CHAPTER 5

SPECTRAL APPROACH FOR EMG FEATURE EXTRACTION

Temporal approach can extract important information from classifying EMG patterns in prosthetic hand control. The experimental results indicate that the temporal features extracted from the 300 ms sequences have significant capability to discriminate the pattern for the grasp types.

Spectrum analysis is also applied to EMG studies. Various feature extraction methods based on the spectral analysis are experimented. The using of information contained in frequency domain could lead to a better solution for encoding the EMG signal. Time-frequency analysis based on short-time Fourier transform is a form of local Fourier analysis that treats time and frequency simultaneously and systematically. The characters of EMG signals in frequency domain are explored and demonstrated in this chapter. The short time variability of spectrum, which is an essential fact for using time-frequency methods in EMG feature extraction, is also discussed in this chapter. The analysis can provide important clues to design feature extraction methods. Wavelets approach is another powerful technique for time-frequency analysis. We will first introduce Fourier theory and then discuss these spectral approaches.

5.1 Spectral Analysis of EMG Signals

5.1.1 Fourier Theory

Fourier transform is a method of expressing a function in terms of the sum of its projections onto a set of basis functions. These functions are sinusoids of different frequency whose sum restores the original function. The idea of Fourier transform is that any waveform can
be made up by a linear combination of the correct sine waves with appropriate amplitudes and phases. The Fourier transform of \( f(x) \) is defined as:

\[
F(\omega) = \int_{-\infty}^{+\infty} f(x) e^{-j2\pi\omega x} dx .
\] (5.1)

The inverse of Fourier transform is defined as:

\[
f(x) = \int_{-\infty}^{+\infty} F(\omega) \cdot e^{j2\pi\omega x} d\omega .
\] (5.2)

A digital computer works only with discrete data. Let \( f[t] \) be sequence of \( N \) sampled values of \( f(x) \). The Discrete-time Fourier transform (DTFT) is a continuous function of the frequency variable. It’s defined as:

\[
F(\Omega) = \sum_{k=0}^{N-1} f[t] \cdot e^{-j\Omega k} .
\] (5.3)

The Discrete Fourier transform (DFT) of \( f[t] \) is obtained by sampling DTFT. It is given by:

\[
F[k] = \sum_{t=0}^{N-1} f[t] \cdot e^{-j2\pi \frac{tk}{N}} .
\] (5.4)

Its inversion is:

\[
f[t] = \frac{1}{N} \sum_{k=0}^{N-1} F[K] \cdot e^{j2\pi \frac{tk}{N}} .
\] (5.5)

Fourier transform is widely used in signal processing. It is an equation to calculate the frequency, amplitude and phase of each sine wave needed to make up any given signal. It’s a linear transform between time and frequency domains and can be used to analyze the spectral component of a signal. An approach name Fast Fourier Transform is employed in practical for the calculation of DFT. In MATLAB DFT is implemented as FFT. The feature extraction
methods are implemented by MATLAB in this thesis and the Fourier transform is actually implemented by the MATLAB function FFT.

5.1.2 Application of FFT on EMG Signals

The following figure shows how DFT works for a raw EMG signal. We can observe that the raw signal in the time domain is converted into a waveform in the frequency domain.

![Figure 5-1: Frequency spectrum of the EMG by Discrete Fourier transform (DFT)](image)

Fourier Transform and its inverse provide a relation between the time domain and the frequency domain. Fourier Transform gives a method to analyze a signal over the entire time. It is an optimal solution when we assume there is no frequency change with time. However, it does not give any information on a time localization of the frequency component of the signal. Time-frequency analysis provides methods to represent a signal by 2-dimensional functions, which can describe the frequency components in time domain.
5.2 Short-Time Fourier Transform

The approach that can give information on the time resolution of the spectrum is Short Time Fourier Transform (STFT). The basic idea of STFT is to divide a signal into short pieces and apply Fourier Transform to each piece.

The STFT of a time sequence $x(n)$ with respect to windowing function $W(n)$ is:

$$STFT(t, k) = \sum_{n=0}^{N-1} x(n) W(n-t) e^{-j2\pi nk/N}$$

(5.6)

The time variant localized frequency contents of EMG signals can be obtained by STFT. The graph in figure 5-2 shows how an EMG signal’s frequency content changes with time. A raw EMG signal is sampled with 1000 Hz for about 1500 millisecond. In MATLAB the function `specgram(a, nfft, fs, window, numoverlap)` computes the windowed discrete-time Fourier transform of a signal by the following steps:

1. It splits the signal into overlapping sections and applies the window specified by the window parameter to each section.

2. It computes the discrete-time Fourier transform of each section with a length `nfft` to produce an estimate of the short-term frequency content of the signal; these transforms make up the columns of resulting matrix. The quantity $(\text{length(window)} - \text{numoverlap})$ specifies by how many samples `specgram` shifts the window.

3. For real input, `specgram` truncates the spectrogram to the first $nfft/2 + 1$ points for even `nfft` and $(nfft + 1)/2$ for odd `nfft`.

We use a window of 1/10 length of the raw signal and slide the window four milliseconds along time axis until the end of the time series. A detail of localization information
about frequency is displayed in the 3D graph in figure 5-2 by MATLAB mesh function. We can observe that the frequency spectrum is changing along the time axis. The changing is of short time variability. The method can produce spectrums for different time locations. The localization information of those frequencies might be essential for the classification of a pattern.

