VoiceTIMES™
Audio Hardware Guidelines
and
Signal Specifications

Version 1.0

Table of Contents

1 About This Document

The goal of this document is to define the hardware and audio signal characteristics that are required to perform reliable deferred speech recognition. The recommendations and details contained in this document should be referenced as guidelines when designing a recording device. If these guidelines are followed, the recorded audio will be suitable for processing by an Automatic Speech Recognition system (ASR) with acceptable accuracy results.

2 Terminology

The terms used to measure the acoustic or audio parameters affecting the ASR technology are discussed in this document.

2.1 Acceptable Accuracy

Acceptable accuracy refers to the expected speech-to-text conversion performance for an ASR system. This can be represented as a percentage of correct words versus incorrect words, or it can be represented as the error rate (percentage of incorrect words). In either case, the acceptable values or how precise the recognition needs to be is application and ASR system dependent.

2.2 Deferred Recognition

Deferred recognition refers to the overall process of recording speech and then using an ASR system to convert the recorded audio into its representative text. The details of the recording and conversion process to text will be application and implementation dependent.

2.3 Dynamic Range

The dynamic range is the difference between the highest and lowest amplitude portions of a detectible signal (or between the highest signal that a device can linearly handle) and the noise level of a device (usually expressed in dB). A wider dynamic range will yield better ASR results.

2.4 Frequency Response

The frequency response is the range of frequency over which the microphone sensitivity is nearly constant to within a specified value. Frequency response is usually described as a +/- dB value over a specified frequency range. A flat frequency response will yield better ASR results.

2.5 Microphone Sensitivity

The microphone sensitivity is the microphone signal amplitude (usually expresses in dB) for a given sound pressure level at a specified distance. Correct specification of microphone sensitivity is application dependent.

2.6 Noise Floor

The noise floor is the audio output level when no input signal is present. A low noise floor will yield better ASR results.

2.7 Signal-to-Noise-Ratio (SNR)

The SNR is the ratio of information carrying signals (speech) to background noise (expressed in dB). A higher SNR will yield better ASR results.

2.8 Sound Pressure Level

The sound pressure level is the expression of air pressure variations that are translated into electrical signals by a microphone. Zero dBspl is often defined as a sound pressure of 20 micropascal (20 micronewtons per meter squared). +94 dBspl corresponds to one Pascal.

2.9 Total Harmonic Distortion + Noise (THD+N)

The THD+N is the ratio of the amplitude (usually expressed as a percentage) of both the signal harmonics and the noise present in the output signal to the test signal amplitude. A lower THD+N will result in better ASR results.

3 Audio Hardware and Signal Specifications Overview

This section provides a general overview of the deferred recognition and its components.

3.1 Deferred Recognition Components

Figure 1 shows the components and data flow through a generic recording device.

