Quality Enhancement of Voice Communication & Audio experience

Application Assignment

Suyash Bagad   Mohak Sahu   Rohan Pathak

Department of Electrical Engineering
Indian Institute of Technology, Bombay

7th April, 2018

Outline

Although we have sufficient resources and already existing technologies in the field of audio and voice enhancement, there are a lot of opportunities for implementing advanced algorithms using Digital Signal Processing for audio and voice enhancements.

Sound Effect Normalization
- SEN Basics and implementation

Acoustic Echo Cancellation
- Background

Mono to Stereo Conversion
- Definitions
- Conversion approach & Results

Outline

Sound Effect Normalization

SEN Basics and implementation

Acoustic Echo Cancellation

Background

Mono to Stereo Conversion

Definitions

Conversion approach & Results

Sound Effect Normalization(SEN)

Sound Effect Normalization(SEN) can be used to improve speech intelligibility in movies through de-emphasis of background sounds and music. Most of the times it is applicable for television shows and sports where we have a lot of background noise due to people shouting and music playing.
Digital Audio Effects:- Audio effects can be classified by the way do their processing. We studied to of them:

- **Basic Filtering :** Lowpass, Highpass filter etc, Equaliser
- **Time Varying Filters :** Wah-wah

Basic Digital Audio Filtering Effects: Equalisers

- **Filters** by definition remove/attenuate audio from the spectrum above or below some cut-off frequency.
  But for many audio applications this a little too restrictive.

Basic Digital Audio Filtering Effects: Equalisers

- **Filters** by definition remove/attenuate audio from the spectrum above or below some cut-off frequency.
  But for many audio applications this a little too restrictive.
- Therefore **Equalisers** are used.
  Equalisers, by contrast, enhance/diminish certain frequency bands whilst leaving others unchanged.

Basic Digital Audio Filtering Effects: Equalisers

- **Filters** by definition remove/attenuate audio from the spectrum above or below some cut-off frequency.
  But for many audio applications this a little too restrictive.
- Therefore **Equalisers** are used.
  Equalisers, by contrast, enhance/diminish certain frequency bands whilst leaving others unchanged.
  They are built using a series of **Shelving** and **Peak** filters.
  Usually second-order filters are employed.

Shelving Filter

Boost or cut the low or high frequency bands with a cut-off frequency, Fc and gain G.

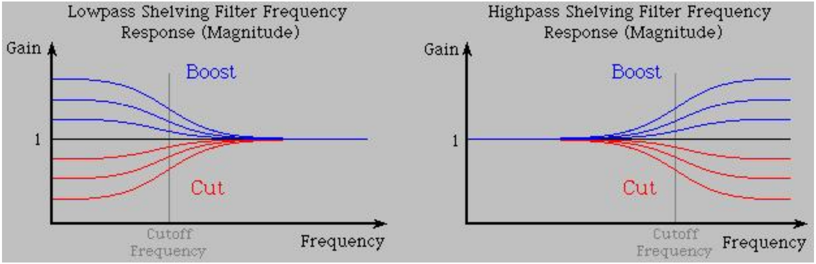


Figure: Shelving Filter Frequency Response

Shelving Filter (Cont.)

A first-order shelving filter may be described by the transfer function:

$$H(z) = 1 + H_0/2 * (1 + A(z)) \qquad .....(ForLowPass)$$
$$H(z) = 1 + H_0/2 * (1 - A(z)) \qquad .....(ForHighPass)$$

Shelving Filter (Cont.)

A first-order shelving filter may be described by the transfer function:

$$H(z) = 1 + H_0/2 * (1 + A(z)) \qquad .....(ForLowPass)$$
$$H(z) = 1 + H_0/2 * (1 - A(z)) \qquad .....(ForHighPass)$$

where A(z) is a first-order all pass filter passes all frequencies but modifies phase:
For Boost:

$$A(z) = \frac{z^{-1} + a_B}{1 + a_B z^{-1}}$$

Similarly, for Cut we use a_C

Shelving Filter (Cont.)

