Detection of In-Air Alphabetical Hand Gesture using Doppler Effect

Utpal Kumar Dey*, Kuntal Patel† and Zhaochen Gu*
*Department of Computer Science and Engineering
University of North Texas, Denton, Texas 76203-5017
Email: utpal-kumardey@my.unt.edu
†Department of Computer Science and Engineering
University of North Texas, Denton, Texas 76203-5017
Email: kuntalbahennarshibhaipatel@my.unt.edu
‡Department of Computer Science and Engineering
University of North Texas, Denton, Texas 76203-5017
Email: zhaochengu@my.unt.edu

***Abstract*—The use of Doppler effect is being evolved in various fields. Hand gesture recognition using Doppler effect has been an important area of research for more than five years and in-air hand gestures producing alphabets is a new edition in this field. This study includes methods for precise identification of alphabets with existing infrastructure. The system can be used to recognize quantitatively detailed information of movements from the signal reflected by the finger. Existing methods use processed Doppler shift and hence cannot precisely detect similar motions whereas our method is able to achieve more accurate results because of extraction of feature primitives in terms of similar motions. This study demonstrates the basic idea to detect and capture the signal having doppler shift. Moreover, each step of processing that signal is explained afterward. That means, how transformation can be applied in the signal, how noise can be reduced, how certain properties can be retrieved from the signal. Then, a technique is applied to fit those measures in a manageable range. Finally, the system can be trained to make specific action using the measurements values with the help of Hidden Markov Model (HMM).**

I. Introduction

The Doppler effect or Doppler shift describes the changes in frequency of any kind of sound or light wave produced by a moving source with respect to an observer. when a signal from a transmitter is obstructed by any moving object then there is a change occurs in frequency which is received at the receiver. This shift in the frequency in the signal is referred as Doppler effect. If transmitted signal is denoted as f_r, received signal is denoted as f_c and Doppler effect is denoted as f_D then

$$f_r = f_c - f_D$$

where $f_D = \frac{v f_c \cos\theta}{c}$, v is the velocity of the object, c is the speed of light and θ is the angle between transceiver and receiver.

Past research has been proved that leveraging Doppler effect is very useful and among them there are two major types can be found which depicts infrastructure based systems and less or no infrastructure based systems.

Gesture detection is to detect movements using computer algorithms. This technique helps humans to naturally communicate with machines. There are several gesture detection methods exist like image processing and piezoelectric gloves. Focus in this project is to use of ultrasonic sound waves for gesture detection based on Doppler effect.

A. Generic gesture recognition system

The general gesture recognition process in any kind of system can be broken down into the following components.

- The first stage is to specify hardware configuration of the system and how data is collected for the recognition process.
- The second stage is a pre-processing stage. Here the gathered data is passed through different filtering process like transformation from one domain to another domain, noise elimination and normalization. This processes prepare the data for the main computational stage, the feature extraction.

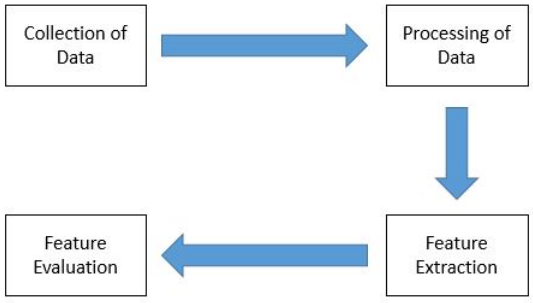

Figure 1: General Flow of Gesture Recognition

- The features of the input signal are extracted in the third stage. In addition, the features are combined to generate feature vectors.
- In the fourth stage, the feature vectors are then evaluated in one or more of several possible ways to make a

decision about which gesture the system is most likely subjected.

II. Related Work

A. Gesture Recognition using DSP

Most of the air gesture sensing technologies use complex algorithms, which require hardware and a large amount of processing power. Also, hardware-dependent gesture sensing is compromising in adverse environment. Garg's [1] project provides better solutions. It can also be used to complement the existing visual gesture detection techniques. His project uses the Doppler Shift principle of change in the frequency of a reflected wave to detect motion and gestures, using digital signal processor (DSP).

The project is inspired by the Microsoft SoundWave [2] gesture detection system, which uses hardware such as a PC, microphone and speaker. The SoundWave gesture detection system utilizes the speaker and the microphone already attached in most common devices to sense in-air gesture. [1] The project uses the Doppler Effect to implement in-air gesture recognition on the BeagleBoards DSP core.

