Beamforming in Acoustic Echo Removal and Noise Reduction

Contents

1 Definitions

Acronym	Description
AIC3204	TLV320AIC3204 Ultra Low Power Stereo Audio Codec
ADC	Analog to Digital Converter
AER	Acoustic Echo Removal
ASNR	Adaptive Spectral Noise Reduction
ASRC	Asynchronous Sample Rate Converter
BF	BeamForming
CAF	Connected Audio Framework
CCS	Code Composer Studio
C5517	TMS320C5517 Digital Signal Processor
dBA	A-Weighted Decibels
dBFS	Decibels relative to Full Scale: in this document, a full scale signal is a 16-bit PCM sin wave
dBSPL	Sound Pressure Level
DMA	Direct Memory Access
DNS	Delay-and-Sum
DRC	Dynamic Range Compressor
DSP	Digital Signal Processor
EMIF	External Memory Interface
EQ	Transducer Equalization
ERL	Echo Return Loss
EVM	Evaluate Module
GPT	General Purpose Timer
HAL	Hardware Abstraction Layer
HID	Human Interface Device
I2C	Inter-Integrated Circuit
I2S	Inter-IC Sound
LMS	Least Mean Squares
MEMS	Microelectromechanical Systems
NLP	Nonlinear Processor
OS	Operating System
PBN	Push Button Network
PCM	Pulse-Code Modulation
PDM	Pulse-Density Modulation
PLL	Phase Lock Loop
PWM	Pulse-Width Modulation
SAR	Successive Approximation ADC
SDRAM	Synchronous Dynamic Random-Access Memory
SLR	Send Loudess Rating
SNR	Signal-to-Noise Ratio
SRAM	Asynchronous Static RAM
THD	Total Harmonic Distortion
UART	Universal Asynchronous Receiver-Transmitter
USB	Universal Serial Bus

2 Introduction

TI's AER software [1] minimizes acoustic echo that is generated by the coupling between the speakers and microphones of a phone or similar device, particularly in hands-free mode. The primary purposes of AER are:

- Canceling acoustic echo so that it is not perceptible.
- Providing full duplex operation.
- Controlling nonlinear distortion that is common at high speaker volumes.
- Noise reduction and control that preserves speech quality.

AER release 17.0 includes an optionally enabled BF feature located prior to the AER transmit side input, as shown in Figure 1. The BF provides spatial filtering, or discrimination of different signals based on the physical location of their sources, using the inputs from a microphone array.

There has been an increasing use of VoIP technology for personal and business communications in recent years. VoIP-based communications services such as Skype provide integrated communications services like audio, video, and text at a relatively low cost. The Skype client software can be used on any PC with a speaker and microphone, and can also be installed on smart phones and tablet computers. The speakers integrated into laptop PCs, smart phones, and tablets, however, are frequently low quality. Call quality can suffer when these speakers are used for hand-free communication. To address this problem, several companies have developed portable speakerphones containing higher quality speakers than those typically integrated into laptops, smart phones, and tablets. Different models exist, with some connecting to the host hardware over USB, and others connecting wirelessly over Bluetooth. Examples of such speakerphones include the CHAT70 by ClearOne (USB), the Yamaha PSG-01S (USB), and the Motorola TX500 (Bluetooth).

This white paper presents an investigation of using BF to assist AER in near-end noise suppression and acoustic echo cancelation. The investigation was conducted on a portable USB speakerphone test platform utilizing the AER software. The speakers and microphones selected for the speakerphone are typical of those found in portable speakerphones currently on the market.

3 AER Software Overview

The AER software includes various signal processing functions which together provide echo cancelation, noise reduction, and volume control. A comprehensive list of these functions is given in [2]. Not all functions provided by the AER software were utilized in the development of the USB speakerphone; a block diagram showing those functions which were used is presented in Figure 1.

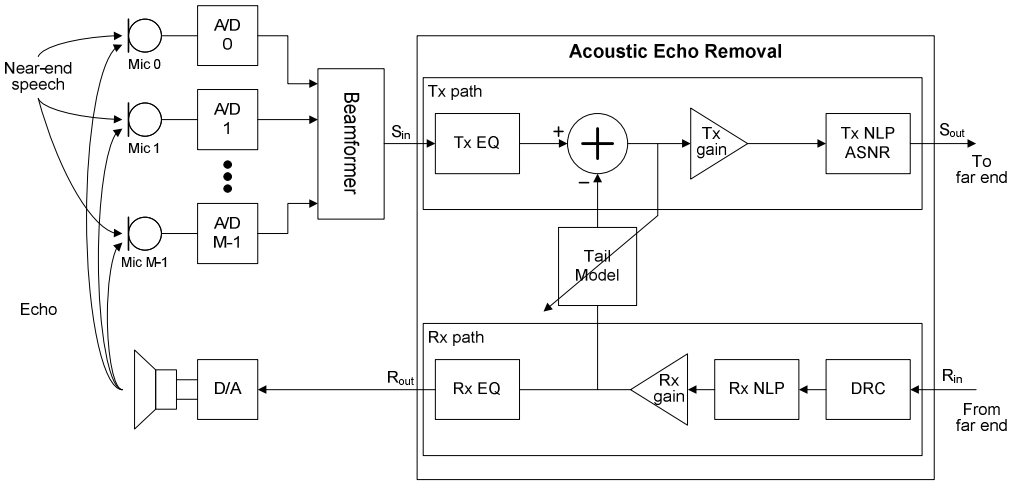

Figure 1. AER/BF Software

The functions shown in Figure 1 include the following:

- Adaptive Spectral Noise Reduction (ASNR): designed to provide SNR improvement in noisy operating environments.
- Dynamic Range Compressor (DRC): designed to provide control of the dynamic range of a speech signal. The DRC may be used to increase the loudness as well as limit excessive signal peaks.
- Transducer Equalization (EQ): designed to correct for minor deficiencies in a transducer frequency response or the overall frequency response of the phone.
- Acoustic Echo Canceller (AEC): designed to cancel acoustic echo and maximize the potential for full-duplex communication. This function includes the following sub-functions:
    - Adaptive filter: implements the normalized LMS in the frequency domain.
    - Bi-directional nonlinear processor (NLP): both the Rx and Tx direction have a nonlinear processor. In the Tx direction, the NLP follows the echo removal by the adaptive filter and is composed of a center clipper and a linear gain.
    - Digital gain units: helps achieve the appropriate loudness rating. These are separate gains from those used within NLP.
- Beamformer (BF): designed to provide spatial filtering using the outputs of a microphone array.

The AER source code is highly portable, and easily adapted to different target architectures. AER is delivered as pre-compiled libraries with C header files, except that source code is delivered for BF. Multiple targets are currently supported, including TI C55x, C64x+/C674, and ARM Cortex-A8.

3.1 BF Software Details

Beamforming is a signal processing technique that uses phased sensor arrays to achieve directional signal transmission or reception. When used for signal reception, beamforming processes the outputs from a sensor array such that the signals incident on the array from desired angles are constructively combined, while those from undesired angles are combined destructively. A basic formulation of beamforming applied to audio signal reception using microphone arrays is presented in [3].

Figure 2 depicts the wave fronts of a plane wave arriving at a four-element microphone array. As shown, the origin of the coordinate system is midway between microphones 1 and 2, the distance between the microphones is denoted by *d*, and the angle of arrival of the plane wave with respect to the *x*-axis is denoted by θ.

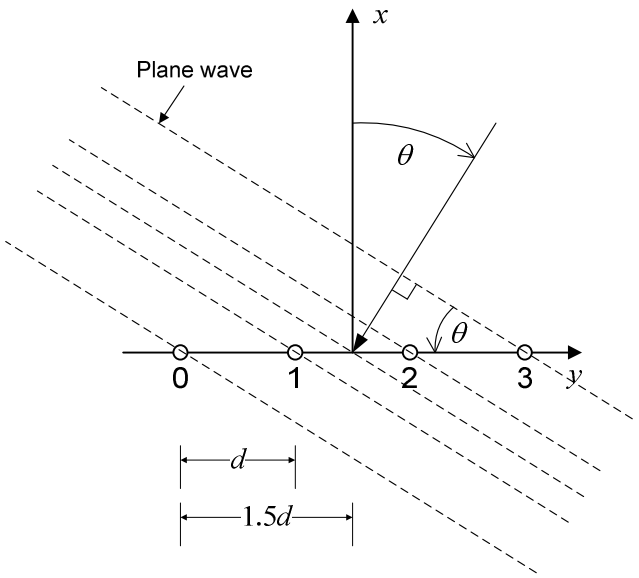

Figure 2. Plane Wave Arrival Times for Microphones in Array

If the speed of sound is represented by *C* and the sampling frequency by f_s, the arrival time difference (delay) of the plane wave between the origin and each of the microphones is given by

$$\tau(d,\theta) = \begin{bmatrix} 1.5 & 0.5 & -0.5 & -1.5 \end{bmatrix} \frac{d \sin(\theta)}{C} f_s \qquad (1)$$

where $\tau(d,\theta)$ is measured in fractional samples. For example, for d = 0.02 m, θ = 60°, C = 340 m/s, and f_s = 16 kHz, the delay vector is give by:

$$\tau = \begin{bmatrix} 1.223 & 0.4075 & -0.4075 & -1.223 \end{bmatrix}$$

The BF function in the AER software implements a Delay-and-Sum (DNS) type beamformer. The DNS beamformer looks in a particular direction θ by: (1) delaying the microphone output signals by $-\tau(d,\theta)$; and (2) summing the delayed signals. By delaying the microphone output signals by $-\tau$, the DNS beamformer aligns the signals for "steering" angle θ; by summing the delayed signals, the DNS beamformer coherently combines the microphone outputs. This is illustrated in Figure 3, where the required delays for microphone *m* are introduced by the filter $g_{D,m}$. Shown in Figure 4 is the case of an interfering signal arriving from an angle other than the steering angle θ. In this case, the microphone signals are unaligned after application of the delays, and the microphone output signals are not coherently combined. Note the summing node output in Figures 3 and 4 is normalized to the output of a single microphone by dividing the summed output by the number of microphones in the array M.

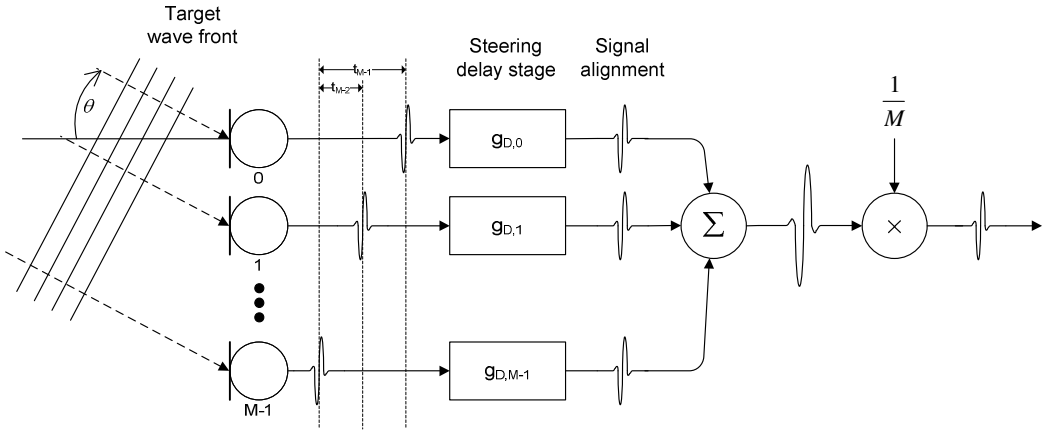

Figure 3. DNS Beamformer – Aligned Signals for Steering Angle θ

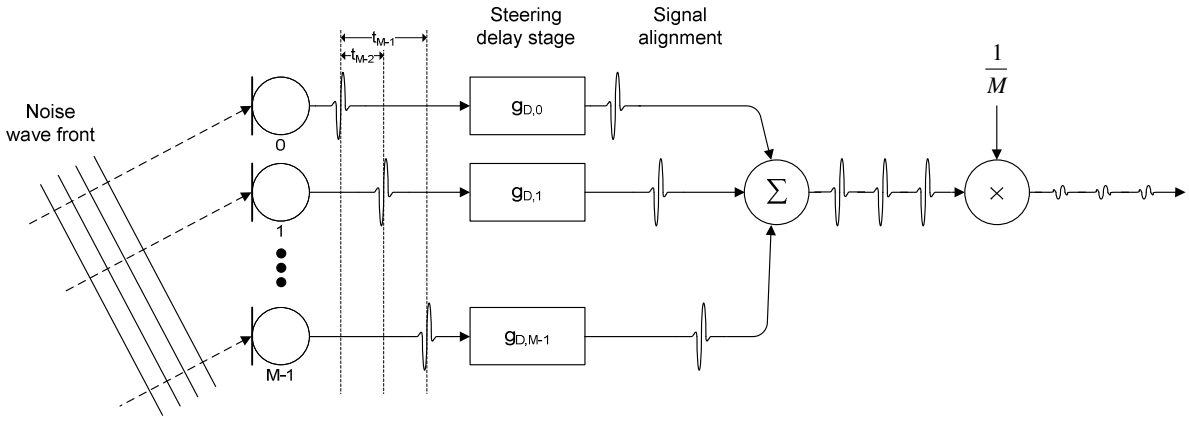

Figure 4. DNS Beamformer – Unaligned Signals for Steering Angle θ

A common method of implementing the fractional sample delays of the filters $g_{D,m}$ is the use of delayed reconstruction filters. The delay for each filter is separated into integer and fractional portions, with the fractional portion in the range [-0.5, 0.5]. The integer portion of the delay is handled using unit delays, and the fractional portion using the delayed reconstruction filter:

$$h(n) = \frac{\sin\left(\pi\left(n - \tau_{m,f}\right)\right)}{\left(n - \tau_{m,f}\right)} \tag{2}$$

where $\tau_{m,f}$ is the fractional delay associated with $g_{D,m}$. The delayed reconstruction filters are windowed (e.g. using a Hanning Window) and delayed by half the filter length to make them realizable:

$$h(n) = w(n - \tau_{m,f}) \frac{\sin\left(\pi\left(n - \tau_{m,f} - (N-1)/2\right)\right)}{\left(n - \tau_{m,f} - (N-1)/2\right)}, \quad 0 \le n < N \tag{3}$$

The filters $g_{D,m}$ are implemented using an FIR structure.

An omni-directional microphone is equally sensitive to signals from all angles. This is shown in Figure 5, which depicts the spatial response of an omni-directional microphone. In contrast, the DNS beamformer is more sensitive to signals arriving at the microphone array from the beamformer steering angle. The spatial response of a DNS beamformer to a 4250 Hz tone for the case of M = 4, d = 2 cm, θ = 0° is shown in Figure 6, where is can be observed that the beamformer is most sensitive to a signal angle-of-arrival which matches the steering angle. The spatial response of the DNS beamformer for M = 4, d = 2 cm, θ = 0° and F_s = 16 kHz is presented in Figure 7. This response shows the beamformer has low spatial selectivity for frequencies below 1 kHz, and is most sensitive to frequencies arriving from the steering angle.

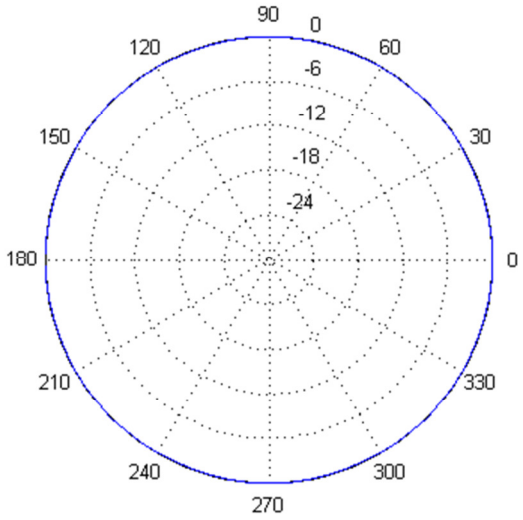

Figure 5. Spatial Response of Omni-Directional Microphone

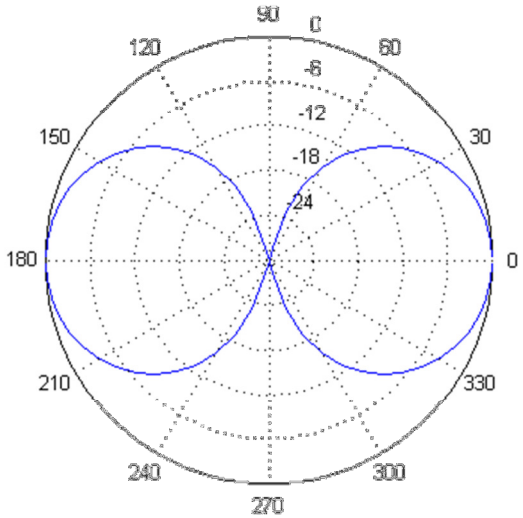

Figure 6. Spatial Response of a Beamformer for M=4, d=2cm, θ=0°, F=4250 Hz

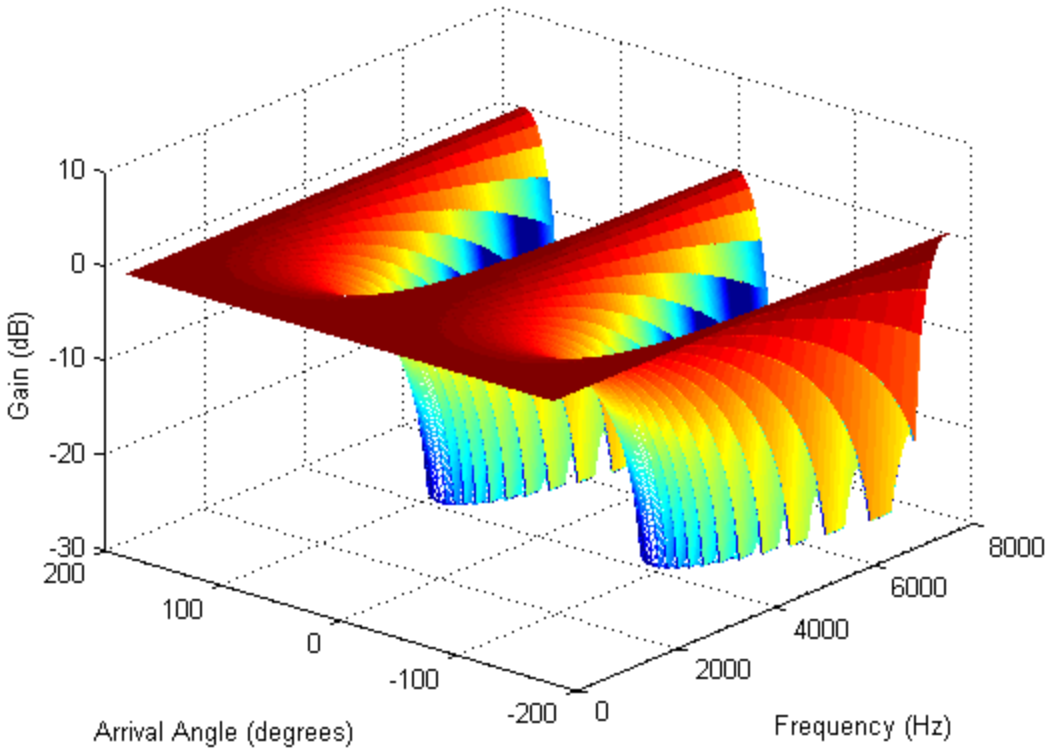

Figure 7. Spatial Response of a Beamformer for M=4, d=2cm, θ=0°

The BF function in the AER software implements a fixed, non-adaptive DNS beamformer with a configurable steering angle. The filter coefficients for a given steering angle are computed offline using a design tool packaged with the AER software. The BF API allows runtime reconfiguration of the steering angle.