5.2.1 Exploration of Hannaford

Blake Hannaford applied STFT to EMG recorded during rapid body movement [35]. He used STFT technique to analyze time-dependent spectrum in EMG. Moments of the spectrum were used to show the variability.

Research on Spectrum analysis of surface EMG shows that spectrum shift exists. The features extracted from the spectrum are central frequency, variance of frequency and moments of frequency.

If we define the $n$th moment of the frequency distribution at time $t$ as:
\[ M_n(t) = \sum_k \omega_k^T |STFT(t,k) |. \tag{5.7} \]

the central frequency and variance can be defined as:

\[ \text{Frequency} \_ \text{EXP}(t) = \frac{M_1}{M_0}. \tag{5.8} \]

\[ \text{Frequency} \_ \text{VAR}(t) = \frac{M_2}{M_0} - \left( \frac{M_1}{M_0} \right)^2. \tag{5.9} \]

Hannaford’s analysis shows these interesting frequency features applied to EMG [35].

EMG spectrogram mean (central) frequency and variance were displayed together with energy estimation. Let \( W(t) \) to be a window and \( x(t) \) to be an input time series, where \( t \) is time. The energy is

\[ EN(t) = \sum_n (W(n-t) \cdot x(n))^2. \tag{5.10} \]

He performed the STFT with windows size of 16, 32, 50, 64, 75 and 128 to EMG data from wrist flexors and extensors during steady contractions at different force level and joint positions.

The STFT with 75 ms windows size is also applied to EMG data from fast head movement. Specifically, EMG spectral content for fast movement changes on both center frequency and variance during the course of three phases of EMG activity— the initial burst in the agonist, a braking burst in the antagonist, and a final climbing burst in the agonist. These three phases were called PA, PB, and PC, respectively. The Mean and variance of frequency distributions changed consistently in the fast head and wrist movement studies. The mean frequency usually declined slightly during PA, and rose to a higher value between PA and PC, again gradually declining thereafter. Hannaford suggested that there is analogy between EMG processing and speech recognition. The fact could be useful in sophisticated movement analysis.
application such as identification of intended limb function from EMG activity in residual muscles of the amputee [35].

5.2.2 Variability of Short Time Spectrum for EMG Patterns

We use a similar approach to Hannaford’s method to analyze the frequency shift property of an EMG signal used in multifunction prosthetic control. The analysis will provide us hints on how to extract features from spectrum, such that they represent the EMG patterns with the best possible way.

The Raw EMG signals of four channels are displayed in the following figure. The signals have length of 800 ms and are taken from four channels after the beginning of a grasp motion. The sampling frequency is 1000 Hz. The data before activation are truncated.

![Figure 5-3: 800 ms EMG signals from four channels](image)

The moments are computed according to equation [5.7]. A moving window of on tenth of the raw signal length is used. The waveforms of the zero, first and second moment for each
channel spectrum are displayed in the following figure. We can observe from the figure that the shapes of the zero, first, and second moment are almost identical to each other.

![Spectrum shift of EMG: moment of spectrum](image)

Figure 5-4: Zero, 1\textsuperscript{st}, 2\textsuperscript{nd} moments of the spectrum change in time domain

The waveform for channel 1, 2, 3 and 4 are displayed as solid, dotted, dash dotted, and dashed lines, respectively. The 0\textsuperscript{th}, 1\textsuperscript{st}, and 2\textsuperscript{nd} moments are very similar. It suggests that there exist redundant information in these waveforms.

The function of central frequency in term of time is defined in (5.8). The standard variance is defined as:

\[
\text{Frequene}_\text{std}(t) = \sqrt{\frac{M_2}{M_0} - (\frac{M_1}{M_0})^2}.
\]  

(5.11)
The energy is defined by (5.10). The central frequency, frequency standard variance and energy are displayed in the following figures respectively.

Figure 5-5: Central frequency, standard variance and energy estimation.

The waveform for channel 1, 2, 3 and 4 are displayed as solid, dotted, dash dotted, and dashed lines, respectively. One interesting property of the waveforms of central (mean) frequency and its standard variance is that they have a similar wave shape. The similarity becomes more significant at the time axis values for local apexes and nadirs. This indicates that a higher frequency leads to higher variance. However, the waveform for energy has different shape. It shows us that the spectrum approach can reveal extra information for classifying EMG signals comparing with the temporal approaches.
5.2.3 Ensemble Average of EMG Patterns

Statistic methods are used to discuss if there is a correlation between the measurements proposed in previous section and grasp types. There are totally 180 raw data files. They belong to six categories of grasping objects--- Large Ball, Large Cylinder, Small Ball, Small Cylinder, Small Disk, and Small Key. There are 30 raw signals for each of the above grasp types.

The ensemble average is average value for all the signals belonging to one category. There are a number of measurements discussed in previous section of this chapter. The ensemble averages for central frequency, frequency standard variance, energy, and the moments are given in the following figures (5-6 to 5-11). The signals have length of 800 ms and are taken from four channels after the beginning of a grasp motion. The sampling frequency is 1000 Hz. The size of windowing function is one tenth of the raw signal size, which is 100 ms in this case.