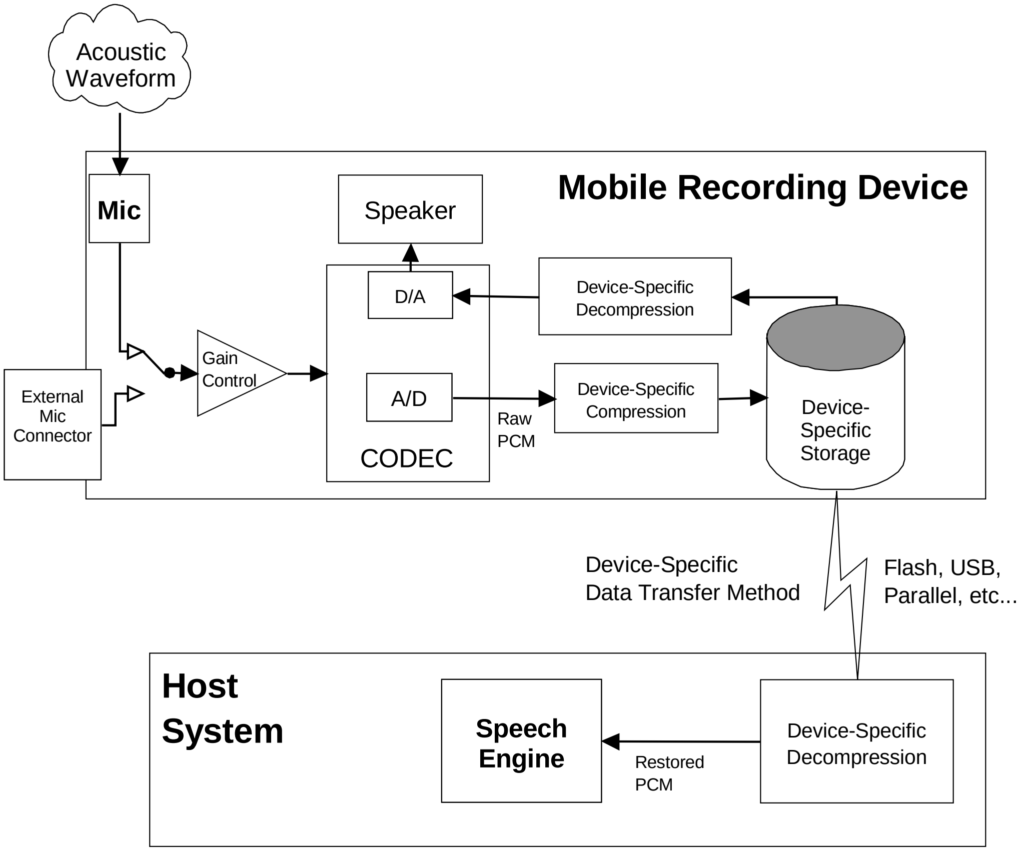

Figure 1: Deferred Recognition Data Flow

When considering the requirements to achieve accurate speech recognition results, the system must be considered as a sum of all the parts. The quality of the restored PCM (as illustrated in Figure 1) is the driving factor that influences the results of the deferred recognition. Therefore, this document provides guidelines and discusses the issues associated with the selection of each component. When the overall effects of all the components are combined, the resulting digital audio characteristics must meet the specifications provided to ensure accurate deferred recognition.

This document provides guidelines and recommendations pertaining to the different components of a generic recording device. It also provides the specifications for the restored PCM required by the ASR system. When these guidelines are followed and the signal quality of the restored PCM falls within the specified tolerances, an ASR system is able to accurately produce text from the recorded audio files.

3.2 Device Pass-Through Modes

A device pass-through mode is one in which a device is used for *live* input to the ASR engine. Figure 2 below shows the basic block diagram of a device configured for pass--through mode.

