feature article

Hearing music in ensembles

Diana Deutsch

Musical illusions created in the laboratory or concert hall provide insight on how the brain regroups auditory inputs.

Diana Deutsch is a professor of psychology at the University of California, San Diego.


SAN FRANCISCO SYMPHONY

Figure 1. Playing Wolfgang Amadeus Mozart's oboe concerto K 314**.**

Imagine you are listening to an orchestral performance in a concert hall (figure 1). The sounds produced by the different instruments are mixed together and subject to numerous distortions as they travel to your ears. Somehow you are able to disentangle the components of the complex, time-varying spectra that impinge on your ears, so that you hear, for example, the first violins playing one set of tones, the flutes another, and the clarinets yet another. You group together the perceived sounds so that you hear melodies, harmonies, timbres, and so on. What algorithms does the auditory system employ to accomplish those difficult tasks? And how successful are those algorithms?

In the 20th century, most research on sound perception was focused on how we perceive single sounds in isolation—or simple combinations of sounds. But there is a growing realization that this approach considers only a limited part of the hearing process. When ongoing streams of sound are presented in parallel, as occurs in the natural environment, powerful higher-level mental mechanisms come into play, mechanisms that can substantially modify what is heard. The perception of music in ensembles provides a particularly good model for investigating those mechanisms. Research on such issues being highly interdisciplinary, physicists, engineers, psychologists, neuroscientists, and musicians have all made major contributions.

Fusion and separation of sound components

The sounds produced by orchestral instruments consist of many Fourier components. So when different sounds are playing simultaneously, the auditory system has to decide which components to link with which. One might surmise that the linking is achieved by grouping together those components that appear to have originated from the same spatial location. However, given the numerous distortions—due to reflections from the walls, floors, and ceilings; the presence of occluding objects; and so on—that occur in sound spectra as they travel from their sources, cues to spatial location that work well under simple laboratory conditions could here mislead us into separating components that should in fact be grouped together (see the article by Bill Hartmann in PHYSICS TODAY, November 1999, page 24). So, rather than attempting to link sound components by commonality of source location, it makes sense to form groupings based primarily on other, more reliable, cues, with location playing a prominent role only under certain circumstances.[1]

What are the relationships between the components of a complex sound spectrum that lead us to fuse them into a unitary sound image, so that they combine to produce, say, the perception of a violin playing a single tone? And what relationships lead us to separate components so that we obtain multiple sound images such as a tone of one pitch from a violin together with a tone of different pitch from a clarinet?

Harmonicity

One prominent cue is harmonicity, the sequence of integral (harmonic) multiples of a fundamental frequency sounded on a musical instrument. Many blown or stringed instruments, as well as the human voice, give rise to overtone components that are in harmonic or close to harmonic relationship to the fundamental. So it's a good bet for the auditory system to fuse together components that stand in harmonic relationship and to attribute to the resultant complex tone a pitch that corresponds to the fundamental. And indeed that is what generally occurs. Furthermore, as expected from that line of reasoning, bells and gongs, whose components are nonharmonic, give rise to more complex pitch sensations.

We can then ask how far a single component of a harmonic complex can be made to deviate from harmonicity and still fuse perceptually with the remaining components to produce the impression of a single tone with one pitch. In 1986, Brian Moore and colleagues at Cambridge University presented listeners with complex tones whose components were either all strictly harmonic or had one component mistuned from its exact harmonic value. In each case, listeners were to judge whether they heard a single sound with one pitch or two sounds: a complex tone together with a pure tone that did not belong to the complex. The researchers found that, with mistunings up to about 2%, the mistuned harmonic was still grouped perceptually with the others. But as the errant harmonic was increasingly mistuned, its perceived contribution to the complex gradually decreased.[2]

What happens when two complex harmonic tones are played simultaneously? When they are in strict unison, so that their tones are built on the same fundamental, there is a good chance that listeners will fuse them perceptually and hear only a single tone. One striking example occurs at the beginning of Franz Schubert's *Unfinished Symphony*. There a clarinet and oboe play in unison, and their sounds blend together to produce the impression of a single instrument with a novel timbre. Timbre is the tone quality that distinguishes one kind of musical instrument from another.