We can observe from these figures that the waveform of Energy demonstrates significant structure properties which can be used to discriminate patterns. The waveforms of other measurements are fairly stable. The fact suggests that there exists limited useful pattern information in these measurements.
Central Frequency:

Figure 5-6: Ensemble of central frequency
Frequency Standard Variance:

Figure 5-7: Ensemble of frequency standard variance
Figure 5-8: Ensemble of energy
The zero moment of spectrum:

Figure 5-9: Ensemble of 0 moment
The first moment of spectrum:

Figure 5-10: Ensemble of 1st moment
The second moment of the spectrum:

Figure 5-11: Ensemble of 2nd moment
5.2.4 Application of STFT on EMG Signals

We use STFT on the sampled raw EMG data. The Fourier Transform gives an optimal result if the data are periodic in time domain which is assumed to be infinite. Since we use a segment of time series to represent the overall time domain and assume this segment repeat periodically, a good spectrum can be achieved when the end of one segment connects smoothly to the beginning of the segment.

Frequency leakage happens if the end of one segment does not connect to the beginning of the next segment smoothly.

The flowing figure shows how a windowing function is applied to a segment of EMG data when grasping small balls. In part A of the figure, an undesired sudden jump in time series will be created if we repeat the segment. The spectrum will be broadened due to the existence of the sudden jump. This is an example of the frequency leakage.
In order to reduce the frequency leakage, a windowing function is applied to each section of time series. The Hanning window in this graph can suppress the beginning and the end of the signal segment and let them connect to each other smoothly. The spectrum analysis concentrates more to the center of the segment. The hanning window is shown in part B of the figure 5-12, while the windowed time sequence is shown in part C. The resulting spectrum has less bandwidth after multiplying the raw signal with the hanning window. The effect is demonstrated in the following figure.
We should notice that windowing functions always put more significance on the middle part of a signal segment and that, despite its good performance on narrowing the bandwidth of the spectrum, it introduces bias by which information contained in the beginning and end of the signal segment might be suppressed. Successive segments are overlapped in order to reduce the bias. The size of the window and the amount of the overlap is determined based on the analysis and time-frequency property of the input signal. The algorithm for STFT is described as:

1. Splits the signal into several overlapping sections.
2. Multiply each segment by a windowing function.
3. For each segment:
   
   Computes the DFT of the segment and get the spectrum a vector.

Figure 5-13: Reduce frequency leakage by a hanning window
The result is a matrix where each column is a vector from step 3. The column index and row index of the result matrix are axis indices of frequency and time respectively.

There is a tradeoff in STFT between time and frequency resolution. Larger windows produce higher frequency resolution but have less time information, while smaller windows can produce a spectrum of higher time resolution but of less detail of frequency. Perfect frequency resolution can be achieved by using a window of infinite length. However, the time information completely lost in this case. Small size window will be suitable for stationary signals. The generating of EMG signals is a random process and it is basically not stationary, as we described in chapter 3.

5.2.5 Feature Extraction Strategies Based on Short Time Fourier Transform

STFT provides a powerful method for time-frequency analysis. There are a number of methods to extract features based on the properties of the spectrums and how they shift with time. A good method can create features which can discriminate the signal category efficiently. An improper method will introduce into the classifier useless information which will introduce unnecessary computation. Irrelevant information becomes noises and it degrades the system’s performance eventually.

Kristin A. Farry [29] and other researchers applied STFT on the EMG feature extraction. Summary of Kristin’s STFT is the following:

1. Divide the initial phase of the signal into six non-overlapping 40 ms segments. Each segment is denoted as \( x[n], n = 1, 2, ..., N \).

2. Apply the Hamming Window \( w[n], n = 1, 2, ..., N \), to each signal segment \( x[n], n = 1, 2, ..., N \).
3. For each windowed segment compute the FFT

\[
P(f) = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \cdot w[n] \cdot e^{-j2\pi fn},
\]

where \( f = \frac{2\pi k}{N}, k = 0, 1, ..., N - 1. \)

4. For each spectrum, compute four spectrum magnitude averages covering four integrals in the range 75~250 Hz.

Totally twenty-four features were extracted from the 240 ms EMG data. The lower frequency 0~75 Hz is much more stationary than the upper part. The 0~75 Hz frequency in the spectrum was not used. The features are inputs to 24-6-2 multi-layer perceptrons for training and classification [29].

Thomson’s multiple windows method (MWM) was also employed in Short Time Thomson Transform (STTT) in a similar way to the STFT. A comparison STTT and STFT was proposed in Kristin’s work [29]. A discussion of STTT is in the next section of this chapter.

The waveform of spectrum is described approximately in the features by sampling the spectrum in Kristin’s approach. Four spectrum magnitude averages covering the 75~250 Hz regions.

An alternative way to describe the waveform of the spectrum is by using the moments of spectrum. Let \( X_i(k) \) is the frequency sequence computed by STFT for the \( i \)th segment, we define the \( n \)th moment of the frequency sequence \( X_i(k) \) as:

\[
M_n = \sum_k k^n X_i(k),\]

(5.13)
where $\omega_k$ is a production of k and frequency resolution-- that is the half of the sampling frequency.

The central frequency and the variance are:

$$\text{Frequency}_\text{EXP}(t) = \frac{M_1}{M_0},$$

(5.14)

$$\text{Frequency}_\text{VAR}(t) = M_2 - \left(\frac{M_1}{M_0}\right)^2.$$  

(5.15)

The STFT is employed in feature extraction in this thesis. Unlike Kristin A. Farry’s approach, the moments of spectrum are taken as features instead the averaged integrals. The performance is discussed based on experimental results. The algorithm for feature extraction method using STFT is described as following.

