Picture This

The sounds of speech lead to novel ways of representing complex data

By IVARS PETERSON

From the faintest whisper to the loudest cry, the human voice has an extraordinary range. It expresses pleasure and pain in innumerable accents and utterances. This remarkable variability, however, makes the analysis of human sounds exceedingly difficult.

Simply recording the pressure variations over time in the acoustic signal generated by human speech produces a complex waveform. A typical signal includes the smoothly corrugated look of vowel-like sounds and the jagged peaks associated with random noise. Significant silences periodically punctuate the waveform. Scattered throughout the signal lie the characteristic signatures of speech sounds known as fricatives and plosives.

Because the human ear alone is not a sufficiently reliable tool for rigorously comparing and characterizing sounds, researchers have been developing methods for seeking out the patterns hidden within speech waveforms. Traditional spectrograms, in which sound intensity is plotted against frequency over time, are valuable for showing the general frequency content of a signal. But sometimes both trained and untrained users have difficulty detecting acoustic differences, clearly evident to the ear, in the corresponding spectrograms.

Such difficulties motivated Clifford A. Pickover of the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y., to explore several novel methods for representing speech sounds. In some cases, his techniques made subtle differences in acoustic signals considerably more obvious to a human analyst. Furthermore, he found that his methods could be extended to other types of data analysis, including the search for patterns in the sequence of nucleotides along a strand of DNA.

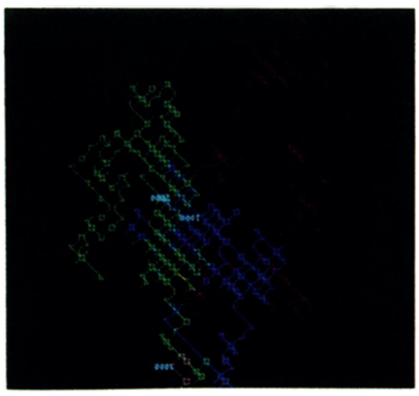
Pickover's research is an instance of the growing interest in finding ways of coping with the vast quantities of data generated by modern instruments and other computer-aided techniques. Particular attention is being paid to identifying patterns or trends. In the past, such efforts have included the use of sound to represent data (SN: 6/1/85, p.348), the use of video and animation, and the use of various computer graphics techniques for manipulating and plotting data.

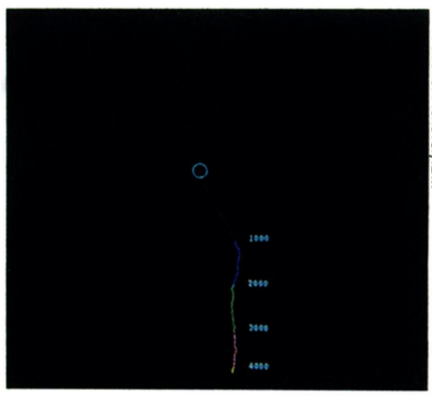
Pickover's innovations are a mixture of old ideas used in new contexts and new ideas brought to bear on longstanding data-analysis problems. A summary of much of his work appears in the IBM research report "Computers, Pattern, Chaos and Beauty." Details of particular techniques are available in other IBM reports. Last week, Pickover presented a "road map" of his "graphics perspectives" at a conference held in New York City on Computer Graphics in the Arts and Sciences.

One of Pickover's most striking and colorful data-display techniques produces figures that have the sixfold symmetry of a snowflake. Although formally termed "symmetrized dot patterns," these forms could just as well be called "speech-flakes."

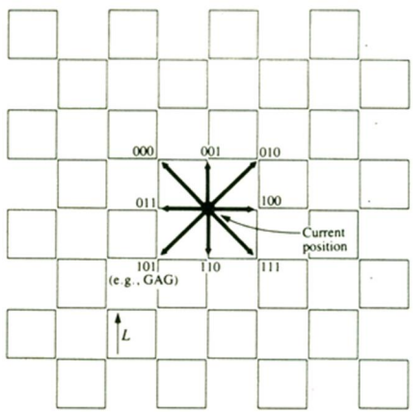
To create a symmetrized dot pattern for a particular sound, Pickover selects a 500-millisecond segment of the corresponding amplitude-vs.-time waveform. He records the amplitudes at 5,000 points along the waveform (every 10,000th of a second). These data are mapped into a snowflake-like pattern.