The gain, G, in dB can be adjusted accordingly:

$$H_0 = V_0 - 1$$

$$\textit{where} \qquad V_0 = 10^{(G/20)}$$

Shelving Filter (Cont.)

The gain, G, in dB can be adjusted accordingly:

$$H_0 = V_0 - 1$$

$$where \qquad V_0 = 10^{(G/20)}$$

and the cut-off frequency for boost or cut are given by:

$$a_B = \frac{tan(2\pi fc/fs) - 1}{tan(2\pi fc/fs) + 1}$$

$$a_C = \frac{tan(2\pi fc/fs) - V_0}{tan(2\pi fc/fs) + V_0}$$

Peak Filter

Boost or cut mid-frequency bands with a cut-off frequency, Fc, a bandwidth, fb and gain G.

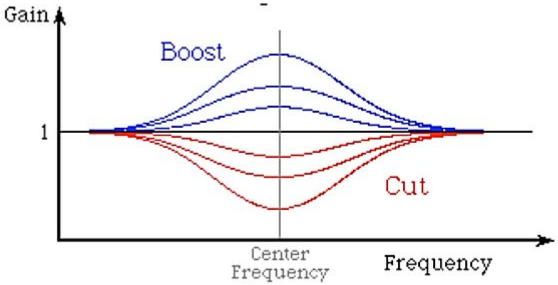


Figure: Peak Filter Frequency Response(Magnitude)

Peak Filter (Contd.)

A second-order peak filter may be described by the transfer function:

$$H(z) = 1 + H_0/2 * (1 + A(z))$$

Peak Filter (Contd.)

A second-order peak filter may be described by the transfer function:

$$H(z) = 1 + H_0/2 * (1 + A(z))$$

where A(z) is a second-order all pass filter
For Boost:

$$A(z) = \frac{-a_B + (d - da_B)z^{-1} + z^{-2}}{1 + (d - da_B)z^{-1} + a_B z^{-2}}$$

Similarly for Cut, we use a_C

Peak Filter (Contd.)

A second-order peak filter may be described by the transfer function:

$$H(z) = 1 + H_0/2 * (1 + A(z))$$

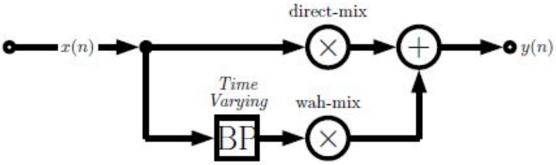
where A(z) is a second-order all pass filter
For Boost:

$$A(z) = \frac{-a_B + (d - da_B)z^{-1} + z^{-2}}{1 + (d - da_B)z^{-1} + a_B z^{-2}}$$

Similarly for Cut, we use a_C
The center/cut-off frequency, d, is given by:
d = -cos($2\pi fc/fs$)

Time-varying Filters

Wah-wah : A bandpass filter with a time varying centre (resonant) frequency and a small bandwidth. Filtered signal mixed with direct signal.
The signal flow for a wah-wah is as follows:

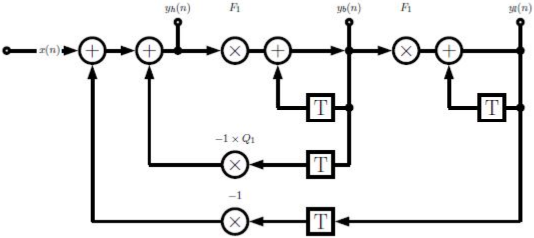


where BP is a time varying frequency bandpass filter.
It alters the tone and frequencies of the audio signal to create a distinctive sound, mimicking the human voice saying the onomatopoeic name "wah-wah". Hence the name.

Time Varying Filter Implementation: State Variable Filter

In time varying filters, independent control over the cut-off frequency and damping factor of a filter is needed.
A **State Variable Filter** is implemented for the same.
One further advantage is that lowpass, bandpass and highpass filter output are obtained simultaneously.
A state variable filter is a type of active filter. It consists of one or more integrators, connected in some feedback configuration.
A state variable filter realizes the state-space model directly.