The system comprises three main features:

1. Provides gesture sensing using just a speaker and microphone.
2. Uses digital signal processing on the BeagleBoards DSP processor core to detect a frequency shift.
3. Provides the visually challenged with a user interface input.

[1] The project provides an affordable and more efficient motion tracker to help the visually impaired people. In this project, a sound wave which is impossible to hear by human being is generated in the air which reflects back after striking any object present in its way. Important parameters, for instance, frequency shift, amplitude, and phase shift of the reflected sound wave can be extracted from the received signal and then processed to detect different motions and gestures. Based on these information, voice instructions can be given to the blind person. This information can be classified into following three types of gestures mainly.

1. Left Swipe
2. Right Swipe
3. Double Throw

B. Efficient Gesture Recognition using PC

Haojun Ai [3] presents an improved gesture recognition system that also uses the Doppler effect of ultrasound to sense in-air hand gestures. The system can identify more accurate gestures than other systems without any alteration to non-commercial laptops. The system can detect detailed and even complex movements from the signals reflected by a moving body. A Hidden Markov Model(HMM) is used to construct a library of independent, discrete gestures. The gestures can be mapped to diverse application actions.

The HMM is a sequence model. A sequence model or sequence classifier is a model whose job is to assign a label or class to each unit in a sequence, thus mapping a sequence of observations to a sequence of labels. [4] An HMM is a probabilistic sequence model: given a sequence of units i.e. gestures, they compute a probability distribution over possible sequences of labels and choose the best label sequence. His method can identify similar gestures but still be able to distinguish them properly. Proposed system reduces false positives caused by causeless motions and is versatile and adaptable to multiple devices. He implemented a proof-of-concept prototype on a laptop and evaluated the system. His results show that the system recognizes six gestures with an average accuracy almost 97% and 18 gestures including similar ones with 95% accuracy.

The experiments were conducted in the conference room of laboratory using desktop PCs with an external USB soundcard and microphone. User faces the PC with the audio interface deployed right in front of him. The device is placed above the user in horizontal direction. The speaker generates a pure sine-wave with 18 kHz and the microphone picks up the signals reflected by the moving body. The PC is used to control the operation of system and respond to the gestures. This system also tested several laptops.

C. Rescue Operation using Wi-Fi Signals

Rescue after accidents becomes a challenge if there is no accessible infrastructure in the area. As an increasing number of people using smart and wearable devices, Yuan-Yao Shih [5] uses the Doppler effect in the rescue system to assist the rescuer to locate the direction of Wi-Fi signals from disaster survivors' mobile phone. A smart device can emit Wi-Fi signals and it can be used as a beacon signal for rescuers to locate without infrastructure. He achieved accuracy in direction-finding by developing algorithm, and hence, the accuracy and sensitivity of relatively small frequency Doppler shift can be improved. Also, he designed an active detector to guarantee that the Wi-Fi signal can be transmitted continuously from the survivors' devices.

He also made a decision logic to minimize energy consumption by the active scheme. He developed the rescue system as a mobile application on Android smart phones and conducted sufficient experiments in real-world environments. Results show that the proposed system can reduce rescue times by significant amount. Shih's future research will extend the system to support distinct kinds of common wireless communication protocols (e.g., Bluetooth, ZigBee, 802.15.6), making it compatible with a wider range of smartphones and wearable devices. We will also make the system operable on aircraft without a human pilot, which can be a helpful tool for Search and Rescue (SAR) operations. In addition, to provide more precise directional guidance, we will consider the possibility of attaching an additional light-weight antenna system to the smartphones.

The scenario of rescue system works in the following way: the survivor is defined as a target who lost in the remote wildlife area. His or her smart devices will emit probe packets on occasion without connect to any Wi-Fi network. Once the

rescuer team received the packets that sent from the target, they can use their smart phones to search the location of the target. The result shows that the signals can arrive faster to rescuers as they move towards the target. On contrary, the signals arrive slower If the rescuer move away from the survivor. Shih's team modified their rescue system using OFDM symbols to extract information from the rescuers smartphone. The purpose of this is to improve accuracy of direction-finding. OFDM is known as Orthogonal frequency-division multiplexing which can convert digital data on multiple carrier frequencies. Another mathematical model is applied to the design of the system is Fourier Transform (FFT). The FFT takes a time based pattern to produce frequency-time Doppler Profile.

