THE PSYCHOLOGY OF HEARING

Sep 1, 1998 12:00 PM, Diana Deutsch

In everyday life, we are continuously bombarded with mixtures of sounds that arise in parallel from many different sources. A major task for our auditory system is to sort out the components of such mixtures so as to reconstruct the originating sound events. As I write this article, I can hear people talking outside my window, a dog barking and the rumbling of construction machinery in the distance. All these sounds are blended together as they reach my ears, and yet I hear each sound as a unified whole, distinct from the others. Somehow my auditory system groups those components of the sound spectrum that have emanated from the same source and separates out those that have emanated from different sources. What are the principles by which such grouping decisions are made?

Perceptual grouping principles Early in the century, the Gestalt psychologist Max Wertheimer proposed that we link elements of perceptual arrays in accordance with a number of simple principles [ref. 1]. One, which he termed proximity, states that we form connections between elements that are closer together in preference to those that are spaced further apart. Figure 1a shows an array of dots that are closer together along the vertical axis than the horizontal one. We group these dots on the basis of proximity, and as a result, we perceive a set of columns rather than rows. A second principle, which he termed similarity, states that we form connections among elements that are similar to each other in some way. In Figure 1b, we group the green dots and separate them from the purple ones with the result that we perceive a green triangle against a purple and white background.

A third principle, termed good continuation, asserts that we form connections among elements that continue smoothly in the same direction. In Figure 1c, we perceive the lines AB and CD, rather than AC and DB. Yet another principle, termed closure, states that we tend to perceive elements of an array as organized into complete units. For example, we interpret the pattern in Figure 1d as a circle that is partially occluded by a rectangle. A further principle, termed common fate, claims that elements that move in the same direction are perceptually linked together. Envisage, for example, four rows of dots, with two of the rows traveling from left to right and the other two traveling from right to left. We link these dots into two different groupings based on their direction of motion.

The perceptual system has presumably evolved to form groupings in accordance with such principles because they enable us to interpret our environment most effectively. Consider, for example, the principle of proximity. In the case of vision, elements that are close together in space are more likely to have emanated from the same object than elements that are further apart. In the case of hearing, sounds that are proximal in pitch or in time are more likely to have arisen from the same source than are sounds that are distant from each other along these dimensions. Analogous arguments can be made for similarity. Regions of the visual field that are similar in color, brightness or texture have probably emanated from the same object, and sounds that are similar in character (a series of thuds or chirps, for example) are likely to have arisen from the same source. A line that follows a smooth

Related Articles

- AMX Names New Chairman of the Board
- Extreme CCTV to Present at 2007 USBX Security Growth Conference
- Da-Lite Introduces Â“Angles of ReflectionÂ” Articles
- NuVision Appoints Geartner Associates to Represent Florida Territory
- HLCA Appoints David Bruce Vice-Chair

RESOURCES

Multimedia
Buyer's Guide
Event Calendar
Source Guide
Blogs
Edit Calendar
Newsletters
SVC Blogroll
Whitepapers
Reader Survey

MOST POPULAR ARTICLES

- Focus on Houses of Worship: Upgrading a Rock 'n'

have arisen from the same source. A line that follows a smooth pattern has probably emanated from a single object, and a sound that changes smoothly in pitch has probably come from a single source. A similar argument holds for common fate. A moving object gives rise to elements that travel across the visual field coherently with each other, and many musical instrument tones are composed of partials that rise and fall in synchrony.

Fusion and separation of a sound's spectral components We now consider the relationships between the components of a sound spectrum that cause us to fuse them into a single perceptual image and those that cause us to separate them into different images. Two types of relationships have been shown to be important here. One is harmonicity. Sounds such as those produced by musical instruments and the human voice are composed of partials that stand in harmonic (or near-harmonic) relationship; that is, their frequencies are integer (or near-integer) multiples of the fundamental. For example, a harmonic series whose fundamental is 100 Hz contains partials at 200 Hz, 300 Hz, 400 Hz, 500 Hz and 600 Hz, and so on. It makes sense, therefore, that the auditory system would make use of this feature so as to combine sound components that stand in harmonic relationship to form a single perceptual image. As an example, when presented with two instrument tones playing together, we perceive two distinct pitches, each resulting from one of the harmonic series that is present in the complex.

