Sensor Fusion for Tangible Acoustic Interfaces for Human Computer Interaction

A thesis submitted to Cardiff University

In candidature for the degree of

Doctor of Philosophy

By

Mostafa Al-Kutubi, B.Sc., M.Sc.

Manufacturing Engineering Centre

Cardiff University

United Kingdom

2007

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

To My family

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor Professor D.T. Pham, Founder Director of the Manufacturing Engineering Centre, and would like to express my warm thanks and sincere gratitude to him for his invaluable assistance and for providing me with all kinds of support I needed throughout this work. I must also thank him for his exceptional patience and tolerance.

During this work I have collaborated with colleagues in the lab and wish to extend my special thanks to Dr. Ming Yang, Mr. Ze Ji, Dr. Zuobin Wang and the visiting student Mr. Olivier Peyroutet for their aid and friendship. The hardware design of electronic circuits by Dr. Ming Yang is acknowledged. I would also like to thank Mrs. Celia Rees for her kind assistance in formality issues and to thank Dr. Michael Packianather for his help in the final corrections of the thesis.

I cannot end without thanking my mother for her encouragement and addressing my appreciation with deep gratitude to my wife Jawahir for her understanding and help. Finally, I would also like to thank my brothers Harith and Omer.

This work was financed by the European FP6 IST Project "Tangible Acoustic Interfaces for Computer-Human Interaction (TAI-CHI)". The support of the European Commission is gratefully acknowledged. The MEC is the coordinator of the EC-funded FP6 I*PROMS NoE.

Abstract

This thesis presents the development of tangible acoustic interfaces for human computer interaction. The method adopted was to position sensors on the surface of a solid object to detect acoustic waves generated during an interaction, process the sensor signals and estimate either the location of a discrete impact or the trajectory of a moving point of contact on the surface. Higher accuracy and reliability were achieved by employing sensor fusion to combine the information collected from redundant sensors electively positioned on the solid object.

Two different localisation approaches are proposed in the thesis. The learning-based approach is employed to detect discrete impact positions. With this approach, a signature vector representation of time-series patterns from a single sensor is matched with database signatures for known impact locations. For improved reliability, a criterion is proposed to extract the location signature from two vectors. The other approach is based on the Time Difference of Arrival (TDOA) of a source signal captured by a spatially distributed array of sensors. Enhanced positioning algorithms that consider near-field scenario, dispersion, optimisation and filtration are proposed to tackle the problems of passive acoustic localisation in solid objects. A computationally efficient algorithm for tracking a continuously moving source is presented. Spatial filtering of the estimated trajectory has been performed using Kalman filtering with automated initialisation.

CONTENTS

DECLARATION AND STATEMENTS ii

ACKNOWLEDGEMENTS iv

ABSTRACT v

LIST OF FIGURES x

LIST OF TABLES xiv

ABBREVIATIONS xv

NOMENCLATURE xvi

CHAPTER 1. INTRODUCTION 1

1.1. Motivation 3

1.2. Research Aim and Objectives 5

1.3. Hypotheses 6

1.4 Methodology 6

1.5. Thesis Outline 7

CHAPTER 2. LITERATURE REVIEW 8

2.1. Commercial Tangible Interfaces 8

2.2. Research Work on Tangible Interfaces 18

2.3. Literature Survey on Source Localisation 22

2.4. Overview of Acoustic Waves in Solids 28

CHAPTER 3. Enhanced Pattern Matching for Impact Localisation 36

3.1. Hypothesis of Pattern-based Localisation 37
3.1.1. Focusing in Time-reversal Theory 39
3.1.2. Realisation of Source Localisation from Time-reversal Focusing 42
3.1.3. Deduced Rules 43

3.2. Criterion of Pattern Matching 43

3.3. Empirical Analysis 47
3.3.1. Evaluation Procedure 50

3.4. Enhanced Resolution 53
3.4.1. Multi-dimensional signal 53
3.4.2. Post Filtering 55
3.4.3. Coherence Filtering 57

3.5. Location Feature Extraction for Enhanced Reliability 60

3.6. Experimentation 64

CHAPTER 4. An Investigation into Localisation Approaches for TAI 67

4.1. Active and Passive Sources 68

4.2. Bearing Estimation 70

4.3. Hyperbolic Localisation Geometry 73
4.3.1. Analytic Solution 74
4.3.2. Iterative Solution 76

4.4. Time Difference Estimation 77
4.4.1. Cross Correlation Based TDOA 78
4.4.1.1. Generalised Cross Correlation 79
4.4.1.2. Criteria of the GCC Filtering Process 80
4.4.2. Beamforming 82

4.5. Signal Analysis 83

4.6. Experimentation 89

4.7. Concluding Remarks 100

CHAPTER 5. Enhanced Acoustic Source Localisation 103

5.1. Spatial Likelihood-based Localisation 104
5.1.1. Theory 107
5.1.2. ELM Algorithm 109
5.1.2.1. Filtering Process in ELM 115
5.1.2.2. Temporal Smoothing 120
5.1.2.3. ELM Outlier Detection 124

5.2. Time Difference-based Localisation 125
5.2.1. Linear Cross Spectral Phase 126
5.2.2. ML Positioning 127
5.2.3. Time Difference Outlier Detection 129

CHAPTER 6. Enhanced Continuous Tracking 133

6.1. Trajectory Prediction with Kalman Filter 134
6.1.1. Automatic Filter Initialisation 139

6.2. Linear Positioning Algorithm 144
6.2.1. Extended Positioning Options 152
6.2.2. Empirical Error Analysis 154

CHAPTER 7. Conclusions and Future Work 158

References. 162

LIST OF FIGURES

Figure		Page
(1.1)	Acoustic HCI model.	02
(2.1)	Resistive touch screen.	10
(2.2)	Optical touch screen.	10
(2.3)	Capacitive touch screen.	11
(2.4)	Infrared touch screen.	12
(2.5)	SAW touch screen.	13
(2.6)	Pegasus PC notes tracker.	15
(2.7)	Mimio interactive whiteboard	16
(2.8)	Virtual laser keyboard.	16
(2.9)	MIT PingPongPlus set up and electronic circuit diagram.	19
(2.10)	MIT Knock localiser on large window.	21
(2.11)	MIT laser tracking interactive screen.	22
(2.12)	Acoustic wave spectrum.	30
(2.13)	Longitudinal and shear body waves.	31
(2.14)	Love and Rayleigh surface waves.	32
(2.15)	Modes of Lamb wave in plate.	33
(2.16)	Transit time measurement for a. group velocity and b. phase velocity using ultrasound signal.	34
(2.17)	Flexural dispersion in unidirectional laminate.	35

