بسم الله الرحمن الرحيم

Speech Enhancement

Dr. Md. Imran Hossain
Assistant Professor
Dept. of ICE, PUST, Pabna

Introduction

- **What is Speech Enhancement (SE)?**

  SE refers to the process of improving speech quality that has been degraded by background noise at the listener side through the use of various audio signal processing techniques and algorithms.

- **What is Noise?**

  Refers to signal that are unpredictable in nature and carry no useful information, it can be stationary, quasi stationary, non-stationary, narrowband, and broadband.

- **Noise Types**

  Additive noise

  Reverberation

  Convolutive channel effects

  Electrical interference

  Codec distortion

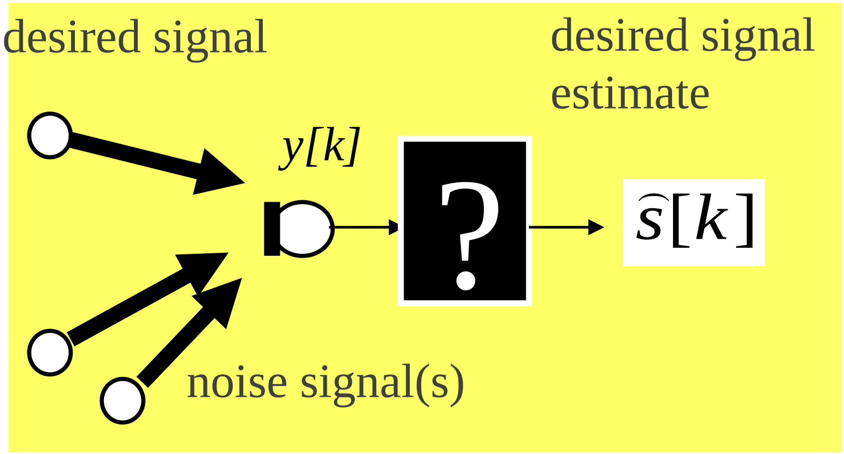
Introduction…

- **Additive Noise Model: (We considered)**

$$y[k] = s[k] + n[k]$$

desired signal contribution

noise contribution



- **Applications of SE**

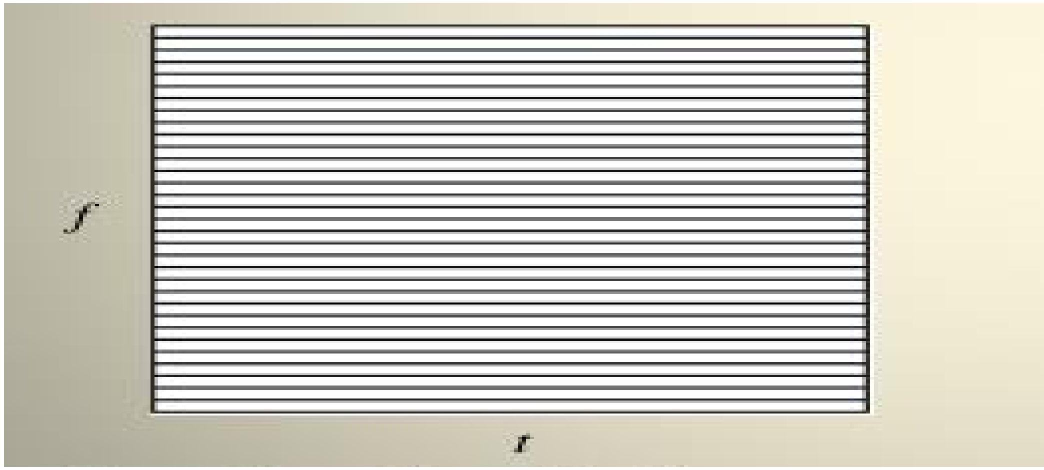
Mobile Phones, VoIP, Teleconferencing Systems, Hearing Aids, Digital Audio Restoration, Speech Recognition, Speech-Based Technology and Air to Ground Communication Between ATC and Pilot.