Another relationship that has been shown to be important here is syn-chronicity of onset. When components of a sound begin to sound simultaneously, it is a good bet that they have come from the same source; if they begin abruptly at different times, it is more likely that they came from different sources. A related issue concerns temporal regularities in the way the components of an ongoing sound fluctuate in frequency or amplitude.

Harmonicity Listening to sounds produced by different musical instruments provides us with many informal examples of grouping by harmonicity. For example, stringed instruments (such as the violin) and blown instruments (such as the flute) produce tones whose partials stand in harmonic (or near-harmonic) relationship. The pitches produced by such instruments are clearly defined. On the other hand, bells and gongs produce tones with nonhar-monic partials, and in listening to such instruments, we easily discern multiple pitches. Experiments using synthesized tones have confirmed this conclusion [ref. 2].

We can then ask: To what extent can a single component of a harmonic complex deviate from harmonicity and still contribute to the perceived pitch of the complex? It has been shown that when a harmonic of a complex tone is mistuned by less than about 3%, it still contributes fully to the tone's perceived pitch. As the degree of mis-tuning increases, however, its contribution to perceived pitch decreases, and at a mistuning of about 8%, the component no longer contributes to the pitch of the complex [ref 3]. Related findings have been obtained with respect to the contribution of a mistuned component to perceived vowel quality [ref. 4].

Another line of research has examined how well we can separate two complex sounds, as a function of the relationships between their fundamental frequencies. For example, as the fundamentals of two complex tones depart from simpl e harmonic relationship, the tones are heard more clearly as distinct entities [ref. 5]. As a related effect, simultaneous speech patterns can be more easily separated perceptually when built on different fundamentals-the amount of perceptual separation has been found to reach its maximum when the fundamentals differ by one to three semitones [ref. 6].

Onset synchronicity The temporal properties of sounds provide us with other cues for grouping. Components that arise from the same source are likely to begin to sound at the same time, and those arriving from different sources are less likely to do so. The brain makes use of such onset relationships in making grouping decisions. This can be demonstrated by presenting a harmonic series in such a way that its components begin at different times. Consider a series that is built on a 300 Hz fundamental. We can start with the 300 Hz component sounding alone, then after one second add the 600 Hz component, then after one more second add the 900 Hz component, and continue in this way until all the components are sounding simultaneously. The perceptual effect is striking. When each component begins to sound, its pitch is first heard distinctly, then it gradually disappears from perception so that finally only the pitch corresponding to the fundamental is perceived.

When two complex tones are played together, they are heard as perceptually more distinct from each other when they begin to sound at different times. An onset difference as small as 10 ms has been found to increase the perceptual salience of the tones in the mixture, and an onset difference of 30 ms has a pronounced effect [ref. 5]. Using recordings of ensemble performances, it was found that values of onset asynchrony for tones that were nominally synchronous ranged from 30 ms to 50 ms-a value that would be expected to be useful in enhancing the perceptual salience of individual voices in a mixture [ref. 7].

Frequency modulation During vibrato, the partials of a complex tone move up and down in synchrony with each other in such a way as to preserve the ratios formed by the different frequencies. One might expect the perceptual system to exploit this feature in determining which components of a sound mixture to link.

The composer John Chowning experimented with this issue in the process of synthesizing a singing voice by computer [ref. 8]. He found that in order to produce the impression of a sung vowel, it was necessary to impose a coherent frequency fluctuation on all the components simultaneously. Chowning then synthesized three simultaneous sung vowels, the first singing "oh" with a fundamental of 400 Hz, the second singing "ah" at 500 Hz, and the third singing "eh" at 600 Hz. When there was no frequency fluctuation, the mixture was heard as a chord consisting of three pitches. However, when the three sets of partials were differentiated from each other by superimposing different patterns of frequency fluctuation on each one, then three sung vowels were clearly heard, each at a different pitch. However, later experiments have shown that the effects of coordinated frequency modulation on perceptual grouping are complex ones, and many issues remain unresolved [ref. 9].

