement (right or left hand or left or right foot). All of these above are positioned on the grid vertically, from the bottom-most space representing the relatively lowest pitched drum and lower limbs, to the topmost space representing the relatively highest pitched cymbal and higher limbs (see Figure 2).

In Experiment 3, we used a sample of professional drummers reading drum-kit notation in order to further examine the mental representation of music notation. Although we expected to find that drummers rely on motor processes including imagery during silent reading, on the basis of the findings of Experiments 1 and 2 we also expected to find some evidence of kinesthetic phonatory involvement—despite the fact that the drum kit does not engage higher order tonological resources.

### Method

*Participants.* The drummers participating in the study were recruited and screened by the drum specialists at the Klei Zemer Yamaha Music store in Tel Aviv, Israel. Initially, 25 drummers were referred and tested, but a full data set was obtained from only 17. Of these, 13 (77%) passed the prerequisite threshold inclusion

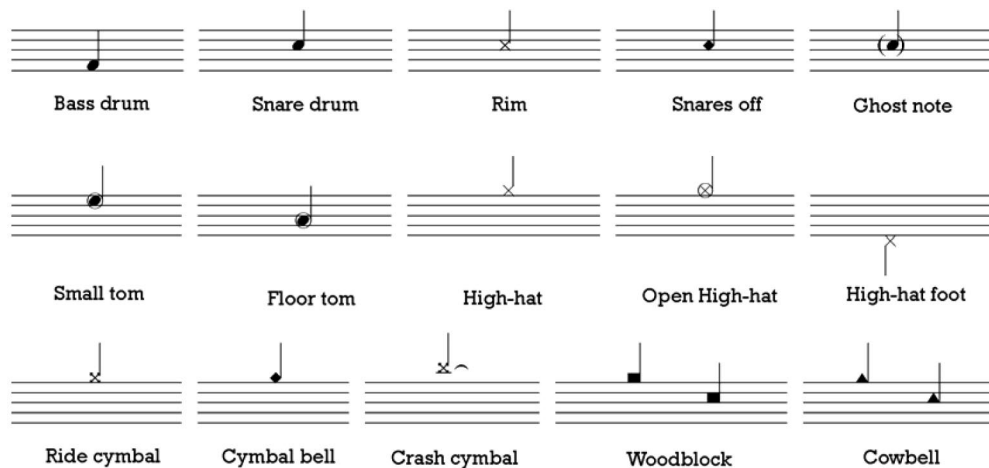


Figure 2. Drum-kit notation key. From *Fifty Ways to Love Your Drumming* (p. 7), by R. Holan, 2000, Kfar Saba, Israel: Or-Tav Publications. Copyright 2000 by Or-Tav Publications. Used with kind permission of Rony Holan and Or-Tav Music Publications.



criterion; 1 participant was dropped from the final data set because of exceptionally high scores on all performance tasks, apparently resulting from his self-reported overfamiliarity with the stimuli because of their commercial availability. The final sample ( $N = 12$ ) was composed of male drummers with high school diplomas; 4 had completed a formal artist's certificate. The mean age of the sample was 32 years ( $SD = 6.39$ , range = 23–42), and participants had an average of 19 years ( $SD = 6.66$ , range = 9–29) of experience playing the drum kit, of which an average of 7 years ( $SD = 3.87$ , range = 1–14) were within the framework of formal lessons, from the average age of 13 ( $SD = 2.54$ , range = 7–16). Further, more than half (67%) had participated in formal music theory or ear training instruction (at private studios, music high schools, and music colleges), but these programs were extremely short-term ( $M = 1.3$  years,  $SD = 1.48$ , range = 1–5). Only 1 drummer claimed to possess absolute perfect pitch. The majority were right-handed (83%) and right-footed (92%) dominant drummers of rock (58%) and world (17%) music genres. Using a 4-point Likert scale (1 = *not at all*, 4 = *highly proficient*), the drummers reported a medium to high level of confidence in their abilities to read new unseen drum-kit notation ( $M = 3.00$ ,  $SD = 0.60$ ), to “hear” the printed page ( $M = 3.17$ ,  $SD = 0.72$ ), to remember their part after a one-time exposure ( $M = 3.42$ ,  $SD = 0.52$ ), and to analyze the music while reading or listening to it ( $M = 3.67$ ,  $SD = 0.49$ ). Finally, the reported learning strategy employed when approaching a new piece was divided between listening to audio recordings (58%) and silently reading through the notated drum part (42%); only 1 drummer reported playing through the piece.