Introduction…

- Fourier Transform (FT)

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt$$

- Time-frequency tile for FT



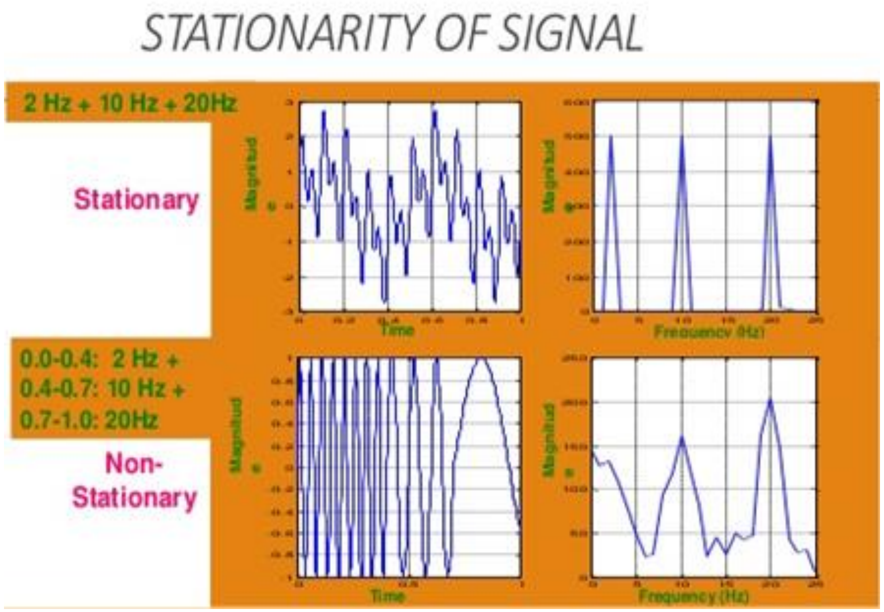
- Different in time but same frequency representation
- FT only gives what frequency components exist in a signal.
- FT cannot tell at what time the frequency components occur.

Introduction…

However, most of transportation signals are non-stationary, so we need to know whether and also when an incident was happened.

Stationary signals consist of spectral components that do not change in time

- All spectral components exist at all time
- No need to know any time information
- FT works well for stationary signals



Non-stationary signals consists of time varying spectral components

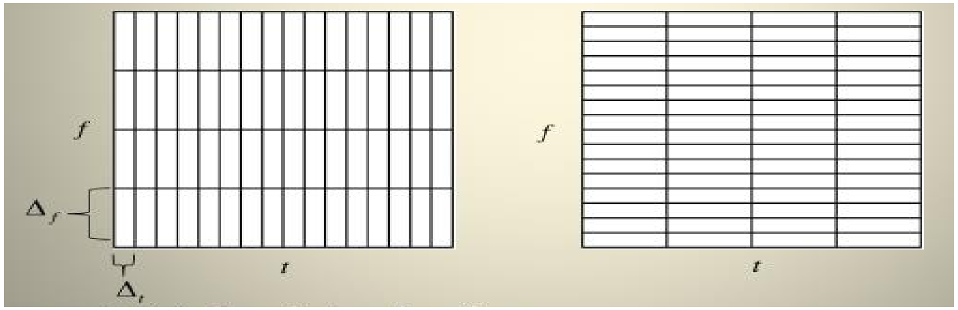
- FT only provides what spectral components exist, not where in time they are located. Need some other way to determine time localization of spectral components

Introduction…

- Short-time Fourier transform (STFT)

  $$F_{STFT}(\omega,\tau) = \int_{-\infty}^{\infty} f(t)w(t-\tau)e^{-jwt}dt$$

- Time-frequency tile for STFT



- STFT provides the time information by computing a different FTs for consecutive time intervals, and then putting them together.
- Selection of width of STFT window

  wide analysis window→ Poor time resolution, Good frequency resolution

  narrow analysis window → Good time resolution, Poor frequency resolution

- We cannot precisely know at what time instance a frequency component is located. We can only know what interval of frequencies are present in which time intervals.

