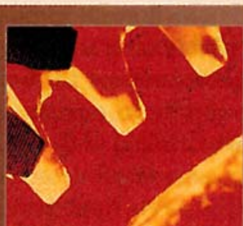
Advanced Noise Reduction for Mobile Telephony

Lloyd Watts, Audience, Inc.



Mobile telephones now contain advanced noise reduction based on the operation of the human hearing system.

Have you ever used your mobile phone to place a call from a noisy place and been asked to call back from someplace quieter? Wouldn't it be great if the phone could just mute the background noise for you?

Developers are now designing a new generation of mobile phones with advanced noise reduction to accommodate the noisy places in which they are used, making it easier for mobile phone users to hear and be heard.

This trend is being fueled by a new generation of advanced multimicrophone noise-suppression algorithms based on the operation of the human hearing system, running on dedicated signal-processing hardware. The result is a better user experience, improved customer satisfaction, reduced customer churn, and improved network efficiency.

A PHONE'S AUDIO PATHWAY

The audio pathway in a traditional mobile telephone contains a microphone to pick up the voice, a mixed-signal data converter chip to digitize the microphone input, a baseband processor chip to encode the voice signal for transmission, and a radio to transmit the signal to the nearest base station.

This basic architecture permits the successful and efficient transmission of voice in quiet environments—successful because the voice encoder works well on isolated speech, and efficient because the radio can be turned off or run at lower data rates when the encoder determines that voice is not present, thus reducing power consumption in the phone and reducing traffic in the network.

THE EFFECT OF NOISE

But what if the environment is noisy? Noise gets into the microphone signal and makes it harder for the voice encoder to operate properly, making the voice harder to understand at the other end of the call. And noise makes it harder for the voice encoder to determine that voice is not present, so the phone wastes energy transmitting noise that adds to the network traffic.

Types of noise

Noise would be easy to eliminate if it were always the same. But there are many types of background noise—the sound of a fan, the babble of a crowd, the rumble of a car engine, unwanted music, or a blaring announcement at an airport. Each of these noise types has different characteristics.

We can classify the types of noise according to how rapidly they change over time:

- *Stationary noise* is defined as noise that does not contain large or rapid changes in its spectrum over time, such as the steady sound of a fan.
- *Quasistationary noise* is noise for which the spectrum is largely constant over time, such as the generally steady babble of a large crowd.
- *Nonstationary noise* is noise that contains large or rapid changes in its spectrum over time, such as a single voice or music containing drums.

Strategies for reducing noise

Because of its relatively constant nature, techniques such as fast Fourier transforms (FFTs) and spectral subtraction can readily recognize and effectively remove stationary noise using conventional digital signal processing. However, nonstationary noise changes too fast and is often too similar to the desired voice to be removed with these conventional methods.

NOISE REDUCTION TECHNIQUES

A new generation of mobile telephones is emerging that uses noise-reduction techniques modeled after the operation of the human auditory pathway, as Figure 1 shows.

Two mics are better than one

In the same way that the human auditory system uses two ears, a handset can use two microphones to capture information about the posi-