1. Segment the EMG signal at initial phase of a grasp motion into $n$ segments with an overlapping portion of $p$ between every 2 consecutive segments.
2. Multiply each segment with a hamming window of the size of the segment.
3. Compute Discrete Fourier transform (DFT) for each of the $n$ windowed segment as defined in (5.4). This results $n$ sequences $X_i(k), i = 1, 2, ..., n$ in frequency domain.
4. Compute the 0, 1st and 2nd moments for $X_i(k)$.
5. Take the zero, first and second moment and integrals as features.
6. Repeat step 1, 2, 3, 4 for each of the four EMG channel.

The nature of EMG suggests that most of its energy is distributed within 0~300 Hz range. We used a similar feature extraction method as Kristin A. Farry’s approach. The resulted 0~300 Hz frequency range is divided into 0~100, 100~200, and 200~300 intervals. The feature values
are integrals of the magnitude of the spectrum in the three frequency ranges respectively. We call this method as Spectrum Integral Method 1 of STFT.

An alternate way of the method is to divide the frequency range into intervals of 0~75, 75~150 and 150~300, and take the integrals as feature values. We call this method as Spectrum Integral Method 2 of STFT.

5.2.6 Experimental Results

As stated in chapter 4, the number of grasp types can be either six or four.

The following values are taken as features from each channel.

- Zero moment of STFT, denoted as M0
- Zero moment and 1st moment, denoted as M0 and M1
- Zero moment, 1st moment and 2nd moment, denoted as M0, M1, and M2

The successful classification rates are recorded as:

- Mean value of the successful classification rate (hit rate) for 100 independent training-testing experiments.
- Standard deviation (Std) of the successful classifications rates for the same 100 independent experiments.

The number of resulting output nodes of DSAM classifier are also recorded as:

- Mean of the number output nodes for 100 independent experiments.
- Standard deviation (Std) of the number of output nodes

The mean and standard deviation (Std) are measured for raw data of the three lengths of 200, 300, and 400 ms. The moments and spectrum integral methods are used for six and four grasp types for the two STFT methods. The mean and Std for the hit rates and number of output nodes are listed in the following tables 5-1 to 5-8.
Experiment 1: Six grasp types, moments of STFT.

Table 5-1: Hit rate for 100 tests for the moments of STFT (six grasp types)

<table>
<thead>
<tr>
<th></th>
<th>M0</th>
<th>M0 and M1</th>
<th>M0, M1, and M2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>200 ms</td>
<td>58.46</td>
<td>4.32</td>
<td>61.01</td>
</tr>
<tr>
<td>300 ms</td>
<td>70.52</td>
<td>3.25</td>
<td>72.61</td>
</tr>
<tr>
<td>400 ms</td>
<td>82.38</td>
<td>3.52</td>
<td>84.52</td>
</tr>
</tbody>
</table>

Table 5-2: Number of output nodes for 100 tests for the moments of STFT (six grasp types)

<table>
<thead>
<tr>
<th></th>
<th>M0</th>
<th>M0 and M1</th>
<th>M0, M1, and M2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>200 ms</td>
<td>6.2</td>
<td>0.4</td>
<td>6.1</td>
</tr>
<tr>
<td>300 ms</td>
<td>6.0</td>
<td>0.2</td>
<td>6.2</td>
</tr>
<tr>
<td>400 ms</td>
<td>6.4</td>
<td>0.6</td>
<td>6.5</td>
</tr>
</tbody>
</table>
Experiment 2: Four grasp types, moments of STFT.

Table 5-3: Hit rate for 100 tests for the moments of STFT (four grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>M0</th>
<th>M0 and M1</th>
<th>M0, M1, and M2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>200 ms</td>
<td>76.01</td>
<td>4.26</td>
<td>68.78</td>
</tr>
<tr>
<td>300 ms</td>
<td>86.60</td>
<td>3.56</td>
<td>83.27</td>
</tr>
<tr>
<td>400 ms</td>
<td>94.11</td>
<td>2.33</td>
<td>91.06</td>
</tr>
</tbody>
</table>

Table 5-4: Number of output nodes for 100 tests for the moments of STFT (four grasps types)

<table>
<thead>
<tr>
<th>%</th>
<th>M0</th>
<th>M0 and M1</th>
<th>M0, M1, and M2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
<td>Mean</td>
</tr>
<tr>
<td>200 ms</td>
<td>6.4</td>
<td>0.6</td>
<td>6.7</td>
</tr>
<tr>
<td>300 ms</td>
<td>7.2</td>
<td>1.0</td>
<td>7.0</td>
</tr>
<tr>
<td>400 ms</td>
<td>6.9</td>
<td>0.8</td>
<td>6.4</td>
</tr>
</tbody>
</table>
Experiment 3: Six grasp types, spectrum integral.