D. Gesture Recognition Using Smart Devices

Gesture recognition is the bridge for interaction between human and computing system. Smart mobile devices like phones or tablets have become more available and popular. Hence, enabling these devices to recognize gesture is very important now a days. The approach in this paper [6] is to recognize gesture based on ultrasonic in low computational complexity and without using extra infrastructure. This approach is not prone to ordinary noise or light. With the flexibility of wider operating range and angle this approach can be applied on portable devices.

This paper [6] uses loudspeaker and microphone embedded smart devices to produce ultrasonic sound and detect the reflected sound. The technique samples the ultrasonic sound continuously while a gesture is being performed as it utilizes the doppler shift of the sound reflected by moving human body. Then the gestures are combined using simple pattern matching technique to differentiate which is supervised by machine learning methods. Smart devices can detect 24 predefined set of gestures with 94% accuracy.

The authors developed a system plug-in for Android platform to support gesture recognition service and validated their approach. Moreover, users can train their system according to defined gestures as needed. They also developed two real time games to prove the low latency and high accuracy of their system. Detected gestures can be mapped to operate variety of controls. For example, the swiping or continuous slapping gesture can be used to scroll web pages, flip e-books, pause movies.

III. Theory of Principle

An ultrasonic sound of around 20KHz should be emitted and reflected sound wave should be processed properly. According to the Doppler formula,

$$f_r = f_t \frac{v_s + v_t}{v_s - v_t}$$

Where f_r is the frequency of received ultrasonic sound, f_t is the frequency of transmitted ultrasonic sound, v_t is the speed of individual gesture, v_s is the speed of sound in air.

After receiving the sound, it should pass through Fast Fourier Transform (FFT) to convert the signal from time domain to frequency domain. The signal is then passed through a noise elimination process and a normalization process. After that, features should be extracted from the filtered signal. Hence, seven feature properties are focused even though more features can be extracted. Because, seven features are enough to differentiate two different gestures. Finally, these features should be compared to pre-assigned values to make proper decision.

IV. Experimental Setup

For doing this experiment, a source is needed to generate the ultrasonic sound. A laptop having a speaker is used as such a source. An Android smart phone having a mic is used to receive the reflected sound. To detect the reflected sound, an app called UltraSound Detector is being used. This app presents sampling rate and also does the FFT of the received signal. The overall setup is presented in figure 2.

Figure 2: Intended Configuration for Experiment.

V. Work Flow

The plan of this work is presented in figure 3.

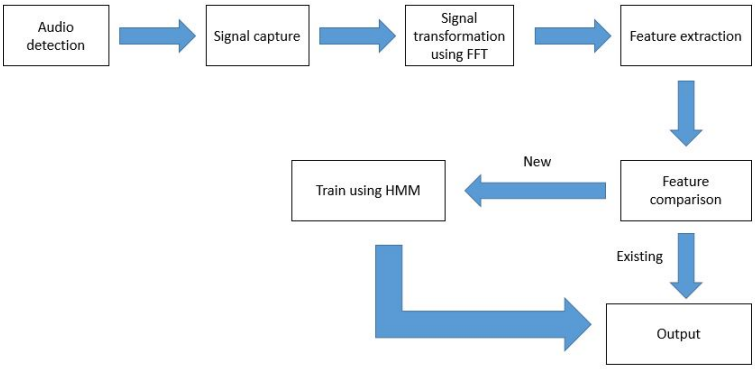

Figure 3: Flow Chart of the Experiment

A pre-recorded tone should be emitted from laptop at the beginning of this process. An Android app UltraSound Detector can be used to detect the reflected sound. In addition, values of few basic properties of the received signal can be extracted from the app.

Figure 4: Detection of Ultrasonic Sound using UltraSound Detector

After detecting the signal Fast Fourier Transform (FFT) is used to convert the signal from time domain to frequency domain.