Many sounds are composed of partials whose amplitudes rise and fall in synchrony with each other, so one might conjecture that coherent amplitude modulation would also be used by the auditory system as a cue for perceptual fusion. However, clear evidence in support of this conjecture has been difficult to obtain [ref. 9].

Ear of input When two different sound components are presented simultaneously, one to each ear, one might at first expect that this difference in ear of input would provide a strong cue for separating the sounds perceptually. Upon reflection, the situation is not that simple. In natural environments, sounds are subjected to numerous distortions as they travel from their sources to each of our ears. Given such distortions, if we were to place heavy reliance on ear differences as cues for perceptual separation, we would risk separating components when they should instead be grouped.

Indeed, the auditory system shows a striking tendency to disregard ear of input as a cue for separating out the components of a complex sound, at least when other supporting cues are absent. In one experiment, listeners identified the pitches of two complex tones when their partials were distributed across the ears in various ways. Pitch identification was only weakly affected by the ways in which the partials were distributed [ref. 10]. Another experiment examined the effect of ear of input on the perception of speech sounds. The first two formants of a phrase were presented, one to each ear. When the two formants were built on the same fundamental, listeners could identify the speech signal and also tended to hear a single voice-that is, they fused the input from the two ears into a single perceptual image [ref. 11].

The fact that spatial location may be disregarded in favor of other cues can be used to produce striking illusions [ref. 12]. These occur when two different streams of tones are presented, one to each ear (or one to each of two spatially separated loudspeakers). The scale illusion and its variants are produced by simultaneously presented ascending and descending scales. These are made to switch from ear to ear such that when a tone from the ascending scale is in one ear, a tone from the descending scale is in the other. In consequence, each ear is presented with a set of tones that leap around in pitch. However, the pattern is not heard this way. Rather, two melodic lines are perceived, a higher one and a lower one, that move in contrary motion. Further, the higher tones are often heard as though coming from one spatial location and the lower tones from the other. Figure 2 shows an example of this illusion, in this case produced by a two-octave chromatic scale [ref. 13].

A different type of spatial reorganization occurs in the glissando illusion [ref. 13]. This is created by an oboe tone that is repeatedly presented together with a sine wave that glides slowly up and down in pitch. The oboe tone and glissando are made to switch from ear to ear (or from loudspeaker to loudspeaker) in such a way that when the oboe tone is to the right, a portion of the glissando is to the left, and vice versa. Most people hear the oboe tone correctly as switching back and forth between locations, but the glissando appears to be joined together quite seamlessly. Sometimes the glissando appears to be consistently in one spatial location, and sometimes, it appears to travel from one side of space to the other as its pitch goes from low to high and then back again as its pitch goes from high to low.

Continuity illusions Information often arrives at our sense organs in fragmented form, and our perceptual system has the task of inferring continuities between the fragments and filling in the gaps appropriately. For example, we generally see branches of trees when they are partly hidden by foliage, and we infer which of the visible segments were derived from the same branch. When we make such inferences, we are employing the principles of good continuation and of closure because we mentally fill in the gaps between the segments of a branch so as to produce a smooth contour.

In the same way, our hearing mechanism constantly "fills in" lost fragments of sound so as to make sense of the world. For example, when two people are conversing near a busy street, they must perceptually restore fragments of speech that are being drowned out by passing traffic. Such perceptual restorations can give rise to intriguing illusions. For example, if a softer sound is briefly replaced by a louder one; this sometimes produces the impression that the softer sound is present without interruption. In an early experiment, a sequence was constructed that consisted of a tone in alternation with a louder noise, with each sound lasting 50 ms. On listening to this sequence, subjects heard the tone as continuing right through the noise [ref. 14].

Continuity effects have also been produced using more complex sounds. In one experiment, a gliding tone was presented that rose and fell repeatedly, and the glide was periodically interrupted and replaced by a loud noise. Listeners did not hear the glide as fragmented, but rather as continuous [ref. 15]. Similar effects have been produced with speech sounds. When sentences were presented with a portion of each sentence deleted and replaced by a louder noise such as cough, subjects heard these sentences as though they were intact [ref. 16]. In an experiment using musical materials, recordings were made of well known piano pieces, and some of the tones were omitted and replaced by noise bursts. Again, listeners heard the pieces as though they were intact [ref. 17]. Perceptual restorations of this type must occur frequently when we listen to music in concert halls where coughs and other loud noises would otherwise cause the music to appear fragmented.

