WEINTRAUB SPEECH SEPARATION SYSTEM

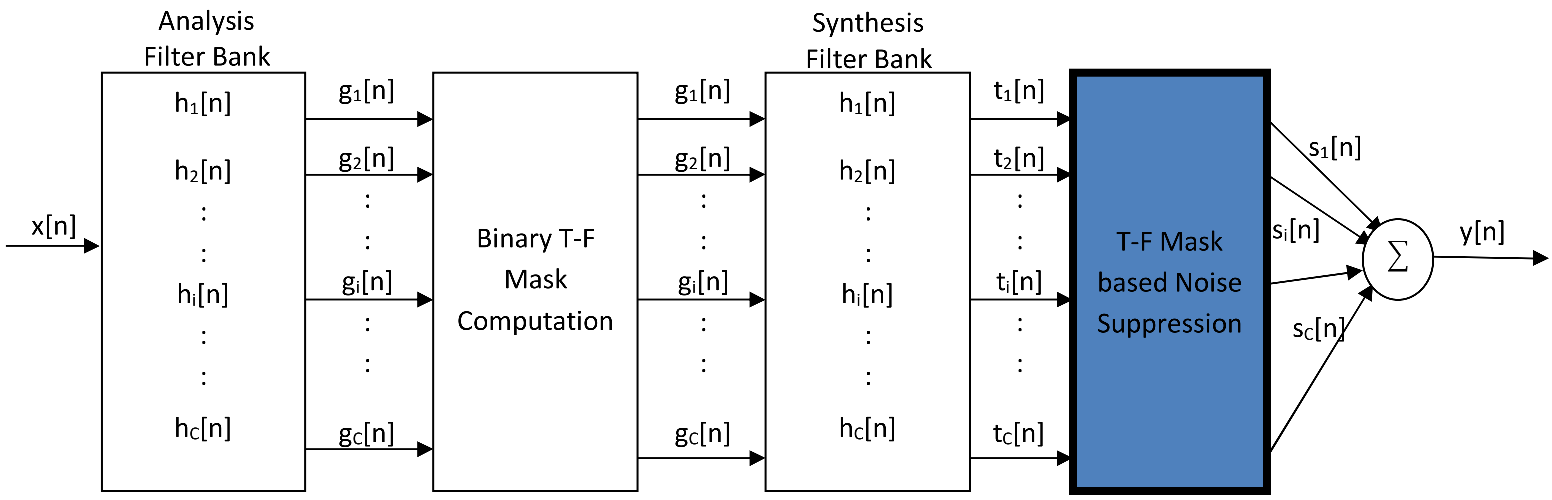

Figure 1: Block Diagram of Weintraub Speech Separation System

x[n] - Input Noisy Speech Signal
y[n] - Enhanced Speech Signal
C - Number of Channels
$h_i[n]$ - Impulse response of the Gammatone Filter; $1 \leq i \leq C$

Speech Analysis

$$g_i[n] = x[n] * h_i[n]$$
$$= \sum_{m=0}^{\infty} x[m].\, h_i[n-m]$$

where $h_i[n] = \quad An^{N-1} \exp(-2\pi b_i n)\cos(2\pi f_i n + \emptyset)u[n];$

A - amplitude,

Ø - phase,

N - filter order,

f_i - center frequency of the i^{th} filter,

b_i - bandwidth of the i^{th} filter.

Binary T-F Mask Computation

Speech Energy : $SE_{i,j} = \sum_{m=jR}^{jR+L-1} (gS_i[m])^2$

Noise Energy :$NE_{,i,j} = \sum_{m=jR}^{jR+L-1} (gN_i[m])^2$

where $SE_{i,j}$ - energy of the speech signal in i^{th} channel, j^{th} frame
$NE_{i,j}$ - energy of the noise signal in i^{th} channel, j^{th} frame
gS_i - filtered response of speech signal in i^{th} channel
gN_i - filtered response of noise signal in i^{th} channel
L - Frame length
R - Window shift (L/2)

The T-F Binary Mask is defined as

$$M(i,j) = \begin{cases} 1 & \text{if } SE_{i,j} > NE_{i,j} \\ 0 & \text{otherwise} \end{cases}$$

Speech Synthesis

$k_i[n] = f_i[n] * h_i[n]$; where $f_i[n] = g_i[-n]$
$= \sum_{m=0}^{\infty} f_i[m]\, h_i[n-m]$

$s_{i,j}[m] = \sum_{m=jR}^{jR+L-1} t_i[m]\, p_{i,j}[jR-m]$ where $t_i[n] = k_i[-n]$

and $p_{i,j} = \begin{cases} w[n] & \text{if } M(i,j) = 1 \\ 0 & \text{otherwise} \end{cases}$

w[n] is the sliding cosine window defined as,

$$w[n] = \begin{cases} 1 + \cos((2\pi(n-1)/L - \pi)/2 & 0 \le n \le L-1 \\ 0 & \text{otherwise} \end{cases}$$