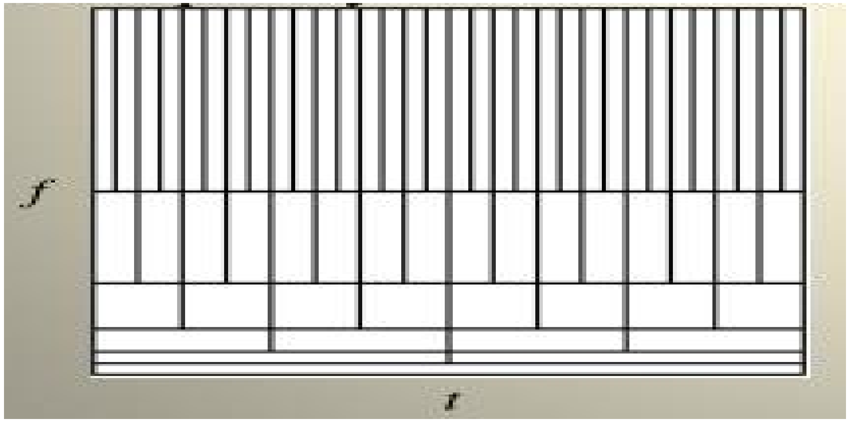
Introduction…

- Wavelet transforms

$$F_{WT}(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{\infty} f(t)\Psi^*\left(\frac{t-\tau}{s}\right) dt$$

where τ is the translation parameter, s is the scale parameter and Ψ is the mother wavelet.

- Time-frequency tile for wavelet transform (WT)



- It overcome the preset resolution problem of the STFT by using a variable length window.

