Use microphone arrays for background acoustic noise suppression in portable devices--Part I

Kenneth Boyce, National Semiconductor Corporation - July 03, 2008

As portable voice communications devices become more ubiquitous, they are increasingly used in such noisy environments as airports, outdoor street traffic situations, and bars and clubs. These conditions make it difficult for the person speaking to be correctly heard and understood on the receiving end of the communications link. Additionally, many communication systems use computer voice recognition, command, and/or response systems. High levels of background acoustic noise can cause high error rates in these types of systems. There is much value, therefore, in being able to improve the speech signal to background acoustic noise ratio.

This two-part article explains the basic principle of using a microphone array for background acoustic noise suppression in a voice communications system.

Microphone arrays
A microphone array is any number of microphones spaced apart from each other in a particular pattern, which work together to produce a resultant output signal or signals. Each microphone is a sensor, or spatial window, for receiving (spatial sampling) the incoming signal. The overall response of the array is a superposition of the responses of each element in the array consistent with the processing algorithm used.

The multiple microphones signals undergo 'array processing' algorithms based on the microphone spacings and patterns, the number and type of microphones, and sound propagation principles. Microphone arrays are typically used to improve speech input signals in the presence of ambient noise in hearing aids, speech recognition equipment, and telecommunication products. But they may also be used to locate the direction and estimate the distance of sounds from the array.

The primary purpose of a microphone array for speech communications is to provide a high quality version of the desired speech signal while at the same time reducing the level of localized and ambient noise signals. The quality aspect means that the resulting speech signal is natural sounding, without any artifacts such as pops and clicks, un-intended muting, frequency distortions, echoes, or aperiodic changes in signal levels associated with the signal processing methods done to achieve the speech enhancement. Therefore, the measure of signal to background noise ratio improvement (SNRI) alone is NOT the only criteria to use in selection of a background noise suppression solution.

Sound Information

**Sound Pressure Level**

Sound Pressure Level (SPL) decreases proportionately with distance "x" from the sound source. Figures 1 and 2 show the SPL drop off, expressed in dB, as a function of distance "x" from the origin.
of the sound. For speech, the reference point is generally accepted as 96dB SPL approximately 1cm from the lips of a person talking. The equation plotted is:

\[ \text{dB} = 96 - 20\log(x/0.01), \text{ or alternatively, } \text{dB} = 96 + 20\log(0.01/x) \]

where the \((0.01/x \text{ or } x/0.01)\) term is the reference value distance of 0.01m, (1cm) relative to distance "x" in meters.

Both curves show a loss of 6dB for every doubling of distance. Figure 1 is for distances out to 200cm. Figure 2 is a magnified portion for distances out to 50cm, and shows how rapidly sound pressure drops as distance increases from the sound source, even for short distances. For example, at 10cm distance, the SPL drops ~20dB, from 96dB SPL to ~76dB SPL.

Near-Field versus Far-Field Sound
The near field of a sound source is generally considered to be within 1 wavelength of the lowest frequency signal of interest.

If the lowest frequency of interest for speech was 300Hz, then the wavelength $\lambda$ is equal to $c/f$ or $331.1/300$, or 1.104 meters, where $c$ is the speed of sound (331.3 meters/second) at sea level and 0 degrees Celsius. For a frequency of 3500Hz, $\lambda$ is equal to $c/f$ or $331.1/3500$, or 0.0946 meters (9.46cm). Therefore, the general near-field range limit of speech signals extends from ∼9.5 cm to ∼1.1 meters from the source. Beyond ∼1 meter, the speech signal is generally considered to be in the far-field of the speech source.

For closely spaced microphones in an array, near-field sound sources present a spherical wavefront with strong signal amplitude, pressure gradient, and frequency dependent differences based upon the distance from the source for each microphone in the array. For example, assume 2 microphones were placed 3cm apart, at a distance of 5cm for the nearest microphone to the sound source.

Figure 2 shows that the first microphone would experience a signal of 82dB SPL while the second microphone (at 8cm distance) would experience a signal of 78dB SPL. Although the difference is 4dB, the overall signals levels are still relatively high.

The near field speech signals in the microphones will be highly correlated with essentially the same spectral content. Compared to the nearest microphone, the signal in the most distant microphone from the source will be reduced in amplitude, as well as delayed by the time it takes for the sound to travel from the nearest microphone to the most distant microphone. Recovery of the speech signal in this case is relatively easy.

Sound signals originating beyond the speech near-field of the microphone array are considered to be in the far-field, presenting essentially planar wavefronts to closely spaced microphones in an array. Each microphone will be sensing almost identical acoustic energy and random phase signals, but those signals are lowly correlated unless the microphones are very close to each other. The further away these signals are from the microphone array, the lower the absolute SPL level at the microphones will be.

As a further example, if the same microphone array was placed at a distance of 150cm (1.5 meters) from the source, the SPL at 150cm would be ∼52.5dB for nearest microphone and at 153cm would be ∼52.3dB for the furthermost microphone. Although the difference is only 0.2dB, the overall signals levels are down ~ 30dB from the sound source to the closest microphone.

The difference signal between the microphone outputs, when appropriately processed and filtered, tends to cancel the far field noise signals while leaving a substantially high-level speech signal in the combined output of the two microphone amplifier and processing circuitry.

**Acoustic Noise Characteristics**

There are three classifications of acoustic noise fields: coherent, incoherent, and diffuse.

*Coherent* noise is characterized as propagating to the microphones without any form of reflections, spreading, or attenuation due to environmental obstacles.

*Incoherent* noise is noise at one location which is uncorrelated with noise at all other locations, and is considered spatially white.

*Diffuse* noise is characterized as noises of relatively equal energy that radiate in all directions simultaneously. Good examples would be office noises, airport lounges, traffic noise, i.e. practically all commonly encountered noisy environments.
There are two acoustic noise types: **stationary**, and **non-stationary**.

*Stationary* noise is characterized as being relatively constant in energy, with known and slowly changing spectral content, and tends to be predictable. Good examples would be engine noise, air conditioning fans, random or "white" noise, and so on. Noise suppression algorithms work well with these types of sounds.

*Non-stationary* noise is characterized by short duration changes in volume and content, such as people speaking loudly, passing traffic sounds, or clapping hands for example, and is basically not predictable. If such sounds occur, they would be gone perhaps before any rational noise recognition and suppression technique could be completely applied. Non-stationary noise is often embedded in a stationary noise field.

The most difficult problem occurs when the noise sources have the same temporal (time) and/or spectral (frequency spectrum), and coherency characteristics as the desired speech signal. This occurs when the background noise is non-stationary and consists primarily of other people talking, as would be the case in restaurants and bars, transportation terminals and parties.

Next: Microphone Array Solutions

**About the Author**

*Kenneth Boyce* is Technology Director for National Semiconductor's Audio products group. Prior to joining National, Boyce served as director of the Audio and Communications Division at Oak Technology, which developed initial implementations of AC-97 Codecs and Digital Audio Controllers. He holds a bachelor of science degree in electronics from West Virginia University.