When the tones played simultaneously by two instruments are built on different fundamentals that stand in simple harmonic relationship, they still tend to blend together perceptually rather than being heard as distinct from each other. The degree of fusion depends on the frequency relationship between the fundamentals. Complex tones fuse best when they are presented in unison (a 1:1 ratio of fundamental frequencies), second best at an octave (a 2:1 ratio), and next best at a fifth (a 3:2 ratio). Rudolf Rasch showed three decades ago that as the relationships between simultaneous complex tones gradually depart from strict harmonicity, the tendency increases for them to be heard as separate sounds.[3]

Composers take account of harmonicity effects in various ways. In polyphonic music, for example, one objective is to maintain the perceptual independence of concurrent voices. For example, David Huron at the Ohio State University has calculated for a sample of Johann Sebastian Bach's polyphonic keyboard works that the composer tended to avoid harmonic intervals to the extent that they promoted fusion.[4] On the other hand, impressionist composers such as Claude Debussy and Maurice Ravel exploited harmonic relationships to create perceptual fusion from different instruments playing simultaneously.

Timing

Another cue that plays a prominent role in promoting fusion or separation of Fourier components is the relative timing of their onsets. The importance of timing can be shown in a striking demonstration, in which a harmonic series is presented with its components entering at different times. Take, for example, a series based on 220 Hz (one octave below "concert A"). The demonstration begins with 220 Hz sounded alone. One second later 440 Hz is added, then one second after that 660 Hz, and so on. As each harmonic component enters, it is briefly heard as standing out clearly. But if all the components continue to sound together for a few seconds, they end up sounding like a single tone with a pitch that corresponds to the 220-Hz fundamental.

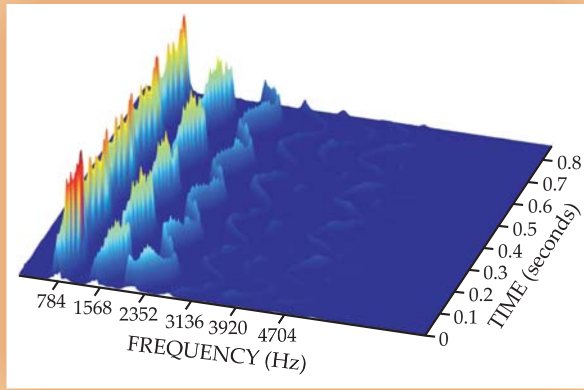

Figure 2. Spectrogram of a single tone and its overtones sung with vibrato. The frequencies of the fundamental tone (784 Hz) and the harmonic series of its overtones all modulate in synchrony at about 6 Hz.

What happens when, instead of combining single Fourier components, we combine two complex tones? Rasch found that an onset difference as small as 10 milliseconds between such tones increased their perceptual separateness, and at an onset difference of 30 ms, the tones were heard as clearly distinct from each other.[3] In a study of trio ensemble performances, Rasch found that onset differences for tones that were nominally synchronous ranged from 30 to 50 ms. That matters because, at such asynchronies, listeners should be able to hear the individual instruments clearly.[5] Indeed, the effect has been exploited in compositional practice. Analyzing performances of Bach's two-part inventions, for which the composer's intention was that the individual parts should stand out clearly, Huron has found that values of onset asynchrony are significantly higher than would happen by chance.[6]

Yet another cue that contributes to perceptual grouping of the components of a complex tone is continuous frequency modulation. Natural sustained sounds generated by musical instruments constantly undergo small frequency fluctuations that preserve the ratios formed by their component frequencies. This is illustrated in figure 2, which displays a spectrogram of a tone sung with vibrato. Building on that knowledge, composer John Chowning created an impressive demonstration at Stanford University. He added to an electronically synthesized singing voice a frequency fluctuation that preserved frequency ratios between the individual harmonics, thereby producing a convincing sense of fusion.

How do source-location differences contribute to producing a sense of fusion or separation? It has been shown that for simultaneous sound complexes, the spatial cue is quite weak. For example, psychologists Chris Darwin and Valter Ciocca at the University of Sussex in England have shown that when a tone plus true harmonics is presented to one ear, the perceived contribution of a mistuned harmonic is about the same, irrespective of whether it's presented to the same ear or the other.[7] But I have found that if there are onset time differences between tones presented to the two ears, listeners are more likely to group the tones by spatial location.[8]