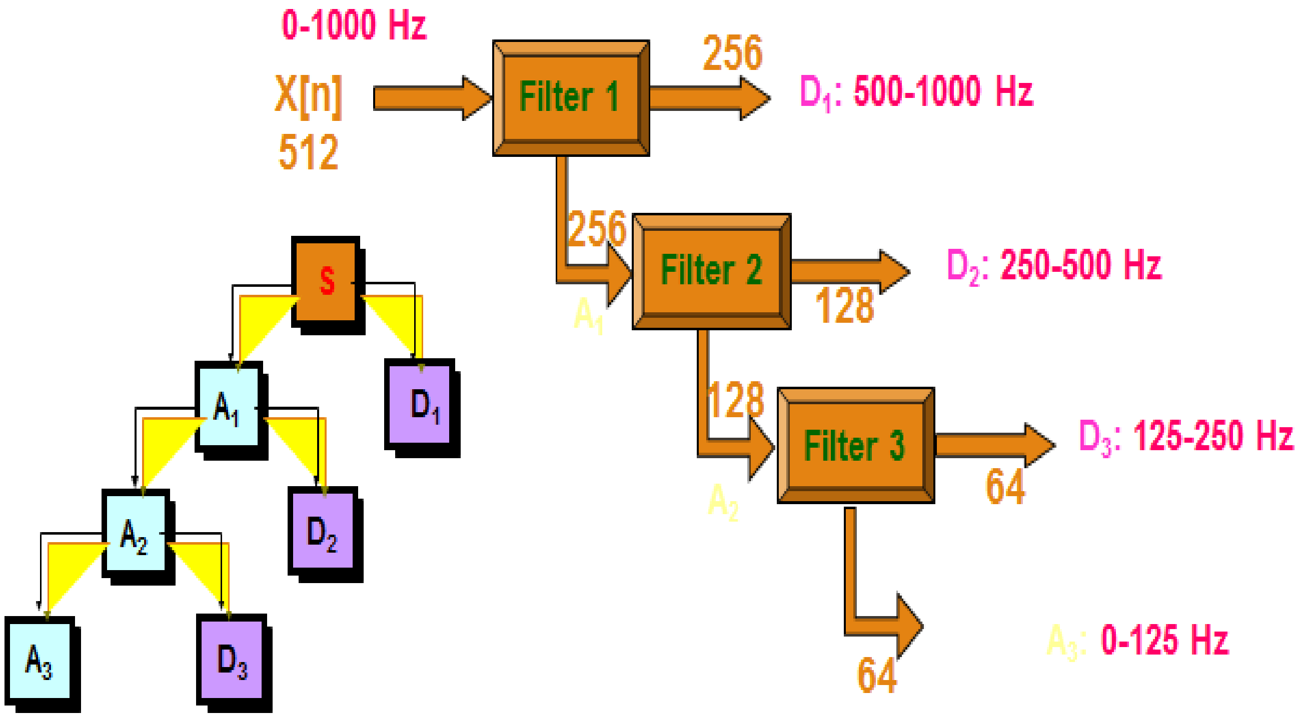
Introduction…

- Analysis windows of different lengths are used for different frequencies

  Analysis of high frequencies → Use narrower windows for better time resolution

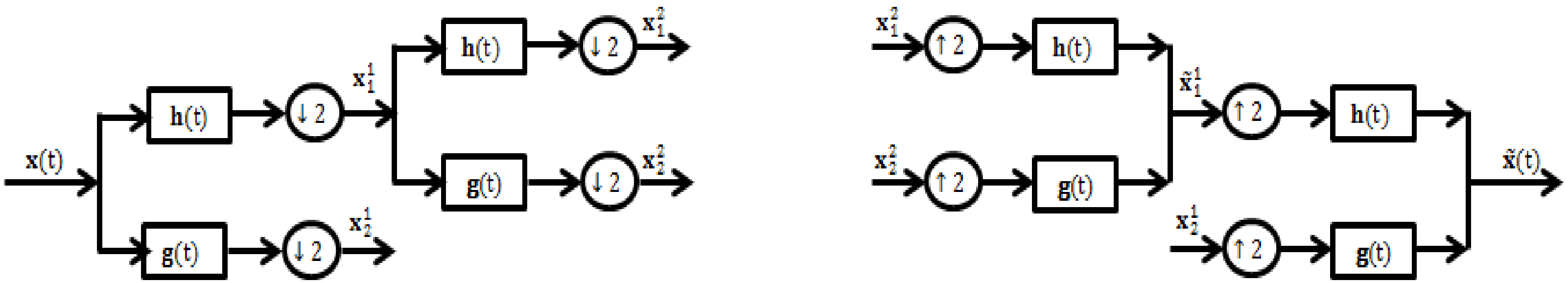
  Analysis of low frequencies → Use wider windows for better frequency resolution
- Provide a way for analyzing waveforms in both frequency and time.
- Representation of functions that have discontinuities and sharp peaks.
- Accurately deconstructing and reconstructing finite, non-periodic and/or non-stationary signals.
- Allow signals to be stored more efficiently than by FT.