A. *Fast Fourier Transform (FFT)*

The prime effect of human hand motions can be found in the frequency-shift occurring in the frequency-domain. Thus it was necessary to transform the original time-domain signal to frequency-domain leveraging the Fast Fourier Transform (FFT). To start with, windowing functions enhance the ability of an FFT to extract spectral data from signals. Windowing functions act on raw data to reduce the effects of the leakage that occurs during an FFT of the data. Leakage amounts to spectral information from an FFT showing up at the wrong frequencies [7]. Hamming window is used to reduce the amount of spectral leakage, it can be described by the following equation.

$$w(n) = 0.54 - 0.46\cos 2\pi\frac{n}{N}, 0 \leq n \leq N$$

Where $N = L - 1$, L represents the width in samples in discrete time [8]. The reason behind using hamming window is, the wave of the reflected signal is closely spaced. The result of using FFT and hamming window is presented in figure 5, 6 and 7 for a sample wave form. Figure 5 shows a sample waveform transformation from time domain to frequency domain, figure 6 shows a sample waveform in frequency domain after using hamming window.

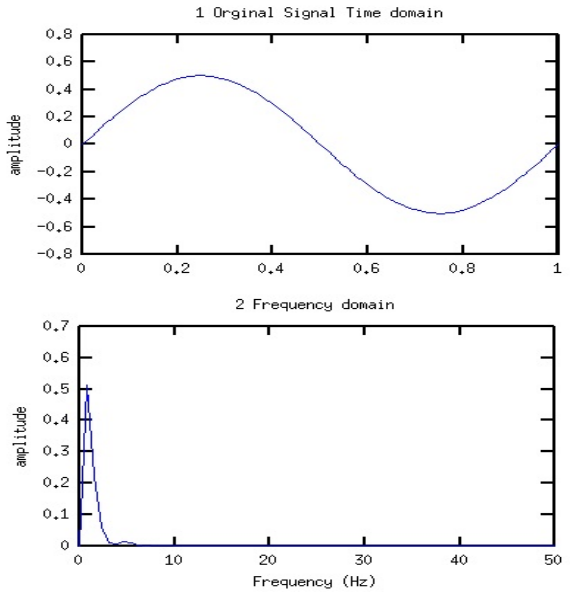

Figure 5: Sample signal transformation from time domain to frequency domain [9].

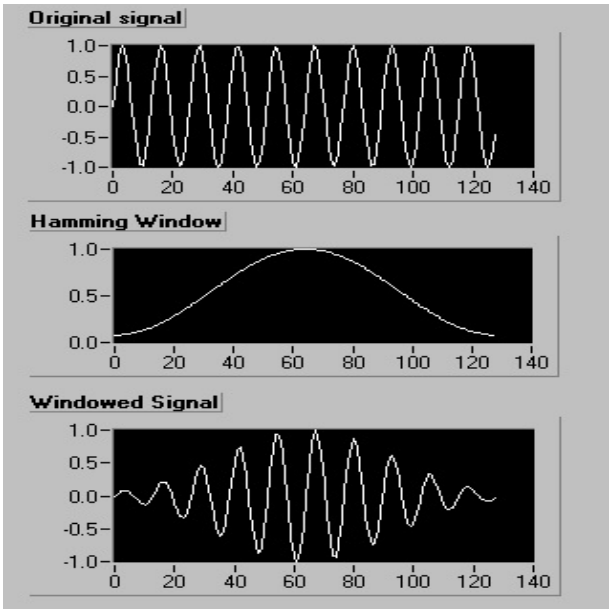

Figure 6: Sample signal in frequency domain after using hamming window [10].

B. *Noise Elimination*

After applying hamming window on the signal there may be chances of noise and to remove this noise the following noise elimination method will be applied. In order to eliminate noise not related to gestures, noise threshold vector N needs to be maintained. The noise threshold vector is initiated during the preparation process and updated during the gesture gap using following equation:

$$N_t = N_{t-1}.(1-\alpha) + E_t.\alpha$$
$$E_t.N_t \epsilon R^{60\times 1}$$
$$A_t^{'} = A_t - N_t$$

Where N_t is the updated noise threshold vector at time t. E_t is the environment noise vector at time t. α is 0.1. A_t is the original data vector and $A_t^{'}$ is noise-eliminated vector. Here, the value of E_t is retrieved during FFT process for each primitive and N_t is initialized to 1 at the beginning for all feature primitives. Then at each iteration noise threshold vector N_t will be updated and original data vector A_t will also be changed accordingly.