4 USB Speakerphone Overview

A portable USB speakerphone test platform was used to investigate possible improvements in (1) noise reduction when using BF with ASNR, and (2) echo cancelation when using BF with AER. The speakers and microphones selected for the speakerphone are representative of those found in portable speakerphones currently on the market. This section presents an overview of the speakerphone hardware and software design.

4.1 USB Speakerphone Hardware

A block diagram of the USB speakerphone hardware is presented in Figure 8. The major hardware blocks used in the speakerphone include:

1. C5517 EVM – C5517 Evaluation Module
2. MKARRAY – 1-D linear microphone array containing eight digital microphones

3. Yamaha PSG-01S – commercial portable USB speakerphone with internal speaker wires routed outside enclosure for connection to C5517 EVM

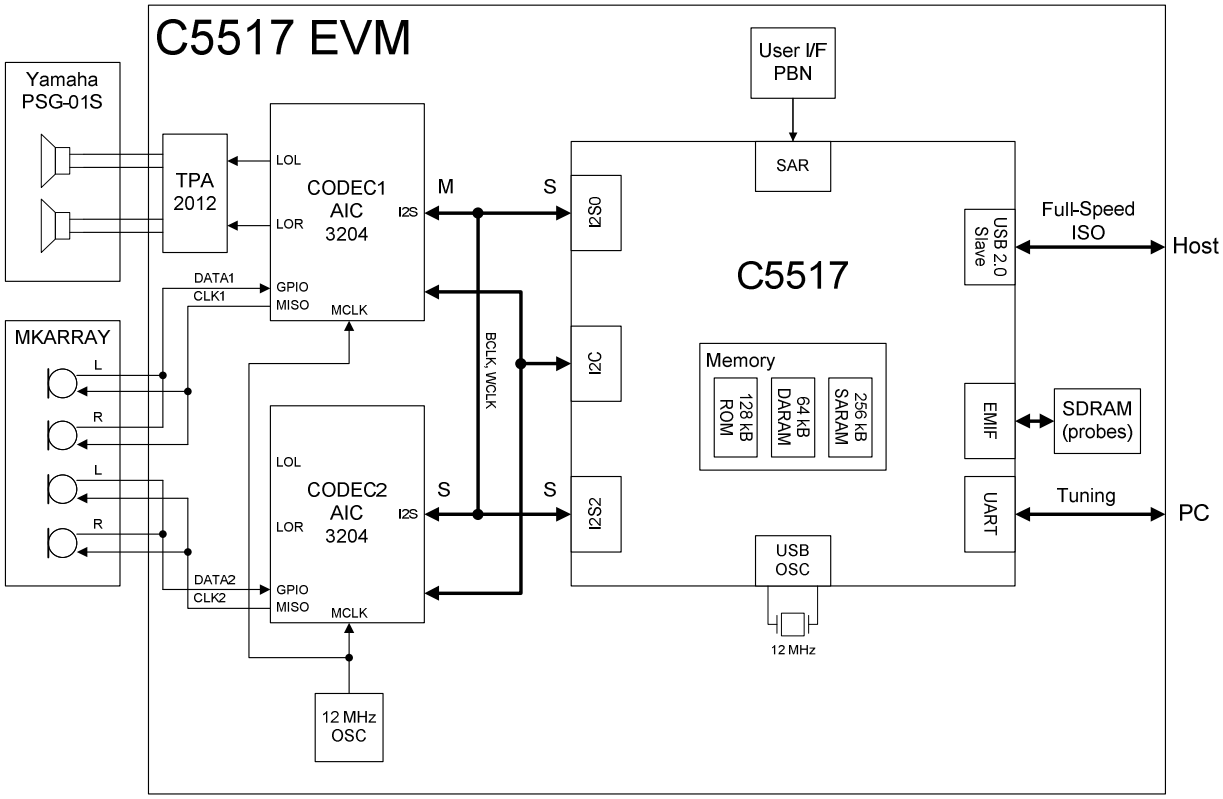

Figure 8. USB Speakerphone Hardware Block Diagram

4.1.1 C5517 EVM

The C5517 EVM [4] contains a C5517 DSP, various peripherals, expansion connectors, and a JTAG interface. The EVM allows software benchmarking, application prototyping, and algorithm debug using the C5517 DSP. The major EVM hardware blocks employed in the USB speakerphone include:

1. C5517 DSP

   The C5517 is a high-performance, low-power DSP based on the C55xx Rev. 3 core [5]. The C5517 features a 75-225 MHz clock rate and 320 KB of zero-wait state on-chip RAM, as well as numerous integrated peripherals including, but not limited to: USB2.0, UART, I2C, 3x I2S, EMIF (supports SDRAM/mSDRAM, NAND/NOR flash, and SRAM), 3x GPT, a 10-bit SAR ADC, and 4x DMA. The C5517 provides hardware and software control in the speakerphone system.

2. 2x AIC3204

   The AIC3204 is a low-power stereo audio codec with programmable inputs and outputs, fixed predefined and parameterized signal processing blocks, an integrated PLL, integrated LDOs, and flexible digital interfaces [6]. The C5517 EVM contains two AIC3204s, labeled Codec1 and

Codec2. The codecs receive their master clock input from a 12 MHz oscillator on board the EVM. The DSP sends commands to, and receives status from, the codecs over I2C. Audio data is exchanged between the DSP and Codec1 using I2S0, and Codec2 using I2S2. Codec1 is the I2S master (indicated by M label in Figure 8) for both I2S busses, and the BCLK and WCLK lines of the I2S busses are connected. The AIC3204 signal processing blocks are used for high-pass filtering in the record path to remove DC, and high-pass filtering in the playback path to reduce nonlinear distortion in the echo path. The LOL/LOR output pins from Codec1 are used to drive the inputs of the TPA2012 Class-D amplifier. The connections between the codecs and the MKARRAY microphone array are described below in Section 4.1.2.

3. TPA2012
   The TPA2012 is a stereo Class-D audio amplifier [7]. The amplifier gain can be selected via two pins as 6, 12, 18, or 24 dB. The amplifier is capable of driving 2.1 W/Ch into 4 Ω at 5 V. The TPA2012 PWM outputs are connected directly to the playback speakers in the Yamaha enclosure.

4. SDRAM
   The SDRAM is a 64 Mbit x 16 Mobile SDRAM [8]. The SDRAM is connected to the DSP via the EMIF, and is used to store AER output data for debug purposes.

5. USB 2.0 Host connector
   The DSP is connected to a USB Host capable of streaming full-speed isochronous audio data. A PC running the Windows 7 OS is an example of such a Host. Windows 7 includes a built-in USB Audio Class driver, and the USB speakerphone correctly enumerates and operates as an echo canceling speakerphone when used with Windows 7. Example applications for streaming audio data include Skype, Audacity, and Adobe Audition.

6. UART connector
   The DSP is connected over UART to a PC executing a terminal program such as Tera Term which takes ASCII commands from the user for AER parameter tuning and debug.

7. User Interface Push-Button Network
   A network of momentary switches is used to implement a user interface for the speakerphone. The network of switches is connected between the SAR supply voltage and ground through a resistor ladder. The voltage level output by the network is determined by which switch(s) are depressed, and this voltage level is periodically sampled using the DSP SAR.

4.1.2 MKARRAY

The MKARRAY is a 1-D linear microphone array designed for beamforming applications containing eight digital microphones equally spaced 2 cm apart. The MKARRAY schematic is contained in [9]. The MKARRAY has the following busses: (1) power; (2) ground; (3) 2x clock input, CLK1 and CLK2; and (4) 2x data output, DATA1 and DATA2. Any four microphones in the array can be selected as active via jumper

connections to the appropriate busses. The four active microphones consist of two stereo microphone pairs: one stereo pair obtains its clock from one clock input bus and drives its output to one data output bus, while the other stereo pair obtains its clock input from the remaining clock input bus and drives its output to the remaining data output bus. Any microphone *n* in the array is configured using jumpers:

- J1_*n*: closed to power microphone, opened to leave microphone unpowered
- J2_*n*: closed to select microphone output left channel, opened to select microphone output right channel
- CLK_*n*: pins 1-2 closed to select CLK1, pins 2-3 closed to select CLK2, opened to select no microphone input clock bus
- DATA_*n*: pins 1-2 closed to select DATA1, pins 2-3 closed to select DATA2, opened to select no microphone output data bus

The MKARRAY is populated with the Knowles SPM1423HM4H-B digital microphone. Digital microphones integrate a MEMS microphone with an oversampling sigma-delta ADC. The oversampling rate of the sigma-delta ADC is determined by the frequency of the input clock to the digital microphone:

$$F_s = f_s \times AOSR \tag{4}$$

where f_s is the baseband sampling rate, *AOSR* is the oversampling ratio, and F_s is the oversampling rate. The digital microphone output format is PDM, which is the 1-bit oversampled data stream from the sigma-delta ADC. Each digital microphone can be configured to drive its output on either the rising or falling clock edge of the input clock, and to tri-state its output on the clock edge opposite from which it drives. This allows two digital microphones to use the same input clock line, and to drive the same output data line. One of the digital microphones thus configured is considered the “left” channel, while the other is considered the “right” channel. This is shown in Figure 8, where each stereo digital microphone pair uses the same input clock line and drives the same output data line. Some of the more important specifications of the Knowles digital microphone include:

- The directivity is omni-directional.
- The nominal sensitivity is -22 dBFS for a 1 kHz tone at 94 dBSPL, with a maximum deviation of 3 dB. The sensitivity vs. frequency curve is relatively flat from 100 Hz to 10 kHz.
- The SNR is 61.5 dBA.
- The THD is 1% for a 1 kHz tone at 100 dBSPL.
- The maximum input clock frequency (i.e. the maximum F_s in Equation 4) is 3.25 MHz.

Additional details on the microphone specifications can be found in the microphone datasheet [10]. The microphones populating the MKARRAY are not specially matched, and normal variation in the sensitivity and phase response of the microphones will result in a loss in BF performance with respect to theoretical performance. Adaptive filters can be used to equalize the microphone outputs and improve the effective matching between the microphones. When BF is used with AER, however, care must be taken concerning when the equalizers are active or AER echo cancelation performance can degrade due to AER observing a changing echo path impulse response.

The AIC3204 can interface directly with up to two digital microphones. When configured in this manner, the AIC3204 produces the digital microphone input clock at the required oversampling rate, and consumes the digital microphone PDM output. The PDM output is routed to the decimation engine inside the AIC3204 where it is low-pass filtered and down sampled to the baseband sampling rate. As shown in Figure 8, Codec1 outputs the clock for two microphones CLK1 on its MISO pin, and inputs the data for the microphones DATA1 on its GPIO pin. Likewise, Codec2 outputs CLK2 and inputs DATA2 for another two microphones. When used with digital microphones, the AIC3204 on-board ADC can be powered-down.

It is advantageous to use the maximum possible oversampling ratio for a digital microphone when configuring the codec. This is because the quantization noise spectrum from the sigma-delta ADC inside the digital microphone will be spread over a wider oversampling bandwidth, and the baseband samples after decimation filtering will therefore have a higher Signal-to-Quantization Noise Ratio. For f_s = 16 kHz and $F_{s,max}$ = 3.25 MHz, $AOSR_{max}$ = 3.25e6/16e3 = 203.125. Hence, a value of *AOSR* = 192 was selected to approach the maximum *AOSR* value. The AIC3204, however, applies an undocumented gain in the decimation engine when using an AOSR other than 32, 64, 128, or 256. For the case of *AOSR* = 192, the decimation filter applies the following gain:

$$4\log_2\left(\frac{48}{32}\right) = 2.34 \text{ bits} \tag{5}$$

$$20\log_{10}\left(2^{2.34}\right) = 14.09 \text{ dB} \tag{6}$$

The nominal output level of the digital microphones for a 1 kHz tone at 94 dBSPL will thus be -22 + 14.09 = -7.91 dBFS. Although this gain increases the observed microphone sensitivity, the gain also increases the microphone output noise level by a corresponding amount, and the SNR remains the same.

It is desirable to synchronously sample the four active microphones in the array so the phase information passed to the beamformer is accurate. However, this is not possible using the codecs on the C5517 EVM because:

1. For each pair of microphones sampled by a single codec, the left and right channel microphones are sampled on separate edges of the oversampling clock. Hence there is a sampling time difference between the channels of $1/(2F_s)$.
2. The clock divider tree of each codec is configured/reset at different times over I2C. The clocks generated by the two clock trees are therefore asynchronous, including the output oversampling clock. Consequently there is a sampling time difference of up to $1/(2F_s)$ for the left channels and right channels output by the two microphone pairs. Note providing the output oversampling clock from a single codec to all four active microphones does not resolve this problem since the decimation engine in the codec with the unused clock will operate

asynchronously with respect to the clock provided to the microphones, resulting in a random left/right output channel swap for the codec with the unused clock.

4.1.3 Yamaha PSG-01S

The Yamaha PSG-01S is a commercially available, portable USB speakerphone designed for use with Skype [11]. The Yamaha features two 3", four ohm speakers with a playback frequency range of 300 Hz to 20 kHz. The speakers contained in the Yamaha were used as the speakers in the USB speakerphone. Two holes were drilled in the Yamaha enclosure, and the positive and negative speaker wires were routed outside the enclosure for direct connection with the TPA2012 Class-D amplifier outputs on the C5517 EVM.

4.2 USB Speakerphone Software

The USB speakerphone software executing on the DSP includes the following basic components: (1) C55xx CAF; (2) AER; and (3) software for AER tuning and debug. An overview of AER is presented in Section 3 above. The C55xx CAF and AER tuning/debug software are discussed below.

4.2.1 C55xx CAF Software

The C55xx CAF provides a software framework which allows the C55xx to operate as a full-speed USB audio peripheral [12]. CAF supports the following USB Device Classes:

- Audio Device Class 1.0
  Employs adaptive synchronization in the playback path. A software ASRC is used to implement the adaptive synchronization.
- USB HID Device 1.11
  Used to implement volume up, volume down, record mute, and hook switch controls. The HID software examines the SAR output to determine which switch has been pressed, and uses this information to construct a HID input report which is sent to the USB Host. The Host determines what action to take based on the input report, and sends a return message over USB as required.

The CAF framework can be extended to incorporate audio processing algorithms in the playback and record paths. For the USB speakerphone software, AER was incorporated into the playback and record paths.

The CAF was used as the basis for developing a mono record (i.e. single microphone) version of the USB speakerphone software which executes on a C5515 EVM [13]. The C5515-based USB speakerphone software was, in turn, ported to the C5517 EVM and modified to support multiple record channels. Because of this, the C5517-based USB speakerphone software inherits its basic software architecture from the CAF software, and reuses much of the low-level HAL software and peripheral drivers utilized by CAF. Details of the CAF software architecture can be found in [14]. The following key modifications were made to the CAF software in developing the C5515-based mono record USB speakerphone software:

- Integrated AER into the Tx/Rx audio processing threads.

- Added a volume table containing codec settings and AER parameters to (1) achieve a pre-determined loudness for each volume level, and (2) optimize echo cancelation performance for each volume level.
- Modified the USB descriptors so the device enumerates as an echo-canceling speakerphone.
- Modified the HID report descriptor to support playback volume up/down, record mute, and hook switch controls.

Additionally, the following key modifications were made to the C5515-based USB speakerphone software during porting to the C5517:

- Added support for digital microphones to the AIC3204 driver.
- Added support for two AIC3204 codecs (one for stereo playback/record, one for stereo record).
- Added support for multiple record channels.
- Optimized the DMA driver by eliminating a single callback function and semaphore post per DMA channel completion, and replacing it with a single post of Rx and Tx semaphores to initiate AER processing for each audio frame.