Suppose the first amplitude is 40 (on a scale from 0 to 50), the second sampled point has an amplitude of 30, and the third has a value of 35. These values are plotted on a graph that looks like a polar view of the earth, with the North Pole at the graph's center. The first number represents an angle, 40°, and the second number the distance, 30 units, from the pole (or origin). That combination marks a spot on the graph. Then the second number becomes the angle and the third represents the distance (30°, 35 units), marking a second spot. This process is repeated until all the amplitude values are accounted for. To create the snowflake-like pattern, the resulting array of dots, which fills roughly one-sixth of the full graph, is reflected as in a kalei-





Illustrations: Pickover/IBM



IBM JOURNAL OF RESEARCH & DEVELOPMENT

A DNA vectorgram converts a long sequence of nucleotide bases into a walk on a square lattice. Each of four bases is given a value of either 0 or 1 to generate a string of binary digits. Taken in groups of three, these digits are assigned specific directions on the lattice (lower left). A random nucleotide-base sequence produces a vectorgram like the one shown above and to the left. The color changes every 1,000 bases. In contrast, a human bladder cancer gene has a different, distinctive nucleotide pattern (above right).

Science Service, Inc. is collaborating with JSTOR to digitize, preserve, and extend access to *Science News*. www.jstor.org

doscope. The addition of color enhances the patterns.

"This algorithm works well for certain sounds," says Pickover. For example, it easily distinguishes the "a" in "father" from the "o" in "mom." The method is particularly sensitive to differences in frequency, producing different curvatures. Further study of the method is needed to determine what the optimal sampling rate and symmetry ought to be for various applications.

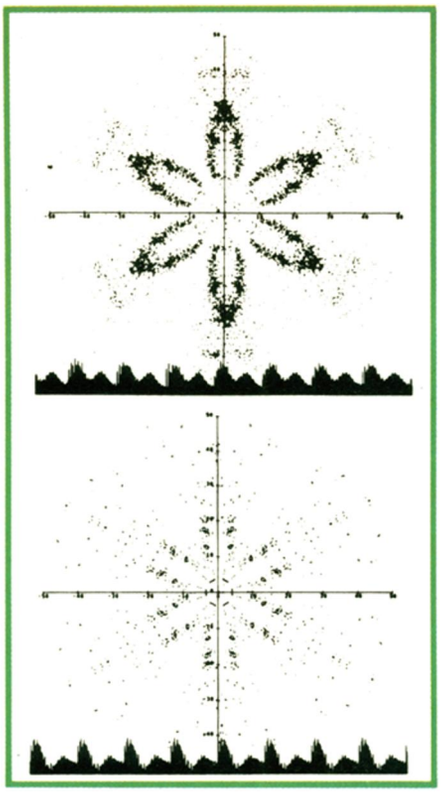
Says Pickover, "Intriguing as an art form, these dot patterns may be a way of visually fingerprinting natural and synthetic speech sounds and allowing researchers to detect patterns in data not easily captured with traditional analyses."

Pickover has tested his algorithm on a variety of sounds, from human and synthetic speech to the full-throated croaking of frogs and the shrill whistles of dolphins. One researcher has already expressed interest in using the technique for studying bird songs. Pickover himself has shown that symmetrized dot patterns may be useful for detecting and characterizing heart abnormalities from the sounds generated by the heart.

Pickover's vectorgrams sometimes look like the steps a drunkard would take in wandering across an open area. They can also be used to search for patterns in the sequences of bases in DNA. Essentially, Pickover translates the location of particular bases into movement on a two-dimensional lattice or, more literally, the movement of a pen across a piece of paper. His technique is an application of previous work by other researchers on tracking bit sequences in digital computers and it follows his own efforts to represent the base sequence of a cancer gene's DNA as a waveform.