(3.1)	Sketch of time reversal focusing in random medium. a. impulse transmission and reception. b. time-reversed transmission and impulse localisation	40
(3.2)	LPM system layout.	45
(3.3)	Spatial cross correlation peaks of impact.	48
(3.4)	Resolution in LPM system.	49
(3.5)	Data recording for evaluation.	51
(3.6)	Evaluation procedure for the data collected as in figure 3.5.	52
(3.7)	Threshold of cross correlation coefficient.	54
(3.8)	Low pass filter.	56
(3.9)	Magnitude squared coherence versus frequency of two nail click signals at the same location on a fibre board.	59
(3.10)	Received signals from two different paths.	62
(3.11)	Black box LPM model.	62
(3.12)	Various experimental equipments for LPM localisation.	66
(4.1)	Active source localisation from time TOA circles' intersection.	69
(4.2)	TDOA hyperbola for passive system.	69
(4.3)	Far-field wave front and beamforming.	70
(4.4)	Far-field source location (x,y) from two AOA's.	71
(4.5)	Near-field TDOA assuming spherical wave front.	72
(4.6)	Hyperbolic intersection geometry.	74
(4.7)	(a) Hann window function (b) mother wavelet from the Symlet family.	86
(4.8)	(a) Time signal example from nail scratch on MDF board (b) power spectrum of the signal.	87

(4.9)	(a) time-frequency representation for the signal in figure (4.9) using (a) STFT and (b) wavelet transform.	88
(4.10)	MDF interactive board.	90
(4.11)	Different types of tested sensors.	92
(4.12)	BU-1771 sensor, connection diagram and frequency response.	93
(4.13)	Signal conditioning board.	94
(4.14)	DAQ-2010 data acquisition board.	95
(4.15)	TAI model diagram.	96
(4.16)	Test signals.	97
(4.17)	Whiteboard interactive object and conditioning circuit.	99
(4.18)	Frequency response of the IIR Low pass digital filter.	99
(5.1)	TAI model with four sensors.	104
(5.2)	Real signals, gi, from the ith sensor generated by (a) nail click (impact) and (b) nail scratch on the MDF board.	105
(5.3)	Cross correlation of impact signals showing multiple peaks.	106
(5.4)	Algorithm diagram for (a) ELM_t and (b) ELM_f.	112
(5. 5)	Theoretical hyperbolic map of time differences.	113
(5.6)	Spatial likelihood for the source at (200,300)mm for each pair of sensors.	114
(5.7)	Spatial Likelihood of the source at (200,300)mm.	115
(5.8)	Elliptic band-pass filter design.	117
(5.9)	Spatial likelihood for the source at (200,300)mm for each pair of sensors after filtering the signals.	118
(5.10)	Spatial Likelihood of the source at (.2,.3)m using filterd signals, peak at (180,280)mm.	119

(5.11)	Spatial likelihood map using PHAT process.	119
(5.12)	ELM of impact signals using (a) raw signals, (b) conditioned signals, (c) as in (b) with Hilbert envelope, (d) PHAT and (e) with Hilbert envelope only.	122
(5.13)	ELM of scratch signals using (a) raw signals, (b) conditioned signals, (c) as in (b) with Hilbert envelope, (d) PHAT and (e) Hilbert envelope only.	123
(5.14)	Contour plot of ML algorithm for (a) impact signal and (b) scratch signal.	130
(5.15)	Geometry of time differences relationship.	131
(6.1)	Tracking continuous nail scratch using (a) ELMt and (b) ML algorithms in black. The plot in red is for the same sketches with a Kalman filter applied.	135
(6.2)	Kalman filter recursive algorithm.	139
(6.3)	Undesired estimations appear as a result of Kalman filter initialisation at the beginning and at discontinuity '+' generated with (a) ELM and (b) ML algorithms.	141
(6.4)	Correction flowchart to solve the problem of unwanted estimations produced by Kalman filer.	142
(6.5)	Undesired estimations at the beginning and at discontinuity 'X' generated with (a) ELM and (b) ML algorithms disappeared with the proposed correction to a Kalman filter initialisation.	143
(6.6)	TDOA localisation from tangent circles.	146
(6.7)	Figure 6.7. Sketching (a) 'X' and (b) 'face' with nail scratch using the proposed algorithm with Kalman filter shown in red for comparison with figure 6.5.	150
(6.8)	Using Paint program as an application interface to the final algorithm. Example of using brush to (a) write mec and (b) draw Sine wave. (c) Live demo in public exhibition.	151
(6.9)	Error analysis.	155

LIST OF TABLES

Table		Page
(3.1)	The percentage of correct estimations from four channels.	56
(3.2)	Results of localising 1080 impacts at 12x9 locations at 20mm resolution for four channels using (a) finger tap and (b) nail click impacts on a glass sheet.	58
(6.1)	Standard deviation of estimated locations for (a) click and (b) scratch test signals	156

ABBREVIATIONS

AOA	Angle Of Arrival.
ELM	Enhanced Likelihood Mapping.
FIR	Finite Impulse response.
GCC	Generalised Cross Correlation.
HCI	Human Computer Interaction.
IIR	Infinite Impulse Response.
LCSP	Linear Cross Spectral Phase.
LED	Light Emitting Diodes.
LP	Linear Positioning.
LPM	Location Pattern Matching.
MDF	Medium Density Fibre.
ML	Maximum Likelihood.
MLF	Maximum Likelihood Filter.
PHAT	Phase Transform.
SAW	Surface Acoustic Wave.
SCOT	Smooth Coherence Transform.
STFT	Short Time Fourier Transform.
TAI	Tangible Acoustic Interface.
TDOA	Time Difference of Arrival.
WT	Wavelet Transform.

NOMENCLATURE

CHAPTER (3)

ρ	Acoustic pressure.
v_p	Phase velocity.
x	Distance in the x-axis direction.
A	Amplitude constant of acoustic wave pressure.
α	Wave propagation attenuation factor.
β	Wave number.
ω	Angular frequency.
t	Time variable.
$H(\omega)$, $h(t)$	Transfer function of transmission medium.
s	Source signal.
g	Received signal.
σ	Standard deviation of Gaussian random variable.
l	Likelihood function.
L	Vector-distance metric.
Q	Tuning factor.
R	Cross-Correlation.
Γ	Cross-Correlation Coefficient.
i	Location index.
K	Number of channels.
τ	Time lag.

γ	Magnitude squared coherence.
P	Cross-Spectral density.
Δ_f	Frequency interval.
S	Fourier Transform of s.
G	Fourier Transform of g.
$\tilde{H}$, $\tilde{h}$	Hypothetical Transfer function.
ψ	Adjustment Filter.
V	Magnitude of $\tilde{H}$.
φ	Phase of $\tilde{H}$.

CHAPTER (4)

ϑ	Angle of arrival.
v	Wave propagation velocity.
τ	Time difference.
d	Two element array spacing.
λ	Wavelength.
(x,y)	Cartesian coordinates of source signal.
(x_i,y_i)	Cartesian coordinates of i^{th} sensor.
(r,θ)	Polar coordinates of source signal.
e	Error.
n, N	Gaussian noise.
t	Time variable.
g	Received signal.
R	Cross-Correlation.

ψ	Weighting function.
E	Beamformer energy signal.
h	Windowing function.
Ψ	Mother Wavelet.
a	Wavelet scale parameter.
b	Wavelet position parameter.