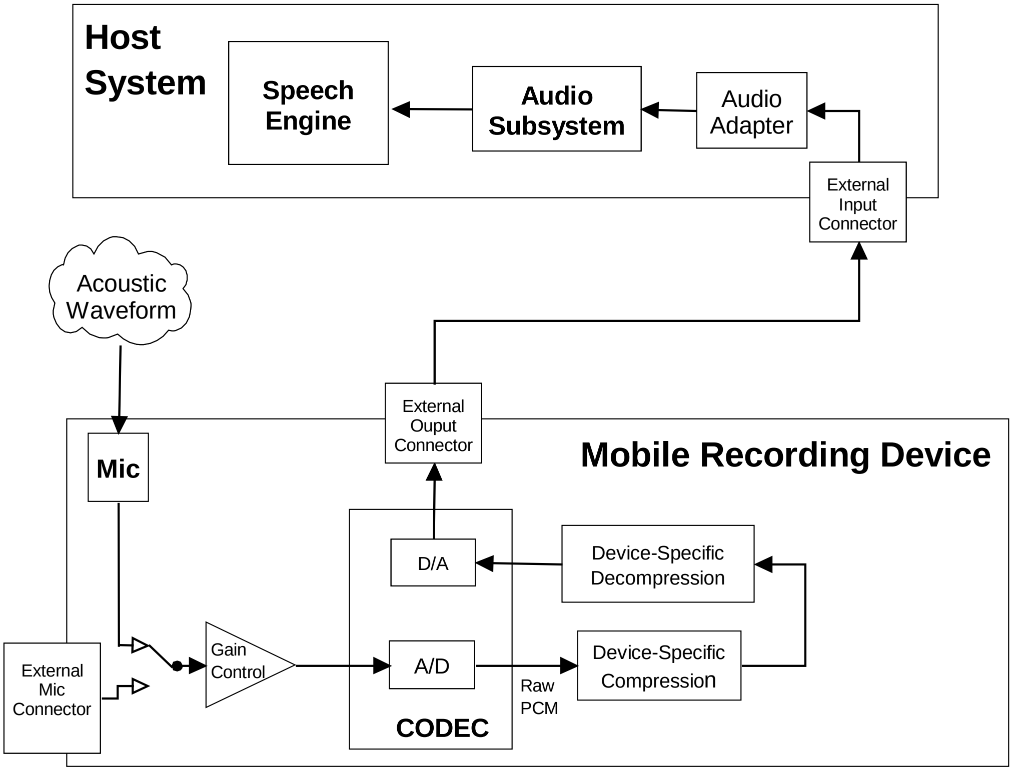

Figure 2: Device Configured for Pass-Through Mode

The types of tasks that require this functionality depend on the ASR implementation. The following is a list of possible scenarios representative of this mode of operation, along with a description of the problems associated with each.

3.2.1 Enrollment/Training Process

The Enrollment/Training process is designed to adapt the system to a specific user and the typical background noise environment. If a device is used as a microphone connected to the host machine for this process, the data being gathered may not be representative of the mobile environment unless specific steps are taken in the product design.

This product design must ensure that any audio the ASR receives is fully processed by all the device's circuitry including: microphone, gain, AGC, A/D, compression, and decompression. This ensures that any signal degradation present in a recorded file is also represented in the training data. The objective is to ensure that the enrollment process replicates the typical usage scenario.

3.2.2 Post-Transcription Correction and Adding New Words

During the editing process after a transcription session, new words may need to be added to the dictionary. At this time, the user may need to speak the word to train the speech recognition system. Some devices may choose to act as the microphone in this scenario and, therefore, will need to ensure the proper electrical connections to interface to a PC. The design must ensure that any audio the ASR receives is fully processed by all the device's circuitry including: microphone, gain, AGC, A/D, compression, and decompression. This ensures any signal degradation present in a recorded file is also represented in the training data. The objective is to ensure that the enrollment process replicates the typical usage scenario.

3.2.3 Providing Audio for Transcription via the Host Analog Input Connectors

If a device is unable to transfer recorded audio files to the host system, or the ASR system is unable to process the stored files, then transcription of the saved audio can still be accomplished if the device includes an audio output. Transcription can be accomplished by connecting the audio output of the device to the audio input of the host system. The speech recognition software is then started, and the user instructs the device to begin playing back the saved audio.

This implementation has many potential caveats. The primary issue results from the conversion of the original speech to and from analog to digital multiple times. A contributing factor to this problem is that the playback volume of the device and the input gain of the host system need to be coordinated to reduce the amount of distortion created during the transfer. A specific mode that provides a line level output (that is, not adjustable by the user) can eliminate concerns over volume control settings.

When considering a fixed gain approach, the goal should be to set the gain so that the peak level of the input signal is half of full scale when the device is used according to its specifications (normal speaking tone, expected distance, expected noise environment, etc.). This allows headroom for the peak signal to increase without clipping, which is a non-recoverable data loss condition. Even if these guidelines can be met in a lab environment, it is difficult to ensure the device will be used correctly. If the input signal does clip, the results will be detrimental to the ASR system's accuracy.

With each of the above scenarios, the device must have the correct connectors and be electronically compatible with the PC audio system. For deferred recognition, this device connection method is not recommended for transmitting the data to the speech recognition system because it is likely to degrade the signal and reduce recognition accuracy.

4 Microphone Considerations

This section discusses the considerations involved when selecting a microphone. Different microphones and microphone characteristics are required depending on the purpose of the device. For the purpose of this document, two modes of operation are considered: dictation mode and conference mode. Both modes are described in Section 5.1, Microphone Interface Circuitry.