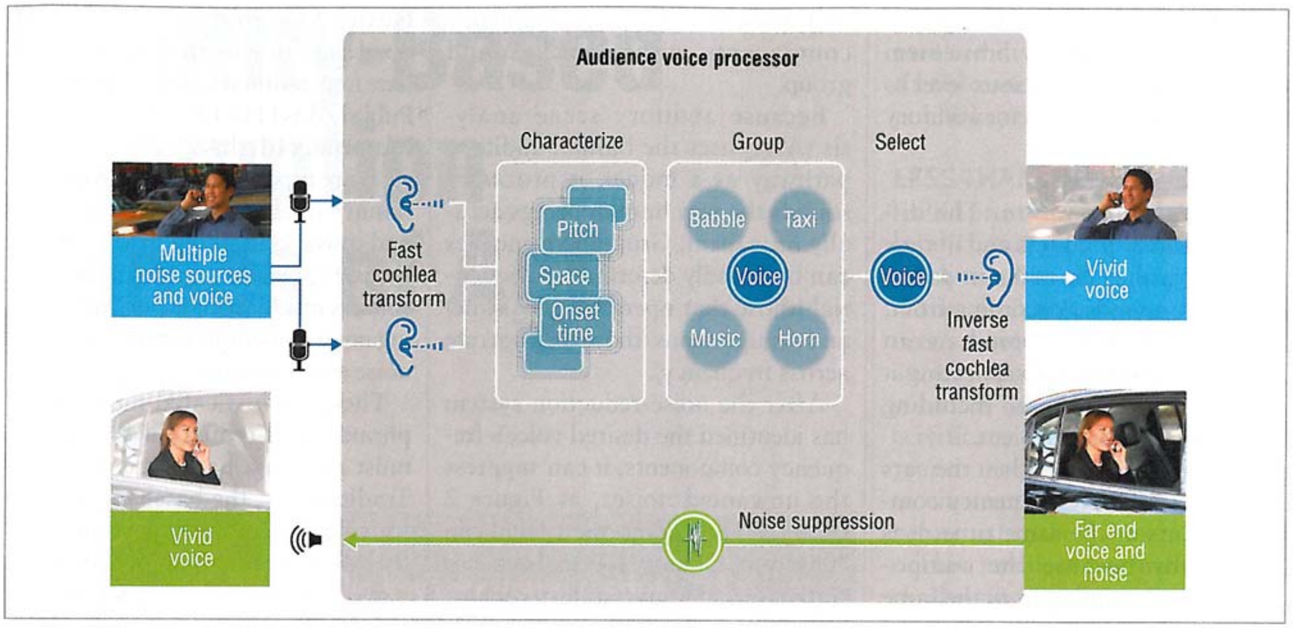


Figure 1. The Audience voice processor employs innovative technology based on auditory scene analysis to characterize and group noise sources and to distinguish and extract a dominant voice from multiple stationary and nonstationary noise sources.

tion of different sound sources in the auditory scene, allowing the phone to lock onto the desired voice and suppress the background sounds. Because of the dramatic benefit to noise suppression and voice quality, handset manufacturers have recently begun to include a second microphone in handset architectures.

Cochlea versus FFT

The first computational element in the human auditory pathway is the cochlea—the sensory organ that performs a sophisticated spectral analysis, decomposing audio signals received at the ears into their frequency components.

The familiar FFT is an efficient and useful algorithm for computing an audio signal's short-term spectrum. However, the FFT has many important limitations compared to the human cochlea.

The FFT performs its analysis on a linear frequency scale, which limits spectral resolution at low frequencies and temporal resolution at high frequencies. In contrast, as Figure 2 shows, Audience's fast cochlea transform operates on a logarithmic frequency scale, providing optimal spectral resolution at low frequencies and optimal temporal resolution at high frequencies, all at minimum possible latency. The result is lower latency, better noise suppression, and higher voice quality, all at lower cost.

Characterization based on the auditory brain stem

Humans are remarkably good at picking out voices in noisy places. We do it by using all the information in the signals arriving at our