4.2.2 AER Tuning/Debug Software

The AER tuning/debug software comprises (1) an interpreter for user commands received over UART for tuning/debug, and (2) a manager for read/write data probes.

Command Interpreter

The command interpreter is provided to simplify the tuning of AER parameters for best AER performance on a given enclosure. AER parameter tuning is discussed in [15]. The command interpreter allows the user to input commands for writing/reading codec registers, AER and BF parameters, and the speaker output volume level.

Data Probes Manager

The data probes manager provides access to read/write data probes. Read/write data probes allow capture/insertion of streaming data to/from memory. Probes may be inserted anywhere in the software, and which probes are enabled is determined at run time. The USB speakerphone software is instrumented with probes at locations which facilitate AER debug. Read probes are located at: (1) AER R_{in}; (2) AER R_{out}; (3) digital microphone outputs; (4) BF output (AER S_{in}); and (5) AER S_{out}. A single write probe is located at AER S_{in}. Data probe destination/source buffers may be located in any memory; currently the buffers are located in SDRAM.

5 BF Testing

A variety of tests were performed using the USB speakerphone test platform to assess the impact of BF on (1) near-end noise suppression when used with ASNR, and (2) echo cancelation when used with AER. The USB speakerphone was used to evaluate the BF impact while using hardware realistic for a portable speakerphone.

5.1 Test Setup

5.1.1 Test Hardware

The hardware listed below was used in performing the BF tests. A diagram of the connections between the various pieces of hardware is presented in Figure 9.

Macbook Pro

A Macbook Pro using Windows 7 was used as the Host for the USB speakerphone, and for controlling the volume level of the Mackie HR624 via the AudioFire8. The Macbook Pro was connected to the AudioFire8 over FireWire, and to the Logitech speakers using the Macbook's headphone output.

Latitude E6420

A Dell Latitude E6420 laptop PC was used as a debug PC. The PC was connected to the C5517 EVM via JTAG, and was used to compile, load, and debug the USB speakerphone program using CCSv5.3. The debug PC was also connected to the C5517 EVM via UART, and was used for AER tuning using Tera Term 4.73.

Echo AudioFire8

The Echo AudioFire8 [16] is high quality digital audio interface which connects to a host PC over FireWire. A mixer console application executing on the host PC allows adjustment of the output gains. The HR624 was connected to the AudioFire8 using Analog Output 1.

Mackie HR624 MK2

The Mackie HR624 is a studio monitor designed for production audio applications. The HR624 features a free-field (anechoic chamber) frequency response of ±1.5 dB from 49 Hz to 20 kHz. Further details concerning the HR624 can be found in [17]. The HR624 was configured as follows for the BF tests:

- Sensitivity set to normal ⇒ +4 dBu input
- Acoustic switch space set to C ⇒ whole, normal
- Low frequency filter set to 49 Hz ⇒ normal
- High frequency filter set to 0 ⇒ normal

Logitech Speakers Z130

Logitech speakers, model Z130, are standard PC speakers with 2.5 W RMS output per speaker.

Radio Shack Sound Level Meter

A Radio Shack sound level meter was used to measure the dBSPL output from the monitor, PC, and Yamaha speakers. The sound pressure meter was configured for A-weighting and slow response.

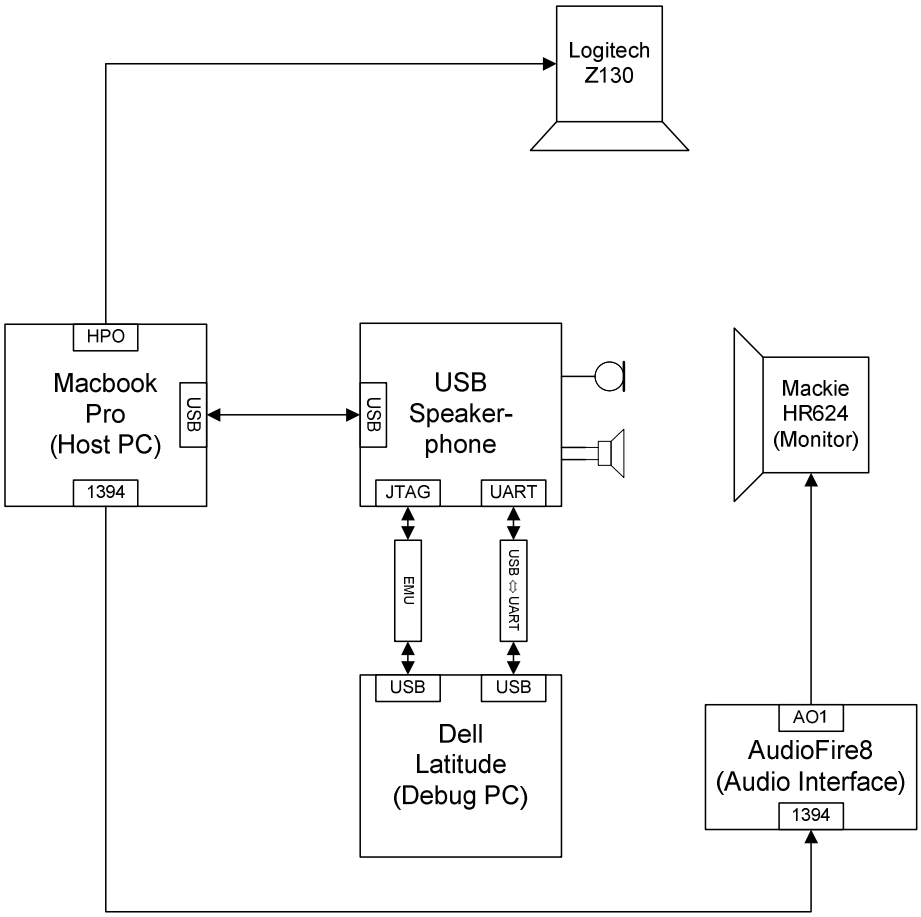

Figure 9. BF Test Hardware Connections

5.1.2 Test Environment

Anechoic chambers are commonly used in acoustic testing to provide free field (no reflective surfaces within the frequency range of interest) conditions. Unfortunately no anechoic chamber was available for BF testing, and the tests were performed in a laboratory with an unknown room response.

The lab in which the tests were performed contained equipment generating audible background noise. The measured background noise level output by the digital microphones was approximately -54 dBFS. Accounting for the gain introduced in the AIC3204 decimation engine discussed in Section 4.1.2, the background noise level in the lab was therefore -54 - 14.09 = -68.09 dBFS. From the digital microphone specifications, the digital microphone noise floor is -22 - 65 = -87 dBFS. Hence the noise in the lab was above the digital microphone noise floor by -68.09 - (-87) = 18.09 dB.

5.1.3 Test Software Configuration

Unless otherwise specified, BF testing was performed with the common AER/BF configuration below:

- Sampling frequency f_s = 16 kHz
- AER full-band echo cancelation, i.e. cancelation in (0-8 kHz)
- AER tail length = 200 msec.
- Tx EQ disabled

- 24 coefficients per BF filter

5.2 BF/ASNR for Noise Reduction

The ASNR is designed to improve the SNR in the AER Tx path [1]. It is tightly integrated within the Tx NLP. Only the stationary component of a background noise is attenuated. Degradation of voice quality may be expected, and the amount of degradation depends on the configured amount of attenuation. Degradation may also be noticeable during the initial estimation of background noise or when the background noise changes abruptly.

Two categories of tests were performed to assess the impact of BF on noise reduction when used with ASNR. For the first category, fan sounds were used as a stationary noise source to be suppressed prior to transmission to the far end. In the second category, keyboard typing sounds were used as a non-stationary noise source. For both test categories, female speech was used as the near-end signal of interest to transmit to the far end.

5.2.1 Fan Noise

The BF applies minimal attenuation to noise in low frequencies regardless of the steering angle (e.g. see Figure 7, where frequencies below approximately 1.5 kHz experience little or no attenuation). Beyond the low frequency range, the BF applies attenuation to noise which depends on the number microphones in the array, microphone spacing, steering angle, and angle of arrival. It is shown in [18] that fixed BF using a four microphone 1-D linear array attenuates car noise by up to 6dB for higher frequencies. The attenuation may be even more depending on the type of noise, e.g. up to 12dB as shown in the tests described in section 5.2.1.2.

The ASNR can distort speech when configured for a high level of noise reduction, particularly the unvoiced sounds in speech. ASNR provides three separate frequency bands with adjustable ranges in which distinct noise attenuation is applied. Using these frequency bands, ASNR can be configured to apply progressively less attenuation with increasing frequency. When configured in this manner, ASNR can provide improved speech quality compared with ASNR configured for high noise reduction across all frequencies. Further, the relaxed ASNR can be combined with BF to provide comparable noise reduction to the aggressively configured ANR alone. The intent of these tests was to determine whether BF and ASNR could together provide as much noise reduction as ASNR alone, while providing improved quality for the speech transmitted to the far end.

5.2.1.1 Test Signals

The signal and noise sources were arranged as given below in Table 1.

Source	Playback source	Angle to mic array (°)	Distance to mic array (cm)	dBSPL at mic array
Signal – female speech	Logitech speakers	0	50	71
Noise – fan sounds	Monitor	90	50	59

Table 1. Test of BF/ASNR Stationary Noise Reduction – Sound Sources

5.2.1.2 Tests

Test1 – no noise attenuation

ASNR/BF configuration:

- ASNR: disabled.
- BF: pass through mic1, i.e. no beamforming, and mic 1 signal passed to the output.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 10.

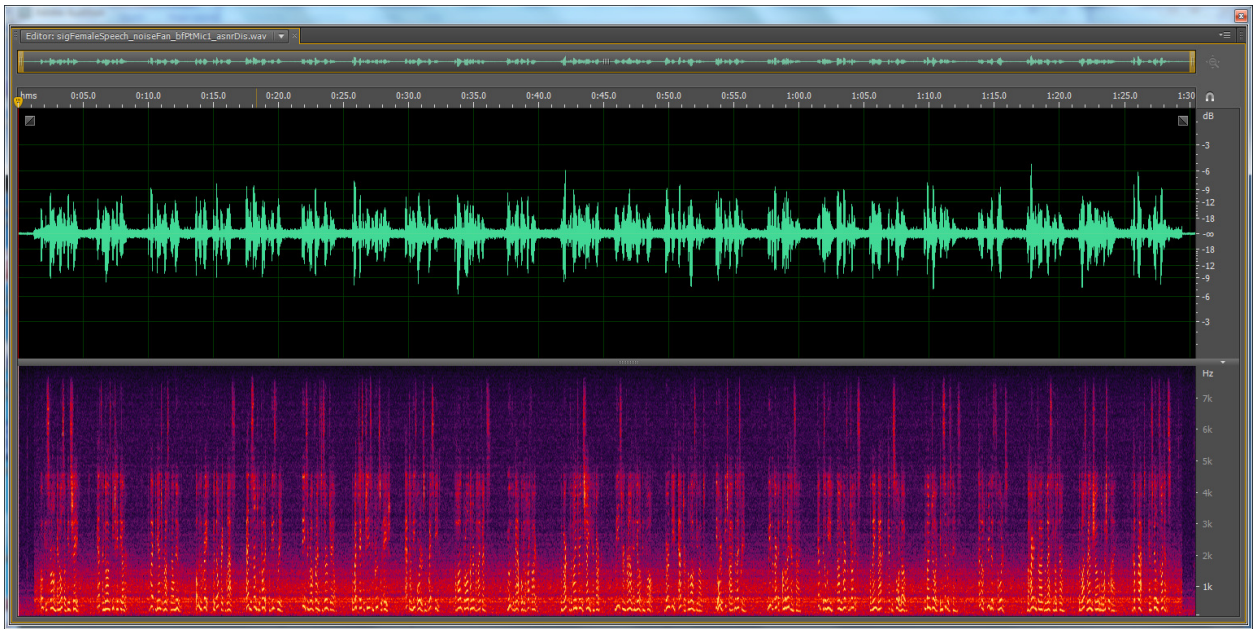

Figure 10. Test 1 - Time Domain and Spectrogram
(BF Disabled; ASNR Disabled)

Test 2 – ASNR high noise attenuation

ASNR/BF configuration:

- ASNR: enabled, {0-8000 Hz: 21 dB}.
- BF: pass through mic1.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 11.
- The noise spectra for the Test 1 and 2 captures are presented in Figure 12. The red line corresponds to the spectrum for Test 1, and the yellow line to the spectrum for Test 2.
- The time domain plot and spectrogram for the 12th speech segment are presented in Figure 13. The circled portion of spectrogram corresponds to a part of the speech segment distorted by ASNR. This is further discussed below in the test results in Section 5.2.1.3.

Comparing Figures 10 and 11 , it can be observed that the stationary noise has been significantly attenuated by the ASNR. As shown in Figure 12, the ASNR applies approximately 21 dB attenuation to the noise.

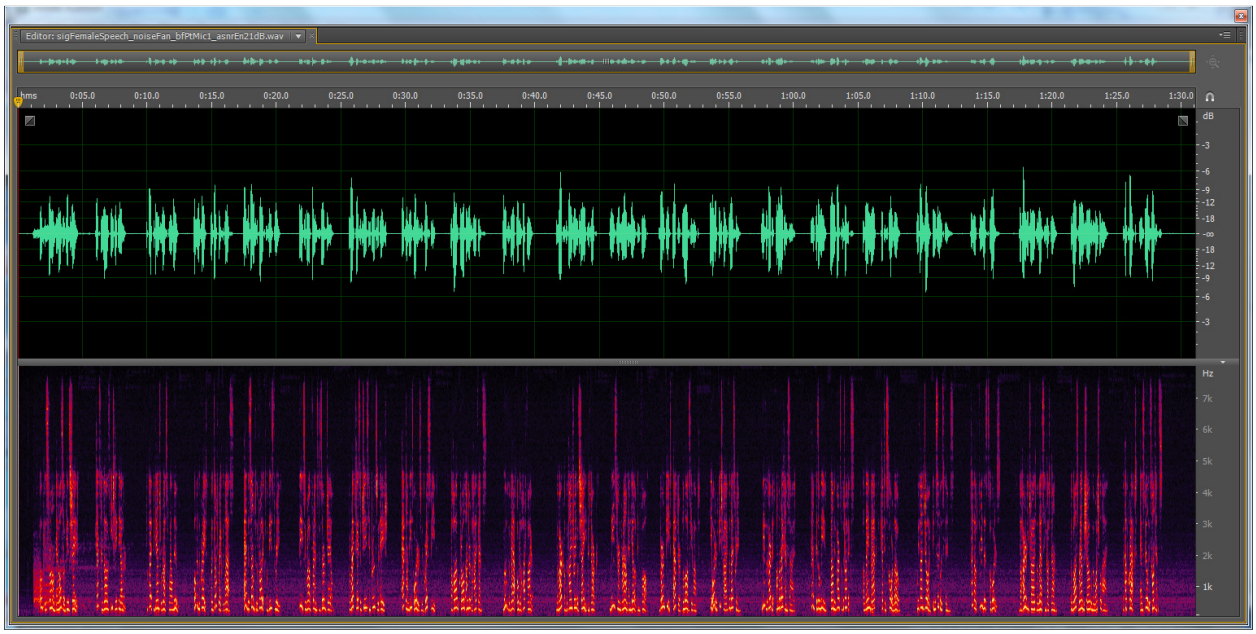

Figure 11. Test 2 - Time Domain and Spectrogram
(BF Disabled; ASNR Enabled, High Noise Attenuation)

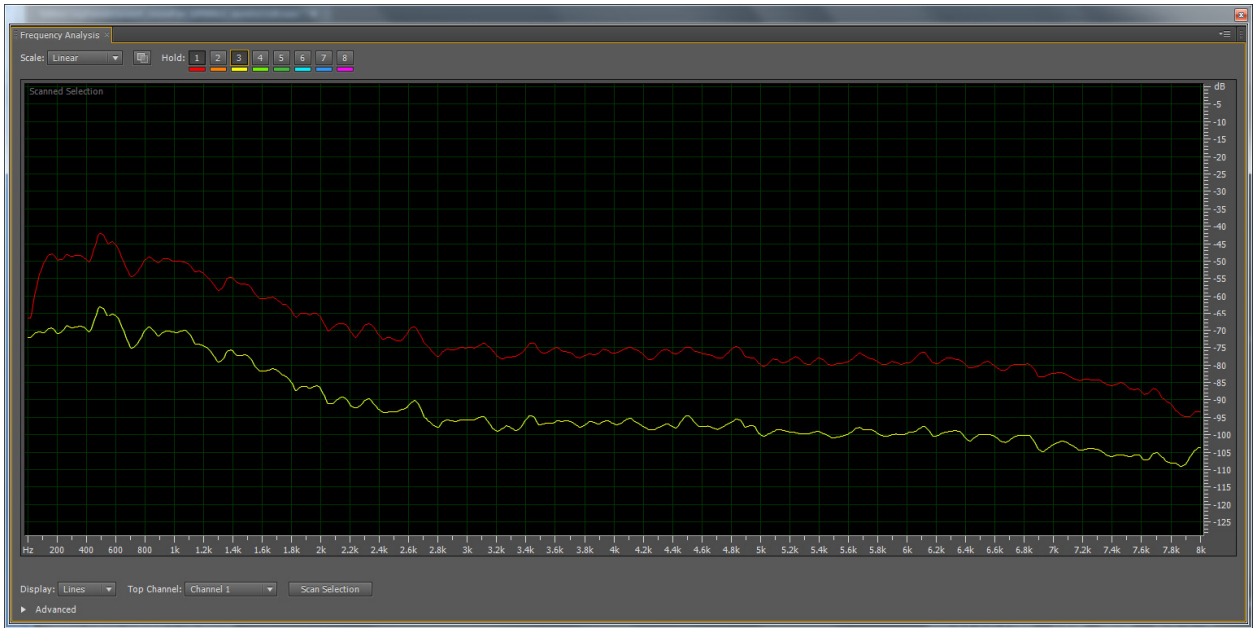

Figure 12. Test 1 and 2 Noise Spectra
(Red: BF & ASNR Disabled, Yellow: BF Disabled & ASNR Enabled)