CHAPTER (5)

τ	Time difference.
s	Source signal.
g	Received signal.
p	Posterior probability.
q	Estimated source postion.
v	Wave propagation velocity.
σ	Standard deviation of Gaussian random variable.
u	Sensor location.
N	Number of sensors.
M	Number of sensor pairs.
Z	Analytic signal.
Θ	Hilbert envelope.
ε	Adjustment factor.
G	Fourier Transform of g.
P	Cross spectral density.

A	Magnitude of P.
ϕ	Phase of P.
f	Frequency.
J	Least error estimator.

CHAPTER (6)

χ	System dynamics.
φ	Transition matrix.
a	Process noise.
p_x, p_y	State variable of x-y position.
v_x, v_y	State variable of x-y velocity.
b	Measurement noise.
A	Covariance of a.
B	Covariance of b.
e	Error.
K	Kalman gain.
I	Identity matrix.
(x,y)	Cartesian coordinates of the source.
(x_i,y_i)	Cartesian coordinates of i^{th} sensor.
τ	Time difference.
v	Wave propagation velocity.
r	Radius of time difference circle.
σ	Standard deviation of estimated location.

Chapter 1

Introduction

Human Computer Interaction (HCI) is of one of two types, tangible (touchable) and intangible (or non-touchable, such as audio or video). The vast majority of input interfaces are tangible. A tangible user interface is a user interface in which a person interacts with digital information through a physical environment, achieving seamless coupling between these two very different worlds of bits and atoms. Currently, the most important forms of such tangible interfaces include keyboards, mice, touch pads and touch screens.

The concept of tangible acoustic interfaces, as depicted in figure 1.1, is that any physical contact with a solid object or a surface (wall, table, etc.) will modify its acoustic pattern by the way acoustic energy is distributed in the object and on its surfaces. Such perturbation of the acoustic pattern can be caused in two ways: passively, by the acoustic vibration generated at the points of contact when tapping or moving a finger on the surface of the object and, actively, by the sound energy that is absorbed at the points of contact (proportional to contact pressure) when the object is acoustically activated.

The ultimate challenge in human computer interaction is to create tangible interfaces that will make the binary world accessible through augmented physical surfaces like

walls and windows, graspable objects like models and equipments, and possibly ambient media like air and water.

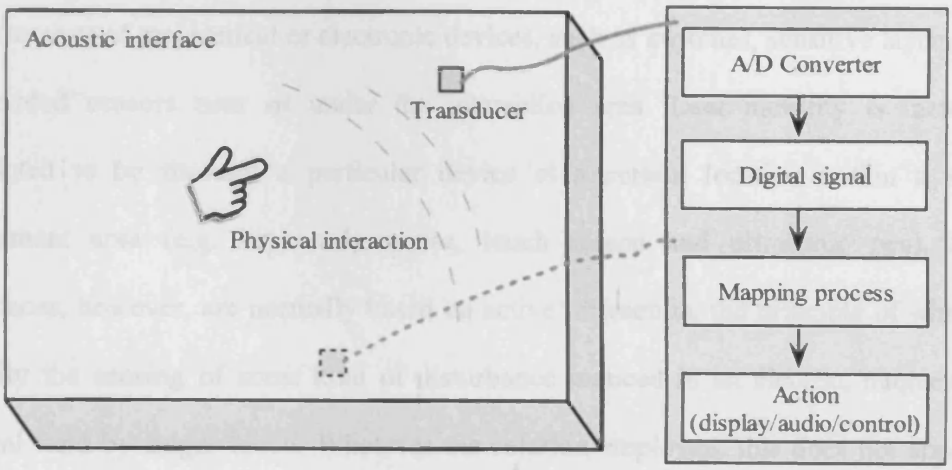


Figure 1.1. Acoustic HCI model.

Different solutions have been suggested to tackle the problem of passive source localisation. They normally rely on the acquisition of time-delayed replicas of a source signal at spatially distributed sensors, or learning and memorising predefined locations from specific features in the signals.

1.1 Motivation

The available commercial interface devices are part of everyday life as they are found in all kinds of electronic devices, from personal computers, to interactive kiosks, to digital personal assistants or bank tellers. A common problem with these interfaces is the presence of mechanical or electronic devices, such as switches, sensitive layers, and embedded sensors near or under the interaction area. User mobility is therefore restricted to be through a particular device at a certain location within a small movement area (e.g. keyboard, mouse, touch screen and ultrasonic pen). Such interfaces, however, are normally based on active interaction, the principle of which is usually the sensing of some kind of disturbance induced in an electric, magnetic or optical field by finger touch. Whatever the solution employed, this does not apply to arbitrary surface materials and sizes, but only to a narrow range of specific materials deliberately made to be an interface device, which usually embed some distributed electronics. An ideal solution would be to convert virtually any tangible object such as table tops, walls and windows into interactive surfaces.

The motivation for embracing passive acoustic based remote sensing technology as the solution is that acoustic vibration is the natural outcome of an interaction and propagates well in most materials. Passive acoustic technology relies on the analysis of the acoustic vibrations generated at the points of contact. These methods are more promising if the requirement is to develop new touch-based interfaces, that have to be scalable in dimensions, cheap, and built with materials and devices that allow them to be suitable for any condition and environment. The advantage of this new sensing

paradigm over other methods of interaction such as computer vision and speech recognition implies significant potential for the computer industry. New applications can include wall-size touch panels, three-dimensional interfaces and robust interactive screens for harsh industrial environments. Potential applications that may benefit from acoustic interface devices are:

- Low cost desktop keyboards and consumer keypads.
- A virtual mouse where the user moves his/her finger directly on the surface of the table.
- Device-free electronic whiteboards and pointing systems, using just conventional pens and fingers.
- Low cost consumer interfaces for PCs, video games, entertainment, etc.
- Large scale interactive screens for academic presentations or educational purposes.
- Interactive windows in shops or public services.
- Interactive interfaces for spectators such as visitors to museums, information centres, and exhibitions.
- Part of the interactive environment of Virtual Reality.
- Sensitive skin for robots and sensitive surfaces in multi-transducer environments.
- Interactive interfaces for security systems (interactive doors, windows, walls and floors).
- Robust interactive screens used in harsh conditions, such as underwater, open fields, high/low temperature environments, and public environments.

- Robust touch sensitive surfaces for industrial condition monitoring, for instance automatic counting and positioning of objects.
- Novel interactive artistic interfaces (interactive painting, photographs and sculptures).
- Novel musical instruments.

1.2 Research Aim and Objectives

The aim was to develop a human computer interaction system capable of using an arbitrary solid object like a wall, table or window as a tangible interface with a natural interaction by a finger or an ordinary device like a wooden stick. Acoustic waves produced from this interaction were utilised as a natural resource conveying source location information to sensors attached to the object surface. Signals acquired by sensors are employed to estimate the location of the impact or the trajectory of the continuous hand movement in an analogous way to conventional PC mouse actions. In this manner, the system does not require any external energy source or embedded electronics.

Although various passive source localisation techniques have been developed, particularly for in-air and underwater applications, not all of them are appropriate for tangible acoustic interfaces because of fundamental differences in the application requirements and the physical properties of the transmission medium. Therefore, an essential task in this work was to study the potential applicability of different techniques.

