

## Gestures Recognition from Sound Waves

Nguyen Dang Binh<sup>1,\*</sup>

<sup>1</sup>Hue University of Sciences, 77 Nguyen Hue street, Hue City, Vietnam

### Abstract

We propose a new method to recognize gestures from sound waves. The main contribution of this paper is to recognize gestures based on the analysis of short-time Fourier transforms (STFT) using the Doppler effect to sense gestures. To do this, we generate an inaudible tone, which gets frequency-shifted when it reflects off moving objects like the hand. We measure this shift with the microphone to infer various gestures. Experimenting method and evaluating results by using the hand gestures of many different people to browse applications such as website, document and images in the browser on the computers in the classroom and library environment for accurate results. In addition, we describe the phenomena and recognition algorithm, demonstrate a variety of gestures, and present an informal evaluation on the robustness of this approach across Laptop device and people.

**Keywords:** In-air gesture sensing; Doppler effect; interaction technique.

Received on 15 April 2016, accepted on 11 May 2016, published on 12 September 2016

Copyright © 2016 *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.12-9-2016.151679

### 1. Introduction

Gestures, a natural language of humans, provide an intuitive and effortless interface for communication with the computers. Gesture recognition is very useful for automation. Gesture is becoming a common means, as technology trends in the management and control interfaces. Gesture recognition based on the variation in sound waves frequency domain approach is a sound wave sensors utilize computer speakers and microphone. It has the advantage of not being affected by light, as the technical language recognition in images, video, voice. For example, the Microsoft's SoundWave and University of Washington [7] or Acoustic Doppler Sonar (ADS) [2]. Currently, this problem is continuing to develop applications, e.g., a gesture recognition system that leverages wireless signals to enable whole-home sensing and recognition of human gestures [6] and bringing gesture to all device [1]. Basically, the gesture recognition systems using more than one characteristic, using machine learning models Hidden markov models, Support vector machine to recognize gestures are really complex

and restrict much of the processing speed of the system. However, the achievements do not satisfy researcher's demands because of the complexity and instability of human gestures. We present a method of recognizing the gesture by dividing the energy levels of short-time Fourier transforms and using Doppler effect to recognize gestures. This method conducted discrete signal on the frequency domain with time into the signal frame of equal length and continue dividing the energy level briefly on each frame signal sequentially. Also, the analysis of energy levels in a short time each signal frame allows detection of noise and remove signal preprocessing steps for reliable results and somewhat reduce the fees charged math. Energy function in short-time been researched much in image recognition, speech recognition [5], [8]. We are inspired by the Doppler effect, the effect of Doppler is used to detect sound waves, voice, gestures [2], [3], [4] and [7]. There have been many technical studies using sound waves or the Doppler effect and motion gestures as sensors to identify gestures and movements in the field of digital signal processing and signal processing as follows : Measure the intensity of echoes received a microphone to detect the presence and

\*Corresponding author. Email:ndbinh@hueuni.edu.vn



**Figure 1.** An illustration of hand gesture recognition via sound waves used to control applications on laptop.

attention of people [10]. Use a continuous sound waves with frequencies of 2.4 GHz to customize the drive. They used the Doppler shifted signals reflect to infer human movement and body movement in an interactive space [11]. More recently, developed a device to recognize hand gestures in 3D space using low-cost ultrasonic transducer emits a 40-kHz tone. They put a transmitter and two receivers in a triangular pattern where gestures can be done and feel [2]. As these research shows the potential low cost due to the sound sensor. Whereas the identification system developed hardware requirements are too high, this is a significant barrier to widespread to other identification technologies. Identification system sound waves focused on a solution that works on a wide range of existing hardware to facilitate the development of mainstream applications and is much simpler.

The paper is organized as follows. In Section 2 we review the main ideas, which build the sound waves gestures recognition system. A method that combines the Doppler effect is based on the division of power levels short-time Fourier transforms to recognize gestures. Section 3 presents empirical evaluation results in sound waves gesture recognition. Finally, we summarize and conclude the paper in Section 4.

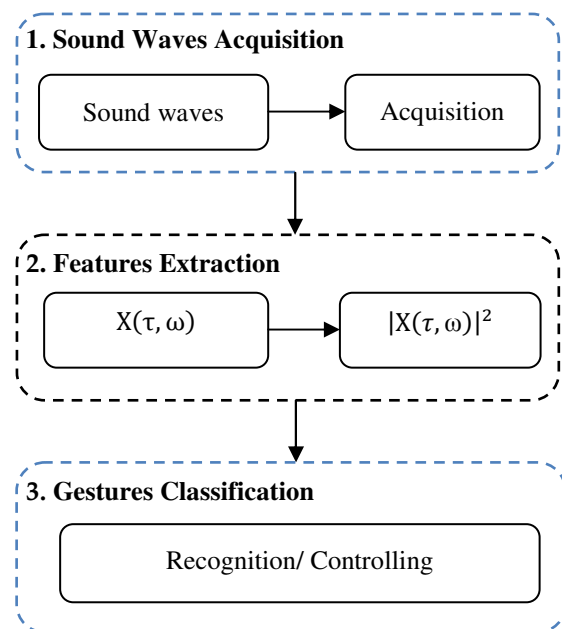
## 2. Sound waves gestures recognition system

### Sound waves

Sound waves are mechanical wave propagation in the physical environment (solid, liquid, gas). Sound waves are transmitted not in a vacuum. Moving objects producing sound are called sound source. The frequency of the sound waves in the oscillation frequency of the sound source. The nature of the sound waves is mechanical wave then it bears basic properties of mechanical waves. The nature and the characteristics of the mechanical waves carry different meanings when considering the problem of sound waves recognition, depending on the individual approach.