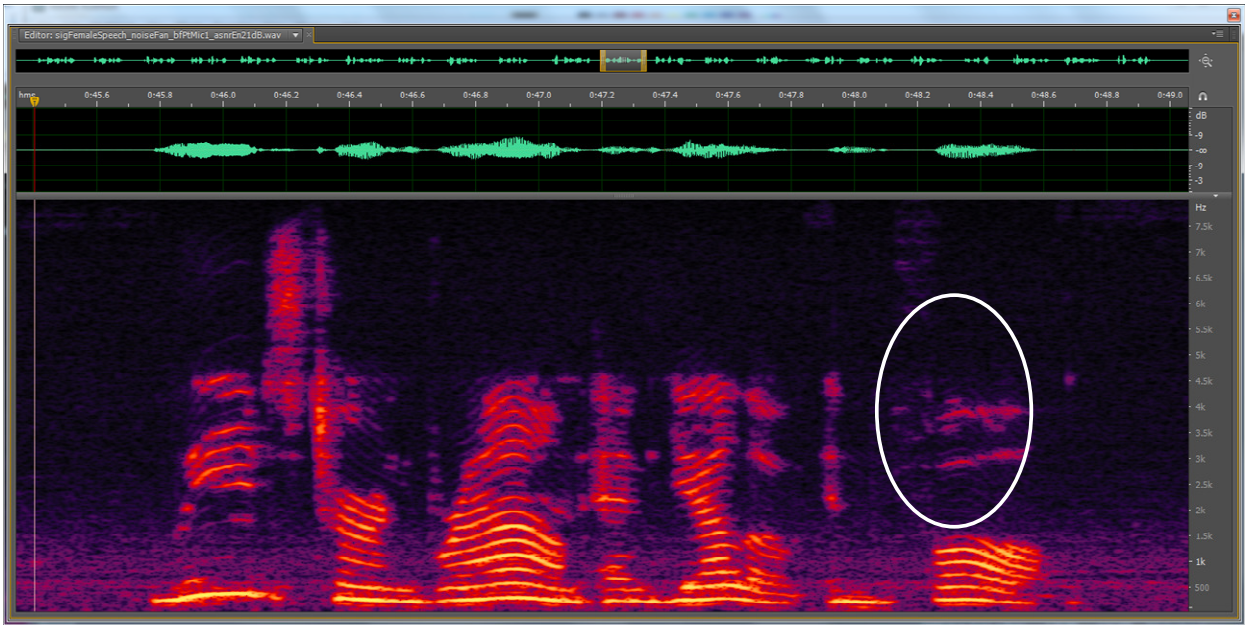

Figure 13. Test2 - Time Domain and Spectrogram, 12th Speech Segment (BF Disabled; ASNR Enabled, High Noise Attenuation)

Test 3 – BF only

ASNR/BF configuration:

- ASNR: disabled.
- BF: enabled, steering angle=0°.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 14.
- The noise spectra for the Test 1 and 3 captures are presented in Figure 15. The red and blue lines correspond to the spectra for Test 1 and 3, respectively.

Comparing the spectrograms in Figures 10 and 14, is can be seen that the stationary noise has been attenuated by BF. Examination of Figure 15 further illustrates this point, where the blue line shows BF begins attenuating frequencies at approximately 1.5 kHz, and applies attenuation of up to 12 dB starting at 4.6 kHz.

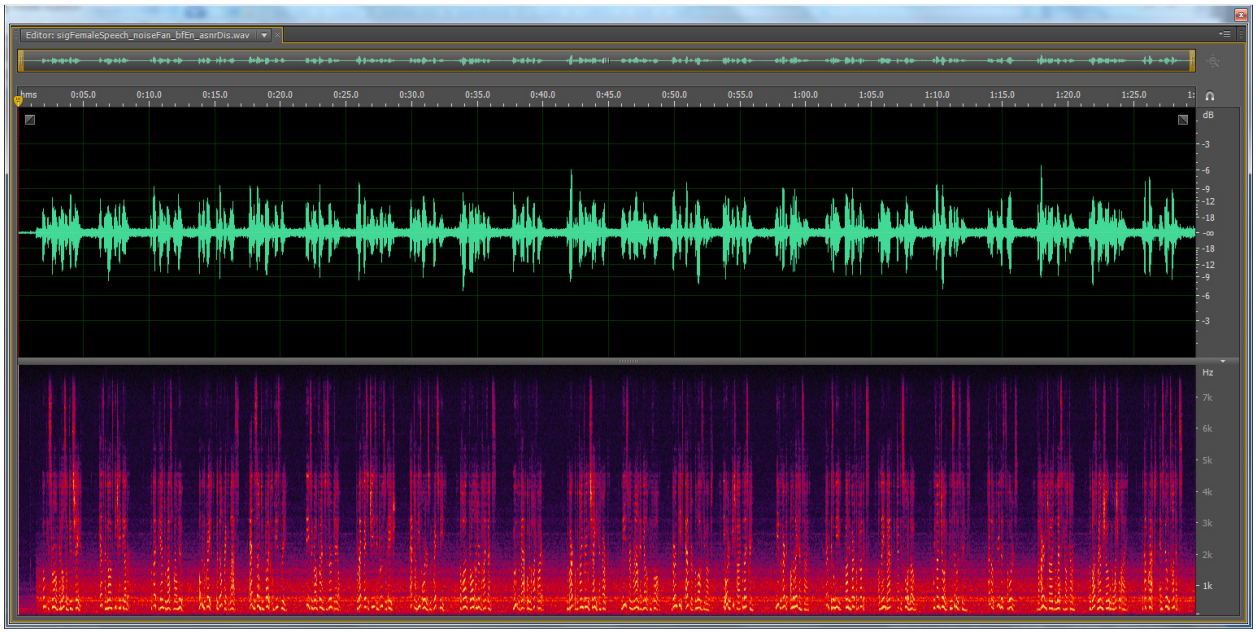

Figure 14. Test 3 - Time Domain and Spectrogram
(BF Enabled; ASNR Disabled)

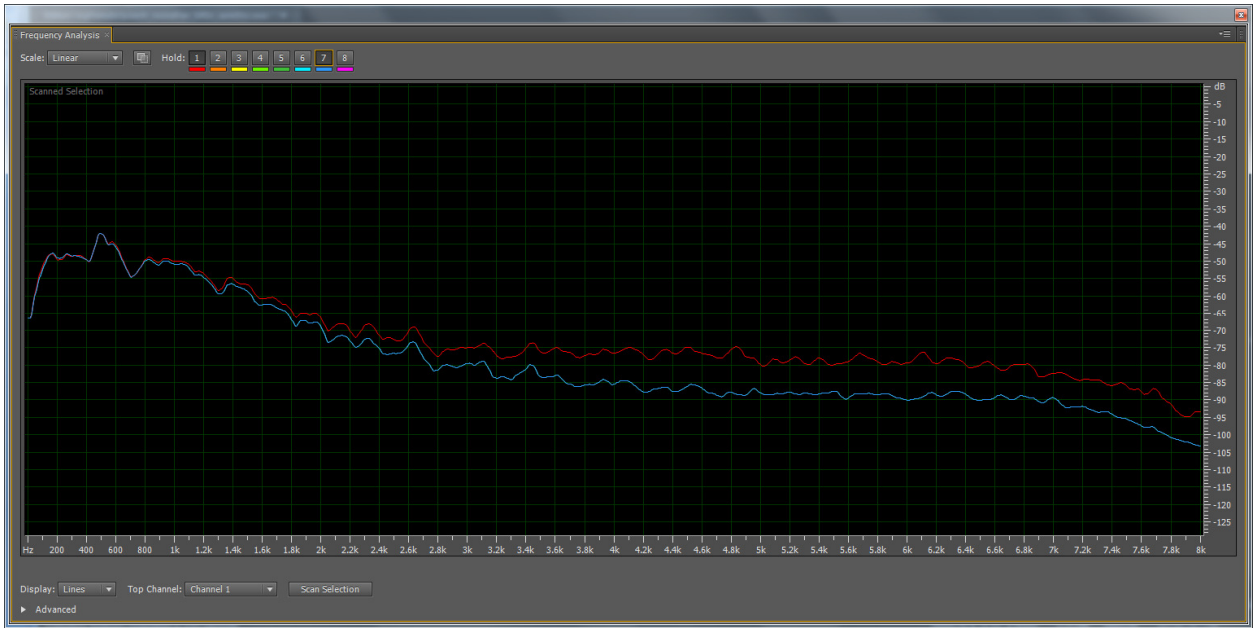

Figure 15. Test 1 and 3 Noise Spectra
(Red: BF & ASNR Disabled, Blue: BF Enabled & ASNR Disabled)

<u>Test 4 – BF/ASNR, ASNR relaxed noise attenuation</u>

ASNR/BF configuration:

- ASNR: enabled, {0-1375 Hz: 21 dB, 1375-3500 Hz: 18 dB, 3500-8000 Hz: 12 dB}.
- BF: enabled, steering angle=0°.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 16.
- The noise spectra for the Test 1-4 captures are presented in Figure 17. The red, blue, yellow and purple lines correspond to the spectra for Test 1, 3, 2, and 4, respectively.
- The time domain plot and spectrogram for the 12th speech segment are presented in Figure 18. The circled portion of spectrogram corresponds to a part of the speech segment distorted by ASNR. This is further discussed below in the test results in Section 5.2.1.3.

Comparing the spectrograms in Figures 10 and 16, is can be seen that the stationary noise has been attenuated by BF and ASNR. Examination of Figure 17 further illustrates this point, where the purple line shows BF used jointly with ASNR applies approximately 21 dB attenuation to the noise.

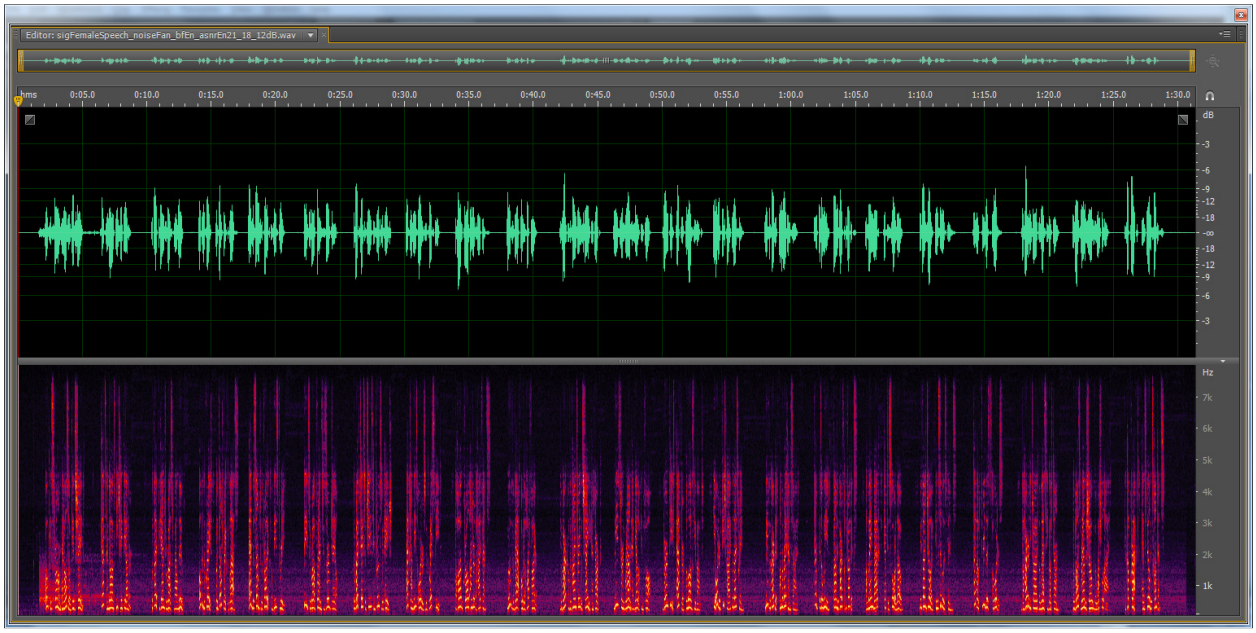

Figure 16. Test 4 - Time Domain and Spectrogram
(BF Enabled; ASNR Enabled, Relaxed Noise Reduction)

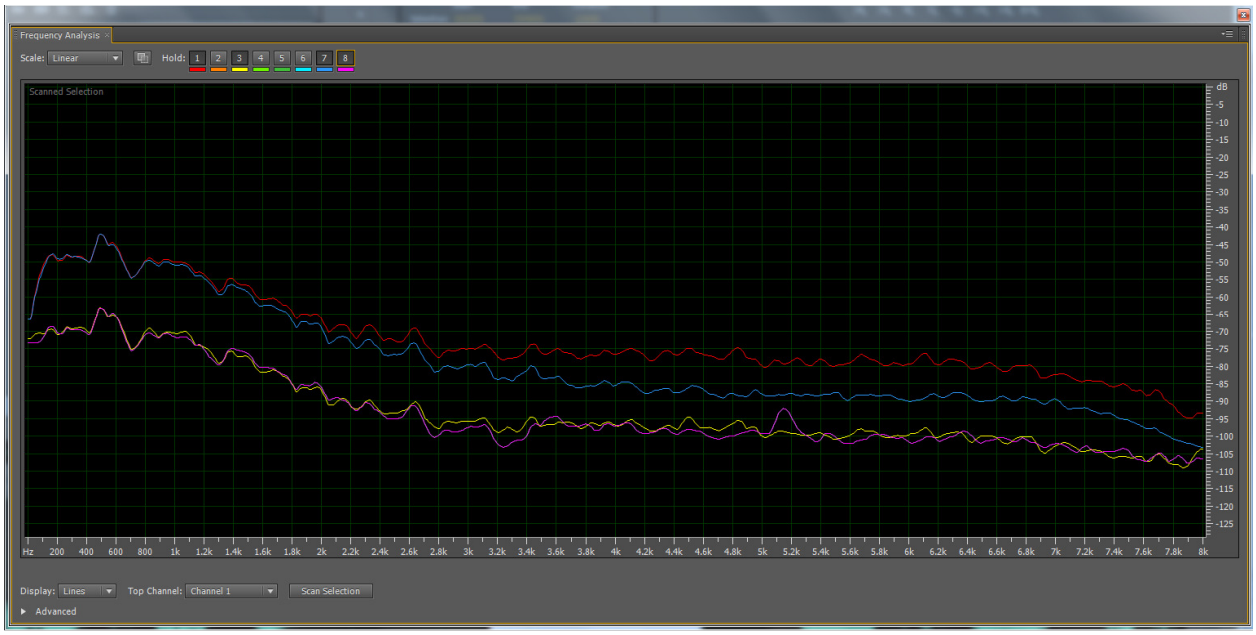

Figure 17. Test 1-4 Noise Spectra
(Red: BF & ASNR Disabled,
Blue: BF Enabled & ASNR Disabled,
Yellow: BF Disabled; ASNR Enabled, High Noise Attenuation
Purple: BF Enabled; ASNR Enabled, Relaxed Noise Reduction)

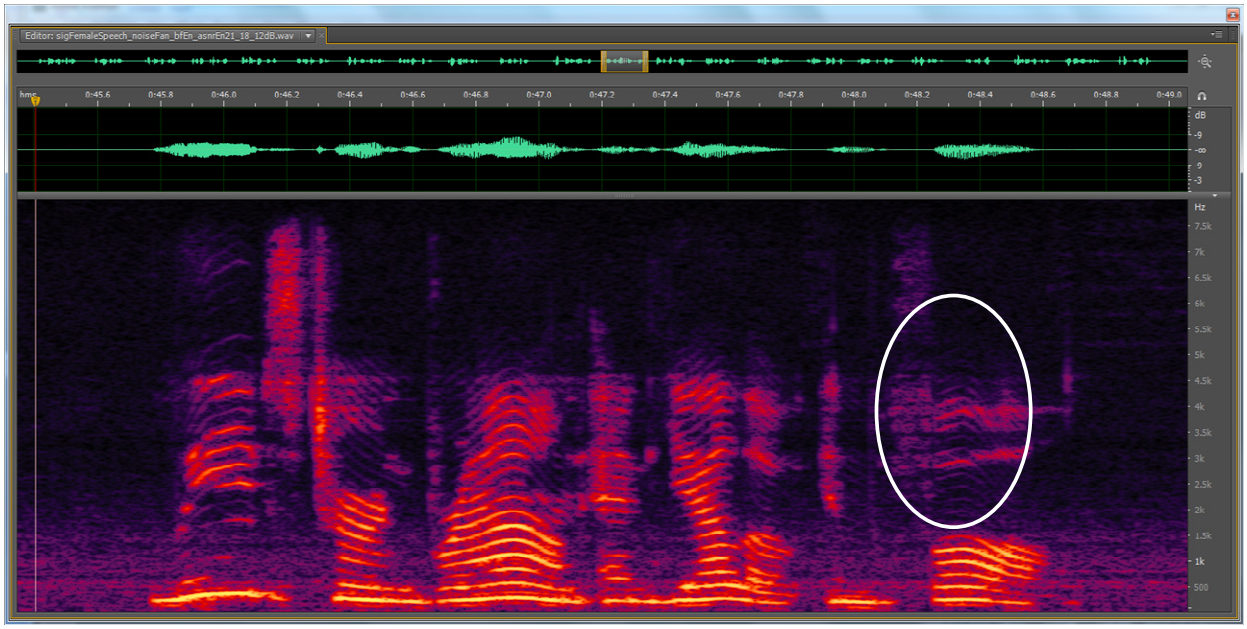

Figure 18. Test 4 - Time Domain and Spectrogram, 12th Speech Segment
(BF Enabled; ASNR Enabled, Relaxed Noise Reduction)

5.2.1.3 Test Results

The Test 2 and 4 results indicate that BF and the relaxed ASNR together provide similar noise reduction to the aggressive ASNR alone. In particular, this is demonstrated by comparing Figures 11 and 16, and the yellow and purple lines in Figure 17.

