COMPARISON OF SPECTROGRAMS USING STFT, SFF and S-Transform

A Project Report

Submitted by

KESAVARAJ - 2022701008

ANUPRABHA - 2022802001

In the partial fulfillment for the completion of the project

of

Speech Signal Processing

	Table of Contents	
Chapter No	**Title**	**Page No**

1.Introduction:

Among various spectral analysis tools arisen in the last years, some were more prominent, such as Fourier transform. Nevertheless, all of them present implementation restrictions for an ideal extraction for low as well as high frequencies, of variant signals in time, as, for example, signals containing time-varying harmonics. This was the reason for the emergence of time-frequency analysis. In this we will discuss some of the famous approaches such as STFT, SFF and S-transform for time-frequency analysis.

2. Short Time Fourier Transform (STFT):

The short-time Fourier transform (STFT), is used to determine the sinusoidal frequency and phase content of local sections of a signal as it changes over time. In practice, the procedure for computing STFTs is to divide a longer time signal into shorter segments of equal length and then compute the Fourier transform separately on each shorter segment.

In the continuous-time case, the function to be transformed is multiplied by window function which is nonzero for only a short period of time. The Fourier transform (a one-dimensional function) of the resulting signal is taken, then the window is slid along the time axis until the end resulting in a two-dimensional representation of the signal. Mathematically, this is written as:

$$\mathbf{STFT}\{x(t)\}(\tau,\omega) \equiv X(\tau,\omega) = \int_{-\infty}^{\infty} x(t)w(t-\tau)e^{-i\omega t}\,dt$$

where w(t) is the window function, commonly a Hann window or Gaussian window centred around zero, and x(t) is the signal to be transformed (the difference between the window function and the frequency)

In the discrete time case, the data to be transformed could be broken up into chunks or frames (which usually overlap each other, to reduce artifacts at the boundary). Each chunk is Fourier transformed, and the complex result is added to a matrix, which records magnitude and phase for each point in time and frequency. This can be expressed as

$$\mathbf{STFT}\{x[n]\}(m,\omega) \equiv X(m,\omega) = \sum_{n=-\infty}^{\infty} x[n]w[n-m]e^{-i\omega n}$$

likewise, with signal $x[n]$ and window $w[n]$. In this case, m is discrete and ω is continuous.

3. Single Frequency Filtering (SFF):

Single frequency filtering (SFF) method gives amplitude envelopes (ek[n]) at each sample n at the selected frequency fk, with a pole close to the unit circle, and extracts information at the highest carrier frequency (i.e., half the sampling frequency). Since the same filter at a fixed frequency is used to derive the amplitude envelopes at different frequencies, it avoids the different gain effects, if separate filters were chosen for each frequency to derive amplitude information. The shape of the filter and its gain vary if different filters are designed to extract information at different frequencies as in the case of filter bank approaches.

The discrete-time speech signal x[n] at the sampling frequency fs is multiplied by a complex sinusoid of a given normalized frequency ωk to give xk[n]. The time domain operation is given by

$$x_k[n] = x[n]e^{j\bar{\omega}_k n},$$

$$\bar{\omega}_k = \frac{2\pi \bar{f}_k}{f_s}.$$

Since x[n] is multiplied with e jωkn to give xk[n], the resulting spectrum of xk[n] is a shifted spectrum of x[n]. That is,

$$X_k(\omega) = X(\omega - \bar{\omega}_k),$$

where Xk(ω) and X(ω) are spectra of xk[n] and x[n], respectively. The signal xk[n] is passed through a single-pole filter, whose transfer function is given by

$$H(z) = \frac{1}{1 + rz^{-1}}.$$

The single-pole filter has a pole on the real axis at a distance of −r from the origin. The root is located at z = −r in the z-plane, which corresponds to half the sampling frequency, i.e., fs/2. The value of r is chosen as 0.99 for most cases in the study. The output yk[n] of the filter is given by

