Handwriting Input Device Using Scratch Sound

メタデータ	言語: eng
	出版者:
	公開日: 2017-10-05
	キーワード (Ja):
	キーワード (En):
	作成者:
	メールアドレス:
	所属:
URL	http://hdl.handle.net/2297/42350

HANDWRITING INPUT DEVICE USING SCRATCH SOUND

摩擦音を用いた手書き入力デバイス

LEONG YENG WENG

平成27年01月6日 提出

博士論文

HANDWRITING INPUT DEVICE USING SCRATCH SOUND

摩擦音を用いた手書き入力デバイス

金沢大学大学院 自然科学研究科

システム創成科学専攻

機能創成システム講座

学籍番号（Student ID Number）：1223122011

氏名（Name）：LEONG YENG WENG

主任指導教員氏名（Chief Adviser）：PROF HIROAKI SEKI

Date of Submission : 2015.01

Table of Contents

List of figures

List of tables

1. INTRODUCTION

Current trend of mobile device adoption and its related services have been steadily increasing. This is evident with world average penetration of mobile-cellular subscriptions in 2013 reaching 96% [1], [2]. Meanwhile, global mobile-broadband penetration is expected to reach 32% by the end of 2014. World population currently stands at 7 billion persons, with smart devices users stand at about 2.2 billion persons and is expected to grow further at a faster rate engulfing the world population in the process[1]. This growth is due to several persistently evolving factors which are interdependent of each other, such as policies, technological advancements and market demands.

A country's policy has a large impact on its national Gross Domestic Product (GDP), the measure of a countries wealth while GDP per-capita is the GDP of a country divided by the number of people in the country. *The Worldbank* reports that a 10% increase in infrastructure investment contributes to a 1% GDP growth [3][4][5]. Infrastructure encompasses fields such as transport, water, energy, Information and Communication Technology (ICT). Studies showing that focused investments into ICT increases GDP per-capita, assuming the other functional basic infrastructures exists which has the international community investing heavily in this field creating a spillover effect to peripheral industries [6], [7], [8], [9].

Meanwhile, rapid advancement technologies, specifically in semiconductors has contributed to the miniaturization of ever increasing processing powers of processors for lower costs due to advanced methods and economies of scale[3]. Initial developments within the last four decades complied with Moore's law, whereby it was predicted that integrated circuits would increase in density every 1-2 years [12][13][14]. This in turn reduced the energy consumption per transistor [12]. Increased miniaturization coupled with the increase of clock speeds allowed more instructions to be completed in every second per unit space. Electronic products such as computers that used to take up large spaces, energy and mass were scaled down, allowing it to be marketed to individuals instead of being limited to the domains of corporations. Input methods then were

previously based on cumbersome punched paper tapes and cards while outputs from computers were just as difficult to work with. These factor required multiple operators who were sufficiently skilled to operate them [15][16].

The 1970's to 1980's saw the successful miniaturization of electronics and development of user friendlier human computer interfaces (HCI) such as keyboards, mouse and graphical user interfaces resulted in the creation of personal computers. Coupled with the software which added value to the hardware of personal computers, the computer industry as a whole witnessed an economic boom well into the present time seeing its utilization and contributions in various fields such as medical, education, military, entertainment etc. Interestingly though personal computers did not capture the Japanese market as the conditions were different such as the keyboards were then customized for roman alphabet inputs, houses had less place to fit a bulky desktop and people spent most of their times commuting to-and-fro in trains[17]. This changed when the service provider *DoCoMo* created the *i-Mode* which provided a locally customized internet based service on a mobile device. This was the turning point where within two years of its inception, it managed to boast a 22 million user portfolio. Other parts of the world tried emulating *i-Mode*, but failed miserably, due to several reasons, one being the packet switching which was used by *DoCoMo* which made it faster and the HCI, where content was arranged according to the user's preference in Japan.

As of the last two decades, exponential progress of various technologies allowed further miniaturization, increased processing power per watt and higher energy density batteries. This in turn allowed the blurring of boundaries between the functions of desktops and mobile devices. This meant more computationally intensive operations which used to be restricted in the domain of desktop computers such as gaming, social networking, internet banking, video, songs, were readily available to the mobile user [18], [19], [20]. This has fueled the sales demand for both smart devices and internet/cellular subscriptions around the world [1]. In 2007, it was the *iPhone* which took the world by storm because it managed to work out the HCI issues which its competitors did not. Customers were demanding better quality pictures (higher resolution, larger screen sizes), larger keys (easier methods of data input) and better range

of inputs (various shapes, alphabets etc) and *iPhone* fulfilled those requirements. By the end of 2010, 73.5 million *iPhones* were sold. The similarity between the *i-Mode* and *iPhone* is that both made HCI a priority in their products, hence making them both an instant success.

At present, the current wave of development is focused on cloud computing enabling infrastructures which allows users to be more mobile by relocating the workload and storage space required away from the user[21], [22]. Current mobile-broadband penetration will reach 32% by 2014 which is 5 times more than it was in 2009. In addition to that, trends in electronic production has reported most electronics produced at present are for automobiles and mobile devices with forecasted USD325Billion in 2015, indicating growth in the mobile industry is poised for greater heights [23]. Aside from such conventional investments, industry giants such as *Google*, *Facebook* have invested heavily into expanding the network connectivity to the other two-thirds of the world population who are now not connected to the internet[17] via the *Loon project* and *Connectivity Lab*. Coupled with high user demands, technological ability and reliable infrastructure, mobile devices are set to be the next crucial device which would encompass our daily lives fully.

To our, surprise though, statistics show that mobile devices are mainly used for simple tasks such as browsing, reading emails, watching videos and reading e-books despite the ability of the mobile devices to do much more in terms of processing ability and connectivity [11]. Closer inspection upon this phenomenon uncovers studies which show that the decreasing size of mobile device's interacting surface (input/output) which increases mobility, results in drop user efficiency and satisfaction [24], [25]. This is further aggravated as interactions with the mobile device becomes a challenge as many applications are competing for an ever shrinking input real estate while increasing its size would reduce its mobility. In short, there exists an inverse relationship between the ease-of-use with mobility. Industries today address this problem by working within the optimal point between the two opposing parameters but face issues such as occlusion of screen by the finger and contact bounce on the screen [26]. Limitations imposed by this optimal balance between size and ease-of-use if addressed could unleash the full potential

of mobile devices as it did for *i-Mode* and *iPhone*. This is seen as an opportunity by the authors to solve the problem from the mobile device's input point of view by eliminating this said relationship instead of trying to find an optimal point between the two opposing parameters.

Hence the objectives of the research are defined as such,

Research Objectives: Create a system which encompasses both methods and devices which breaks the inverse relationship between mobility and user-friendliness of input of mobile devices and yet allowing for wide input ranges by capturing acoustic signals released from tribological interactions during natural human finger tracing gestures on various surfaces via the usage of microphones.

This theses is arranged where chapter 2 is the literature review to develop general concept to be built to be further developed via experiments, followed by chapter 3 where the individual technologies within the system is explained, this is then followed by the explanation of the proposed methods by the author in chapter 4, which ends in the building and testing of the first prototype. Chapter 5 solves the issues seen in chapter 4 by first simulating various possible causes of errors and then solving and verifying its effectiveness in real-life. Chapter 6 increases the accuracy and speed of the localizing algorithm by effectively merging two known algorithms which were verified through simulations and real-life(offline and online) experiments. Chapter 7 introduces a new algorithm which utilizes spatial cues that allows the device to adapt to different acoustic backgrounds to make it even more versatile, this claim was simulated and tested offline. Chapter 8 was an upgrade of chapter 7 where it included both spatial and frequency cues to segregate the background noise from the TES and was verified in real-time in various acoustic conditions and surfaces. In addition to the advanced algorithms included, the hardware was greatly miniaturized where it could be worn on the hand or placed on various surfaces easily. Various letters, shapes captured by the device is also shown in this chapter. Last but not least, chapter 9 contains the conclusions and future works of this research.

2. LITERATURE REVIEW

Miniaturization and network infrastructure has allowed for the 'internet of things' to emerge driving the 'big data' research. The success of these concepts is determined by the willingness of the users to carry around an array of sensors in which they can interact with constantly.

2.1. WHAT IS HCI

Human Computer Interface actually covers a very broad area which encompasses multiple disciplines such as Computer Human Interaction (CHI) which refers to computer science while Ergonomics (E) relates to human related structure [27]. The Human is an evolving system by itself while the computer too is an evolving separate system. The interfacing of these two systems are akin to shooting a moving target while the shooter was also moving at a different velocity from the target. It is therefore obvious that much effort and ingenuity would be required to interface the two systems.

2.2. EVOLUTION OF HCI

In 1911, Fredrick Taylor a mechanical engineer sought to increase productivity using science and mathematics. He sought to understand the two systems, the human and the machine so as to make the production process repeatable processes which could generate higher yield. The world wars saw these ideas further refined.

Early designs of computers such as the *Eniac* was ten feet tall, encompassed an area of 1k square feet. Programs were loaded into it via switches, dials and cable connections. Outputs were in the form of punched cards. In addition to that, since it ran on unreliable vacuum tubes, the computer required constant supply of replacement tubes. Running a computer required multiple persons. Hence driving the research in HCI to make the interactions with the computer as simple as possible so that the user can be freed to do other more productive things[16], [28].

In the 1950's two major push within HCI, one was on the technological region which made strides in improving input and output devices such as CRT displays, magnetic tapes while the second push was from the ergonomics field which looked into training of the staffs to increase efficiency and such[28].

In 1960, J.C.R. Licklider identified that despite the availability of many man-machine-systems, there were very few good examples of man machine symbiosis. This led to the identification of some capabilities which required more development to enhance the uses of computer to humans such as electronic input-output surface for the display of various types of information to the user, interactivity, real-time operations and large storage capabilities. Ivan Sutherland's PhD involved the creation of the *Sketchpad* making the computer then to be the first complete GUI which ran in real-time. The computer used was a TX-2 which normally ran in batch mode. Douglas Engelbart created and demonstrated the input devices such as mouse and keyboards working in unison with a multidisplay environment. This was created to police the Arpa net which was to be used to link the few computers available in that country at that time[27].

The progress of the computer science field and the ergonomics did not always go hand in hand, instead they grew in spurts when the other was saturated. As of today, many devices such as desktops, the internet, laptops, mobile devices have been built based upon the guiding principles developed in the field of HCI. The inputs of these devices have evolved to function within the constraints of the devices such as current smart phones have touchscreens as a input and output device which saves space and allows for a large range of inputs and outputs. Examples of some current HCI's on mobile devices are as such Figure 1 [29], [30].

Figure 1: (a) *iPhone* (b) *Nexus6*

The specification of the *iPhone 6 plus* as stated on their website has a (158.1 x77.8)mm wide display touch screen and a software for voice control named *Siri* weights at 172 g. The *Nexus 6* on the other hand has the dimensions of (159.26 x82.98)mm with voice commands enabled weights at 184g. Both have a raft of specifications which make it very useful to users who are on the go or within confined spaces.

But strangely enough, after the smart phones were unveiled, tablet computers were also introduced into the market which has larger size in general. This seems to be inconsistent with the miniaturization trend which was set by the smart phones. Figure 2 shows the slightly larger versions of smart phones called tablets in the market right now[31], [32].

Figure 2: (a) *iPad Air2* (b) *Nexus 9*

The *iPad air 2* has a dimension of 169.5 x 240 mm with a mass of 444g while the *Nexus9* has a dimension of 153.68 x 228.25 mm weighting 425g both come equipped with touch screens and voice command availability. The most obvious difference between the smart phone and the tablets are their difference in mass and size where the tablets are bigger and heavier. This as mentioned earlier seems contrary to the miniaturization trend. This can be reasoned that the larger the screen, hence large the elements within the screen, therefore making it easier (accurate and faster) for the user to interact with it despite the loss of mobility[33][34].

Right after the tablet wars started, another strange trend began where touch screen smart watches emerged with both supporting voice input commands as shown in Figure 3 [35],[36]. The functions of these devices standalone are much less as compared to the tablets and smart phones. The full potential is unleashed when tethered to the tablets or smart phones. The input surfaces of these devices are the smallest followed by the smart phones and lastly the tablets.

8

Figure 3: (a) *Microsoft band* (b) *Sony smart watch 3*

As one notices all the devices here have the touch input as the primary input method with other input methods such as voice activation as a secondary. This basically indicates that the touch method of input is the most reliable and at the same time users feel it to be more interactive and natural as they utilize their tactile feeling. Research shows that the usage of an actual keyboard versus a virtual keyboard, a keyboards yields a higher rate of words per minute as compared to a virtual keyboard indicating tactile feed back as an important factor to be considered in usability [37][34]. In addition to that the error rate of the touch screen is affected by the target size within the screen in an inverse manner. Despite touch screen's imperfections as compared to traditional input devices such as keyboard and mice, its mobility and high tactile feedback to users makes it practical to be used as a mobile device input.

The touchscreen is the advanced model of the touchpad which was used on desktop set ups. The touchpad's ability includes accepting a wide range of inputs from the user with natural tactile feedback to the user via the user's mechanoreceptors but suffers the issue of mobility with ease of use as the touchscreens do. The input area of

the touchpad alike the touchscreen is limited by the size of the input area. The increasing size would therefore reduce the mobility of the device. Hence the scalability of the device is very practical for fixed positions devices like desktops where mobility is not an issue. Unfortunately since the touchpad's which were designed to work very well for desktops, and the touch screens technology is derived from it, is not a practical solution for mobility as its size is not scalable without sacrificing the mobility of the device it is being used on or with.

It was also noticed that the voice based inputs was always used as a secondary input method as compared to the touch screen method. The verbal command based input is extremely mobile due to its low power, small size and usage, where the user can input messages verbally when other limbs are busy such as driving a car. This method offsets the input medium to the environment. Hence is has great mobility but the medium which it offsets is extremely noisy. This is because the medium in which the acoustic signal travels through is normally quite noisy. These noises comprises of non-verbal and verbal noises. In addition to poor robustness, the voice recognition based system have a rather narrow input. For example the users' will find it very difficult to create characters or pictures which do not exist in the database as the touchpad can.

Another point of interest if one were to follow the chronology of devices introduced to the market, the sizes of the input first introduced was small with the smart phones, then it became large with the tablets and then became even smaller with the unveiling of the wearable devices. This basically shows that the industry tries to segment the market based on different levels of mobility and input accuracy (user ability) and cannot decide which is the best input size. This therefore gives researchers an opportunity to search for the ideal method which fits all mobility factors, accuracy and user ability.

2.3. RESEARCH BASED HCI

Observing trends related to internet of things which encompasses mobile devices, we noticed that mobility is achieved by the trend is to shift as much as possible the

limitations such as large computing loads to an offsite location, hence rendering the device mobile but at the same time able to perform the required computing tasks for the user such as collecting user data. Taking a leaf out of this trend, we would like to profit from a large but yet mobile input surface by delegating the input surface to the environment. Hence, breaking the relationship between the screen size and the mobility. Subsequent literature are reviews of some of the methods which have been tried by other researchers, we would try to build upon such revolutionary ideas by evaluating pros and cons of each research. As the adage goes, ' I have seen further it is by standing on the shoulder of giants'.

Table I, shows the list of researchers and the general methods employed in an attempt to shift the input surface to the environment hence attaining mobility.

Table I Comparison of methods used

Method of input to the computer	Visual	Acoustic	Kinesthetic	Magnetic
Hand Menu system	✓			
Omni Touch		✓		
Hambone		✓		
Skinput		✓	✓	
Wearable handwriting input device			✓	✓

Figure 4: (a) *OmniTouch* usage (b) *OmniTouch's Kinect* (c) *Hand-menu system* usage (d) *hand menu system*

Example of such an innovation is the *Hand-Menu-System* which utilizes visual capture device such as a camera to capture user's gestures and unique shapes but unfortunately suffers from instabilities due to unstable lighting and overlapping [38]. A predecessor to this technology is the *OmniTouch* which utilizes the *Microsoft Kinect* allowing it to mitigate the depth issues related to standard visual systems, as it contains within it a depth camera [39]. It also allows for unique shape inputs and gestures. Despite the benefits of being small and light, visual based systems are computationally expensive

and also require large amounts of energy, making it a poor choice for mobile devices which are required to be ubiquitous.

Figure 5: (a) *Skinput* usage (b) *Skinput* (c) *Hambone* usage (d) *Hambone*

This therefore leads us to acoustic or vibration based sensors which requires much less power as compared to visual ones, examples of innovations which utilizes such sensors are the *Hambone* which transfers voice through the human body to the ear whereby eliminating noise from the environment. The *Skinput* which utilizes the human body as the medium for conducting the vibrations of taps via the specialized cantilever system developed [40], [41]. This method has high noise immunity but requires training prior to usage. This system is unique as it utilizes human preprioception as one of the feedback which does not require additional design or cost and at the same time increasing

user ability of the device. Despite both these systems being able to achieve higher mobility than visual based systems, the input ranges were limited to pre-defined gestures or taps with correspondingly limited outcomes.

Figure 6: Wearable handwriting input device

The *wearable handwriting input device* on the other hand utilized the localization of a permanent magnet which was placed in a finger glove of the writing finger [42]. This method was able to achieve a wide input range akin to that of visual systems, had relatively low energy consumption which was akin to acoustic systems and was mobile. In additional to that, when used on the surface of the skin, leverages upon the free and ever-present mechanoreceptors for human feedback similar to that of *skinput*. The drawback of this system was that it required a specialized finger glove akin to a stylus and it could not differentiate whether the user was actually sketching on the surface or just hovering the finger above the work area.

Chapter 3 would discuss about the concept of the system which would be proposed as the first prototype while chapter 4 would describe the workings of the prototype and the results from the testing based on the observations collected in chapter 3 and by other researchers in chapter 2.

3. AN INTRODUCTION TO SYSTEM CONCEPT

From the literature review, it is found that despite visual systems being mobile and have readily large input ranges during the input phase, its large energy consumption requires it to be charged more often hence reducing its stand-by time, thus reducing its operational mobility. Acoustic systems on the other hand requires less energy during its input phase, hence extending its stand-by time, subsequently its operational mobility but with limited input ranges. The magnetic input addressed the problems of both the acoustic and visual systems but had reduced mobility due to the finger glove and ambiguity regarding the user's state of writing[42].

3.1. SYSTEM CONCEPT

The ideal solution for the new system would be if the wearable *handwriting input device* could function without the finger glove and could differentiate whether the user was actually tracing a shape on a surface or just had the finger hovering over surface. If the tracing of the finger was done on a surface without the finger glove, observable effects due to this tribological process would be heat and sound [43]. Heat signatures from overlapping body parts are difficult to differentiate, especially between the finger and the palm. Besides that, sudden temperature changes in the environment could result in fluctuation of results. This leaves acoustic detection as the most viable solution.

Hence, by adopting and merging the various ideas together, we propose the replacement of magnetic hall sensors from *wearable handwriting input device* with acoustic sensors and the removal of the finger glove containing the permanent magnet resulting in higher mobility. In theory, this results in a device which would be versatile in multiple fields as it can be used on any surface which only generates sounds when two surfaces rub against each other, able to handle a large range of unique inputs, consumes low amounts of energy as it relies on acoustic sensors, small in size and has the benefit of leveraging upon the human skin's mechanoreceptors for better tactile feedback.

With these in mind, the authors propose a system which utilizes the triboacoustical emitted signals (TES) generated by the user's finger tracing a shape onto a bare skinned

or covered area on the body. This method inherently filters out errors such as two surfaces overlapping but not touching scenarios which traditional visual methods suffer from. Energy efficient small form factor microphones which are cheaply available should be incorporated into the system allowing for high mobility, captures the resultant TES which is subsequently localized upon using time-difference-of-arrival (TDOA). The TES are broken into predefined blocks where a coordinate is calculated for each block. These resultant coordinates when joined together form the shape traced by the user. This localization method allows for a wide range of inputs as any shape or gestures can be drawn freely on a surface.

Existing systems only utilize transducers for feedback which are bulky and energy inefficient while this system has a choice to leverage upon the existing rich sensory system (mechanoreceptors) present within the human skin and proprioception. Tracing of the bare finger done on the bare palm is the best, due to the high density of mechanoreceptors sensitive to vibrotactile stimulus present on the human palm and finger [44], [45]. Nevertheless tracing can also be done on parts of the body covered with clothing which can still generate triboacoustic localizable signals and at the same time can be felt by the user as depicted in Figure 1(a). In the worst case scenario, the tracing can be done on any rough surface present around the user, but with the loss of the rich natural sensory feedback of the human skin. Examples are shown in Figure 7.