4.1 Operating Mode

The input source for a hand-held device could vary depending on the intended use of the device at any given time. It is possible for a device operating in one mode to require completely different microphone parameters than the same device operating in another mode.

4.1.1 Conference Mode

Some devices are intended for use in a conference or meeting mode. In this case, two or more people may be gathered around a device located on a conference table with the intent to record the proceedings, no matter who is speaking and regardless of the positioning between the device and the people speaking.

This mode yields the best results when all the people speaking are roughly the same distance from the device and speaking at about the same volume. Speech and background noise are both recorded without bias or distinction. This mode is not recommended for ASR applications because of variations in the different people speaking versus training data and the person's relative position to the microphone.

4.1.2 Dictation (ASR) Mode

Large vocabulary ASR systems require user enrollment so that the software can adapt various data sets and acoustic models to a particular individual's manner of speaking. Continuous recognition systems are therefore speaker dependant; that is, only one person is the intended source and any other speech should be considered a component of the background noise.

4.2 Sensitivity

Microphone sensitivity should be selected to match the microphone input circuitry when used in the intended mode of operation.

The following sections explain the relationship between microphone sensitivity and other device characteristics or features.

4.2.1 Distance from the Source

One of the important design considerations to ensure the best performance of an ASR system is the intended distance from the person speaking to the microphone. This is because the sound pressure levels of speech at the microphone decrease as a function of the square of the distance between the source of the speech and the microphone.

Depending on the type of microphone used and the placement relative to the person speaking, small changes in distance can have profound affects on performance.

To determine the optimum distance, it is necessary to consider the type of microphone being used and the ambient (or background) noise.

4.2.1.1 Type of Microphone

The type of microphone can determine the optimum distance from source to the device. For the noise cancellation techniques to be effective, the microphone must be used within approximately 12 to 40mm (.5 to 1.5 inches) of the speaking person's lips. Noise canceling microphones are also known as *close-speaking* microphones. Other types of microphones can be used at different distances. The distance between the source and the microphone depends on the following parameters:

- How loudly or softly the person is speaking
- Microphone sensitivity
- Acoustic background noise typical of the intended use
- Product package design

4.2.1.2 Background Noise

All ASR systems work best with a large signal-to-noise ratio (SNR). One way to obtain good SNR characteristics is to manage the background noise. In general, an effective technique to manage background noise is the use of a noise-canceling microphone. A disadvantage of noise-canceling microphones is that they are very sensitive to the location of the source of the speech. Therefore, they may not be the best solution for a hand-held device. See Section 4.4.4, Bi-directional (Noise Canceling) Microphones for a description of these issues.

If a noise-canceling microphone is not desirable because of one of several constraints, an alternative strategy could be employed to reduce the effects of the acoustic background noise on ASR performance. Reducing the microphone sensitivity allows the user to move the microphone closer to the their mouth to take advantage of the higher signal levels present. Lowering the microphone sensitivity and moving the microphone closer to the source of the signal reduces the level of background noise and increases the source signal level. This results in a higher SNR and better performance in a high-noise environment.

4.3 Omni-directional Microphones

Omni-directional microphones work best for sources that may be in any direction relative to the recording device. Conference mode applications often benefit from omni-directional microphones.

These microphones are not recommended for use with ASR. The drawback of an omni-directional microphone is that sound is picked up from all directions making it difficult to distinguish between the source of interest and background noise.

4.4 Unidirectional Microphones

This section discusses unidirectional microphones, which fall into one of the following main categories: cardioid, supercardioid, hypercardioid, and bi-directional/noise canceling.

4.4.1 Cardioid Microphones

A cardioid microphone is more sensitive to sound sources in front (on axis) of the microphone and is much less sensitive to sources behind the microphone. Since the source of interest is normally located in front, a cardioid microphone is much better at discriminating between the source of interest and background noise located behind the microphone compared to an omni-directional microphone.