*Stimuli.* Forty-eight drum-kit rhythms (often referred to as “groove tracks”) were taken from Holan (2000; see Figure 3). The stimuli reflect generic patterns associated with particular stylized dance rhythms; they do not represent the rhythmic frames of an individual melody or of well-known music pieces. Hence, although most drummers have an elementary familiarity with such beats, the exact exemplars employed here were not known to them. Further, it should be pointed out that these stimuli were exclusively rhythmic in character; they were not formatted into embellished phrases (EMs) as were the melodic themes used in Experiments 1 and 2. The 48 grooves were rock, funk, and country beats ( $n = 10$ ); Latin beats ( $n = 8$ ); Brazilian beats ( $n = 7$ ); jazz beats ( $n = 7$ ); Middle Eastern beats ( $n = 4$ ); and dance beats ( $n = 12$ ). Each target groove in the form of a notated score (see Figure 3A) and corresponding audio track was matched to another rhythm from the pool as a lure groove (see Figure 3B). The criteria used in choosing the lures were thematic or visual similarities in contour, texture, rhythmic pattern, phrasing, meter, and music style. The scores were either four or eight bars long (each repeated twice for a total of 8–16 measures in length), and the audio tracks were roughly 20 s ( $SD = 5.41$  s, range = 13–32 s) long. The 48 items were randomly assigned to one of four blocks (one block per experimental condition), each controlled for an equal number of targets and lures. All audio tracks (ripped from the accompanying CD formatted as 16-bit wave files) were cropped with the Soundforge XP4.5 (RealNetworks) audio-editing package and were standardized for volume (i.e., smoothed or enhanced where necessary). The graphic notation, as supplied by the publisher, was formatted as 24-bit picture files.



Figure 3. Groove tracks. A: Rock beat. B: Funk rock beat. From *Fifty Ways to Love Your Drumming* (pp. 8, 15), by R. Holan, 2000, Kfar Saba, Israel: Or-Tav Publications. Copyright 2000 by Or-Tav Publications. Used with kind permission of Rony Holan and Or-Tav Music Publications.

The pretest BC task, apparatus (including collection of audio and EMG output), design and test presentation, and procedure were the same as in Experiment 1, but with three adjustments. First, a fourth condition was added to the previous three-condition (NR-RD-PI) format: Virtual drumming (VD) consisted of a virtual music performance on an imaginary drum kit and involved overt motions of both arms and legs but without sound production. Annett (1995) referred to this variety of imaginary action as “voluntary manipulation of an imaginary object” (p. 1395). Second, taking into account a lower level of mental effort required to decode the groove figures compared with the EMs used in Experiments 1 and 2, we halved the allotted music notation reading time per item from 60 s to 30 s. Third, although the location of testing was somewhat isolated, it was still within the confines of a commercial setting (i.e., a music store). Therefore, AKG K271 Studio (AKG Acoustics) circumaural closed-back professional headphones were used. This fixed-field exposure format resulted in the subsequent use of a Rhythm Watch RW105 (TAMA) to supply the concurrent rhythmic distraction stimuli for the RD condition and a Samson Q2 (Samson Audio) neodymium hypercardioid vocal microphone to supply the phonatory interference stimuli for the PI condition. The TAMA RW105 is essentially a digital metronome with additional features such as a “tap-in” capability; this function served as the means for the experimenter to input (via percussive tapping on a small rubber pad) the irrelevant cross-rhythms required by the RD condition. During the PI condition, the participant used the Q2 vocal microphone to sing the required interfering folk song. Both of these devices were channeled to the headphones through an S · AMP (Samson Audio) five-channel mini headphone amplifier linked to an S · MIX (Samson Audio) five-channel mini-mixer; the overall output volume was adjusted per participant.

## Results

For each participant in each music-reading condition (NR, VD, RD, PI), the responses were analyzed for PC, hits, FAs,  $d'$ , and RTs. As can be seen in Table 9, across the conditions there was a general decreasing level of PC, a decrease in percentage of hits with an increase in percentage of FAs and hence a decrease in level of  $d'$ , as well as an increasing level of RTs. We note that the scores indicate slightly better performances for the drummers in the RD condition (i.e., higher hits and hence higher  $d'$  scores); although the reasons for such effects remain to be seen, we might assume

that the tapping task may have facilitated in some visuospatial processing, as explained earlier (see Aleman & Wout, 2004). Taken together, these results seem to indicate an escalating level of impairment in the sequenced order NR-VD-RD-PI.

Each outcome variable with music reading condition as a within-subjects variable was entered into a repeated measures ANOVA. There was a significant effect of condition for RTs,  $F(3, 33) = 2.89$ ,  $MSE = 3.4982$ ,  $\eta_p^2 = .21$ ,  $p = .05$ ; no effects for PC, hits, FAs, or  $d'$  surfaced. As can be seen in Table 9, a comparison analysis for RTs showed that effects were due to significant differences between music reading under PI versus nondistracted music reading as well as versus music reading with simultaneous VD. In general, these findings indicate a more serious level of PI effects, as seen in longer RTs, in comparison to the other conditions (NR, VD, RD).

Finally, audio and EMG output of correct responses for each participant in each condition was calculated for all segments involving music reading; measurements taken during the pretest control task (BC) representing baseline-level data were also calculated (see Table 10).