Figure 7: Writing with (a)(b) and without (c) natural sensor feedback

Besides working as a standalone input device, this proposed system could also be merged with other devices such as smart phones or head mounted display which enhances its functionality. The microphone arrays could be connected to the headmounted display and earpiece via the smartphone and worn on the wrist to detect shapes or gestures traced by the finger on the opposing palm. The additional feedback is then relayed back to the user acoustically and/or visually. This idea is depicted in Figure 8.

Figure 8: Acoustic input system melding

Additionally, this idea can also be implemented with tactile feedback devices strapped to various parts of the body for individuals who are visually and hearing impaired, or the elderly as they rely very heavily on their tactile senses to interact with their environment. It is believed that with the onset of visual deprivation and the inherent plasticity of neural cells develop enhanced tactile abilities[46]. Hence the usage of mobile tactile systems as in created in the research can be used in tandem with this research after the information has been converted to digitally legible data[47].

Figure 9: Tactile feedback device

Besides natural occurring tactile feedback from tracing the finger on the surface of the skin, an unbalanced motor can be used as a modular tactile feedback device. This device could inform the user of a successfully localized TES signal by vibrating. This indicates to the user using tactile /kinesthetic means that the input which the user traced on a surface has been accepted by the mobile device. This is similar to the functions which currently exist for the smart phone touchscreen inputs which vibrate when a character is successfully accepted by the virtual keyboard. This method of feedback is rudimentary. It can be seen though that with more advanced tactile feedback devices, the range of feedbacks can be wider therefore allowing the proposed acoustic device to be more useful to visually impaired, acoustically impaired and elderly people who suffer the similar symptoms due to degradation.

Besides tactile feedback, other forms of feedback such as audio and visual kinds of feedback modules can be created for specific needs and disabilities. This thesis evaluates the feasibility of localizing upon TES by first discussing the localizing equations, TES characteristics, realizing all the ideas through the design considerations section and finally verifying the functionality of the developed system by benchmarking it against the visual localization system.

3.2. SOUND LOCALIZATION CONCEPT

Localization is in fact an endearing concept; it has been revived time and time again as technology advanced allowing different permutations each time. In nature, animals such as dolphins, bats use acoustics for localization be it for prey hunting or for obstacle avoidance. In human society today, localizations are utilized every day for things such as global positioning (GPS) or radar technology for tracking planes in the sky both utilizing radiowaves. Regardless of the medium used, the general concept for localization is the same.

By using cross-correlation, the TDOA between sensors can be attained, for use either with the hyperbolic localization equations or angle-of-arrival (AOA) method. The first method is illustrated as shown in Figure 10.

Figure 10: hyperbolic localization method

$$R_n = \sqrt{(x_n - x_s)^2 + (y_n - y_s)^2} \tag{1}$$

R_n represents the distance of separation between the sound source and the nth sensor. x_n and y_n represents the coordinates of the nth sensor.

Both AOA and hyperbolic localization method rely heavily on TDOA for its localization ability. The time difference of arrival TDOA is needed as sound emission times of TES are unknown and uncontrolled. TDOA is attained using (2) [48].

$$t_{n,d\,tdoa} = t_n - t_d = \frac{R_n - R_d}{V_s} \qquad n \neq d \tag{2}$$

n and d in (2) both represent the sensors, the reason that n is not equal d is so that all the possible combinations of TDOA can be attained. Variables t_n and t_d represents the time taken for sound to travel from the sound source to the sensors, V_s represents the speed of sound.

The number of sensors used defines the number of equations able to be derived from (2) which in turn dictates whether the set of equations are undetermined or overdetermined. Assuming the numbers of sensors are more than the unknown variables, the system reduces to an elegant overdetermined system. As elegant as the equation might seem, an analytical solution is highly unattainable due to the fact that the TDOA

20

measured is wrought with errors. This can be solved by implementing the said equations numerically via the gradient descent method to get approximate towards the solution as shown in (3).

$$x^{i+1}_s = x^i_s - \alpha\nabla F(x^i_s, y^i_s)$$
$$y^{i+1}_s = y^i_s - \alpha \nabla F(x^i_s, y^i_s)$$

(3)

The velocity of sound is a constant within the equation, as this velocity is the speed of sound within the assumed homogenous properties of air shared by the three sensors. α is the step size for each correction of (x^i_s, y^i_s).

The output of the error function $F(x^i_s, y^i_s)$ is evaluated at each iteration, utilizing the guessed values of coordinates x^i_s and y^i_s. If this error value is higher than a user defined value, the system will try to guess the next improved coordinates' x^{i+1}_s and y^{i+1}_s values by using the gradient ∇ of the function $F(x^i_s, y^i_s)$. This process will continue until the error value set by the user has been achieved or when the number of maximum iterations set by the user has been reached.

Meanwhile the second method utilizes the TDOA attained in (2) is used to calculate the angle of which the sound source is arriving from in reference to an arbitrary sensor center. Utilizing the TDOA from (2), the AOA can be found with the help of (4).

$$\phi = \sin^{-1}\left(\frac{t_{n,d\ tdoa}\cdot V_s}{L_d}\right)$$

(4)

ϕ describes the calculated angle of arrival of the sound source with reference to the axis perpendicular with the base plane, L_d is the distance between the two sensors. Localization of a sound source point in 2 dimensional space via the intersection of the angles of arrivals is attainable using a minimum of three sensors of known location as illustrated in Figure 11. This intersection is achieved by using the straight line equation of the Cartesian coordinate system as in (5).

$$y_s = m_{n,d}\cdot x_s + c_{n,d}$$
$$m_{n,d} = \tan^{-1}[(90 - \phi) + (\beta\ )]$$

(5)

$m_{n,d}$ stands for the slope of the line which can be attained from the angles attained in (4) with slight modifications. While $c_{n,d}$ represents the intersect of the vector with the y-axis. This generalized term when applied to localization as in Figure 11, produces two equations for each set of sensor pairs. Assuming that the sensor base planes are the axes, results in the sensor pair base plane angle, β to $0°$ yielding (6).

$$y_s = m_{1,2}.x_s + c_{1,2}, \ y_s = m_{1,3}.x_s + c_{1,3} \qquad (6)$$
$$x_s = \frac{c_{1,3} - c_{1,2}}{m_{1,2} - m_{1,3}}$$

Despite the availability of the slope for both sensors pairs, the unknowns such as the sound source coordinates and the y -intersect constant results in (6) to be unsolvable. The solution to this is to first attain the y - intersect constant values. These two straight line equations intersect at a unique middle point between their respective sensor pairs. The equation in (6) has its sound source coordinate (x_s, y_s) modified to the midpoint $(y_{n,d\ mid}, x_{n,d\ mid})$ as shown as in (7).

$$y_{1,2\ mid} = m_{1,2}.x_{1,2\ mid} + c_{1,2} \qquad (7)$$
$$y_{1,3\ mid} = m_{1,3}.x_{1,3\ mid} + c_{1,3}$$

By solving for the midpoint first, y-intersect constant can be attained and subsequently used to solve for the sound source using (6). Solving for the midpoint requires (8).

$$\frac{\sqrt{(x_n - x_d)^2 + (y_n - y_d)^2}}{2} = \sqrt{(x_n - x_{n,d\ mid})^2 + (y_n - y_{n,d\ mid})^2} \qquad (8)$$

In this equation, the unknowns are the midpoint coordinates, while the constants are the sensor coordinates. The variables $x_{n,d\ mid}$ or $y_{n,d\ mid}$ are solved using substitution within (8).

Equation (4) has inherent localization accuracy errors as it is an approximation equation. Hence, despite using ideal TDOA values, (4) will produce an angle of arrival smaller than that of the actual angle of arrival. As a result, the intersection of the two angle's vector will occur sooner when closer to the base planes as opposed to the actual sound source location. This is illustrated in Figure 11.

Figure 11: Localization

The 'star' represents the localization derived from the AOA method and the 'triangle' represents the actual sound source location. Between the AOA and hyperbolic gradient descent localization method, the AOA method was chosen to be used in the first prototype as it is mathematically simple and also requires less computer resources to implement despite its known accuracy deficiencies.

Regardless of the method used another factor which is of the outmost importance when designing a wave based localization system is the wavelength or the frequency of interest which is to be localized. This determines the spatial separation between the sensors which disambiguates detected phase differences and in turn prevents spatial aliasing. Errors in capturing the phase lags and its polarity would cause the AOA to either intersect at an erroneous point or not intersect at all. Spatial aliasing occurs when the a same set of phase lag represents two different angle of arrival [49]. Rotation/ polarity of the signal is used to determine its arrival quadrant while the signal delay is used to determine it's arrival angle.

$$\pi \frac{L_d}{\lambda} \sin \phi_1 = \pi \frac{L_d}{\lambda} \sin \phi_2 + n2\pi, \quad \phi_1 \neq \phi_2 \tag{9}$$

For (9) conditions $1 \geq \sin(\phi_1) \geq -1$, multiple solutions exists when $d \geq \lambda/2$. Hence to prevent this, distances between the microphones have to be kept smaller than $\lambda/2$. Figure 12 illustrates the equation using sensor separation of $\lambda/2$ and λ.

Figure 12: Sensor separation distance (a) $\lambda/2$, s1 direction, (b) λ, s1 direction, (c) $\lambda/2$,s2 direction, (d) λ, s2 direction

Figure 12 shows the two different conditions, one being the sensor separation, the other the direction of the signal. A periodic sine wave is used to depict the sound waves being captured. The graphs below the sensors indicate the electrical signals which are produced by the respective sensors. The dotted box indicates the evaluation window for the cross-correlation algorithm, which is normally twice that of the sensor separation distance L_d. The distance L_d is represented in terms of wavelengths $\lambda = V_s/F$, where V_s is the velocity of the wave, in this case sound. F is the frequency of the wave being evaluated. Cross-correlation is also included in this discussion as the sensor separation has a large impact upon the performance of the cross-correlation, and in turn the localization. The y-axes of the graphs in Figure 12 are magnitudes of the signals. The x axis of graphs depicting the outputs from the sensors represents time. While the x axis for the graphs showing the product of cross-correlation depicts the sample number. The arrow pointing perpendicular to the sample axis depicts the correlation center, which brings the meaning that if a peak occurs at this point, the two signals are neither leading or lagging one another, while if a peak appears on the right of this point, one signal is leading the reference and vice versa. The differently coloured lines indicate signals which occur at different times. The signal of interest is depicted as a solid black line.

In Figure 12(a), it can be seen that the signal impinges upon sensor 1(S1) first before reaching sensor 2(S2). This causes the electrical signals of the wave front to be first produced in S1 graph and then π later appears on the S2 graph. It can be seen that the second half of the graph on S1 is clipped by the dotted box indicating that it will not be evaluated in the cross-correlation. Instead the blue dotted trace on S1 which was not part of the original signal to be evaluated would be evaluated in the cross-correlation. The results from the cross-correlation shows that the peak of the correlation process happens on the left of the correlation center which indicates that S2 lags S1 by a certain number of samples.

Meanwhile, in Figure 12(c), the exact situation with the only difference of the signal direction was re-enacted. It can be seen that the correlation peak was offset by the same amount as it was in Figure 12(a), but in this case it was leading instead of lagging. This clearly indicates that the polarity of the signal defining the leading or lagging is used to

define the quadrant in which the signals are arriving while the magnitude of the phase differences provides the angle of arrival via equation (4), producing +90˚ or -90 ˚.

In Figure 12(b) and (d), the distance of separation was fixed at λ of the frequency used. The directions of the signals are similar with that to Figure 12(a) and (c). Inspecting the cross-correlation results show that the displacements of the peaks are exactly the same, which is 0. By using (4) with this data, the angle of arrival yielded was 0˚ which was erroneous. In addition to that, the quadrant in which the sound source arrives from cannot be identified either.

From the examples given, it is clear that the distance between two sensors must be kept within λ/2 of the frequency of the signal being localized upon. Distances less than λ/2 such as λ/4 can be used but this limits the available resolution. Hence to maintain the maximum resolution without spatial aliasing, the sensor distance should be kept at λ/2. In addition to that, despite keeping the distance of sensor separation to λ/2, the sensor pair method only works in the half-plane as it is unable to differentiate the occurrence of sound sources with a direction symmetric to the microphone pair. Hence it only works of two quadrants.

Cross-correlation is a method used to find when a match between two signals occurs. By implementing a cross-correlation and detecting the location of the highest peak, the lag and polarity can be detected and subsequently fed to the AOA. The equation for cross-correlation is given as (10).

$$(f.g)[n] = \sum_{m=-\infty}^{m=\infty} f[m].g[m+n] \tag{10}$$

This equation when realized could look like Figure 13.

Figure 13: Cross- correlation process

Figure 13 shows two discrete signals, f[n] and g[n], whereby their similarity to each other is to be tested via cross-correlation. The signal g[n] is incrementally moved towards the static f[n] where each overlapping elements are multiplied and summed together. This process continues until no elements within f[n] and g[n] are overlapping. The resultant graph would have the largest peaks at the [n] sample number where the two graphs are the most similar. In addition to that, the number of samples created from this process would double the original [n] becoming [2n]. If the graph were to be cross-correlated with itself, it would generate a peak at the center, [n] of the graph containing [2n] elements.

Figure 14 (a) shows and example of two periodic signals comprising of 73 samples which are similar to each other but shifted by 6 samples. Figure 14 (b) depicts the correlation process when the reference signal is correlated with itself producing a peak at sample 73 which is at the center of the 146 element graph. End result as shown in Figure 14 (c) was produced when the reference signal was cross-correlated with the evaluated signal. The peak resided at sample 67 which was 6 samples to the left of the center. This

indicated that the evaluated signal was lagging the reference signal by 6 samples. Imagine if each of the samples represents the sampling rate of a DAQ of 1us, it would mean that the evaluated signal is lagging behind the reference signal by 6us. From the figure it can be deduced that the frequency of the two signals are about 27KHz. Speed of sound assumed to be 330m/s. If the rule for sensor separation is adhered to be $\lambda/2$, this would yield a separation of 0.006m. Using equation (4), the angle of arrival of the evaluated signal is +19˚.

Figure 14: Cross-correlation(a) raw signals (b) no lag (c) lag

3.3. SOUND OF INTEREST - SCRATCH SOUND CHARACTERISTICS

TES is defined as the acoustic signals generated by means of tribology. Acoustic signals are defined as longitudinal waves due to their mode of propagation. Acoustic wave propagation requires medium of transmission. Consequently, the state of the medium which it travels through affects the propagation speed. The approximate speed of sound in dry air (RH = 0%) at 20℃ is 343 ms-1.

Tribology is a complex science which involves the interaction between two or more surfaces with a net motion larger than zero. In this case, we are interested in the byproduct of this interaction at the interface in the form of sound produced for localization purposes [50]. Researchers [43], [50] show that tribological interactions between surfaces rigid and elastic alike do generate sound. Acoustic magnitude and frequencies measured by the researchers appear to be white noise dependant on parameters such as materials, surface roughness, roughness wavelength, contact force, surface conditions (oil, Rh% etc) [43], [50]. General characteristics discovered by researchers [51] show that the signals generated triboacoustically are quasi-periodic and non-stationary implying that these signals in time domain attained within a specified time frame are unique from the signals collected in the other time frames [51]. Cursory inspection by the authors as shown in Figure 15 yields the observation that naturally occurring acoustical signals are neither strictly periodical nor stationary.

Figure 15: Spectrogram (a) scratch sound+ background sound, (b) background noise, (c) voice + background noise

Experiments were conducted to better understand the effects of the surface material, force applied by the finger and the speed of finger moving across the surface has on the acoustic signals generated. The combinational table as in Table II was created.

Table II: Combinational table

Combination prefix	Surface material	Force(user defined)	Speed(m/s)
A(1,2)	Paper	Soft	Slow
B(1,2)	Paper	Soft	Fast
C(1,2)	Paper	Hard	Slow
D(1,2)	Paper	Hard	Fast
E(1,2)	Cloth	Soft	Slow
F(1,2)	Cloth	Soft	Fast
G(1,2)	Cloth	Hard	Slow
H(1,2)	Cloth	Hard	Fast
I(1,2)	Wood	Soft	Slow
J(1,2)	Wood	Soft	Fast
K(1,2)	Wood	Hard	Slow
L(1,2)	Wood	Hard	Fast
M(1,2)	Skin	Soft	Slow
N(1,2)	Skin	Soft	Fast
O(1,2)	Skin	Hard	Slow
P(1,2)	Skin	Hard	Fast

An experimental setup was created to test out the items on the combinational table as shown in Figure 16.

Figure 16: TES characteristics setup

The setup consists of a softboard which acts as a platform in which the sensors and workplane resides on. Two types of sensors exists in the setup, the SPM 0408LE5H which is generally used for voice based applications and the SPM0404UD5 which is used to capture ultrasonic sounds. The amplifiers used on all the modules were the LM4562 which has a gain bandwidth product of 55MHz. Outputs from the sensors are channeled into the amplifiers which in turn are channeled into the DAQ running at 1Msa/s for digital format conversion. The frequency response curves of the sensors are as shown in Figure 17.

Figure 17: Frequency response curve (a) SPM 0408LE5H (b) SPM0404UD5

The SPM0408LE5H has a rather flat response curve from 100 Hz to 10 KHz which is also similar to that of SPM0404UD5. Unfortunately, the response curve declared in the data sheet does not show the response for SPM0408LE5H after 10 KHz. Despite that, it is expected that the response would be much higher for higher frequencies due to the increasing trend seen just before the graph ended at 10 KHz. A peak response at frequencies 40 KHz to 50 KHz in the frequency response graph of SPM0404UD5 was

found. TES as other researchers have attested is a white noise, therefore the energy should be distributed across the frequencies evenly provided that the devices used to measure it have a flat frequency response with a large enough bandwidth. The amplifiers have a large bandwidth but the sensors do not. In addition to that, the sensors have an uneven response curve, which means in, certain frequencies of TES will be accentuated due to this.

The experimental procedure involved the activation of the DAQ and sensor modules which collects data for duration of two seconds in which the experiment of tracing the finger over various surfaces with varying speeds and forces are conducted. This amounts to eight million data points for each combination. The speed and force which the user is supposed to be applied is based on the user's perception and no measurement device was used to verify this. The experiment was conducted twice for each experiment as the two second window is rather short for human actions, but rather long for computer acquisition at high sampling rates. There is a possibility that the human was unable to complete or even start the experiment before the two second sampling window closed. Only one of the results was chosen from the two experiments conducted from each combination.

The data was then analyzed in the frequency domain to produce the spectrograms and the accompanying raw signals as shown in Figure 18 and Figure 19. The envelope of the raw signals seem to signify the amount of energy which was used to create the TES, As area under the envelope is the product of voltage and time, with the assumption of circuit resistance of 1Ω, which gives the product of voltage and time. $V^2T = \text{Energy}$.

Figure 18: Spectrograms of combination table elements

Figure 19: Raw signal of combination table elements

The even numbered graphs generated in Figure 18 and Figure 19 are results captured from the high frequency sensor SPM0404UD5 while the odd numbered graphs are results captured by the sensor SPM0408LE5H. The conditions of tracing the finger fast and with large normal force against all surfaces resulted in all the detectable frequencies being activated as shown in Figure 18(d1) (h1) (l1) (p1) (d2) (l2) (p2) (h2). This trend was also detectable but to a lesser degree in the events where the finger was traced slowly but with large normal force against the workplane. The raw signal in Figure 19(d1) (h1) (l1) (p1) (d2) (l2) (p2) (h2) have higher magnitudes due to the larger force and speed applied by the finger. It can be seen that the spectrogram was activated in all the frequencies for a longer period of time due to the slow tracing speed as shown in Figure 18 (c1) (k1) (g1) (o1) (c2) (k2) (o2) (g2). Meanwhile tracing the finger lightly and slowly on any surface slowly usually yielded low magnitude response on all the frequencies in the spectrogram for all the tested materials as shown in Figure 18 (a1) (e1) (i1) (m1) (a2) (e2) (i2) (m2). Slightly more frequencies are activated with higher magnitude when the finger was traced lightly but faster. This can be explained that the more energy is used by the finger in the form of speed and normal force, the higher the probability that this energy would be converted to other forms such as heat and sound due to friction which exists between the two contacting surfaces. This is also translated as higher voltage magnitudes for the detected raw signals in Figure 19. Regardless of that, the postulate that SPM0404UD5 would accentuate the frequencies from 40 to 50 KHz

46

was found to be true for all the conditions regardless of the surface, speed or normal force applied by the finger, provided that there were some detectable sounds created from the triboacoustic trace. Where the sensor SPM0408LE5H kept accentuating upon frequencies around 25 KHz for all conditions tested was detected. This was further proven with an acoustic sample of finger tracing on palm sound captured by a microphone (Earthworks - M30BX) with a relatively flat frequency response curve from 10 Hz to 30 KHz as shown in Figure 20(b). Comparing the frequency response of Figure 18, Figure 19, Figure 20(a), shows no additional amplification of certain frequencies regardless of surface by the sensors SPM0408LE5H in Figure 18(m) - (p). The TES is truly a white noise, the selective frequency amplification seen in Figure 18 are caused by the sensor's frequency response.