In most cases, the more directional a microphone is, the better the signal-to-noise ratio will be; assuming that the person speaking is properly positioned relative to the microphone.

4.4.2 Supercardioid Microphones

A supercardioid microphone is similar to a cardioid microphone, but the off axis sensitivity is lower, while the rear sensitivity is slightly higher.

4.4.3 Hypercardioid Microphones

A hypercardioid microphone offers the most directional (on axis) sensitivity, but has a higher sensitivity to background noise from behind than either a supercardioid or cardioid microphone. A hypercardioid microphone works best when noise sources are from the side.

4.4.4 Bi-directional (Noise Canceling) Microphones

A bi-directional microphone is used to employ noise-canceling characteristics. It is equally sensitive to sound arriving at the microphone from the front or back. In this case, the sound from the front is 180 degrees out of phase with the sound from the back. When sounds arrive from both directions (which are equal in all aspects, but opposite in phase), they cancel each other, resulting in very low microphone sensitivity. Since far-away noise sources arrive at both sides of the microphone at nearly the same time, they are canceled out. This type of microphone is referred to as a *noise-canceling* microphone.

A bi-directional microphone is also referred to as a *close-speaking* microphone because it must be used within approximately 12 to 40mm (.5 to 1.5 inches) of the speaking person's mouth to be effective. If used farther away, the user’s voice is canceled in the same manner as the background noise.

This type of microphone is very effective in getting good signal-to-noise ratios in the presence of medium to high noise levels. The drawback of a bi-directional microphone is a very high sensitivity to its position relative to the person's mouth. Therefore, a bi-directional microphone is usually associated with headset products and should not be employed when a stable, close position relative to the speaking person cannot be guaranteed; that is, most users prefer to hold a device 50 to 75mm (2 to 3 inches) from their lips and are not able to maintain a constant distance.

4.5 External Microphone Input

An external microphone connector allows a user-selected microphone—possibly a close-talking head set microphone—to be attached to a device. A headset with a bi-directional microphone yields a higher quality signal than that possible with a microphone built into the device.

4.5.1 Microphone Jack

The most widely-accepted definition of an external microphone jack for PC-like devices is contained in Chapter 17, of the *PC-99 System Design Guide*, available at: http://www.intel.com/design/desguide.

5 Analog Signal Processing Considerations

Analog signal processing encompasses the circuitry between the microphone and the digital-to-analog conversion in the CODEC.

5.1 Microphone Interface Circuitry

The purpose of the interface circuitry is to:

- Satisfy the connectivity and power requirements of the microphone.
- Provide signal amplification of low-level microphone signals at the first opportunity to minimize exposure of low-level signals to contamination caused by electromagnetic interference from other sources.
- Provide adequate signal amplification to match the input signal requirements of the CODEC circuit.
- Provide gain adjustment (input volume control) in order to provide an adequate SNR without saturating the CODEC input. Please see Section 5.2, Automatic Gain Control.

It is important that the interface circuitry be designed to match the characteristics of both the microphone and the CODEC circuitry over the intended range of operation. Note that the microphone amplification circuitry could be internal to the CODEC, or it could be a separate device. In either case, the characteristics of the circuitry must match the components being used.

5.2 Automatic Gain Control

Recording devices often employ a method of automatically optimizing the amount of gain applied to the audio input signal based on the signal level being monitored at a specific part of the audio circuit. This technique is commonly known as *Automatic Gain Control* or AGC.

Simple AGC functions well enough to provide a simple level of optimization, but lack the sophistication to properly optimize the gain for ASR applications. Most AGC circuits have the following drawbacks, which affect ASR applications.

- Most AGC circuits do not distinguish between speech and silence. The result is that when a person pauses in the middle of speaking, the AGC misinterprets the silence as a low signal and increases gain accordingly. When the user resumes speaking, the first utterance is over-amplified. This condition continues until the circuit again adjusts to the speech signal level. The time required to readjust and the number of words affected are determined by the controlling time constants of the particular AGC. Over-amplification usually leads to a condition called *clipping*, or the saturation of the CODEC input, which causes poor ASR performance.

