Phoenix Audio Technologies
Algorithm Portfolio

Phoenix Audio Technologies offers a large menu of algorithms and techniques to extract a wanted audio signal from a noisy signal. These technologies clean up interfering noises while maintaining the integrity of the wanted signal. Phoenix also offers algorithms and techniques to obtain other information like the direction and nature of the emitting source. The recommended technique and related parameters are selected based on the particular problem, the required performance, the nature of the desired signals, the characteristics of the noise, and the expected cost of the solution.

The first consideration in selecting the right solution is determining the number of elements (microphones) that will take part in recording the audio signals. Typically, the more elements there are, the better the performance. But any addition of sensors increases cost, raises the complexity of installation, and creates a larger physical size. Phoenix arrives at the optimum design taking into account the particular challenge at hand, physical considerations, and cost / performance issues.

The required implementation is a consideration that opens a host of questions and issues that demand resolution. Run the algorithms on-line in realtime, or record the signals and run the algorithms off-line? If we run on-line, what is the amount of processing power and electrical power we can allocate to the tasks?

The Phoenix team will map your audio tasks and provide a solution that addresses all program requirements. The following executive summary is designed to overview our algorithms already successfully implemented with mature application examples.

The following table contains a short summary of some of the algorithms and their availability. More detailed and elaborated list follows.

Algorithm	Uniqueness	Implemented
Adaptive Beamforming	Transients Auto focusing in reverberant environment Non coherent residual suppression	PC, DSP
Acoustic Echo canceller	Adaptation during full duplex Residual suppression with no pumping noise Fast adaptation True full duplex with no degradation (167 allows degradation)	PC, DSP
Line echo canceller		PC, DSP
Blind Source Separation	First to implement a real time system	PC, DSP
Speech enhancement	Unique noise estimator (during continuous speech) Noise estimation under low SNR Non casual algorithms to manage dynamic noises Non speech transients Maximal noise reduction with minimal speech distortion	PC, DSP Some elements are PC only
Voice level compensation	No pumping noise, low computation load	PC, DSP
Sporadic array	Unique algorithm	PC
Reference channel reduction	Same as echo canceller	PC, DSP
Direction finding	Multi sensor DF, dual sensor DF, small aperture, high immunity to noise	PC
Tracking	Unable to compare – very few commercialized solutions	PC

I. Multi Sensor Processing

Multi sensor processing techniques utilize the acoustic information from more than one sensor to come up with information that does not exist in the single element. The relationship between the different sensors, the correlation between their data, the Time Of Arrival, and other such information can help improve the signal to noise ratio, reduce acoustic information that arrives from sources we want eliminated, find the direction of a source, and more.

A. Array Processing

Array processing is a subset of Multi Sensor Processing. It utilizes an array of microphones that are arranged in a determined position relative to each other. The acoustic signal hits the different microphones at different times. If the sensors are not too far apart from each other the microphone outputs will be correlated. The signals and the correlation between them can be used to derive information relating to the directions from which they originated, and also to create a spatial filter that will reject signals based on their direction information rather than on their statistical characteristics. The array can be structured as a linear array, or in any other desired configuration. The aperture of the array relative to the source of interest will determine the performance of the array, and more specifically the resolution that can be obtained with regards to this source.

1. System Identification

In an ideal free space environment, the acoustic signals would travel with no interference and no diversion, and arrive at the microphones with time differences that depend on geometry only. In a real world environment, the signals are reflected from surfaces such as walls, the floor, different objects and even from the microphone’s own packaging. The microphones receive the signals through the direct path, but also through several indirect paths. This effects the phase and amplitude information, creating confusion. In addition, the microphones themselves have variations in their phase and amplitude response.
The multi-sensor algorithms assume that the signals travel in free space and they are all matched in phase and amplitude. The mis-alignments in the signals will cause degradation in the performance and it is necessary to compensate for it.

The Phoenix’ “System Identification” algorithm estimates the acoustic transfer function between the array elements with regard for the desired source signal, and focuses the signal to obtain the optimum performance. This process improves the performance of array algorithms, but also allows the use of more aggressive adaptive filters that are very sensitive to channel errors and mis-matches in the elements’ sensitivity.

a. Features

- Robust and performs well in noisy and reverberant environments.
- Automatically tracks in a time varying scenario; e.g. movement of the desired source or non-stationary noise.
- Efficient recursive algorithm for real time application.

b. Maturity

- The algorithm was fully implemented on a PC environment.
- Tested using recorded signals.