The data for each participant with condition (baseline and the four music-reading conditions) as a within-subjects variable were entered into a repeated measures ANOVA. There were significant effects of condition for audio output,  $F(4, 44) = 17.26$ ,  $MSE = 215.54$ ,  $\eta_p^2 = .61$ ,  $p < .0001$ ; as well as for EMG output,  $F(4, 44) = 4.78$ ,  $MSE = 1.0898$ ,  $\eta_p^2 = .30$ ,  $p < .01$ . As can be seen in Table 11, a comparison analysis for audio data found no significant differences between the silent-reading conditions, but did show significant differences when comparing all three silent-reading conditions with the vocal condition. Further, the analysis of audio data found no significant differences between the BC tasks and either silent-reading condition, but did show significant differences between the control tasks and the vocal condition. In addition, comparison analysis of EMG data showed no significant differences between the silent-reading conditions, but did reveal significant differences when comparing the two silent-reading conditions with the vocal condition. It should be pointed out that finding no significant difference between music-reading with VD (considered a silent-reading condition) and music reading under PI is especially important as this result indicates the subvocal nature of virtual performance (i.e., mental rehearsal). The analysis of EMG data also showed significant differences between the BC tasks and all four music-reading conditions. This latter finding is indicative

Table 9

*Experiment 3: Descriptive Statistics of Behavioral Measures by Sight Reading Condition and Contrasts Between Sight Reading Conditions for Behavioral Measures*

Condition	% PC		% hits		% FAs		$d'$		Median RTs (s)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
NR	88.2	9.70	84.7	13.22	8.3	8.07	4.09	2.45	7.22	2.84
VD	86.1	10.26	80.6	11.96	8.3	15.08	3.97	1.80	7.02	3.06
RD	86.1	18.23	86.1	19.89	13.9	19.89	4.55	3.18	7.61	3.41
PI	81.3	11.31	79.2	21.47	16.7	15.89	3.50	2.02	9.05	4.22

*Note.* For RTs, PI vs. NR,  $F(1, 11) = 5.90$ ,  $MSE = 3.4145$ ,  $\eta_p^2 = .35$ ,  $p < .05$ ; PI vs. VD,  $F(1, 11) = 5.33$ ,  $MSE = 4.648$ ,  $\eta_p^2 = .33$ ,  $p < .05$ . PC = percentage correct; FA = false alarm; RT = response time; NR = nondistracted reading; VD = virtual drumming; RD = rhythmic distraction; PI = phonatory interference.

Table 10  
Experiment 3: Descriptive Statistics of Audio and EMG Output by Sight Reading Condition

Condition	Audio ( $\mu\text{V}$ )		EMG ( $\mu\text{V}$ )	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
BC	9.23	11.31	1.76	0.37
NR	4.28	0.24	2.58	1.31
VD	6.10	5.44	2.85	1.71
RD	8.50	8.76	2.60	0.96
PI	46.16	28.92	3.60	1.28

Note. EMG = electromyography; BC = baseline control; NR = nondistracted reading; VD = virtual drumming; RD = rhythmic distraction; PI = phonatory interference.

of the fact that phonatory involvement reaches levels of involvement during the reading of music notation for the drum kit that are significantly higher than those demonstrated during silent language text reading or mathematical computation.

Discussion

Drum-kit musicians are often the butt of other players’ mockery, even within the popular music genre. Clearly, one reason for such harassment is that drummers undergo a unique regime of training, which generally involves years of practice but fewer average years of formal lessons than other musicians undertake. In addition, drummers most often learn to play the drum kit in private studios and institutes that are not accredited to offer academic degrees and that implement programs of study targeting the development of intricate motor skills, while subjects related to general music theory, structural harmony, and ear-training procedures are for the most part absent. Furthermore, drum-kit performance requires a distinct set of commands providing instructions for operating four limbs, and these use a unique array of symbols and a notation system that is unfamiliar to players of fixed-pitch tonal instruments. Subsequently, the reality of the situation is that drummers do in fact read (and perhaps speak) a vernacular that is foreign to most other players, and hence, unfortunately, they are often regarded as deficient musicians. Only a few influential drummers have been recognized for their deep understanding of music theory or for their performance skills on a second tonal instrument. Accordingly, Spagnardi (2003) mentions Jack DeJohnette and Philly Jo Jones (jazz piano), Elvin Jones (jazz guitar), Joe Morello (classical violin), Max Roach (theory and harmony), Louie Bellson and Tony Williams (composing and arranging), and Phil Collins (songwriting). Therefore, we feel that a highly pertinent finding of Experiment 3 is that 77% of the professional drummers (i.e., the proportion of those meeting the inclusion criterion) not only proved to be proficient drum-kit music readers, but demonstrated highly developed cognitive processing skills, including the ability to generate music imagery from the printed page.