- The AGC adjustments are controlled by time constants built into the system software and hardware. If the time constant is too short or too long, the AGC adjustments can cause an interaction with the dynamic adjustments in the ASR software and result in poor ASR performance. AGC time constants on the order of 3 to 5 seconds are recommended.

Products that utilize a correctly-designed AGC can function properly over a wider range of sound pressure levels than would be possible for the same product equipped with only a fixed audio gain.

6 CODEC Considerations

A CODEC provides a conversion from analog (microphone) inputs to digital signals for later processing by the ASR software. Playback of the digitized signals (digital-to-analog conversion) is also provided along with various other features.

Several types of CODECs are available. However, not all types are well suited to ASR. In general, Telephony CODECs—often distinguished by 8 kHz sampling rates and bandwidth limiting—are NOT recommended for ASR applications. CODECs designed for personal computer applications or consumer devices (for example, portable tape recorders) are more likely to include the features and specifications that are required for high-accuracy ASR.

Delta-sigma converters that utilize over-sampling to minimize quantitative errors are preferred for ASR applications. Since these types of devices run the ADC at a frequency several times the base sampling rate, some thought must be given to power management issues.

There are several interface alternatives to connect the CODEC to system processing elements. Parallel interfaces and proprietary serial connection schemes may be used. To provide for maximum flexibility, CODECs with I2C or I2S interfaces are recommended. Many micro-controllers offer on-board I2C or I2S interfaces that simplify the hardware connections to CODECs and allow various types of devices to be supported by a common hardware platform. In addition to data transfers to or from the CODEC, the interface is used to access registers within the CODEC to control functions, such as power management, gain controls, source selection, and so on.

By their very nature, CODECs straddle the analog and digital sections of the circuit board. This requires some degree of care in order to prevent spurious noise from degrading the audio signal and/or the system's EMI performance. Specific design recommendations are generally available from the CODEC manufacturer. Following those recommendations closely yields optimum audio performance and generally results in the best available ASR accuracy.

Presently, the typical portable device has a restricted range of function that limits the number of inputs and outputs. Therefore, mixers (other than simple source selection switches within the CODEC) are beyond the scope of this document.

6.1 Analog input

Many CODECs include inputs for either direct microphone attachment or a line-level input source. The choice of inputs is determined by the physical properties of the device.

Line-level input sources are characterized by a maximum voltage swing of 0 to +1 volts. Line-level inputs are used when the microphone pre-amp and amplifier stages are in separate circuits within the system. This approach is preferred when the physical design of the product prevents the CODEC from being located close to the microphone. In that event, the signals should be amplified first and then routed through the product as line-level signals that are less prone to the introduction of noise.

Microphone inputs generally include a 20dB boost preamp and an additional, programmable stage that should include at least 10 steps of 1 or 1.5dB per step. AGC software can then use the programmable stage to provide a consistent level into the analog-to-digital conversion elements to improve SNR and prevent clipping.

6.2 Resolution

The ASR software assumes a 16-bit linear analog-to-digital converter for optimum recognition accuracy. This is based on several assumptions related to circuit performance, background noise, and so on, which could allow for a lower-resolution CODEC being employed in circuits that have demonstrated very low THD+N. In general, these techniques are not recommended for speech data collection where recognition accuracy is important.

The 16-bit samples are also assumed to be linear; that is, 1 bit equals 6 dB for a total theoretical range of 96 dB. Lower resolution devices with U-Law, A-law, or other built-in compression schemes intended to approximate a 16-bit linear sample are not recommended.

In some cases an audio memo application with the ability to temporarily select a lower resolution and/or a compression scheme could be useful to minimize storage space for speech recordings where ASR is not likely to be employed.

6.3 Sampling Rate

The minimum sampling rate recommended for digitized speech is 11.025 kHz for continuous ASR applications. CODECs limited to sampling rates 8 kHz (generally Telephony-oriented devices) are not recommended for ASR applications, but are still supported with reduced accuracy for the transcribed text. Over-sampling—in particular the Delta-sigma technique—is recommended to maximize the SNR within the band of interest.