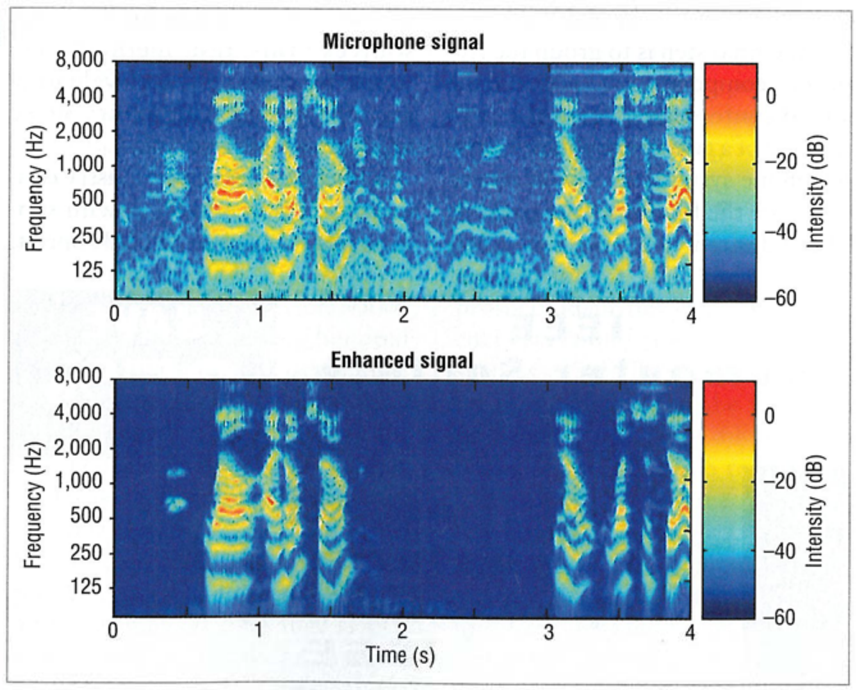


Figure 2. The fast cochlea transform representation of a microphone signal, before and after noise reduction. Notice the logarithmic frequency scale: each tic mark is a doubling in frequency. The signal is a voice recorded on a busy street with traffic noise, and a nearby talker and cell phone ring tone.

ears. Specialized neural circuits in the human auditory brain stem operate at the subconscious level to compute the major cues for auditory scene analysis:

- *Spatial information.* The differences in loudness and in time of arrival at the two ears tell us where sounds are coming from.
- *Pitch.* A voice's pitch is an important cue for separating it from other sounds, including other competing voices.
- *Common onset.* When the ears detect multiple frequency components at the same time, it is usually because the components are coming from the same source.

These powerful cues for source localization and separation are now incorporated into modern noise-reduction algorithms.

AUDITORY SCENE ANALYSIS FOR GROUPING AND SELECTION

Once the noise reduction system has characterized the auditory spectra, the final step is to group the frequency components according to the sound source that created them. In the case of a two-microphone mobile telephone system, we would like to associate the frequency components of the desired voice into one group and associate all other frequency components with a background group.

Because auditory scene analysis (ASA) uses the human auditory pathway as a model, it processes sounds the way human beings actually hear them. Grouping principles can be broadly described as sequential (those that operate across time) and simultaneous (those that operate across frequency).

After the noise-reduction system has identified the desired voice's frequency components, it can suppress the unwanted noises, as Figure 2 shows. At this point, the system can convert the auditory spectra back into sounds via the inverse fast cochlea transform, as Figure 1 shows.

TEST STANDARDS

In 2006, it became clear a new technology was emerging that would have an impact on the mobile telephone industry. At that time, common industry practice was to test noise-suppression algorithms with pink noise, car noise, street noise, and cafeteria babble noise. Because all these are quasistationary noise sources, this test methodology would not be useful for evaluating an advanced nonstationary noise suppressor's performance.

In October 2007, an industry consortium led by Audience, with support from T-Mobile, AT&T, Sprint, Nextel, Motorola, and Nokia, developed a new test methodology to evaluate nonstationary noise suppressors. Published as ITU-T P.835 Amendment I Appendix III, this standard specifies six noise types, including two nonstationary ones: single-voice distractor and music containing drums. It also requires physical motion of the noise sources in a four-loudspeaker recording environment to simulate realistic noise environments.

These sophisticated multimicrophone noise-reduction algorithms must run on a hardware platform. Traditionally, the baseband processor chip in a mobile telephone runs the voice encoder and, occasionally, it also runs a single-microphone stationary noise-reduction algorithm.

In 2008, we are seeing the deployment of a new class of chip in mobile telephone architectures: the *voice processor.* Residing between the microphones and the baseband processor, these chips run sophisticated algorithms with very low power consumption and an increasing suite of dramatic new features, including advanced echo cancellation, receive-side noise reduction, speech time-stretching, and many others.

The next generation of voice processors promises to bring exciting new functionality to mobile telephones, PCs, and converged entertainment devices.

Lloyd Watts is the founder, chairman, and chief technology officer at Audience, Inc. Contact him at lwatts@audience.com.

Computer welcomes your submissions to this bimonthly column. For additional information, or to suggest topics that you would like to see explained, contact column editor Alf Weaver at weaver@cs.virginia.edu.