and finally,

$y[n] = \sum_{i=1}^{C} s_i[n]$

PROPOSED SPEECH SEPARATION SYSTEM

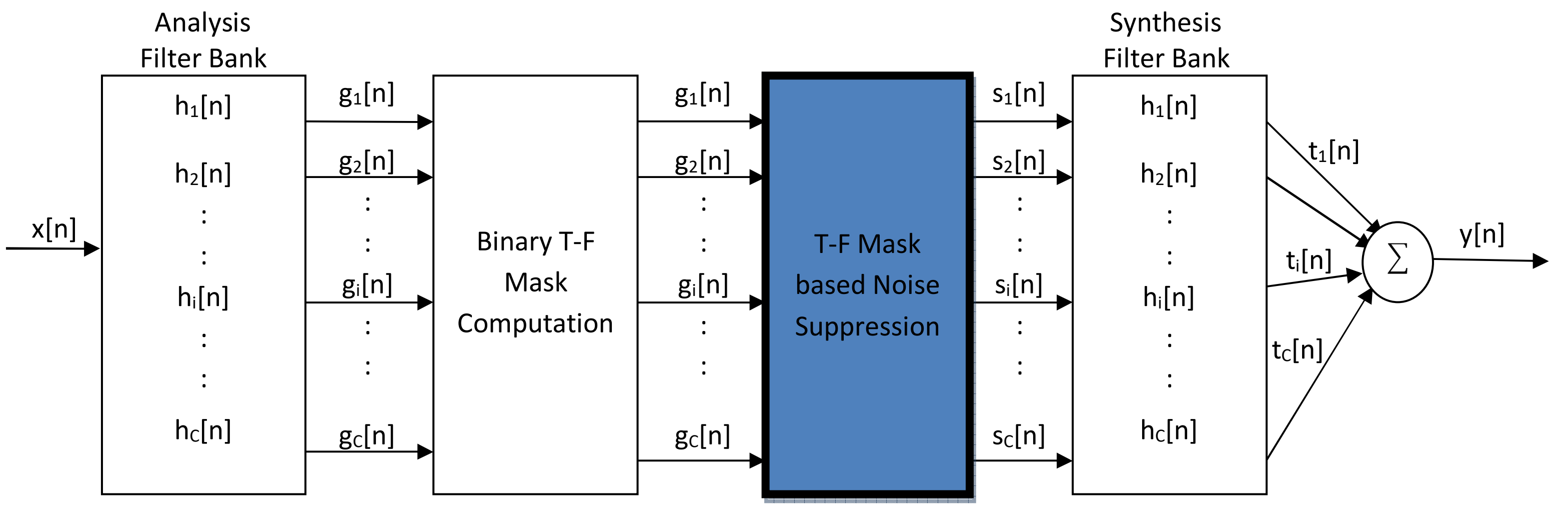

Figure 2: Block Diagram of Proposed Speech Separation System

x[n] - Input Noisy Speech Signal
y[n] - Enhanced Speech Signal
C - Number of Channels
$h_i[n]$ - Impulse response of the Gammatone Filter; $1 \leq i \leq C$

Speech Analysis

$$g_i[n] = x[n] * h_i[n]$$

where $h_i[n] = An^{N-1} \exp(-2\pi b_i n)\cos(2\pi f_i n + \emptyset)u[n]$;

A - amplitude,

$\emptyset$ - phase,

N - filter order,

f_i - center frequency of the i^{th} filter,

b_i - bandwidth of the i^{th} filter.

$$g_i[n] = x[n] * h_i[n]$$
$$= \sum_{m=0}^{\infty} x[m]\, h_i[n-m]$$

Binary T-F Mask Computation

Speech Energy : $SE_{i,j} = \sum_{m=jR}^{jR+L-1} (gS_i[m])^2$

Noise Energy :$NE_{,i,j} = \sum_{m=jR}^{jR+L-1} (gN_i[m])^2$

where $SE_{i,j}$ - energy of the speech signal in i^{th} channel, j^{th} frame
$NE_{i,j}$ - energy of the noise signal in i^{th} channel, j^{th} frame
gS_i - filtered response of speech signal in i^{th} channel
gN_i - filtered response of noise signal in i^{th} channel
L - Frame length
R - Window shift (L/2)

The T-F Binary Mask is defined as

$$M(i,j) = \begin{cases} 1 & \text{if } SE_{i,j} > NE_{i,j} \\ 0 & \text{otherwise} \end{cases}$$

T-F Mask based Noise Suppression

$s_{i,j}[m] = \sum_{m=jR}^{jR+L-1} g_i[m]\, p_{i,j}[jR-m]$ where $p_{i,j} = \begin{cases} w[n] & \text{if } M(i,j) = 1 \\ 0 & \text{if } M(i,j) = 0 \end{cases}$

Here w[n] is the sliding synthesis window (cosine window)

$$w[n] = \begin{cases} 1 + \cos(2\pi(n-1)/L - \pi)/2 & 0 \le n \le L-1 \\ 0 & \text{otherwise} \end{cases}$$

Speech Synthesis

$k_i[n] = f_i[n] * h_i[n]$
$= \sum_{m=0}^{\infty} f_i[m]\, h_i[n-m]$; where $f_i[n] = s_i[-n]$

$t_i[n] = k_i[-n]$

and finally,

$y[n] = \sum_{i=1}^{C} t_i[n]$