$$y_k[n] = -ry_k[n-1] + x_k[n].$$

The envelope of the signal ek[n] is given by

$$e_k[n] = \sqrt{y_{kr}^2[n] + y_{ki}^2[n]},$$

where ykr[n] and yki[n] are the real and imaginary components of yk[n]. Since filtering of xk[n] is done at fs/2, the envelope ek[n] corresponds to the envelope of the signal xk[n] at the desired frequency given by

$$f_k = \frac{f_s}{2} - \bar{f}_k.$$

4. S-Transform:

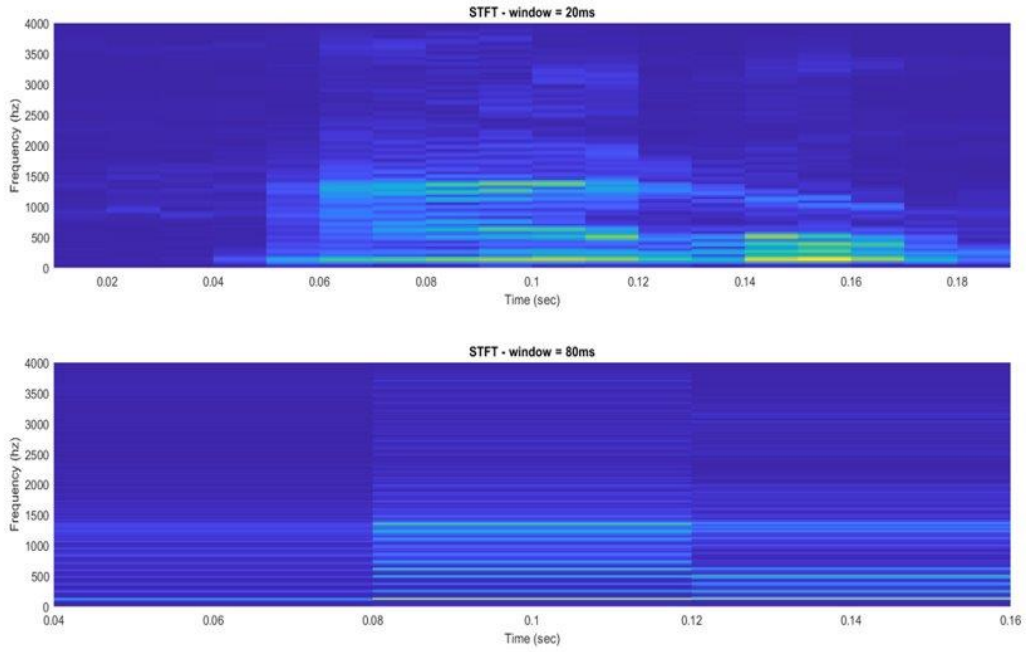
ST can be considered a hybrid method, between STFT and WT. This tool can be classified as a frequency-dependent STFT or a WT with corrected phase. ST can be defined as

$$S(\tau, f) = \int_{-\infty}^{+\infty} h(t) \underbrace{\frac{|f|}{\sqrt{2\pi}} e^{-(\tau - t)^2 f^2/2}}_{\text{window}} e^{-j2\pi f t}\, dt\,.$$

The frequency-dependent window enables a frequency resolution with narrower ranges for higher frequencies and wider range ones for lower frequencies. Opposite to WT, phase information provided by ST is connected to origin in time, using FT as a basis, which is not possible with continuous wavelet transform, where phase information is locally referenced