2. Adaptive Beamforming

A simple delay-and-sum technique operating on an array of microphones (beamforming) will create a spatial filter. The “width” of the listening beam depends on the signal’s frequency and on the overall aperture of the array. In many cases, the frequency of the signal is such that a huge and non-realistic aperture is required to get reasonable performance.

In a more aggressive approach the array is used to create a “signal free” channel by placing a null towards the wanted signal. This “signal free” channel can be utilized as a reference channel with an adaptive filter creating a much higher resolution beamformer output. The adaptive beamformer is very effective in attenuating strong interfering sources. It is typically very sensitive to channel mis-match which might cause a distortion of the desired signal.
Phoenix offers a robust adaptive algorithm that overcomes these problems and performs well in real-life situations without causing any signal distortion.

a. Features

- Automatic compensation for microphone mismatch, (automatic focusing and calibration).
- Protection mechanisms to prevent signal distortion
- Fast adaptation.
- High attenuation of the interfering signal.
- Efficiently adapts to handle transient noises

b. Maturity

- Implemented in fixed point and floating point arithmetic
- Used in operational systems

3. Multichannel Post filtering

The adaptive beamformer is not efficient with non-stationary noises. Phoenix has developed an algorithm that uses the direction information from an array to track and associate transient noises with sources. This algorithm will identify transients that "do not belong" to the wanted source and eliminate them. This post-filtering process allows additional noise reduction at a beamformer output.

a. Features

- A significantly reduced level of non-stationary noise is achieved without further distorting wanted signal components.

B. Blind Source Separation

Beamforming algorithms use the TOA (Time Of Arrival) information to filter out signal, based on their relative direction. Another approach known in literature as Blind Source Separation is to try to separate a signal into its sources based on a statistical independence criterion. If, for example, there are two dominant sources transmitting audio signals and these signals are convoluted and received by two microphones, the algorithm can de-convolute the microphone outputs and reconstruct the sources. The advantage of this algorithm over beamforming is that it does not assume any geometry and does not use any direction related information.

a. Features

- Performs in a real environment e.g. two sources in a normal environment and reasonable background noise.
- Propriety algorithm that can handle a convoluted mixture with a relatively long transfer function.
- Does not introduce scaling or permutation problems.

b. Maturity

- Implemented on floating and fixed point arithmetic
- Tested on real life scenarios using a hardware demo unit.

II. Echo and Reference Channel Canceling

When we have the direct measurement of the noise signal, we can use an adaptive filter to filter out this noise from the polluted signal. Echo canceling is a private case of such an application. Phoenix offers several algorithms that have been developed and optimized under this category.

A. Line Echo Canceller

Line Echo is the reflection of the audio signal that occurs in communication lines due to impedance differences at the end of the line. The signal we are sending hits the end of the line and part of it returns back. The signal we are sending is the reference to the echo we will receive after a delay that equals the travel time. Line echo represent a relatively simple scenario - short impulse response and fixed transfer function.

a. Features

- Propriety adaptive filter update algorithm.
- Fast adaptation
- True full duplex operation – no perceptual distortion of the near-end signal during full duplex scenario.
- Integrated noise reduction algorithm.

b. Maturity

- Implemented in fixed point and floating point arithmetic

B. Acoustic Echo Canceller

Acoustic echo occurs when the microphone receives both our speech and also the signal coming out of the loudspeaker containing the "other side's speech" (far-end). Without an echo canceller, the other side will hear their own voice returning back. We will use the electrical signal as it is measured at the loudspeakers as the reference channel.

a. Features

- Long tail length (up to 200 ms).
- Proprietary adaptive update algorithm.
- Fast convergence.
- Tracking transfer function variations even during full duplex scenarios.
- True full duplex operation, i.e. no perceptual distortion of the near-end signal during full duplex scenario.
- Insensitive to background noise both in the near-end and the far-end signals.
- Integrated noise reduction algorithm.