Figure 20: (a) Handscratch frequency response -Earthworks M30BX   (b) Frequency response curve of Earthworks M30BX (c) Polar response Earthworks M30BX - omnidirectional

47

3.4. DESIGN CONSIDERATIONS

The choice of workspace materials which interacts with the finger of the author is important for this experiment as these mutual interactions defines the values of TES generated. This therefore implies the parameters that affect the finger's vibrational outputs would also most likely effect the acoustic frequency and magnitude. The tribological process in this chapter requires two surfaces to be in contact, human finger (bare or covered) and a generic workplane surface. The generic workplane in this particular paper is represented by the glabrous skin (palm), cloth and paper (book). These surfaces were chosen as palm and cloth represented locations which the user can trace a shape on their body with a finger while paper (book) represents the generic surfaces the user can acquire to trace on if the latter surfaces are unavailable. Skin rheology varies greatly between individuals which are also affected by environmental conditions, hence producing varying magnitudes and frequency of acoustic signals.

As it was seen in Figure 18, 25 KHz kept yielding the highest magnitude for various surfaces and conditions. Due to this, and the fact that higher frequencies give better cross-correlation accuracy, 25 KHz was chosen as the fundamental frequency. Despite the advantages attained from using higher frequency as the fundamental frequency, literature as discussed in Figure 12 dictates that the distance of sensor separation for single frequency sound as $\lambda/2$ [48], which greatly reduces the angle ϕ resolution as well as increasing the challenge of fabrication of microphone distance separation with the assumption that the sampling rate and bit resolution remains constant. The reduced angle resolution can be explained using (4), by reducing L_d, the finest available unit $t_{n,d\ tdoa}$ will represent a larger steps of angle ϕ. Popular solutions utilizing controlled periodic signals would be increasing the sampling rate which would increase the cost or extrapolation which would increase the processing time with non-guaranteed results.

Due to the fact that the signal of interest, TES is a quasi-periodic and non-stationary signal, the authors proposed using the microphone pair distance of separation of the 8th subharmonic of 25 KHz. This solution retains the accuracy advantages from using an assumed fundamental frequency and at the same time addresses the low angular resolution deficiency. Additionally, the increase in sensor separation reduces the complexity of fabrication. This can be described with the postulated fundamental

frequency of 25 KHz at the assumed speed of sound of 340ms-1, resulting in the distance of sensor separation based on [48], as 0.0068m. If the 8th subharmonic of 25 KHz is used with the same assumptions as above, the distance of separation would be 0.0544m. Referring to (4), it is evident that by increasing the sensor pair distance of separation, the TDOA resolution would increase which in turn would increase the angular resolution L_d, hence increasing the localization accuracy and resolution without requiring any increase to the sampling rate of the DAQ or extrapolation of measured data. This system therefore had a sensor separation designed for a signal of 3.125 KHz but instead measured the TDOA of a 25 KHz signal.

If the signal localized upon is a periodic and stationary signal of 3.125 KHZ, the localization system would fail due to the erroneous TDOA's captured using cross-correlation. This phenomenon is shown as in Figure 21

Figure 21: Errors caused by periodic signals interacting with cross-correlation (a)
periodic signal (b) TES

Figure 21 (a) shows a periodic signal captured by the microphones where the sensor pair arrangement used to capture it is spaced at 8λ/2 distance from each other. λ in this context refers to 25 KHz's λ. This indicates that the maximum possible lag or lead can be more than ± 180. As it can be seen the signals represented in the graphs for both S1 and S2, three waveforms appear within the sensor separation distances. This distance was based on the argument that the TES signal can utilize this spacing but the periodic signal can't. The evaluation window is represented by the square dotted box encompassing both the signals in S1 and S2 graphs which in this case has the width of 4λ. The signal of interest is actually unknown, but in this case we can see from the solid lines that the S2 leads S1 by about 100 samples. The hardware uses a sensor separation which is more than λ/2 and larger than λ evaluation window. This therefore causes more than one peak to appear from the cross-correlation as compared to the case as shown in Figure 12(a) and (c). An ideal periodic and non-stationary signals inherently the same for every 2π. Ideally, the cross-correlated results would yield a unique maximum point. This site in which the highest peak appears corresponds to the correct TDOA. Unfortunately, this is not true for a non-ideal case where errors are present in the creation, transmission and detection of the signal. The cross-correlated values are very close to one another due to the similarity in periodic signal magnitudes and phase. This unstable system when introduced to noise can easily yield maximum peaks at sites other than the true peak. This results in erroneous TDOA being collected.

Figure 21 (b) used all the same parameters except for the TES signal which was quasi-periodic and non-stationary. The evaluation window used was many times larger than 4λ. This TES signal was an actual captured data by the hardware which was cross-correlated offline to produce the cross-correlation data. It is difficult to identify where the signals actually match each other visually. The cross-correlation process conducted found that the signals S2 lead S1 by 94 samples which is equivalent to 94us. The cross-correlation produced very clear peaks as the input signals were quasi-periodic and non-stationary making the magnitudes and phase of the signals to differ stochastically hence resulting in very sharp cross correlating peaks isolated from one another. From the frequency domain point of view, this sensor arrangement and design allows for more frequencies to participate in the TDOA whereby increasing the odds that the results

attained would be more accurate. This allows it to have better interference immunity as that of pure periodic signals. Hence, in short the more erratic the sound sources signal, the better its TDOA accuracy.

With the algorithm for localization and the distance between sensors decided, the next item to be considered would be the actual placement of the sensors. The angles between the sensors have to be determined through analysis as its placement might affect the accuracy of localization. Another parameter which had to be considered would be the device's application which dictates the placement of the sensors. This device is intended to be worn on the wrist of the user; hence placing sensors at large angles from one another would actually impede the user's mobility

To save time, a simulation tool written in C language was written to calculate the errors yielded from the different shapes in which the sensors can be arranged in at distance of 0.0544m from each other. This software's imaginary workplane has an x axis ranging from 0.04m to 0.13m and a y axis ranging from 0.04m to 0.18m as shown in Figure 22. An imaginary sound source was moved around this imaginary workplane in an incremental manner of 0.01m where each step (superscript g). The simulation is stopped when all this point has finally traversed each point within the imaginary plane. Each time the imaginary sound source was moved to a new position, a set of TDOAs were created based on the ideal coordinates $(x_{s\,real}{}^g, y_{s\,real}{}^g)$ as shown as in (11).

Figure 22: Flowchart of simulation

$$t_{n,d} = \frac{\sqrt{(x_n - x_{s\,real}{}^g)^2 + (y_n - y_{s\,real}{}^g)^2} - \sqrt{(x_d - x_{s\,real}{}^g)^2 + (y_d - y_{s\,real}{}^g)^2}}{V_s} \quad (11)$$

The attained TDOA are then fed to the AOA algorithm (4) (5) (6) (7) (8)which was programmed into the simulation program to create AOA calculated coordinates $(x_{s\,AOA}{}^g, y_{s\,AOA}{}^g)$. The difference between the ideal coordinate and the AOA calculated coordinates were calculated for each point on the imaginary plane thus used to create an error map for the imaginary plane. The errors calculated on the contour map are defined as (12).

$$xy_{error} = \sqrt{(x_{s\,real}{}^g - x_{s\,AOA}{}^g)^2 + (y_{s\,real}{}^g - y_{s\,AOA}{}^g)^2} \quad (12)$$

Figure 23: Error contour map

54

As it can be seen in Figure 23, the occurrence of errors were not uniform for the arrangement of sensors in a triangle fashion S1 (0.02, 0.02), S2 (0.075, 0.02) S3 (0.048,-0.028). This implied that some arrangements could have yielded different accuracies.

Hence, some arrangements for the sensors were proposed based on literature reviews and the intended usage of the device. They were: a straight line, where each sensor was spaced 0.0544m from each other, except the last and the first, a triangle which ensured each sensor was equidistance from each other and finally the letter L-shape. These arrangements are illustrated in Figure 24.

Figure 24: (a) Straight line (b) L- shape (c) triangle

As seen in Figure 23 and Figure 25(a)-(k), it is difficult to decide which method has lower error based solely on the contour map. Hence the mean of the errors x_{em}, y_{em} for each axis is used to calculate the hypotenuse of all the errors points xy_h within the contour map which was used as a determinant as in (13). Equations (12) and are (13) essentially the same with (13) giving more detailed data for analysis while (12) gives a better visual overview.

$$x_{em} = \frac{\sum_{g=0}^{g=q} \sqrt[2]{(x_{s\ real}{}^g - x_{s\ AOA}{}^g)^2}}{q} \ , \ y_{em} = \frac{\sum_{g=0}^{g=q} \sqrt[2]{(y_{s\ real}{}^g - y_{s\ AOA}{}^g)^2}}{q} \tag{13}$$

$$xy_h = \sqrt[2]{(x_{em} + y_{em})^2}$$

The results of xy_h for different arrangements are as shown in Figure 25(l). While (12) produced the error maps in Figure 25 (a) - (k).

Figure 25: Error maps - L shape (a) 80˚ (b) 70˚ (c) 60˚ (d) 50˚ (e) 40˚ (f) 30˚ (g) 20˚ (h) 10˚ (i) 0˚ (j) 90˚-line (k) triangle (l) mean cumulative errors

As it can be seen from Figure 25, the L-shape sensor configuration yielded the best results. But the L-shape has a right angle, which means it might not be able to fit certain

palm sizes. Hence the right angle needs to be replaced with at slightly larger angle to accommodate as many palm sizes as possible.

Figure 26: Proposed sensor arrangement

Based on Figure 26, the positions of S1 and S2 would be fixed but the position of S3 would be changed by rotating it counter-clockwise with S1 as the axis of rotation in the increments of 10° from 10° to 90°. The mean cumulative error attained for each S3 placement is as shown in Figure 27.

Figure 27: Cumulative mean error - angles

By referring to Figure 27, it can be seen that the angle of S3 of 50° yields the lowest cumulative mean error for all the angle variations. Despite that, the angle 50° is too wide and would hinder with the user when worn on the wrist. Hence as a compromise, the

authors chose the angle 30° which had relatively higher error of 0.0005m but was reasonable as it could accommodate many hand sizes without being a hindrance. It was therefore decided the sensors should be arranged at coordinates S2 (0.02, 0.02), S2 (0.075, 0.02) S3 (-0.008, 0.067).

3.4.1. DETECTION OF TRIBOACOUSTICAL EMITTED SIGNALS(D-TES)

This system was designed to only respond to scratch sounds. Voice in this particular system is considered as a noise source which occupies the lower frequency bands as shown in Figure 15. Comparing spectrograms in Figure 15 it is easy to discern voice as its energy is mostly contained below the 7 KHz band. Incidentally the signal of interest, the scratch sound has similar characteristics with that of voice + background noise where energy is spread across the whole detectable frequency band including below the 7 KHz level. The slight difference is that voice + background noise has higher ratio of energy in its lower frequency as compared to scratch sound which has higher concentration of energy in the higher frequency bands.

Hence using a simple highpass filter cannot correctly determine whether the sound accepted is that of a scratch sound or some high frequency based environmental noise. Using FFT we calculate the Decibel ratio of the sum of high frequency versus the sum of the low frequency as shown in (14).

$$R = \frac{\sum_a^b M}{\sum_c^d M} \tag{14}$$

R is the ratio of sums between the high and low frequencies components. M describes the magnitude of the frequencies attained, while a represents the lower limit of the high frequencies and b represents the upper limit of high frequencies. In this case, it was set to 8 KHz and 30 KHz. c represents the lower limit of low frequencies and d reprsents the upper limit of low frequencies which was defined as 1 KHz and 7 KHz.

The ratio should be set heuristically as it varies from system to system. Scratch sounds were found to have a very distinct ratio in this set up.

Despite the best of efforts to keep the resultant noise low, noise will still be picked up by the system, hence a simple voltage thresholding method is used to counter this

noise floor. Besides setting the noise floor threshold, a maximum limit voltage setting was also set, to avoid saturation signals from being accepted for evaluation. Saturated signals have flat peaks which basically produces very poor TDOA. This happens as the generation of TES is difficult to control; hence many factors can drive the amplifiers into saturation. Saturation limit was set to 97% of the actual saturation value. These evaluations were conducted upon samples collected within a time frame.

Hence the D-TES is a combination of the voltage range between the noise floor and the saturation voltage and frequency ratio method, any signals which qualifies these three criteria's were then allowed to proceed to the next level of processing which is the cross-correlation process to find the TDOA.

To validate this idea, a test was conducted where a single scratch sound event was conducted and the signals collected were with and without D-TES filtering in real-time. The results are as shown in Figure 28. Figure 28(a) shows the raw signal before D-TES and the corresponding Ratio calculation between high frequency and low frequency components. The threshold specification was set to 5 heuristically. It can be seen that 3 regions exist in the raw signal, where the first when the hardware was not turned on, the output voltage measured was below the noise floor voltage but managed to achieve high frequency ratios due to self noise. The second region was when the output voltage measured was above the noise floor and below the saturation voltage but did not satisfy the requirement of having the R above 5. This was due to the first contact of the finger with any surface was a low frequency based interaction. The subsequent motion of the finger moving lightly on the surface of the workplane generated signals which were above the noise floor and above ratio value of 5 which is shown in Figure 28(b). Unfortunately, a significant number of signals were in saturation. Despite that there were signals in this region which satisfied all the three conditions although it was not apparent with the naked eye. The total signal length prior to D-TES comprised of 40000 samples. Of this total number, about 18000 samples were extracted using D-TES. As a comparison these two sets of data were plotted in a single spectrogram to show the difference as shown in Figure 28 (c) and (d). The first part of the signal Figure 28 (c) comprised of the original signal prior to D-TES. Figure 28 (d) comprised of the signal which was filtered by D-TES. It can be seen visually the signal processed by D-TES is between the noise

floor and the threshold voltage. In addition to that the spectrogram shows that the Figure 28 (d) has evenly distributed energy throughout the frequency band, since the number of frequency bands for the high frequency is higher the ration would results in a larger value than 5. This therefore shows the effectiveness of the D-TES in sorting out the TES signals.

Figure 28: (a) raw signal with ratio (b) FFT of raw signal and ratio (c) raw signal and FFT of raw signal (d) signal after D-TES 'filter and FFT of D-TES filtered signal

4. PROTOTYPE

A hardware device can in theory be built from the cumulation of all the ideas discussed in previous chapters. Despite that, real-life problems which were not foreseen during the conceptualizing stage could creep into the hardware. This therefore spurred the need for a prototype stage to implement the ideas in real-life to study the weaknesses which arise and subsequently improving upon them as needed.

4.1.   HARDWARE

Design consideration merges the objectives and limitations of the known technology to establish a tentative design. Premises of arguments within the tentative design were proven by building and subsequently experimenting upon it to avoid fallacies. In this case, the experimental prototype was created using hardware and software. The creation of the prototype allowed the software to interact with the environment through the hardware. It is therefore imperative that the hardware design and assembly had to be compatible with the research objectives. The implemented system in this case comprises of components as shown in Figure 29.

Figure 29: system overview

The TES generated from the interactions of the finger with the passive material of the workspace were caught by three spatially displaced microphones (SPM040LE5H) which were subsequently converted into electrical signals to be fed to the amplifier. It

had to be taken into account that such acoustical signals are extremely small and therefore require large amplification. In this particular set up the authors used a 2.5k, two stage amplification to amplify the signals from the microphones which have already been pre-amplified within its surface mount devices (SMD) body. These amplified analog signals were then channeled to a 12 bit DAQ sampling at the rate of 1MSa/s which were then converted into digital data and processed by the algorithms written in the computer for localization.

This section describes the hardware of Figure 29. The microphone required two stage amplification and was based on the design as shown in Figure 30.

Figure 30: Circuit Diagram of sensor modules

The limit for biasing voltage of the Lm4562 amplifier is $\pm$ 2.5V to $\pm$17V, but in this case is $\pm$ 9V was used although $\pm$ 17V would give a wider margin it was found to be not practical spatially and economically. Instead, this research relied on off the shelf 9V batteries which are readily available. The microphone on the other hand was powered by a 3.3v coin battery which again was readily available. The microphone used, came with inbuilt amplification which allowed for noise mitigation, in addition to that, decoupling capacitors were placed along power rails and also at output ports to reduce noise. This

design could be further miniaturized by eliminating the amplifiers altogether but with the condition that the low voltage output signals are properly shielded from noise.

Figure 31: Realized sensor module

The outputs of these sensors were inputs to the DAQ which had four inputs, each sampled at 1Msa/s with the resolution of 12bits. Assuming that the amplifiers were able to attain ideal rail to rail voltages, $\pm$ 9V, this produced the voltage resolution of 18V/ 4096 = 4.4mv per 12 bit combination. Better resolution could have been achieved if the sensors utilized the full range of the amplifiers at 34V. The voltage resolution which

theoretically could be achieved was 8.3mv per 12 bit combination. The data was fed from the DAQ into the computer using PCIE slot. The computer used was a Mini ATX with a (Intel(R) Core(TM) I5-750s)@2.4Ghz) processor, running on a Japanese flavored Linux system called Vine version 4.4.5-6VL6, kernel Linux 2.6.35-20VL6 with GCC compiler version 4.4.5, an integrated graphic card, a 17 inch monitor, 4GB amount of ram and 500GB amount of hard drive space.

The sensors modules were checked for frequency response similarity as it is of outmost importance for the cross-correlation method which depends on signal matching. The sensor modules as seen in Figure 30, primarily consists of a sensor and an amplifier. Hence the frequency responses of the amplifiers were first measured experimentally. Unfortunately the speakers available in market were unable to broadcast signals which included signals higher than 20 KHz which are the characteristics of TES. The solution was to disconnect the jumpers which connected the individual microphones to their respective amplifiers except for module 2 which was used as an input source as shown in Figure 31. The output from the microphone at the input of the amplifier of module 2 was connected to the input of the amplifiers of module 1 and module 2. This therefore allowed the test of real TES signals effect to the amplifiers frequency response similarity. The designs for each of the amplifier circuit were identical. Hence the signals which were to be collected from ch1 through ch3 were supposed to be of the same magnitude and phase with each other. The sounds that were introduced to it were handscratch sounds and impulse sound such as pencil lead break. These signals represent the TES of long durations and short durations respectively. The schematic of the test is as shown in Figure 32.

Figure 32: Amplifier frequency response similarity test set up

The data collected from the DAQ, were then analyzed first using a simple visual method of Lissajous diagram where the output of 1 channel is used as the x axis while the output of the following channel was used as the y axis. If the signals were inphase, a 45° line would appear in the first quadrant.

Figure 33: Results handscratch (a) raw signal (b) Lissajous s1-s2 (c) Lissajous s1-s3 (d) Lissajous s2-s3

As can be seen in Figure 33(a) the signal is really noisy, quasi-periodic and non-stationary, despite that, the Lissajous for all the combinations for the sensors available were found to be inphase with each other. But close inspection finds the thickness of the line for Figure 33 (c) and (d) are much more than that of Figure 33 (b). This is most likely caused by difference in amplitude and not the phase. This was tested by simply subtracting the sensors of interest from one another. The results are shown as in Figure 34.

Figure 34: Magnitude difference (a) chan1 - chan2 (b) chan2- chan3 (c) chan4-chan3

It is noted that the magnitude difference is larger for the channel pairs which created slightly broader lines in the Lissajous diagram. Regardless, this small error did not affect the cross-correlation process in attaining the TDOA, as shown in Figure 33.

The impulse based sound pencil break which was conducted in the similar manner yielded results as shown in Figure 35.

Figure 35: Results pencil break (a) raw signal (b) Lissajous s1-s2 (c) Lissajous s1-s3 (d) Lissajous s2-s3

Similar response was garnered from the impulse based input. Proving the amplifiers were inphase. Hence they should yield accurate TDOA.

The next test was to verify whether the sensors were inphase. As the signals which exit the sensors were rather weak a direct method of evaluating the sensors was unavailable. The signals had to be amplified prior to being tested which introduced a probable source of error to the test results. With each sensor being equidistance from the sound source, the TDOA was expected to be 0 for all combinations. Hence a method as shown in Figure 36 was devised to verify that all the sensors modules used had the same frequency response.

Figure 36: calibration setup

It was expected that there would be high deviation due to acoustic reflections being picked up by different sensors at different times.

As it can be seen in Figure 36, four sensors were attached to a metal plate at equidistance from the hole in the center of the plate. The fourth sensor was a backup sensor and therefore its results are not discussed. An acoustic transducer which broadcasted a mixed acoustic signal comprising of (5 KHz, 8 KHz and 15 KHz) sine signals was affixed to the hole in the plate. The cross-correlation from each pair of sensors should yield 0s. Another method is by using Lissajous to detect phase shift by plotting two sensors output, one on the x axis and the other on the y axis. This yielded results as in Figure 37.