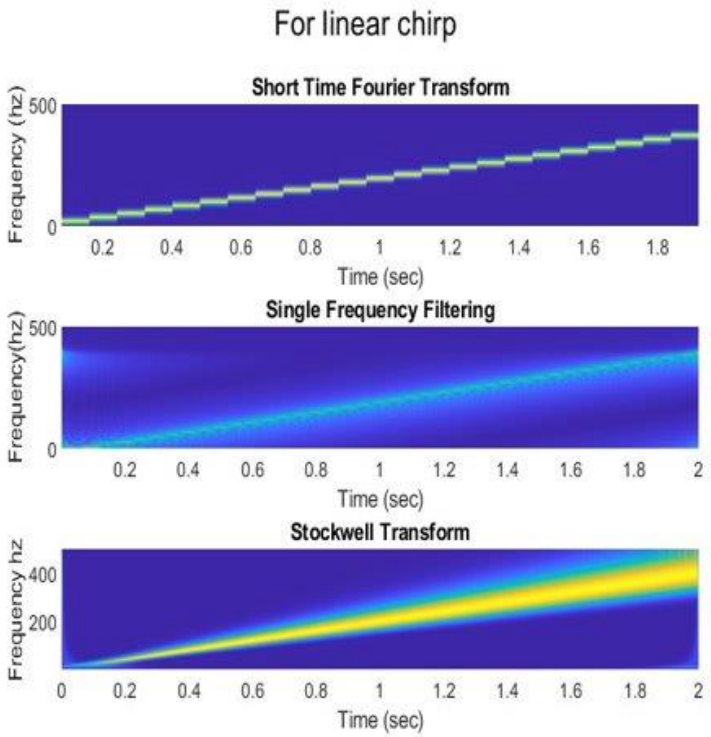
5. Results & Discussion:

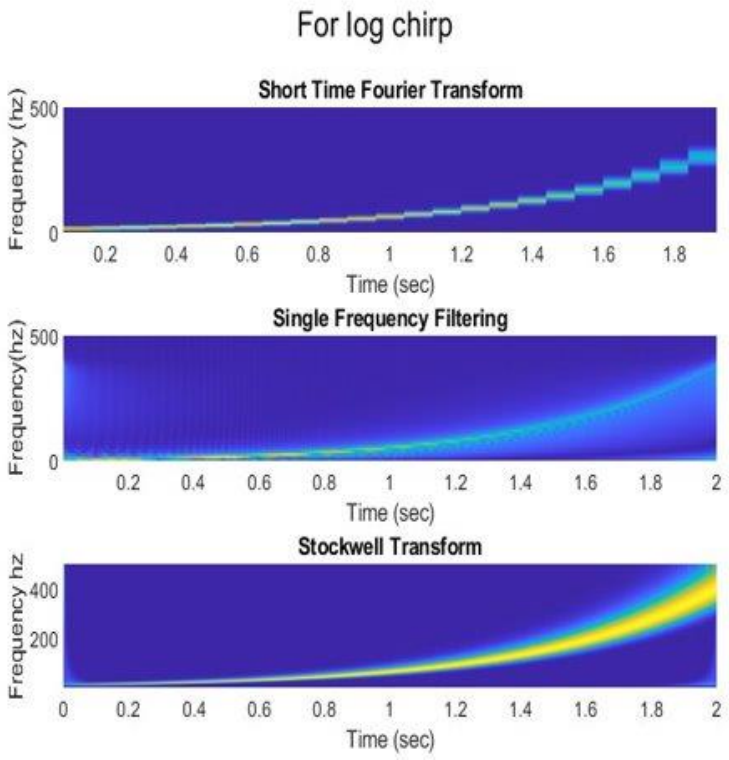
5.1 STFT-Changing window width:



