

**WHITE PAPER**

# Voice Processing for Home Automation Systems

**Andre Coucopoulos**

Product Marketing Manager, Voice Processing  
Zarlink Semiconductor

As published in Network Systems Design Line  
January 2007



Home networking has evolved from linked personal computers to a more complex system that encompasses advanced security and automation applications. Once just reserved for high-end luxury homes, home networks are now a regular feature in residences. These networks allow users to consolidate heating, air conditioning, lighting, appliances, entertainment, intercom, telecommunication, surveillance and security systems into an easy-to-operate unified network.

Interactive applications operated by voice recognition, for example integrated door security systems and the ability to control home appliances, are key features of home automation networks. This interactive capability depends on high-quality voice processing technology, including acoustic echo cancellation, low signal distortion and noise reduction techniques. A home automation system must also be scalable to allow future evolution, flexible to support field upgrades, interactive, easy-to-use, cost-efficient and reliable.

This article introduces some of the voice quality performance issues and design challenges unique to home automation systems. It will discuss home automation network applications that rely on voice processing, and examine some of the critical features and functionality that can help ease design complexity and cost to deliver enhanced performance.

## **VOICE PROCESSING IN HOME AUTOMATION**

The home automation market is moving beyond high-end luxury homes to target the mainstream consumer. Even in its infancy, researchers estimate the market is worth over \$1 billion. In Asia, Europe and North America the home automation market is growing at an average of 10% per year. In Europe alone, demand for home automation systems is expected to double by 2009 to create a \$0.5 billion market. While the market grows, home automation systems themselves are evolving to incorporate technologies such as Bluetooth, Wi-Fi, X10, ZigBee and TCP/IP.

As the market and technology matures, high-quality voice processing performance becomes increasingly important for home automation and security applications. Voice is an enabling technology that unifies the home network and is used to control appliances, telecommunication, security and entertainment equipment. End-users are also more comfortable communicating with a human voice, rather than interacting with a machine.

Poor acoustic echo cancellation, ambient noise and signal distortion make it increasingly difficult for a home automation system to perform reliably. If impeded by poor voice performance, voice recognition cannot easily detect commands to turn on/off the appliances and voice authentication has difficulty verifying the user to allow access to the residence.

### **Integration of Telephony and Intercom**

Home automation systems are increasingly integrating telephony communication and intercom functionality. In some system designs, the intercom panel becomes the main communication tool and serves as both the residence access monitor and a hands-free telephone. Therefore, the quality of voice communication becomes an important aspect of the overall system.

An important feature of a high-quality speakerphone is full-duplex operation and good double-talk performance. A full-duplex system allows two-way voice signals to pass simultaneously, enabling a natural conversation. A system with half-duplex fallback will only allow the signal with the highest power level to pass and

cuts off the other signal during double-talk. As a result only one side of the conversation can be heard and the conversation becomes choppy and unnatural.

In a full-duplex system, good acoustic echo cancellation is achieved when the algorithm continuously converges, even during double-talk situations. When the algorithm continuously converges it tracks changes in the echo path as the signal source moves. If the algorithm stops to converge during double-talk and resumes convergence when double-talk ceases, users will hear an audible burst of echo while the algorithm re-adapts to the new echo environment.

For digital speakerphones, another aspect of achieving a good quality call is the choice of codecs (coders/decoders) that perform the signal conversion from digital to analog and vice-versa. Traditional mid-quality digital phones use narrowband codecs and a sampling frequency of 8KHz. Wideband codecs enable higher speech quality by doubling the sampling frequency of narrowband codecs (16KHz vs. 8KHz) and capture more components of the human voice to enhance speech quality.

## DESIGNING A HIGH-PERFORMANCE SYSTEM

### Overcoming Physical limitations of plastic enclosures

Smaller intercom panels are now being designed to be less obtrusive as part of an overall home architecture. The restricted panel size creates limitations on the design of the plastic enclosure, the choice of microphones, speakers, their placement and separation.

In an ideal system the separation between speakers and microphones should be maximized to optimize the performance. In a small plastic enclosure, the reduced separation between the microphone and the speaker leads to excessive acoustic coupling and results in louder echo.

Designers are faced with the challenge of selecting smaller speakers to fit the limited space within the plastic enclosure while having to meet high volume audio requirements. The designer is then forced to drive a small speaker into a non-linear range. This results in an increase of total harmonic distortion in the plastic enclosure.

To solve this problem the designer needs a voice processing solution that can handle more distortion in the echo path to compensate for poor speaker performance. An algorithm that can cancel linear and non-linear echo allows the designer to drive the speaker volume higher while minimizing distortion. The non-linear echo canceller compares the residual echo signal – from the linear echo canceller – with the reference signal and subtracts the estimated frequency components.