Table 5-5: Hit rate for 100 tests for the spectrum integral of STFT (six grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>Spectrum Integral Method 1</th>
<th>Spectrum Integral Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>200 ms</td>
<td>54.19</td>
<td>4.09</td>
</tr>
<tr>
<td>300 ms</td>
<td>66.04</td>
<td>4.59</td>
</tr>
<tr>
<td>400 ms</td>
<td>74.91</td>
<td>3.14</td>
</tr>
</tbody>
</table>

Table 5-6: Number of output nodes for 100 tests for the spectrum Integral of STFT (six grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>Spectrum Integral Method 1</th>
<th>Spectrum Integral Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>200 ms</td>
<td>6.1</td>
<td>0.3</td>
</tr>
<tr>
<td>300 ms</td>
<td>6.0</td>
<td>0.1</td>
</tr>
<tr>
<td>400 ms</td>
<td>6.0</td>
<td>0.2</td>
</tr>
</tbody>
</table>
Experiment 4: Four grasp types, spectrum integral

Table 5-7: Hit rate for 100 tests for the spectrum integral of STFT (four grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>Spectrum Integral Method 1</th>
<th>Spectrum Integral Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>200 ms</td>
<td>57.60</td>
<td>3.90</td>
</tr>
<tr>
<td>300 ms</td>
<td>70.77</td>
<td>3.86</td>
</tr>
<tr>
<td>400 ms</td>
<td>77.86</td>
<td>3.66</td>
</tr>
</tbody>
</table>

Table 5-8: Number of output nodes for 100 tests for the spectrum integral of STFT (four grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>Spectrum Integral Method 1</th>
<th>Spectrum Integral Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>200 ms</td>
<td>6.0</td>
<td>0.2</td>
</tr>
<tr>
<td>300 ms</td>
<td>6.1</td>
<td>0.2</td>
</tr>
<tr>
<td>400 ms</td>
<td>6.5</td>
<td>0.7</td>
</tr>
</tbody>
</table>
a. Multi-taper Approach (Thomson Transform)

In order to discuss the Thomson Transform, we will first introduce some necessary background facts.

5.3.1 Power Spectral Density (PSD) of EMG Signal

EMG Signals cannot be described by a well-defined formula. The distributions for the various grasp types can be however described with the probability laws. EMG signal is a random process whose value at each time is a random variable.

The Fourier transform we used in the previous section views nonrandom signals as weighted integral of sinusoidal functions. Since a sample function of random process can be viewed as being selected from an ensemble of allowable time functions, the weighting function for a random process must refer in some way to the average rate of change of the ensemble of allowable time functions [12].

The power spectral density (PSD) of a wide sense stationary random process $X(t)$ is computed from the Fourier transform of the autocorrelation function $R(\tau)$:

$$S_X(f) = \int_{-\infty}^{\infty} R(\tau) \cdot e^{-j2\pi ft} d\tau,$$

where the autocorrelation function

$$R(\tau) = E\left[ X(t+\tau)X(t) \right].$$

The nonparametric methods are methods in which the estimate of PSD is made directly from a signal itself. One type of such methods is called periodogram.

The periodogram estimate for PSD for discrete time sequence $x_1, x_2, x_3, \ldots, x_k$ is defined as square magnitude of the Fourier transform of data:
An improved nonparametric estimator of the PSD is proposed by Welch P.D. The method consists of dividing the time series data into (possibly overlapping) segments, computing a modified (windowed) periodogram of each segment, and then averaging the PSD estimates. The result is Welch's PSD estimate.

The multitaper method (MTM) is also a nonparametric PSD estimation technique which uses multiple orthogonal windows.

5.3.2 Leakage Versus Variance for PSD

The periodogram is a biased estimator. The bias becomes more significant for EMG signals with smaller length.

The estimated PSD converges to the true PSD as \( k \to \infty \). However, there are some limitations for time sequences of finite length:

- **Bias**: an incorrect spectrum caused by limited number of time samples.
- **Leakage**: another form of bias.
- **Resolution**: the ability to discriminate spectral features, a form of bias.
- **Variance**: The variance of periodogram estimate is proportional to the square of the power spectral density and does not approach zero as the length of time sequence increases.

One method to reduce the variance and smooth the estimation is by taking the average of \( N \) independent periodograms. Bartlett’s smoothing procedure cuts the data into \( N \) blocks, computes the periodogram for each separated block and takes an average of them.

\[
\hat{S}(f) = \frac{1}{k} \left| \sum_{m=1}^{n=k} x_m \cdot e^{-j2\pi fm} \right|^2.
\] (5.18)
There is always a trade-off between variance and bias. The bias increases as the length of block decreases. The limitation of periodogram due to finite length of data usually includes bias, variance, leakage and resolution. Increasing block number introduces less variance, while using time sequences of smaller length always has poorer frequency resolution, larger bias and frequency leakage.

5.3.3 Discrete Prolate Spheroidal Sequences (DSPP)

The discrete prolate spheroidal sequences (DPSS) are employed to balance the trade-off between variance and bias which are mentioned above. This section will introduce the DPSS.