The results of Experiment 3 verified that the drummers were proficient in task performance during all music-reading conditions but also that they were slightly less capable of rejecting lures in both RD–PI conditions than in the NR–VD conditions and were worse at matching targets in the PI condition. Moreover, the PI condition seriously interfered with music imagery, as shown by

statistically significantly increased RTs. This is in line with our previously reported findings (Brodsky et al., 2003). Further, this study found that the VD condition—reading notation with concurrent overt motions as though performing on a drum kit—did in fact hamper drummers to much the same extent as the RD condition, although not in quite the same way—qualitatively speaking. That is, whereas the RD condition facilitated participants’ ability to choose correct targets (hits) but hindered their ability to correctly reject lures (FAs), by contrast the VD condition facilitated participants’ ability to correctly reject lures (FAs) but hindered their ability to choose correct targets (hits). Such differences, akin to the results of Experiment 2, support the notion that overt performance movements compensate for the cognitive disruption supplied by concurrent tapping with irrelevant rhythmic distraction. Therefore, it would appear that the combination of RD plus virtual music performance (as implemented in Experiment 2) offsets the effects of RD, allowing for results that are similar to those obtained in the nondistracted normal music-reading condition.

The results of Experiment 3 include interesting physiological findings that are similar to the findings of Experiment 1. That is, the audio and EMG output levels of all three silent-reading conditions (NR, VD, and RD) were not statistically significantly different from each other; yet, as expected, all of them were statistically significantly lower than in the vocal condition (PI). Further, in all three silent music-reading conditions, VFA levels were higher than those seen in the combined BC control tasks (i.e., sitting quietly, language text reading, and mathematical computation).

However, we view the most important finding of Experiment 3 to be its demonstration of reliance on phonatory resources among drummers. Most musicians, including drummers themselves, would tend to view the drum kit as an instrument essentially based on manual and podalic motor skills. Yet anecdotal evidence points to the fact that all drummers learn their instrument primarily via the repetition of vocal patterns and verbal cues representing the

Table 11  
Experiment 3: Contrasts Between Sight Reading Conditions for Audio and EMG Output

Conditions	<i>F</i> (1, 11)	<i>MSE</i>	$\eta_p^2$	<i>p</i>
Audio output				
NR vs. PI	25.31	415.81	.70	<.001
VD vs. PI	20.89	460.81	.70	<.001
RD vs. PI	16.51	515.33	.60	<.01
BC vs. PI	16.29	502.29	.60	<.01
EMG output				
NR vs. PI	5.33	1.1601	.33	<.05
RD vs. PI	5.39	1.1039	.33	<.05
VD vs. PI	1.27	2.6047	.10	.28
BC vs. NR	5.20	0.7820	.32	<.05
BC vs. VD	4.57	1.1731	.29	.06
BC vs. RD	10.22	0.4175	.48	<.01
BC vs. PI	38.11	0.5325	.78	<.001

Note. NR = nondistracted reading; PI = phonatory interference; VD = virtual drumming; RD = rhythmic distraction; BC = baseline control; EMG = electromyography.

anticipated sound bites of motor performance. Moreover, all drummers are continuously exposed to basic units of notation via rhythmic patterns presented phonetically; even the most complex figures are analyzed through articulatory channels. Hence, it is quite probable that the aural reliance on kinesthetic phonatory and manual–podalic motor processing during subvocalization is developed among drummers to an even higher extent than in other instrumentalists. Therefore, we view the current results as providing empirical evidence for the impression that drummers internalize their performance as a phonatory–motor image and that such a representation is easily cued when drummers view the relevant graphic drum-kit notation. Nonetheless, within the framework of the study, Experiment 3 can be seen as confirmation of the cognitive mechanisms that appear to be recruited when reading music notation—regardless of instrument or notational system.

### General Discussion

The current study explored the notion that reading music notation could activate or generate a corresponding mental image. The idea that such a skill exists has been around for over 200 years, but as yet, no valid empirical demonstration of this expertise has been reported, nor have previous efforts been able to target the cognitive processes involved. The current study refined the EM task in conjunction with a distraction paradigm, as a method of assessing and demonstrating notational audiation. The original study (Brodsky et al., 2003) was replicated with two samples of highly trained classical-music players, and then a group of professional jazz-rock drummers confirmed the related conceptual underpinnings. In general, the study found no cultural biases of the music stimuli employed nor of the experimental task itself. That is, no difference of task performance was found between the samples recruited in Israel (using Hebrew directions and folk song) and the samples recruited in Britain (using English directions and folk song). Further, the study found no demonstrable significant advantages for musicians of a particular gender or age range, nor were superior performances seen for participants who (by self-report) possessed absolute perfect pitch. Finally, music literacy was ruled out as a contributing factor in the generation of music imagery from notation as there were no effects or interactions for stimulus type (i.e., well-known, newly composed, or hybrid stimuli). That is, the findings show that a proficient music reader is just as proficient even when reading newly composed, previously unseen notation. Thus far, the only descriptive predictor of notational audition skill that surfaced was the self-reported initial strategy used by musicians when learning new music: 54% of those musicians who could demonstrate notational audiation skill reported that they first silently read through the piece before playing it (whereas 35% play through the piece, and 11% listen to a recording).