In some cases, for instance an audio memo application, the ability to temporarily select a lower sample rate could be useful to minimize storage space for speech recordings where ASR is not likely to be employed.

6.4 Filtering

The most suitable primary frequencies for ASR are 100 Hz to 8 kHz. Low-pass, band-pass and high-pass filters must be selected with care to minimize the effects within the desired band, flat to +/- 1dB. CODECs typically have provisions for the required filters built into the device with pins for the connection of discrete components, in some cases to complete the circuit. Discrete, external filters are generally not required unless all of the microphone amplification stages are outside of the CODEC.

Delta-sigma converters require several stages of filtering to yield an accurate digital representation of the analog signal. The anti-aliasing filter should be built into the device with pins to attach appropriate resistors and/or capacitors, as per the manufacturer's guidelines. Filters to remove quantitative noise and decimation of the digital output should be provided in the CODEC.

7 Audio Signal Processing Considerations

This section describes the techniques that can be applied to the digitized audio signals and how they affect the speech recognition process.

7.1 Noise Cancellation (Reduction) Techniques

Simply stated, noise cancellation and noise reduction techniques are methods of reducing the effects of audio information from all sources other than the desired source. The goal of these techniques is an enhanced SNR. An improved SNR with negligible increases in THD usually translate into better ASR performance.

7.1.1 Acoustic Noise Cancellation

Several noise cancellation techniques are used at the microphone element. All use a similar approach and are known as *bi-directional microphones*. For information on acoustic noise cancellation, see Section 4.4.4, Bi-directional (Noise Canceling) Microphones.

7.1.2 Adaptive Noise Reduction (ANR)

For the purpose of this document, Adaptive Noise Reduction (ANR) is defined as a technique used to enhance the SNR of an audio signal by passing the digital audio data through a sophisticated filtering algorithm. The effectiveness of this technique depends to a large extent on the audio content and the design of the algorithm. In most cases, signal-to-noise is noticeably improved, but recognition accuracy is not. In these cases, ANR improves the SNR at the expense of a higher THD+N. In some cases, this type of compromise has unpredictable results; where ASR performance is not enhanced, but actually degraded.

7.1.3 Multiple Microphone Array

Multiple microphone arrays can be employed alone or with sophisticated software algorithms, which process the digital audio information from the microphone array. The purpose of the array is to improve the SNR in one of two ways.

The microphone array can be designed to produce a narrowly-focused beam of high microphone sensitivity that is directed at the source of interest and excludes all other audio sources. Alternatively, the opposite effect can be created using a null beam of low sensitivity that is directed at the source of noise thereby reducing its effects.

7.2 Speech Detection

Speech detection occurs when a device decides whether to record audio data based on whether the data is considered to be silence or speech. This is also referred to as *voice-activated recording*, *silence detection*, or *silence compression*. This technique increases the perceived amount of audio a device can store because silence is ignored and not stored

For deferred recognition, these techniques can reduce the overall speech recognition accuracy under the following conditions:

- Words following silence can be clipped at the beginning while the device determines the signal is no longer silence and starts the recording process.
- If the algorithm is too aggressive, it can clip the trailing end of words that end in soft sounds.

To avoid accuracy degradation, silence detection schemes should allow at least 0.2 seconds of silence at the beginning of a recorded phrase and 0.5 seconds of silence at the end of a recorded phrase. Also, the transitions of the recorded audio when stopped and restarted should be smooth and not contain any spikes or artificial audio characteristics introduced by the circuitry or algorithm

8 Digital Data Compression Considerations

Many devices implement some form of digital audio data compression to increase the amount of audio data that can be saved on the device. In order for this data to be processed by ASR systems, these implementations require the compressed data to be restored to the PCM format expected by the ASR system. Depending on the algorithm used, the compression/decompression process could degrade the signal by adversely affecting the THD+N, thus reducing the accuracy of the ASR system.

When selecting a compression algorithm, it is required that the restored PCM data meet the signal requirements specified in Section 9 Restored PCM Specifications.