Using sound waves to sense motion that is the change in the frequency of sound waves in response to a moving

object, an effect called the Doppler effect. This frequency change is proportional to frequency resources and the speed at which the object moves. In this approach, the source (speaker) and receiver (microphone) are the hardware devices which has been integrated in mobile devices and laptops, so in the absence of any movement , no change in frequency. The system of sound waves gestures recognition and controls often include: signal sound waves acquisition, features extraction, finally classify, recognize and control interfaces. General diagram of gesture recognition system is shown on Figure. 2 sound waves uses existing speakers on commodity devices to generate tones between 18-22 kHz, which are inaudible. We then use the existing microphones on these same devices to pick up the reflected signal and estimate motion and gesture through the observed frequency shifts.



**Figure 2.** The diagram of the basic stages of gestures recognition system and controlling.

**Sound waves acquisition:** Use the built-in microphone on your computer to capture sound wave signal of gesture Doppler effect to change when it has changeable location of the two sources of waves.

**Features extraction:** Typical short time energy is extracted Windowing of Fourier transforms in a short time. Sound waves gestures and noise classification: positive and negative energy based on the distribution of energy levels in the short time-frequency domain to classify gestures and noise Doppler effect.

**Gesture recognition and controlling:** The user selected control functions, through the state classification system gestures to recognize gestures that execute commands from the controlling interface application browsers use.

## 2.1. Sound waves acquisition based on doppler effect

### Dropler effect

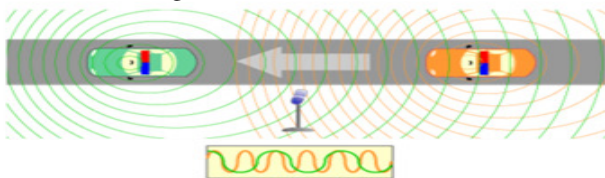
The principle is the Doppler frequency shift of sound waves obtained in microphone ( $f_r$ ) upon the relative shift of position with hands in the air compared with the computer speakers. The relationship between observed frequency  $f_r$  and emitted frequency  $f_t$  is given by

$$f_r = f_t \cdot \left( \frac{c + v}{c - v} \right)$$

where  $f_r$  is perceived frequency at microphone;  $f_t$  is original frequency from speaker;  $c$  is the velocity of sound waves in the air (speed of sound in air) and  $v$  is the velocity of hand in air; if hand is moving towards the source then positive (and negative in the other direction). Any motion nearby computer (about 1 meter depending on speed), integrated microphone in your computer will obtain the frequency shift of the reflected Doppler effect.

In our approach, the computer speakers act as a broadcast source, microphone as receiver. Speakers will continuously emit a signal whose frequency from 18kHz-22kHz constant (adjusted by the user), although the sound waves can operate outside our scope but we are consistent over the range from 18kHz -22kHz because it matches most hardware devices on the computer and we do not need higher frequencies to sense gestures in the air [2], [7], [10]. Then, use the built-in microphone on the device to capture and digitize signals through the recording (sampling frequency is 44,1kHz) signal observed Doppler principle. It is a combination of two separate effects induced by two sound wave sources (hand moves and speakers), frequency will increase as the hand observer moving closer to the computer and will decrease when the hand moves away.

We will easily see this phenomenon through the howling whistle of a train, or ambulance. When they are away from us, smaller the whistle is heard and when come nearer the vehicles the bigger and more dissonant is the whistle. Changing phenomenon is called the Doppler effect. The Doppler effect is not only true on the sound wave but also on electromagnetic waves (including microwave and light wave).



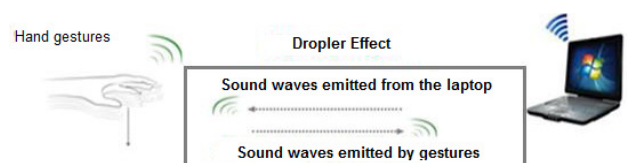
**Figure 3.** Description of the Doppler effect is obtained at the microphone for two waves approach each sources.

The Doppler effect is a physical effect, Christian Johann Doppler named, including the frequency and wavelength of the sound waves, electromagnetic waves or

the general wave is altered when the broadcast makes the relative motion with observe. For wave motion in an environment, such as sound waves, the wave source and the observer are moving relative to the environment. This time the Doppler effect is a combination of two separate effects caused by these two movements.

The frequency of the sound waves will increase when the source and the receiver move closer and it will decrease when the source and receiver moving apart. This is the idea that we rely upon to build sound waves identification system for the subject, by researching the Fourier transform signals that sound waves; then forming and building detection & identification algorithms based on two basic conditions- that are hand approaching and moving far from the computer through the consideration of the nature of the reflected waves changes according to the Doppler effect.

The sound waves identification system using the Doppler effect to hand moving gestures in front of laptop computer in this paper is described as follows: the computer's speakers will emit a constant frequency waves, whilst the gestures of hands moving before the computer will create a second sound wave. Micro Computer acquires the frequency change when the wave source and the computer from changing hands is collected and calculated from which detect direction of motion of the human hand. We can see more clearly through the description and picture schematically in Figure 4 intuitive.