Drost, Rieger, Brass, Gunter, and Prinz (2005) claimed that for musicians, notes are usually directly associated with playing an instrument. Accordingly, “music-reading already involves sensory-motor translation processes of notes into adequate responses” (p. 1382). They identified this phenomenon as *music-learning coupling*, which takes place in two stages: First, associations between action codes and effect codes are established, and then, simply by imagining a desired effect, the associated action takes place. This *ideomotor* view of music skills suggests that highly trained expert musicians must only imagine or anticipate a

music sequence, and an associated sequence of related actions will subsequently be automatically activated—there is no more need of direct conscious control of movements. Although such a conception might explain the automaticity of music performance, one might inquire whether the same is true about music reading. Drost et al. further surmised that the ability to generate music imagery from graphic notation is no more than associative learning coupling in which mental representations are activated involuntarily. Yet they offered no further insights into the nature of the representation.

It is important to ask how the human brain represents music information. For example, if mental representations for music are defined as “hypothetical entities that guide our processing of music” (Schröger, 2005, p. 98), then it would be clear that how we perceive, understand, and appreciate music is determined not by the nature of the input but by what we do with it. For example, Halpern and Zatorre (1999) concluded that the SMA is activated in the generation of auditory imagery because of its contribution to the organization of the motor codes; this would imply a close relationship between auditory and motor memory systems. Halpern and Zatorre’s study was based on earlier research by Smith et al. (1995), who distinguished between the *inner ear* and the *inner voice* on the basis of evidence that the phonological loop is subdivided into two constituents. Accordingly, the activation of the SMA during music imagery may actually imply a “singing to oneself” strategy during auditory imagery tasks, reflecting motor planning associated with subvocal singing or humming during the generation process.

Whereas the roles of inner ear and inner voice seem to be difficult to disentangle behaviorally, functional neuroimaging studies can potentially shed more light on the issue. Both Smith et al. (1995) and Aleman and Wout (2004) proposed that whereas the inner ear would be mediated by temporal areas such as superior temporal gyri, the inner voice would be mediated by structures involving articulation, including the SMA and left inferior frontal cortex, Broca’s area, the superior parietal lobe, and the superior temporal sulcus—all in the left hemisphere. Yet, in later studies, Aleman and Wout (2004) found contradicting evidence showing that articulatory suppression interfered with processes mediated by Broca’s area, whereas finger tapping interfered with processes mediated by the SMA and superior parietal lobe—both considered areas involved in phonological decoding (thought to be the seat of the inner ear). Finally, in an investigation of mental rehearsal (which is a highly refined form of music imagery representing covert music performance), Langheim et al. (2002) found involvement of cortical pathways recruited for the integration of auditory information with the temporal- and pitch-related aspects of the music rehearsal itself. They suggested that this network functions independently of primary sensorimotor and auditory cortices, as a coordinating agent for the complex spatial and timing components of music performance. Most specifically, Langheim et al. found involvement of several pathways: the right superior parietal lobule in the spatial aspects of motor and music pitch representation, the bilateral lateral cerebellum in music and motor timing, and the right inferior gyrus in integrating the motor and music-auditory maps necessary (perhaps via premotor and supplementary motor planning areas) for playing an instrument.

We feel that the mental representation of music information cannot be understood by mapping the neurocognitive architecture



of music knowledge in isolation, apart from empirically valid behavioral measures. That is, we believe that the juxtaposition of neuropsychological approaches and brain imaging tools with behavioral measures from experimental psychology and psychoacoustics is the only way forward in investigating how music is represented in the human brain and mind. Therefore, we consider the current exploration of notational audiation an appropriate step toward understanding the musical mind and, based on the results reported above, suggest that the methodological archetype with which to assess such covert processes is the EM task.

In a recent study, Zatorre and Halpern (2005) raised the question as to whether there is evidence that auditory and motor imagery are integrated in the brain. We feel that the current findings provide a preliminary answer. We designed our study to uncover the kinesthetic-like phonatory-linked processes used during notational audiation. We exploited Smith et al.'s (1995) assumption that imagery tasks requiring participants to make judgments about auditory stimuli currently not present must employ an inner-ear/inner-voice partnership as a platform for the necessary processes and judgments to take place. Accordingly, participants would use a strategy whereby they produced a subvocal repetition (inner voice) and listen to themselves (inner ear) in order to interpret and/or judge the auditory or phonological stream. We thus developed the EM task. Then we considered Smith et al.'s second assumption, that when an empirical task requires analytic judgments or the comparison of novel melodic fragments, a reliance on the phonological loop is predicted, and consequently, performance deficits under articulatory suppression can be expected. We therefore used a distraction paradigm. Nonetheless, after refining the EMs, the results of Experiment 1 could not demonstrate that imagery generated from the reading of music notation was exclusively phonatory in nature. In fact, the behavioral results showed that effects of the PI condition were not significantly different from effects of the RD condition.