Pickover assigns the value one to the bases cytosine (C) and guanine (G) and the value zero to the bases adenine (A) and thymine (T). A typical strand of DNA, consisting of 4,000 bases, becomes a long string of ones and zeros. Then one of eight possible directions is assigned to each different set of three consecutive digits. For example, from a starting position on a square grid, the string 001 indicates a movement directly upward to an adjacent corner, while the string 101 corresponds to a movement diagonally downward and to the left.

In this way, a string of ones and zeros is converted into a zigzagging path across a checkerboard. The pattern seen depends on the base sequence found in the DNA strand. In such vectorgrams, sequences that have a high guanine and cytosine content tend to move downward. A repeating sequence, such as . . . GGGGAAGAATACGAGGGGAA . . . , generates a trace that returns to its starting point. Various characteristic short sequences, found within DNA strands and which



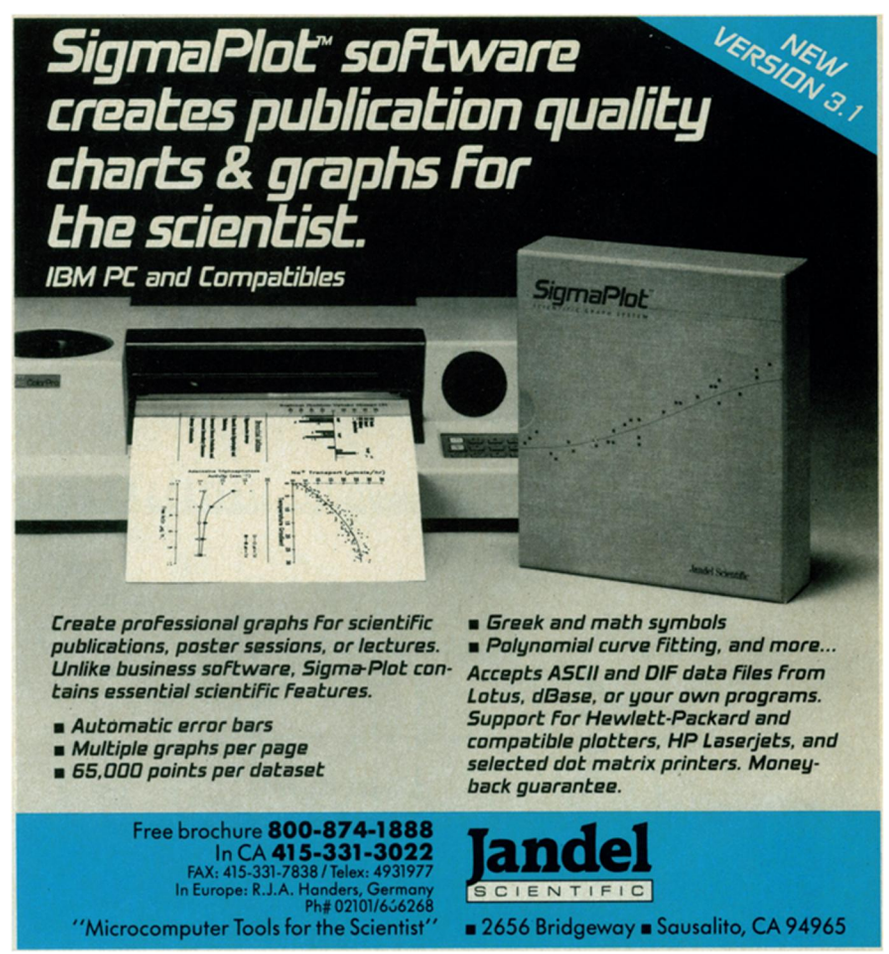
Despite similar waveforms, symmetrized dot patterns clearly show the difference between a human-made "ee" sound (upper diagram) and a synthesized version of the same sound (lower).

