Technical White Paper on Fortemedia Voice Processing Technologies

Introduction

Fortemedia’s voice processing IC is a unique solution that enables one piece, full duplex speakerphones to be designed into a single package previously thought impossible. Microphone to loudspeaker separation of only 40mm is now a reality, allowing a full duplex speakerphone to be built in a 25x40mm package (see figure 1). The purpose of this white paper is to introduce this revolutionary voice processing technology which combines Fortemedia’s patented AS3NLF (Adaptive Step Sized Sub-band with Non-Linear Filter) technology with AMBIN (Array Microphone Beam-forming Integrated Noise suppression) small array microphone into a single chip solution. Fortemedia’s single IC solution can be applied for all speakerphone applications, and can achieve up to 65dB of acoustic echo cancellation and up 18dB noise suppression. Fortemedia’s single IC includes ADCs, DACs, and digital interface for control, voice, and data.

Figure 1. Comparison of Fortemedia Speakerphone and Conventional Speakerphone

Background

Full duplex communication is the natural way of communicating because both parties on either side of the communication channel can talk at the same time without having their voice cut off. This is how people talk face to face, naturally such that either party can speak at anytime during the conversation. Therefore, it is unnatural to communicate in half duplex mode where one side has to wait for the other side to finish talking. In half duplex mode, if doubletalk happens, the voice will be cut off on one side and some of the conversation will be lost.

With full duplex communication, echo, howling, and acoustic feedback become problematic if the headset or speakerphone used cannot isolate the speaker from the microphone acoustically. The traditional method is to separate the speaker and the microphone by keeping them far apart to minimize acoustic leakage and to also avoid putting them in the same enclosure. However, the size of the device cannot be smaller than the minimum distance required for the separation of the speaker and microphone to achieve full duplex. Although DSP algorithms can be used to overcome the issues mentioned above, most of the algorithms used cannot provide sufficient echo cancellation to bring the speaker and microphone closer while still driving the speaker to full volume. Most of them also cannot suppress non-linear echo to allow the speaker and microphone to be designed in the same enclosure.
Today, the demand of the electronics world is to make devices smaller, more portable, and still exhibit the same performance as a larger device. As such, there’s a need to find a new way to overcome these problems of placing the speaker and microphone close to each other in the same enclosure, but still achieve good full duplex without echo and howling. Using the traditional methods of isolating the speaker and microphone or lowering the performance by turning down speaker volume are simply not enough for the new demand in portable electronics. Fortemedia’s voice processing technologies are designed to overcome such obstacles.

The Fortemedia solution can achieve superior full duplex communication with a loud speaker and microphone in the same housing only 4cm apart from each other. This is made possible by using our unique two microphones solution and our patent pending DSP algorithm. One of the microphones is used to sample the voice while the other samples the sound from the loud speaker. The DSP algorithm then does the computation to cancel the echo and eliminate the howling. No other solution on the market can achieve this kind of performance and allow designers to fit the speaker and microphone(s) in the same small enclosure. The Fortemedia IC will also work with a single microphone; however, the performance will not be as good as the dual microphone approach.

**Why Array Microphone**

Seamless handsfree full-duplex communication without requiring the users to wear or hold special devices is very desirable for natural human-human or human-machine interaction. For handsfree communications, acoustic echoes and howling are the two major problems, particularly at a big volume for loudspeakers. Noise/interference is the other problem. In daily life, people are always exposed to background noise, which may come from car engine, traffic noise, other speakers present in the same space (cocktail party noise), computer fun, and audio equipment, etc. Therefore, to achieve the real natural handsfree communications, acoustic signal processing is employed to cancel acoustic echoes, control acoustic howling, reduce noise and suppress interference. Such handsfree full-duplex communication technologies may be tailored for incorporation into a wide variety of communication terminals, including handsfree car kits, car information systems, mobile phones, teleconferencing systems, telephones, computers, VoIP equipments, home entertainment equipments, etc.