5.2 Chirp Signal:

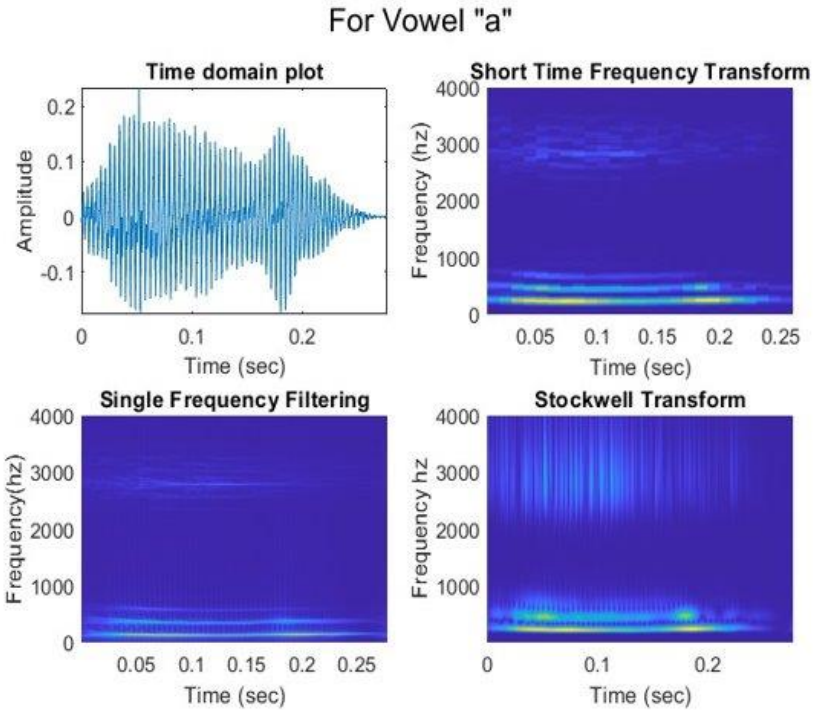
The chirp signal covers all the frequencies in the given range.