**Figure 4.** Human-Computer interaction through hand gestures using the Dropler effect.

## 2.2. Features extraction

Signals are often analyzed in the frequency domain, this method is applied to all signal types, including continuous signal or discrete-time one; when a signal passes through a linear system unchangeable over time, the frequency spectrum of the output signal will equal the product of the frequency spectrum of the input signal and the impulse response of the system. Some methods of signal convert between the time domain and frequency domain is based on the Fourier transform, Laplace, Hibert, change Z, Wavelet ... In this part we introduce transformations Short-time Fourier transform (STFT), one of the improvements of the discrete Fourier transform traditional DFT (discrete Fourier transform) in signal processing. STFT, to my point of view, is appropriate and a particular advantage for the idea of recognition of sound waves in combination with the Doppler effect of this research. Also it is used to extract selected STFT characteristics in order

to better cope with the challenges that we mentioned above.

In the problem of sound waves recognition, we used Fourier transform short time (STFT) to conduct a review of the energy change in a short time of the sound wave to each gesture state under the effects Doppler. From that the recognition algorithm is built based on the idea of detecting energy reflected wave Doppler. Why do we chose transform STFT? What are the purposes, advantages and disadvantages of STFT than other methods used to extract selected characteristic for the sound wave identification problem within this paper? The problems are to be presented more clearly in the following sections.

In fact, the gesture moves continuously in a certain period of time and often not be consistent between the times of performance (in speed, direction and time travel) as well as depending on the user (Campaign airborne sound transmission speed of 343 m / sec, speed hand gestures about 0.25m / s - 4m / sec) [10]. Another factor, the acquisition of sound waves from the moving gesture is rather much affected by noise (interference) which is available or random existence in their surroundings (as a hardware device or operating emitted program with active sound created). Therefore, the signal processing and extracting features selected to recognize which is a gesture or an environmental noise in the frequency domain is really complex.

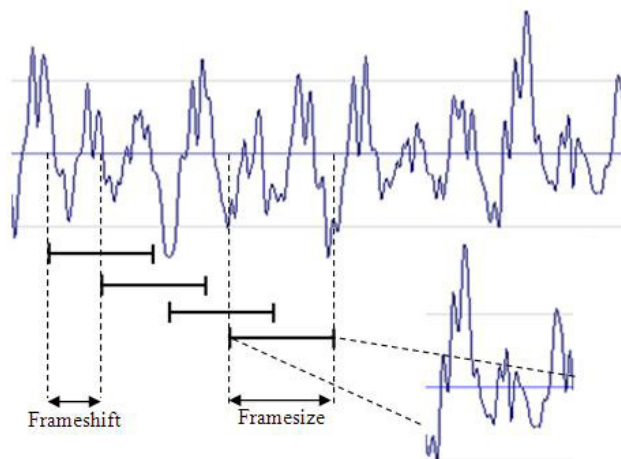
Following the research content of this paper, with an aim to improving the accuracy of the identification of the sound waves emitted from hand gestures, instead of identifying each individual gesture is very difficult and very complicated real, every gesture is analyzed into the energy content of the signal shortly after the sound waves and transform STFT system identifies each specific gesture through a series of transformations of the energy content A short time using the Doppler principle. This is the idea of this paper conducted intensive research to select specific criticism and build a suitable identification algorithm for ideas mentioned in the research scope of this paper.

As such, a gesture, from start to finish can be considered as the transformation of energy according to the Doppler effect from the source of computer speakers. For that reason, characteristic of the energy content should be extracted on each short time frame instead of a long signal segments from the beginning to the end of a gesture. This is the advantage of STFT transformation in selecting specific criticism for short time energy content in comparison to other Fourier transformations.

Another reason for the necessity of the digitizing the continuous signal passage of input sound waves into signal samples with such small amount of time because the sound signal emitted gestures changes rapidly so the properties as amplitude, frequency will not stabilize. When the audio signal to be considered in small sections, then in each section, which can be considered as a stable signal, characteristic of the signal energy content can be

considered constant over time. This treatment process is also known as the window handle (Windowing).

**Windows of signal:** Signal in the short period of time can be seen relatively stable and unchanged over time. For a signal of gesture, this can be done by windowing of a signal  $x(n)$  into an unbroken chain of sequential  $x(t)$  window,  $t = 1, 2, \dots T$  call is the signal frame.



**Figure 5.** Describe the process of dividing the discrete signals.

To do this, we use a window along the audio signal and cut out signal segments within that window. A window is defined by these parameters:

- Frame size: the width of the window, as well as the magnitude of the signal frame will be cut out.
- Frame shift: moves the window- the length of which the window will slide to cut out the next frame.
- Each frame will then be multiplied by a coefficient, the value of this factor depends on the type of window.

$$y[n]=w[n]*x[n]$$

Where  $x(n)$  is the value of the  $n$ th sample,  $y(n)$  is the value of the  $n$ th sample after when multiplied by the coefficient,  $w(n)$  is the coefficient for the  $n$ -th sample in that frame. The simplest type of window is the rectangular window, the value of the coefficient  $w(n)$  is given by the following formula:

$$w(n) = \begin{cases} 1 & \text{with } 0 \leq n \leq N, N \text{ is number of sample in a frame} \\ 0 & \text{otherwise} \end{cases}$$

The selection window Hamming [7], for discrete signals into the signal frame (with sample points in 2048) we considered suitable for the energy spectrum will be concentrated in the middle of the frame signal:

$$w(n) = \begin{cases} 0.54 - 0.46\cos\left(\frac{2\pi n}{N}\right) & \text{with } 0 \leq n \leq N \\ 0 & \text{otherwise} \end{cases}$$

Where  $n$  is the number of samples on a window ( $n$  is an even number),  $N$  is the number of signal frames.