However, the results of Experiment 2 showed that movement representations of music performance could facilitate task performance and even overcompensate for the specific interference-inducing errors during RD. From this we might infer two assumptions: (a) There is a profound reliance on kinesthetic phonatory and manual motor processing during subvocalization (which is the seat of music imagery generated by the inner voice when one reads music notation); and (b) the mental representation of music notation entails a dual-route stratagem (i.e., the generation of aural-oral subvocalization perceived as the internal kinesthetic image of the inner voice and aural-motor impressions perceived as the internal kinesthetic image of music performance). It is of interest to note that we are not the first to raise such possibilities. For example, in her extensive review of working memory, Wilson (2001) highlighted the fact that many central cognitive abilities seem to depend on perceptual and motor processes. Accordingly, off-line embodied cognition involves sensorimotor processes that run covertly to assist with the representation and manipulation of information in the absence of task-relevant input. However, in reference to music notation, future research is needed to probe further into these processes. Nevertheless, we partially confirmed these assumptions in Experiment 3 by examining the music reading of drum-kit notation—the drum kit being an instrument assumed more often than not to require exclusive motor action as fixed-pitch tonal features are not present. The results of Experi-

ment 3 show an equal reliance on both phonatory and motor resources among drummers. It is thus our opinion that the results provide evidence that clearly indicates that auditory and motor imagery are integrated in the brain. Notational audiation skill, then, is the engagement of kinesthetic-like covert excitation of the vocal folds with concurrently cued motor imagery.

Finally, we would like to consider the idea that silent reading of music notation is essentially an issue highlighting cross-modal encoding of a fundamentally unisensory input. Clearly, most everyday people with naive music experience, as well as all those with formal music training, see the spatial layout on the staff of the auditory array. However, our study plainly shows that only a third of all highly trained expert musicians are proficient enough to hear the temporal, tonal, and harmonic structure of the portrayed visual changes. Guttman, Gilroy, and Blake (2005) claimed that “obligatory cross-modal encoding may be one type of sensory interaction that, though often overlooked, plays a role in shaping people’s perceived reality” (p. 233). In their nonmusic-based study exploring how people hear what their eyes see, they found that the human cognitive system is more than capable of encoding visual rhythm in an essentially auditory manner. However, concerning a more music-specific context, they presumed that such experiences should rather be termed cross-modal *recoding*. That is, they proposed that auditory imagery by music notation develops only after explicit learning, effortful processing, and considerable practice take place. According to Guttman et al., the generation of kinesthetic phonatory and manual motor imagery during music reading is exclusively strategic—not automatic or obligatory.

Although it was not within the objectives of the present study to explore the above issue, our impression from the current results is similar to the assumption made by Drost et al. (2005) and quite the opposite of that advocated by Guttman et al. (2005). That is, within the current study we observed that among musicians who have demonstrable notational audiation skills, music notation appears to be quite automatically and effortlessly transformed from its inherently visual form into an accurate, covert, aural-temporal stream perceived as kinesthetic phonatory and manual motor imagery. We therefore conclude that both kinesthetic-like covert excitation of the vocal folds and concurrently cued manual motor imagery are equally vital components that operate as requisite codependent cognitive strategies toward the interpretation and/or judgment of the visual score—a skill referred to as notational audiation.

## References

- Aleman, A., & Wout, M. (2004). Subvocalization in auditory-verbal imagery: Just a form of motor imagery? *Cognitive Processing*, 5, 228–231.
- Annett, J. (1995). Motor imagery: Perception or action? *Neuropsychologia*, 33, 1395–1417.
- Baddeley, A. D. (1986). *Working memory*. Oxford, England: Oxford University Press.
- Barlow, H., & Morgenstern, S. (1975). *A dictionary of musical themes: The music of more than 10,000 themes* (rev. ed.). New York: Crown.
- Benward, B., & Carr, M. (1999). *Sightsinging complete* (6th ed.). Boston: McGraw-Hill.
- Benward, B., & Kolosic, J. T. (1996). *Ear training: A technique for listening* (5th ed.) [instructor’s ed.]. Dubuque, IA: Brown and Benchmark.
- Brodsky, W. (2002). The effects of music tempo on simulated driving performance and vehicular control. *Transportation Research, Part F: Traffic Psychology and Behaviour*, 4, 219–241.