What are requirements for good natural handsfree communications? The best natural handsfree communications is just let people talk like face to face. To achieve this purpose, ideally we need a) totally echo free and howling free at any situations, b) truly full-duplex without cutting off voices for both far-end and near end, c) big loudspeaker volume with good intelligibility, d) noise free voice without distortion, even in noisy environments, e) interference free without distortion, and f) perfect voice recognition rate, even in adverse environment, for human-machine interaction.

Could single microphone solution get what we need for good natural handsfree communications? Single microphone solution is the most popular one for current handsfree communications. The most reasons behind are single microphone solution is simple and has been available for a long time. Single microphone acoustic echo cancellation (AEC) uses the signal sent to the loudspeaker as reference information to cancel acoustic echoes. In such a case, nonlinearity will destroy the correlation between the echoes received by microphone and the reference signal, which will directly degrade echo cancellation performance by single microphone. The nonlinearity may be caused by loudspeaker saturation due to high volume and low-end loudspeaker, vibration of the device, echo path change, or microphone saturation, etc. How long echo tail can be cancelled for one microphone AEC is very dependent on the filter length, which will cause much more computations and memory consumption. Double talk is not very easily detected for single microphone AEC, particularly on high loudspeaker volume and short distance between microphone and loudspeaker. In such a case, the signal to echo ratio may be less than –20dB to –30dB. So definitely full-duplex will be affected by inaccurate double talk detection. In addition, single microphone noise suppression techniques can only differentiate between signals that have
different temporal and spectral characteristics. As a result, they cannot improve the local SNR in each frequency band, but only the global SNR. One microphone noise suppression techniques also cause artifacts introduced by the nonlinear filtering process and an inaccurate estimation of the noise characteristics. The most popular information to differentiate speech and noise is how stationary it is. In general, for one microphone noise suppression schemes, it is assumed that noise characteristic is 10 times more stationary than that of speech. The speech intelligibility is then only improved if the frequency bands that are attenuated contain so much noise that are mask out the useful speech information in other bands. However, the frequency spectra of everyday noises, such as car noise and cocktail party noise, are often similar to the average spectrum of speech, which makes it difficult for one microphone noise suppression schemes to effectively eliminate most of the noise without reducing speech intelligibility at the same time. In a word, one microphone is very difficult to get very good performance for handsfree full-duplex communications.

How can two or multi-microphone solutions get much better performance? Two or multi-microphones can form an array microphone, which can allow beamforming with spatial filtering of arriving signals and, thus, desired signals inside the beam can be enhanced and, noise and interferences outside beam can be suppressed. With full-duplex communication, echoes will join local interferences to corrupt the desired speech signals. Beamforming, however, doesn’t exploit the signal sent to loudspeaker as reference information to cancel acoustic echoes. Instead, it uses spatial filtering to remove echoes. The advantages of the spatial filtering include a) it can cancel nonlinear echoes because it is independent on the nonlinearity caused by loudspeaker saturation, echo path change, and even vibration, etc. b) it can make big echo cancellation length without increasing much computations because it is also independent on echo tail length, and c) it can easily get accurate double talk detection. All above items can enhance AEC and full-duplex performance. In most scenarios, the desired talker and its disturbing sources are located at different positions in space. Two or multi-microphone noise suppression techniques are able to exploit the spatial diversity and statistical characteristics difference in addition to spectral and temporal difference. They can form a beam to enhance the desired speech and suppress other interferences outside beam, and also can use statistical information of noise and speech to extract speech from noisy speech. These can not only improve local SNR, but also improve the global SNR, which will increase speech intelligibility under environment with both stationary noise and nonstationary noise. Fortemedia’s array microphone technology can fully realize the advantages on AEC and noise/interference suppression of multi-microphone technology with small array microphone.

History

First let’s take a look at the history of the adaptive filter since it is a big part of Fortemedia’s echo cancellation algorithm and it will help to have a deeper understanding of Fortemedia’s technology.