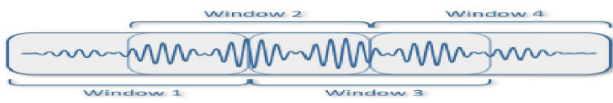


Figure 5. Signal window

**Featured short time energy**

Energy shortly is determined by calculating the average of the total area of the sample (sample) single in each frame. With a window ends at the  $m^{th}$  sample, short time energy function  $E(m)$ :

$$E(m) = \sum_{-\infty}^{\infty} [x(n)w(m - n)]^2$$

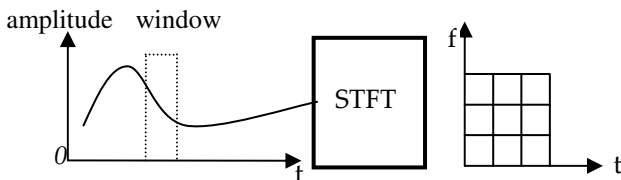


Figure 6. Description of short time Fourier transform

Frequency signals emitted from the hand gestures or environmental noise signal with non-stop (non-stationary signals) means a periodic signal over time. Short-time energy analysis after each frame signal is unstable, or undetailed process of moving gesture or noise already existing or appearing at random. So we continue to separate into discrete signal on each frame (framing) using short-time Fourier transforms.

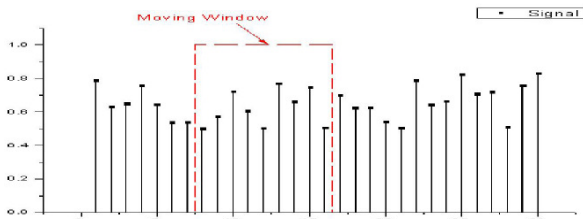


Figure 7. Shift window

Short time Fourier transform (STFT) is the division of time into overlapping blocks or non-overlapping equal length and apply fast Fourier transform (FFT) for each block in sequence.

$$STFT\{x(t)\}(\tau, \omega) \equiv X(\tau, \omega) = \sum_{-\infty}^{\infty} x(t)\omega(t - \tau)e^{-j\omega t}$$

with the signal  $x(t)$  and the window  $w(n)$ . In this case,  $m$  is discrete and  $\omega$  is continuous, but in most typical

applications STFT is performed on a computer using the FFT, so both variables are discrete.

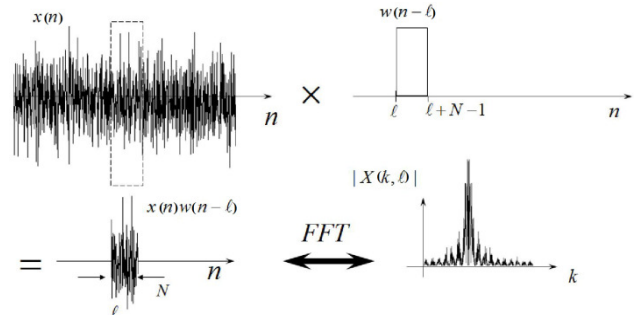


Figure 8. The diagram transformed into FFT signal

From discrete STFT transformation, the development of the intensity with time is called spectrogram. Spectrogram can be calculated using the formula:

$$spectrogram\{x(t)\}(\tau, \omega) \equiv |X(\tau, \omega)|^2 \equiv |STFT(\tau, \omega)|^2$$

where,  $x(t)$  is the signal to change at time  $t$  ( $1 \leq t \leq \tau$ ),  $X(\tau, \omega)$  represents the phase and amplitude of the signal in time and frequency. Spectral energy distribution function is the result of the transformation process STFT, featured short time energy is calculated after each step sliding window (windowing) of STFT:

$$Energy = \sum_{t_i}^{t_{i+1}} x^2(t) = x^2(t_i) + x^2(t_{i+1})$$

in which energy is featured in short time,  $x^2(t_i)$  is the energy spectrum signals in frame  $t_i$  ( $1 \leq i \leq N$ ).

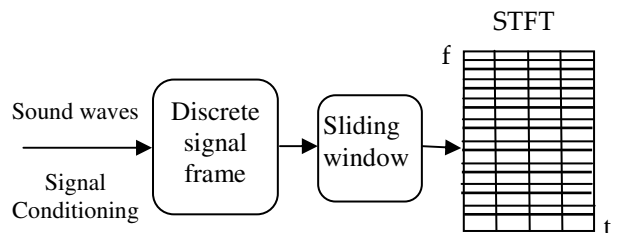
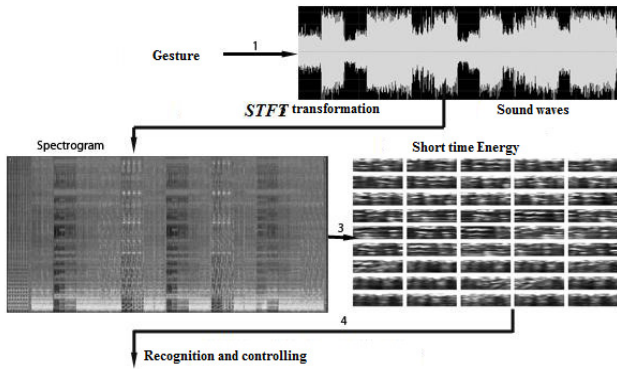


Figure 9. Describe the process selected feature extraction.