 Proposed by D. Slepian in 1978 [17], the discrete prolate spheroidal wave function (DPSWF) $U_k(N,W; f')$ and their eigenvalues $\lambda_k(N,W)$ are defined as the solutions of integral equation:

$$\int_{-\infty}^{\infty} \sin N\pi(f-f') U_k(N,W; f') df' = \lambda_k(N,W) U_k(N,W; f)$$

$$-\infty < f' < \infty, k = 0, 1, 2, ..., N-1,$$

The function has $N$ non-zero eigenvalues $\lambda_0(N,W), \lambda_1(N,W), ..., \lambda_{N-1}(N,W)$ such that $\lambda_0(N,W) > \lambda_1(N,W) > ... > \lambda_{N-1}(N,W) > 0$. It is normalized so that

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} [U_k(N,W; f)]^2 df = 1,$$

where

$$U_k(N,W; 0) \geq 0, \frac{dU_k(N,W; 0)}{df} \geq 0, k = 0, 1, ..., N-1.$$ (5.21)

For each $k = 0, 1, ..., N-1$, the DPSS is defined as the real solution of the system of equations
\[\sum_{m=0}^{N-1} \sin 2\pi \frac{W(n-m)}{n-m} v_m^{(k)} (N,W) = \hat{\lambda}_k (N,W) v_m^{(k)} (N,W),\] 
\[n = 0, \pm 1, \pm 2, \ldots\]

We normalize it as:
\[\sum_{m=0}^{N-1} v_j^{(k)} (N,W)^2 = 1,\]

where
\[\sum_{j=0}^{N-1} v_j^{(k)} (N,W) \geq 0, \sum_{j=0}^{N-1} (N-1-2j) v_j^{(k)} (N,W) \geq 0.\]

Both DPSWF and DPSS are doubly orthogonal. We have
\[U_k (N,W; f) = \varepsilon_k \sum_{n=0}^{N-1} v_n^{(k)} (N,W) e^{-i\pi (N-1-2n)} f,\]

where \(k = 0, 1, \ldots, N-1\), and \(\varepsilon_k\) is 1 for \(k\) even and \(i\) for \(k\) odd.

The DPSS are the Fourier transforms of the DPSWF [17], [34], that is
\[v_n^{(k)} (N,W) = \frac{1}{\varepsilon_k \hat{\lambda}_k (N,W)} \int_{-W}^{W} U_k (N,W; f) e^{-i2\pi f[n-(N-1)/2]} df,\]

where \(k = 0, 1, \ldots, N-1\).

The following figure illustrate an example of DPSS:
5.3.4 Thomson’s Method

Thomson suggested a multiple windows spectrum estimation scheme for low bias and low variance estimation [34]. Instead of cutting the time data into a number of blocks, Thomson’s method computes the periodogram for entire windowed time sequence, and then averages the results. A set of orthogonal time windows is employed. These windows are DSPPs.

We let $v_k[n], n = 0, 1, 2, ..., N-1; k = 0, 1, 2, ..., 2NW - 1$ be a set of $2NW$ orthogonal DSPP windows. Each window is a time sequence of $N$ points. The Thomson’s MWM power spectral density estimator for an $N$-point time sequence is:

$$
\hat{S}(f) = \frac{1}{2NW} \sum_{k=0}^{2NW-1} \frac{1}{\lambda_k} \left| \sum_{n=0}^{N-1} x[n] v_k[n] e^{-j2\pi fn} \right|^2 .
$$

(5.27)
Thomson’s MWM method gives dramatically lower bias and variance than the periodogram [34]. It balances the variance and the resolution very well. There are $2NW$ tapers used to form the estimate. The variance decreases as the value of $NW$ increases. The bias increases and after employing a larger number of windows. More spectral leakage is also introduced after the number of windows increases. An optimal value of $NW$ depends on the nature of data input. We will find the best $NW$ by experiment.

![Power Spectral Density Estimation of EMG signal](image)

Figure 5-15: Power spectral density (PSD) estimate of the EMG signal

The three graphs are the PSD by periodogram, Welch’s method and Thomson’s Multiple Taper Method (MTM) with eight tapers.

We can observe from the figure that the variance of MTM is smaller than that of periodogram, while the local peaks and nadirs are still preserved very well. The spectrum of
Welch’s method has smooth shape and the variance is also small, which is obtained by sacrificing the frequency resolution.

5.3.5 Short Time Thomson Transform in EMG Feature Extraction

Kristin A. Farry and other researchers extended Thomson’s Method to short-time Thomson’s transform (STTT), by using the same idea of STFT for time-frequency analysis. Three non-overlapping 64 ms time sequences were used for STTT. Each 64 ms time sequence is a segment of the entire 192 ms EMG signal, which is recorded during the initiation of grasp motion. A good classification result by STTT was obtained. It deadly outperformed the STFT. This method is less sensitive to windows size comparing with the periodogram method. The EMG signal used in STTT has smaller size than that used in STFT at comparable big rate. This is an important fact relevant to the prosthetic hand control in real-time.

We use the STTT in feature extraction methods in a similar way we used in section 5.2. The algorithm is summarized as follows:

Step 1: Segment the EMG signal into $n$ overlapping sequences.

Step 2: Compute Thomson’s transform for each segment. Employ a set of Slepian windows (DPSS). For each segment multiply the segment with each of the Slepian widows, get its periogram, then average the resulting spectrums to get $X_i(k), i = 1, 2, ..., n$.

Step 3: Compute the 0, 1$\text{st}$ and 2$\text{nd}$ moments for $X_i(k)$. Take the moments as features.

Repeat step 1, 2, 3 for each of the four EMG channel.
The Spectrum Integral is also experimented in STTT. As a result of STTT the 0~300 Hz frequency range is divided into 0~100, 100~200, and 200~300 Hz (integral method 1), or 0~75, 75~150 and 150~300 Hz (integral method 2).