Adaptive Filter History
1940s - Wiener filter
1960s - Kalman filter
1960s - LMS (Least Mean Square)
1970s - NLMS (Normalized LMS)
1980s - Affine Projection
2000s – Fortemedia’s Adaptive Step Size Sub-band, Non-Linear Filter (AS3NLF)

Echo cancellation has been through significant changes since the Wiener filter technology became available in 1940. The Kalman filter and LMS (Least Mean Square) algorithm replaced the Wiener filter during 1960s, because the Wiener filter was not suitable for real time applications. Due to the increased computational requirements for the calculations in the embedded world, the cost of implementation of the embedded system is very high with the Kalman algorithm. The LMS algorithm has less stability while applied with real human voice communication applications. A modified and more stable algorithm, NLMS (Normalized LMS),
was a big milestone and became prevalent in the ‘70s, while the Affine Projection algorithm was introduced in the 80’s to accelerate the convergence times. Today, Fortemedia’s advanced AS3NLF adaptive filter provides superior echo cancellation.

Fortemedia AS3NLF Technology

Fortemedia’s AS3NLF technology is implemented in our state of the art single-chip echo canceller, AMBIN FM1073. With echo cancellation in both time and frequency domain and innovative non-linear processing, AS3NLF is able to handle extremely large echoes. The FM1073 also incorporates AMBIN Channel Control, Small Array Microphone with Beam-forming, and Non-Artificial Noise Smoother.

Fortemedia’s AS3NLF technology has two major portions: Adaptive Step Sized Sub-band NLMS and AMBIN Non-Linear Filter, as shown in figure 2. With AS3NLF, FM1073 achieves 65dB (Adaptive Step-Sized Sub-band NLMS contributes 30dB and AMBIN Non-Linear Filter contributes 35dB) of Acoustic Echo Cancellation performance in a mostly linear system. In its worst case where a system is operating in a non-linear mode, the FM1073 can cancel 52dB. The convergence time for the echo path change is less than 50ms as compared to other systems at 60-80ms. It also provides excellent noise suppression of 13-18dB for both stationary noise and non-stationary noise, enabling exceptional voice quality and superior full duplex capability.

Figure 2. AMBIN Technology Block Diagram

AS3 NLMS (Adaptive Step-Sized Sub-band Normalized Least Mean Square Algorithm)

AS3 NLMS does not use noise to train the echo canceller. Rather, it continuously trains on the normal exchange of speech during the conversation. This results in a more natural flow of conversation from the beginning of the conversation. Fortemedia technology also allows talkers to move around.

AEC adaptation can be roughly divided into two phases: large, rapid changes are required to adapt to major acoustical changes (such as moving to a new room); smaller changes are required to adapt to minor perturbations or echo path changes (people moving, doors opening, etc.). When an AEC is first operated in a room/car or moved to a new location, it needs to adapt to the new acoustics of its surroundings. A good AEC approaches this level of acoustical change quickly and unobtrusively by determining when it is in the receive state and adapting rapidly.
during that state. **AS3 NLMS** uses adaptive step-size to automatically adjust the change rate based on the echo environment. When there is only near end talk, the step size is minimized, and becomes medium when there is double-talk. The step size is maximized when far end talks. This method allows the system to quickly switch among the various states and still maintain stability. Hence it performs extremely well on the echo path change.

**AMBIN Nonlinear Filter**

Non-linear echo is caused by mechanical vibration, speaker and microphone saturation, and echo path changes. If non-linear echo is not dealt with properly, it will result in howling and echo from the speaker, or the system would revert to half duplex.

Traditionally, center clipping is widely used to suppress non-linear echo. The major disadvantage of center clipping is that it does not suppress echo without significantly degrading the full duplex performance and voice quality. **AMBIN Non-Linear Filter** uses very different techniques to suppress non-linear echo. There are three major elements in the AMBIN Non-Linear Filter: Big Echo Cancellation, Post Filter, and Frequency Domain Non-Linear Filter.