9 Restored PCM Specifications

The restored PCM is the data that is actually being input to the ASR system. This data contains the cumulative signal degradation of all the components of the deferred recognition data path. (See Figure 1: Deferred Recognition Data Flow.) The following table defines the boundaries for the signal characteristics for the restored PCM.

Parameter	**Minimum**	**Recommended**	**Maximum**
Signal-to-Noise (SNR)	15 dB	25 dB	N/A
Sampling Rate (Samples/Sec)	8 K	(see notes below)	22.05 K
Frequency Response (+/-5 dB):			
8 k Samples per sec	150 to 3.4 kHz	100 to 3.4 kHz	100 to 4 kHz
11.025 k Samples per sec	150 to 4.7 kHz	100 to 4.7 kHz	100 to 5.5 kHz
22.05 k Samples per sec	150 to 7.5 kHz	100 to 8 kHz	100 to 11 kHz
Dynamic Range, Usable Signal	10 Bits	12 Bits	16 Bits
THD+N @ 1Khz	N/A	< 1%	< 3%

Notes:

The usable signal range refers to the nominal speech signal at the analog-to-digital converter input. Background noise, speech amplitude variations, and AGC influences are not included in the value. To achieve excellent accuracy results, a 16-bit linear CODEC is recommended (given typical environments, noise floors, and AGC implementations).

The recommended sampling rates are 22.05kHz, 16kHz, and 11.025kHz. These are the most common sampling rates accepted by ASR systems. In general, higher sampling rates improve ASR accuracy and in particular for users with higher frequency content. For example, analog-to-digital converters employing 22.05kHz sampling rates better represent the voices of women or children.

The sampling rate parameter of an ASR system is implementation dependent. Therefore, the audio data generated by a device may need to have its sampling rate modified in order to meet the requirements of a specific implementation. This conversion can be done in software and requires that the appropriate filters be applied to ensure the resulting data is representative of the target sampling rate.

10 EMI Considerations

Recording devices are subject to the same EMI requirements as PC and consumer hand-held electronic devices. Standard circuit board and enclosure design practices intended to minimize radiated emissions should be followed. For example, grounded guard traces or clear channels around all clock signals, high frequency bypass capacitors close to the power supply pins of all digital modules, and so on.

Because recording products include analog and digital sub systems (often with separate power and ground circuits), care must be taken to avoid creating radiating antennas. Generally, this involves encapsulating high-frequency digital signals within digital ground structures. High-frequency signals should not switch between digital and analog ground return paths or pass within 1 mm of sensitive analog signals, such as microphone inputs. CODEC and related analog module manufacturers typically provide application notes to assist in minimizing local interference (for example, cross talk), and to describe suitable grounding schemes that yield acceptable analog performance in a primarily digital environment.

Cables for optional external microphones, headphones, and host system interconnects represent another potential radiated EMI source. Proper grounding—in addition to low-pass filtering at each connector—help to minimize high-frequency energy from coupling onto cables. Connector shields that would nominally attach to cable shields should generally not be connected directly to digital ground to prevent system switching noise and clock harmonics from coupling onto the cable shield. Provisions should be made in the design to allow for isolating or blocking high-frequency paths to the connector shields from digital ground. Provisions should also be made for low-pass filtering or high-frequency blocking on all signals that could be carried by an external cable. Note that some industry standard interface specifications may limit the designer's freedom to filter specific signals.

Cables—the microphone input in particular—also provide a means for externally-generated noise to couple into the device and interfere with its intended operation. The cable management techniques above also help to curtail sources of potential interference. The microphone input cable—because of the low signal levels generated by microphones and the higher levels of amplification within the device—represent an ideal entry point for external noise. Where possible, shielded cables should be employed with low-pass filtering at the location where the cable connects to the system electronics board. Un-amplified microphone input circuit traces should be shielded by analog ground traces and/or have 1 mm clear channels adjacent to the signals. Clear channels are also recommended around control signals with high-impedance inputs (for example, device push button or switch inputs and resets) to prevent cross talk that could generate false edges.