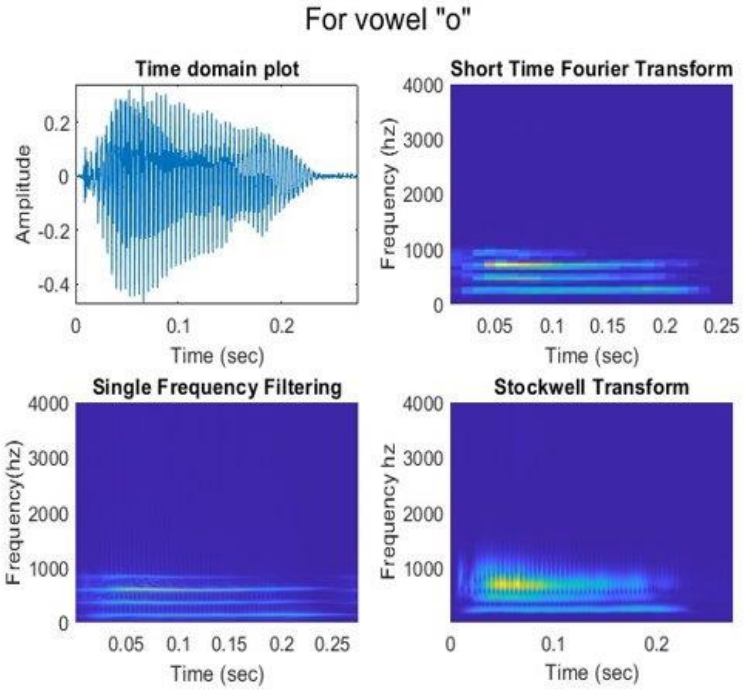




5.3 Vowels:

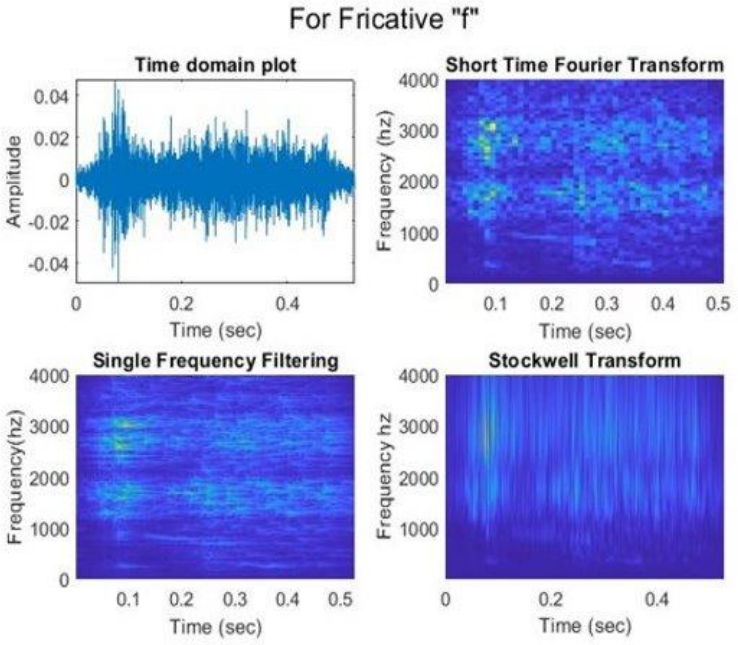
Both SFF and S-Transform capture the vowel regions (low frequency components) effectively.