Grouping of rapid sequences of sounds Grouping by pitch proximity emerges strongly in sequences of tones that are presented in rapid succession. Composers frequently exploit this phenomenon in the technique of pseudopo-lyphony, or compound melodic line. Here, a series of tones is played at a fast tempo, and the tones are drawn from different pitch ranges; as a result, listeners hear two or more melodic lines in parallel, each in a different pitch range. Such passages occur frequently in twentieth-century guitar music. In the example shown in Figure 3, taken from Tarrega's Recuerdos de la Alhambra, the lower tones form the melodic line while the repeating higher tones form a background, which is heard separately from the melody.

One consequence of such perceptual splitting of tones into different streams is that temporal relationships across streams become difficult to judge. In one experiment, a repeating series of six tones was presented, three from a high pitch range and three from a low one. When the rate of presentation was as fast as 10 per second, subjects were unable to judge the orders in which tones in the different streams occurred [ref. 18]. At slower rates of presentation, there is still a gradual breakdown of temporal resolution as the presentation rate increases, and also as the pitch distance between successive tones increases [ref. 19].

Another factor that influences perceptual grouping is sound quality, or musical timbre-an example of grouping by similarity. Composers often place different instrument tones in overlapping pitch ranges, recognizing that listeners will group these tones on the basis of instrument type provided that the difference in timbre is sufficiently large. Many examples can be found in Schubert's songs in which the piano accompaniment often overlaps in pitch with the singer's voice, yet the pitch patterns that are heard correspond to those produced by each instrument separately.

In a demonstration of this effect, a three-tone ascending pitch line was repeatedly presented in which successive tones were made up of alternating timbres. When the difference in timbre was small, listeners heard the ascending pitch lines, as expected. However, when this difference was large, they instead grouped the tones on the basis of timbre, and so perceived two, interwoven descending pitch lines [ref. 20].

Just as with grouping by pitch range, so grouping by timbre can affect a listener's ability to make order judgments concerning sequentially presented sounds. In one experiment, listeners were presented with a repeating sequence consisting of four unrelated sounds (a high tone, a low tone, a hiss and a buzz). When this sequence was played at a sufficiently fast rate, listeners were unable to name the order in which the sounds appeared [ref. 21].

Conclusion In the past, approaches to sound perception have tended to focus on low-level factors, such as thresholds for pitch and loudness and masking functions. Recently, however, the importance of higher-level, cognitive factors has become increasingly evident, and there is growing recognition that the auditory system of the brain contains some remarkably ingenious circuitry-perhaps the most ingenious of all the sensory modalities. The phenomena we have been examining illustrate the operation of some of this circuitry, which has evolved to enable us to interpret our sound environment most effectively [ref. 22].

1. Wertheimer, M. Untersuchung zur Lehre von der Gestalt II. Psychologische Forschung, 1923, 4, 30l-350.

2. De Boer, E. On the 'residue' and auditory pitch perception. In W. D. Keidel & W. D. Neff (Eds.) Handbook of sensory physiology. Vol 5, part 3, 479-583.New York: Springer-Verlag, 1976.

3. Moore, B. C. J., Glasberg, B. R., & Peters, R. W. Thresholds for hearing mistuned partials as separate tones in harmonic complexes. Journal of tbe Acoustical Society of America, 1986, 80, 479-483.

4. Darwin, C. J. & Gardner, R. B. Mistuning a harmonic of a vowel: Grouping and phase effects on vowel quality. Journal of the Acoustical Society of America 1986, 79, 838-845.

5. Rasch, R. A. The perception of simultaneous notes such as in polyphonic music. Acustica, 1978, 40, 1-72.

6. Assman, P. F. & Summerfield,A Q.Modelling the perception of concurrent vowels: Vowels with different fundamental frequencies. Journal of the Acoustical Society of America, 1990, 88, 680-697.