Figure 37: Results speaker mixed frequency (a) raw signal (b) Lissajous s1-s2 (c) Lissajous s1-s3 (d) Lissajous s4-s3

It can be seen that the Lissajous diagram consists of many lines in the general angle of 45 ˚. This shows the existence of phase angles other than 0°. This was further analyzed by breaking up the two million data points into segments of 1000 samples and cross-correlated creating a list of 2000 TDOA's to analyze the distribution.

Figure 38: Distribution plot of TDOA (a) S1 versus S2 (b) S1 versus S3 (c) S2versus S3

It can be seen that S1 versus S2 's mean is at 0us, while the mean of S1 versus S3 is at 1us and the mean of S2 versus S3 is at 2us. Based on this TDOA deviation, AOA would yield multiple intersecting points instead of a single point. Revisiting Figure 11 with the new information, we would see that the interactions are now different as shown in Figure 39.

Figure 39: AOA with TDOA errors

Figure 39, shows that the intersection lines have expanded into intersection cones due to the TDOA errors expected from the interaction of the TES with the microphones. Assuming a spread of 3us for each sensors pair, two sensor pair being used at any one time, produces nine possible coordinates. Of the nine possible coordinates, one represents the true coordinate. This yielded the probability that for every localized coordinate there was a 0.111 chance that it is correct. In addition to that, as the sound source moved further from the base planes of the sensors, the width of the cones increased causing the spread of the coordinates to worsen.

This phenomenon was simulated by assuming the worst case where the error from the cross-correlation was ±10us for the pre-defined conditions such as the arrangement and spacing of the sensors. Two ideal coordinates were chosen, (0.07, 0.08) m and (0.16, 0.16) m. They were then both added with the same normally distributed noise source with a range from -10us to +10us. These were then inserted into the AOA algorithm to be evaluated. The results attained were then plotted into a scatter plot as in Figure 40.

Figure 40: Scatter plot of two coordinates with TDOA noise

The plot 'correct1' represents the ideal coordinate of (0.16, 0.16) m while 'correct2' represented the ideal coordinate (0.07, 0.08) m. Plots 'spread1' tries to approximate upon 'correct 1' while 'spread2' tries to approximate to 'correct 2'. The area of intersection between the TDOA tainted AOA cones contained all the possible coordinates.

It was known that AOA posses some errors in it due to its approximation nature. But the presence of TDOA errors exacerbates the problem by creating a cloud of probable coordinates. When the sound source was near the base plane, the usage of mean of attained coordinates could be used to approximate the correct coordinates. Assuming normally distributed TDOA error, when the sound source was moved further from the base plane of the sensors, the spread of the coordinates increased. Despite the coordinates spread far apart from each other, the mean could still be used to approximate to the correct answer. The mean of 'spread 1' was found to be (0.17, 0.17) m while the mean of 'spread2' was found to be (0.07, 0.07) m.

An experiment in the hardware was conducted where the sensors were placed at coordinates (0.16, 0.16) m and (0.07, 0.08) m Figure 41.

Items(0.16,0.16)	x(0.16,0.16)	y(0.16,0.16)	T12(0.16,0.16)	T13(0.16,0.16)	T23(0.16,0.16)
Stdev	0.030	0.027	1.494	1.396	1.452
Median	0.254	0.242	109.000	21.000	-88.000
Mean	0.255	0.242	108.535	21.010	-87.525
Items(0.07,0.08)	X(0.07,0.08)	y(0.07,0.08)	T12(0.07,0.08)	T13(0.07,0.08)	T23(0.07,0.08)
Stdev	0.000	0.001	0.643	0.722	0.680
Median	0.067	0.085	46.000	9.000	-36.000
Mean	0.067	0.085	45.490	9.380	-36.110

Figure 41: Hardware test of localization error mixed signal source, locations (0.07, 0.08) m and (0.16, 0.16) m

75

Figure 41 shows the localization error which happens in a real-life set up using the recommended spacing between sensor elements and mixed frequency broadcast from an earphone. The results confirm what the simulation suggested would happen where the coordinate far from the base plane had its coordinates scattered widely while the coordinates near the base plane suffered less scattering. The error introduced to both the ideal coordinates of (0.16, 0.16) m and (0.07, 0.08) m were exactly the same but yielded different ranges of coordinate scattering in the simulations. In this case however, based on the standard deviation of the time delay, it shows the coordinate (0.16, 0.16) m suffered more severe TDOA errors as compared to the (0.07, 0.08) m coordinate. The actual coordinate was not contained within the cloud of coordinates calculated using the AOA method. In addition to that, the scattering of the coordinates of (0.16, 0.16) was way more severe as it was shown in the simulation while the scattering of the coordinates of point (0.07, 0.08) m was little as was expected though the simulation. Figure 41(b) and (c) show the difference between the AOA calculated with the actual coordinate for both coordinates distribution. The bins for the (0.07, 0.08) m are in the 3 decimal value while point (0.16, 0.16) m show bin values of the 1st decimal point. Such large deviations were expected as the signal used although mixed, it is to a certain degree periodic and stationary. As the subharmonic spacing of the sensors matched with the signal detected determines the quality of the attained TDOA, a periodic signal causes ambiguity. This proposal is tested out with a digitally generated white noise and played back through a speaker to produce results as in Figure 42.

Figure 42: Hardware test of localization error white noise signal source, locations (0.07, 0.08) m and (0.16, 0.16) m

In this particular where white noise was broadcasted, the mean for both the points were very close to the actual coordinate value. Both the actual coordinates were found

within the coordinates calculated using AOA. The standard deviation for the coordinates for both the points reduced, but was more noticeably for coordinate (0.16, 0.16) m. Interestingly though, the standard deviation for the TDOA however increased for both coordinates. Despite that, it was still within ± 3us which as the error margin of this device. This test also shows that the system was designed for white noise applications, and localizes upon such sounds very well as compared to periodic signals. The more stochastic the signal, the more unique it is, hence making it easier for the cross-correlation algorithm to produce an accurate TDOA.

To disambiguate the possibility that the cause was not the signal's frequency value which caused the problem for such sensor pairs, a subsequent test was conducted with the same setup. A signal was created using Audacity v2.0.3 which swept from 2 KHz to 20 KHz at 1KHz steps at a fixed duration for each step. The amplitude was set to 80% for all frequencies. The signal was then played back on the earphone where the Fourier Transform's magnitudes were recorded with the TDOA attained from the sensor pair. This signal was then compared to white noise play back at the same location (0.16, 0.16) m. The white noise was recorded for 100 TDOA samples and the results were repeated five more times as a comparison. The result of the test is shown in Figure 43.

The amplitude of the individual frequencies was set to at the same values. Despite that, the values detected by the sensors showed that they were different. This was due to the earphone and microphone's frequency response. As one can see, the earphone's frequency response gave the most amplification to the signal at 9 KHz. The amplitude of the signal does play a role in the stability and accuracy of the TDOA collected. Frequencies 4 KHz, 13 KHz, 14 KHz, 17 KHz and 18 KHz exhibited rather low amplitudes as compared to the other detected frequencies also showed an unstable TDOA representation. The TDOA, T12 (white noise) was the ideal data with an average of 109us. It can be seen from the graph, the TDOA attained by using periodic signal in spite of high magnitude and frequency is unable to achieve this value precisely and accurately.

Figure 43: Frequency sweep with resulting TDOA

4.2.   SOFTWARE

The hardware requires algorithms to operate. The algorithm in the end requires to be realized in software form. The flowchart of the said algorithm is shown as in Figure 44.

Figure 44: Flow chart of program

The initial part of the software requires the initialization of parameters from the DAQ and GLUT such as the sampling rate, screen resolution, screen positioning. This is then followed by the calling of the *Idle* function within the GLUT. This function will only be activated when there are no requests to the graphic card to render graphics. The

DAQ captures raw signals from the sensors and converts it into digital signals. These digital signals are then decoded to floating point format. Subsequently, the digital signals are then converted using Fourier Transform. Prior to the decision to localize upon the signal, the ratio of the high frequency to low frequency was calculated.

The digital signals captured are checked whether they are within the range of the minimum voltage or maximum voltage as previously mentioned in D-TES. These digital signals are then cross-correlated to produce TDOA, these TDOA are then fed to the AOA to attain the coordinates. The coordinates are then sent to the *GLUT* to be rendered. Both the hardware and software when merged together forms the complete device as shown in Figure 45.

4.2.1. TEST-BED SETUP

The basic idea of this prototype is to test the viability of using scratch sound as a computer input medium. The abilities and constraints as mentioned in the previously are to be tested with some simple tests. The most important feature is its ability to decipher and localize upon TES generated by tracing a finger (bare or covered) on human skin and on some random surface. Experimental hardware setup and the camera specifications are as in Figure 45 were configured based upon the evaluations done in the previous sections.

Figure 45: Hardware setup

The TES generated from the interactions of the finger with the active material in the workspace were caught by three spatially displaced microphones (SPM040LE5H) which were subsequently converted into electrical signals to be fed to the amplifier. It has to be taken into account that such acoustical signals are extremely small and therefore require large amplification. In this particular set up the authors used a 2.5k, 2 stage amplification to amplify the signals from the microphones which have already been pre-amplified within its SMD body. These amplified analog signals are then channeled to a 12 bit DAQ sampling at the rate of 1MSa/s which are then converted into digital data and processed by the algorithms written in the computer for localization. Derived localized points are then displayed and stored on the computer in real time. When the finger was being traced upon a surface, a camera (*Casio Exilim EX-F1*) was used to simultaneously take video of the action for accuracy verification. In addition to that, a graph paper was kept in its field of view (FOV) for scaling purposes. The FOV from the camera is shown in Figure 46. The settings of the camera were: 300 frames per second, manual white balance, and spot focus.

Figure 46: Visual Tracking post-processing

In this particular case in Figure 46, the surface used is made from denim material. The finger was marked with red and black to assist the software to better discriminate it from the background. Purple lines on the left and bottom mark the declaration of the y-axis and x-axis. The sensors could not be affixed at the exact coordinates previously calculated due to the limitations of the workplane panel, the angles were not preserved but the distances were kept as close to 0.054m as possible. The cumulative red arrows mark the passage of the finger at every frame detected by the software. The actual motion during contact with the cloth/work surface was a circle, while the additional lines were caused by the entry and exit of the finger into and from the FOV during non-contact times. This data was then used as the benchmark for comparison against data attained from the acoustic based localization. The surfaces prepared are, human palm (glabrous skin), cloth and book surface. As for the finger conditions, tests were done bare skinned or with covers (plastic or paper covers) as shown in Figure 47.

Figure 47: (a) Bare Finger (b) paper covered finger (c) plastic covered finger

Preparations are as follows, the author's finger was washed with hand soap. It was then dried with tissues and subsequently left to be air dried in the experimentation room with humidity and temperature of RH60% and 25.6℃ for 5 minutes prior to experimentation.

4.3. RESULTS

Experiments on the first prototype were conducted in multiple conditions as previously stated. Results of localization are shown as in Figure 48.

Figure 48:(a) Bare finger scratch palm – localization (b) Bare finger scratch book – localization (c) Bare finger scratch cloth – localization (d) Plastic covered finger scratch palm - localization (e) Paper covered finger scratch palm - localization

4.4. DISCUSSIONS

Figure 48 shows that, the objective of drawing a shape/circle using TES generated on different surfaces was possible despite the fact that the size of the resultant trace was smaller and shifted towards the origin as compared to the shape/circle visually obtained. Although the visually obtained coordinates are more accurate, it suffers from the inability to discriminate between a contact trace from a no-contact trace. This is evident in all the visually derived plots where extra 'tails' exists due to the entry and exit of the finger from the camera's FOV. During the entry and exit of the finger into the FOV, the finger did not contact the workplane surface, creating phantom coordinates with no TES data produced and captured. Figure 49 shows the comparison of time delay between the visual system and acoustic system where the shaded region on the graph indicates the region where the visual system detected the 'finger' but the acoustical system did not pick up any acoustic signals.

Figure 49: Visual time delay & acoustic time delay

The unshaded regions time delay graphs in Figure 49 have the similar shape, which supports the previous results that indicated the preservation of shape traced. This implies that despite accuracy issues, it is very likely that it can approximate letters/numbers and even gestures. Despite the similarities, the TDOA values attained acoustically are smaller than that attained visually by approximately 20-30us. The source of this error was unknown. An acoustic time delay attained was offset by +20us and re-plotted using AOA as shown in Figure 50, 'Acoustic +20us offset' plot exhibits similar characteristics to that of the visually localized plots.

Figure 50: Localization plots

This further supports the claim that the difference between the idealized localization to the acoustically attained output is mainly affected by the missing 20us. The shape wasn't exactly the same as the visually localized points as some TES did not meet the minimum requirements of D-TES and hence dropped from being considered, reducing the density of localized points. Further improvements needed to be done to increase the efficiency of D-TES to increase the data density.

The plane of the microphones was 0.03m higher than that of the workplane which could have been one of the sources of the missing 20us. Comparing the theoretical 3 dimensional TDOA with the theoretical 2 dimensional TDOA with said sensor positions to a sound source of coordinate (0.05, 0.06, -0.03) m, yields a TDOA12 of 10us and TDOA13 of 5us. As an approximate, the original acoustic TDOA was offset by 10us and re-plotted as shown in Figure 50. It was observed that the shape was retained while the size increased. Despite that, the size still did not match that of visually detected plot. This weakness was addressed in future works to gain higher precision and accuracy. The remaining missing 10us was unaccounted for and was assumed that it could be caused by the calibration of the camera video which did not take into account for lens curvature. Furthermore, best possible localization that could be achieved by using AOA method is shown in Figure 50 ' Visual AOA' plot, which is smaller and slightly shifted towards the origin as compared to the ideal, 'Visual actual' plot. This phenomenon was attributed to the AOA and was expected during the design phase.

4.5. IMPROVEMENTS

As the discussions have pointed out, the errors could either be from 1 of the 2 sources, the difference between the sensor plane to the workplane or the verification device used, the camera. By process of elimination, the authors decided that it was best to solve the problem regarding the mismatch between the sensor plane and the workplane in the next prototype. Instead of relying only on experiments to test many combinations in real-life which takes up a lot of time, simulations should be written to simulate the possible errors which can be caused by such mismatches and compared against that of the experimental values. The simulation results which were found to be promising are then

realized in real-life experiments. Not only would the time used to verification be shortened, but the combination of both simulation and experimental results would be more conclusive in measuring the effectiveness of the solution.

5. SIMULATIONS TO FIND ERRORS

Based on the observations of the basic prototype, a few parameters could be the cause of the inaccuracies. The normal approach to verify this was to conduct troubleshooting. Although this proven method is effective, it puts the prototype under risks of being damaged. Hence the approach used was to simulate such errors computationally. Candidate parameters which caused the most significant errors were then chosen to be improved in the prototype. This is then followed by actual experimentation of the prototype with the said upgrades for verification of the upgrade's effectiveness.

5.1. SIMULATIONS OF BASELINE

Prior to searching for prototype's assembly related caused errors to the localization, the errors contributed by the AOA was first taken into account as the baseline. Hence the first subsection discusses in detail and simulates the effect of the AOA on the localization of the coordinates.

With the contribution of error from the AOA being identified, the validity of possible errors caused by prototype's assembly factors were evaluated using simulations. They were namely the erroneous relative displacement of sensor's plane from the workplane or from each other. Another possible contributor to errors was the discrepancies between ambient temperature and the software declared temperature.

The evaluation method for the simulation consisted of a grid 0.01X0.01 m boxes arranged to create a 0.06 X 0.06 m box created by points spaced in 0.001m intervals. These ideal sound source coordinates were exposed to the simulated erroneous environment thus creating the resultant TDOA which was processed by AOA to attain the final coordinates. The resultant coordinates were compared against the ideal grid coordinates.

In this particular test, the ideal coordinate points were generated arithmetically. 834 pseudo-coordinates were   arranged to from a square grid as shown in   Figure 51. A software which emulated the travel of sound through the atmosphere prior to reaching

the individual sensors calculated the TDOA between the sensors on a homogenous flat plane for a sound source on the same plane. The resultant TDOA calculated data were then inserted into the AOA algorithm to produce a set of coordinates for evaluation. The results of simulation yielded the data as shown in   Figure 51.

Figure 51: Results - AOA (baseline) versus Ideal

It can be visually seen that the AOA method is able to approximate the ideal data in terms of shape, albeit imperfectly. There exists an 'error free region' area which approximates the ideal grid coordinates closely as compared to the other regions of the calculated grid. Also, the error trend in coordinate displacement and angle deviation becomes more pronounced as the coordinates re-created are further from the 'error free region'. This could be explained by the fact that AOA is an angled based localization system. Any small angle deviations become more pronounced as the distance from the sensors increases. In addition to that, the error trends tend to reverse as they past the 'error free region'. For example, if the coordinate re-created has an erroneous positive offset prior to the 'error free region', after that it would have an erroneous negative offset. Closer inspection revealed that the 'error free region' is a coordinate which is equidistance from all three sensors. This fact was used advantageously in subsequent sections.

89

5.2. SIMULATIONS : EFFECT OF SENSOR DISPLACEMENTS

This particular test was to evaluate the effect of sensors in a homogenous plane being displaced from the workplane in the z direction and also the displacement of individual sensors in the xy-plane which are not accounted for by the software. Such undeclared displacements could cause TDOA errors which in turn result in localization errors when they are fed into the AOA equation. It has to be noted that AOA weakness does not in any way affect the quality of the attained TDOA.

5.2.1. SENSOR PLANE OFFSET IN THE Z-PLANE

This simulation utilized the ideal pseudo data which was then processed to produce TDOA. Different z-axis displacements between the sensor plane and workplane conditions were simulated. The displacements were in the increments of 0.01m starting from 0.01 to 0.03m. The TDOA were re-calculated at each increment for the same xy-coordinates and then introduced to the AOA algorithm for evaluation. This simulation was chosen as it simulates the effect of the using this prototype on different flat and homogenous surfaces which might be in relative state of offset from the sensor plane. The results of the simulations are shown visually in Figure 52(a)-(c).

It is visually evident from the visual results that the grid generated by the AOA localized points via the captured TDOA shrinks in comparison to the ideal grid coordinate as the plane of the sound source is moved further away from the sensor plane. The 'error free region' is strangely not affected by the z -offset. This phenomenon can be utilized as a calibration spot for comparison between the ideal TDOA and the experimental TDOA (collected from the prototype) of the same shape traced. If the TDOA of the experimental data is the same to that of the ideal TDOA on the 'error free region', but the extremums are different, it can be deduced that a z offset of the sensor plane in relation to the workplane could have occurred in the prototype's set up.

Figure 52: Z displacement simulation (a) Z displacement 0.01m (b) Z displacement 0.02m (c) Z displacement 0.03

5.2.2. INDIVIDUAL SENSOR OFFSET (XY)

In this particular test, it was assumed that the error of translocation of the cumulative coordinates to produce a grid was caused by the unaccounted shifting of sensors along the x-axis or y-axis. The set up had three sensors, and in this simulation, it assumed that only one of the sensors were susceptible to erroneous shifting in the x or y axis of 0.005m per shift, the results are shown as in Figure 53.

Figure 53: Simulated shapes due to displaced sensor 2 (a) x +0.005m (b) x-0.005m

(c) y+0.005m (d) y-0.005m

The grids re-created by the shifting of the sensor in its x or y position were very different from that of the ideal grid in terms of size and shape. The coordinate of the 'error free region' in this particular test was also affected which would indicate that the TDOA trend between the ideal and that of the experimental at this point did not hold true anymore. As a conclusion, any offset of sensors whether in the x or y axis is highly detrimental to the final outcome of the localization system.

5.3. EFFECT OF TEMPERATURE

Temperature affects the speed of sound. This system relies upon the propagation delay of sound waves moving though the air. This therefore indicated that, the ambient temperature of air intuitively could be a factor which affected the accuracy of the system.

The simulation software was configured so that the AOA would localize with the resultant speed of sound at 26 °C while the simulated environment was set to varying temperatures of 26°C, 16°C and 6°C. The temperature difference simulation tried to simulate the effect of the user using the AOA system without calibrating it to the current environmental temperature. The results of the simulated prototype are shown as in Figure 54.