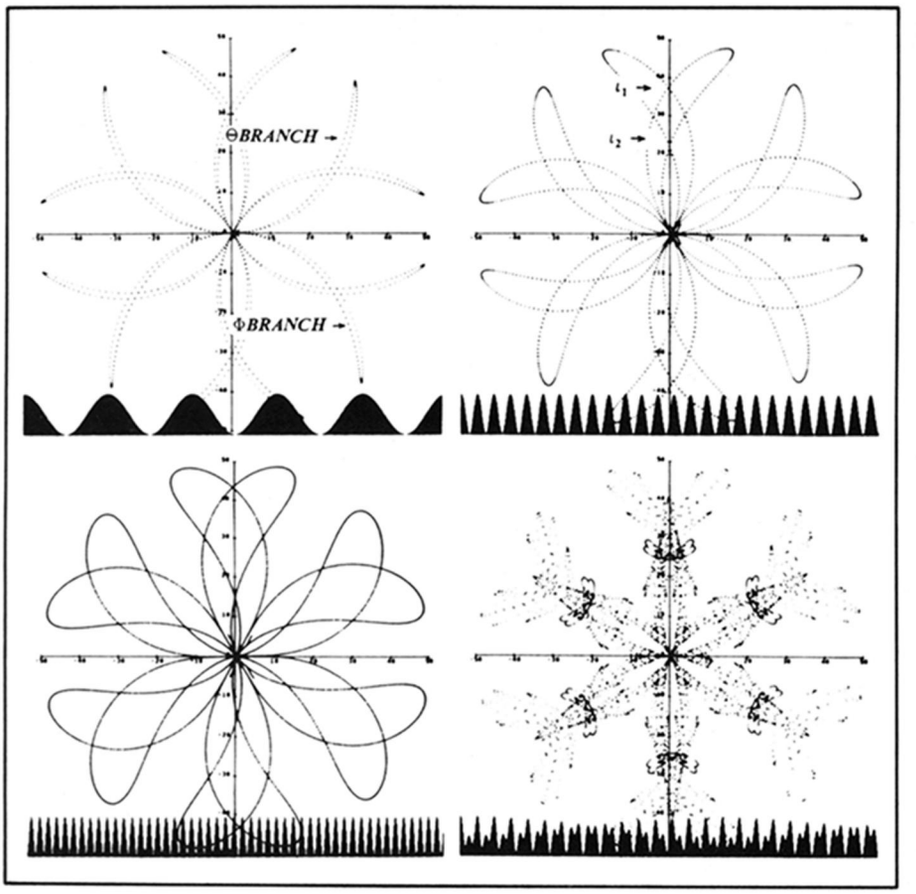
appear to control particular DNA properties, turn out to have distinctive signature patterns.

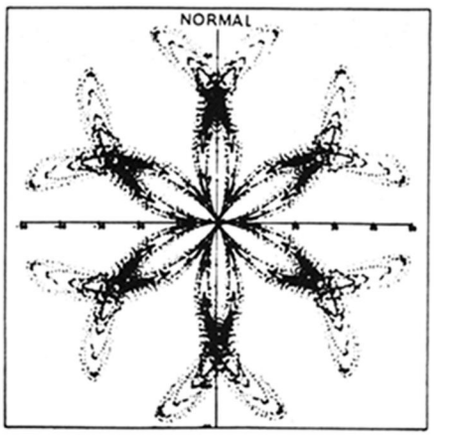
"This permits the human observer," says Pickover, "to visually detect some important sequence structural properties and patterns not easily captured by traditional means." However, the method, by failing to distinguish between sequences such as ATATAT . . . and AAAAAA . . . , which have different folding and bending properties, can sometimes obscure patterns.