The objectives of this work were:

- Convert tangible objects into interactive interfaces.
- Employ passive sensing technology suitable for various materials.
- Develop localisation algorithms that can respond to discrete (as impact) and continuous (as scratch) type of interactions.
- Resolve problems associated with acoustic propagation in solids.

1.3 Hypotheses

The main hypothesis to be proved was that the adoption of redundant sensors and sensor fusion techniques improved the accuracy of localisation. Another hypothesis was that in-air acoustic localisation algorithms can be adopted for use in in-solid applications.

1.4 Methodology

The above hypotheses were proved using a combination of theory and experimentation. Different mathematical models were developed and implemented in algorithms which were then tested. The techniques employed included:

- The adoption of the coherence function as an alternative to cross correlation for better performance with a single sensor system and the use of a hypothetical transfer function to resolve reliability problems using two sensors.

- The application of Hilbert envelope filtering in addition to phase transform filtering to achieve enhanced impact localisation.
- The computation of the linear cross spectral phase to estimate time differences of arrival.
- The smoothing of estimated trajectories using Kalman filtering.

1.4 Thesis Outline

This thesis comprises into seven chapters and is organised as follows.

In the following chapter, the state of the art of commercial tangible interfaces is presented and this is followed by a description of the research work carried out on large interactive surfaces. A literature review on the theory of source localisation and the basic physics of acoustic propagation in solids is then provided. Chapter three is devoted to the topic of localisation based on pattern matching including cross correlation, coherence and the conceptual transfer function. In chapter four, popular localisation techniques are analysed and compared to identify methods that are not suitable for tangible acoustic interfaces based on their physical properties and practical limitations. Enhanced impact localisation using one-step and two-step techniques is proposed in chapter five. A computationally less expensive algorithm is proposed for tracking a continuous source with position filtering using a Kalman prediction algorithm is presented in chapter six. Finally, chapter seven summarises the conclusions and contributions of the research, and gives suggestions for further investigation.

Chapter 2

Literature Review

The vast majority of HCI systems are tangible, and different detection methods of interaction have been developed which can be divided into two groups, passive and active. Most of the commercial HCI products are active while most of the localisation techniques used in modern applications for in-air and underwater acoustics are passive. The attributes of passive systems make them valuable and attractive for HCI applications, which have triggered some ideas for employing them in this way. There is a wide range of techniques for passive localisation, particularly for in-air applications but as yet they have not been comprehensively evaluated for HCI development, which is one of the tasks in this work. In this chapter, the state of the art of commercial tangible interfaces is first presented followed by a review of the research work carried out on acoustic interactive objects. A comprehensive literature review on the techniques used in solving the source localisation problem is provided and finally, a basic physics of acoustic wave propagation in solids is given.

2.1 Commercial Tangible Interfaces

Tangible interfaces can be divided into two broad categories: passive and active. Passive systems do not require a special pointer to inject energy or a source to activate

the touch surface. The user's finger or a solid object is sufficient. Active systems, on the other hand, require a special device which can emit some kind of energy either at the point of contact or by exciting the area of interest of the tangible surface.

The popular commercial tangible interfaces are touch screens, meaning that glass material is used as a substance. The technologies used in passive systems are mainly resistive and optical, and the technologies used in active systems include capacitive, infrared, ultrasonic and Surface Acoustic Wave. The principles of these technologies can be described as follows [1,2,3],

- Resistive

Resistive is the most common type of touchscreen technology. It is a low-cost solution found in many touchscreen applications, including hand-held computers, PDAs, consumer electronics, and point-of-sale applications. It is ideal for screen sizes up to 12.1". A resistive touch screen uses a controller and a specially coated glass overlay on the display face to produce the touch connection as in figure 2.1. Resistive touch screens substantially reduce light throughput and thus provide lower image clarity.

- Optical

Optical technology used to develop tangible interfaces based on using four cameras at the corners of the tangible object as shown in figure 2.2. Practically, this visual system may not suite all surfaces, particularly those with reflections.

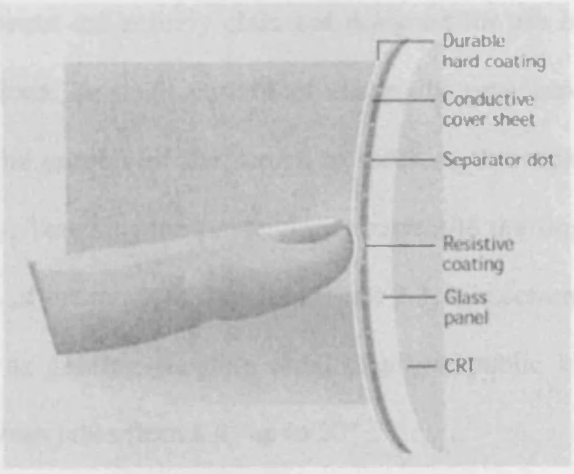


Figure 2.1. Resistive touch screen.

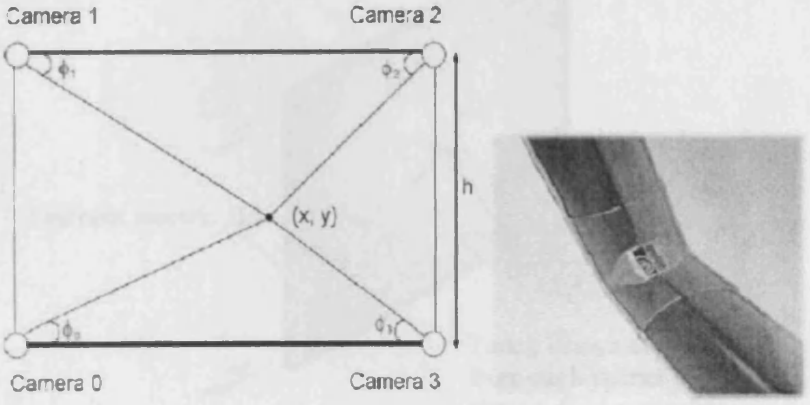


figure 2.2 Optical touch screen.

➢ Capacitive

Capacitive touch screens are entirely glass and designed for use in ATMs and similar kiosk type applications. A small current of electricity runs across the screen with circuits located at the corners of the screen to measure the capacitance of a person touching the overlay. Touching the screen draws current to the finger and the software calculates a position of the touch as shown in figure 2.3. It is commonly used in harsher environments such as gaming, vending retail displays, public kiosks and industrial applications with screen sizes from 8.4″ up to 20″.

