

## A Review of Daniel Reisberg (Ed.), *Auditory Imagery*\*

Bruno H. Repp

This book appears to be the first in the psychological literature to be devoted entirely to the topic of auditory imagery. Research on visual imagery has been going on for some time, spearheaded by such authors as Roger Shepard, Stephen Kosslyn, and Ronald Finke. Characteristically, research on the analogous phenomenon in audition has lagged behind, and it is fair to say that even now it is not an area teeming with activity. The purpose of the present volume is evidently to stimulate interest in the topic, as well as to review whatever pertinent findings have been obtained so far. A perusal of the 10 chapters reveals that these findings are still very limited and leads one to wonder whether auditory imagery is going to be as fertile an area of investigation as visual imagery has proved to be.

The relative paucity of empirical data is compensated by the diversity of angles from which the topic is illuminated in this book. Three of the ten chapters deal with musical imagery, two with simple sounds, five with speech. Among the latter, there are discussions of inner speech in the deaf and of auditory hallucinations in schizophrenics. Most of the authors are well-established researchers, though not necessarily in areas primarily concerned with auditory imagery. While some were able to simply summarize their own research, others had the more difficult task of deriving implications for auditory imagery from their ideas and findings on related topics. All contributions, however, are well-written and interesting.

The articles on speech will perhaps be of greater interest to the readers of this journal (i.e., LANGUAGE AND SPEECH) than those on music and other nonspeech sounds. Nevertheless, the issues addressed in the music research are quite

pertinent to speech also. Imagery is that facet of memory which retains or regenerates the analog characteristics of the original, modality-specific perceptual experience—it is “surface memory,” as it were. Being such a general function, it is equally relevant to speech, music, and environmental sounds.

The order of the chapters is, in editor Reisberg's own words, “somewhat arbitrary,” though he does mention the loose organizational principle he had in mind. My summary follows a different but not necessarily better order.

Margaret Jean Intons-Peterson (Chapter 3) acknowledges the theoretical debt of auditory imagery research to visual imagery research; according to her, there are no specific models of auditory imagery as yet, and the models borrowed from vision focus primarily on the relation of imagery to perception. She reviews some of the theoretical concepts as well as the results of several experiments, most of which employed simple stimuli and used reaction time as the dependent variable. In her own research of more than a decade ago, for example, she showed that the time to compare two imagined environmental sounds with respect to their loudness increases with the difference in loudness between them, as assessed by previously having subjects rate the typical loudnesses of these sounds on a scale. This suggests that loudness is a property that is represented in auditory images. Intons-Peterson reviews similar findings suggesting that pitch and timbre are represented literally in images.<sup>1</sup>

The evidence for timbre comes from the work of Robert Crowder who, with Mark Pitt (Chapter 2), reports original data that extend his earlier findings. Despite its narrow focus, this research is significant because it demonstrates, perhaps more convincingly than any other research reported so far, that subjects are able to generate specific auditory images from verbal instructions. Crowder and Pitt asked listeners to make same/different

---

I am grateful for the support of Haskins Laboratories during the preparation of this review.

judgments about the pitches of two successive tones. When these tones differed in timbre, subjects were slower in making "same pitch" judgments than when the tones had the same timbre as well as pitch. The crucial finding was that the same effect emerged when the first tone in each pair was a sine wave plus a verbal instruction to imagine a particular instrument timbre: Subjects responded faster and/or more accurately to same-pitch pairs when the second tone had the same timbre as the *imagined* timbre of the first tone. In their chapter, Crowder and Pitt report a replication of this finding for plucked versus bowed cello tones. In a subsequent attempt to separate the static and dynamic aspects of this timbre contrast, they used synthetic sounds differing in either spectrum or rise time. They obtained the desired effect with the former variation, but not with the latter. Their tentative conclusion was that dynamic cues to timbre are not represented in imagery, whereas static spectral properties are. Thus, besides providing an elegant experimental demonstration of imagery, these results also suggest that there are aspects of the original perceptual experience that cannot be recreated faithfully.<sup>2</sup>

Some of the evidence for the representation of pitch in auditory imagery comes from Andrea Halpern's research on imagery for songs, which she reviews in Chapter 1. Since songs essentially are a form of speech produced with a particular prescribed rhythmic and intonational pattern, the questions she asked are equally applicable to imagery for memorized stretches of nonmusical speech, such as poems or literary texts. (Psychological research has, unfortunately, neglected these more artistic forms of language use.) Halpern was interested in the temporal layout of imagined songs, and in whether their temporal extent matches that observed in actual singing. She used a probe task similar to one used in studies of the spatial layout of visual images: Subjects were given the initial word of a familiar song and had to decide whether a second word, presented shortly afterwards, occurred in the text of that song. Reaction times for correct responses increased monotonically with the distance between the first and second words in the song, whether or not subjects were instructed to imagine it. In another study, subjects had to make judgments about the relative pitch heights of two monosyllabic words in familiar songs. Again, reaction times increased with the distance between these words. Thus, the subjects (nonmusicians) seemed to scan through a

temporal representation of the songs in their heads, as one should expect if auditory imagery is veridical. In further studies, Halpern showed that subjects' imagined tempo was comparable to their preferred tempo when listening to the same song.<sup>3</sup> Halpern also reports some data suggesting that listeners, regardless of musical training, have a long-term memory for the approximate starting pitch of familiar songs and are able to reproduce it by singing, choosing a tone on a keyboard, or giving ratings to presented pitches, though not nearly with the accuracy that possessors of "absolute pitch" might display.