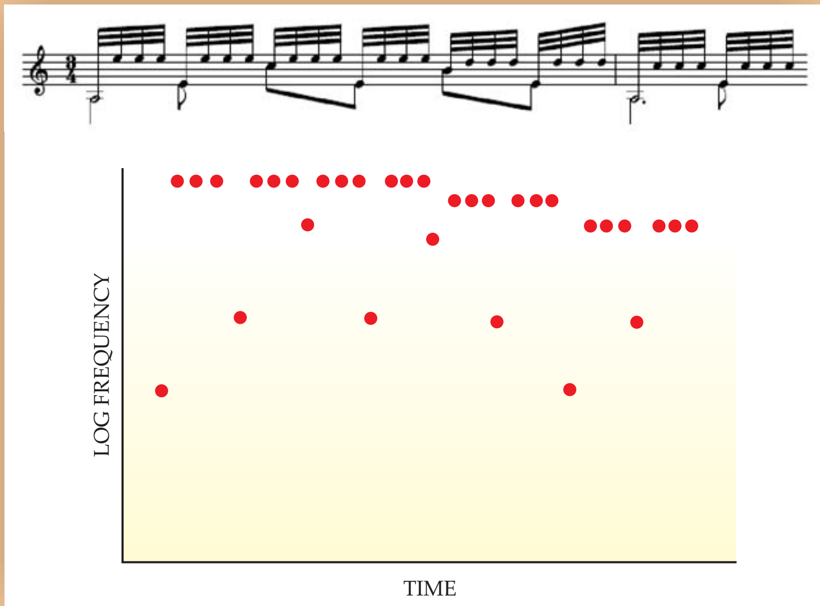

Figure 3. A guitar passage from the beginning of *Recuerdos de la Alhambra*, by Francisco Tarrega. The hearer perceives parallel melodic lines organized by pitch proximity. (Adapted from ref. 1.)

Larger-scale groupings

What happens when, instead of single tone complexes, sequences of tones are presented to a hearer? At that level, the auditory system abstracts further relationships between tones and uses them as additional grouping cues. One prominent cue is pitch proximity: The hearer tends to link successive tones that are close in pitch and to separate those that are further apart. That's useful in nature because sounds that are in the same pitch neighborhood are likely to be coming from the same source.

The pitch-proximity effect is particularly salient when tones are presented at a rapid tempo—around 8 to10 per second. At such tempi, when the tones are drawn from different pitch ranges, listeners do not form perceptual relationships between the temporally adjacent tones. Instead, they perceive two melodic lines in parallel, one corresponding to the higher-pitch tones and the other to the lower ones.

That effect was frequently employed by baroque composers such as Bach and Georg Philipp Telemann. Striking examples can be found in Bach's cello suites. Even more dramatic examples occur in guitar music of the subsequent classical and romantic periods. Figure 3, showing a guitar passage from *Recuerdos de la Alhambra*, by the late-19th-century Spanish composer Francisco Tarrega, provides a good example. The listener perceives two simultaneous but separate musical streams, corresponding to the two pitch lines in the passage.

A considerable body of findings has been amassed concerning the parameters governing the pitch-proximity effect. For example, psychologist Albert Bregman at McGill University in Montreal has performed experiments in which listeners are presented with two tones, A and B, in a continuing pattern . . . ABAB . . . . He then asks them to determine whether they hear one stream of tones or two. Figure 4 shows typical results for small and large pitch differences between A and B. As the pitch difference increases, one begins to hear two separate streams. The faster the tempo, the smaller is the threshold pitch difference at which one perceives separate streams.[9]

An interesting consequence of this streaming effect is that the greater the pitch disparity between the alternating tones, the more difficult it is to perceive temporal relationships between them. Physicist Leon van Noorden explored that effect in his influential 1975 doctoral research at Eindhoven Polytechnic in the Netherlands. He presented tones in the continuing pattern

. . . —ABA—ABA— . . . ,

where the dashes represent silent intervals. He found that when the frequency difference between the tones was small, listeners perceived a galloping rhythm. With increased frequency disparity between the tones, however, the galloping rhythm was no longer perceived. Instead, listeners heard two isochronous but unrelated streams of tones.

Neural underpinnings

There is considerable interest in the neural underpinnings of the streaming effect.[10] Six years ago Yonatan Fishman and coworkers at the Albert Einstein College of Medicine in New York City recorded cortical neural responses to ABAB sequences in awake macaque monkeys.[11] Fishman and his colleagues adjusted frequency A to elicit the strongest neural response at the cortical recording site and then varied frequency B. At slow tempi, the cortex showed marked re-