Figure 54: Effect of Temperature mismatch

It can be seen from visual inspection that the effect of temperature is less as compared to that of plane displacement and sensor displacement. Also evident is that despite AOA's exposure to large steps in temperature mismatch, the differences in results between each temperature step was low. It can therefore be assumed that the expected errors calculated should be lower. Similarly with that of the z displacement, it was expected that as the displacement increased so would the localization errors, strangely the larger the difference the temperature, the closer in which the cumulative localized points approximated that of the ideal. Closer observation yielded the fact that the temperature gradient is inversely proportional to the squared error of the localized points. This strangely opposing idea of inverse relationship between errors to increasing temperature gradient can be explained by the fact that, the increase in temperature difference caused

the TDOA to be higher than the ideal therefore increasing the size of the shape which in turn opposed the effect of the AOA which shrank the shape. The temperature in general had little effect to the final outcome of the shapes or the accuracy of the localized points. Hence it was concluded that temperature in which humans find comfortable operating this system in would not affect the final outcome of the localization points.

5.4. EXPERIMENT AND DISCUSSIONS OR PROTOTYPE

The observations seen in the simulations were verified in this section by comparing the captured the coordinates of the moving finger with the help of a high speed camera affixed perpendicularly to the workplane. The results of the acoustically captured coordinates versus that visually captured in various settings are shown in Figure 55.

The result of eliminating the environmental factors causing errors produced the graph as shown in Figure 55 (b) which indicates very close approximation to the visual data. Simulation results of Figure 53(a) and experimental results in Figure 55 (c) which share similar key settings which is the offset of sensor 2 by 0.005m in the x axis, shows similar characteristics such as shape mismatch, and shape offset. Similarly, simulation conditions were re-created and tested for the condition where the sensor plane was offset by 0.03m from the workplane produced results as shown in Figure 55(d) which also showed same characteristics as the graph in Figure 52(c) where the shapes re-created were smaller than the ideal version.

Extreme care had been taken in setting up the camera for the z-axis distortion experiment as any slight error will be compounded and cause the 'ideal' data to be erroneous. Experiments related to temperature were not done as the simulations indicate that its effects were negligible. Nevertheless, future improvements such as inserting a temperature sensor into the design which can measure the ambient temperature, simultaneously updating the software could improve the accuracy of the system when used in extreme temperature conditions.

94

Figure 55: (a) set up (b) improved experiment, (c) sensor2 offset in x axis by +0.01m (d) sensor plane offset by +0.03m

5.5. IMPROVEMENTS

This chapter successfully simulated, identified, verified, and improved the errors contributed by prototype's environmental setting factors. As a conclusion, conditions such as sensor plane displacement contributes to size change of the re-created shape which is still recognizable, while sensor displacement causes shapes to be deformed greatly to the point recognition could be difficult, and lastly the temperature mismatch causes little or no effect to the final shape re-created. AOA contributes significant errors to the localization despite being fast. This accuracy issue therefore needs to be improved in the next chapter but at the same time retain the fast processing capabilities inherent to the AOA.

6. LOCALIZATION ACCURACY IMPROVEMENT ALGORITHM

As the previous chapter has highlighted, the AOA is inaccurate as it is an approximation equation. Hence if we were to apply gradient descent method onto an approximation equation, we would be unable to attain the accurate answer as the equation itself is flawed. To improve accuracy, the equation has to be crafted carefully by taking consideration of the limitations in the real world such as available hardware. This chapter focuses on creating and verifying accurate, fast and practical equations for localization.

6.1. EQUATION FORMULATION

The gradient method is an optimization method for a carefully crafted equation. Hence looking back at the problem first posed in Figure 10. Merging (1) and (2) for all sensor combinations yield the equation (15).

$$(t_{12} \ . Vs) - (R_1 - R_2) = (t_{12} \ . Vs) - (R_{12}) = 0 \qquad (15)$$
$$(t_{13} \ . Vs) - (R_1 - R_3) = (t_{13} \ . Vs) - (R_{13}) = 0$$
$$(t_{23} \ . Vs) - (R_2 - R_3) = (t_{23} \ . Vs) - (R_{23}) = 0$$

yields

$$|(t_{12} \ . V_s) - [((x_1 - x_s)^2 + (y_1 - y_s)^2)^{1/2} - ((x_2 - x_s)^2 + (y_2 - y_s)^2)^{1/2}]| + |(t_{13} \ . V_s) - [((x_1 - x_s)^2 + (y_1 - y_s)^2)^{1/2} - ((x_3 - x_s)^2 + (y_3 - y_s)^2)^{1/2}]| + |(t_{23} \ . V_s) - [((x_2 - x_s)^2 + (y_2 - y_s)^2)^{1/2} - ((x_3 - x_s)^2 + (y_3 - y_s)^2)^{1/2}]| = 0$$

There were two unknowns with three balancing equations in this particular case. The velocity of sound is a constant within the equation. Environment's temperature could vary making this assumption false. This could then be a source of errors for the system. Temperature from (15) was removed by means of dividing the equation (15) with each other providing (16).

$$(t_{12} \,.\, Vs) - (R_1 - R_2) = (t_{12} \,.\, Vs) - (R_{12}) = 0 \qquad (16)$$

$$(t_{13} \,.\, Vs) - (R_1 - R_3) = (t_{13} \,.\, Vs) - (R_{13}) = 0$$

$$(t_{23} \,.\, Vs) - (R_2 - R_3) = (t_{23} \,.\, Vs) - (R_{23}) = 0$$

yields

$$\frac{(t_{12}.V_{\bar{s}})}{(t_{13}.V_{\bar{s}})} = \frac{(R_{12})}{(R_{13})}$$

$$\frac{(t_{12}.V_{\bar{s}})}{(t_{23}.V_{\bar{s}})} = \frac{(R_{12})}{(R_{23})}$$

$$\frac{(t_{13}.V_{\bar{s}})}{(t_{23}.V_{\bar{s}})} = \frac{(R_{13})}{(R_{23})}$$

yields

$$|[((x_1 - x_s)^2 + (y_1 - y_s)^2)^{1/2} - ((x_2 - x_s)^2 + (y_2 - y_s)^2)^{1/2}] \,.\, t_{13} - [((x_1 - x_s)^2 +$$
$$(y_1 - y_s)^2)^{1/2} - ((x_3 - x_s)^2 + (y_3 - y_s)^2)^{1/2}] \,.\, t_{12} | + |[((x_1 - x_s)^2 + (y_1 - y_s)^2)^{1/2} - ((x_2$$
$$- x_s)^2 + (y_2 - y_s)^2)^{1/2}] \,.\, t_{23} - [((x_2 - x_s)^2 + (y_2 - y_s)^2)^{1/2} - ((x_3 - x_s)^2 + (y_3 - y_s)^2)^{1/2}]$$
$$.\, t_{12} | + |[((x_1 - x_s)^2 + (y_1 - y_s)^2)^{1/2} - ((x_3 - x_s)^2 + (y_3 - y_s)^2)^{1/2}] \,.\, t_{23} - [((x_2 - x_s v$$
$$+ (y_2 - y_s)^2)^{1/2} - ((x_3 - x_s)^2 + (y_3 - y_s)^2)^{1/2}] \,.\, t_{13} | = 0$$

Both, (15) and (16) should have reduced to 0 if the guessed x_s and y_s were the correct coordinates. Hence by visual inspection, the equations seemed convergeable.

6.2. VERIFICATION

The main purpose of this section was to test the ability of the said equations to converge to a global minimum. This particular mathematical problem is represented as in Figure 56. The vertical and horizontal arrows represent the x and y axes while the U represents the universal set which comprises of all the feasible coordinates which can exist within the x-and y-axes. Meanwhile Q represents the subsets which exist within the universal set. The elements that exists throughout the entire set is t_{12}, t_{13}, t_{23} respectively. $t^*{}_{12}$, $t^*{}_{13}$, $t^*{}_{23}$ are the element which yields the lowest results in the subset Q.

Figure 56: Set of feasible regions

Figure 56 yields (17) which define the local minimum.

$$f(t_{12}, t_{13}, t_{23}) \geq f(x^*_s, y^*_s) \, , \, (t^*_{12}, t^*_{13}, t^*_{23}) \in Q \quad , \, (t_{12}, t_{13}, t_{23}) \in U \qquad (17)$$

The point $f(t^*_{12}, t^*_{13}, t^*_{23})$ is not unique as there exists a possibility that $f(t_{12}, t_{13}, t_{23})$ can be the same. If the equality is removed, the equation converts to a strict local minimum. Since the global minimum is the objective of this search, the search for minimum values with the same rules within Q is instead expanded to the size of U. Therefore the $f(t^*_{12}, t^*_{13}, t^*_{23})$ will become the strict global minimum as shown in (18).

$$f(t_{12}, t_{13}, t_{23}) > f(t^*_{12}, t^*_{13}, t^*_{23}) \qquad , \, (t_{12}, t_{13}, t_{23}) \in Q \, , \, \lim_Q U \qquad (18)$$

The following experiments will have the U defined as x_s and y_s [0>0, 20) meters.

6.2.1. OFFLINE TESTS OF EQUATION'S CONVERGEABILITY TO GLOBAL MINIMUM

The typical method for testing the quality of convergence of an equation is to numerically minimize the function. This is a direct and simple method, but it does not offer an intuitive feel of the error plane which is important to the user to judge on reasons why the equations are not minimizing toward the global minimum. The area of interest was declared as all coordinates within 0 m to 0.20 m of both the x and y axis in this chapter. Since the area was relatively small, it was therefore practical to propose the evaluation of the error functions as in (15) and (16) for the contained coordinates. A set of pseudo-coordinates of imaginary sound sources were generated with their complimenting theoretical TDOA's between microphones at an assumed temperature of 26 ℃ for this particular set of tests. The set of pseudo-coordinates comprised of 396 individual coordinates when displayed on the scatter plot resembles a circle as shown in Figure 57.

Figure 57: Pseudo ideal coordinates (Ideal)

The accompanying TDOA is as shown as in Figure 58.

Figure 58: TDOA - pseudo ideal coordinates

Only the coordinate's corresponding set of TDOA is introduced into the error function which is evaluated within the area of interest at 0.0001m steps to create a 3-D error mesh plot. In the meantime, the program will also search for the smallest error and location within each 3-D mesh plot to be compiled in the lowest-error location table. This processes was repeated for all the sets of pseudo TDOA's. Hence each pseudo-coordinate will generate its own 3-D mesh plot and a table of lowest-error location. The location of the lowest error point defines the most likely coordinate from which the pseudo sound source was generated from as shown in Figure 59.

Figure 59: Error mesh plot - single set of TDOA

This test data was used numerous times in subsequent experiments related to simulations. The theoretical TDOA sets calculated were represented in micro-seconds (us) and were divided into two types, the integer type and the decimal type. The decimal type basically has six digits after the decimal point which represents the precision of 1×10^{-12} while the integer type has a precision of 1×10^{-6}.

6.2.1.1. WITHOUT SOUND VELOCITY

The simulation as described in the previous section was used with (16) as the minimization error function. The TDOA of decimal type pseudo ideal coordinates were fed into the simulation software to calculate the errors. The mesh plot of pseudo ideal coordinates point 0 is shown as in Figure 60.

Figure 60: 3-D wire mesh plot of point 0 (3 sensors)

The x and y axes represent the Cartesian coordinate system in which in the area of interest was tested. The z axis represents the error. In addition to that, the colours follow the jet colouring convention where red is represents high value of error and blue represents low value of error. It can be observed that the error plane created a trench instead of an expected inverse cone. This implies that many solutions which seem to be right might be generated from such an equation as in (16) if it were to be used in conjunction with gradient descent methods. An aerial view of the 3-D wire mesh plot for randomly chosen points 0, 65,131,197,263,329 were generated and evaluated as shown in Figure 61.

Figure 61: Contour plot for selected points (three sensors) (a) point 0 (b) point 65 (c) point 131 (d) point 197 (e) point 263 (f) point 329

Corresponding pseudo ideal coordinates were also indicated as white circles and its values labeled on each contour map. Notice that the pseudo ideal coordinates were always within the trough of contour map. This indicates that the existence of the minima point within the equation is very likely. Nevertheless there might be many local minima's as the trough indicates many solutions which would lead the error function astray. To confirm this suspicion, the location of the lowest error value table was extracted from

each 3-D mesh plot for all 396 points to be plotted against the ideal pseudo-coordinates in a scatter plot graph to produce Figure 62.

Figure 62: Lowest error - derived coordinates (three sensors)

Despite having a fine resolution of 0.0001m, the search for the lowest error value yielded very poor localization results. Hence the equation fell into the category of an underdetermined system despite having more equations than unknowns. Careful re-analysis of (16) showed that despite the clever method of making the equations independent of the velocity of sound, the number of independent equations were reduced to a smaller number than that of the variables. Having said that, an underdetermined system has infinite number of solutions and although the ideal solution exists it cannot be clearly differentiated from the other solutions. This will also pose a problem of determining the uniqueness of the solution attained. The solutions showed in Figure 62 are local minimum points which were not strict global minimum points which were required.

The number of independent equations had to be more than the number of unknowns to solve this problem. A possible method was to increase the number of sensors from the

current three to four sensors. Figure 63 illustrates the uniqueness of the global minimum which exists such an error function.

Figure 63: 3-D wire mesh plot of point 0 (4 sensors)

The point indicated by the global minimum in Figure 63 implies that this equation provides a unique coordinate for this set of TDOA. The 2-D contour plots for a sampled number of TDOA were then evaluated as evaluating every single plot was impractical. Points 0, 65,131,197,263,329 were plotted as shown in Figure 64.

Figure 64: Contour plot for selected points (4 sensors) (a) point 0 (b) point 65 (c) point 131 (d) point 197 (e) point 263 (f) point 329

In addition to using sample data of contour plots, the coordinates of the lowest error attained from the mesh plots were plotted in the same graph as the pseudo-coordinates to produce Figure 65. This was also done in Figure 61.

Figure 65: Lowest error - derived coordinates (four sensors)
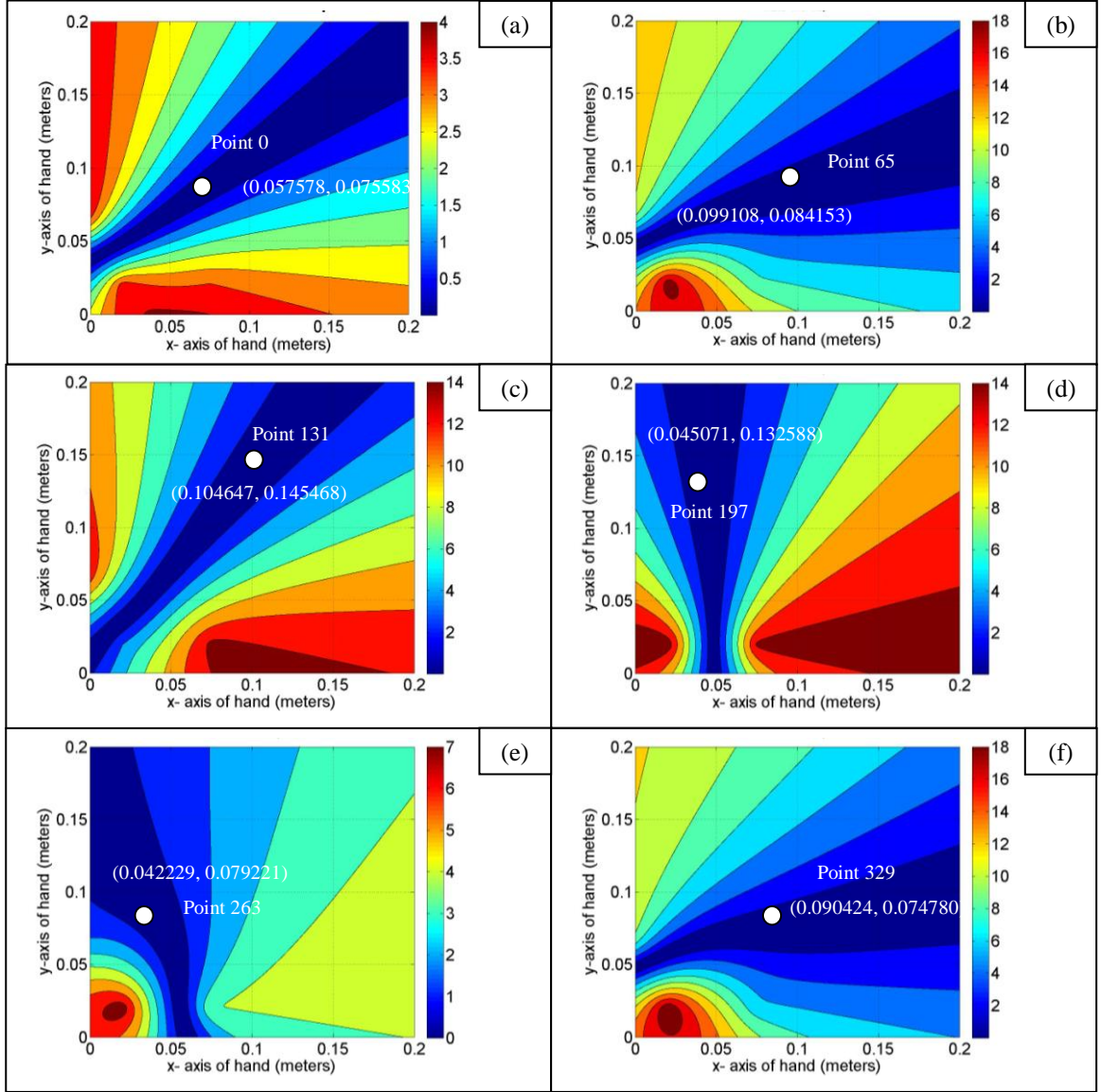
Figure 65 shows the ability of the error plane software to correctly detect the location of the lowest errors when utilizing four sensors. Both the pseudo ideal coordinates and the lowest points were overlapping each other which strongly proved the convergeability and accuracy of the method. This indicated that minimization methods could work upon the derived equation to solve for the location of sound source. Nevertheless, the usage of another sensor although beneficial to the convergeability of the algorithm, was deemed disadvantageous to mobility of the device due to the additional hardware, mass, volume and energy consumption.

It was therefore proposed the usage of (15) where the velocity of sound was assumed constant within the calculations, but in reality it fluctuated. This caused irregularities in the results calculated. It was therefore evaluated in detail to wight its negative and positive effects it has to the system being developed.

6.2.1.2. WITH SOUND VELOCITY

The simulation temperature was set to 26℃ and 0℃ where 26℃ was the same temperature in which the pseudo data were created in. The results are as shown in Figure 66 and Figure 67.

Figure 66: Lowest point plane 26 ℃ (a) point 0 (b) point 65 (c) point 131 (d) point 197 (e) point 263 (f) point 329

Both Figure 66 and Figure 67 do not differ much visually. Instead they look quite similar with dark blue points indicating the existence of a minima point.

Figure 67: Lowest Point 0 ℃ (a) point 0 (b) point 65 (c) point 131 (d) point 197 (e) point 263 (f) point 329

The mesh plot coordinates and the pseudo ideal coordinates were plotted in the same graph and two things become clear, the first being that the equation can converge under such conditions and second, the 0 ℃ converges to form a smaller shape as initially defined by the ideal data. This can be explained by the fact that (15) depends on the velocity of sound multiplied with the TDOA to balance out the physical distance difference between the sound source to sensor sets. Sound velocity is directly proportional to temperature. Hence, when the temperature is declared lower, the velocity

of sound is lower causing the calculated distance difference between the sounds source to sensor sets to be lower, to balance this; the equation accepts the location which yields smaller difference between sound source to sensor sets. This makes the expected sound source to be close to the base planes of the sensors as shown in Figure 68.

Figure 68: Lowest error - coordinate points (0 and 26) ℃

The evaluation was made more detailed by comparing each point attained via the lowest error of the mesh plot coordinates with the pseudo ideal coordinates as shown in Figure 69, maximum error due to a 26 ℃ mismatch is approximately 0.02m. This will become an issue and the error would be quite large if the user is totally unaware of a 26 ℃ difference in temperature.

Figure 69: Difference between coordinates of 0 and 26 ℃

To avoid this large disparity, the user will be required to set the approximate temperature into the system. Temperature differences of 5 ℃ will cause an approximately 0.004m error which would be acceptable.

6.2.2. OFFLINE TESTS TO VERIFY THE ACCURACY OF EQUATIONS USING PLANE SOFTWARE

It was established in the previous tests that the equations were able to converge using pseudo ideal coordinates TDOA sets which were represented in micro seconds of six decimal point precision. The sampling rate of the actual hardware that was designed was instead represented in micro seconds of integer format. This low precision might cause low accuracies which might lead to no convergence of the equation. The experiment will therefore focus on the convergeability, followed by the accuracy by comparing it to the ideal pseudo-coordinates. This was therefore verified using the same pseudo TDOA utilized in previous experiments. The slight difference was that the TDOA data sets were rounded up from six decimal to integer format of zero decimal. These rounded up TDOA sets were then introduced to the plane software with the same setting described earlier in the main section.