- Brodsky, W. (2003). Joseph Schillinger (1895–1943): Music science promethean. *American Music*, 21, 45–73.
- Brodsky, W., Henik, A., Rubinstein, B., & Zorman, M. (1998). Demonstrating inner-hearing among highly-trained expert musicians. In S. W. Yi (Ed.), *Proceedings of the 5th International Congress of the International Conference on Music Perception and Cognition* (pp. 237–242). Seoul, Korea: Western Music Research Institute, Seoul National University.
- Brodsky, W., Henik, A., Rubinstein, B., & Zorman, M. (1999). Inner-hearing among symphony orchestra musicians: Intersectional differences of string-players versus wind-players. In S. W. Yi (Ed.), *Music, mind, and science* (pp. 374–396). Seoul, Korea: Western Music Research Institute, Seoul National University.
- Brodsky, W., Henik, A., Rubinstein, B., & Zorman, M. (2003). Auditory imagery from music notation in expert musicians. *Perception & Psychophysics*, 65, 602–612.
- Drost, U. C., Rieger, M., Brass, M., Gunter, T. C., & Prinz, W. (2005). When hearing turns into playing: Movement induction by auditory stimuli in pianists. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 58A, 1376–1389.
- Elliott, C. A. (1982). The relationships among instrumental sight-reading ability and seven selected predictor variables. *Journal of Research in Music Education*, 30, 5–14.
- Gathercole, S. E., & Martin, A. J. (1996). Interactive processes in phonological memory. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 73–100). Hove, England: Psychology Press.
- Gerhardstein, R. C. (2002). The historical roots and development of audiation. In B. Hanley & T. W. Goolsby (Eds.), *Music learning: Perspectives in theory and practice* (pp. 103–118). Victoria, British Columbia: Canadian Music Educators Association.
- Gordon, E. E. (1975). *Learning theory, patterns, and music*. Buffalo, NY: Tometic Associates.
- Gordon, E. E. (1993). *Learning sequences in music: Skill, content, and patterns*. Chicago: GIA Publications.
- Guttman, S. E., Gilroy, L. A., & Blake, R. (2005). Hearing what the eyes see: Auditory encoding of visual temporal sequences. *Psychological Science*, 16, 228–235.
- Halpern, A. R. (2001). Cerebral substrates of music imagery. In R. J. Zatorre & I. Peretz (Eds.), *Annals of the New York Academy of Science: Vol. 930. The biological foundations of music* (pp. 179–192). New York: New York Academy of Science.
- 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.
- Holan, R. (2000). *Fifty ways to love your drumming*. Kfar Saba, Israel: Or-Tav Publications.
- Hubbard, T. L., & Stoeckig, K. (1992). The representation of pitch in music imagery. In D. Reisberg (Ed.), *Auditory imagery* (pp. 199–236). Hillsdale, NJ: Erlbaum.
- Intons-Peterson, M. J. (1992). Components of auditory imagery. In D. Reisberg (Ed.), *Auditory imagery* (pp. 45–72). Hillsdale, NJ: Erlbaum.
- Jacques-Dalcroze, E. (1921). *Rhythm, meter, and education*. New York: Putnam.
- Kalamoski, V. (2001). Music imagery and working memory. In R. I. Godoy & H. Jorgensen (Eds.), *Music imagery* (pp. 43–56). Lisse, The Netherlands: Swets & Zeitlinger.
- Karpinski, G. S. (2000). *Aural skills acquisition: The development of listening, reading, and performing skills in college-level musicians*. New York: Oxford University Press.
- Kopiez, R., Weihs, C., Ligges, U., & Lee, J. I. (2006). Classification of high and low achievers in a music sight-reading task. *Psychology of Music*, 34, 5–26.
- Langheim, F. J. P., Callicott, J. H., Mattay, V. S., Duyn, J. H., & Weinberger, D. R. (2002). Cortical systems associated with covert music rehearsal. *Neuroimage*, 16, 901–908.
- Larson, S. (1993). Scale-degree function: A theory of expressive meaning and its application to aural skills pedagogy. *Journal of Music Theory Pedagogy*, 7, 69–84.
- Lee, J. I. (2004). *Component skills involved in sight-reading*. Unpublished doctoral dissertation, Hanover University of Music and Drama, Hanover, Germany.
- Lehmann, A. C., & Ericsson, A. (1993). Sight-reading ability of expert pianists in the context of piano accompanying. *Psychomusicology*, 12, 182–195.
- Lehmann, A. C., & Ericsson, A. (1996). Performance without preparation: Structure and acquisition of expert sight-reading and accompanying performance. *Psychomusicology*, 15, 1–29.
- MacKay, D. G. (1992). Constraints on theories of inner speech. In D. Reisberg (Ed.), *Auditory imagery* (pp. 121–150). Hillsdale, NJ: Erlbaum.
- Martin, D. W. (1952). Do you auralize? *Journal of the Acoustical Society of America*, 24, 416.
- Mikumo, M. (1994). Motor encoding strategy for pitches and melodies. *Music Perception*, 12, 175–197.
- Petsche, H. J., von Stein, A., & Filz, O. (1996). EEG aspects of mentally playing an instrument. *Cognitive Brain Research*, 3, 115–123.
- Raffman, D. (1993). *Language, music, and mind*. Cambridge, MA: MIT Press.
- Reisberg, D. (Ed.). (1992). *Auditory imagery*. Hillsdale, NJ: Erlbaum.
- Schneider, A., & Godoy, R. I. (2001). Perspectives and challenges of music imagery. In R. I. Godoy & H. Jorgensen (Eds.), *Music imagery* (pp. 5–26). Lisse, The Netherlands: Swets & Zeitlinger.
- Schröger, E. (2005). Mental representations of music—combining behavioral and neuroscience tools. In G. Avanzini, S. Koelsch, L. Lopez, & M. Majno (Eds.), *Annals of the New York Academy of Science: Vol. 1060. The neurosciences and music II: From perception to performance* (pp. 98–99). New York: New York Academy of Sciences.
- Schumann, R. (1967). Musikalische Haus—und Lebensregeln [Musical house and the rules of life]. In W. Reich (Ed.), *Im eigenen Wort* [In his own words] (pp. 400–414). Zurich, Germany: Manesse Verlag. (Original work published 1848)
- Schurmann, M., Raij, T., Fujiki, N., & Hari, R. (2002). Mind's ear in a musician: Where and when in the brain. *Neuroimage*, 16, 434–440.
- Seashore, C. (1919). *The psychology of music talent*. Boston: Silver Burdett.
- Seashore, C. (1938). *The psychology of music*. New York: Dover Publications.
- Sloboda, J. (1981). The use of space in music notation. *Visible Language*, 15, 86–110.
- Smith, J., Reisberg, D., & Wilson, E. (1992). Subvocalization and auditory imagery: Interactions between inner ear and inner voice. In D. Reisberg (Ed.), *Auditory imagery* (pp. 95–120). Hillsdale, NJ: Erlbaum.
- Smith, J., Wilson, M., & Reisberg, D. (1995). The role of subvocalisation in auditory imagery. *Neuropsychology*, 33, 1422–1454.
- Sokolov, A. N. (1972). *Inner speech and thought*. New York: Plenum Press.
- Spagnardi, R. (2003). *Understanding the language of music; a drummer's guide to theory and harmony*. Cedar Grove, NJ: Modern Drummer Publications.
- Vygotsky, L. S. (1986). *Thought and language*. Cambridge, MA: MIT Press.
- Walters, D. L. (1989). Audiation: The term and the process. In D. L. Walters & C. C. Taggart (Eds.), *Readings in music learning theory* (pp. 3–11). Chicago: GIA Publishers.
- Waters, A. J., Townsend, E., & Underwood, G. (1998). Expertise in music sight reading: A study of pianists. *British Journal of Psychology*, 89, 123–149.
- Waters, A. J., & Underwood, G. (1999). Processing pitch and temporal