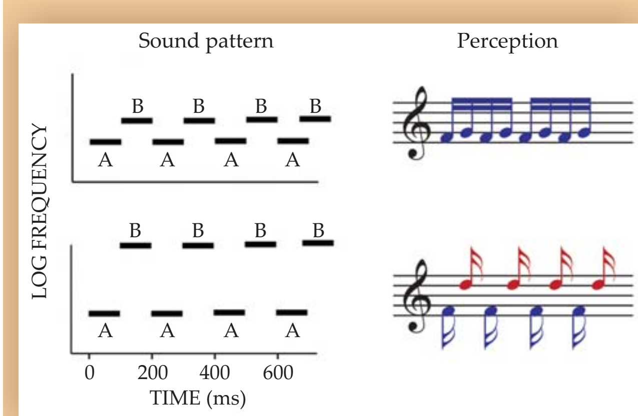

Figure 4. Perceptual segregation of tone sequences is studied by presenting listeners with a sequence of tones that alternate between two frequencies, A and B. When the frequency difference between them is small (top row), listeners perceive a single coherent stream of tones. But when it is large (bottom row), the sequence breaks apart perceptually into two unrelated streams, indicated by the colors.

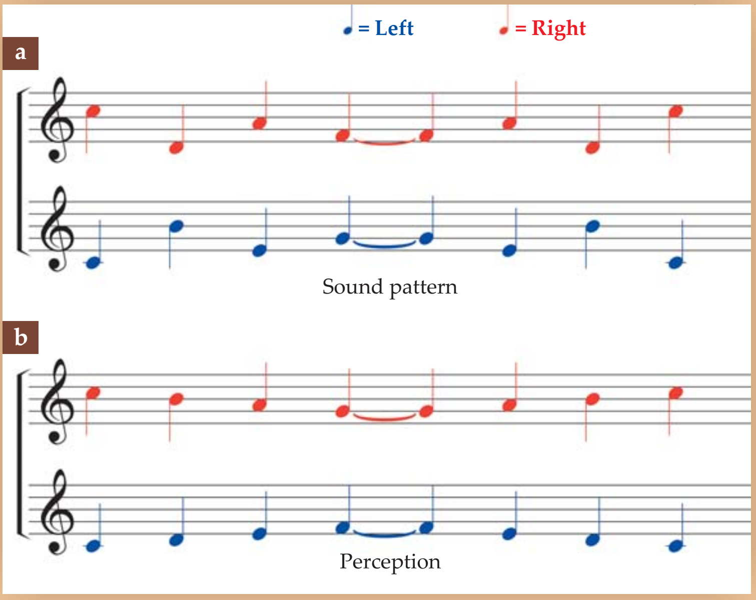

Figure 5. In the scale illusion, two scales an octave apart are presented simultaneously through a headphone in ascending and descending form. **(a)** As presented, the tones in each earphone (left or right, coded by color) alternate between descending notes of one scale and ascending notes of the other. The pattern is played continuously without pause. **(b)** As perceived by most right-handers, the right ear hears a melody formed by the higher tones while the left ear hears another formed by the lower tones. (Adapted from ref. 1.)

sponses to both A and B. But at fast tempi, the responses to B tones became increasingly weaker as the frequency difference between A and B was increased. At large frequency separations and fast tempi, the neural responses were predominantly to the A tones. The findings parallel those obtained in perceptual studies on human listeners. Fishman hypothesized that the stream segregation reported in the human studies is based, at least in part, on the neural interactions observed in their experiment.

Perceptual groupings also depend on differences in instrument timbre. Composers make considerable use of timbre so that the listener can follow the lines produced by different instruments, even when their pitch ranges overlap. One example occurs at the beginning of the second movement of Ludwig van Beethoven's *Spring Sonata* for violin and piano, Opus 24. Although the tones played by the two instruments overlap substantially in pitch, the listener clearly hears two melodic lines in parallel, each played by a different instrument.[1]

That timbre effect was studied in detail at McGill by Caroline Bey and Stephen McAdams.[12] They presented subjects with a "target" melody interleaved with a "distractor" sequence. They then presented a test melody and asked listeners to decide whether or not it occurred in the composite target–distractor sequence. Listeners performed badly when the target and distractor melodies had both the same pitch range and the same timbre. But their performance improved substantially as the difference in timbre between the target and distractor melodies was increased, even when their pitch ranges remained the same.

The scale illusion

Clear illustrations of grouping by pitch proximity occur when two simultaneous streams of tones are presented, each from a different position in space. The so-called scale illusion provides a particularly striking example. Figure 5 shows the scale illusion as I originally reported it 35 years ago.[13] Simultaneous ascending and descending scales of electronically generated tones are presented to the listener via headphone, with successive tones in each scale alternating between the left and right ears, as shown in the figure. So the right ear receives one disjoint, jumpy sequence of pitches, while the left ear simultaneously receives a different but equally disjoint sequence. The tones are purely sinusoidal and sustained, with no amplitude drops at the transitions between them, and the pattern is played repeatedly without pause.