C. Feature Extraction

A feature vector is one method to represent an object, by finding measurements on a set of features. The feature vector is an K-dimensional vector that contains these measurements. In some applications it is not sufficient to extract only one type of feature to obtain the relevant information from the object data, instead two or more different features are extracted. A common practice is to organize the information provided by all these descriptors as the elements of one single vector, commonly referred to as a feature vector.

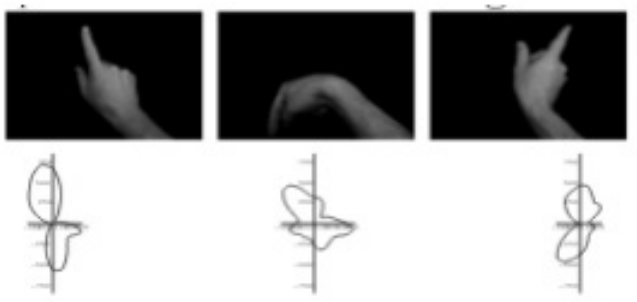

Figure 7: Features for Hand Gesture.

Seven feature properties are extracted from frequency-domain vector A to quantize the Doppler shift. Feature primitives are defined as $P_1, P_2, P_3, P_4, P_5, P_6, P_7$ where P_1 is the quantitative property based on motion characteristics. The reason behind extracting seven primitives is, seven properties are enough to differentiate between two individual gestures.

(a) P_1 is amplitude of the processed signal after FFT which is usually represents the strength of overall signal. P_1 is related to the size and proximity of the target. If the signal is reflected by human body, there will be a larger amplitude than a signal reflected only from hands. The amplitude increases when user moves closer to the device. It is captured by the following equation.

$$P_1 = A(f_0)$$

where A is this shift data processed with fft, $A(x_i)$ is the amplitude of frequency x_i, and f_0 is original frequency from speaker.

(b) P_2 represents the energy on the left side of the emitted peak, and P_3 represents the energy on the opposite side. When the frequency occurs a positive shift, P_3 will be increased, and a negative shift will increase the value of P_2 as:

$$P_2 = \sum_{i=a}^{b} A(i)$$

$$P_3 = \sum_{i=b}^{c} A(i)$$

where $a = f_0 - \Delta f$, $b = f_0$, and $c = f_0 + \Delta f$, Δf is the measurement of Doppler shift.

(c) P_4 is used to measure the velocity of movement and is computed by the bandwidth at amplitude θ_v. Change in this property is proportional to the absolute velocity of the target. It is given by $P_4 = \beta - \alpha$, where α is left margin of the signal and β is right margin of the signal. Here, $\alpha\epsilon(f_0 - \Delta f, f_0)$ and $\beta\epsilon(f_0, f_0 + \Delta f)$.

(d) P_5 represents the direction of movement in this time frame and defined by the following equation.

$$P_5 = \sum_{i=b}^{c} A(i) - \sum_{j=a}^{b} A(j), if P_4 \geq \sigma, else, P_5 = 0$$

where σ is the slowest velocity $(0.25m/sec)$ that can be detected.

(e) P_6 is used to measure the time duration of a gesture indirectly and the sequence number of the primitive in a gesture sample.

(f) P_7 is defined to quantize slight changes in gestures based on correlations between human body parts. Hand movements also put the arm in motion and the arm often performs different motions in similar gestures, especially the elbow joint. The size of an arm is larger than a hand but the speed of an arm movement is slower. Hence, more signal can be reflected in air but the frequency shifts slightly. Therefore, the bins near the emitted peak have higher amplitudes. We define the bandwidth at amplitude θ_d (approximately 90% of the maximal amplitude) as the value of this property.

D. Determining a Gesture Completion

One of the important point in gesture detection is how system can determine whether a gesture is being generated or not and when the gesture is completed. This decision can be made by judging feature primitives. Whether there is a shift occurred in frequency or not, can be found from feature primitive. Primitives can be converted to strings where three types of value will be stored. The values are, positive sign for right shift, negative sign for left shift and zero for no shift. The beginning of a gesture can be determined by a positive sign or negative sign. But the completion of a gesture can be determined when continuous four zeros will be detected.

Finally, the system can go to sleep mode if no gesture is detected for a certain period of time.

E. Normalization

Normalization is a basic statistical operation. It is used to scale heterogeneous sets of data, so that they could be compared relevantly [11]. That means, signal normalization brings signal values into manageable range by avoiding uneven values. Also normalization facilitates defining thresholds in different system.