Introduction…

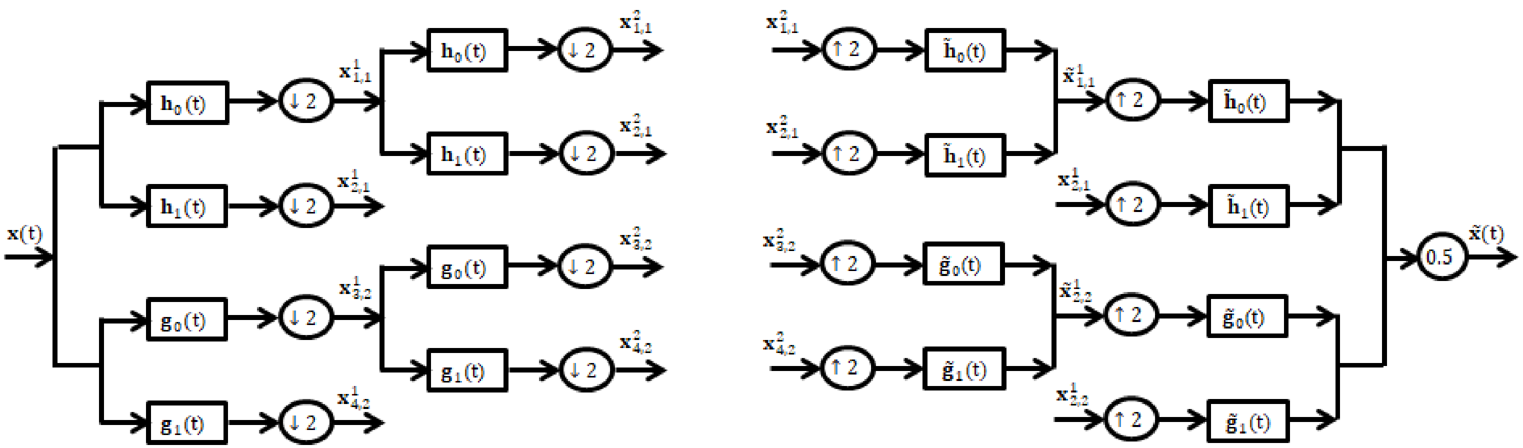


Introduction…

Filter bank (FB) implementation of discrete wavelet transform (DWT):



FB implementation of dual-tree complex transform (DTCWT):