That presentation can produce a number of different illusions, which vary from one listener to another. Figure 5b shows the illusion most frequently reported by right-handers. A melody corresponding to the higher tones appears to be coming from the right earphone, while another melody, corresponding to the lower tones, is perceived as coming from the left one. Even when the earphone positions are reversed, the higher tones still seem to come from the right and the lower tones from the left. The illusion gives the bizarre and sometimes unnerving impression that reversing the headphone has caused the higher and lower scales to switch earphones. But other listeners, particularly left-handers, experience different illusions. Sometimes they are just mirror images of the typical right-hander's illusion; but sometimes they are quite complex.

The scale illusion can also occur with tones generated by natural instruments and presented in normal room environments. Its localization effect is generally not as strong as with headphones and pure electronic tones, but the perceptual reorganization of the pitch patterns into higher and lower melodic lines still occurs.

Similar illusions can be created with different musical patterns. One example comes in the beginning of the last movement of Pyotr Ilyich Tchaikovsky's Sixth Symphony, the *Pathetique*. There the theme and accompaniment are alternated between the first and second violin sections, so that each section plays every other note of the theme. In Tchaikovsky's day, the two violin sections were to the extreme left and right of the orchestra, so that the audience might be thought to have perceived the theme as wafting back and forth across the stage.

Nowadays, the two violin sections are usually adjacent on the conductor's left. But conductor Thomas Nee at the University of California, San Diego, where I was teaching, tried an informal experiment. He seated the university orchestra as it would have been in Tchaikovsky's day, and we found that when the passage from the *Pathetique* was played, the scale illusion emerged strongly: The theme seemed to be coming from one set of instruments on one side of the orchestra, with the accompaniment seeming to come from the other side.

There's an interesting story concerning the *Pathetique*

Hermann von Helmholtz

Hermann Ludwig Ferdinand von Helmholtz (1821–1894) was one of the great scientists of his age. Known primarily as a physicist, he also made important contributions to physiology and psychology. His 1859 book *On the Sensations of Tone as a Physiological Basis for the Theory of Music* is still valuable reading.[16] Helmholtz was a strong proponent of the theory of unconscious inference in perception. The theory asserts that our perceptions are heavily influenced by our unconscious knowledge and experience of the environment. He applied the theory to the perception of music by arguing, for example, that we employ our unconscious knowledge of the relationships between components of a complex sound to fuse together components that are likely to have emanated from the same source and to separate those that are likely to have emanated from separate sources. His ideas have inspired present-day work on the sometimes illusory perceptual fusion and separation of musical sounds—work that he was unable to carry out with the limited technology of his time.

AIP EMILIO SEGRÈ VISUAL ARCHIVES

passage. In the summer of 1893, Tchaikovsky met with conductor Arthur Nikisch to discuss the symphony, which was soon to have its premiere. Nikisch tried to persuade the composer to rescore the piece so that one set of instruments would play the theme and the other the accompaniment. But Tchaikovsky refused and the piece premiered as originally written. Nevertheless, Nikisch felt so strongly about the matter that he later initiated a second tradition of performing the symphony with the passage rescored as he wanted. A few conductors still adhere to Nikisch's version, though most follow Tchaikovsky's original. It's not known why the two great musicians disagreed so strongly. There's no evidence that either of them realized that an illusion was involved, but it may well be that their argument was based on differing perceptions of the passage—as can happen with the scale illusion.

What becomes of the scale illusion when the sounds from the two sources differ in timbre? When that difference is small, listeners continue to hear the illusion, but it creates a new tone quality that appears to emanate from both spatial locations.[14] But when the difference in timbre is pronounced, the illusion can break down and listeners perceive melodic lines based on timbre.[15]

Why does the scale illusion occur? The 19th-century physicist Hermann von Helmholtz argued that when faced with a complex configuration, our perceptual system adopts the most plausible interpretation in terms of our knowledge of the environment.[16] (See the box above.) Similar sounds are likely to be coming from the same source, and different sounds from different sources. Unless a trick is being purposefully played, it's most unlikely that a source producing sounds in one pitch range should be leaping around between two widely separated locations and that another source producing sounds in a different pitch range should constantly be leaping around in opposite directions. The mind seems to prefer the more plausible illusion that one source produces the higher tones and another source the lower ones.

