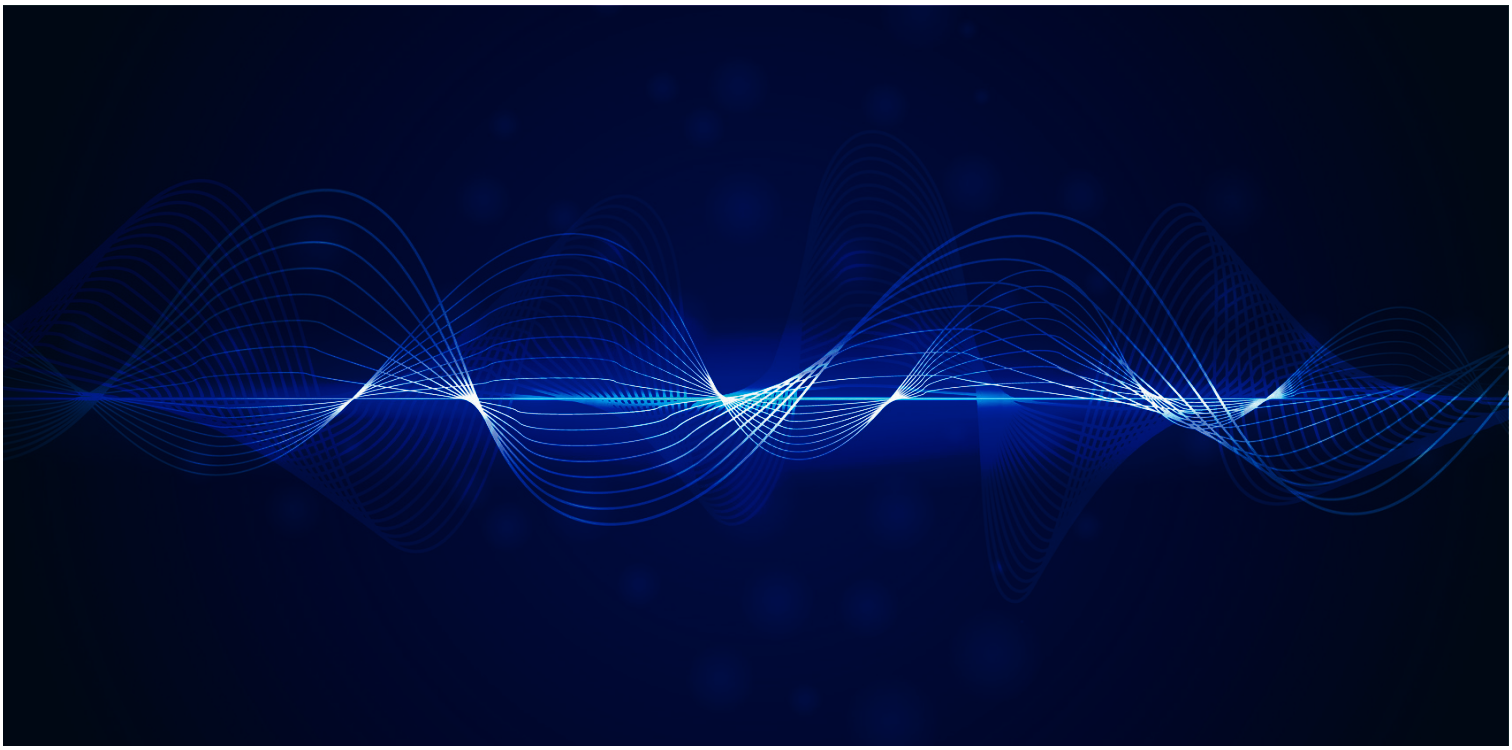

From a Cry in the Dark to the Forensic Voiceprint

BY DAVID KALAT

With the aggressive pace of technological change and the onslaught of news regarding data breaches, cyber-attacks, and technological threats to privacy and security, it is easy to assume these are fundamentally new threats. The pace of technological change is slower than it feels, and many seemingly new categories of threats have been with us longer than we remember.

Nervous System *is a monthly series that approaches issues of data privacy and cyber security from the context of history—to look to the past for clues about how to interpret the present and prepare for the future.*

In 1932, the kidnapping and murder of the infant son of Charles Lindbergh and his wife, Anne, gripped the nation. During a secret payoff at a cemetery, Lindbergh heard the voice of a man claiming to be the kidnapper. Almost three years later, Lindbergh testified at the trial of Bruno Hauptmann that he recognized Hauptmann's voice from the cemetery.