Fortemedia’s **Big Echo Cancellation** process first detects and classifies the non-linear echo element by correlating the reference signal from the reference microphone and the main signal from the voice pickup microphone. The resultant correlated signal process can handle very large and severe non-linear echo. Fortemedia’s nonlinear process also suppresses the non-linear component accordingly based on the degree of non-linearity; therefore it doesn’t affect full-duplex performance.

Fortemedia’s **Post Filter** utilizes sophisticated suppression factors to eliminate residual echo. Depending on the proportion of the linear and non-linear echo, the Post Filter can achieve an additional 25 to 35dB of echo cancellation with a faster convergence time than other echo cancellers. This allows the FM1073 to remove echo tail and adapt rapidly and unobtrusively to changing acoustic conditions. Users can move about freely during the conversation without degrading communications quality.

The **Frequency Domain Non-Linear Filter** utilizes the difference in acoustic characteristics between two microphones, preferably an omni-directional and a unidirectional microphone, to further identify and suppress specific non-linear echo elements. This mechanism takes advantages of correlated acoustic information from the two microphones and effectively eliminates echo created by a loudspeaker operating in non-linear range. A much better full duplex performance is achieved this way while suppressing non-linear echo compared to other non-linear filters.

The implementation is described as follow. Non-linear echo in the integrated speakerphone is caused by (1) mechanical vibration from the speaker to microphone (2) over driven speaker (see below).
By having both microphones receiving non-linear echo, the non-linear echo can be removed after the process.

With Fortemedia’s patent pending mechanical design, only the reference microphone (omni microphone) is able to receive echo from internal path (see below). The main microphone (either omni or uni directional microphone) has to be completely isolated from internal feedback. This mechanism enables Fortemedia’s solution to detect double talk more accurately and therefore significantly increase the full duplex performance.

The combined **AMBIN Non-Linear Filter** automatically adapts to changes in the placement of both the loudspeaker and microphone and to changes in loudspeaker volume. The system integrator is freed from the design constraints required of other echo cancellers, and end users can arrange their space as they desire.

The 65dB AEC performance from Fortemedia (compared to 35dB AEC performance with conventional acoustic echo cancellers) offers significantly greater acoustic power output. This permits great flexibility in microphone and loudspeaker placement and volume adjustment. With the **AS3NLF** technologies, FM1073 achieves between 25dB to 35dB side tone reduction.

**AMBIN Channel Control**
A traditional AEC uses “Voice Activity Detection” to converge echoes. But using “Voice Activity Detection” is not as decisive or may even make incorrect decisions during this critical phase due to the fact that Voice Activity Detection is hardly accurate by nature. As a result, the AEC remains un-converged for a long time, and in extreme cases, never properly adapts. Some echo cancellers force the speakerphone to store the room characteristics after the initial convergence. This compensates for the fact that the echo canceller is not capable of converging quickly to major acoustical changes. In this scenario, the AEC must undergo a rapid training procedure to learn the interior environment from its un-initialized state. Once trained, it adapts to small acoustical changes, but major changes require retraining. This training usually takes the form of a loud burst of noise or a sequence of tones, which the AEC uses to adapt to the gross acoustical characteristics of its environment. Fortemedia does not use “Voice Activity Detection” in its AEC algorithm.

Fortemedia’s AMBIN Channel Control compensates for inadequacies in the conventional AEC by restricting the rate of change and the amount of change allowed in the adaptive filters. This prevents the AEC from going too far out of convergence by adapting too rapidly when it is confused by a major disturbance, while allowing the AEC to track relatively minor changes such as a door opening or slow movement of a driver in a car. The most complex and difficult task of an effective AEC is reliably determining when to permit its internal acoustic model to adapt to changes in the acoustic character of the car/room. Such changes occur when volume levels are changed, people move about, doors are opened or closed, the loudspeaker or the microphone is moved, etc. Adaptation should only occur when in receive mode (for example, when the near end party is silent and the far end party is talking). Inaccurate mode decisions cause the AEC’s internal model to diverge resulting in echoes which are not effectively canceled. Inaccuracies in this decision process can cause another significant problem, one that is handled very elegantly with Fortemedia’s AMBIN Channel Control. In addition to reducing the effectiveness of the echo cancellation, inaccurate state or mode decisions may introduce artifacts into the transmitted or received speech signals. Words may be clipped or exhibit dropouts. Switch loss or center clipping, which are nonlinear processes, may be applied to a speech signal in the wrong mode. This causes chirping or warbling artifacts that can be annoying and distracting. This can be noticeable in any mode, but particularly in doubletalk (both parties talking simultaneously).