The State Variable Filter



where:

$$\begin{aligned} \text{x(n)} &= \text{input signal} \\ \text{yl(n)} &= \text{lowpass signal} \\ \text{yb(n)} &= \text{bandpass signal} \\ \text{yh(n)} &= \text{highpass signal} \end{aligned}$$

and tuning coefficients F1 and Q1 related to the cut-off frequency, fc, and damping, d:

$$\text{F1} = 2\sin(2\pi fc/fs), and Q1 = 2d$$

Outline

Sound Effect Normalization
SEN Basics and implementation

Acoustic Echo Cancellation
Background

Mono to Stereo Conversion
Definitions
Conversion approach & Results

Background & Motivation

- Echo is a phenomenon where a delayed and distorted version of an original sound or electrical signal is reflected back to the source. Since the advent of telephony echoes have been a problem in communication networks
- Electrical echo due to impedance mismatch
- Acoustic echo is due to the coupling between the loudspeaker and microphone
- Did you know?
  - The range of human hearing is generally considered to be 20 Hz to 20 kHz, but it is far more sensitive to sounds between 1 kHz and 4 kHz. For example, listeners can detect sounds as low as 0 dB SPL at 3 kHz, but require 40 dB SPL at 100 hertz (an amplitude increase of 100)

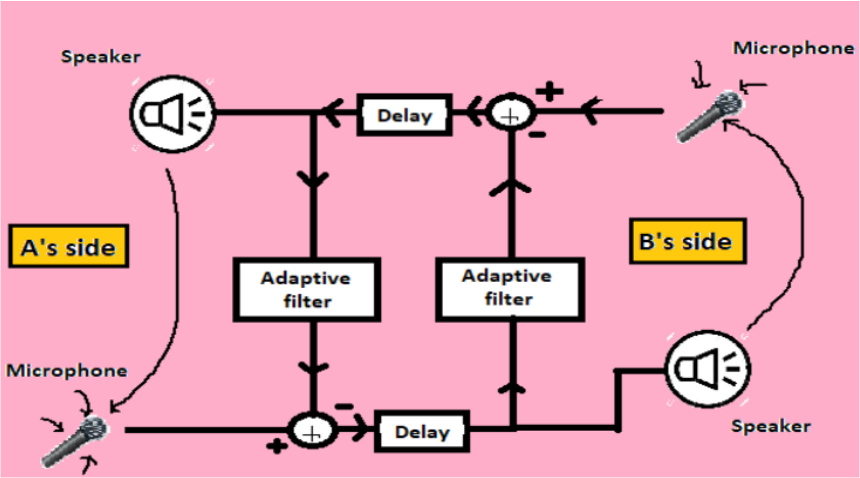


Figure: Basic block diagram of AEC Implementation

AEC Generation block

- A measured microphone signal d(n) contains two signals: The near-end speech signal r(n) The far-end echoed speech signal d(n)
- The goal is to remove the far-end echoed speech signal from the microphone signal.

$$e(n) = d(n) - \hat{y}(n)$$

e(n) in turns adjust the coefficients

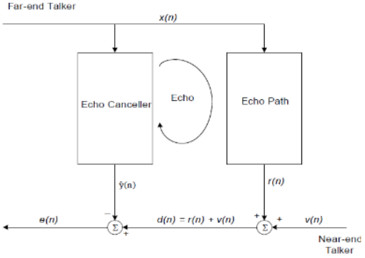


Figure: AEC Generation

Adaptive filter

- The closed loop digital adaptive filter uses feedback in the form of an error signal to refine its transfer function
- Least mean squares (LMS) algorithms are a class of adaptive filter used to mimic a desired filter by finding the filter coefficients that relate to producing the least mean square of the error signal
- [Here] the adaptive filter tends to minimize

$$[e(n)]^2$$

and accordingly adjust

$$h^*[n]$$

Outline

Sound Effect Normalization
- SEN Basics and implementation

Acoustic Echo Cancellation
- Background

Mono to Stereo Conversion
- Definitions
- Conversion approach & Results

What is Mono & Stereo sound?