After feature extraction a normalization is performed to avoid the uncertainty of ultrasonic intensity. This is done using following equation.

$$S_t = \sum_{i=1}^{m} P_t(i)$$

$$F_t(i) = \frac{P_t(i)}{S_t}$$

Where F_t is a feature vector at time t, P_t is feature primitive at time t, and m is the number of using properties. Using the above equations a strategy is defined for parameter selection for the threshold of θ_v and θ_d. This normalization approach can be performed to explore the relevance between the recognition accuracy and parameters. A strategy of optimization can be followed to find out the highest accuracy for the both parameters. This strategy can be started by initializing the parameters based on common practices which is similar to binary search algorithm.

F. Training System using Hidden Markov Model

The normalized feature primitives are used to compute the probability of each gesture type with Hidden Markov Model (HMM). The feature vector is used describe the features and as the input of the HMM [12]. There are three main steps for HMM: evaluation, decoding, and training. HMM recognizer chooses a model with the best likelihood, any pattern can not be guaranteed as similar to the reference gesture unless the likelihood is high enough. After training the system, a searching algorithm can be used to find out primitives in a faster way. Binary search algorithm is half interval search algorithm which finds the position of a target value within a sorted array. The time complexity of the algorithm is $O(\log_2 n)$ (where $n =$ number of primitives) which is faster than other searching algorithm. This algorithm is proper for this system because it works better on array or vector.

VI. Future Enhancements and Conclusion

This study presents a hand gesture recognition system that can detect in-air alphabets using Doppler effect based ultrasonic sound. This system can be able to differentiate similar gestures efficiently. After careful consideration, we can conclude that, future primitive 1 (P_1) becomes very small for in-air finger gesture and hence challenging to capture it and analyze it. Also, to capture subsequent gestures, only one algorithm is not enough, so improvised algorithms need to be studied in future. There is also possibility to extract more than 7 features and make system more efficient.

References

[1] Shashank Garg, Rohit Kumar Singh, and Ravi Raj Saxena, *Doppler Effect: UI Input Method Using Gestures for the Visually Impaired*, Texas Instruments India Educators' Conference, Bangalore, India, 2014.

[2] Sidhant Gupta, Dan Morris, Shwetak N Patel, and Desney Tan, *SoundWave: Using the Doppler Effect to Sense Gestures*, CHI '12 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Pages 1911-1914, Austin, Texas, USA May 05 - 10, 2012.

[3] Haojun Ai, Yifang Men, Liangliang Han, Zuchao Li, and Mengyun Liu, *High Precision Gesture Sensing via Quantitative Characterization of the Doppler Effect*, 23rd International Conference on Pattern Recognition (ICPR), Cancn Center, Cancn, Mxico, December 4-8, 2016.

[4] "Hidden Markov Models (HMM)". [Online] Available: *https://www.mathworks.com/help/stats/hidden-markov-models-hmm.html*

[5] Yuan-Yao Shih, Ai-Chun Pang, and Pi-Cheng Hsiu, *A DopplerEffect-Based Framework for Wi-Fi Signal Tracking in Search and Rescue Operations*, IEEE Transactions on Vehicular Technology, 2017.

[6] Qifan, Y., Hao, T., Xuebing, Z., Yin, L., and Sanfeng, Z., *Dolphin: Ultrasonic-based gesture recognition on smartphone platform*, Computational Science and Engineering (CSE), 2014 IEEE 17th International Conference on (pp. 1461-1468). IEEE.

[7] "Fast Fourier Transform (FFT)". [Online] Available: *https://www.edn.com/electronics-news/4383713/Windowing-Functions-Improve-FFT-Results-Part-I*

[8] "Hamming Window". [Online] Available: *https://en.wikipedia.org/wiki/Window_function#Hamming_window*

[9] "Signal Transformation Using FFT". [Online] Available: *https://i.stack.imgur.com/NziJO.png*

[10] "Signal Processing Using Hamming Window". [Online] Available: *http://zone.ni.com/images/reference/enXX/help/371361G-01/loc_fp_timesigwindowed.gif*

[11] "Normalization". [Online] Available: *http://www.edaboard.com/thread117303.html*

[12] "System training using HMM". [Online] Available: *https://www.hindawi.com/journals/mpe/2012/986134/*