**AS3NLF** and **AMBIN Channel Control** create a smooth transition between states in double talk mode, permit natural, undistorted doubletalk, and enable a fast and accurate adaptation.

**Small Array Microphone and Cone Shaped Beam-forming**

Fortemedia systems work best with 2 or more microphones, but it can also work with just one microphone. With 2 microphones, a small array microphone configuration can be utilized, and beam-forming becomes an option if one unidirectional microphone and one omni-directional microphone are used.

<table>
<thead>
<tr>
<th>Overall System Performance</th>
<th>Microphone Configuration</th>
<th>AEC and Convergence Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Good</td>
<td>One omni-directional microphone</td>
<td>45dB, 30ms</td>
</tr>
<tr>
<td>Better</td>
<td>Two omni-directional microphones</td>
<td>65dB, 15ms</td>
</tr>
<tr>
<td>Best with Beam-forming</td>
<td>One unidirectional microphone and one omni-directional microphone</td>
<td>65dB plus 15-20dB non-stationary noise, 15ms</td>
</tr>
</tbody>
</table>
Best with gated microphones

| Two unidirectional microphones (gated) and one omni-directional microphone | 65dB plus 15-20dB non-stationary noise, 15ms |

With Fortemedia’s patent pending AMBIN small array microphone technology, the distance between 2 microphones may be shortened to a few millimeters using both a unidirectional (main) microphone and an omni-directional (reference) microphone. The resulting unique cone-shaped beam is able to suppress noise right above and below the microphone array (see figure 3).

By processing acoustic information based on the different characteristic between the unidirectional (main) microphone and omni-directional (reference) microphone, AMBIN technology achieves an exceptional beam-forming effect (see figure 4) to suppress non-stationary noise by 20-25 dB and to enhance the voice quality, while still maintaining a very small form factor.

*Figure 3. Beam-forming Effect of Small Array Microphones*
Non-Artificial Noise Smoother

A CDMA handset has its own noise suppression to (1) suppress the background noise, but it requires a long convergence time, and (2) cut off the background completely when there is no voice from the near end. The problem occurs in the scenario where there is a hands-free car kit at the near end: the noise is suppressed twice, first by the hands-free car kit, then by the CDMA handsets. When the near end is in a noisy environment, the far end will hear unstable noise coupled with the near end’s distorted voice due to the long convergence time from the CDMA handset’s noise suppression. This effect reduces the intelligibility and deteriorates the quality of communication. A noise smoother adds some noise back to the background to shorten the convergence time for the handset noise suppression, therefore improving the communication quality. However, conventional noise smoothers are artificial, which introduces un-natural background noise and creates a very unpleasant user experience.

Fortemedia’s Non-Artificial Noise Smoother uses the actual noise as the base. While applied to the hands-free applications, the noise level can be adjusted to the handsets to maximize the comfort with a soothing background.

Conclusion

Combining AS3NLF, powerful AMBIN Channel Control, and small array microphones, Fortemedia’s echo cancellation technologies deliver a superior AEC performance of up to 65dB. Fortemedia then embodies these technologies into a very small single chip package: the FM 1073/FM1072LP/FM1093 with integrated DSP, memory, and CODEC.
FM1073/FM1072LP/FM1093 are the ideal single-chip solutions for hands free communication applications such as hands free car kit, speaker phone, or other embedded devices.