Researcher Frances McGehee was fascinated by the spectacle, and the obvious challenges of accurate speaker recognition, after so much passage of time and under such emotional pressure. She set out to measure the problem scientifically, and thereby opened an enduring debate regarding the forensic validity of voice recognition.

McGehee's 1936 doctoral dissertation, "The Reliability of the Identification of the Human Voice," explored the question of voice recognition scientifically. Her experiments exposed groups of subjects to the voice of an unfamiliar speaker and then tested the groups' ability to recognize that speaker at various time intervals later. Her test subjects had a high degree of accurate recall and voice recognition after one day, but the ability to correctly identify an unfamiliar speaker dropped off precipitously over time. In other words, the natural human ability to recognize voices is imperfect and prone to deterioration. Crafting an improved method of speaker recognition would require something more.

The next major figure in the development of speaker recognition technology was Bell Laboratories physicist Lawrence Kersta. At the behest of law enforcement, he had been investigating whether persons could be identified, scientifically and reliably, by their voices (the question had arisen in the wake of various anonymous bomb threats made over the telephone).

Kersta's 1962 article "Voiceprint Identification" coined the term "voiceprint" and detailed a methodology that Kersta had been developing since 1943. Although derided by some as "more Dick Tracy than scientific," Kersta's notion of voiceprints proposed identifying individual people from the way their unique biological characteristics manifested as patterns in the

sounds they made with their vocal tracts. Kersta's method used a spectrogram to graph the frequency and intensity of an audio signal visually over time; a trained human expert then examined that visual record to look for patterns. Problematically, though, the high degree of accuracy that Kersta enjoyed in controlled laboratory settings proved to be elusive when the technique was attempted in the field. In the wake of the academic debate over the scientific basis of the Kersta voiceprint and the inability of other researchers to replicate Kersta's success, the National Academy of Sciences conducted its own tests and concluded that Kersta's voiceprints were too uncertain to be reliable in any legal, forensic application.

Despite the evocative nature of the term, there is no scientific consensus on what constitutes a "voiceprint," nor on whether such a thing can reliably identify unique speakers. Instead, a range of competing techniques and technologies exists, with distinct academic debates and divisions among them.

One division is between "spectrum analysis" and "cepstrum analysis." Kersta's approach was a form of spectrum analysis, mapping how acoustic energy is distributed across frequencies. Subsequent researchers refined Kersta's ideas by automating the analysis using computers. This line of research itself branched into two general approaches: mapping the energy distribution in the actual sound (using a mathematical technique called "fast Fourier transform") versus estimating the vocal tract filter that shaped the sound (using a mathematical technique called "linear prediction").

By contrast, cepstrum analysis decomposes the components of the sound mathematically into what can be characterized as an anagram of the spectral analysis. The term "cepstrum" is simply an anagram of "spectrum," and that wordplay is meant to signify the relationship between the two approaches.

Spoken human language is the product of two factors working in combination. The vocal chords act as a sound source, generating acoustic energy that is then filtered and shaped by the vocal tract. The individual sounds that make up spoken human language are phonemes that arise from different combinations of one or more sources with different filters (such as tongue placement)—or, put another way, different combinations of sources and filters result in different phonemes. Consequently, research into how to decompose spoken speech into its constituent phonemes is another side of the same coin as research into how to synthesize artificial speech from individual phonemes.

Swedish researcher Gunnar Fant emerged as the next major researcher of voice recognition, with critical discoveries in the 1950s and 1960s. Fant has been celebrated as one of the most influential and pivotal figures in speech science for his work on modeling the interaction, or "convolution," of the vocal chord source and vocal tract filter. From Fant's source-filter model, researchers developed ways of representing the effects of the vocal tract filter as a linear equation. Such equations then could be used to model the production of speech phonemes. Fant's landmark studies about how speech is physically constructed remain standard references for the field of speech processing.

Modern voice recognition technologies build on the pioneering work of these and other important researchers. The technologies are as diverse as the varied approaches of the scientists who inspired them.

As for the case that inspired this research in the first place, Charles Lindbergh testified before a grand jury, some two-and-a-half years after his furtive encounter with the ransomer, that "[i]t would be very difficult for me to sit here today and say I could pick a man by that voice." Nevertheless, prosecutors brought Lindbergh to the district attorney's office the following day and asked him to do just that. Lindbergh said he recognized the German-accented voice of immigrant carpenter Bruno Hauptmann. Based largely on this highly subjective earwitness testimony, Hauptmann was found guilty of the kidnapping and murder of Lindbergh's baby. Hauptmann was electrocuted in 1936.

The views and opinions expressed in this article are those of the author and do not necessarily reflect the opinions, position, or policy of Berkeley Research Group, LLC or its other employees and affiliates.

Legaltech® news