

Suppressing nonstationary noise in mobile handsets

MOBILE CARRIERS ARE INTIMATELY AWARE OF THE ROLE THAT VOICE QUALITY PLAYS IN CUSTOMER RETENTION. ONE OF THE PRIMARY FACTORS AFFECTING VOICE QUALITY IS ENVIRONMENTAL NOISE, SO ANY MEANS OF SUPPRESSING NOISE PROVIDES A POTENTIAL DIFFERENTIATOR FOR HANDSET MANUFACTURERS.

Until recently, noise-suppression technology focused on reducing slow-changing stationary noise sources. However, nonstationary noise sources are fast-changing, and the current technology does not suppress them. As a result, subscribers cannot reliably use their handsets on busy streets, in crowded restaurants, or even at home.

Suppressing nonstationary noise brings substantial benefits to both subscribers and carriers. Users gain the freedom to speak and hear clearly wherever and whenever they want, enjoy increased privacy by being able to speak softly in noisy environments, and need not leave important conference calls. Carriers will see a reduction in customer churn, increased air-time usage, more efficient use of network bandwidth, and significant savings of capital and operational expenses.

You can readily recognize stationary noise, such as a loud fan in the background, because of its relatively constant nature, and you can effectively subtract this noise through conventional signal-processing techniques (Figure 1). Nonstationary noise, in contrast, involves rapid or random change, such as a person talking, background music, or keyboard typing (Figure 2). By the time you recognize nonstationary noise as noise, it has already passed, so it requires more sophisticated noise-suppression techniques.

MULTIPLE-MICROPHONE NOISE SUPPRESSION

Next-generation noise-suppression techniques, such as ASA (auditory scene analysis), beam forming, and BSS (blind source separation), use multiple microphones to more accurately identify, locate, and suppress noise sources than is possible with a single microphone. ASA uses psychoacoustic grouping principles to separate noise sources from the voice of interest. ASA's developers based the technology on the human auditory pathway; ASA processes noise in the same way

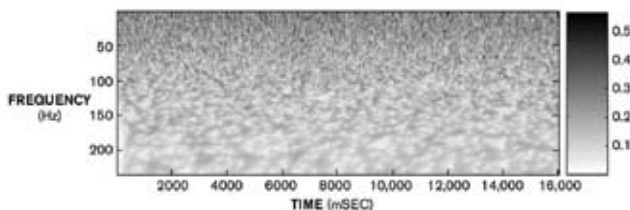


Figure 1 Stationary white noise is patterned and slow-changing.

that people listen to sound, using powerful cues such as pitch and spatial location of sound sources.

Beam forming uses multiple microphones to track a cone-shaped area of interest to locate and identify noise sources. Because the voice of interest lies within the cone, listeners can quickly differentiate as noise any sound sources outside the cone. Today's handset manufacturers recognize the trend toward multiple microphones and have begun to introduce second microphones into handset architectures.

BSS uses a linear unmixing technique to decompose the input sound mixtures into independent sources. The challenge in using BSS is that its linear unmixing technique requires as many microphones as there are noise sources, and it can suffer from convergence problems in the presence of reverberation when too many simultaneous noise sources are present.

USING MULTIPLE CUES FOR GROUPING

The human auditory system can hear voices in noisy environments because it uses all the information available in the signals arriving at the two ears. Like the human auditory system, ASA uses many methods to analyze the signals, resulting in multiple cues that can group the spectral energy into the corresponding sound sources. Some of the more important cues include pitch, spatial location, and common onset time. *Pitch* refers to the harmonics that a pitched sound source generates.

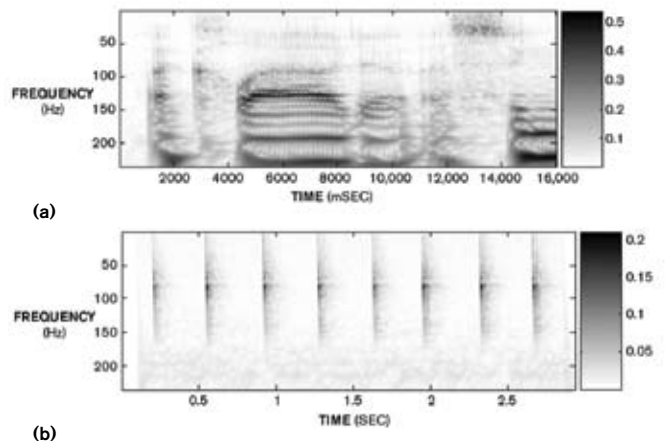


Figure 2 Speech (a) and pen-tap noise (b) are examples of nonstationary noise: rapidly changing, random, and often also containing harmonics across a wide frequency range.

These harmonics form distinct frequency patterns, so you can use them to distinguish one sound from another. Pitch is one of the primary cues, for example, in distinguishing between male and female voices. *Spatial location* refers to the location of a sound, based on its distance and direction; you can use spatial location to group sounds, thereby differentiating them from the voice of interest. *Common onset time* refers to the fact that, when two bursts of sound energy and their corresponding harmonics simultaneously occur, they are likely from the same source.