The Test 2 and 4 results also show that the BF and relaxed ASNR provide higher speech quality than the aggressive ASNR. This is confirmed by comparing the spectrograms in Figures 13 and 18, where it can be observed that the BF and relaxed ASNR preserve more of the speech spectrum. The circled portions of the spectrograms in these figures correspond with a region of the speech segment ("F" sound in "Fall") in which the improvement in speech quality improvement is audible.

5.2.2 Typing Noise

The ASNR attenuates the stationary component of background noise, and non-stationary components are passed through without attenuation. Examples of non-stationary noise include background chatter and typing noise. The BF attenuates sound sources arriving at the microphone array from angles other than the steering angle. Thus BF can attenuate non-stationary noise components unaffected by the ASNR. The purpose of these tests was to judge how well non-stationary noise is attenuated by BF, the ASNR, and BF with ASNR.

5.2.2.1 Test Signals

The signal and noise sources were arranged as given below in Table 2.

Source	Playback source	Angle to mic array (°)	Distance to mic array (cm)	dBSPL at mic array
Signal – female speech	Logitech speakers	0	50	71
Noise – typing sounds	Monitor	90	50	*

(*) Same volume level as fan sounds in Table 1.

Table 2. Test of BF/ASNR Non-Stationary Noise Reduction – Sound Sources

5.2.2.2 Tests

Test 1 – noise-only capture; no noise attenuation

ASNR/BF configuration:

- ASNR: disabled.
- BF: pass through mic1.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram for a noise-only capture are presented in Figure 19.

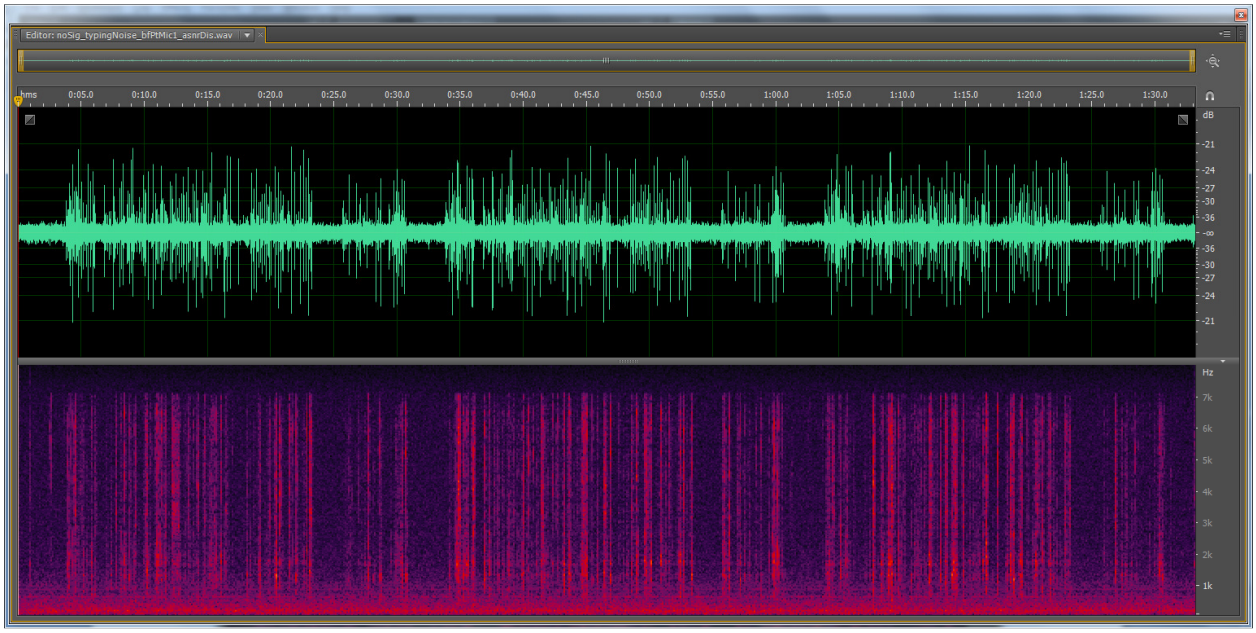

Figure 19. Test 1 – Noise-Only Time Domain and Spectrogram (BF Disabled; ASNR Disabled)

Test 2 – noise-only capture; ASNR high noise attenuation

ASNR/BF configuration:

- ASNR: enabled, {0-8000 Hz: 18 dB}.
- BF: pass through mic1.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 20.

Comparing Figures 19 and 20, it is evident the ASNR has applied the expected attenuation to the stationary background noise. However, the ASNR has applied little attenuation to the non-stationary typing noise.

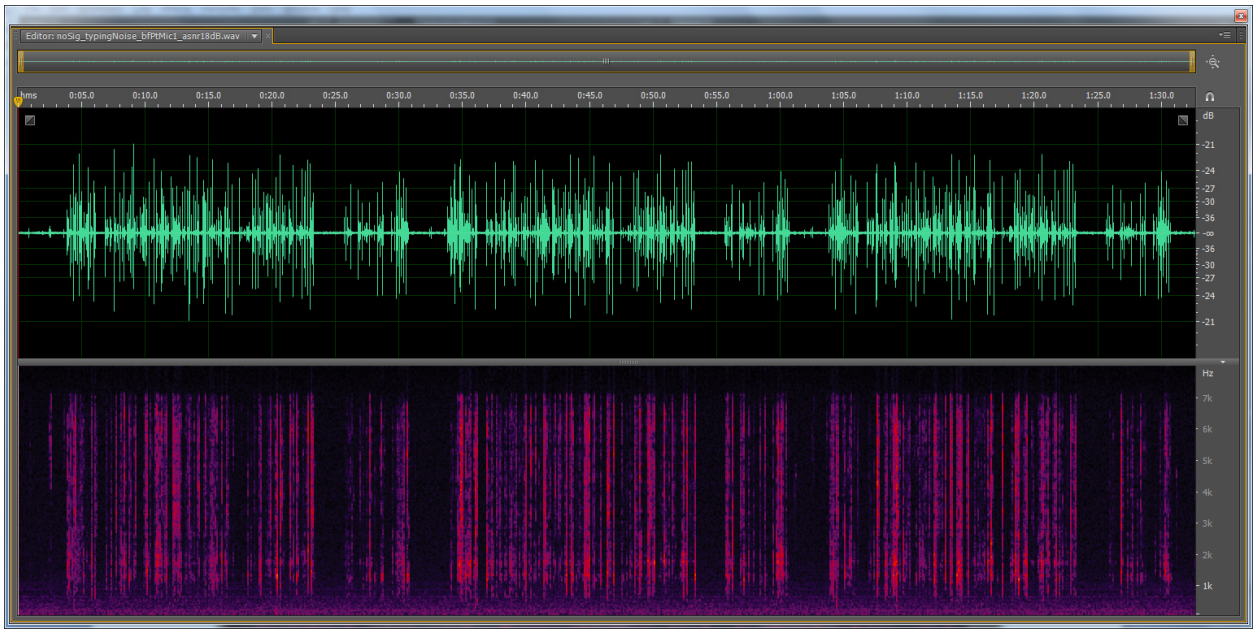

Figure 20. Test 2 – Noise-Only Time Domain and Spectrogram (BF Disabled; ASNR Enabled, High Noise Attenuation)

Test 3 – noise-only capture; BF only

ASNR/BF configuration:

- ASNR: disabled.
- BF: enabled, steering angle=0°.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 21.

Comparing Figures 19 and 21, it can be seen that BF has applied the expected attenuation to the stationary background noise. Unlike ASNR, however, BF has also attenuated the non-stationary typing noise.

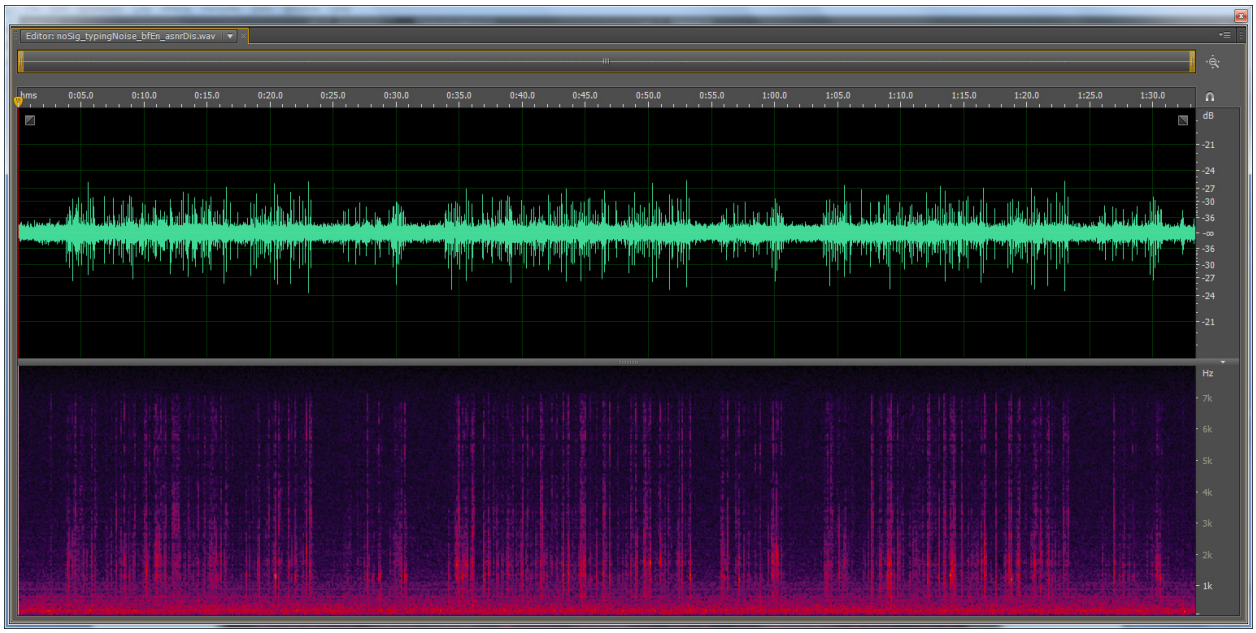

Figure 21. Test 3 – Noise-Only Time Domain and Spectrogram (BF Enabled; ASNR Disabled)

Test 4 – noise-only capture; BF/ASNR, ASNR relaxed noise attenuation

ASNR/BF configuration:

- ASNR: enabled, {0-1375 Hz: 18 dB, 1375-2000 Hz: 15 dB, 2000-4000 Hz: 12 dB}.
- BF: enabled, steering angle=0°.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 22.

Comparing Figures 19 and 22, it is apparent that BF and ASNR have simultaneously attenuated the stationary background noise and the non-stationary typing noise.

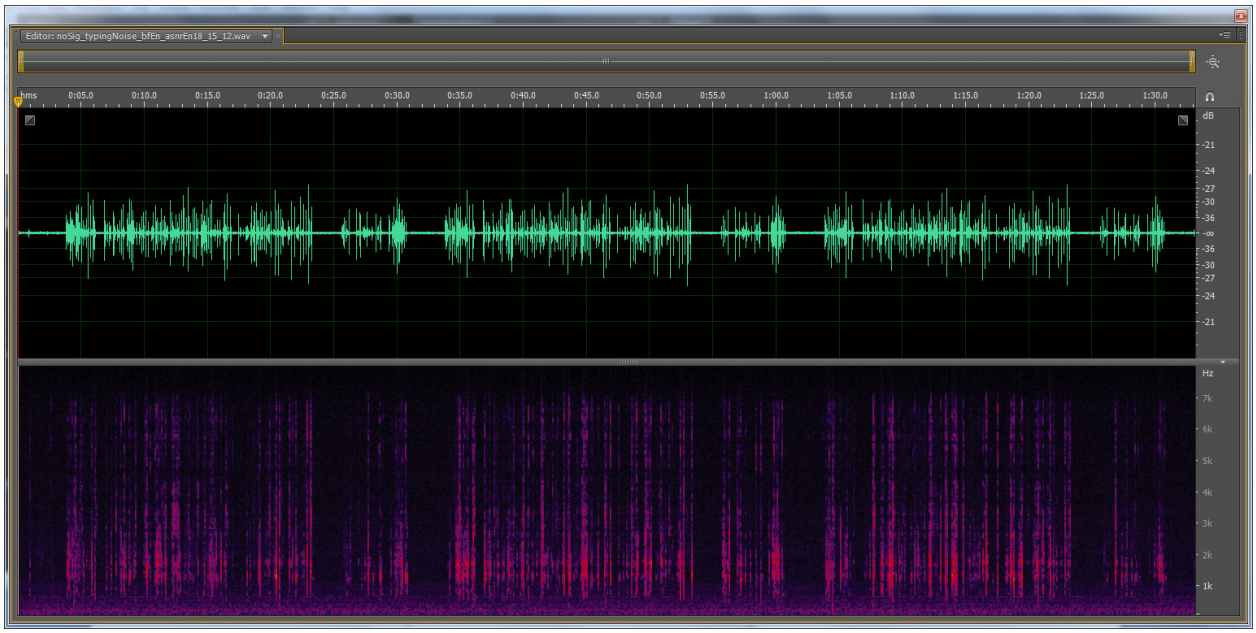

Figure 22. Test 4 – Noise-Only Time Domain and Spectrogram (BF Enabled; ASNR Enabled, Relaxed Noise Attenuation)

Test 5 – signal-plus-noise capture; no noise attenuation

ASNR/BF configuration:

- ASNR: disabled.
- BF: pass through mic1.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram for the 10th speech segment are presented in Figure 23. The earlier parts of the plots through about 0:40.1 min:sec show a time span during which the speech and noise overlap, while the later parts show a time span during which only noise is present.

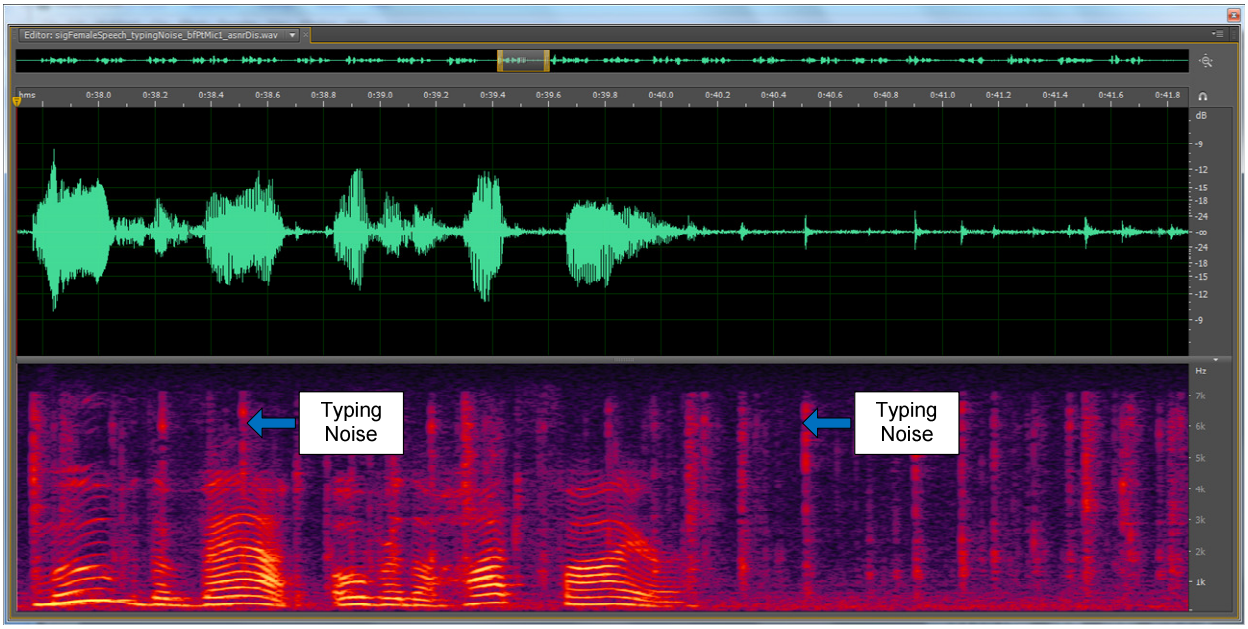

Figure 23. Test 5 – Signal-Plus-Noise Time Domain and Spectrogram, 10th Speech Segment (BF Disabled; ASNR Disabled)

Test 6 – signal-plus-noise capture; BF only

ASNR/BF configuration:

- ASNR: disabled.
- BF: enabled, steering angle=0°.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 24. The plots are time-aligned with the plots from Test 5.

Comparing Figures 23 and 24, it can be seen that BF has attenuated the non-stationary typing noise. The typing noise is attenuated whether speech is present or absent. When speech is present, the speech spectrum is preserved.