The results of the experiment is as shown in Figure 70 where the dark blue surrounded by other brighter colors indicates an inverse conical mesh plot with a higher chance for a single convergence.

111

Figure 70: Plane lowest point – convergeable (a) point 0 (b) point 65 (c) point 131 (d) point 197 (e) point 263 (f) point 329

The actual lowest error point for each TDOA sets indicating the existence of unique actual coordinate was plotted against the pseudo ideal coordinates as shown in Figure 71.

In addition to that the coordinates attained from the lowest error plane was compared against that attained using velocity of sound at 26 ℃ which is the ideal data for the experiment. These differences between pseudo-coordinates versus that generated from TDOA with integer level precision were small as shown in Figure 72.

112

Figure 71: Effect of sampling rate

The coordinates attained from the lowest error of the mesh plane was compared to that of the pseudo-coordinate using absolute squared error to produce a graph as shown in Figure 72.

Figure 72: Absolute error - sampling error effect

Despite differences exhibited, the cumulative errors of both x and y were below 0.004m which is really low considering the fact that the precision used for the mesh plot

calculation was 0.0001m. The error could have been smaller if smaller steps were used. Comparing with the experiment before, temperature has higher influence to the accuracy of the coordinates as compared to the precision of the TDOA. But both temperature and sampling rate used did not affect the convergeability of this algorithm. The cost of increasing the sampling rate from 1Msa/s to 1Tsa/s was too high that it out weighted its accuracy contribution.

Hence, the usage of the hardware of lower sampling rate, 1Msa/s was justified in terms of accuracy versus costs.

6.2.3. OFFLINE ITERATIVE METHOD

Up until now, the tests utilizing mesh plots had successfully verified that the set of equations utilizing three sensors and velocity of sound was tentatively convergeable, producing good localizations. In spite of temperature mismatches and reduced sampling rate. The equation in this particular section was modified to be implemented as a minimizing function where the coordinates were attained numerically through a process called gradient descent method. The modified equation is written as in (19).

$$\|(t_{12}. V) - (R^i{}_1 - R^i{}_2)\| + \|(t_{13}. V) - (R^i{}_1 - R^i{}_3)\| + \|(t_{23}. V) - (R^i{}_2 - R^i{}_3)\| = F(x^i{}_s, y^i{}_s) \qquad (19)$$

$$x^{i+1}{}_s = x^i{}_s - \alpha \nabla F(x^i{}_s, y^i{}_s)$$
$$y^{i+1}{}_s = y^i{}_s - \alpha \nabla F(x^i{}_s, y^i{}_s)$$

The velocity of sound is a constant within the equation. This velocity is the speed of sound within the assumed homogenous properties of air shared by the three sensors. α is the step size for each correction of $(x^i{}_s, y^i{}_s)$.

Utilizing the guessed values of coordinates $x^i{}_s$ and $y^i{}_s$ at each iteration, the output of the error function $F(x^i{}_s, y^i{}_s)$ is evaluated. If this error value is higher than a user defined value, the system will try to guess the next improved coordinates' $x^{i+1}{}_s$ and $y^{i+1}{}_s$ values by using the gradient ∇ of the function $F(x^i{}_s, y^i{}_s)$. This process will continue until the error value set by the user has been achieved or when the number of maximum iterations set by the user has been reached. The target error was defined at 0.0001m while the maximum iteration was set to 800.

114

This section tested the ability of the modified equation to self-correct its erroneous initial guess of the sound source location. If the modified equation was able to yield correct sound source location despite being supplied with the initially erroneous sound source location with corresponding sets of correct TDOA's, then the modified equation was deemed to be effective.

This particular simulation set up utilized the pseudo-coordinate as the ideal coordinate. A set of normally distributed offsets which were independent of each other were generated and added to the pseudo-coordinate to produce the noisy coordinates. The TDOA of both the pseudo-coordinates and noisy coordinates were the same, with the only difference being their coordinates. The minimization was configured to stop its iterative process when the minimizing error function produced a value below 0.0001 m.

The noisy coordinates were then introduced to the iterative process of the gradient descent method, and the results yielded are as shown Figure 73. The noisy data is very obviously scattered in the x-y workplane while the corrected data plots are obscured by the overlapping pseudo-coordinates.

Figure 73: Scatter plot - ideal, noisy, and corrected

Figure 74: Absolute error (noisy coordinates)

Figure 74 displays the absolute error between the pseudo ideal coordinates versus the noisy data. The absolute average error for both x and y was 0.0078m.

Figure 75: Absolute error (corrected coordinates)

Figure 75 displays the absolute error between the ideal coordinate and the corrected coordinates for both x and y. The average absolute error on x axis after correction was 0.00062m while the absolute error on y after correction was 0.00072 m. On average, the absolute error of both x and y after correction was 0.0009 m which was much smaller

than its initial value of 0.0078m. This was a reduction of 88%. This proved that the equation thus far was able to correct for initially supplied positional errors.

Figure 76: Minimization error versus coordinate point

Despite that, the number of iterations required to achieve the target minimization of 0.0001 m was sometimes very high as shown at coordinate point 282 in Figure 77 which required 13342 iterations. The average number of iterations per coordinate point was about 1256 iterations.

Figure 77: Iterations to correction

Analysis of point 282 found that the reason that it took so many iterations to finally reach its global minimum was due to the fact that the initial guess position was far away from the final position. These was verified by changing the different values of initialization points for the associated TDOA and watch the number of iterations needed

to achieve the minimization error of 0.0001 m. This data is shown as in Table III where as it can be seen as the difference of the positional difference increases, so does the number of iterations required for the equation to reach the corrected answer.

Table III: Initial positions versus iterations

Minimization error (m)	Initial - X(m)	Initial - Y(m)	Corrected - X(m)	Corrected - Y(m)	Iterations	Positional Difference(m)
0.0001	0.045	0.070	0.052	0.072	2836	0.007
0.0001	0.0751	0.070	0.052	0.072	4036	0.023
0.0001	0.0751	0.104	0.052	0.072	6883	0.040
0.0001	0.2493	0.104	0.052	0.072	37416	0.199

Offline processes can afford the luxury of large number of iterations, but if this algorithm were to be implemented in real-time, the number of iterations would translate to long delay time. For example, if each iteration requires 1ms, an average of 1256 iterations would translate to approximately 1 second to attain per dot (coordinate) on the workplane. This was not deemed practical. Despite increasing the accuracy of the sound source localization, the delay had increased drastically. Analysis of the data showed that this was due to the large displacement of the guessed location from the real sound source location as shown in Table III. This resulted in the time taken for the equation to iteratively reach the real location was long. The AOA could quickly produce an approximate answer. Meanwhile, the iterative method utilizing gradient descent method onto hyperbolic localization was very accurate. But unfortunately suffered a drawback where it was slow to converge and was sensitive to initial values. It was therefore suggested that the two localization algorithms should be merged. The merger of these two would result in a GDM's sensitivity to initial values to be eliminated as the AOA would produce initial point which are close the final answer. This in turn reduces the time to convergence of the GDM as it was seen in Table III. In addition to that the accuracy problem faced by AOA can be mitigated by the GDM's ability to iteratively approximate the erroneous initial answer to the final accurate answer as seen in Figure 73.

6.2.4. OFFLINE HYBRID METHOD

The previous set up has proven that the gradient descent was able to correct for initial positional errors, but the larger the error the, longer it takes for it to converge onto its final target. By utilizing the AOA together with the gradient descent method, it was hoped that the average number of iterations per point would be reduced.

Figure 78: 3D Error mesh plot of one set of TDOA (t12=25us, t13=3us, t23=-23us)

The Gradient Descent Method (GDM) used was accurate but slow. The equation was designed to have a prominent single global minimum. Figure 78 shows the error mesh plot derived from the error function (19) for the TDOA set t12 = 25us, t13= 3us, t23= -23us for an evaluation range of 0.20m (x-axis) by 0.20m(y-axis) with an incremental step of 0.0001m.

The correct coordinate would yield the lowest error as seen in the mesh plot. As the correct coordinates are not known, the GDM equation (19) used alone could randomly start at a coordinate far from the correct coordinate resulting in a large number of iterations required to reach the correct coordinate.

This sub-chapter proposed the merging of the AOA and GDM methods whereby the AOA calculates the first guessed coordinate, which was an approximation of the correct coordinate, which was then supplied to the GDM, which refined the solution via multiple iterations as illustrated in Figure 78.

The TDOA generated from the pseudo ideal coordinates were used to calculate the AOA in this particular set up. This was then followed by evaluating the said calculated AOA as the initial points together with the TDOA. Accuracy of the algorithm was evaluated by comparing the AOA evaluated coordinates against the gradient descent method corrected AOA coordinates. Besides the accuracy of the algorithm, the overheads in terms of number of iterations required to achieve such accuracies were also considered.

396 AOA derived data points and their corresponding TDOA's were introduced to the algorithm producing outputs as shown in Figure 79. As expected, the pseudo ideal coordinates (ideal) overlapped the Corrected AOA data which also indicated that the corrected data was very similar to that of the pseudo ideal coordinates which should be around 0.0001m difference. The AOA method by itself on the other hand displayed an expected weakness where it utilized exact same TDOA but reproduced a smaller geometry especially when further away from the base planes.

Figure 79: Corrected AOA output

Nevertheless, data shows that this error from the AOA could easily be corrected to be within acceptable error limits of 0.0001m. The number of iterations required to achieve this is as shown in Figure 80.

Figure 80: Iterations for hybrid algorithm

The average number of iterations required per point was 1029. Which were approximately 200 iterations less per coordinate point. This supported the idea that by placing the initial guessed coordinates close to the actual answer, the number of iterations required would be less. Strangely though, despite the general number of iterations to achieve a minimum error of 0.0001m reduced, certain coordinates despite their close proximity of the initial guessed coordinate via AOA to the ideal coordinates, required a disproportionately large number of iterations to achieve the same minimum error of 0.0001m. They are marked in dotted boxes in Figure 79 and Figure 80. The most likely cause for this is high learning rate which forces the guessed coordinates to be oscillating around the ideal coordinate and not being able to converge. The algorithm was executed with its learning rate halved to prove this idea, producing Figure 81. This idea managed to reduce the average iterations required per coordinate point to 807.

Figure 81: Iterations for reduced learning rate of hybrid algorithm

Comparing Figure 80 with Figure 81, the coordinate points which required highest iterations in Figure 80 required the least iterations in Figure 81 to achieve accurate results. This clearly proves the initial assumption that the points were very close to the actual answer took longer time to converge was due to the large learning rate was true. On the flip side, if the initially guessed position was very far away from the answer, using a small learning rate would also cause the hybrid algorithm to require many iterations to finally converge. There isn't a perfect learning rate hence a compromise was made where the learning rate varied accordingly. In most iterative strategies, the learning rate was normally set high as the initially guessed data is normally arbitrary and could be quite far away from the convergence point. The learning rate was first set to low rate as it was being assisted by the AOA, therefore the probability that the correct answer would lie near the initial guess was high. For cases where the AOA's guess was far away from the actual answer, the initially set low learning rate would take a very long time to converge, hence a method involving the doubling of the learning rate every 200 iterations was added to the algorithm. A maximum iteration limit was set at 3000 iterations to avoid run away conditions where the calculations oscillated forever. The results of the test are as in Figure 82.

Figure 82: Iterations per Coordinate (step alpha reduction) hybrid

The average number of iterations per coordinate was further reduced to 495. Nevertheless despite that, there were some coordinate points which were unable to converge within such constraints. The maximum limit of iterations was reduced from 3000 to 800, with the points which required 800 and more iterations discarded, the total number of average iterations per coordinate point were 373 as shown in Figure 83.

Figure 83: Iterations per Coordinate (step alpha reduction) hybrid max limit

The hybrid method was proven to be faster in convergence. Hence, the TDOA and AOA calculated coordinates were collected using the AOA system previously developed Figure 11 [52] to be introduced to this hybrid algorithm offline.

The visual data was used as the ideal data was plotted together as shown in Figure 84.

Figure 84: Corrected real acoustic AOA

This data shows that the acoustic data had fewer samples as compared to the ideal data which was visual in nature. This sampling rate difference was one of the main factors which inhibited direct point-to-point comparisons which left the accuracy evaluation to be done visually. The visual inspection showed that the final outcome between the corrected AOA coordinates and visually attained coordinates were very close. One coordinate point was removed from the plot as its iteration rates reached 800.

6.3. ONLINE HYBRID METHOD TESTED ON WORKPLANE

Method proposed in the previous test was deemed to be promising as it could achieve high accuracies of localizations within a small number of iterations. All the ideas generated from all the previous tests were merged in this final test to build a hardware system to verify the practicality of the system to converge to an accurate answer within a

reasonable time. It was speculated that the time taken for processing was linearly dependent upon the iteration times. The data collected in this particular test not only would show the coordinate and TDOA but also the time needed for the computer to process each coordinate point in real-time. The algorithms were written into the set up shown in Figure 45 with a slight improvement to the workplane and sensor plane displacement as dictated in pg95 resulting in the set up shown in Figure 85.

Figure 85 : Sensors with minimal Z displacement

The base for the workplane was replaced with a softboard which allowed for more precise placement of the sensors as compared to the previous base which had predefined holes. Despite providing rigid placement, the previous base had poor placement resolution and hence precision. In addition to that, the z-displacement between the sensor plane and the workplane were kept to a minimum by placing the sensors directly onto the workplane.

The temperature within the software was set to 26℃ which was the measured temperature of the experimental environment. The author's bare finger was used to trace upon the workplane with paper. The results of the localizations are as shown Figure 86.

Figure 86: Real-time TES localization

The real-time test had also proven that it is able to utilize the AOA localized points as an initial guess point and improve upon it. The main difference when comparing Figure 84 with Figure 86 are the number of AOA localized points successfully captured in Figure 84 was found to be much higher. This could be caused by the fact that the iterative method consumed more CPU resources hence restraining the computer from capturing data. The number of iterations required to iteratively attain each coordinate is shown in Figure 87. When compared to the corresponding time duration in Figure 88, no obvious direct relations were found.

Figure 87: Iteration per coordinate point

Figure 88: Time duration per coordinate point

Assuming the time duration to attain a single point was 0.05s, 20 points could be attained in 1s. Meanwhile, the average minimization error attained during the real-time capture of the coordinates was 0.00008. Figure 89 shows the detailed minimization error for each coordinate point captured. Points with minimization errors larger than 0.0001m were automatically rejected from being evaluated.

Figure 89: Minimization error per coordinate point

The absolute error xy in Figure 90 was attained by comparing the first guess coordinate provided by the AOA with that of the final coordinate attained iteratively. The difference fluctuates between 0.002m to 0.006m, this was a rather small value.

Figure 90: Absolute xy error per coordinate point

6.3.1. ONLINE HYBRID METHOD TESTED ON HAND

Tests on surfaces without mechanoreceptors yielded promising results. This therefore prompted tests to be conducted on the human skin. In this particular test, the hardware was modified slightly for it to be wearable as shown in Figure 91.

Figure 91: Tracing on back of hand

A semi stiff, transparent graphing plastic sheet was used as the platform to mount the sensors. The graphing plastic sheet was chosen as the sensors can be kept on the same plane and the lines on the platform can be used for calibration by the visual system. Additionally it acted as a platform allowing for the attachment of the Velcro strap which was used to fasten the hardware to various surfaces.

Figure 92: Tracing on back of hand coordinates

129

Figure 92 depicts the coordinates attained from processing the TES and also the visual data. As it can be visually seen, the acoustically attained coordinates approximates the shape traced by the visual system. The corrected AOA on the other hand approximated the visual system better except for the portion on the x axis from 0.08m to 0.10m. This difference of shape was probably due to the effect of different height of the hand surface, in this case the hand surface at that portion drops, hence making the detected coordinates seem closer to the sensors.

6.4. DISCUSSIONS

The error difference between the iteratively corrected coordinate and that attained by AOA was very low. Nevertheless, the iterative method was important in self determining erroneous coordinate points in which its removal subsequently improved the general accuracy of the localization system. The surface of the workplane was kept as flat as possible in this chapter to match the plane of the sensors ensuring errors due to different dimensions do not happen. Despite the iterative method's self correcting abilities, it was still unable to differentiate between the workplane which was in the same plane as the sensors or different, resulting in a possibility to converge at a wrong answer. This is because the equation was modeled as 2-D function, hence all TDOA data fed to the iterative method is assumed to be a 2-D function. Hence, any lengthening or shortening of TDOA due to change of height on the workplane will instead be translated as relative closer and further in the 2-D plane which will result in possible successful convergence of answers in the wrong 2-D location.

Errors in localizations between the corrected AOA versus the visually collected coordinates, was about 0.01m. This could be attributed to the surface of the finger in contact with the workplane. The visual system locks onto a particular location of the finger throughout the entire experiment, but the contact between the finger and the workplane surface changes even in a single motion [53]. The nominal contact surface of the pulp of the finger is about 0.01m but the amount of force applied on the pulp of the finger would change the effective contact surface.

6.5. IMPROVEMENTS

Despite addressing accuracy problems stemming from the algorithm covered in this chapter, the device required more improvements to be environmentally reliable to fully exploit the ability of the device to be used on any surface. Tests so far have been conducted in an environment with the lowest possible noise. In real-life environments however, such conditions are difficult to achieve without rendering the device difficult to use. As mentioned in pg 29 the TES generated is a white noise, hence isolating it from diverse environmental noises is a challenge due to its similarity to environmental noises. In addition to that, creating a database of various environmental would severely limit the versatility of this device. These highlighted issues were addressed to a certain degree in the subsequent chapters.

7. IMPROVING ACCURACY & NOISE CLUSTERING

This chapter focuses on technology which allows the device to be used in noisy environments was created and tested in the lab. This technology like chapters before this are upgrades which are built upon the ideas from the previous chapters.

7.1. PROPOSED NOISE SEPARATION METHOD

As described in pg 19 the localization process of the system hinges heavily upon the TDOA extracted via cross-correlation which in turn is dependent upon the raw acoustic signals captured by the sensors. The sound sensed by the sensors which had the highest amplitude among the various acoustic sources would have its TDOA calculated.

The occurrences and amplitude of the TES are uncontrolled, as it is a byproduct of frictional forces when user's finger is traced on a surface. Acoustical interference or noise can also be defined similarly as the occurrences, amplitude and frequency of noise is also uncontrolled. Sources of interferences for the localization system can be divided to two characteristics, first where the signal occupies all the detected frequencies such as rustling of clothes and second where the sound only occupies a limited band of frequencies such as voice. It is assumed that neither the detected TES nor interference, individually or together will drive the sensors to saturation as this would nullify the ability of the system. With this assumption, it can be deduced three possible localizations probabilities can exist when both sources are active, the first is when the noise dominates the sensors, which results in the system continuously localizing upon the noise source. The second probability is that the TES dominates the sensors, which results in the continuous generation of coordinates which are centered on the TES location. The third probability is when neither TES nor the acoustical interference dominates the sensors, this result in the mixture of noise and TES coordinates occurring within a time frame in no particular order.

The third probability was of interest as it poses a complicated yet common issue for localization problems. The method proposed must be able to work in all the active

frequency ranges, able to deal with signals with low SNR continuously or intermittently which are spatially separated from the noise source.

Despite being a bane to researchers, interference sources which are spatially displaced an active within a short span of time gives the illusion that the human's finger moved extremely fast across the workplane. This scenario being unlikely was used as a rule to segregate apparent fast occurring coordinates as compared to the previous coordinate. The proposed concept is illustrated as in Figure 93 [54].

Figure 93: Coordinate arrivals

The first ever signal which is detected and successfully localized upon is represented as c_1. As there are no groups in existence, the coordinate of c_1 is used to initialize M_a which is the first group center. The subsequent coordinate is detected and labeled as c_2, group segregation process is defined as follows; the velocity between c_1 and c_2 was calculated, v_{12}. If v_{12} is larger than the specified threshold V_{th} of 0.30ms-1, a new group would be produced whereby the coordinates of c_2 is used to initialize the new group, M_b. This sis due to the fact that it is highly unlikely that the finger moved very quickly to a new position on the workspace, hence there is a high chance that the localized point could be interference. Additionally since there is a possibility that the interfering signal could be sustained intermittently or constantly throughout the entire process the finger tracing process, a new group was created to group all the potential noise localized coordinates together. The more erratic and inconsistent the interference signal's location, the easier it is for the system to recognize and cluster it without eliminating it.