The representation of pitch in musical images is discussed in much more detail by Timothy Hubbard and Keiko Stoeckig (Chapter 9). In fact, their treatment is perhaps somewhat too broad and abstract, lending a certain turgidity to their long chapter. They use the term "qualia" to refer to "a sensory quality or 'raw feel' that makes the experience of imaging similar to the experience of perceiving in a way that abstract representation is not" (p. 199). They also point out that the presence of these qualia has rarely been the focus of the work they review, which includes various models of the mental representation of music (psychoacoustic, rule-based, connectionist), studies of memory for isolated pitches and melodies, theories of representational form, the issue of cognitive penetrability, and questions of methodology. The chapter is valuable in that it provides a broad framework within which to view research on music imagery; however, it also raises the question (in my mind) of whether the subjective "qualia" are really all that important. The more significant question is perhaps how people use their memories and images of auditory properties to accomplish various tasks that are important in their lives. Hubbard and Stoeckig's discussion occasionally gives the impression that their primary aim is the pragmatic one of providing experimental psychologists with grist for their laboratory mills.

The final chapter on music (Chapter 10), by Diana Deutsch and John Pierce, clarifies what I mean by that comment. This unusual and decidedly iconoclastic contribution starts with a series of historical quotes that document the essential and unquestionable role of auditory imagery in composing. The historical survey goes on to bolster the authors' contention that scientists of earlier centuries were well informed about musical phenomena and usually took them into account when theorizing about human auditory capabilities, whereas in this century scientific reductionism has led to a musically uninformed

tradition of laboratory psychoacoustics. Deutsch and Pierce argue forcefully against the exclusive focus on low-level explanations in much current auditory research and go so far as to suggest that "most [psychoacoustic] models are a distraction from thought and an inspiration to certitude for the uncertain" (p. 250). They go on to make another controversial point, namely that "good demonstrations are more convincing than experiments," and go on to cite several of these demonstrations, which predictably include Deutsch's well-known octave and scale illusions. These phenomena, however, are of a fairly simple nature, and their relevance to actual music experience is not always clear, as the authors themselves realize. At the end of their chapter, Deutsch and Pierce express the hope that modern music technology will lead to insights that traditional psychoacoustics has failed to provide. Although they come across as favoring the insights of "amateurs at work" (the title of one of their sections) over those of experimental psychologists, their bottom line is that better communication is needed between practicing musicians and ivory tower scientists. The connection with the specific theme of the book becomes very loose by the end of this provocative chapter.

The remaining five chapters deal with various manifestations of linguistic imagery, or "inner speech." Alan Baddeley and Robert Logie (Chapter 8) review some of Baddeley's well-known research on the "phonological loop" of working memory. They speculate that this component may "represent the seat of auditory imagery" (p. 180). Although its function in the temporary storage and maintenance of speech materials has been extensively investigated, the phenomenology of the accompanying imagery, if any, has not been studied. Baddeley and Logie feel that there is an urgent need for such studies. A related question that has been addressed is whether subvocal articulation is involved in phonological working memory. There is evidence that concurrent speech input, as well as speech produced by subjects themselves interferes with some tasks that rely on phonological short-term memory; other tasks, however, seem unaffected. Moreover, there are reports of neurological patients who have normal speech but show a phonological memory deficit, as well as of children who have never learned to speak but show relatively normal phonological function. These findings suggest that phonological processing may be relatively abstract and not necessarily connected with vivid auditory imagery.

This issue is pursued further by Ruth Campbell (Chapter 4) in her discussion of inner speech in deaf individuals. She reviews evidence that individuals without hearing can develop adequate phonological skills, including the ability to read in an alphabetic orthography. Campbell also cites her own, by now well-known finding that lipread words are retained in phonological form, just like auditorily presented speech. This suggests that phonological representations are not necessarily (perhaps even: necessarily not) auditory.