Proposed DTCWT-NMF SE Method…

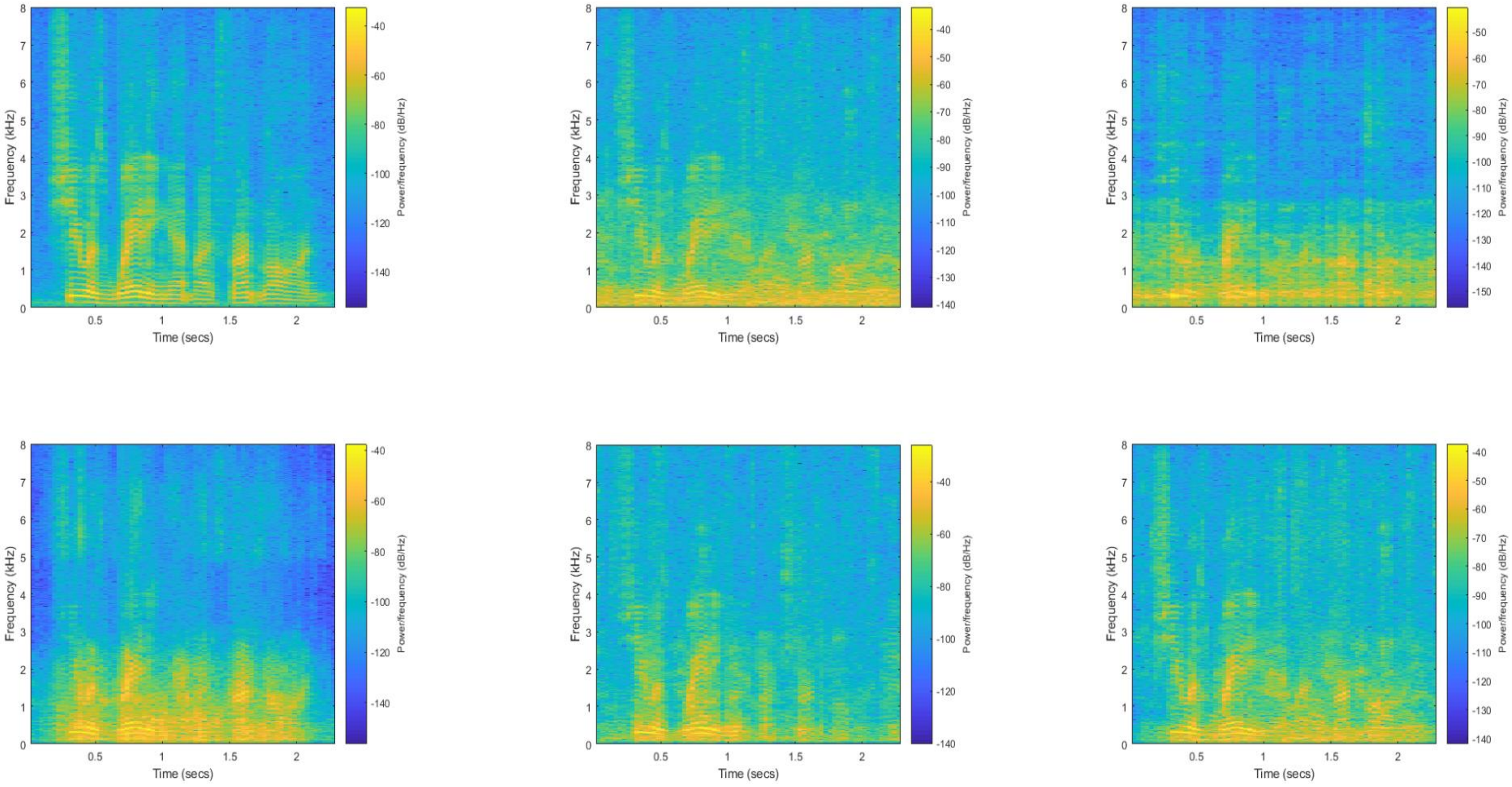


Figure 9. Spectrogram of Clean speech, Noisy, STFT-NMF, DWPT-NMF, DNN-IRM, and DTCWT-NMF

Proposed DTCWT-STFT-SNMF SE Method…

Table 16. Comparison of PESQ values of eight methods at five SNR conditions

Method	-10	-5	0	5	10
STFT-SNMF	1.529	1.776	2.148	2.483	2.782
STFT-SNMFSE	1.541	1.975	2.22	2.528	2.791
MLD-STFT-SNMF	1.571	1.953	2.277	2.532	2.800
STFT-GDL	1.562	1.938	2.260	2.514	2.725
STFT-CJSR	1.518	1.906	2.253	2.525	2.754
DTCWT-SNMF	1.526	1.918	2.268	2.519	2.748
DWPT-STFT-SNMF	1.588	1.987	2.301	2.544	2.742
DTCWT-STFT-SNMF	**1.598**	**2.039**	**2.414**	**2.692**	**2.900**

Table 17. Comparison of STOI values of eight methods at five SNR conditions

Method	-10	-5	0	5	10
STFT-SNMF	0.538	0.649	0.759	0.845	0.906
STFT-SNMFSE	0.533	0.662	0.778	0.812	0.889
MLD-STFT-SNMF	0.561	0.680	0.785	0.844	0.904
STFT-GDL	0.529	0.660	0.770	0.848	0.899
STFT-CJSR	0.547	0.669	0.774	0.851	0.906
DTCWT-SNMF	0.555	0.677	0.780	0.849	0.903
DWPT-STFT-SNMF	0.546	0.657	0.740	0.800	0.838
DTCWT-STFT-SNMF	**0.587**	**0.706**	**0.803**	**0.873**	**0.920**