After digitizing the signal, we started extract features of a sound wave moving hand gesture at a computer through STFT transformations to calculate the energy in a short time frame of each state of a gesture . Work on all handling in the short time analysis (Short-Time Analysis) (Sau khi đã số hóa tín hiệu, chúng ta bắt đầu trích lấy trung sóng âm của một cử chỉ tay di chuyển trước máy tính thông qua phép biến đổi STFT để tính năng lượng trong thời gian ngắn của từng khung trạng thái của một cử chỉ. Công việc trên đều xử lý trong phép phân tích thời gian ngắn (Short-time Analysis))



**Figure 10.** Describe the process of features extraction of sound waves

**Algorithm 1: Features extraction**

**Input:** Processed signal recording sampling (data).

**Output:** Short time energy features

**Methods:**

1. FFT transform input signals  
 $f = \text{FFT}(\text{data}) ;$
2. STFT transformation from the domain to the frequency domain and time.

$$\text{stft} = \text{STFT}(f) ;$$

3. Browsing through each frame to extract features signal.

For  $t=0$  to  $N-1$  do

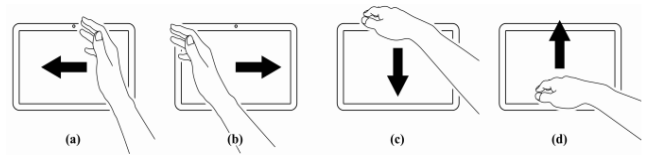
$$E = x_t^2 + x_{t+1}^2 ;$$

3. **End.**

The analysis of the energy content in a short time signal of the sound waves that is emitted from the simultaneous gestures combined with the advantages of Hamming window shows that the use of shot-time Fourier transform in extracting sound waves signals characteristic is an advantage and entirely consistent with this approach.

**2.3. Gestures classification**

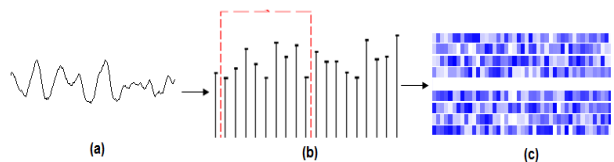
We limit the waving gesture (palms facing the computer, limit the distance in 1m depending on the speed of movement of the hand than the computer and audio hardware of the computer) move from top-down "up to down", from right to left "right to left" is called the "state" move closer or move away from the computer including gestures moving from left to right "left to right", from bottom to top "down to up". The moving gesture is shown clearly in Figure 11.



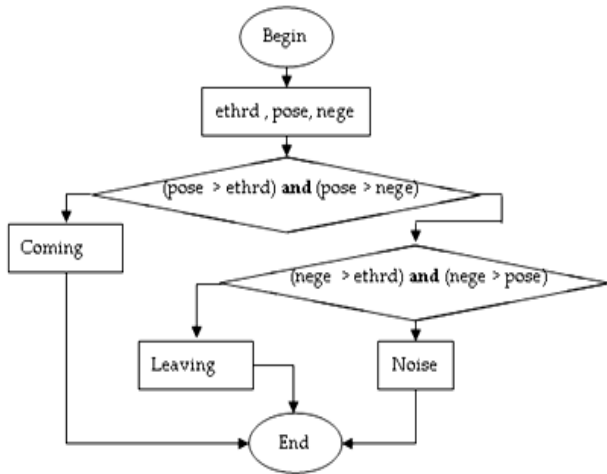
**Figure 11.** The hand gestures. (a) and (b) is approaching "Coming". (b) and (d) is carried away "Leaving".

The idea of gestures classification is based on the transformation of the energy of positive or negative on the frequency domain (oy axis is frequency-domain, time-domain is the ox axis) through the distribution of energy levels Shortly after wave change on every frame STFT signal Doppler effect principle. When there is not any movement, short time energy to wave signal emitted from the speakers (whose frequency is  $cf$ ) called threshold ( $ethrd$ ) will remain unchanged. When a gesture is moving towards or away from it a short time energy component increased or decreased distribution around the threshold of the Doppler principle. We call the energy increase in the short period of time is positive energy (pose) similar to the energy reduction (nege) is negative energy (this is why we calculated the energy function for the amount of the short time period after the sliding window step change STFT).

The analysis of the energy positive or negative in scope (range) of the near threshold energy level ( $cf - oy \leq \text{range} \leq cf + \text{range}$ ) is not specific to the "state" moving gesture hand because it changes so fast around the threshold. In this case, may be due to the noise of the hardware or the program that uses sound triggers. We are interested in the range ( $oy < cf - \text{range}$  and  $oy > cf + \text{range}$ ) to change the energy section briefly relatively clear than the threshold to classify gestures into advanced classes closer "Coming" or move away from the "Leaving". Similarly, in a range (exrange) larger ( $oy < cf - \text{exrange}$  and  $oy > cf + \text{exrange}$ ) existence of the energy available to a positive or negative than the threshold, then it is sure to be noise "Noise "which is not the state of motion of hand gestures, because it is not fast enough to spread the energy distribution around the threshold.