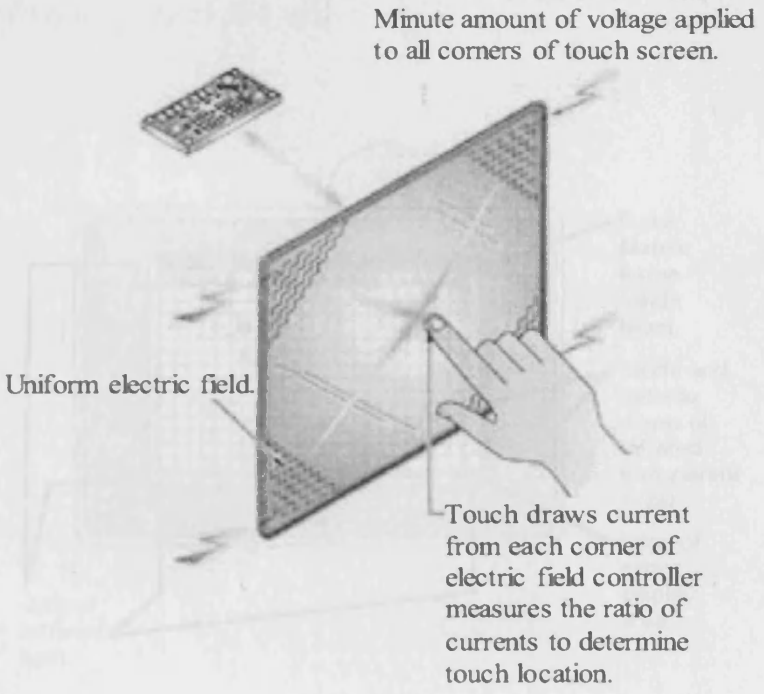


Figure 2.3 Capacitive touch screen.

➢ Infrared (IR)

IR technology relies on the interruption of an IR light grid in front of the display screen. The touch frame contains a row of IR light emitting diodes (LEDs) and phototransistors, each mounted on opposite sides in both X and Y dimensions, to create a grid of invisible infrared light. The IR controller sequentially pulses the LEDs in each dimension and when a stylus, such as a finger, enters the grid, it obstructs one or more of the beams in each dimension. The phototransistors, as in figure 2.4, detect the absence of IR light and transmit a signal that identifies the X and Y coordinates.

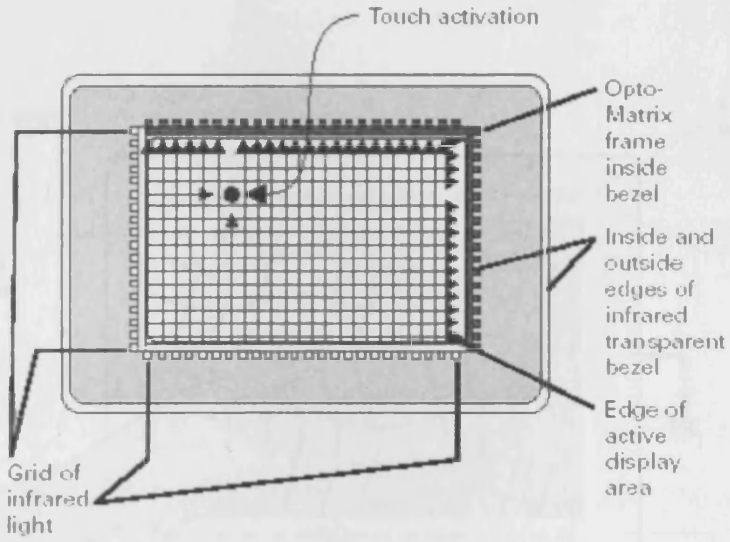


Figure 2.4 Infrared touch screen

➢ Surface Acoustic Wave (SAW)

SAW technology provides better image clarity because it uses a pure glass construction. A SAW touch screen uses a glass display overlay and is suitable for screen sizes from 8.4" up to 20". When sound waves are transmitted across the surface of the display, each wave is spread across the screen by bouncing off reflector arrays along the edges of the overlay. Two receivers detect the waves. When the user touches the glass surface, as in figure 2.5, the user's finger absorbs some of the energy of the acoustic wave and the controller circuitry measures the touch location. SAW Touchscreen technology is used in ATMs, Amusement Parks, Banking and Financial Applications and kiosks. Compared to resistive and capacitive technologies, it provides superior image clarity, resolution, and higher light transmission.

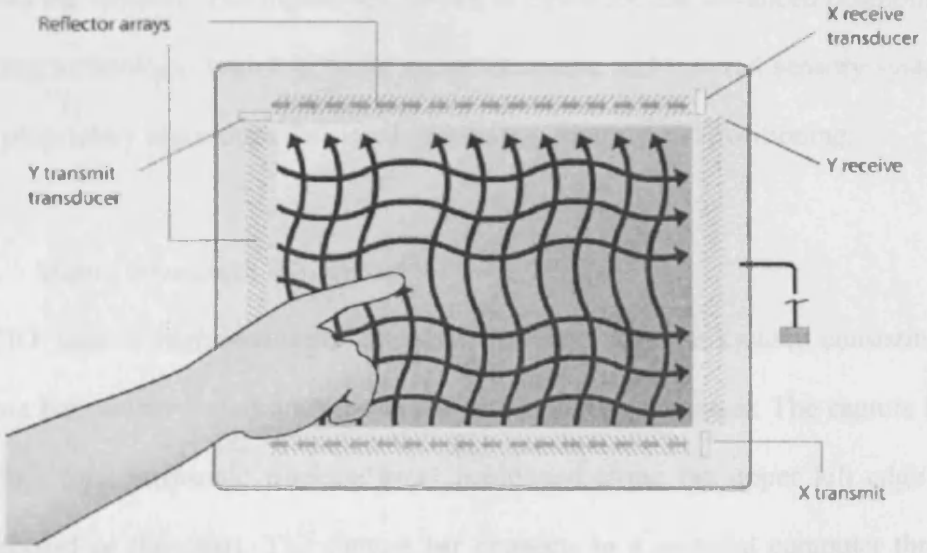


Figure 2.5 SAW touch screen

Other hybrid technologies also exist, such as using infrared with optical. NextWindow's optical touch screen technology, [4], uses two line scanning cameras located at the corners of the screen. The cameras track the movement of any object close to the surface by detecting the interruption of an infra-red light source.

Apart from touch screens, other interactive objects are also of interest. These are active systems developed for specific applications, such as Pegasus PC Notes Tracker [5], Mimio Whiteboard [6] and Virtual Laser Keyboard [7].

- Pegasus PC Notes Tracker

PC Notes Taker is a device that captures natural handwriting on any surface onto a PC in real time. Pegasus technology utilises ultrasonic transmission, from the tip of the pen, and time measurement of the duration from when the pulse leaves the pen until it reaches the receiver. The digital pen, shown in figure 2.6 has advanced positioning and tracking technology, which is based on an ultrasonic and infrared sensory system and uses proprietary algorithms for signal processing, filtering and positioning.

- Mimio Interactive Whiteboard

MIMIO uses a high-resolution ultrasonic position capture system consisting of a capture bar, colour-coded marker sleeves and an electronic eraser. The capture bar is a two-foot long ultrasonic tracking array positioned along the upper left edge of the whiteboard or flip chart. The capture bar connects to a personal computer through a serial or USB interface cable. The electronic marker sleeves transmit an ultrasonic signal to the capture bar, which triangulates the pen's position on the board as the user

writes. The only change users must make is to be sure they use the electronic eraser to make corrections. The system, as pictured in figure 2.7, captures each move of a marker or stylus on the whiteboard or flip chart surface as digital data that expresses vector strokes over time, which are then interpreted by software.