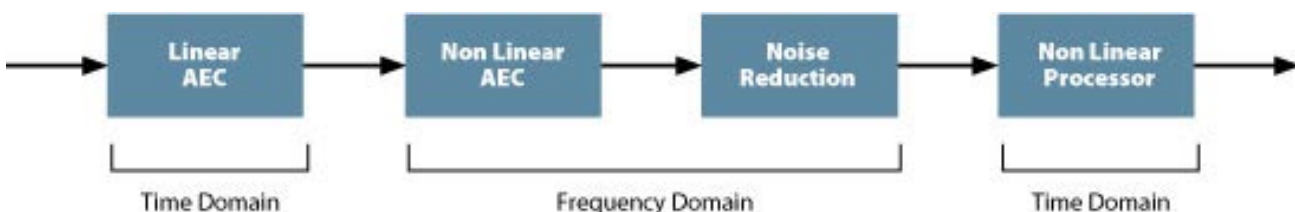


Figure 1: Block diagram of linear and non-linear echo cancellation.

Figure 1 shows a block diagram of linear and non-linear echo cancellation. This circuit removes linear echo and non-linear distortion components. It allows the system to overcome poor acoustic echo cancellation caused by distortion in the echo path.

If the designer is forced to use a non-optimized plastic enclosure design with a small speaker, a multi-band equalizer can be used in the speaker path to boost the performance. The equalizer attenuates low frequencies where excessive distortion is caused due to the poor frequency response of the small speaker at low frequency bands (Figure 2).

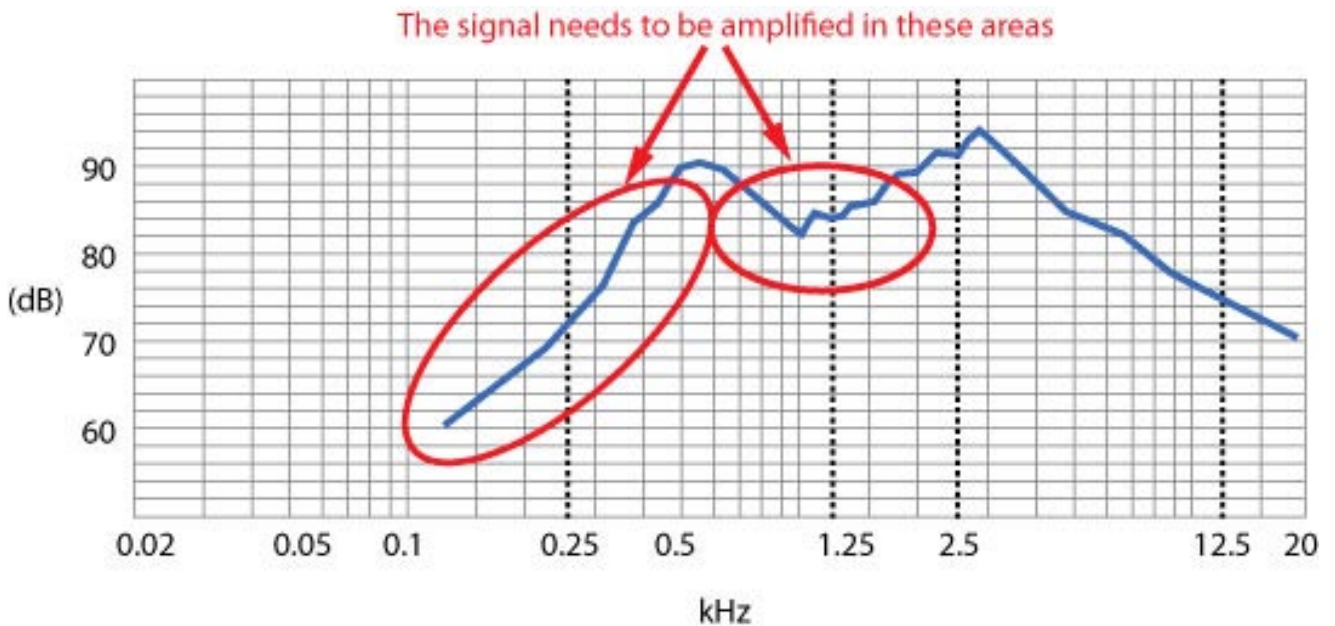


Figure 2: Typical frequency response of a small speaker. Small speakers attenuate lower frequencies. An equalizer placed in the speaker path can be used to amplify those frequency bands to improve voice quality.

A good acoustic echo canceller is measured by its ability to handle gain in the echo path, while also continuing to track changes in the echo source without falling back to half-duplex. Tolerating sufficient gain in the echo path (at least 10 dB), allows the designer to use a louder speaker and a more sensitive microphone. It also reduces the restrictions on plastic design by allowing more coupling.