**Figure 12.** Distribution of share positive/negative energy. (a) the signal of gesture has been digitized. (b) The modified signal STFT using Hamming window. (c) The distribution of the energy of negative/positive energy around the threshold when motions are shown colored boxes with value increases from white to dark blue.



**Figure 13.** Flowchart gestures classifier algorithm and noise.

**Algorithm 2: Classification algorithm for gesture and noise**

**Input:** Feature vectors are extracted from short time energy function of the sound waves.

**Output:** one of 5 Classes gesture:

"Coming" = {Go down; shift into left},

"Leaving" = {Move up; shift to right},

"Noise" = {other cases}.

**Methods:**

1. Browse the energy levels in the frequency domain (oy <cf - range and oy> cf + range) to calculate the input parameters (ethrd, pose, nege) by the following formula:

$$\text{Energy} = \sum_{t_i}^{t_{i+1}} x^2(t) = x^2(t_i) + x^2(t_{i+1})$$

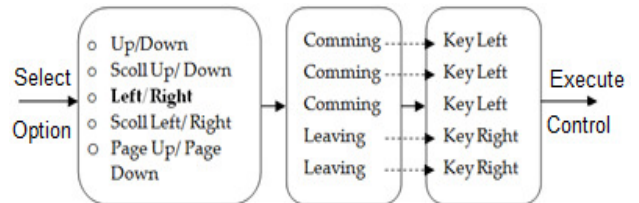
2. Comparison of the parameters are calculated in step 1 with the threshold to classify gestures to the state of motion "Coming", "Leaving", or "Noise".

3. **End.**

**2.4. Gesture recognition and control application interface**

We built a system to recognize and control the selection gesture function using virtual keys to browse applications including: Browsing the document horizontally, browsing the document vertically, transforming a document page, scrolling up or scrolling down the document page in the way that user selects options virtual keys in Figure 9.

When appearing a moving gesture of waving in front of the computer, the system will classify it into class gesture "Coming" (approached), including two gestures: "shift to the left" and "go down" (the celebration only (a) and (c) in Figure 6) or belonging to the class "Leaving" (moved away) include gesture "translated to" and "go down" (the gesture (b) and (d) in Figure 9).



**Figure 14.** Selecting functional diagram of the control system.

Through the selection of options to browse virtual key document by the gesture introduced, the system will accurately recognize either gesture in each class from which it executes control commands instead of direct interaction with the computer keyboard or mouse.

**3. Experiments and Results**

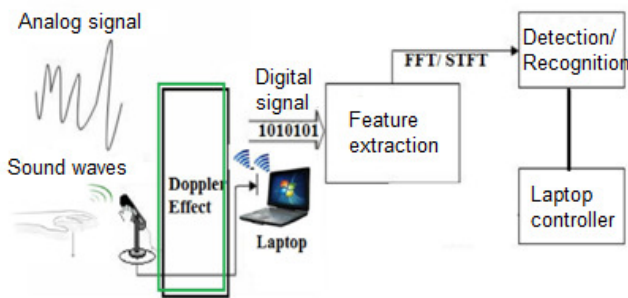
By using short-time Fourier transformation to extract short time energy features presented in Section 2.2; we focus on the construction of the experimental program to identify the sound waves from hand gestures in laptop computer controlling. With these applications, we review the results, evaluate the effectiveness and applicability of the method that was developed. Since then we point out the limitations of the methodology and the innovative direction to develop better applications.

**3.1. Sound waves gesture recognition system**

The aim to build the application as a motion recognition system of the laptop based on sounder speaker and microphone. That's touch less sensor can recognize movement of the hand. Users can use gestures to control programs like Flip Slide PPT, moving in the photo browser, browse PDF documents, Word, Excel, or surf websites ... This application is similar to sound waves Microsoft in researching how to handle it completely. The system is operating on sound waves at 18kHz-22kHz frequency and can be adjusted in the user interface. Some techniques are built in the programming to make the results reliable recognition, including short time Fourier transforms, Doppler effect, gesture recognition algorithm, the human voice and recognize ambient noise, as well as enhanced recognition algorithms strong identity.

### 3.2. Environmental construction and operation of the application

Application is installed on Microsoft Visual C ++ environment, so it may be fine on computers using Windows operating systems understand. Application allows to run on any laptop computers have integrated at least a microphone and speakers. The application uses sound waves to the use of the day or night absolutely not affected by other applications that use the camera identification. The application uses the built-in microphone and speakers on the laptop computer so the gap between users and computers in applications is only 0.7m, the closer the distance, the higher the accuracy.