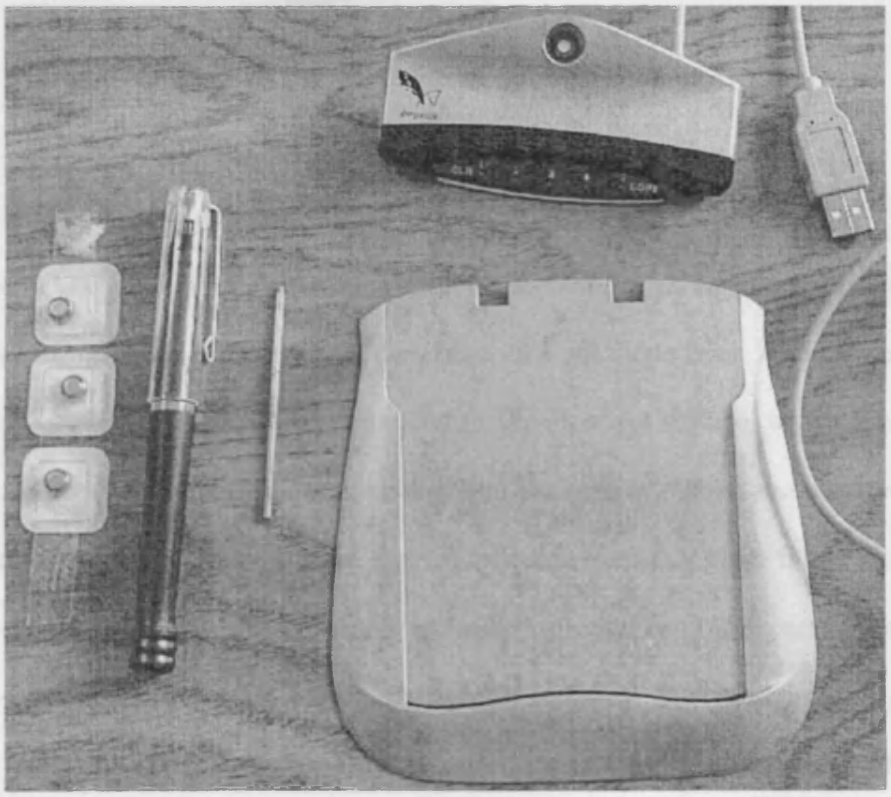


Figure 2.6. Pegaus PC Notes Tracker.

Figure 2.7., MIMIO Interactive Whiteboard [6].

➢ Virtual Laser Keyboard

The virtual keyboard, as shown in figure 2.8, consists of projecting a keyboard template by illuminating a holographic optical element with a red diode laser. An infra-red plane of light is generated just above, and parallel to, the interface surface, which is invisible to the user and hovers a few millimetres above the surface. When the user touches a key position on the interface surface, light is reflected from this plane in the vicinity of the key and directed towards the sensor module. Reflected light from user interactions with the interface surface is passed through an infra-red filter and imaged on to a CMOS image sensor in the sensor module. Embedded hardware in the sensor chip then makes a real-time determination of the location of the reflected light. The processing core can track multiple reflection events simultaneously and can thus support both multiple keystrokes and overlapping cursor control inputs.

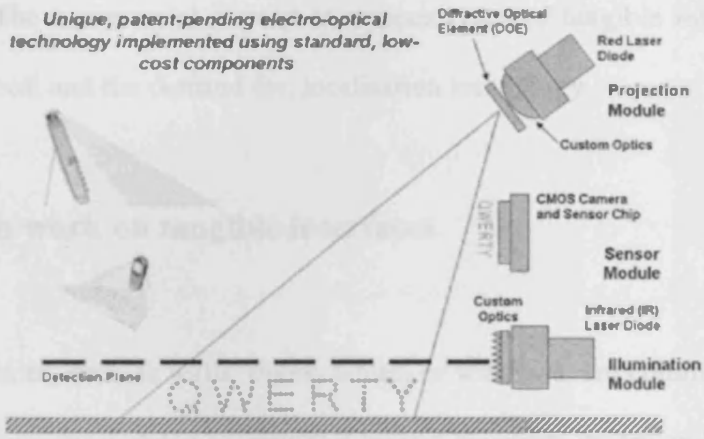


Figure 2.8 Virtual Laser Keyboard.

Recently, after this work began, passive acoustic technologies have emerged on the market based on utilising the vibration caused by a touch to determine its location. The technology presented by SensetiveObject [8] is based on the recognition of sound waves propagated in an object when the user touches a defined zone. A tap on an object produces a pattern of sound waves through the material. This pattern creates an acoustic signature that is unique to the zone of the impact. Acoustic sensors linked to a computer are used to capture the audio vibrations within an object and generate acoustic signatures. The zone of a hit can then be determined by mapping the best matched signature. The technology is applicable to different materials and demonstrated with keyboard and control switch applications but all zones of interest on the object surface must be trained first. Another product presented by 3M [9] and i-vibrations [10] calculates the coordinates of a nail click on a glass surface from the vibration signals received by sensors at different times. The applications include shop windows and display cases. So far, no technical information is available about these

technologies. The commercial interest in various types of tangible interfaces reflects the importance of, and the demand for, localisation technology.

2.2 Research work on tangible interfaces

Large flat surfaces, such as walls, floors, tables, or windows, are common structures in everyday life, usually dictated by practical human necessity or driven by architectural design. At present, these surfaces are used to display decorative items such as paintings, photographs, and rugs. It is unusual to see large portions of the walls, floors, or windows themselves used directly as interactive surfaces. Therefore, it is desired to develop new technologies that will enable such architectural surfaces to become sensate. User interaction with large surfaces is a topic of considerable interest in HCI and among the ubiquitous computing communities [11,12]. In contrast with other interactive object technologies, sensitive skin is a large area with a flexible array of sensors having data processing capabilities, which can be used to cover the entire surface of a machine or even a part of a human body or a robot [13].

In late nineties a group of researchers at MIT developed a passive interactive ping pong table called PingPongPlus (PP+) [14]. The goal is to visualise each ball impact location on the table by certain graphics, such as water waves, via video projector which is controlled by a PC that determines the impact location. As shown in figure 2.9, when a ball hits the table, the sound travels through the table to eight microphones mounted on the underside of the table. The picked-up signals are passed through an operational amplifier to a set of comparators and OR gates, where hardware peak thresholding on

signals is performed by comparing the signal's absolute value against a fixed threshold voltage. The comparator pair returns true to a PIC chip if there is an impact. If there is a hit, the PIC chip assigns a time value to that microphone input and sends a microphone number along with its associated time value to the host PC where the received information is evaluated by a time difference algorithm that determines the location of the hit. The hyperbolic intersection algorithm, which is calibration-free, has been tried then replaced by a training-based algorithm. The model of the latter algorithm is given by AX = Y, where Y is the ball landing coordinate vector, X is the time difference information and A is the model parameter that needs to be performed once for a given table, unless the microphone placement is changed. Training data is acquired by dropping a ping-pong ball on certain known spots on the table a number of times then matrix A was calculated through a least-squares fit to this data. When an impact occurs, the time differences are multiplied by the model parameters which returns a ball landing coordinate.