5.3.6 Experimental Results

We have similar experimental results for STTT as for STFT shown in the previous section. The mean and standard deviation (Std) are measured for raw data of the 200, 300, 400 ms sequences for each implemented method. The moments and spectrum integral methods are used for six and four grasp types. The hit rate and the number of output nodes are also measured. The results are shown in the table 5-9 to 5-16.
Experiment 1: Six grasp types, moments of STTT.

Table 5-9: Hit rate for 100 tests for the moments of STTT (six grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>M0</th>
<th>Std</th>
<th>M0 and M1</th>
<th>Std</th>
<th>M0, M1, and M2</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 ms</td>
<td>67.84</td>
<td>3.88</td>
<td>69.13</td>
<td>4.01</td>
<td>70.43</td>
<td>3.88</td>
</tr>
<tr>
<td>300 ms</td>
<td>81.89</td>
<td>3.22</td>
<td>83.69</td>
<td>3.09</td>
<td>84.96</td>
<td>2.99</td>
</tr>
<tr>
<td>400 ms</td>
<td>88.97</td>
<td>2.22</td>
<td>91.03</td>
<td>2.01</td>
<td>92.09</td>
<td>2.20</td>
</tr>
</tbody>
</table>

Table 5-10: Number of output nodes for 100 tests for the moments of STTT (six grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>M0</th>
<th>Std</th>
<th>M0 and M1</th>
<th>Std</th>
<th>M0, M1, and M2</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 ms</td>
<td>6.0</td>
<td>0.1</td>
<td>6.2</td>
<td>0.4</td>
<td>6.3</td>
<td>0.5</td>
</tr>
<tr>
<td>300 ms</td>
<td>6.2</td>
<td>0.5</td>
<td>6.3</td>
<td>0.5</td>
<td>6.1</td>
<td>0.2</td>
</tr>
<tr>
<td>400 ms</td>
<td>6.3</td>
<td>0.5</td>
<td>6.5</td>
<td>0.7</td>
<td>6.8</td>
<td>0.8</td>
</tr>
</tbody>
</table>
Experiment 2: Four grasp types, moments of STTT.

Table 5-11: Hit rate of 100 tests for the moments of STTT (four grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>M0</th>
<th>Mean</th>
<th>Std</th>
<th>M0 and M1</th>
<th>Mean</th>
<th>Std</th>
<th>M0, M1, and M2</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 ms</td>
<td>88.37</td>
<td>4.26</td>
<td>M0</td>
<td>82.04</td>
<td>4.55</td>
<td>M0</td>
<td>79.40</td>
<td>4.18</td>
<td></td>
</tr>
<tr>
<td>300 ms</td>
<td>94.18</td>
<td>1.92</td>
<td>M0</td>
<td>90.89</td>
<td>2.38</td>
<td>M0</td>
<td>90.41</td>
<td>2.67</td>
<td></td>
</tr>
<tr>
<td>400 ms</td>
<td>96.61</td>
<td>1.73</td>
<td>M0</td>
<td>94.67</td>
<td>2.38</td>
<td>M0</td>
<td>95.78</td>
<td>1.70</td>
<td></td>
</tr>
</tbody>
</table>

Table 5-12: Number of output nodes for 100 tests for the moments of STTT (four grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>M0</th>
<th>Mean</th>
<th>Std</th>
<th>M0 and M1</th>
<th>Mean</th>
<th>Std</th>
<th>M0, M1, and M2</th>
<th>Mean</th>
<th>Std</th>
</tr>
</thead>
<tbody>
<tr>
<td>200 ms</td>
<td>6.8</td>
<td>0.8</td>
<td>M0</td>
<td>6.9</td>
<td>0.8</td>
<td>M0</td>
<td>6.7</td>
<td>0.7</td>
<td></td>
</tr>
<tr>
<td>300 ms</td>
<td>6.4</td>
<td>0.6</td>
<td>M0</td>
<td>7.4</td>
<td>0.9</td>
<td>M0</td>
<td>7.0</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>400 ms</td>
<td>6.8</td>
<td>0.8</td>
<td>M0</td>
<td>9.3</td>
<td>1.3</td>
<td>M0</td>
<td>6.0</td>
<td>0.0</td>
<td></td>
</tr>
</tbody>
</table>
Experiment 3: Six grasp types, spectrum integral of STTT.

Table 5-13: Hit rate of 100 tests for the spectrum integral of STTT (six grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>Spectrum Integral Method 1</th>
<th>Spectrum Integral Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>200 ms</td>
<td>68.96</td>
<td>3.52</td>
</tr>
<tr>
<td>300 ms</td>
<td>85.43</td>
<td>2.53</td>
</tr>
<tr>
<td>400 ms</td>
<td>92.44</td>
<td>1.91</td>
</tr>
</tbody>
</table>

Table 5-14: Number of output nodes for 100 tests for the spectrum integral of STTT (six grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>Spectrum Integral Method 1</th>
<th>Spectrum Integral Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>200 ms</td>
<td>6.3</td>
<td>0.5</td>
</tr>
<tr>
<td>300 ms</td>
<td>6.3</td>
<td>0.5</td>
</tr>
<tr>
<td>400 ms</td>
<td>6.6</td>
<td>0.7</td>
</tr>
</tbody>
</table>
Experiment 4: Four grasp types, spectrum integral of STTT.