**Figure 15.** Our sound waves gesture recognition system for controlling applications on Laptop.

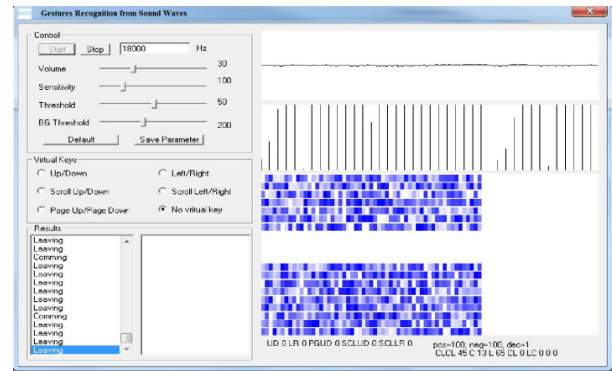
#### The outstanding advantage of the application

This application brings many advantages compared to other gesture identification techniques; uses probably the cheapest touchless sensor at present, and is an effective solution for notebook computers and other embedded systems. The computational complexity is relatively low compared to other control applications using popular image and video available today. Control the user's computer becomes easy and fun by waving. We can use it to browse documents and turn pages, images ... Thus the computer control becomes simpler to everyone. Compared to speech recognition with some language barriers, these obstacles are seen common in the international platform. Also compared with face or gesture recognition based on camera, it is much easier to learn and use for novice computer users. Human voice and environment noise do not affect the program. The music player can run simultaneously on the same computer when using this application program.

#### The composition and operation of the application program

The interface of the program is divided into four main parts:

- **Group calibration parameters:** On the left side of the interface we can press the "Start" button to start the program. "Stop" to stop the play echoes



**Figure 16.** Our application program interface. This program provides a simple user interface.

from loudspeakers and stop recording from the microphone. There are a number of sliders available in the interface, we have made the default parameters when starting the application, these parameters are suitable for different types of computer. we chose the parameters just fit my computer. Frequency signal transmission provides echoes of speakers which can be adjusted. The default is 18000 Hz and can be changed from 18,000 Hz Hz-22,000. Lower frequency does not affect the findings; however, we will hear the sound coming from the speaker unpleasant, high-frequency sound will be filtered out by the computer sound card. The meaning of the above parameters as follows:

- Volume:** The volume control of the transmission (the volume of the echo emitted from the speaker).
- Sensitivity:** Sensitivity Control Slider Gain to amplify the received waveform.
- Threshold:** Provides an energy threshold to filter out unwanted noise.
- BG Threshold:** Used to provide an energy threshold for filtering platform to avoid the human voice and other audio sources.

If we want to start it again, please close and restart the program because of some errors to be fixed but we have not yet solved in this application.

- **The graph displays waveforms and spectrum:** There are several graphs on the right side of the interface, top chart is the time-domain waveform showing the signal frame recorded from microphones. It varies according to the microphone recording signals from the surrounding environment. The middle graph is a graph of the sound wave spectrum after Fourier transform (FFT) from the time domain to the frequency domain, which shows a part of the spectral components near the resonance frequencies emitted by the speaker. Bottom graph is a graph of short time Fourier transform (STFT) from time to time and spectrum.
- **Virtual keys control:** There is a choice of virtual keys which we can use gestures to control waving



applications rather than ordinary interact directly via the navigation keys on the keyboard or move the mouse as Up / Down (U / D); Left / Right (L / R); Scroll Up / Down, Scroll Left / Right, Pages Up / Down.

- **Results:** Two states approached (coming) and moving away (Leaving) by hand are displayed simultaneously on the last frame when making the movement of the hand in front of computer status. The remaining frame display control functions used our virtual keys option under virtual keys control group.

### Activity program

Install the first to run the application to make the software run on your computer and get a good result, a number of set-up steps are necessary.

- (v) Control Panel / Sound / Recording: double click the microphone use, select the Levels tab, set all to the maximum level.
- (vi) Some sound cards have an advanced microphone marked tab, deselect this mark, because it will filter out the high frequency signal.
- (vii) Adjust the volume of the volume control to a value large enough.

The main working status is described as follows: When there is no transfer and no noise, no graph spectrum surrounding components, which means most of the energy is roughly equal to the frequency found by the echo from the microphones out. If we waved, then there is a change of frequency spectrum in all three charts. The third graph (STFT) has a white background, if we wave some waving blue frames are shown in the graph. The list box of status decision in the left will show you exactly our motion at the same time. Note: For the first time, using the program on the laptop computer can meet a few errors leading to inaccurate results. The most common is the graph STFT bearing several random blue dots and line graphs spectrum with more click random flashing; in this case, please check the machine's microphone if it was set as a maximum or not. Then, change the parameters to find out a set of parameters that best suits your type of computer used.

### 3.3. Experimental results

In general, applications can run stably on most laptops, obtained very high results in the identification of two basic states of the hand moving in front of computer. From that we control the functions of the navigation keys. To achieve good results, it takes a reasonable time to adjust the initial configuration parameters to suit the environment or different kinds of laptops. We performed tests to evaluate sound waves working through people and laptops and to estimate accuracy and robustness in

different environments (in the library and in the classroom).

We conduct empirical methods in quiet place (environment), considered easy to recognize gestures such as a library or a calm place. In these locations, usually at survival signals emitted sound waves or less random noise existed to affect the classification process and gestures recognition. At the same time, we also experimentally do the same in noisy places such as classroom, locations are often available randomly or other source of sound waves that are not only caused by. Through noisy places or quiet ones, we evaluate the effectiveness of methods for environmental use. Also, in both quiet environment (library) and noisy (classrooms) for many different users to assess the impact (in terms of speed, direction of movement) of the election different only for the method. To review the stability and effectiveness of the methods implemented, we use a number of different types of computers. The computers operate in different modes but (allowing applications to run simultaneously in multiple time difference) to control a number of applications such as web surfing, browse PDF documents, PPT slides, or browse Photos in the browser application using four hand gestures which were introduced instead of using direct or mouse navigation keys on the keyboard. By aggregating data in the evaluation method mentioned above, we found that the method can recognize gestures resulting noise and very reliable.