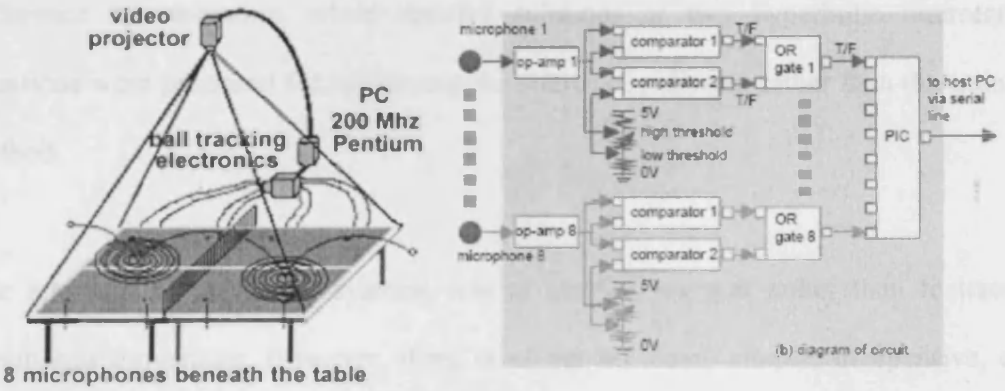


Figure 2.9. MIT PingPongPlus set up and electronic circuit diagram.

Another research for tracking knocks on large windows as depicted in figure 2.10 was carried out for two Master degree projects at MIT which was motivated by the PP+ project. However, the development of this application required more accurate time differences than PP+ as impacts on glass are much less distinct than those of a ping-pong ball on wood and the dispersion creates significant distortion in the received audio waveforms. The modifications introduced by [15] include hardware signal filtration and software time difference calculations by sending the conditioned signals to a PC via a data acquisition board, which provides flexible signal processing such as normalising the signals before thresholding. The arrival time determination algorithm locates the first peak that is above a certain level, rather than a fixed, threshold. The average of two cross correlation peaks was also suggested for timing. The training method as in the PP+ was used with higher order polynomial data fitting. Further improvements to the knock tracking system on a large window followed in [16], where the threshold used for measuring the time differences was defined as a function of the signal peak. Cross correlation on part of the signal with few heuristics was also proposed for the time difference determination, while iterative solutions of two hyperbolic intersection equations were proposed for calculating the source coordinates rather than the training method.

The intention of the above systems was to localise impacts rather than to trace a continuous movement. However, there is as yet no clean, simple, inexpensive, and general means of robustly tracing the movement of bare hands near or on the surface of large objects. However, a solution to this problem that has been developed by MIT researchers [17] is to place a scanning laser rangefinder, as illustrated in figure 2.11, at

one corner of the display surface to determine the polar coordinates of hands in a clearly defined plane above the projection surface. This would ideally be a compact device, enabling a simple retrofit to make any flat surface interactive. One unit, in a single corner of the screen is able to scan the entire display surface, and because it produces two coordinates simultaneously, there is no correspondence problem with multiple hands. However, there are still occultation issues and also the system is not receptive to the tapping type of interaction.

As can be observed from the above review, most of the interactive systems are active. Passive acoustic technology, whether developed for a commercial product or for research work is limited to user interaction with single impacts and not for tracing a continuous movement and mainly intended for a glass substance.

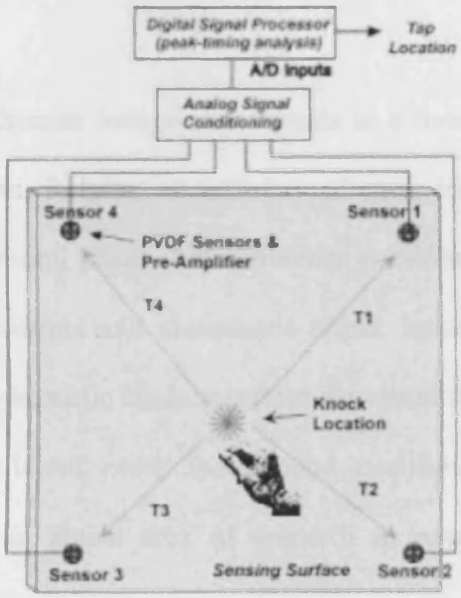


Figure 2.10. MIT Knock localiser on large window

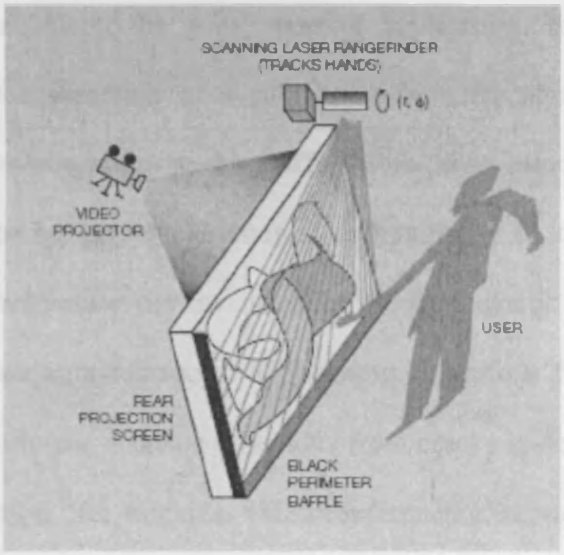


Figure 2.11. MIT Laser tracing interactive screen

2.3 Literature Survey on Source Localisation

Binaural localisation in human beings and animals is a live example of an efficient source localisation system. Relying on a variety of cues, including interaural power level difference and interaural phase shift difference between the two ears, as well as information from the spectrum and precedence effect, human brains create a three-dimensional image of the acoustic landscape from the sound they hear. After more than a century of work, there is still much about sound localisation in humans that is not understood. It remains an active area of research in psychoacoustics and in the physiology of hearing [18].

The technology of source localisation in general has already received significant attention for decades due to its wide areas of application, including: underwater acoustics, such as the detection of a submarine from its noise or via ultrasound scanning since electromagnetic waves suffer from high attenuation in seawater; seismology, to locate an earthquake source; and military, to detect airplanes with passive acoustic radar before the invention of electromagnetic radar. Nevertheless, source localisation has attracted recent applications in various fields, including: non destructive testing, to locate acoustic emissions from cracks in solids such as airplane wings; voice localisation, for example, video conferencing, service robotics, e-textiles and hands-free speech communication; wireless sensors network, e.g. environmental monitoring, surveillance and security; impulsive noise, e.g. machinery fault diagnosis; and radio frequency such as aerospace, cellular phones, electronic warfare and electromagnetic compatibility testing.

In most active systems, as RF sources, the waveforms are narrowband since the ratio of the highest to the lowest frequency is usually very close to unity (e.g., for the 802.11b wireless LAN system, the ratio is 2.4835 GHz/2.GHz=1.03) which means the signals have a well defined nominal wavelength and therefore the time delay can be compensated for by a simple phase shift. On the other hand, audio waveforms in the range of 100 Hz to 10 KHz having a ratio of 100 are considered wideband signals and interpolation approaches are therefore required to estimate the time delay.