Table 5-15: Hit rate of 100 tests for the spectrum integral of STTT (four grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>Spectrum Integral Method 1</th>
<th>Spectrum Integral Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>200 ms</td>
<td>74.73</td>
<td>3.82</td>
</tr>
<tr>
<td>300 ms</td>
<td>89.91</td>
<td>2.42</td>
</tr>
<tr>
<td>400 ms</td>
<td>94.07</td>
<td>2.01</td>
</tr>
</tbody>
</table>

Table 5-16: Number of output nodes for 100 tests for the spectrum integral of STTT (four grasp types)

<table>
<thead>
<tr>
<th>%</th>
<th>Spectrum Integral Method 1</th>
<th>Spectrum Integral Method 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Std</td>
</tr>
<tr>
<td>200 ms</td>
<td>6.1</td>
<td>0.3</td>
</tr>
<tr>
<td>300 ms</td>
<td>6.3</td>
<td>0.6</td>
</tr>
<tr>
<td>400 ms</td>
<td>7.1</td>
<td>0.8</td>
</tr>
</tbody>
</table>
5.4 Wavelet Approach

5.4.1 Introduction to Wavelets

A signal or function \( f(t) \) can often be better analyzed if it is expressed as a linear decomposition like

\[
f(t) = \sum_l a_l \psi_l(t).
\]  
(5.28)

where \( a_l \) are the expansion coefficients, while \( \psi_l(t) \) form a set of functions called wavelet basis.

If the basis is orthogonal, we have

\[
\langle \psi_k(t), \psi_l(t) \rangle = \int \psi_k(t) \psi_l(t) \, dt = 0, k \neq l.
\]  
(5.29)

Coefficients can be computed as inner product:

\[
a_k = \langle f(t), \psi_k(t) \rangle = \int f(t) \psi_k(t) \, dt.
\]  
(5.30)

Fourier analysis consists of breaking up a signal into sine waves of various frequencies. The orthogonal basis functions \( \psi_k(t) \) are \( \sin(k\omega_0 t) \) and \( \cos(k\omega_0 t) \) with frequency of \( k\omega_0 \).

Similarly, wavelet analysis is the breaking up of a signal into shifted and scaled versions of the original (or mother) wavelet \( \psi(t) \).

The continuous wavelet transform (CWT) is defined as

\[
W_\psi(b,a) = \int_{-\infty}^{\infty} f(t) \overline{\psi_{b,a}(t)} \, dt,
\]  
(5.31)

where

\[
\psi_{b,a}(t) = \frac{1}{\sqrt{a}} \psi \left( \frac{t-b}{a} \right), a > 0.
\]  
(5.32)

Here \( a \) and \( b \) are parameters that specify the scale and the shift, respectively.
The discrete wavelet transform (DWT) is obtained by choosing scales and shifts based on powers of two: \( a = 2^{-s} \) and \( b = k 2^{-s} \). Then we have

\[
W_\psi(k 2^{-s}, 2^{-s}) = 2^s \int_{-\infty}^{\infty} f(t) \psi(2^s t - k) \, dt,
\]

It is can be rewritten as the following if \( f(t) \) is sampled by 1.

\[
W_\psi(k 2^{-s}, 2^{-s}) = 2^s \sum_n f(n) \psi(2^s n - k).
\]

We can compare time-frequency localization of STFT and DWT by the following figure:

![Figure 5-16: The tiling of the time-frequency plane of STFT and DWT](image)

To compute the wavelet transform of a function at some point in the time-scale plane, we do not need to know the function values for the entire time axis. It is can be done almost in real time [39]. In wavelet analysis, we often speak of approximations and details. The approximations are the high scale, low frequency components of the signal. The details are the low scale, high frequency components. The signal is decomposed into successive approximation and details in the computation of the wavelet transform. The approximation is repeatedly decomposed into a higher level of approximation and detail, and the details are kept at each stage. This process is called the wavelet decomposition tree.

The wavelet packet transform (WPT) is a generalization of wavelet decomposition in which the details as well as the approximations can be decomposed. The most suitable
decomposition of a given signal can be selected. Therefore there is an adaptive time-frequency tiling in WPT instead of having a fixed time-frequency tiling as in DWT.

5.4.2 Wavelet-based Approach Applied to EMG Feature Extraction

Wavelet technique is suitable for time-frequency analysis. A lot of EMG feature extraction methods are explored based on the wavelet technique [40][41][42][43].

Wavelet approach is used in research work at Hokkaido University, Japan. The Gabor mother wavelet is used in some work because this mother wavelet uses complex exponential function, which also plays a main role in the Fourier transform [42]. Spectra summation is calculated and the local peak points to the summation is detected after a wavelet transform. The features are the scale coefficients those consist of the spectral in the peak points.

Kevin Englehart evaluated various mother wavelets such as the Daubechies, Coiflets, Symmlets, Biorthogonal, and etc. No single wavelet can be identified clearly as a superior [27]. It is also proposed that the wavelet transform local extrema is a suitable method for the feature values, because it can deal with the time-variant nature of wavelets very well. The spectral at local extrema is invariant to time shift. The dimension of feature space is usually high in this case. There are many extremas for a wavelet transform, and each extrema needs a number of parameters to represent it. The feature selection and feature projection methods such as class separability (CS) and principal components analysis (PCA) are employed for the dimensionality reduction.

We already have good results by the short time Thomson’s transform and multiple windows methods in the presented thesis. The wavelet approach is not necessary for the presented work. Therefore it is beyond the scope of presented work.