In Chapter 5, David Smith, Daniel Reisberg, and Meg Wilson distinguish between an "inner voice" and an "inner ear" that listens to the inner voice. They examined the effects of concurrent auditory input and/or concurrent articulatory activity on performance on several phonological tasks. In each case, both types of interference were effective, leading the authors to conclude that both processes were involved. (That the articulatory activity also produced auditory input in several instances is a complication that the authors do not discuss.) In a version of Crowder's timbre imaging task, it appeared that only auditory input interfered, so that only the inner ear, but not the inner voice, seemed to be involved—a reasonable conclusion. Some of the other evidence these authors discuss converges with that discussed by Baddeley and Logie (whose phonological store and loop indeed correspond to the inner ear and voice, respectively), and the existence of more abstract phonological processes is also acknowledged by reference to a "lexical ear."

A more critical stance is taken by Donald MacKay (Chapter 6), whose research on internal phonological processes spans more than two decades. With considerably more self-confidence than most other authors in this book, he states right at the outset that inner speech is nonarticulatory and nonauditory. The first claim agrees with the evidence for the abstractness of phonology discussed in other chapters. With regard to the latter claim, MacKay points out that "what seems phenomenally to be auditory often is not" (p. 126), thus revealing a major problem for any phenomenological approach to auditory imagery. MacKay points out (relying on introspection, it seems) that inner speech usually lacks loudness and fundamental frequency variation; he attributes the processing of these qualities to a separate "auditory concept system," distinct from the phonological system. Imagining a concrete voice or words spoken with a specific intonation requires both of these systems. MacKay

is critical of some of Reisberg's and Baddeley's conclusions, particularly of the "inner ear" concept: "...the internal listener concept is functionally questionable: The 'double agent' approach to comprehension of internal speech must address the fundamental issue of why speakers must independently 'listen to' the meaning and sound of what they are saying internally when they know all along the meaning and sound of what they are saying" (p. 140). A number of additional related issues are discussed in this chapter, probably the most thought-provoking in the collection.

The last chapter to be mentioned is that by David Smith (Chapter 7) on the auditory hallucinations of schizophrenia. Halpern and MacKay both briefly alluded to the sometimes involuntary and persistent nature of auditory images (such as a "haunting" melody), a phenomenon that is found most dramatically in the illusory voices reported by schizophrenics, which often seem to come from within and seem to speak intelligibly. Smith, revitalizing a neglected theory of these hallucinations, argues that they are a form of inner speech contingent on subvocalization. He cites a variety of reports, some anecdotal, that engagement of the articulators results in a reduction of the vocal hallucinations. The evidence remains merely intriguing but points to a more specific hypothesis about the nature of these hallucinations.

In summary, this is a stimulating collection of articles on a relatively neglected topic. It demonstrates that there are a variety of activities in which auditory imagery plays a role, though some of the most obvious (musical composition and performance) are barely mentioned. In other cases, internal processes rather more abstract than images seem to be involved. The tasks employed in most of this research are fairly distant from real life, and it is not so clear whether further laboratory demonstrations of this kind will contribute any important insights about auditory imagery. It would be worthwhile, perhaps, to look more closely at auditory

properties that *cannot* be imagined, such as hinted at by Crowder and Pitt, rather than at those that can. More generally, imagery is perhaps better viewed merely as one end of the memory continuum, the one dealing with analog information, rather than as a special phenomenon on the basis of its subjective "qualia." Its relative neglect in the past may be attributed to experimental psychologists' characteristic preoccupation with discrete, symbolic (and if analog, then visual) processes, and with activities that are technologically rather than culturally significant. I conclude by paraphrasing what I take to be one of Deutsch and Pierce's messages: If you want to learn about auditory imagery, look at what composers do.

## REFERENCE

- Crowder, R. G. (1973). Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology*, 98, 14-24.

## FOOTNOTES

- \*Hillsdale, NJ: Lawrence Erlbaum Associates, 1992. 274 pp. \$45.00.  
This review appears in *Language and Speech*, 35, 341-346 (1992).
- <sup>1</sup>Intons-Peterson also reports that the time needed to generate the image of a single sound does not depend on its loudness. However, it is not clear why it should; this point is also made by Hubbard and Stoeckig in Chapter 9 (p. 221, Footnote 2). The same comment applies to Intons-Peterson's analogous results for pitch comparisons. Her conclusion that loudness and pitch are sometimes not represented in images may not be justified.
- <sup>2</sup>Although the authors do not elaborate on this point, the failure to find evidence for dynamic properties in auditory images is clearly reminiscent of Crowder's well-known finding that stop consonants are poorly retained in precategorical auditory memory (see, e.g., Crowder, 1973). It would be interesting to see a demonstration that stop consonants are likewise difficult to imagine.
- <sup>3</sup>It should not be concluded from these data, however, that a memorized song (or stretch of speech) needs to be scanned from beginning to end in order to determine that some word or phrase occurs in it; surely, there must be multiple "access points" in longer structures such as Handel oratorios or plays by Shakespeare. Scanning such as demonstrated by Halpern may be mandatory, however, within small structural units such as sentences or clauses. The size of the units that must be exhaustively scanned would be a worthwhile topic for further investigation, as it bears on the memory representation of large-scale musical and linguistic structures.