- Stereo uses two (or more) independent audio channels to reproduce sound creating an illusion of multi-directional audible perspective
- Mono-aural sound is a single channel sound intended to reproduce sound perceived as coming from single position
- The practical difference is that mono has only one spatial dimension; something can be either close to (loud) og far away (quiet) from the listener. Stereo has two dimensions: close/far and left/right, giving a sense of space while listening to a stereo track

Outline

Sound Effect Normalization
- SEN Basics and implementation

Acoustic Echo Cancellation
- Background

Mono to Stereo Conversion
- Definitions
- Conversion approach & Results

Mono to Stereo Conversion approach

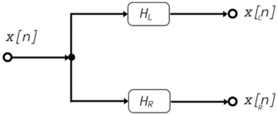


Figure: Basic blockscheme of the pseudo-stereo system

- Down-mix compatibility

$$
\begin{aligned}
x_M[n] &= x_L[n] + x_R[n] \\
&= x[n] * h_L[n] + x[n] * h_R[n] \\
&= x[n-D] \\
\Rightarrow \quad X_M(e^{j\Omega}) &= X_{(}e^{j\Omega})H_L(e^{j\Omega}) + X_{(}e^{j\Omega})H_R(e^{j\Omega}) \\
&= X(e^{j\Omega})e^{-j\Omega D}
\end{aligned}
$$

$$
\Rightarrow \quad H_L(e^{j\Omega}) + H_R(e^{j\Omega}) = e^{-j\Omega D}
$$

Continued...

- For a neutral sounding stereo output without coloration artifacts the sum of the magnitude frequency responses must be constant

$$\left|H_L(e^{j\Omega})\right| + \left|H_R(e^{j\Omega})\right| = 1$$

- For obtaining real-valued impulse responses both pseudo-stereo filters must feature conjugate symmetry

$$H_{L/R}(e^{j\Omega}) = H^*_{L/R}(e^{-j\Omega})$$

- Left channel decorrelation filter

$$H_L[k] = (1/2 + (1/\pi)\arctan(\omega^2 R[k]))e^{-j(2\pi kD/N)}$$

where R[k] is DFT of standard Gaussian Noise.

Time domain implementation

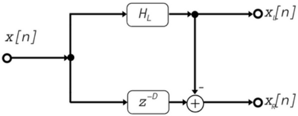


Figure: Blockscheme of the time-domain realization

- To get the left channel FIR filter impulse response, we have

$$h_L[n] = \mathcal{F}^{-1}(H_L[k])$$

- For the right channel,

$$x_R[n] = x[n-D] - x_L[n]$$

where

$$D = (N+1)/2$$

is the group delay of the FIR filter of length N.

Frequency domain implementation

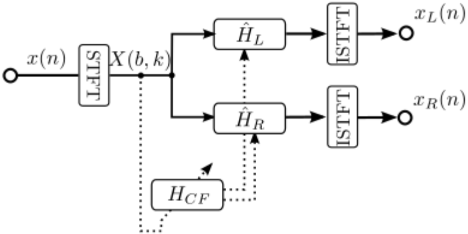


Figure: Blockscheme of the frequency-domain realization

- The time-frequency representations of the output channels are computed with a dot-wise product and a following over-lapping inverse STFT synthesis of the left and right output channel.
$$X_{L/R}(b,k) = X(b,k).H_{L/R}[k]$$

References I

Dave Marshall
Digital Audio Effects
Dave.Marshall_CM0268_PDF_10_CM0268_Audio_FX

Markus Christoph
Acoustic echo cancellation
United States Patent US12710092 (2009)

Marco Fink, Sebastian Kraft, Udo Zolzer,
Downmmix-compatible conversion from mono to stereo in time- and Frequency-domain
Proc. of the 18 th Int. Conference on Digital Audio Effects (DAFx-15), Trondheim, Norway, Nov 30 - Dec 3, 2015

References I

- MathWorks Inc.
  AEC in MATLAB
  MATLAB Examples
- Wikipedia
  Adaptive Filtering
- Some interesting forums for discussion on DSP
  Music Stack Exchange
  Science Encyclopedia
  DSP Stack Exchange