**WEINTRAUB SPEECH SEPARATION SYSTEM**

s1[n]

si[n]

sC[n]

y[n]

x[n]

g1[n]

g2[n]

:

:

gi[n]

:

:

gC[n]

g1[n]

g2[n]

:

:

gi[n]

:

:

gC[n]

t1[n]

t2[n]

:

:

ti[n]

:

:

tC[n]

T-F Mask based Noise Suppression

h1[n]

h2[n]

:

:

hi[n]

:

:

hC[n]

Binary T-F Mask Computation

h1[n]

h2[n]

:

:

hi[n]

:

:

hC[n]

Synthesis Filter Bank

Analysis Filter Bank

**Figure 1: Block Diagram of Weintraub Speech Separation System**

x[n] - Input Noisy Speech Signal

y[n] - Enhanced Speech Signal

C - Number of Channels

hi[n] - Impulse response of the Gammatone Filter; 1 ≤ i ≤ C

**Speech Analysis**

g**i**[n] = x[n] \* hi[n]

= n-m]

where hi[n] = AnN-1 exp(-2πbin)cos(2fin + )u[n];

A - amplitude,

- phase,

N - filter order,

fi - center frequency of the ith filter,

bi - bandwidth of the ith filter.

**Binary T-F Mask Computation**

Speech Energy : SEi,j = i[m])2

Noise Energy :NE,i,j i[m])2

where SEi,j - energy of the speech signal in ith channel, jth frame

NEi,j - energy of the noise signal in ith channel, jth frame

gSi - filtered response of speech signal in ith channel

gNi - filtered response of noise signal in ith channel

L - Frame length

R - Window shift (L/2)

The T-F Binary Mask is defined as

M(i,j) = 1 if SEi,j > NEi,j

0 otherwise

**Speech Synthesis**

ki[n] = fi[n] \* hi[n] ; where fi[n] = gi[-n]

= n-m]

si, i[m] pi,j[jR-m] where ti[n] = ki[-n]

and pi,j = w[n] if M(i,j) = 1

0 otherwise

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

w[n] = 1 + cos((2π(n-1)/L – π)/2 0 ≤ n ≤ L-1

0 otherwise

and finally,

n]

**PROPOSED SPEECH SEPARATION SYSTEM**

Analysis Filter Bank

Synthesis Filter Bank

y[n]

x[n]

g1[n]

g2[n]

:

:

gi[n]

:

:

gC[n]

g1[n]

g2[n]

:

:

gi[n]

:

:

gC[n]

s1[n]

s2[n]

:

:

si[n]

:

:

sC[n]

h1[n]

h2[n]

:

:

hi[n]

:

:

hC[n]

T-F Mask based Noise Suppression

Binary T-F Mask Computation

h1[n]

h2[n]

:

:

hi[n]

:

:

hC[n]

t1[n]

ti[n]

tC[n]

**Figure 2: Block Diagram of Proposed Speech Separation System**

x[n] - Input Noisy Speech Signal

y[n] - Enhanced Speech Signal

C - Number of Channels

hi[n] - Impulse response of the Gammatone Filter; 1 ≤ i ≤ C

**Speech Analysis**

g**i**[n] = x[n] \* hi[n]

where hi[n] = AnN-1 exp(-2πbin)cos(2fin + )u[n];

A - amplitude,

- phase,

N - filter order,

fi - center frequency of the ith filter,

bi - bandwidth of the ith filter.

g**i**[n] = x[n] \* hi[n]

= n-m]

**Binary T-F Mask Computation**

Speech Energy : SEi,j = i[m])2

Noise Energy :NE,i,j i[m])2

where SEi,j - energy of the speech signal in ith channel, jth frame

NEi,j - energy of the noise signal in ith channel, jth frame

gSi - filtered response of speech signal in ith channel

gNi - filtered response of noise signal in ith channel

L - Frame length

R - Window shift (L/2)

The T-F Binary Mask is defined as

M(i,j) = 1 if SEi,j > NEi,j

0 otherwise

**T-F Mask based Noise Suppression**

si, i[m] pi,j[jR-m] where pi,j = w[n] if M(i,j) = 1

0 if M(i,j) = 0

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

w[n] = 1 + cos(2π(n-1)/L – π)/2 0 ≤ n ≤ L-1

0 otherwise

**Speech Synthesis**

ki[n] = fi[n] \* hi[n]

= n-m]; where fi[n] = si[-n]

ti[n] = ki[-n]

and finally,

n]