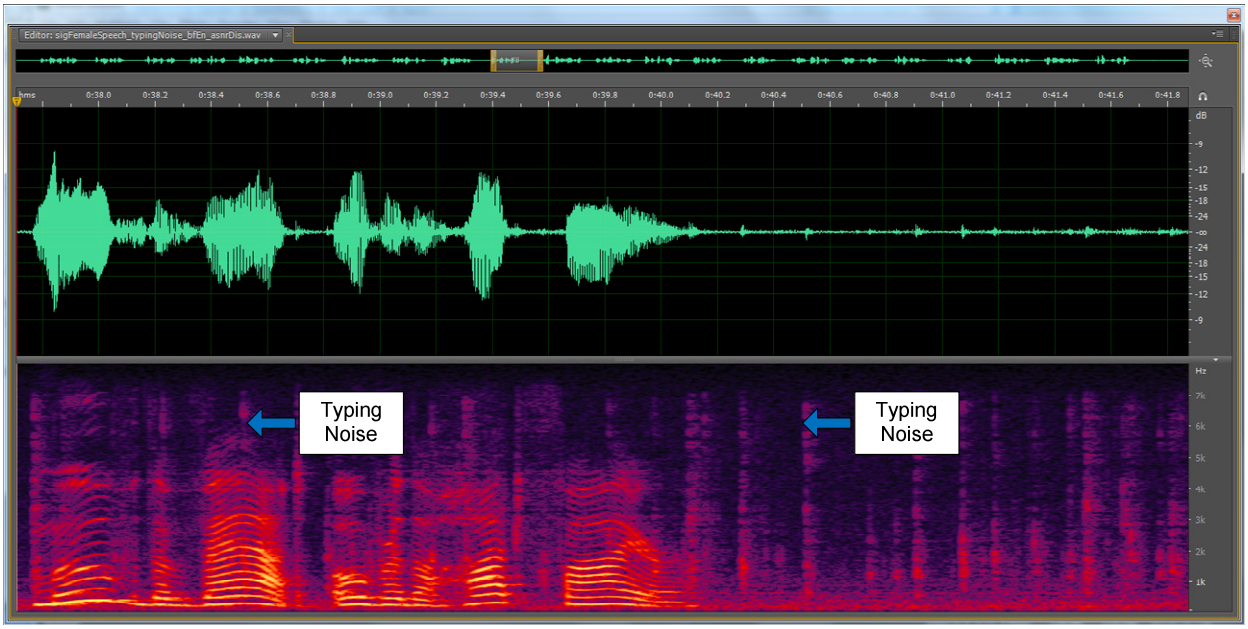

Figure 24. Test 6 – Signal-Plus-Noise Time Domain and Spectrogram, 10th Speech Segment (BF Enabled; ASNR Disabled)

5.2.2.3 Test Results

The Test 1-4 results indicate BF attenuates non-stationary typing noise, while the ASNR applies minimal attenuation to this type of noise. The BF and a relaxed ASNR together provide the same level of stationary noise reduction as an aggressive ASNR (see Section 5.3.2), but the BF also attenuates the non-stationary typing noise.

The Test 5-6 results indicate the BF attenuates non-stationary typing noise without degrading the quality of the speech transmitted to the far end.

5.3 BF/AER for Echo Cancelation

The echo produced by the speakers in Figure 1 contains linear and nonlinear components. The LMS adaptive filter employed by AER is able to cancel linear components in the echo, but unable to cancel nonlinear components. The BF attenuates frequencies which arrive at the microphone array from outside the BF steering angle. The BF can therefore attenuate frequencies in the nonlinear components of the echo, provided the echo arrives at the microphone array from an angle other than the BF steering angle.

Two categories of tests were performed to investigate the BF capacity to assist AER in canceling echo by reducing the nonlinear echo components. In the first category of tests, AER convergence data was examined to see if utilizing BF with AER would result in lower residual echo after cancelation by the LMS. In the second category of tests, double-talk testing was performed to see if employing BF with AER would result in less nonlinear echo leak during double talk. The female speech used in Section 5.2 was used for the far-end (echo generating) signal in the first category of tests. Male speech was used in the second category of tests for the near-end signal.

5.3.1 Yamaha Speaker Setup and AER Tuning

The Yamaha enclosure was placed at 90° to the microphone array at a distance of 10.75 cm to the nearest microphone, with the speakers facing 0° with respect to the array. The volume levels of an unmodified Yamaha PSG-S01 were observed as a reference for the volume levels for the Yamaha speakers. For playback of pink noise at -18 dBFS, the following volume levels were observed 50 cm in front of the unmodified Yamaha:

- Nominal volume (Win7 volume slider at 50%): 68 dBSPL
- Maximum volume (Win7 volume slider at 100%): 77 dBSPL

In addition to the above two volume levels, the USB speakerphone was provided an 83 dBSPL volume level. This purpose of this volume level was to generate nonlinear echo higher than that produced by an unmodified Yamaha at its maximum volume level.

Several of the AER tuning steps discussed in [15] were performed to improve AER performance for the USB speakerphone, including:

1. Tx gain calibration
2. Rx gain calibration
3. Initial volume table design for 68 and 77 dBSPL output volume levels
4. Adjustment of ADC analog gain to avoid ADC saturation for echo produced at 77 dBSPL volume level, and corresponding adjustment of AER Tx digital gain to maintain SLR.
5. Nonlinear analysis of echo path for nominal and maximum volume levels as described in [19].
6. DRC design to reduce nonlinear distortion identified in Step 5.

Additionally, a DAC HPF and Rx EQ were designed to high-pass filter the Rx signal and reduce low-frequency (< 250 Hz) nonlinear distortion identified in Step 5.

5.3.2 BF Nonlinear Echo Reduction

Testing was performed to evaluate the BF ability to reduce nonlinear echo at AER S_{in}. Echo was produced using the R20 tone sweeps described in [19] at the 83 dBSPL volume level.

5.3.2.1 Test Signals

The far-end signal configuration used to generate echo is shown in Table 4.

Source	Playback source	Angle to mic array (°)	Distance to mic array (cm)	dBSPL at 50 cm (*)
Far-end signal – sweep16kr20	Yamaha speakers	90	10.75	83

(*) Volume level for -18 dBFS pink noise playback

Table 3. Test of BF Nonlinear Echo Reduction – Far-End Signal

5.3.2.2 Tests

Test 1 – raw echo

AER/BF configuration:

- AER: disabled.

- AER Tx digital gain=0 dB.
- BF: pass through mic1.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 25.

Nonlinear distortion may be observed in the output for all frequencies in the tone sweep. The distortion is most prevalent in the 500-2000 Hz range. Frequencies below 250 Hz are significantly attenuated by the DAC HPF and Rx EQ. The initial transients in the output for each tone are due to the DRC.

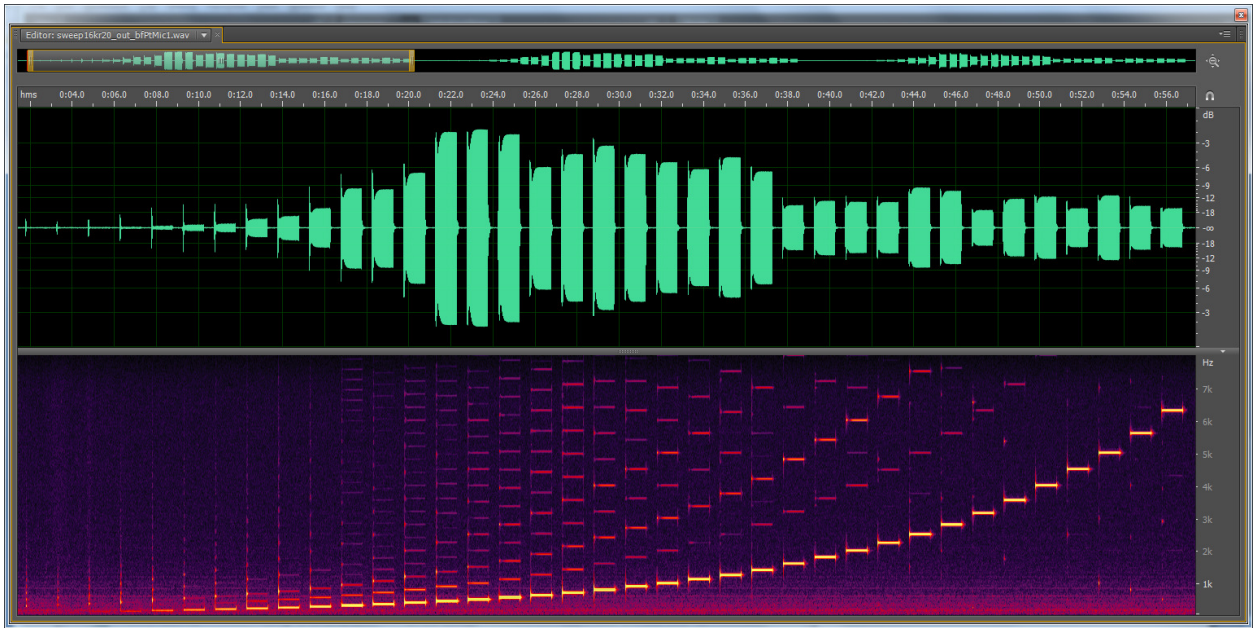

Figure 25. Test 1 – Tone Sweep Raw Echo, Time Domain and Spectrogram (BF Disabled; AER Disabled)

<u>Test 2 – BF effect on raw echo</u>

AER/BF configuration:

- AER: disabled.
- AER Tx digital gain=0 dB.
- BF: enabled, steering angle=0°.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 26.
- The spectra of the echo associated with the 800 Hz tone for the first tone sweeps in Test 1 and 2 are presented in Figure 27. The red line shows the spectrum for the BF in pass through, while the blue line shows the spectrum for the BF enabled.
- The spectra of the echo associated with the 3550 Hz tone for the first tone sweeps in Test 1 and 2 are presented in Figure 28. The red line shows the spectrum for the BF in pass through, while the blue line shows the spectrum for the BF enabled.

Comparing Figures 25 and 26, it can be seen the BF attenuates frequencies in the tone sweep echo beginning at 1.5 kHz. Both the fundamental and nonlinear distortion components are attenuated. This is further illustrated in Figures 27 and 28 as follows:

- Figure 27: The fundamental frequency is not considerably attenuated by the BF since is falls below 1.5 kHz. The nonlinear distortion above 1.5 kHz, however, is more significantly attenuated.
- Figure 26: The fundamental frequency and nonlinear distortion components are attenuated by the BF since they fall above 1.5 kHz.

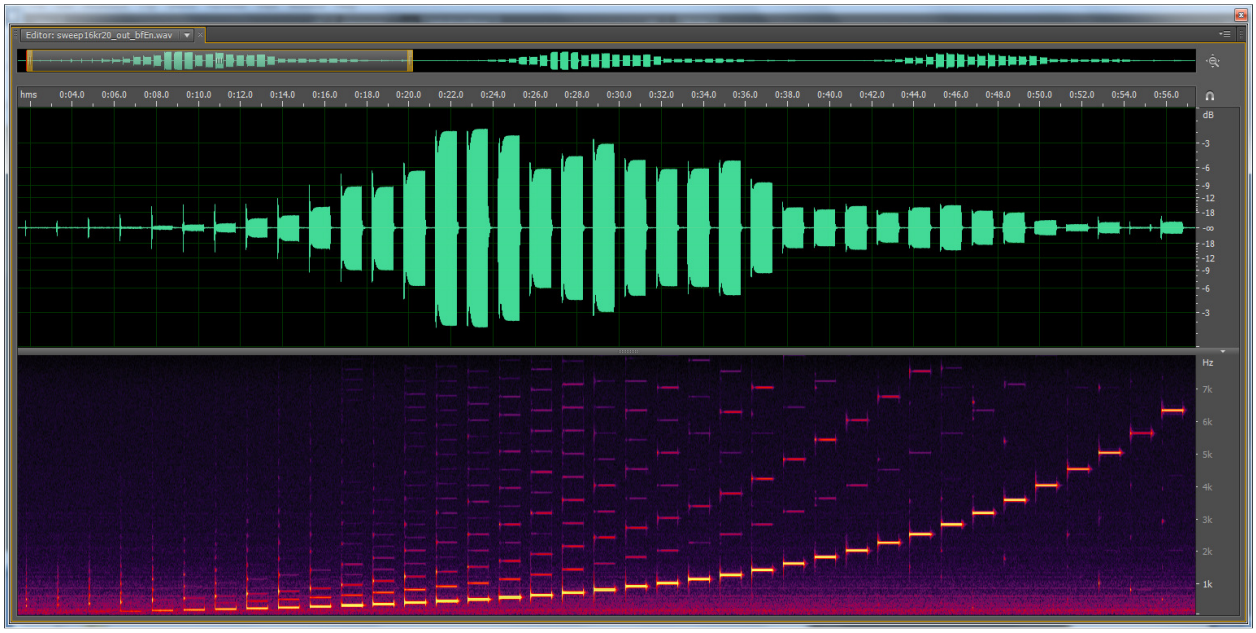

Figure 26. Test 2 – BF Effect on Tone Sweep Raw Echo, Time Domain and Spectrogram (BF Disabled; AER Disabled)

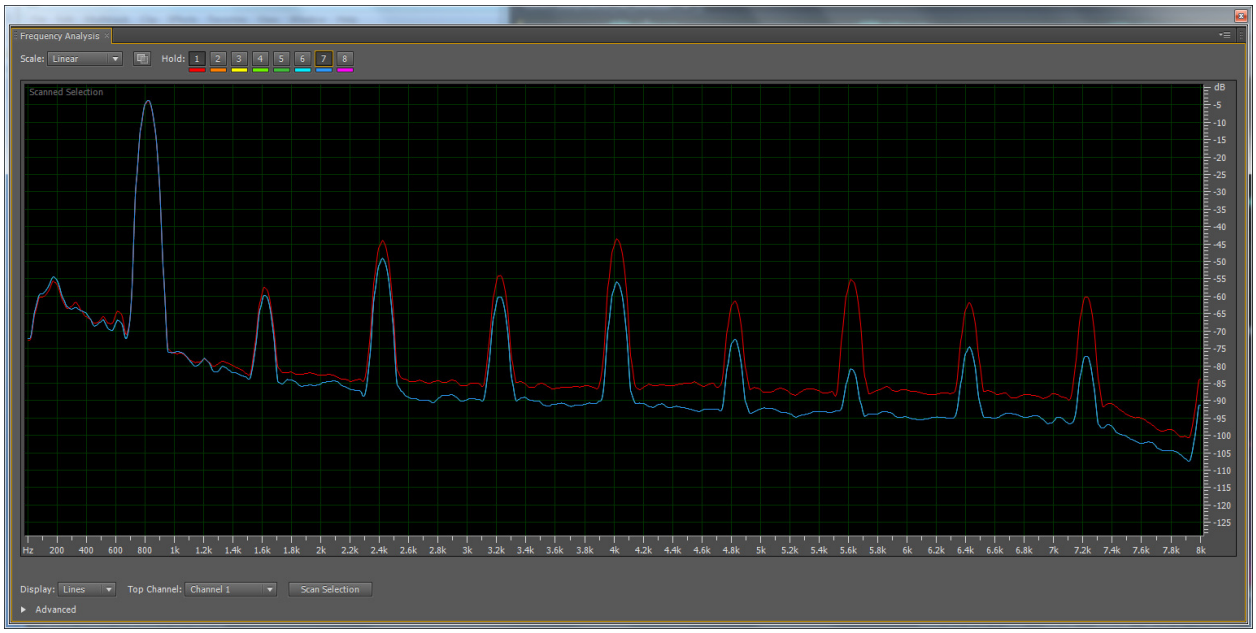

Figure 27. Test 1 and 2 Echo Spectra for 800 Hz Tone
(Red: BF Disabled, Blue: BF Enabled)

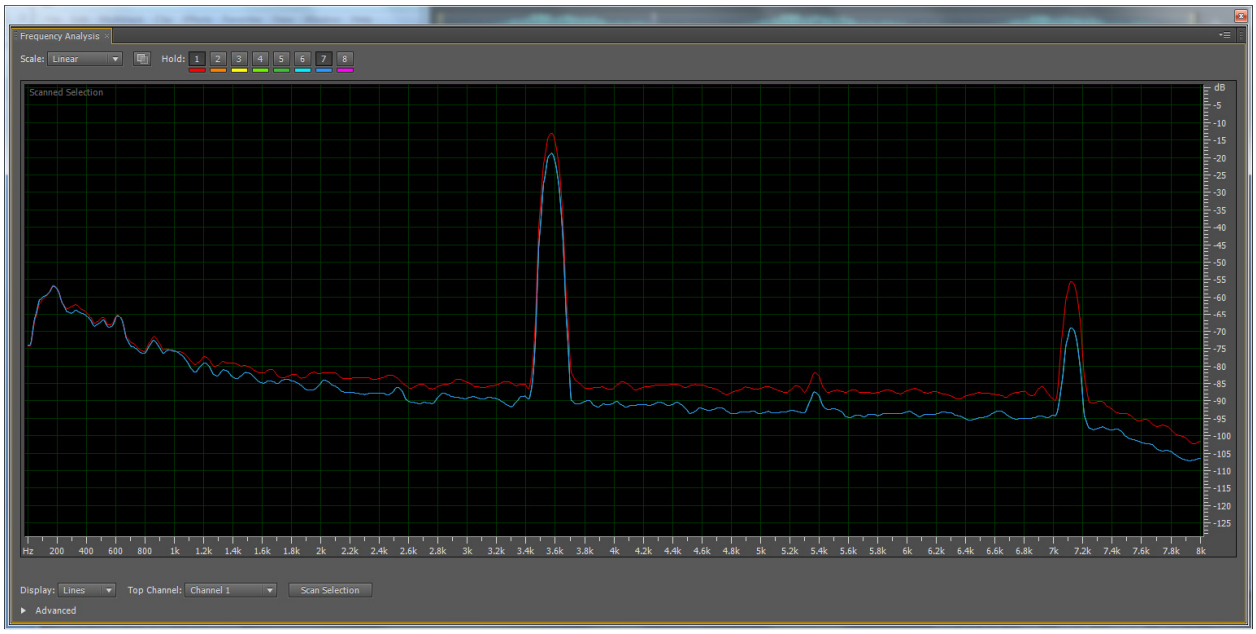

Figure 28. Test 1 and 2 Echo Spectra for 3550 Hz Tone
(Red: BF Disabled, Blue: BF Enabled)

5.3.2.3 Test Results

The Test 1-2 results indicate the BF attenuates echo frequency components above 1.5 kHz. Frequency components associated with linear echo are attenuated, as are components associated with nonlinear distortion. The BF thus reduces nonlinear echo sent to AER S_{in}. Since nonlinear distortion is highest in the 500-2000 Hz range, the BF contribution to nonlinear echo reduction will be correspondingly highest in this frequency range.