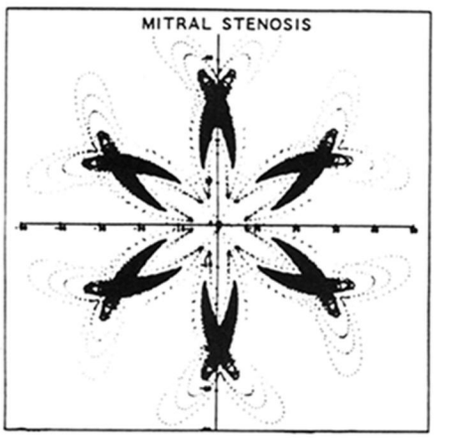
More work is needed on identifying which DNA parameters should be plotted and how those characteristics ought to be represented. Better results may be obtained by using a hexagonal grid rather than a square grid, by inspecting pairs or quadruplets of digits instead of triplets, and by assigning values to bases in different ways.

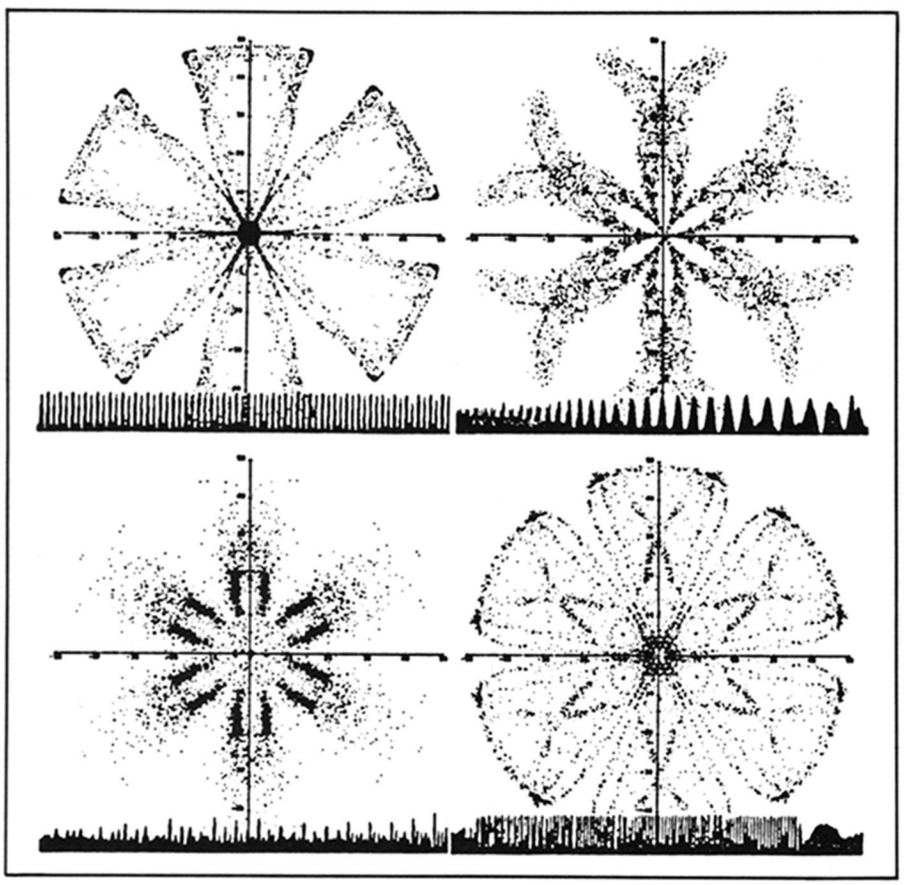
"The exploration of this large parameter space provides a provocative area for future research," Pickover writes in the January IBM JOURNAL OF RESEARCH AND DEVELOPMENT. "It may be possible to discover interesting properties and periodicities in the DNA sequence by having the [computer] program produce many vectorgrams by automatically iterating through a large number of input parameters and mappings. In this way, the program may suggest to the human analyst important features and parameters