b. Maturity

- Implemented in fixed point and floating point arithmetic
- Used in operational systems

C. Stereo Acoustic Echo Canceller

When the signal coming out of the communication line is stereo we need to cancel two reference channels that are different, but still highly correlated to each other. The unique problem of the stereo acoustic echo canceller is fundamentally associated with the time varying correlation between the reference channels. This generally prohibits the algorithm from converging to the correct solution. A known solution in literature introduces nonlinearity in one of the channels (before the signals are played through the loudspeakers), in order to reduce the correlation between them. However, this solution degrades the audio quality. We developed an algorithm that does not reduce the audio quality.

a. Features

- Tracking transfer function variations even during full duplex scenarios.
- True full duplex operation, i.e. No perceptual distortion of the near-end signal during full duplex scenario.
- Insensitive to background noise both in the near-end and the far-end signals.
- Integrated noise reduction algorithm.

b. Maturity

- Implemented in fixed point and floating point arithmetic
- Used in operational systems

III. Single Sensor Processing

A. Single channel speech enhancement.

The goal is to improve the SNR of a signal received by a single sensor (microphone) while maintaining the integrity of the wanted information (speech embedded in noise). When we try to extract a wanted signal (typically voice) from noise, the challenge is to detect the noise elements within the signal. This challenge is particularly difficult when the algorithms are required to work on-line so that we cannot manually mark the noise segments. In many cases, the assumption is that the noise is stationary and the wanted signal (speech) is not stationary. In more advanced solutions, we try to address non-stationary noises (transients) and we use additional and finer statistical differences between speech and non-speech signals.
Once the noise elements are detected and analyzed, they are filtered out from the signal. The two processes, the detection of the noise and the filtering, are done even in the presence of noise.

The techniques we have developed can be divided into two categories:

Low delay, low complexity algorithm

Pros:

- Small footprint - very low computation load and low memory requirements
- Short signal delay (10-20 ms)
- Fast tracking of non stationary noise components.
- No musical artifact

Cons:

- This algorithm performs very well when the SNR of the input signal is not too low. For very low SNRs, the algorithm introduces some signal distortion.

The algorithm uses a unique technique for the noise detection and estimation process that we refer to as a *"Minima Control Recursive Average (MCRA)"*. This technique achieves good results even in the presence of continuous speech and does it in real time, on the fly.

b. Maturity

- This algorithm has been implemented both in fixed point and floating point arithmetic.
- Extensively used in operational systems – both embedded applications and PC applications.

High performance Algorithms

Pros:

- The algorithm performs well for very low SNRs with minimal signal distortion.
- Fast tracking of non stationary noise components.
- No residual musical artifact.

Cons:

- Larger footprint and higher computation load and memory requirements compared to our "low delay algorithms".

The Phoenix high performance algorithms include several elements that differentiate us from commonly used techniques:

The a- priori SNR estimation
This is a critical parameter in modern speech enhancement algorithms. Existing methods (e.g. the Decision Directed approach) are heuristic. Phoenix developed an estimation approach with solid mathematical foundations which yield better performance. This approach, based on speech models, performs estimation under speech presence uncertainty and utilizes *Noncausal estimation.*

The Filtering
The common way to perform noise filtering is referred to in the literature as the "*Log Spectral Amplitude (LSA) estimation*". This method assumes that speech is present at all times and it estimates the filter coefficient accordingly. Phoenix developed an optimized technique that we call "*Optimally Modified Log Spectral Amplitude (OM-LSA) estimation*". This technique takes into account the probability that speech is present while estimating the filter

coefficient. The optimized method reduces further the musical noise artifacts and achieves better overall performance.

Maturity:

- These algorithms were implemented in floating point arithmetic for the PC environment.
- Performance was validated both by informal listening and by formal objective criteria, such as Segmental SNR and Perseptual Evaluation Of Speech Quality Scours (PESQ) –IDUT p.862. Also, the efficiency of the algorithm as a preprocessor for Voice Recognition algorithms, was proven by using a variety of Voice Recognition packages.