When the new coordinate of c_3 is detected, its velocities with respect to c_1 and c_2 are calculated as v_{13} and v_{23} followed by the calculation of distance between c_3 and the center of 2 groups to produce d_{a3} and d_{b3}. Provided that the v_{13} or v_{23} are not larger than V_{th}, the equation as shown in (20) is used [55].

$$M_a \in \{v_{13} + d_{a3} < v_{23} + d_{b3}\} \tag{20}$$
$$M_b \in \{v_{13} + d_{a3} \geq v_{23} + d_{b3}\}$$
$$M_c \in \{v_{13} \geq v_{th}, v_{23} \geq v_{th}\}$$

In this case, the maximum numbers of groups are limited to three, but it can easily be scaled upwards to handle with more sources of interferences. The center of the group is updated every time a new coordinate is added to the group. Or it could be scaled downwards which allows generalizing of parameters within each group. It is defined by the mean of all the coordinates within its group. This implies that the group's center is dynamic resulting in a scalable and versatile method.

7.2. EXPERIMENTAL SET UP

The set up for this research comprised of two components first being the hardware and second being the software. The hardware set up was constructed as shown in Figure 94 (a). D-TES's frequency ratio function was disabled (14).

134

Figure 94: (a) Physical experimental setup (b) Camera field of view (FOV)

A notice board with a piece of paper placed on top of it was used as the surface in which the user would trace with their finger upon. The TES signals generated were detected by the sensors which were fed to the computer as information via the DAQ. An earphone was placed at coordinates (-0.01, 0.18)m and was used to introduce pre-recorded noises to the experiment. The camera was used in this experimental set up to trace the finger's motion in order to verify the acoustic localization's accuracy and will not be used in normal operations. The view from the camera is as shown in Figure 94(b). Visually, two colored traces were produced via the visual method, red for the center of

the finger and blue for the right side of the finger. This is due to the fact that the finger's contact surface changes hence the acoustic localization would change and the visual system needs to account for such occurrences which could result in a ±0.01m displacement. The FOV of the camera included the sensors as well as three strategically placed LEDs. Each of these LEDs had a function. The blue LED's Acoustic Indicator (AI) brightness response was linear with the amplitude of the acoustic signals captured. The green LED, Successful Localization Indicator (SLI), was used by the localization software to indicate to the visual capturing device that an acoustic localization has been successfully calculated. The third LED, Noise Emission Indicator (NEI), lighted up when the earphone was actively broadcasting an acoustic signal (interference). The SLI was used by to synchronize the data captured via both the acoustic and visual systems. The acoustic system localizes upon pseudo noise, finger generated TES and real background noises (occurs occasionally despite best efforts to eliminate them). Indistinguishable real-background noise localized by the acoustic system before detection of NEI and AI (also used in tandem with the presence of the finger in the visual feed) caused the inability of the visual system to be synchronized with that of the acoustic system. They are synchronized by matching the first eligible SLI detected with the first acoustic signal localized by the acoustic system. The definition of the first eligible SLI was defined as: if it occurred with NEI or AI indicating that the acoustic system localized upon the experimental sounds instead of natural background noises. Figure 95(a)-(c) are examples of graphs which show how these LEDs were used. When the SLI was activated as shown in Figure 95(b), the corresponding acoustic data were stored to a file for post processing Figure 95(c).

Figure 95: Synchronizing parameters (a) visual data (b) acoustic wavelet spectrogram

The acoustic system had localized many times prior to the beginning of the experiment, hence by using this LED indication system, it was possible to define when localizations were caused by environmental or experimental factors. Notice that the acoustic system in this case had already successfully localized a sound source three times prior to the introduction of the pre-recorded noise and also the finger scratch. These first three coordinates attained can hence be removed. The average localization duration was found to be 50ms. The camera frame rate was set to 300frames/sec. Consequently the first four acoustically localized coordinates can be removed as the sound sources detected are not within the scope of this experiment. According to these conditions, if both the pseudo noise and natural background noise are not present, the visual system's time of trigger and sampling rate can be reduced exactly to that of the acoustic system's just by using the SLI. This method of setup assisted in determining the source of the natural interference and the effectiveness of the implemented countermeasures by extracting the coordinates of the finger visually only when the SLI was triggered.

As mentioned earlier, the AOA and GDM were incorporated as a software and implemented together to improve accuracy which is required by the proposed spatially dependant algorithm (20). In addition to that, each successful localized coordinate is stored in a file together with its calculated TDOA and raw data from the DAQ. The proposed method was implemented offline, hence it required such files for calculation and verifications.

7.3. EXPERIMENTAL RESULTS AND DISCUSSIONS

Pre-recorded sounds of a human voice reading "the quick brown fox jumped over the lazy dog", the opening and closing of a door and white noise (generated using Audacity v2.0.3) were prepared.

The acoustic data collected and TES were introduced to the proposed method simultaneously which emulated the arrival of coordinates as if it were implemented in real-time. The coordinates were grouped automatically by the method proposed (20). Each group was plotted using different plot shapes and colors to clearly indicate their memberships.

The coordinates of the finger tracked visually were extracted using the active times of SLI, hence reducing its sampling rate. As shown in the results in Figure 96, the definitions of the groups are arbitrary. It was also apparent that the proposed method could easily segregate the two main sound sources and also sometimes a third sound source from actual environmental disturbances. In addition to that, experiments Figure 96(b) and(c) were not only able to segregate the different sounds but also localize accurately the location of the pseudo noise was at (-0.01, 0.18) m.

As mentioned previously the acoustic and visual coordinates could be synchronized perfectly with SLI if no noise is present. Since pseudo noise was present in the system, it triggered some of the SLI when AI was active

Figure 96: (a) voice interference (b) door open close interference (c) white noise interference

The visual system was unable to differentiate whether the SLI triggered was caused by the acoustic input from finger trace or from the pseudo noise. Hence a direct

139

comparison between each finger trace's coordinates with TES was unavailable. Despite that, it can be evaluated visually that the TES were successfully segregated from the pseudo noise. The comparison of the mean and standard deviation attained by the assumed trace acoustic group with that of the visual finger trace data triggered by SLI is conducted and shown in Table IV.

The experiment involving the door opening and closing yielded the smallest difference of mean with that of the mean of the visual data with the difference of 0.019m followed by the voice interference 0.023 m and lastly that of the white noise interference 0.035 m. The deviation and difference of the mean of the groups post experiment, as defined by the user as belonging to the earphone, were compared to the ideal earphone coordinate of (-0.01, 0.18)m, voice interference was found to be the worst based on Table IV and Figure 96(a).

Table IV: Statistics of coordinates

Parameter	Voice Interference		
	Group 1 (trace)	*Visual*	*Group2 (earphone)*
X mean(m)	0.080	0.100	0.015
Y mean(m)	0.086	0.098	0.136
X Std Deviation(m)	0.031	0.026	0.028
Y Std Deviation(m)	0.040	0.027	0.049

Parameter	Door Open Close Interference		
	Group 2 (trace)	*Visual*	*Group 1 (earphone)*
X mean (m)	0.068	0.078	-0.022
Y mean(m)	0.100	0.115	0.194
X Std Deviation(m)	0.035	0.031	0.005
Y Std Deviation(m)	0.037	0.030	0.044

Parameter	White Noise Interference		
	Group 1 (trace)	*Visual*	*Group 2 (earphone)*
X mean(m)	0.061	0.092	-0.045
Y mean(m)	0.085	0.101	0.203
X Std Deviation(m)	0.035	0.033	0.012
Y Std Deviation(m)	0.041	0.0341	0.023

This was most likely caused by the fact that voice is dominated by low frequency components which causes the localization to be unstable and hence the large deviation. As the grouping of the system is very dependent on the coordinates and time of occurrence of the detected sound, an unstable localization could also result in more groups being created as seen in Figure 96(a) where the maximum of three groups were created. Interestingly though, the standard deviation of the groups for TES shows that voice interference gave the lowest deviation, which means voice can be segregated from TES well therefore allowing TES to be localized upon well. Comparing means of

coordinates by TES as compared to means of visual coordinates, 'door opening and closing' did better than white noise. This is most likely due to the fact that 'door opening and closing' although played back many times throughout the experiment, was an impulse based interference, which allowed TES to be opportunistically detected and localized upon during the quiet times. White noise interference as Group 2 had the lowest deviation where the coordinates grouped around the earphone very tightly, mainly because it has similar characteristics with TES in which the system is optimized to localize upon.

7.4. IMPROVEMENTS

This chapter was generally about the additional improvements done to the previously proposed unique triboacoustic based mobile device input system by the authors. The improvements suggested in this chapter, hybridizing AOA with GDM and the usage of velocity and distance between attained coordinates is able to, to a certain degree, correctly segregate various interferences from the coordinates of TES. This verification was made possible with the innovative usage of LED's to synchronize the disparate sampling rates and different device's activation times to verify.

Future work would include an upgrade to the system's ability to not only segregate based on coordinates but also consider frequency of acoustic signals detected. In addition to that, the system should be made real-time with an easy method for users to label the groups of coordinates segregated by the proposed algorithm.

8. ADVANCED ROBUST NOISE CLUSTERING METHOD

This chapter paper builds upon current knowledge to study the reliability of this method when it is assembled using miniaturized hardware and introduced to various natural acoustic environments. The fabrication process of the device was changed from discrete to SMD with reduced size, mass and power consumption but at the same time providing higher levels of performance. As the interconnections between the devices were shorter and came within it an in-built electrical and magnetic noise protection, induced noise into the said circuit were expected to be less and insignificant. The algorithm from the previous chapter was upgraded to be implemented in the new hardware.

8.1. ROBUST DETECTION METHOD

The algorithm used to separate noise was the same as (20) with a difference. Instead of calculating the group's distance from the group's center, the magnitude of the frequencies attained from the signals were evaluated instead. 29 parameters were used to calculate the Euclidean distance where the 29 frequencies consisted of intervals of 1 KHz starting from 1 KHz. The magnitudes of the individual frequencies were attained via Fourier Transform process. The Euclidean distance is the distance between two points in Euclidean n-space. Typical Cartesian coordinates have a maximum dimension of three while there is no limit to the Euclidean n-space where n determines the number of dimensions. The coordinates of points in Euclidean n-spaces are represented with n-number of parameters and are called Euclidean vectors as shown in (21).

$$\vec{E_c} = (p_1, p_2, p_3, \dots p_n) \tag{21}$$

$\vec{E_c}$ represents the Euclidean vector while c differentiates the Vectors from one another. p_1 signified the magnitude for each frequency of the acoustic sources with the subscript representing the number of dimensions.

$$Ed_c = \sqrt{\left(p_{1,c} - p_{1,m}\right)^2 + \left(p_{2,c} - p_{2,m}\right)^2 + \dots\dots\left(p_{n,c} - p_{n,m}\right)^2} \tag{22}$$

The distances between two Euclidean vectors in an Euclidean *n*-spaces are calculated using Pythagorean equation where the number of dimensions evaluated are *n* as shown in (22). $p_{1,m}$ and $p_{2,m}$ represents the vector of the mean of the groups. Ed_c represents the Euclidean distance of the current Euclidean vector to a particular group's mean Euclidean vector. For the application used, the number of dimensions were 29 hence each acoustically calculated coordinate had a corresponding Euclidean vectors with 29 parameters representing its acoustic frequencies. This modified algorithm was incorporated into the miniaturized prototype to be operated in real -time.

This particular chapter only used two groups. Each group was assigned a colour in which its coordinates would be plotted in visually on the computer screen as shown in Figure 97. Yellow coordinates in this case were assigned as group 1 and the red coordinates were assigned as group 2.

Figure 97: (a) two groups of localized points (b) one group of localized points

Two groups are plotted in Figure 97(a), while only the coordinates of group 2 are shown in Figure 97(b) as the coordinates of group 1 were evaluated but not displayed. The user selected the group which represented the noise group, from the randomly system grouped coordinates based on their characteristics. This prompted the system to subsequently avoid displaying coordinates from the chosen group.

The user was provided with two methods of indicating the noise group to the computer. The first was by using the mouse to point and click on the coordinate which

144

the user had decided represents the noise group. This method required high precision and good hand-eye-coordination from the user as the coordinates appeared rather quickly due to the continuous occurrences of background noise sounds. Despite that, it gave more control to the users who were already familiar which was the background noise to be removed. The second option required simple but specialized hardware to realize the interaction to and from the user. This option relied on two LED's and two switches, each set representing a group. When a coordinate of that particular group was detected, the respective LED flashed. Since the background noise was continuously occurring, the group which had its LED flashing continuously in the absence of TES could easily be identified as background noise. All the user had to do was to press the corresponding switch to indicate to the computer on which group to ignore. Once the user had assigned the noise group, both the LED's were turned off until the next initialization which served two purposes, the first to save energy and second as a feedback to the user that the noise group input by the user had been successfully accepted.

It can be envisioned that the second option can easily be modified for disabled users where high luminance LED's can be used to indicate the groupings or some kinesthetic based feedback device such as unbalanced motors could be used in the place of LED's. Hence, this prototype was not only versatile in its location of usage but also its ability to cater to wider needs of various users.

8.2. MINIATURIZATION OF PROTOTYPE DEVICE

One of the main objectives of this research was to have high mobility, which implied that it has to be light, small and long lasting on a single charge. With this in mind, the new prototype was designed using similar components as mentioned in the previous chapters but of smaller packages were used [52], [56]. In addition to that, the circuits were designed and implemented onto customized PCB's as shown Figure 98(a).

Figure 98: (a) Sensor module and (b) circuits

Figure 98 shows that the sensor modules are small and the measured mass was less than 2g. Three of such modules were required for the implementation of the system. The modules were connected using Flat Printed Cables (FPC) where they were then grouped together and connected to the computer via a HDMI cable as shown in Figure 98(b).

The device's combinational mass was 120g with the thickness at the sensor area of about 5mm, and the maximum thickness of 15mm at the HDMI connector. The base of the device was flat hence allowed it to be assembled flush against any flat surface therefore avoiding the pitfalls as highlighted in chapter 5. Uneven surfaces were avoided to attain the best accuracy possible. Studies as in [56] found that an uneven surface or a surface with a different plane height with that of the sensor plane would result in low accuracy although the shape traced could be re-created.

Due to its light weight and low profile it could be assembled on almost any surface or any orientation with little effort. Arrangement in Figure 99 depicts the various arrangements.

Figure 99: various surfaces (a) surface of a paper (b) surface of cupboard

8.3. EXPERIMENTS IN VARIOUS ENVIRONMENTS AND SURFACES

All ideas mentioned were realized into a new prototype which was subsequently tested to evaluate its performance. Each experimental combination involved the user tracing a circle on a particular surface in a particular environment. The device was tested on various surfaces in various environments to have its output evaluated. Evaluation in this sense was to measure the ability of the traced circles on various surfaces to be re-created and to be correctly segregated into a different group from that of the background noise.

The most pertinent technology evaluated was the system's real-time segregation ability of attained acoustic coordinates which gave rise to its ability to adapt to different environments in real-time. Experimental setup which was used to evaluate the performance of the device in various environments and surfaces is shown as in Figure 100.

Figure 100: Mobile experimental unit

Table V shows the experimental schedule which is non-exhaustive. But it gives a general idea of the workability of the system.

Table V: Experimental combination schedule

Experimental combination	Experimental surface	Experimental location
1	cloth	Laboratory ✠ 1
2	paper	Laboratory ✠ 1
3	table	Laboratory ✠ 1
4	skin	Laboratory ✠ 1
5	cloth	Hall ✠ 2
6	paper	Hall ✠ 2
7	table	Hall ✠ 2
8	skin	Hall ✠ 2

✠ 1: Exhaust fan (45 dB)

✠ 2: Hall leading to the cafeteria during lunch time = extreme noise with reflections (60 dB)

Each of the combinations was tested using the same procedure where the system was activated on in the presence of background noise. D-TES's frequency ratio function was disabled (14). As soon as the system was turned on, the localization and display function were actively processing sounds detected. The user then traced some random patterns onto the surface of interest, which then allowed the system to randomly categorize it. It was hoped that the TES was in a different group from that of the background noise. This learning phase only required 10 points at most as the method used as explained in the earlier portion explained has quick learning ability. Since only two groups were available in the system, it was possible that the background noise monopolized all the acoustic parameters. In such an unfortunate case, the system would be rendered unavailable. Assuming the self-categorizing of the system was successful, the user then proceeded to select the group which is felt to be background noise.

This was then followed by the tracing of the finger in a circular motion onto the experimental surface. This motion was captured by a high-speed camera for post verification of the accuracy of the localization of the system[52], [56]. Acoustic raw data and localized acoustic data were stored in the hard drive for post verification.

Figure 101 depicts the results of conducting the experiment on various surfaces in a laboratory with a background noise of 45dB. The scatter plots Figure 101(a-1),(b-1), (c-1), (d-1) contain three different plots each, where Groups 1 and 2 plots are acoustically detected coordinates but were separated automatically by the computer based on their respective parameters. The third plot is the visually attained coordinates from the motion of the finger used to trace the shape on the respective surfaces. The graphs Figure 101(a-2), (b-2), (c-2), (d-2) represent the average magnitude of the frequency from 1-29 KHz in 1 KHz increments for the respective groups. Figure 101(a-1), (b-1), (c-1) show good results where groups which had been successfully segregated by the computer to able to overlap with that data attained visually. Meanwhile, Figure 101 (d-1) where the surface of the experiment was that of skin was found to have rather good accuracy. The points coincided with that of the visually attained points but had poor resolution which resulted in poor re-creation of the shape traced.

Figure 101: Results (a-1), (a-2) cloth in lab (b-1), (b-2) paper in lab (c-1), (c-2) table in lab (d-1), (d-2) skin in lab

As mentioned in the development of this prototype, the segregation of acoustically detected sounds was done automatically by the algorithm. This uncontrolled method of segregation resulted in the groups which represented the signal of interest to be random for each experiment. The signal of interest, TES is a white noise makes it difficult to define a cut-off frequency to filter out noise. Hence the localization evaluations of the acoustic signals measured are the combination of the background noise and that of the signal of interest. This therefore yields the graphs (a-2), (b-2), (c-2), (d-2) of Figure 101. The characteristic between each group looks similar with magnitude as the main distinguishing factor. Since the trace was done close to the sensors, it stands to reason that the magnitude of its signal of interest, TES would be larger than that of the background noise. Skin elasticity determines the amount of energy converted to TES or deformation of the dermis. This could explain the reason why the recreation of shape using the skin as a workplane in a noisy environment did not yield high resolution of coordinates. There were times within the process of tracing that the energy was released as skin deformation instead of acoustic signal production [43]. Regardless of that, it can be assumed that the despite that most of the energy was released as deformation, small amounts would still be released as acoustic signals.

Figure 102: Results (a-1), (a-2) cloth in hall (b-1), (b-2) paper in hall (c-1), (c-2) table in hall (d-1), and (d-2) skin in hall

It can be seen from Figure 102(a-1), (b-1) that cloth and paper surfaces when used, were able to re-create the shapes traced by the finger well while the utilization of surfaces table and skin were unable to do so Figure 102 (c-1), (d-1). Looking at the graphs (a-2), (b-2), (c-2), (d-2) of Figure 102, it can be seen that the signal of interest usually had significantly higher magnitude for every frequency. Despite that, experiments involving the surface of table and skin, exhibit small differences indicating that the background signal had drowned out the signal of interest and the grouping were just that of the background noises.

8.4. EFFECT OF BARRIER AT MICROPHONES

Results show that the algorithm was able to automatically segregate the signals as long as the signals were non-saturating, as in Figure 101(d) and Figure 102(c) & (d) where the background signals sends the sensors into saturation mode. A suggested solution to this problem was to create a physical barrier at the sensors which attenuates the background noise while at the same time amplifies the signal of interest. It is known that the source of TES is in the same plane as that of the sensors. The barrier erected had to consider the area in which the TES are generated. This influenced the design of the barrier to have a larger angle.

The barrier was 0.025m long and 0.01m high. The longest edge of the barrier was 0.025m. This barrier dimension was be able to keep frequencies above 13 KHz arriving from behind the sensor from being detected and at the same time allow for sound frequencies above 13 KHz arriving from the workplane to be concentrated into the microphone opening. These barriers were incorporated onto the prototype as shown in Figure 103.

Figure 103: (a) Prototype with barrier side view, (b) top view, (c) overall view

Tests with this prototype in noisy conditions of the hall were conducted again for the surfaces which previously failed to yield accurate results. The results are as in Figure 104.

Figure 104: Mitigation results (a) table surface in the hall (b) skin in the hall (c) skin in the lab

As seen from Figure 104, the barriers used were able to help the prototype to effectively capture the TES of interest and recreate the shape which was traced. Despite being able to approximate the shape traced to a certain degree, the shapes recreated were smaller and shifted. In addition to that, the frequency graph shows that the barrier created a more pronounced frequency pattern where the higher frequencies are of higher magnitude as compared to the background noise. This is due to the fact that the barrier was about 1cm by 2.5cm which only allowed it to reflect high frequency waves from the surface and attenuate the high frequencies coming from the wrong direction. This helped the localization of the sound source, but the shape of the barriers despite being able to increase the magnitude of the signal of interest via concentration had a possible side effect of shifting the coordinates due to complex interactions or reflections. This therefore calls for focus into the design of the barriers so that high signal selectivity can be attained without the drawbacks of coordinate shifting.