- structure in music-reading: Independent or interactive processing mechanisms. *European Journal of Cognitive Psychology*, *11*, 531–533.
- Waters, A. J., Underwood, G., & Findlay, J. M. (1997). Studying expertise in music reading: Use of a pattern-matching paradigm. *Perception & Psychophysics*, *59*, 477–488.
- Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin & Review*, *8*, 44–57.
- Wöllner, C., Halfpenny, E., Ho, S., & Kurosawa, K. (2003). The effects of distracted inner-hearing on sight-reading. *Psychology of Music*, *31*, 377–390.
- Zatorre, R. J., & Halpern, A. R. (1993). Effect of unilateral temporal-lobe excision on perception and imagery of songs. *Neuropsychologia*, *31*, 221–232.
- Zatorre, R. J., & Halpern, A. R. (2005). Mental concerts: Music imagery and auditory cortex. *Neuron*, *47*, 9–12.
- Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of music imagery and perception. *Journal of Cognitive Neuroscience*, *8*, 29–46.

Received October 20, 2006

Revision received May 16, 2007

Accepted May 23, 2007 ■

### Low Publication Prices for APA Members and Affiliates

**Keeping you up-to-date.** All APA Fellows, Members, Associates, and Student Affiliates receive—as part of their annual dues—subscriptions to the *American Psychologist* and *APA Monitor*. High School Teacher and International Affiliates receive subscriptions to the *APA Monitor*, and they may subscribe to the *American Psychologist* at a significantly reduced rate. In addition, all Members and Student Affiliates are eligible for savings of up to 60% (plus a journal credit) on all other APA journals, as well as significant discounts on subscriptions from cooperating societies and publishers (e.g., the American Association for Counseling and Development, Academic Press, and Human Sciences Press).

**Essential resources.** APA members and affiliates receive special rates for purchases of APA books, including the *Publication Manual of the American Psychological Association*, and on dozens of new topical books each year.

**Other benefits of membership.** Membership in APA also provides eligibility for competitive insurance plans, continuing education programs, reduced APA convention fees, and specialty divisions.

**More information.** Write to American Psychological Association, Membership Services, 750 First Street, NE, Washington, DC 20002-4242.