We tested gestures recognition from sound waves on 3 different laptops: a Dell inspiron 3420 Computer (Intel (R) Core (TM) i3-2328M CPU @2.20GHz; a Intel(R) Core (TM), i7-2460M CPU @2.80Ghz; and an IBM Thinkpad T43. Environment normal activities like working in an office, a room school, Library... distance movement of arms smaller than 0.5m computers. The initial configuration parameters are set by default when the program starts and do not add any calibration parameters. This also included a sound card and microphone. To ensure that gestures recognition from sound waves across people, we tested it with 3 individuals. We asked them to control four gestures: (1) Scrolling a webpage or page screen of an application; using single hand motion up/down corresponding to the state Coming and (2) Shift to the left/Shift to the right a screen of an opening application corresponding to the state Leaving. The experimental results are averaged in each gesture shown in Table 1.

Table 1. The average percentage of four gesture recognition.

Environment experiment	Results percentage ratio of hand gestures recognition for 100 times per gesture.			
	Move up	Go down	Shift to the left	Shift to the right
In the classroom	80,7%	74,6%	82,4%	73,5%
In the library	89,3%	77,7%	88,1%	79,9%

Through our testing evaluation on 2 gestures detection we have got very good application results which were almost perfect “Hand shift to the left” and “Hand move up” because this time the computer detects a change of energy very well. The remaining 2 cases (Hand go down and Shift to the right) by programming techniques divide energy level was not good, what we call the noise, does not achieve high results. We understand fully this matter and will overcome in the near future. Through the development, implementation and evaluation methods, we found that the analysis of the short-time energy levels through the positive / negative energy of reflects well the movement of the hand gestures on the frequency domain with time. Besides that, it helps to eliminate the impact of environmental noise interference. The method focuses on a specific analysis of energy having no combination with many other features, it kind of makes findings is somewhat restricted compared with other methods. In addition, the energy spectrum analysis is only considered the basic hand gestures through two states approaching and moving away. The more complex gestures have not been considered in the energy spectral features shortly.

#### 4. Conclusion and future work

We describe a novel gestures recognition from sound waves system. We showed the robustness and accuracy of the approach through different Laptops, users and environment. The basic idea is to recognize gestures based on the analysis of short-time Fourier transforms (STFT). Our method that combines the Doppler effect and the division of power levels short-time Fourier transforms on the frequency domain to recognize gestures. The method is based on a single feature to recognize gestures and noise removal of ambient devices that makes cost in terms of computation and processing improved somewhat. Also, the method shows the simplicity and easy application deployment for leverage spacious sound hardware often built on the device from which the cost price is minimized. Next time, we will research and develop towards detection and gestures control completely automatic no longer depending on the selection of the control functions of the user. Using machine learning models (HMM, SVM) based on the energy characteristic short time is to train and discovered many more gestures. The use of filters and sound waves combining this method with other methods of detecting gestures from images through the camera in a system to improve the accuracy and detecting multiple complex gestures in future work.

#### References

- [1] Bryce, K., Vamsi, T., and Shym, G. (2014) Bringing Gesture Recognition To All Devices. *NSDI*.
- [2] Kalgaonkar, K., Raj, B. (2008) Ultrasonic Doppler Sensor For Speaker Recognition, Acoustics, Speech and Signal Processing, *ICASSP*.
- [3] Kalgaonkar, K., Raj, B. (2007) *Ultrasonic Doppler Sensor for Voice Activity Detection*, Signal Processing Letters, IEEE.
- [4] Kalgaonkar, K., Raj, B. (2009) One-handed gesture recognition using ultrasonic Doppler sonar. In *Proc. IEEE Acoustics, Speech and Signal Processing*.
- [5] Liu, D., Tian, J., Yang, B., Sun, J. (2012) *Time-Frequency Analysis Based Motor Fault Detection Using Deconvolutive STFT Spectrogram*. Journal of Convergence Information Technology (JCIT).
- [6] Pu, Q., Gupta, S., Gollakota, S., Patel, S. (2013) Whole-home gesture recognition using wireless signals. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*.
- [7] Pu, Q., Gupta, S., Gollakota, S., Patel, S. Soundwave: Using the doppler effect to sense gestures. In *Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems* (2012)
- [8] Sharat, C., Venu G., Alexander N. C. (2005) Fingerprint Image Enhancement Using STFT Analysis. In *Proceedings of the Third international conference on Pattern Recognition and Image Analysis*.
- [9] Tarzia, S. P., Dick, R. P., Dinda, P. A., Memik, G. (2009) Sonar-based measurement of user presence and attention. In *Proceedings of ACM UbiComp*.
- [10] Stephen P. T., Robert P. D., Peter A. D., Gokhan M. (2009) Sonar-based measurement of user presence and attention. In *Proceedings of the 11th international conference on Ubiquitous computing*, pp.89-92.
- [11] Joseph P., Craig A., Kai-yuh H., Matthew R. (1997) The magic carpet: physical sensing for immersive environments. In *Proc. ACM CHI 1997*, pp.277-278.