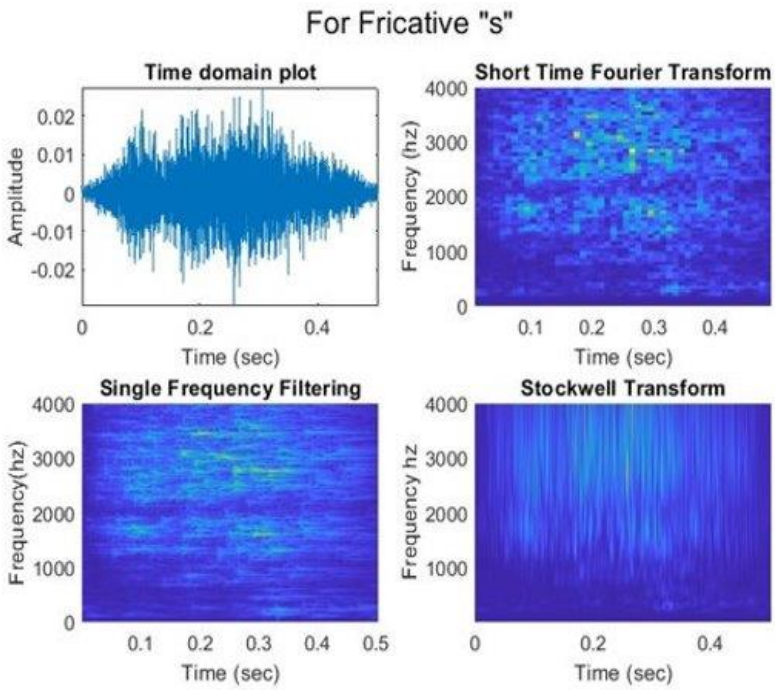




5.4 Fricatives:

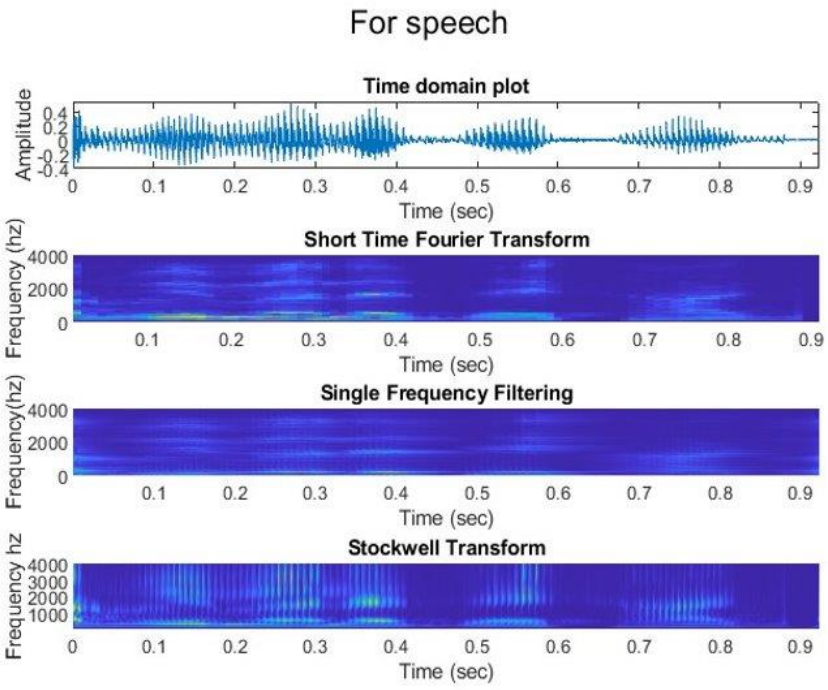
S-Transform gives better resolution at high frequencies and works well for fricatives.





5.5 Speech Signal:

SFF works well for low frequencies. S-Transform works well for representing for both low and high frequency components.



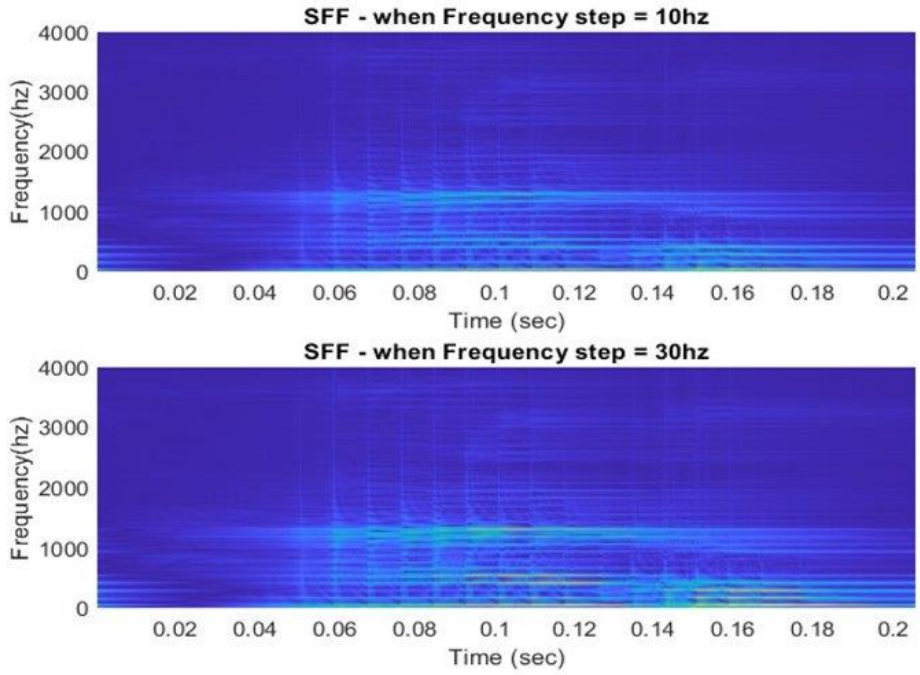
5.6 SFF- Resolution Controlling Parameters:

1.Frequency shift:

As frequency shift increases, the plot starts to smear out and the resolution decreases. Optimum values = 5, 10, 15, 20.