8.5. WRITING USING DEVICE

Up until now, the only shape draws was a circle as it was easy and repeatable but this does not mean that the system is only limited to accepting one shape. Figure 105 shows examples of various shapes, letters and numbers of various orientations which it can re-create.

Figure 105: (a) line '---' (b) plus ' +' (c) nine '9' (d) triangle ' Δ '

Figure 106: Corresponding examples of sensor orientation (a) hung from a cubicle (b) surface of a file (c) surface of a paper (d) surface of a table

The characters have been rotated in Figure 105 as the set up itself was rotated as shown in Figure 106, showing that the input device is versatile and can handle any orientation or shapes presented to it. The results by themselves are self-explanatory showing the ability of the device to re-create the shapes traced by the human finger.

8.6. IMPROVEMENTS

The experiment conclusively proves that the system proposed is highly robust against non-saturating background noises. It also has the ability to be used on various surfaces and adaptable to the needs of the user making it highly versatile due to its adaptability to any environment and user. It is proposed that using improved directional barriers could solve the problem of saturating background noise without the side effect of shrinking re-created shapes.

9. CONCLUSION AND FUTURE WORK

This thesis introduced a new method of mobile device input which maintains the mobility of mobile device via its usability on various surfaces, lightweight and low power consumption. While at the same time affords users a large range of inputs and ease of use through the detection and localization of TES from the action of users tracing on various surfaces. Subsequent paragraphs lists the conclusions garnered from body of the thesis.

This thesis pioneered the art of detecting and localizing upon continuous white noise (TES) via the usage of sensor spacing of the 8th subharmonic of 25 KHz. This method provided stability to the TDOA results attained through cross-correlation as it allowed the participation of multiple frequencies of various phases and magnitudes. Existing localization methodology emphasizes on the taming naturally detected signal to suit the design or utilizing artificially created signals which is compatible to the design. The method created by this thesis instead adapted the designs to the unruly characteristics of TES.

The accuracy and stability of the design was increased by utilizing the L-shape sensor arrangement which was shown to have superior accuracy as compared to other arrangements. In addition to that, the stability of the TDOA was controlled by regulating the input signals which were beyond the range of the sensors and emphasizing upon signals which have larger ratios of high frequency. This ensured that the low frequencies signals which are dominated by clutter such as voice do not interfere with the localization.

Due to the inaccuracies inherent in hardware used, the AOA algorithm which was supposed to yield a single coordinate from a set of TDOA instead yielded multiple coordinates. Magnitude of the signal measured was directly proportional to the precision of the TDOA yielded, with the limit that the magnitude was within the saturation limit. Due to the sensor spacing, the frequency characteristics signals which can be accurately and precisely cross-correlated to produce TDOA were also important. The less periodic and non-stationary the signal measured, the more stable and accurate the TDOA which can be attained from it.

Environmental issues such as the offset of the sensors from each other, from the plane and the temperature offsets which were not declared to the system were detrimental to the accuracy of the TDOA and hence the localization. These issues were inversely proportional to the accuracy of the system save for temperature. Temperature offsets were countered by AOA's opposing natural occurring error.

This thesis established a new method of solving GDM's weakness, where the initial position was the main source of the no-convergence and oscillating problems. This problem can be further exacerbated if the error function was crafted to have many local minima. Normal methods to solve this problem include crafting equations with few local minima or to create algorithms which can effectively traverse across the error function. This thesis utilized AOA to place the initial point of the GDM close to the final answer hence avoiding the local minima issue altogether.

Due to the improvements of using AOA and GDM hybrid, more accurate results coordinates were produced at a faster rate which allowed for the following improvement. This ingenious improvement utilized the accurate coordinates to create a self-learning segregation algorithm which was based on spatial methods. This method could in real-time adapt to various acoustic environments to segregate TES originated coordinates from background noise caused coordinates. Not only could it segregate the noise from TES but also pinpoint the source of the interference.

This innovative methodology was further improved by utilizing not only spatial methods for evaluation but also frequency profile of each coordinate collected. This self-learning segregation method was tested in real-acoustic environments and was found to be effective for all conditions except when the interference saturated the sensors. Directional acoustic barriers used to counter this problem were effective in amplifying TES and at the same time attenuating interference. This came not without a negative side effect of shrinking of shapes re-created. More work has to be done in this area

Future works into this research could go into two directions, applications or further research into the fundamental workings of the system. Regarding applications, here are some possible directions:

a) Usage of microprocessors to replace the computer.

b) Improved back end software to increase the usability.

c) Locating multiple leaks within a pipe system[57].

d) Locating and verifying fluid droplets dispersions onto a surface.

e) Locating wildlife such as frogs[58]/ insects (crickets)[59].

f) Abnormal noise detection within a machine[60].

Meanwhile, Future works in research can go in the direction of:

a) Study of ears in detail[61].

b) Study of the factors which effect TES such as surface, materials[62].

c) Use of HRTF concept to filter out certain frequencies.

d) The study of TES using capacitive type touch pads.

REFERENCES

[1] B. Sanou, "2014 ICT Facts and Figures,"International Communication Union, 2014.

[2] Cisco, "Cisco Visual Networking Index : Global Mobile Data Traffic Forecast Update , 2013 – 2018,"Cisco Systems, 2014.

[3] The World Bank, "Infrastructure and Growth." [Online]. Available: http://web.worldbank.org/WBSITE/EXTERNAL/TOPICS/EXTINFRA/0,,contentMDK:2 3154473~pagePK:64168445~piPK:64168309~theSitePK:8430730,00.html. [Accessed: 09-Sep-2014].

[4] C. Papageoriou, L. Kolovich, G. Ho, K. Wang, N. Gigineishvili, A. Pitt, T. S. Choi, N. Duma, S. Rosa, C. Henn, T. Eicher, and D. Kuenzel, "IMF Policy Paper - Sustaining Long-Run Growth and Macroeconomic Stability in Low-Income Countries-The Role of Structural Transformation and Diversification," International Monetary Fund, 2013.

[5] N. Vandycke, S. Anees, L. Andres, A. Baietti, D. Biller, C. M. Briceon-Garmendia, O. Chaudry, and Et.al, "Transformation Through Infrastructure," World Bank Infrastructure Strategy Committee, 2015.

[6] N. Czernich and O. Falck, "Broadband Infrastructure and Economic Growth Broadband Infrastructure and Economic Growth Abstract," Center for Economic Studies ifo Institute, 2009.

[7] Roller.L and Waverman.L, "Telecommunications Infrastructure and Economic Development : A Simultaneous Approach," Jul. 1996.

[8] P. Thomas, "World Economy and ICT," Intel, London, 2012.

[9] C. Calderón and L. Servén, *The Effects of Infrastructure Development on Growth and Income Distribution*. The World Bank, 2004.

[10] A. N. Sokolov, M. E. Roberts, and Z. Bao, "Fabrication of low-cost electronic biosensors," *Mater. Today*, vol. 12, no. 9, pp. 12–20, Sep. 2009.

[11] M. D. P. Baquero Forero, "Mobile communication networks and Internet technologies as drivers of technical efficiency improvement," *Inf. Econ. Policy*, vol. 25, no. 3, pp. 126–141, Sep. 2013.

[12] R. A. Robison, "Moore's Law: predictor and driver of the silicon era.," *World Neurosurg.*, vol. 78, no. 5, pp. 399–403, Nov. 2012.

[13] G. E. Moore and L. Fellow, "Cramming More Components onto Integrated Circuits," *Proc. IEEE*, vol. 86, no. 1, pp. 82–85, Jan. 1998.

[14] "Intel Chips - poster." Intel, 2012.

[15] A. W. Forrey, *Organic Scintillators and Scintillation Counting*. Elsevier, 1971, pp. 835–848.

[16] M. V. WILKES, *Computing Perspectives*. Elsevier, 1995, pp. 13–20.

[17] M. Ahmad, *The Next Web of 50 Billion Devices: Mobile Internet's Past, Present and Future (Smartphone Chronicle)*. North Charleston: CreateSpace Independent Publishing Platform, 2014.

[18] C. A. H. Richard and J. E. Bryant, "Pharmacy student perceptions on the introduction of clinical case studies solved with Apple mobile devices into a basic health science laboratory," *Curr. Pharm. Teach. Learn.*, vol. 6, no. 5, pp. 659–666, Sep. 2014.

[19] S. Sumriddetchkajorn, K. Chaitavon, and Y. Intaravanne, "Mobile device-based self-referencing colorimeter for monitoring chlorine concentration in water," *Sensors Actuators B Chem.*, vol. 182, pp. 592–597, Jun. 2013.

[20] Susan Teltscher, V. Gray, E. Magpantay, D. Olaya, I. Vallejo, L. Kreuzenbeck, M. Best, S. Morray, F. Callorda, D. Korka, C. Stork, and S. Zuhley, "Measuring the Information Society," 2013.

[21] S. Ou, K. Yang, and J. Zhang, "An effective offloading middleware for pervasive services on mobile devices," *Pervasive Mob. Comput.*, vol. 3, no. 4, pp. 362–385, Aug. 2007.

[22] A. Gani, G. M. Nayeem, M. Shiraz, M. Sookhak, M. Whaiduzzaman, and S. Khan, "A review on interworking and mobility techniques for seamless connectivity in mobile cloud computing," *J. Netw. Comput. Appl.*, vol. 43, pp. 84–102, Aug. 2014.

[23] World Semiconductor Trade Statistics, "Worldwide Semiconductor Market is forecasted to be US$325 billion in 2014, up 6.5% from 2013," *2014..6.3*, 2014. [Online]. Available: https://www.wsts.org/PRESS/Recent-News-Release. [Accessed: 05-Sep-2014].

[24] C.-C. Lai and C.-F. Wu, "Display and device size effects on the usability of mini-notebooks (netbooks)/ultraportables as small form-factor Mobile PCs.," *Appl. Ergon.*, vol. 45, no. 4, pp. 1106–15, Jul. 2014.

[25] A. A. Ozok, D. Benson, J. Chakraborty, and A. F.Norcio, "A Comparative Study Between Tablet and Laptop," *Int. J. Hum. Comput. Interact.*, vol. 24, no. 3, pp. 329–352, 2008.

[26] W. Xu, C. Yu, J. Liu, and Y. Shi, "RegionalSliding: Facilitating small target selection with marking menu for one-handed thumb use on touchscreen-based mobile devices," *Pervasive Mob. Comput.*, vol. 1, pp. 1–16, Mar. 2014.

[27] A. Sears and J. A. Jacko, *The Human-Computer Interaction Handbook*, Second. Taylor & Francis Group, 2008.

[28] W. B. Fritz, M. K. Mcnulty, F. Bilas, and B. J. Jennings, "The Women of ENIAC," *IEEE Ann. Hist. Comput.*, vol. 18, no. 3, pp. 13–28, 1996.

[29] "Apple - iPhone." [Online]. Available: https://www.apple.com/jp/iphone/. [Accessed: 13-Dec-2014].

[30] "Nexus 6（32 GB、ダークブルー） - Google Playの端末." [Online]. Available: https://play.google.com/store/devices/details?id=nexus_6_blue_32gb. [Accessed: 13-Dec-2014].

[31] "Apple - iPad Air 2 - Technical Specifications." [Online]. Available: http://www.apple.com/ipad/. [Accessed: 13-Dec-2014].

[32] "Nexus 9（16 GB、Wi-Fi、インディゴ ブラック） - Google Playの端末." [Online]. Available: https://play.google.com/store/devices/details?id=nexus_9_black_16gb_wifi. [Accessed: 13-Dec-2014].

[33] H. Kim, S. Kwon, J. Heo, H. Lee, and M. K. Chung, "The effect of touch-key size on the usability of In-Vehicle Information Systems and driving safety during simulated driving.," *Appl. Ergon.*, vol. 45, no. 3, pp. 379–88, May 2014.

[34] A. Sears and B. E. N. Shneiderman, "High precision touchscreens : comparisons with a mouse design strategies and comparisons with a mouse," *Int. J. Man. Mach. Stud.*, vol. 34, no. 4, pp. 593–613, 1991.

[35] "Microsoft Band Official Site." [Online]. Available: http://www.microsoft.com/microsoft-band/en-us. [Accessed: 13-Dec-2014].

[36] "SmartWatch3 SWR50 Sony Mobile Communications." [Online]. Available: http://www.sonymobile.co.jp/product/smartwear/swr50/. [Accessed: 13-Nov-2014].

[37] A. Sears, "Improving touchscreen keyboards: design issues and a comparison with other devices," *Interact. Comput.*, vol. 3, no. 3, pp. 253–269, Dec. 1991.

[38] H. Sasaki, T. Kuroda, P. Antoniac, Y. Manabe, and K. Chihara, "HAND-MENU SYSTEM : A DEVICELESS VIRTUAL INPUT INTERFACE FOR WEARABLE COMPUTERS," *J. Control Eng. Appl. Informatics*, vol. 8, no. 2, pp. 44–53, 2006.

[39] C. Harrison and A. D. Wilson, "OmniTouch : Wearable Multit ouch Interaction Everywhere," in *Proceedings of the 24th annual ACM symposium on User interface software and technology - UIST '11*, 2011, pp. 441–450.

[40] C. Harrison, D. Tan, and D. Morris, "Skinput : Appropriating the Body as an Input Surface," in *28th Annual SIGCHI Conference on Human Factors in Computing Systems*, 2010, pp. 453–462.

[41] T. Deyle, S. Palinko, E. S. Poole, and T. Starner, "Hambone: A Bio-Acoustic Gesture Interface," in *2007 11th IEEE International Symposium on Wearable Computers*, 2007, pp. 3–10.

[42] X. Han, H. Seki, Y. Kamiya, and M. Hikizu, "Wearable handwriting input device using magnetic field," *Precis. Eng.*, vol. 33, no. 1, pp. 37–43, Jan. 2009.

[43] H. Zahouani, G. Boyer, C. Pailler-Mattei, M. Ben Tkaya, and R. Vargiolu, "Effect of human ageing on skin rheology and tribology," *Wear*, vol. 271, no. 9–10, pp. 2364–2369, Jul. 2011.

[44] R. Fagiani, F. Massi, E. Chatelet, Y. Berthier, and A. Akay, "Tactile perception by friction induced vibrations," *Tribol. Int.*, vol. 44, no. 10, pp. 1100–1110, Sep. 2011.

[45] L. Skedung, K. Danerlöv, U. Olofsson, C. Michael Johannesson, M. Aikala, J. Kettle, M. Arvidsson, B. Berglund, and M. W. Rutland, "Tactile perception: Finger friction, surface roughness and perceived coarseness," *Tribol. Int.*, vol. 44, no. 5, pp. 505–512, May 2011.

[46] C. Bauer, L. Yazzolino, G. Hirsch, Z. Cattaneo, T. Vecchi, and L. B. Merabet, "Neural correlates associated with superior tactile symmetry perception in the early blind.," *Cortex.*, vol. 63C, pp. 104–117, Aug. 2014.

[47] H. Ishizuka and N. Miki, "MEMS-based tactile displays," *Displays*, pp. 1–8, Nov. 2014.

[48] K. C. HO and Y. T. Chan, "Solution and performance analaysis of geolocation by TDOA," *Aerosp. Electron. Syst.*, vol. 29, no. 4, pp. 1311–1322, 1993.

[49] I. J. Tashev, *Sound Capture and Processing: Practical Approaches*. Wiley, 2009.

[50] H. Zahouani, R. Vargiolu, G. Boyer, C. Pailler-Mattei, L. Laquièze, and a. Mavon, "Friction noise of human skin in vivo," *Wear*, vol. 267, no. 5–8, pp. 1274–1280, Jun. 2009.

[51] K. Asamene and M. Sundaresan, "Analysis of experimentally generated friction related acoustic emission signals," *Wear*, vol. 296, no. 1–2, pp. 607–618, Aug. 2012.

[52] Y. W. Leong, H. Seki, Y. Kamiya, and M. Hikizu, "A feasibility study of utilizing tribo-acoustics for mobile user interface," in *2013 Seventh International Conference on Sensing Technology (ICST)*, 2013, pp. 804–809.

[53]   Y. W. Leong, H. Seki, Y. Kamiya, and M. Hikizu, "Verification of finger contact area on paper," in *The 20th International Symposium on Artificial life and Robotics*, 2015, pp. 276–281.

[54]   J. Zhang, M. Walpola, D. Roelant, H. Zhu, and K. Yen, "Self-organization of unattended wireless acoustic sensor networks for ground target tracking," *Pervasive Mob. Comput.*, vol. 5, no. 2, pp. 148–164, Apr. 2009.

[55]   S. Okada and O. Hasegawa, "Incremental Learning , Recognition , and Generation of Time-Series Patterns Based on Self-Organizing Segmentation," *J. Adv. Comput. Intell. Intell. Informatics*, vol. 10, no. 3, pp. 395–408, 2006.

[56]   Y. W. Leong, H. Seki, Y. Kamiya, and M. Hikizu, "TRIBOACOUSTIC LOCALIZATION SYSTEM FOR MOBILE DEVICE - ENVIRONMENTAL EFFECTS TO ACCURACY," *Int. J. Smart Sens. Intell. Syst.*, vol. 7, no. 2, pp. 658–673, Jun. 2014.

[57]   D. Ozevin and J. Harding, "Novel leak localization in pressurized pipeline networks using acoustic emission and geometric connectivity," *Int. J. Press. Vessel. Pip.*, vol. 92, pp. 63–69, Apr. 2012.

[58]   R. a. Page and M. J. Ryan, "The effect of signal complexity on localization performance in bats that localize frog calls," *Anim. Behav.*, vol. 76, no. 3, pp. 761–769, Sep. 2008.

[59]   R. a. Costello and L. B. Symes, "Effects of anthropogenic noise on male signalling behaviour and female phonotaxis in Oecanthus tree crickets," *Anim. Behav.*, vol. 95, pp. 15–22, Sep. 2014.

[60]   T. He, D. Xiao, Q. Pan, X. Liu, and Y. Shan, "Analysis on accuracy improvement of rotor-stator rubbing localization based on acoustic emission beamforming method.," *Ultrasonics*, vol. 54, no. 1, pp. 318–29, Jan. 2014.

[61]   Y. Kahana and P. a. Nelson, "Numerical modelling of the spatial acoustic response of the human pinna," *J. Sound Vib.*, vol. 292, no. 1–2, pp. 148–178, Apr. 2006.

[62]   K. Mizuhara, H. Hatano, and K. Washio, "The effect of friction on the usability of touchpad," *Tribol. Int.*, vol. 65, pp. 326–335, Sep. 2013.

ACKNOWLEDGEMENTS

First and foremost I would like to thank my main supervisor Professor Hiroaki Seki for being incredibly supportive, helpful and kind to me in my various endeavors which allowed me to complete this fulfilling research and at the same time enjoy my stay in Japan.

I would like to also thank Professor Dr. Yoshitsugu Kamiya and Mr. Masatoshi Hikizu for their suggestions and support throughout the entire course of my studies .

I am greatly indebted to Mitani Scholarship Foundation for providing me financial assistance and also interesting monthly discussions throughout the entire PhD course.

My life in Kanazawa University was made all the more fulfilling with great lab mates past and present. To name a few Hiroki .S, Akihiro .F, Junya .K, Mikami, Ono, Kato, Zhang, Oh. I would also like to thank Ishikawa International House staff for being great hosts during my entire stay in Japan. My life in Kanazawa was interesting due to the camaraderie offered by the members Zaimoku Badminton club which allowed me to compete in prefectural level competitions and also learn more about local life. The Malaysian and International community here in Kanazawa helped me grow as a person and to see things from a broader perspective.

The office staff , to name a few Ms Oyama , Mdm Akiko .T and Ms Hisada who have been greatly co-operative and helpful every time I dealt with them.

I would also like to thank Assoc Prof Dr. Khairul Salleh for his assistance and suggestions, most importantly, that I should complete my PhD in Kanazawa University and also his introduction of Karaoke to me. I am also appreciative of the encouragement showered onto me by Assoc Prof Dr. Ong H.S, Prof Dr. Norashidah, Dr. Chua K.H, Assoc Prof Dr. Farrukh.N and Dr. Tan C.S. I would like to thank Dr. Fazreena.A for supporting my decision to come to Japan. I would like to thank the Uniten University Management Committee for approving my PhD study leave. Ms Siti Fatimah.Z for being an awesome HR staff to liaison with.

There are many people whom have helped me throughout this journey which space does not permit me to name them, I thank them all the same.

My family for being supportive of me regardless of the challenges faced in doing so. We can choose our friends but not our family, and I count myself lucky on having a great family.