5.3.3 BF Nonlinear Echo Reduction – Effect on AER Convergence Depth

Testing of AER convergence depth using speech was performed to examine the BF capability to assist AER in canceling echo by reducing nonlinear echo at AER S_{in}. For the 68 dBSPL volume level, the ERL was high and the BF contribution was negligible. For the 77 dBSPL volume level, the BF contribution was discernible, but still small. Testing was therefore performed at the 83 dBSPL volume level to increase nonlinear echo.

5.3.3.1 Test Signals

The far-end signal configuration used to generate echo is shown in Table 4.

Source	Playback source	Angle to mic array (°)	Distance to mic array (cm)	dBSPL at 50 cm (*)
Far-end signal – female speech	Yamaha speakers	90	10.75	83

(*) Volume level for -18 dBFS pink noise playback

Table 4. Test of BF Effect on AER Convergence Depth – Far-End Signal

5.3.3.2 Tests

Test 1 – raw echo

AER/BF configuration:

- AER: disabled.
- BF: pass through mic1.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 29.

Figure 29. Test 1 – Raw Echo, Time Domain and Spectrogram (BF Disabled; AER Disabled)

Test 2 – BF effect on raw echo

AER/BF configuration:

- AER: disabled.
- BF: enabled, steering angle=0°.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 30.

Comparing Figures 29 and 30, it can be seen that BF has attenuated the echo presented to AER S_{in}.

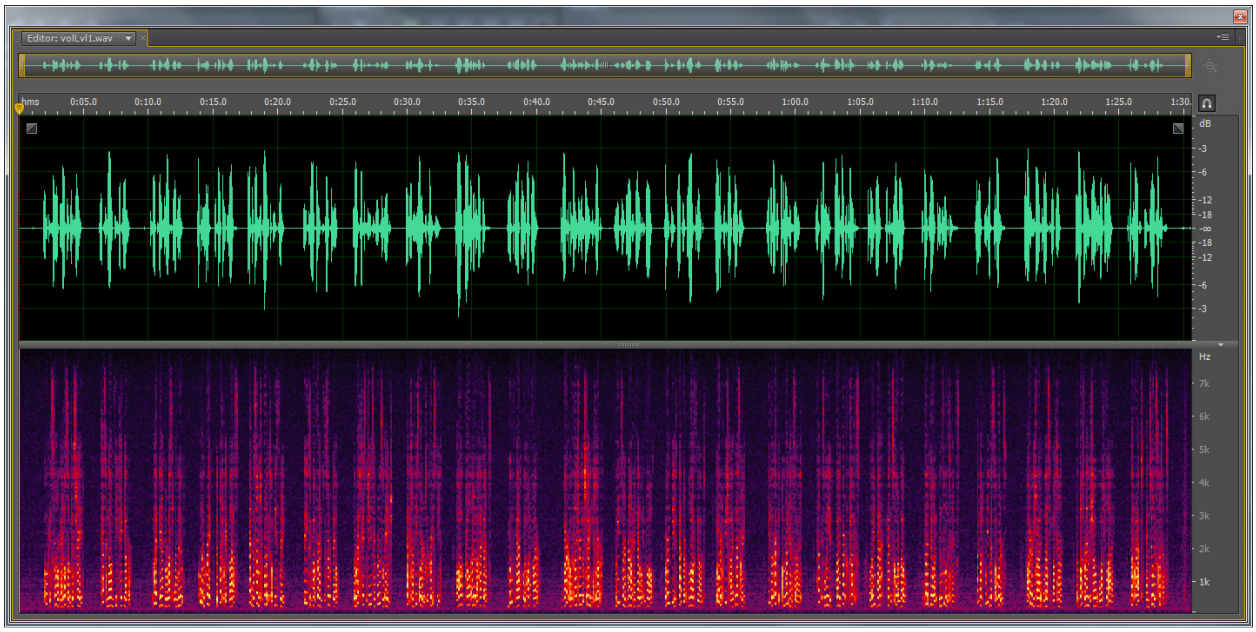

Figure 30. Test 2 – BF Effect on Raw Echo, Time Domain and Spectrogram (BF Enabled; AER Disabled)

Test 3 – AER convergence

AER/BF configuration:

- AER: disabled, enabled after 3rd speech segment (but NLP remains disabled)
- BF: pass through, mic1.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 31.

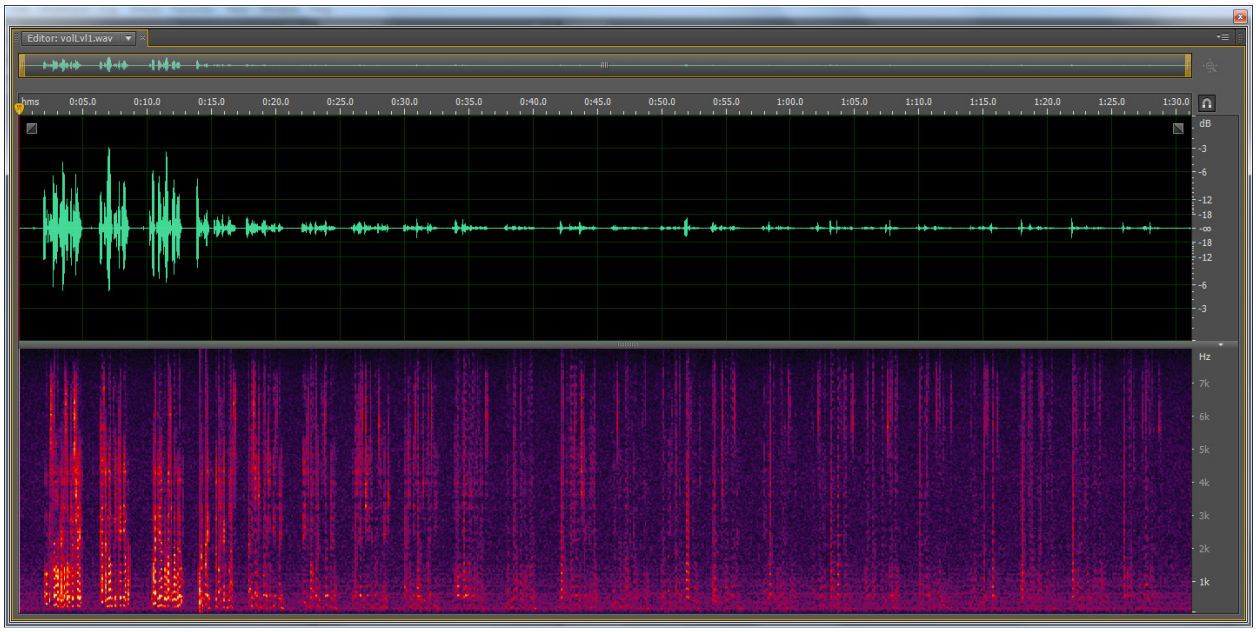

Figure 31. Test 3 – AER Convergence Depth, Time Domain and Spectrogram (BF Disabled; AER Enabled After 3rd Speech Segment)

Test 4 – BF effect on AER convergence

AER/BF configuration:

- AER: disabled, enabled after 3rd speech segment (but NLP remains disabled)
- BF: enabled, steering angle=0°.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram are presented in Figure 32.

Comparing Figures 31 and 32, it can be seen the residual echo after the LMS filter in AER is less when using BF. This is further illustrated in Figure 33, which plots the Test 3 and 4 residual echo power, and shows that BF can help reduce residual echo by up to 6dB for certain speech portion.

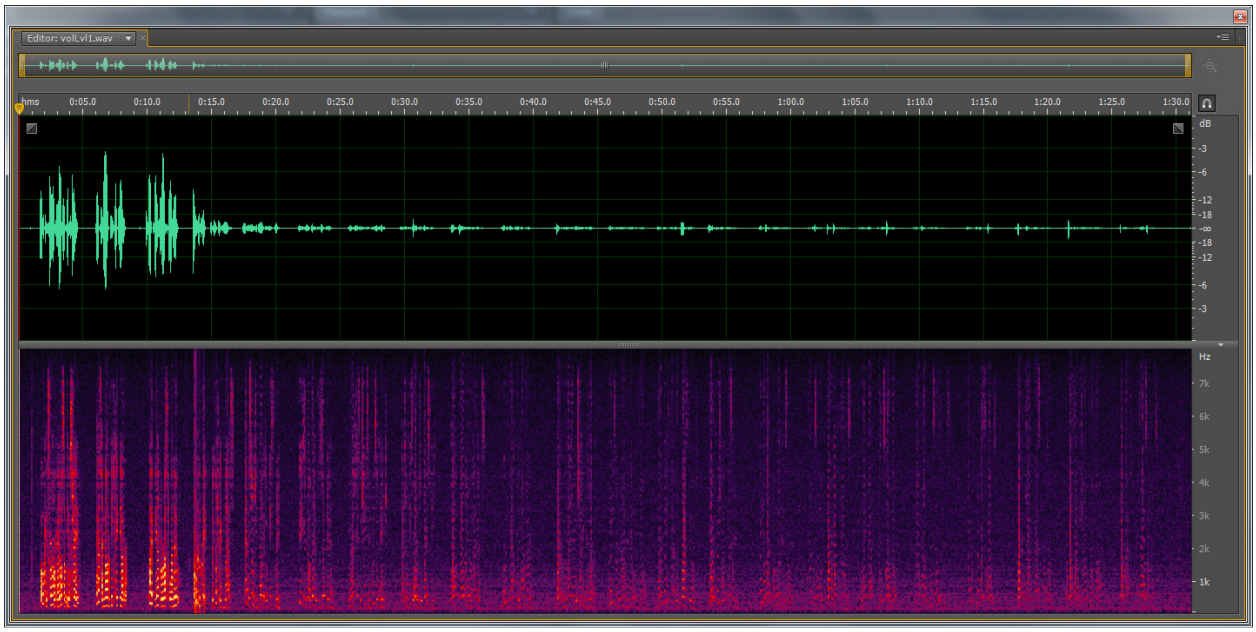

Figure 32. Test 4 – BF Effect on Convergence Depth, Time Domain and Spectrogram (BF Enabled, AER Enabled After 3rd Speech Segment)

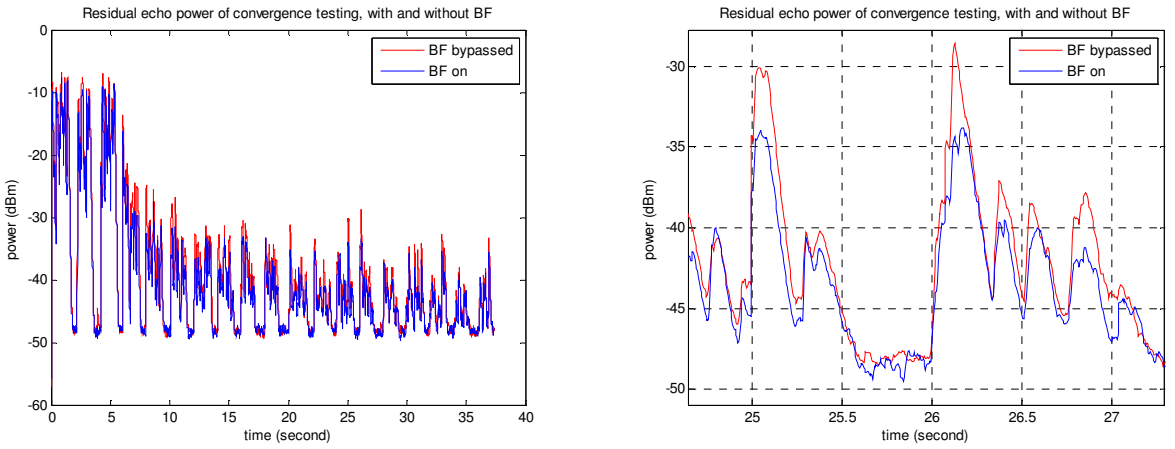

Figure 33. Test 3 and 4 – BF Effect on Convergence Depth, Power Measurement for Entire Test (Left) and Two Zoomed in Segments (Right)

5.3.3.3 Test Results

The Test 1-4 results indicate BF helps AER cancel echo by attenuating the nonlinear echo sent to AER S_{in}.

5.3.4 BF Nonlinear Echo Reduction – Effect on Double-Talk Echo Leak

Testing of echo leak during double talk was performed to examine the BF capability to assist AER in canceling echo by reducing nonlinear echo at AER S_{in}. For the 68 dBSPL volume level, echo leak during double talk was insignificant. For the 77 dBSPL volume level, echo leak during double talk was perceptible, but small. Testing was therefore performed at the 83 dBSPL volume level to increase nonlinear echo.

5.3.4.1 Test Signals

The far- and near-end signal configuration used to generate double-talk is shown below in Table 5.

Source	Playback source	Angle to mic array (°)	Distance to mic array (cm)	dBSPL at 50 cm (*)
Far-end signal – female speech	Yamaha speakers	90	10.75	83
Near-end signal – male speech	Monitor	0	5	80

(*) Volume level for -18 dBFS pink noise playback

Table 5. Test of BF Effect on Double-Talk Echo Leak - Near- and Far-End Signals

The near- and far-end signals were arranged to overlap approximately 75%, with the far-end speech occurring first. Figure 34 presents the time domain plot and spectrogram of the near-end speech in the proximity of the 11th near- and far-end speech segments, while the time domain plot and spectrogram for the far-end speech in the same vicinity is presented in Figure 35. The circled locations in the spectrograms correspond to regions which contain nonlinear echo in the Tests 1-3 results below.

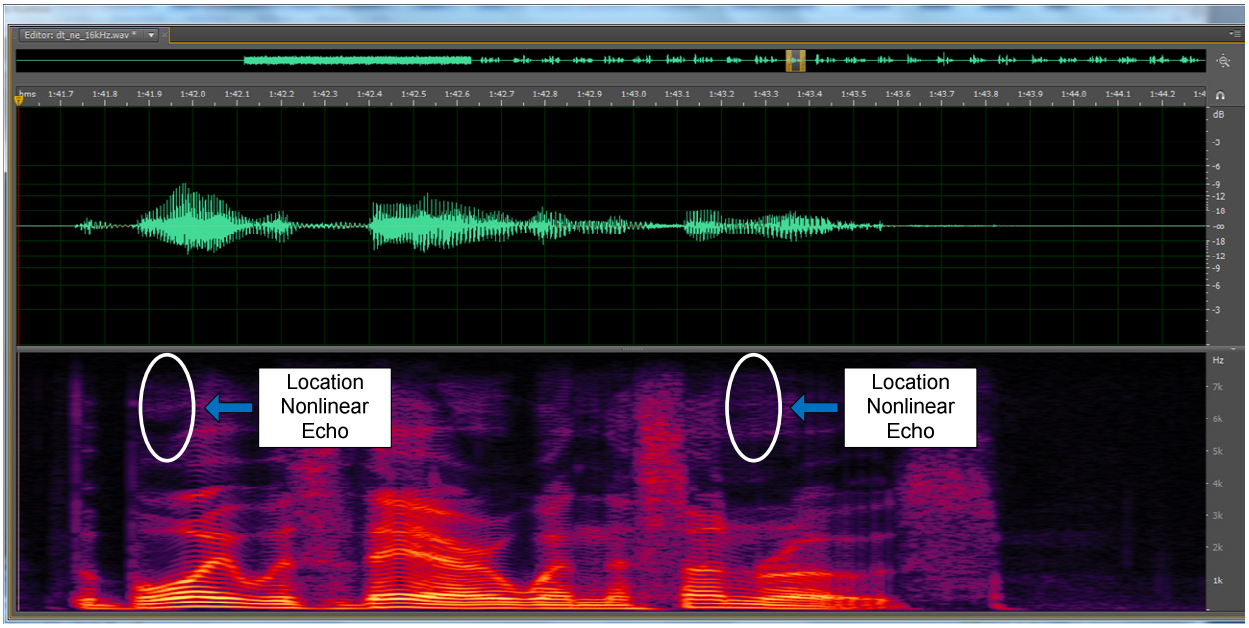

Figure 34 – Time Domain Plot and Spectrogram of Original Near-End Speech Around 11th Speech Segments

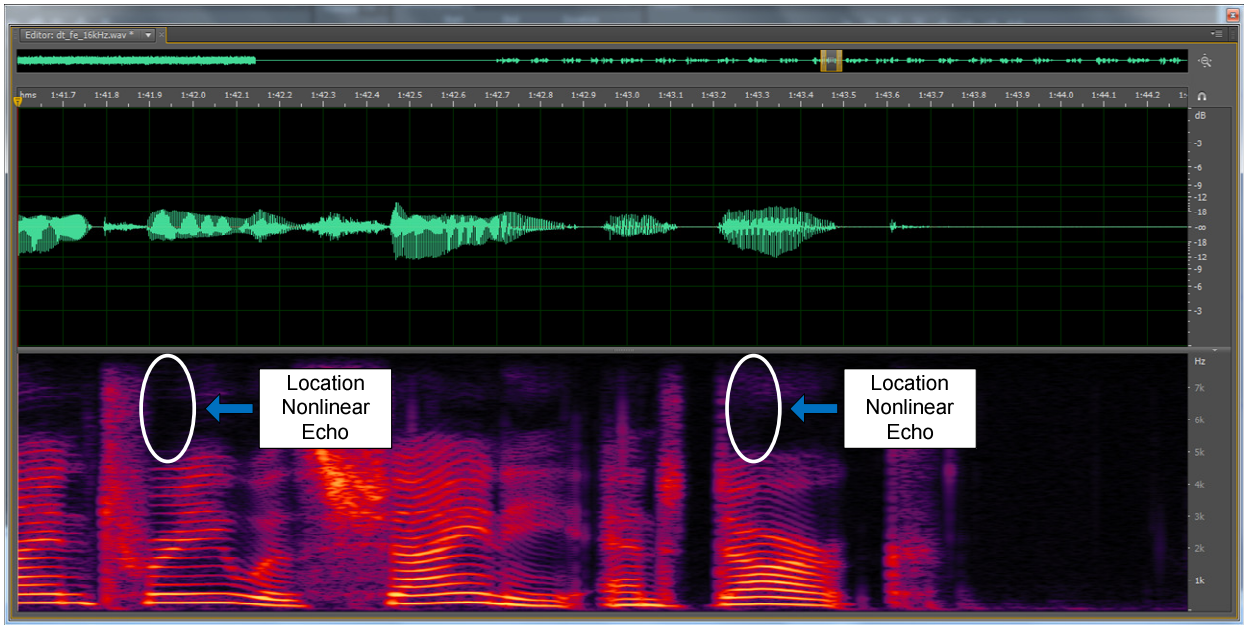

Figure 35 – Time Domain Plot and Spectrogram of Original Far-End Speech Around 11th Speech Segments

5.3.4.2 Tests

Test 1 – double talk, BF & AER disabled

AER/BF configuration:

- AER: disabled.
- BF: disabled.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram around the 11th speech segments are presented in Figure 36. Spectrum content not present in either the original near- or far-end speech is circled in the spectrogram. This must be nonlinear echo.