B. Multi Resolution Analysis (Wavelets)

Classical enhancement systems are based on uniformly spaced frequency resolutions. However, a non-uniform frequency resolution, which reflects the human auditory system, often leads to improved intelligibility and quality of the enhanced signal when combined with an appropriate spectral gain function. We have developed an enhancement method, which integrates a Bark-Scaled Wavelet Packet Decomposition (BS-WPD) and a soft-decision gain modification technique.

Features:

- The BS-WPD provides an efficient auditory representation.
- The algorithm achieves lower residual noise and higher intelligibility and quality of enhanced speech, compared to classical enhancement systems, which are based on uniform spectral decompositions

C. Signal Power Compression

Voice signals occupy a large dynamic range. Different speakers at different times, under different noise environments, and from highly different distances to the microphone will excite the microphone to produce a very large dynamic range. The nature of speech is that it also includes a lot of high level transients. Trying to avoid clipping by pushing the signal downwards will leave very little room for most of the signal, and soft speakers will find themselves below the noise floor.

Phoenix offers a unique algorithm that addresses these issues. It allows comfort listening in scenarios of multi speakers with significantly different power (for example: a scenario of room conferencing where one participant is located near the microphone and another may be located far from the microphone). It also avoids clipping while maintaining enough “bits” for the rest of the signal. The algorithm excels in its very fast response and its low complexity. It has no pumping noise that is typical to simple AGC techniques.

Maturity:
- Implemented both in floating point and fixed point arithmetic.
- Widely used in operational systems.

IV. Direction Finding, Data Tracking and Data Fusion

A. Bearing Only Tracking systems

We have developed multi target tracking for a passive low frequency sonar system (flank array).

Features:
- Multi Component tracking – Broadband; Narrowband and Demon
- Optimal direction finding of each component and for the combined data
- $\alpha\beta$ tracker (steady state Kalman filters)
- Maintains tracking at a very low SNR (-40 dB)

Maturity:
- Fixed point implementation.
- Used in operational submarine.

B. Data Fusion Algorithms

We developed a data fusion algorithm for sources with explosive nature.
The information was collected from distributed sensors (the total aperture was over 20Km), where each sensor composed of a small aperture microphone array.
Data association was resolve by using Bearing, Time Of Arrival and acoustic features. The system achieves accurate source localization and good rejection of false detection.

C. Direction Finding

We have developed Direction Finding algorithm for three applications: speech signals localization, sniper localization, artillery localization.

Features:
- Insensitive to errors in the sound speed
- Supplies measurement for quality estimate

- Both azimuth and elevation can be estimated using planar array.
- Achieves the Cramer Rao Bounds
- Very low computation load

Maturity
- Implemented both in fixed point and floating point arithmetic
- Used in operational system.

D. Dominant Speaker Detection

The need for dominant speaker selection arises in some applications such as multi-point video conferencing or multi-channel surveillance recording.

Features:
- Operates under adverse conditions such as noise, interferences, non uniform SNR at the different channels.
- Distinguishes between non speech transient and speech transient
- Low complexity

Maturity:
- The algorithm is implemented on PC environment.
- Used in operational systems

E. Automatic Mixer for Audio Streams

In this field we developed two kinds of algorithms (for different applications)

1. Algorithm #1:
- The mixer selects only the "most desired" channel to the output
- The amplitudes of the channels are automatically balanced (signal power compression)
- The noise is removed from each channel.

2. Algorithm #2:
- The mixer optimally mixes the input channels
- Channels are automatically balanced.
- No saturation when several signals are active simultaneously.

Maturity:
- The algorithms are implemented both in fixed point and floating arithmetic
- Utilized in operational systems.

V. Anomaly Detection

Anomaly detection is the process of detecting samples, which differ from their environment in some statistical sense. We have developed an anomaly detection approach for images and multi-dimensional data, and a wavelet-based iterative procedure for feature extraction.

Features:

- The anomaly detection does not rely on an exhaustive statistical model of the targets, but rather on the local statistics of the data and optionally on some limited a priori information regarding the targets' subspace.
- The wavelet-based feature extraction generates multi-resolution features, which enable to characterize different anomalies and backgrounds in multiple scales.
- Low complexity.
- The false alarm rate is iteratively reduced, while maintaining a high probability of detection.