Proposed DTCWT-STFT-SNMF SE Method…

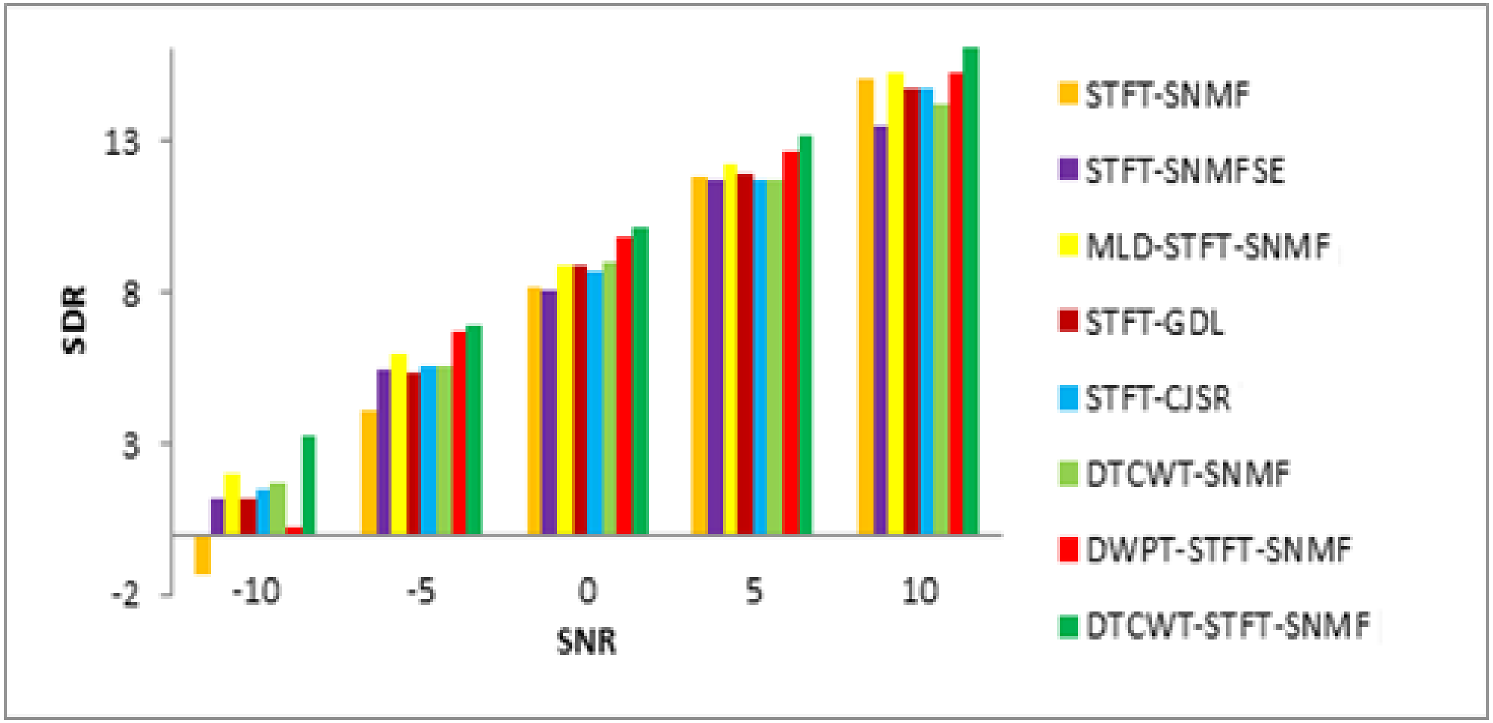
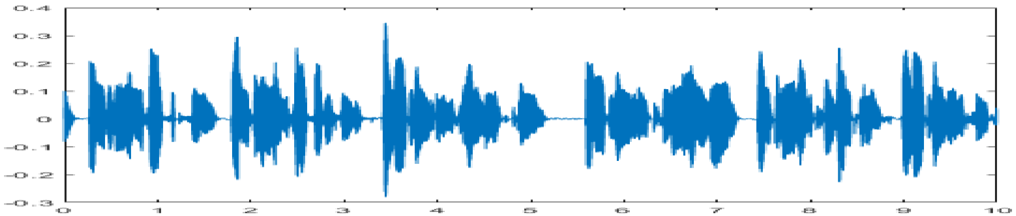
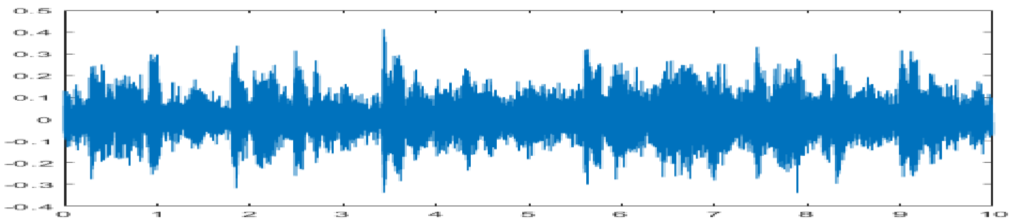