2. Pole location of the filter:

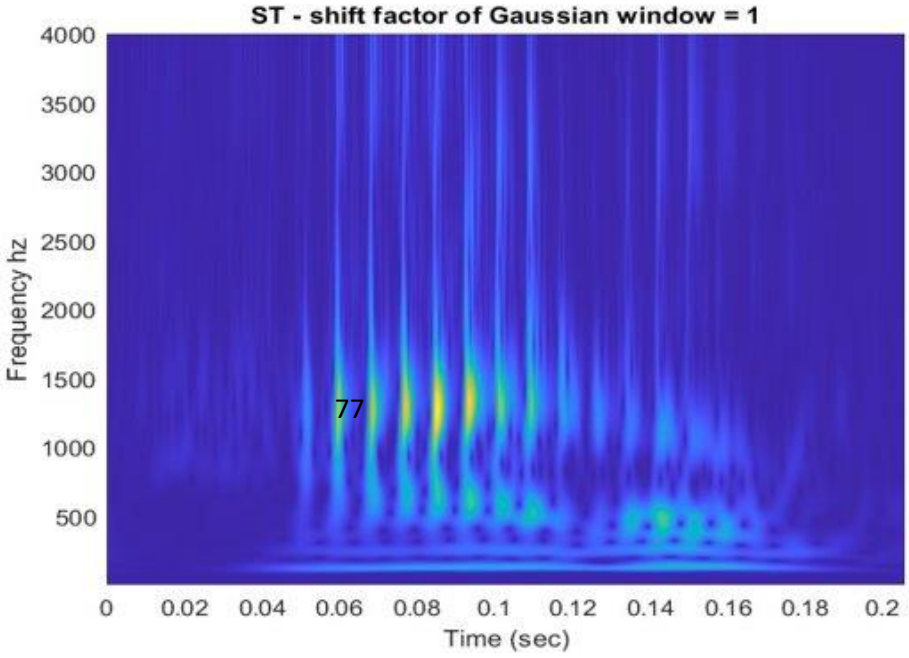
As the pole moves away from the unit circle, the resolution decreases.Ideal value of r = 0.995.

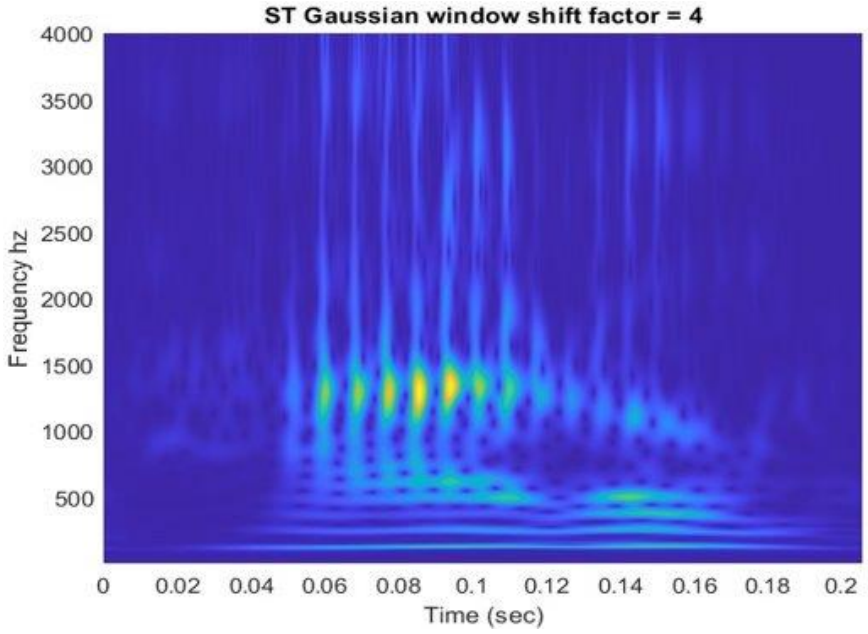


5.7 ST – Resolution controlling parameter:

Window shift:

As window shift factor increases, resolution increases





6. Conclusion:

In this project, we studied the fundamental concepts of STFT, SFF & S-Transform and implemented it with some time-varying signals like vowels, fricatives, etc. We have performed visual comparisons to find out which gives a better spectral resolution. From our analysis, S-Transform which extracts both low and high frequency components distinctively seems to be the best fit for time-varying signals like speech.