Another illusion, known as the precedence effect, can also be viewed from Helmholtz's perspective of unconscious inference. To obtain the precedence effect, the listener sits facing two loudspeakers; one to his right and the other to his left. Both speakers present identical sound patterns, but the sounds differ slightly in their onset times. When the two onsets are separated by less than 30 ms—though that threshold value can differ with sound pattern or room acoustics—the listener perceives only a single sound pattern, which appears to be coming from the source of the first-arriving signal; the other loudspeaker appears to be silent, even if it is producing a louder sound.[17] We can surmise that this illusion serves the useful purpose of suppressing unwanted echoes and reverberation from conscious perception. Work on the precedence effect can be traced back to another 19th-century physicist—Joseph Henry.

A related illusion is named after its discoverer Nico Franssen. To obtain the Franssen effect, one employs the same seating arrangement that demonstrates the precedence effect, except that a reverberant environment is particularly important. One loudspeaker presents a tone of abrupt onset, brief duration, and slow falloff, while the other, in parallel, presents a tone with gradual onset and long duration. The listener perceives the two tones as a single entity coming from the speaker that produces the abrupt tone, even though that speaker is completely silent for most of the other tone's duration.[18] Once again, an illusion yields the most plausible perceptual interpretation of the environment. In the real world without psychological tricksters, the abrupt onset of a complex of sounds usually serves to locate the single source.

Interdisciplinary research on how we perceive music in ensembles has yielded rewarding insights concerning the psychological nature of hearing. It has shown that we cannot think of auditory signals as simply moving up from our ears to higher brain centers where they are analyzed in a straightforward fashion. Rather, the brain acts on auditory input in a complex way, so that the sounds we end up hearing are often quite different from what we might naively expect.

Most of the time, our perceptions are useful for figuring out what's happening around us. But under certain circumstances, those usually reliable perceptual mechanisms yield striking illusions. From a musical perspective, research on such illusions shows that we do not necessarily perceive music as it is written in a score—or as one might imagine it from reading the score. Instead, the brain reshapes the auditory signals so that the music we end up hearing has been substantially altered by our own perceptual machinery.

On the author's web site (http://philomel.com) readers can hear various musical illusions, including some of those discussed above to which the online version of this article (at www.physicstoday.org) will provide links.

References

1. D. Deutsch, in *The Psychology of Music*, 2nd ed., D. Deutsch, ed., Academic Press, San Diego, CA (1999), p. 299.
2. B. C. J. Moore, B. R. Glasberg, R. W. Peters, *J. Acoust. Soc. Am.* **80**, 479 (1986).
3. R. A. Rasch, *Acustica* **40**, 21 (1978).
4. D. Huron, *Music Percept.* **9**, 135 (1991).
5. R. A. Rasch, in *Generative Processes in Music: The Psychology of Performance, Improvisation, and Composition*, J. A. Sloboda, ed., Oxford U. Press, New York (1988), p. 71.
6. D. Huron, *Music Percept.* **10**, 435 (1993).
7. C. J. Darwin, in *Auditory Perception of Sound Sources*, W. A. Yost, A. N. Popper, R. R. Fay, eds., Springer, New York (2008), p. 215.
8. D. Deutsch, *Percept. Psychophys.* **25**, 399 (1979).
9. A. S. Bregman, *Auditory Scene Analysis: The Perceptual Organization of Sound*, 2nd ed., MIT Press, Cambridge, MA (1994).
10. C. Micheyl et al., *Hear. Res.* **229**, 116 (2007).
11. Y. I. Fishman, J. C. Arezzo, M. Steinschneider, *J. Acoust. Soc. Am.* **116**, 1656 (2004).
12. C. Bey, S. McAdams, *J. Exp. Psychol. Hum. Percept. Perform.* **29**, 267 (2003).
13. D. Deutsch, *J. Acoust. Soc. Am.* **57**, 1156 (1975).
14. D. Butler, *Percept. Psychophys.* **25**, 264 (1979).
15. W. M. Hartmann, D. Johnson, *Music Percept.* **9**, 155 (1991).
16. H. von Helmholtz, *On the Sensations of Tone as a Physiological Basis for the Theory of Music*, 2nd English ed., A. Ellis, trans., Dover, New York (1954).
17. R. Y. Litovsky et al., *J. Acoust. Soc. Am.* **106**, 1633 (1999).
18. W. A. Yost, D. Mapes-Riordan, S. J. Guzman, *J. Acoust. Soc. Am.* **101**, 2994 (1997). ■