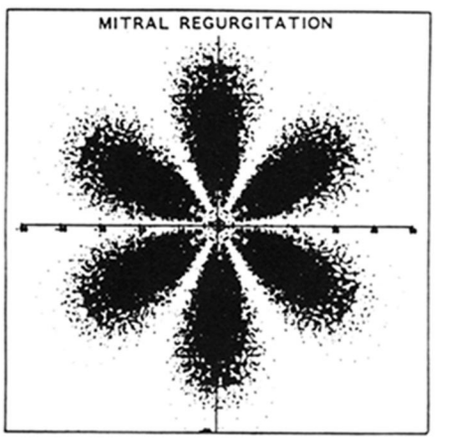












Symmetrized dot patterns computed from acoustic waveforms are sensitive to frequency variations. The first set of four diagrams (upper left) demonstrates how the patterns depend on input frequency. The actual waveform is shown at the bottom of each diagram. The lower set of four diagrams shows the differences evident in the sounds of (clockwise, starting at upper left) a rooster, a dolphin, a frog and a cat. A normal heart sound (above, top) can be contrasted with cardiac sounds associated with various pathological conditions (middle and bottom).

which would not even be considered otherwise."

Cartoon faces – some grinning, some sad, some cryptically blank – certainly attract attention. They can also be used to represent complicated data. Introduced in 1973 by statistician Herman Chernoff of Harvard University, the technique has piqued the interest of many data analysts. Using the characteristics of various facial features, such as the face's shape or the mouth's curvature, a single face can convey the value of 10 different variables at the same time. Its effectiveness depends on the human ability to integrate facial features into a meaningful image.

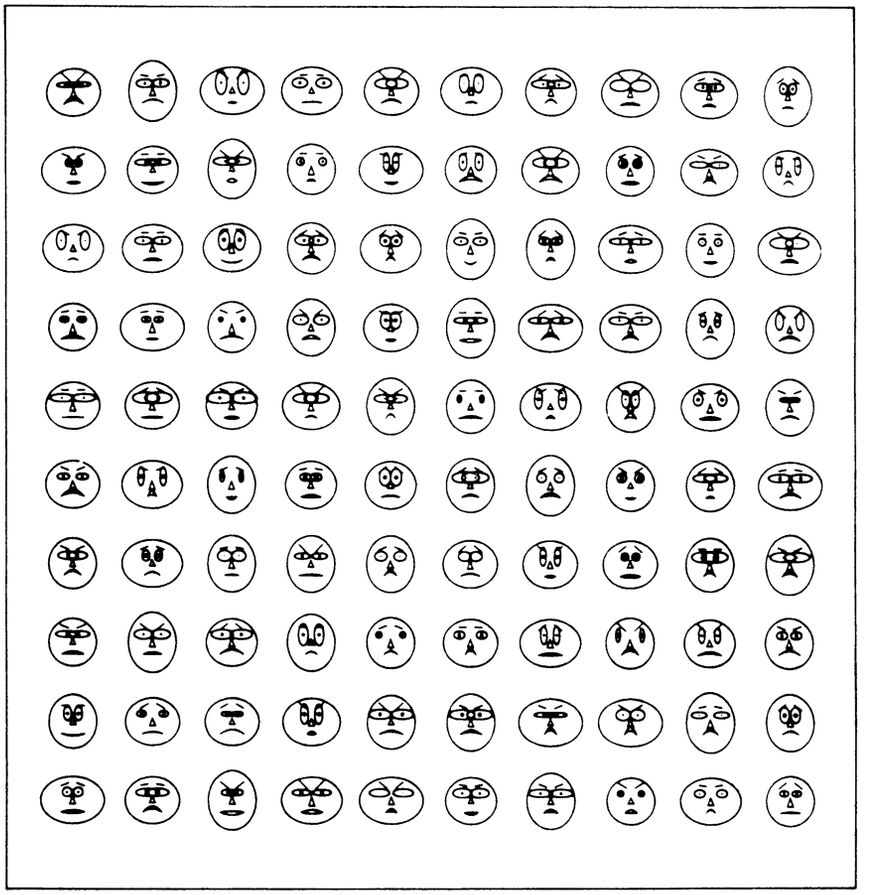
"Such faces have been shown to be more reliable and more memorable than other tested icons," says Pickover, "and allow the human analyst to grasp many of the essential regularities and irregularities in the data."