Traditional noise-suppression techniques must first converge before they can remove noise, making them ineffective in suppressing nonstationary noise sources. By using fast-acting cues to characterize sound, you can identify and remove even instantaneous events, such as a finger snap.

LOGARITHMIC VERSUS LINEAR SCALES

The familiar FFT (fast Fourier transform) decomposes frequency components on a linear scale that limits spectral resolution at low frequencies; it also uses a constant frame size and frequency-independent bandwidth. In contrast, an approach such as the FCT (fast cochlea transform) mimics characteristics of the human cochlea and operates on a logarithmic frequency scale. As a result, the FCT does not limit spectral resolution. By operating continuously instead of in frames, FCTs also reduce processing

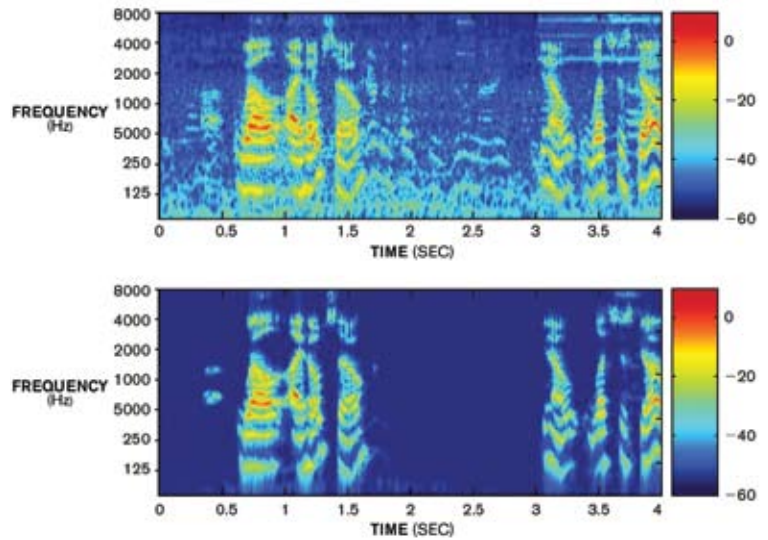


Figure 3 Next-generation techniques can dramatically improve the signal-versus-noise characteristics of captured audio.

latency, making them appropriate for identifying nonstationary noise sources. Additionally, FCTs operate with frequency-dependent bandwidth, so you can more precisely match the time-versus-frequency trade-off at each frequency of the human hearing range.

Certain techniques, such as beam forming, require a specialized cardioid unidirectional microphone. Cardioid microphones cost more and have tighter tolerances than do omnidirectional microphones. They also require individual calibration and matching to within 1 dB, introduce restrictions on spacing, and add as much as 12 dB of noise because of sensitivity to wind and breath. Beam forming is also limited in that any

distractions in the beam of interest will incorrectly pass through as part of the voice of interest.

You traditionally remove echoes using separate echo-cancellation techniques. Such techniques can be computationally intensive because

they must calculate echo reflections, and they offer poor performance in the presence of rapidly changing noise sources. Grouping cues enable you to treat echoes as simply another noise source. Instantaneous suppression becomes possible because you need neither to calculate nor to track the changes of echoes, providing echo-suppression performance to 46 dB.

NEW TESTING STANDARDS

The mobile-equipment industry continues to drive test standards to reflect higher levels of voice quality through innovations in noise suppression. To ensure the best quality for products, the recently amended ITU (International Telecommunication Union) P.835 specification provides a consistent test method for measuring and reporting voice quality with

active-noise-suppression technology.

Effective suppression of both stationary and nonstationary environmental noise is essential if handset manufacturers and carriers are to keep pace with their competitors. By employing next-generation noise-suppression techniques, developers can reduce noise levels in handsets by as much as 35 dB under a range of operating conditions (**Figure 3**).**EDN**

[+](#) Go to www.edn.com/ms4303 and click on **Feedback Loop** to post a comment on this article.

AUTHOR'S BIOGRAPHY



Lloyd Watts is the founder and chief technology officer of Audience and, as such, provides ongoing guidance and impetus for the company's core technology direction as well as the vision of neuromorphic computing for voice systems. Before joining Audience, he was principal researcher at Paul Allen's Interval Research Corp, where he collaborated with leading auditory neuroscientists on his vision of a machine that could hear as people do. Before joining Interval, Watts developed ICs and software for satellite-communications systems, telephony systems, optical-character-recognition systems, and LCDs for Microtel Pacific Research, Synaptics, and Arithmos. He also invented a low-delay digital-speech-coding algorithm that was sold to Cisco in 1999. Watts is the author of three issued and seven pending patents. He holds a doctorate from the California Institute of Technology (Pasadena, CA), a master's degree in digital-speech coding from Simon Fraser University (Burnaby, BC, Canada), and a bachelor's degree in engineering physics from Queen's University (Kingston, ON, Canada).