### Overcoming Excessive Noise

Some designers believe dual microphone systems are necessary to achieve high performance. To maintain a low bill-of-material, minimize system cost, and simplify the system setup and algorithm tuning, the majority of designs today use a single microphone. The designer should find an algorithm that provides good noise reduction while minimizing distortion.

Psychoacoustic noise reduction provides improved noise reduction over traditional methods and allows the system designer to achieve high performance at a lower cost by using a single microphone solution.

Traditional noise reduction methods model noise across the entire signal on the frequency spectrum. The modeled signal is then removed from the total signal, resulting in reduced noise but also lower signal integrity.

Psychoacoustic noise reduction relies on the human perception of noise and cancels the noise elements of the signal that are more noticeable to the human ear. The algorithm differentiates between pure noise signal and a mixed noise with voice signal. It considerably attenuates noise components from frequency bands that are far from voice components and leaves noise components unattenuated from frequency bands that are near voice components. The unattenuated noise components near voice components are normally masked by voice, meaning the end-user can barely notice the noise. The Figure 3 diagram illustrates psychoacoustic noise reduction.

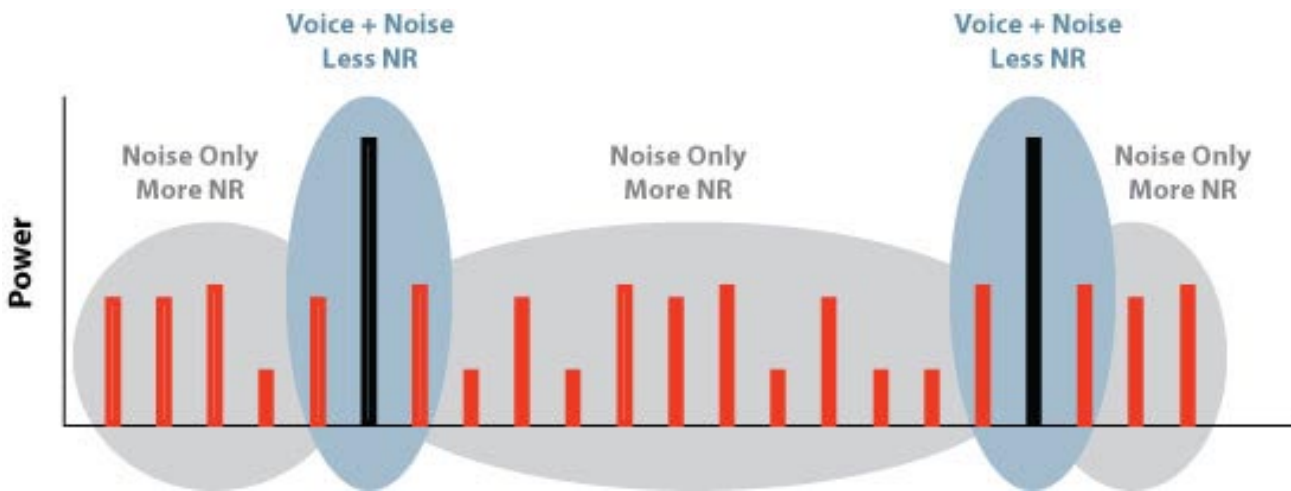


Figure 3: Psychoacoustic noise reduction cancels the noise elements of the signal that are more noticeable to the human ear.

## DESIGNING A SCALABLE SYSTEM

A good home automation system is scalable, allowing the end-user to build on the hardware as their needs evolve. It also allows the designer to develop a single platform and spin-off multiple variants to the system without having to redesign the hardware every time.

To design a scalable system, designers need programmable and field upgradeable voice processing solutions. These are key requirements to keep pace with the rapid evolution of systems and the continuous need for new features and enhanced performance. Figure 4 shows an example of a field upgradeable voice processing solution.

As home networking evolves and aims at higher adoption within the mass residential market, it becomes more reliant on voice technology to provide a user-friendly interface. A flexible and field-upgradeable voice processing solution enables future system upgrades with new features such as voice recording, verification, recognition, messaging and prompting.

Advanced home automation systems have an interactive interface and use voice/speech recognition and prompting to operate home appliances and provide a more sophisticated home security system.

Speech recognition converts acoustic signals captured by a microphone to a set of signaling commands. Speech recognition systems vary in complexity and can be characterized by mode of speaking (words vs. continuous speech), training (speaker dependant or independent) and vocabulary (small or large).