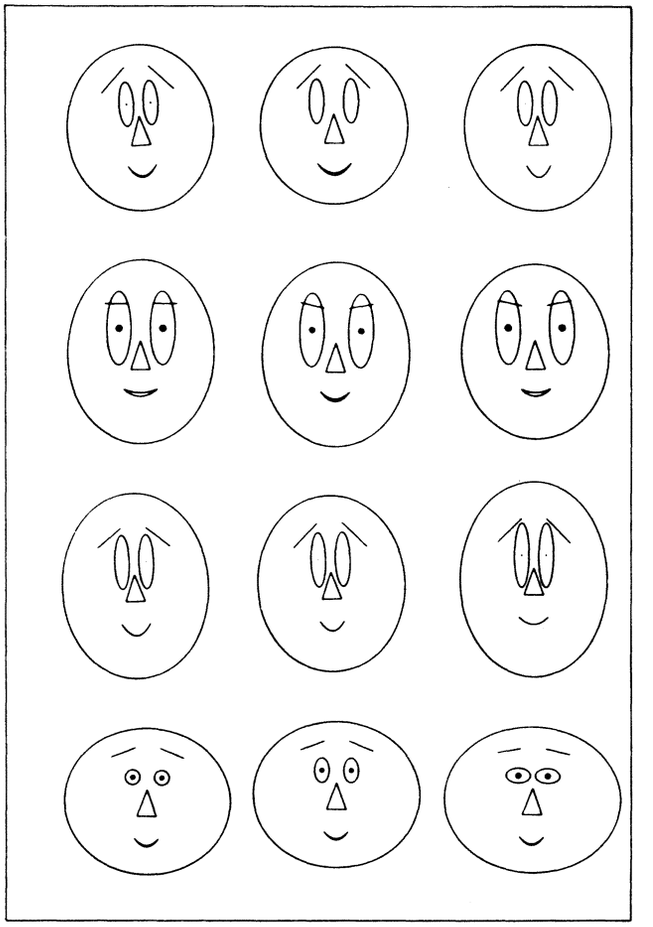
Pickover has used cartoon faces to characterize broad classes of sounds. Different types of sound, when analyzed in this way, appear to generate distinctive types of faces. Such "speech-faces," says Pickover, may be a useful aid in teaching deaf or near-deaf children how to modify the sounds they make. Although the expressions on the cartoon faces have little to do with facial movements that a person makes when speaking, the cartoon faces provide a vivid feedback target. Students, on seeing a particular face after uttering a sound, would have a better sense of how close they are to the required sound.

IBM researchers have also proposed the use of cartoon faces on control panels. Pilots in military aircraft, for example, are sometimes overwhelmed by the number of dials and indicators they must monitor to keep an airplane safely in the air. Displaying several key pieces of information in one place is a simple way to reduce this data overload. By learning to recognize certain faces as danger signals, pilots would be able to react more quickly. Cartoon faces also bring together signals that by themselves may not indicate a threatening situation but taken together reveal a potential hazard.

Symmetrized dot patterns, cartoon faces and DNA vectorgrams represent only a few of the display techniques that Pickover has explored. "For most of these things," he says, "there really hasn't been a lot of work done on them." Many techniques look promising, but the conditions under which they work best are not yet known.

Eventually, researchers sitting at their computer work stations will be able to call on a broad palette of data-analysis tools. Ironically, the availability in the future of many different ways of displaying data may itself contribute to the data-overload problem. □





Cartoon faces can be used to represent the values of as many as 10 variables, each variable corresponding to a facial feature. In the upper illustrations, a computer-generated array of faces, in which facial parameters were computed using a random number generator, demonstrates the diversity of faces available for data representation. Pickover has used cartoon faces to characterize sounds (right). The top row represents the fricative sound "s"; the second, a "sh"; the third, a "z"; and the fourth, a "v".