Figure 15. Comparison of SDR values of seven methods at five SNR conditions

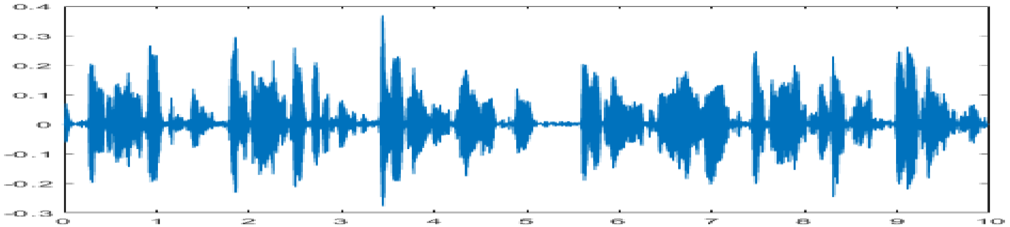
Proposed DTCWT-STFT-SNMF SE Method…



(a) Clean speech signal

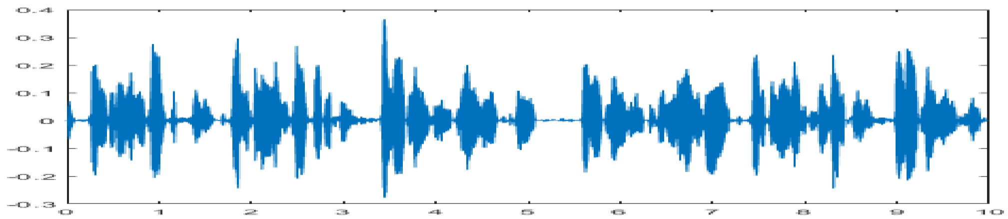


(b) Noisy (0dB, leopard) speech signal

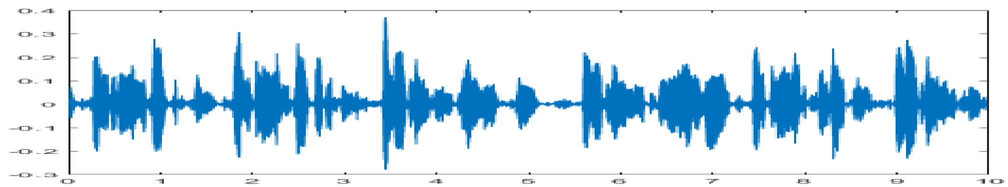


(c) Estimated speech signal using STFT-SNMF method

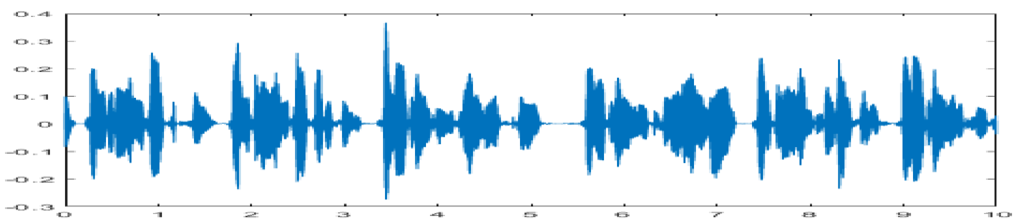
Third Proposed DTCWT-STFT-SNMF SE Method…



(d) Estimated speech signal using STFT-GDL method



(e) Estimated speech signal using STFT-CJSR method



(f) Estimated speech signal using DTCWT-STFT-SNMF method

Figure 16. The time-domain waveform of speech, where x-axis corresponds to a time in second and the y-axis corresponds to amplitude in dB

DTCWT-STFT-SNMF SE Method…

Clean Signal	Noisy Signal Mixed with Noise at 0dB	Enhanced Signal
	m109	
	ssn_ieee	
	Volvo	
	White	