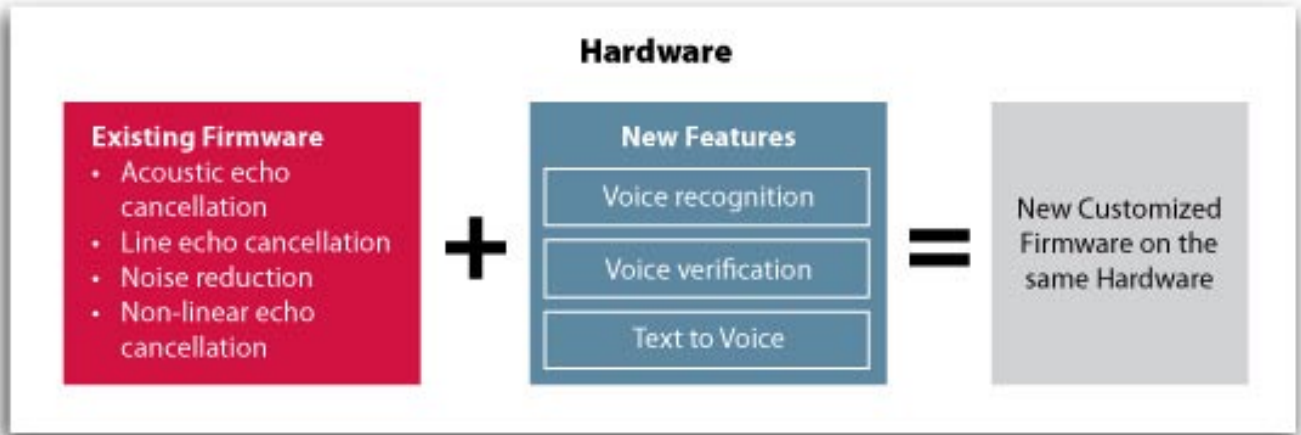


Figure 4: Example of a field-upgradeable voice processing solution that allows a designer to build on existing hardware to support a wider range of features.

Some systems require training, and adapt to the user's voice, style of speech and vocabulary to increase accuracy. These systems have a higher probability of predicting the right function over speaker independent systems. The trade-offs of a voice recognition algorithm are response time versus the size of the vocabulary and the overall quality versus the memory size and the speed of the processor. An algorithm that supports a large vocabulary takes longer to respond. A high-quality algorithm also requires more memory and processing power.

Voice prompting initiates the authentication process by prompting the user to say a phrase that has been programmed in the system. After verifying the user's voice, the system prompts the user to enter a password. The authentication process varies from system to system. Some systems rely on just voice verification techniques, however these systems are more vulnerable to break-in as voice verification alone can be recorded. Other systems rely on a combination of biometrics and user information verification to provide a higher level of security.

High-performance voice processing is critical to ensuring these systems work properly. Echo, ambient noise and distortion greatly impact the performance ability of voice recognition and verification applications.

## SUPPORTING HIGHER LEVELS OF INTEGRATION

As the functionality of terminals increases and the plastic enclosure becomes smaller and more sophisticated, integration becomes imperative. As designers look for increased functionality in a smaller footprint, saving board real estate becomes a high priority.

A programmable solution that integrates codecs and replaces external components with firmware features such as telephony signaling (DTMF, caller ID, tone generation, etc.) is critical for designing a good system. This type of solution replaces many of the standalone components used in system designs today.

Zarlink Semiconductor has developed a family of single-chip devices targeting high-performance hands-free communications, including home automation systems.

The company's newest hands-free solution, now commercially available, is a dedicated voice processor that combines integrated dual-channel wideband codecs with multiple interfaces. The device supports advanced echo cancellation, psychoacoustic noise reduction, full-duplex operation and is flexible and field-upgradeable. The chip's patented software algorithms continuously track changes in the echo path, even during double-talk conditions, and reduce background noise while preserving high voice quality.

The solution is based on a flexible platform that targets introductory systems with basic features and high-end systems with advanced features. It's a field upgradeable solution that allows the designer to build on platform designs with future system upgrades without re-designing the hardware.

## CONCLUSION

Once just a feature of luxurious high-end residences, home automation is now bridging to mainstream residences and will become a standard for new and existing homes.

Traditionally, home automation systems have included very basic voice processing techniques that provide half-duplex speakerphone performance. As these terminals integrate speakerphone functionality, and home security systems rely on voice verification and recognition technologies, high-performing voice processing solutions become a key element in home automation system design.

To maintain a low bill-of-material and achieve high performance, single-microphone systems need a voice processing solution with advanced noise reduction techniques (e.g. psychoacoustic noise reduction) that provide an improved performance over traditional noise reduction and limit distortion.

To achieve high performance in a small plastic enclosure, the designer has to drive a small speaker in a non-linear range to meet audio requirements. An algorithm that can cancel non-linear echo and handle gain and distortion in the echo path solves the designer's problem.

High integration, field upgradability and flexibility are also key criteria for a viable voice processing solution. Integration eases the design by reducing the complexity of interfacing multiple components and reduces the bill-of-material cost. Field upgradability and flexibility allow the designer to continuously enhance the feature set and add functionality without changing hardware.



[www.ZARLINK.com](http://www.ZARLINK.com)