Brokx, J. P. L. & Nootebohm, S. G. Intonation and the perceptual separation of simultaneus voices. Journal of Phonetics, 1982, 10, 23-36. Sheffers, M. T. M. Sifting vowels: auditory pitch analysis and sound segregation. Doctoral thesis, Groningen University, The Netherlands. 1983.

7. Rasch, R. A. Timing and synchronization in ensemble performance. In J. A. Sloboda (Ed.) Generative processes in music: The psychology of performance, improvization, and composition. Oxford: Oxford University Press, 1988.

8. Chowning, J. M. Computer synthesis of the singing voice. In Sound Generation in Winds, Strings, Computers, Stockholm: Royal Swedish Academy of Music Stockholm, Publ. No. 29., 1980. 4-13.

9. Darwin, C. J. & Carlyon, R. P. Auditory grouping. In B. C. J. Moore (Ed.) Hearing. San Diego, Academic Press, 1995, 387-424.

10. Beerends, J. G. & Houtsma, A. J. M. Pitch identification of simultaneous dichotic two-tone complexes.Journal of the Acoustical Society of America, 1989, 85, 813-819.

11. Broadbent, D. E. & Ladefoged, P. On the fusion of sounds reaching the different sense organs. Journal of the Acoustical Society of America, 1957, 29, 708-710.

12. Deutsch, D. Two-channel listening to musical scales. Journal of the Acoustical Society of America, 1975, 57, 1156-1160; Deutsch, D. Auditory illusions, handedness, and the spatial environment. Journal of the Audio Engineering Society, 1983, 31, 607-618.

13. Deutsch, D. Musical illusions and paradoxes.Philomel Records, P.O.Box 12189-2189, La Jolla, CA 1995.(CD)

14. Miller, G. A., & Licklider, J. C. R. The intelligibility of interrupted speech. Journal of the Acoustical Society of America, 1950, 22, 167-173.

15. Dannenbring, G. L. Perceived auditory continuity with alternately rising and falling frequency transitions. Canadian Journal of Psychology, 1976, 30, 99-114.

16. Warren, R. M. Auditory illusions and their relation to mechanisms normally enhancing accuracy of perception. Journal of the Audio Engineering Society 1983, 31, 623-629.

17. Sasaki, T. Sound restoration and temporal localization of noise in speech and music sounds. Tohuku Pschologica Folia, 1980, 39, 79-88.

18. Bregman, A. S., & Campbell, J. Primary auditory stream segregation and perception of order in rapid sequences of tones. Journal of Experimental Psychology, 1971, 89, 244-249.

19. Van Noorden, L. P. A. S. Temporal Coherence in the Perception of Tone Sequences. Unpublished doctoral dissertation. Technische Hogeschoel Eindhoven, The Netherlands, 1975.

20. Wessel, D. L. Timbre space as a musical control structure. Computer Music Journal, 1979, 3, 45-52.

21. Warren, R. M., Obusek, C. J., Farmer, R. M., & Warren, R. P. Auditory sequence: Confusions of patterns other than speech or music. Science, 1969, 164, 586587.

22. For further reading,see Bregman, A. S. Auditory scene analysis: The perceptual organization of sound. Cambridge: MIT press, 1990; Darwin, C. J. & Carlyon, R. P. Auditory grouping. In B. C. J. Moore (Ed.) Hearing. San Diego, Academic Press, 1995, 387-424; and Deutsch, D. Grouping mechanisms in music. In D. Deutsch (Ed.) The psychology of music (2nd edition) San Diego, Academic Press, 299-348, in press.

- Logout from DISQUS

- Logged in as
- using Facebook Connect (Logout)

Type your comment here.

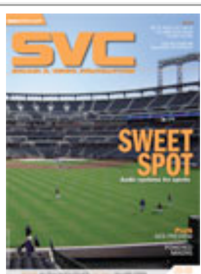
Community Page   Subscribe by email

blog comments powered by Disqus

Want to use this article? Click here for options!

Browse Back Issues   Select an Issue

BROWSE ISSUES



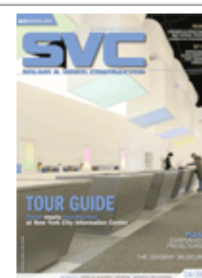
September 2009

August 2009

July 2009

June 2009

May 2009



April 2009