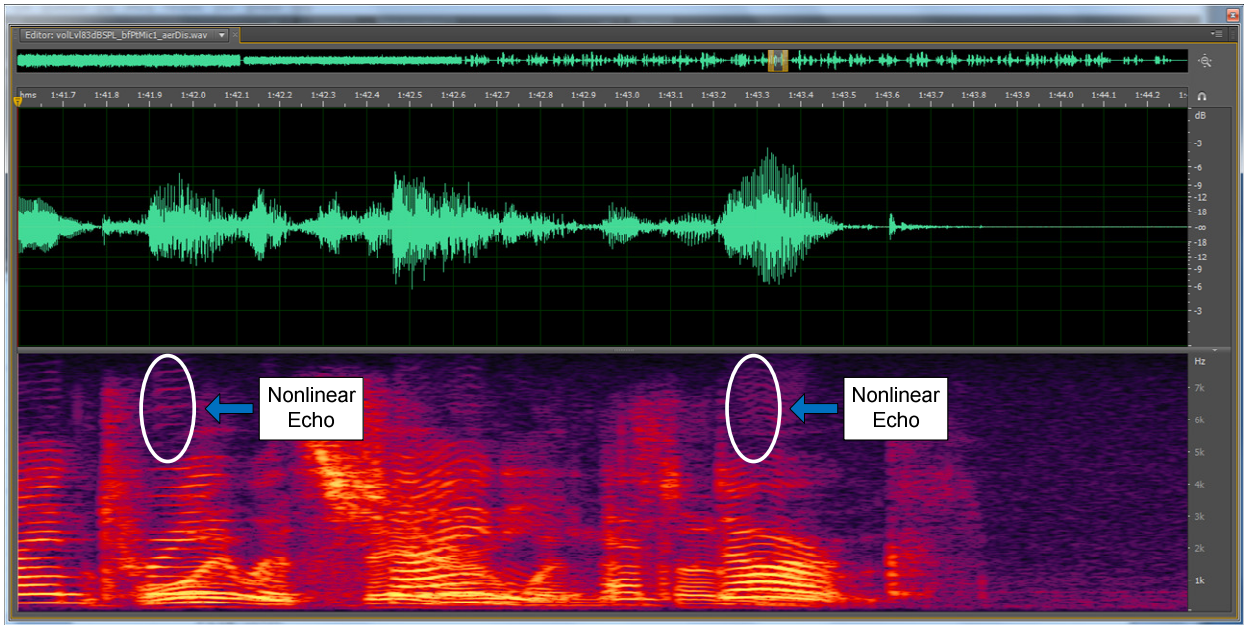

Figure 36. Test 1 – AER Send Path Output (AER S_{out}), BF & AER Disabled (Mixed Near-End Speech and Far End Echo)

Test 2 – double talk, AER enabled

AER/BF configuration:

- AER: enabled.
- NLP enabled and configured to have good duplex but allow slight echo leak
- BF: disabled.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram around the 11th speech segments are presented in Figure 37. Nonlinear echo is circled in the spectrogram.

Comparing Figures 36 and 37, it can be seen AER has removed most of the far-end echo, but some nonlinear echo has leaked through. NLP could be configured more aggressive to attack the nonlinear echo, but that would damage near end speech more.

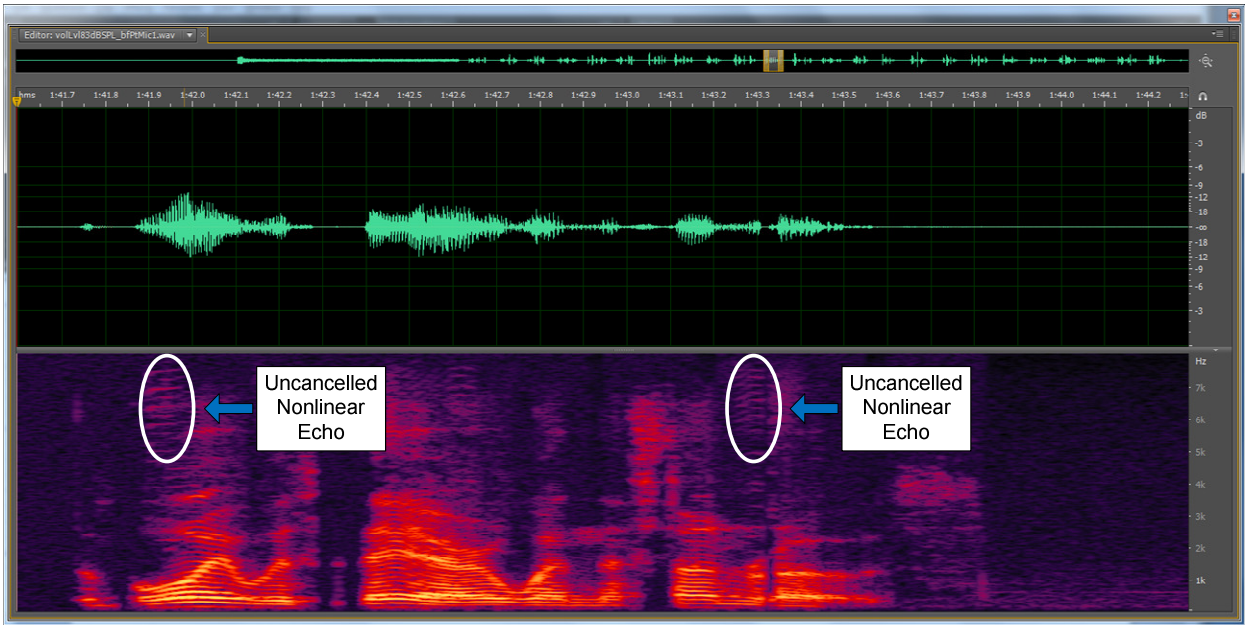

Figure 37. Test 2 – AER Send Path Output (AER S_{out}), BF Disabled & AER Enabled

Test 3 – double talk, BF & AER enabled

AER/BF configuration:

- AER: enabled.
- NLP enabled. nlp_clip_agg_l2=3, nlp_combloss_target=989.
- BF: enabled, steering angle=0°.

Test(s):

- Capture AER Tx output.

Capture(s):

- The time domain plot and spectrogram around the 11th speech segments are presented in Figure 38. Nonlinear echo is circled in the spectrogram.

Comparing Figures 36 and 38, it can be seen AER has removed most of the far-end echo, but some nonlinear echo has leaked through. Comparing Figures 37 and 38, it can be seen using BF with AER has resulted in reduced nonlinear echo leak.

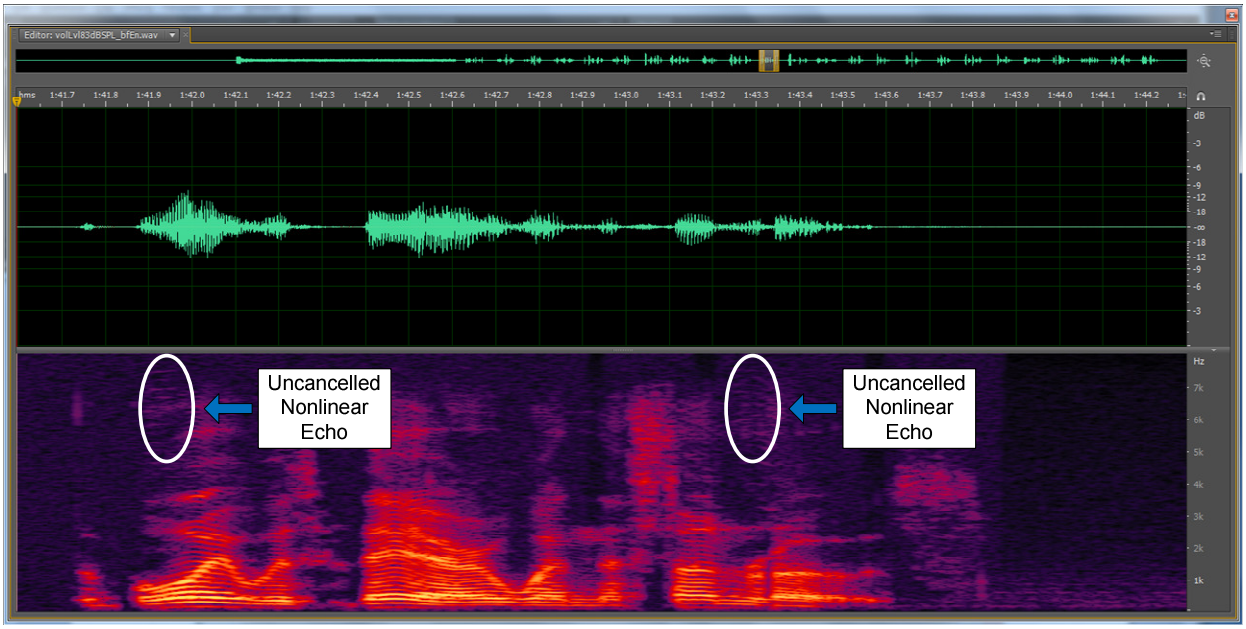

Figure 38. Test 4 – AER Send Path Output (AER S_{out}), BF & AER Enabled

5.3.4.3 Test Results

The Test 1-4 results indicate using BF with AER reduces nonlinear echo leak during double talk. It is a fact that many consumer communications devices have low quality speakers, and it is also desirable to have high volumes during conversations. Therefore, nonlinear distortion caused by overdriving low quality speakers can be frequent and may be severe at times. To control nonlinear echo, the NLP will have to be configured aggressively and that may severely damage near end speech. With the help of BF, nonlinear echo, especially at higher frequencies, can be reduced (by up to 6dB as shown in the specific tests shown above), which allows NLP to be configured less aggressively. This will in turn improve the duplex performance with the same speaker and at the same volume level.

6 Conclusions

The BF ability to assist ASNR with near-end noise reduction and AER with acoustic echo cancelation was investigated. The study was performed using a USB speakerphone test platform which employs hardware, speakers, and microphones characteristic of those found in commercially available portable speakerphones. The results of the study indicate the BF can help ASNR with near-end noise reduction: (1) BF with a relaxed ASNR provides the same level of noise reduction and improved speech quality vs. an aggressive ASNR alone; and (2) BF is able to attenuate non-stationary noise like typing noise, while ASNR is unable to attenuate this type of noise. The test results also indicate BF can help AER with echo cancelation by reducing nonlinear echo at the AER send input.

7 Bibliography

[1] Texas Instruments, Inc., "AER_16_1_0_3," 14 05 2013. [Online]. Available: http://software-dl.ti.com/libs/aer/latest/index_FDS.html. [Accessed 01 08 2013].

[2] Texas Instruments, Inc., "AER_Fact_Sheet.pdf," Texas instruments, Inc., 2013.

[3] A. Geensted, "The Lab Book Pages," 11 29 2010. [Online]. Available: http://www.labbookpages.co.uk/audio/beamforming.html. [Accessed 1 08 2013].

[4] Spectrum Digital, Inc., "C5517 EVM Support Page (Revision B)," 2013. [Online]. Available: http://support.spectrumdigital.com/boards/evm5517/revb/. [Accessed 01 08 2013].

[5] Texas Instruments, Inc., "TMS320C5517 Fixed-Point Digital Signal Processor Datasheet," 2012.

[6] Texas Instruments, Inc., "TLV320AIC3204 - Texas Instruments," 10 2012. [Online]. Available: http://www.ti.com/lit/ds/symlink/tlv320aic3204.pdf. [Accessed 06 08 2013].

[7] Texas Instruments, "TPA2012D2," 06 2008. [Online]. Available: http://www.ti.com/lit/ds/symlink/tpa2012d2.pdf. [Accessed 06 08 2013].

[8] Micron, "MT48H4M16LF - Micron," 2004. [Online]. Available: http://download.micron.com/pdf/datasheets/dram/mobile/Y25L_64Mb.pdf. [Accessed 05 08 2013].

[9] Texas Instruments, Inc., "MKARRAY-A_RevA.pdf," Texas Instruments, Inc., 2012.

[10] "SPM1423HM4H-B - DigiKey," 2012. [Online]. Available: http://media.digikey.com/pdf/Data%20Sheets/Knowles%20Acoustics%20PDFs/SPM1423HM4H-B.pdf. [Accessed 06 08 2013].

[11] Yamaha, "PSG-01S User's Manual - Yamaha," [Online]. Available: http://www.yamaha.co.jp/soundgadget/download/pdf/PSG-01S_en.pdf. [Accessed 01 08 2013].

[12] Texas Instruments, Inc., "C55x Connected Audio Framework," 15 05 2013. [Online]. Available: http://www.ti.com/tool/c55x-audioframework. [Accessed 01 08 2013].

[13] Spectrum Digital, Inc., "EVM5515 Support Home (Revision B)," 2013. [Online]. Available: http://support.spectrumdigital.com/boards/evm5515/revb/. [Accessed 01 08 2013].

[14] Texas Instruments, Inc., "connected-audio-documentation.pdf," 15 05 2015. [Online]. Available: http://www.ti.com/tool/c55x-audioframework. [Accessed 01 08 2013].

[15] Texas Instruments, Inc., "AER_Quick_Tuning_Guide.pdf," Texas Instruments, Inc., 2013.

[16] Echo Audio, "Downloads," 2013. [Online]. Available: http://files.echoaudio.com/manuals/audiofire_windows_manual_v2.2.pdf. [Accessed 08 08 2013].

[17] Mackie, "HR624 MK2 - Mackie," 2007. [Online]. Available: http://www.mackie.com/products/hrmk2series/downloads/hr624mk2_om.pdf. [Accessed 08 08 2013].

[18] E. Hänsler and G. Schmidt, Acoustic Echo and Noise Control: A Practical Approach, Wiley, 2005.

[19] Texas Instruments, Inc., "NonlinearTool.pdf," Texas instruments, Inc., 2013.

8 Appendix A: Summary of Test Captures

BF/ASNR, Fan Noise	
Test	**Capture (bfAsnr_fanNoise\...)**
1	sigFemaleSpeech_noiseFan_bfPtMic1_asnrDis.wav
2	sigFemaleSpeech_noiseFan_bfPtMic1_asnrEn21dB.wav
3	sigFemaleSpeech_noiseFan_bfEn_asnrDis.wav
4	sigFemaleSpeech_noiseFan_bfEn_asnrEn21_18_12dB.wav
BF/ASNR, Typing Noise	
Test	**Capture (bfAsnr_typingNoise\...)**
1	sigNone_noiseTyping_bfPtMic1_asnrDis.wav
2	sigNone_noiseTyping_bfPtMic1_asnr18dB.wav
3	sigNone_noiseTyping_bfEn_asnrDis.wav
4	sigNone_noiseTyping_bfEn_asnrEn18_15_12.wav
5	sigFemaleSpeech_noiseTyping_bfPtMic1_asnrDis.wav
6	sigFemaleSpeech_noiseTyping_bfEn_asnrDis.wav
BF/AER, Nonlinear Echo Reduction	
Test	**Capture (bfAer_bfNLEchoReduction\…)**
1	sweep16kr20_out_bfPtMic1.wav
2	sweep16kr20_out_bfEn.wav
BF/AER, BF Effect on AER Convergence	
Test	**Capture (bfAer_bfEffectAerConvergence\...)**
1	aerDis_bfPtMic1.wav
2	aerDis_bfEn.wav
3	aerEn_bfPtMic1.wav
4	aerEn_bfEn.wav
BF/AER, BF Effect on DT Echo Leak	
Test	**Capture (bfAer_bfEffectDTEchoLeak\...)**
1	volLvl83dBSPL_bfPtMic1_aerDis.wav
2	volLvl83dBSPL_bfPtMic1.wav
3	volLvl83dBSPL_bfEn.wav