When an acoustic source is located close to the sensors, the wave front of the received signal is curved and the curvature depends on the source distance, then the source is in

the near-field. As the distances become large, all the wave fronts are planar and parallel, and the source is in the far field. For a far-field source, only the direction-of-arrival angle in the coordinate system of the sensors is observable to characterise the source. A simple example is when the sensors are placed in a line with uniform inter-sensor spacing, then all adjacent sensors have the same time delay and the direction of arrival of the far-field source can be estimated readily from the time delay. For a near-field source, the collection of all relative time delays and the propagation speed of the source can be used to determine the source location.

Classical source localisation can be categorised into three major methods; time of arrival, time difference of arrival (TDOA) and angle (or direction) of arrival. Time-of-arrival methods are based on measuring the time delay between the transmitter and the receiver, hence the source has to be active. On the other hand, TDOA is obtained from passive sources by measuring the time difference of signals arriving at multiple receivers. Readings of multiple times of arrival or TDOA's can be integrated and the source can be located by triangulation. Accordingly, two hot areas in research were initiated, one that deals with the improvement of TDOA estimation and another that deals with the triangulation and data fusion to improve the estimation of the source coordinates. Basically, the angle of arrival can be obtained from the TDOA between two sensors using the concept of array beamforming, which is the most popular technique in radar systems but has also been employed in recent applications.

The source in most applications is located in the far field, hence beamforming is the most popular and well established approach for estimating the angle of arrival [19]

particularly for narrowband signals. It is an array signal processing technique that has been well studied for more than two decades [20]. Beamforming is based on a one-step procedure by finding the bearing of the maximum energy driven from time delayed, filtered, weighted and summed versions of the received signals, which form a single output signal. The estimator virtually steers the beam of the array to various locations and searches for a peak in output power. It has the advantage of accuracy and has the potential for detecting multiple sources but is time consuming and normally requires at least eight sensors. For example, in [21] eight microphones are used in beamforming to localise sound sources as a complement to the vision system in a mobile robot. In [22], a PhD work is carried out for developing angle-of-arrival estimation algorithms for wideband signals.

The technique for most passive sound source localisation systems using a microphone array is a two-step procedure. First the TDOA in microphone pairs of the sensor array are estimated, usually by cross-correlation-based technique. In a second step, these TDOAs are used together with the microphone array geometry to determine the position of the sound source. TDOA estimation has been very profoundly described in an IEEE special issue [23]. The most common technique used to estimate the TDOAs is the Generalised Cross Correlation method [24]. It is a computationally efficient method which involves performing a cross correlation process in the frequency domain with various filtering criteria. A popular filtering process used in room acoustics to overcome the reverberation problem is the Cross-Power Spectrum Phase [25,26]. An alternative but less common technique is the Adaptive Eigenvalue Decomposition [27]

proposed for room acoustics. It outperforms the Generalised Cross Correlation technique in moderately reverberant rooms but the convergence time is considerable.

Various positioning techniques have been developed to be used in the second step to locate the source based on the given TDOAs. Interesting to mention here is the positioning technique used in the Great War by British troops on the battlefield to find the location of the enemy's gunfire by measuring TDOAs from recorded signals received from microphones using a kinematograph. Target location was then found from manually plotting the TDOA triangulations [28].

The fundamental positioning method is based on the intersection of two hyperbolas defined by two TDOA information obtained from three sensors. Solving hyperbolic equations is a highly nonlinear problem and sensitive to the errors in TDOAs. Research has been active to find a closed form solution and also to develop techniques that utilise more than two TDOA measurements to reduce the error in the estimated location. Taylor-Serious [29] and spherical interpolation [30] have been proposed to make the triangulations of the localisation equations linear. Maximum Likelihood is a least-error estimate computed from the measured TDOAs and the true TDOAs. This method has been implemented to find a speaker location in a room [31,32].

The two-step procedure has been extensively implemented in algorithms for various in-air applications where a common problem with these applications of in-room acoustic localisation is the reverberation. A framework for designing a wearable microphone array for localising speech source or vehicle is presented in a piece of Master research

[33] where the triangulation of two angles of arrival from two arrays is used to locate the source. Sound localisation in a network of sensors has been designed using two arrays each of two microphones and different localisation algorithms were compared [34]. Acoustic localisation algorithms were proposed for service robots working in real conditions. One of the main uses of these algorithms in a mobile service robot is that the robot can localise a human operator and eventually interact with them by means of verbal commands [35]. In [36] the two-step procedure is proposed for passive source localisation to steer a video camera towards a speaker in a conference room. Using a cross correlation and hyperbolic intersection, the method has also been used to determine the source location in thin plates [37].

Recent advances in acoustic localisation have combined the advantages of the traditional methods of beamforming and TDOA, leading to techniques that are both accurate and fast. A procedure proposed for room acoustics estimates the location in one-step [38]. This method implies maximising the likelihood of the source location based on the data received from multiple sensor pairs, where the TDOA is associated with the estimated location rather than having been computed individually. The method has the advantage of including filtering criterion when implemented in the frequency domain [39]. Least-error is another localisation technique based on minimising the error between the measured and the ideal time delay [40].

Sensor fusion is an important concept in source localisation since it provides a mechanism for integrating extra information from redundant sensors to improve the accuracy of the estimated location. Least-error and maximum likelihood are examples

of sensor fusion but have the disadvantage of being nonlinear. Linear least square is an optimisation technique that can be used with a linear positioning algorithm and one that is popular in wireless localisation [41].

The problem of the passive tracking of a moving source is encountered in many diverse applications. In passive tracking, sensors "listen" for the signal emitted by the source in order to determine its location. Unlike radar systems, passive tracking systems have a stealthy operation capability. Passive tracking is achievable by employing TDOA between signals received at multiple receivers or from bearing measurements. However, because outliers in the estimated locations crucially affect the trajectory, it is essential to use a location prediction algorithm. Among the algorithms that have been proposed for tracking a moving source such as sound and mobile phones are recursive smoothing [42], particle filter [43,44] and the popular Kalman prediction [45].

2.4 Overview of Acoustic Wave in Solids

Sound is an acoustic wave pattern which is caused when a sound source disturbs the normal random pattern of the molecules in air or in any other molecular medium, such as a liquid or solid. Sound waves propagate in a variety of media, including gas, liquids, organic and inorganic solids, in plasmas and superconductors, and in interplanetary, interstellar, and intergalactic media. They range in frequency from billions of cycles per second to a single cycle within a period of years. Human hearing is only a small fraction of the frequency range of the full acoustic wave spectrum. These waves propagate in a variety of media at different speeds, and have different vibration

characteristics depending on their source. They include traveling waves, standing waves, internal waves, surface waves, trapped waves, thermal waves, shock waves, and plasma waves which are included in the acoustic wave spectrum represented in the chart shown in figure 2.12 [46].

In air, sound travels by the compression and rarefaction of air molecules in the direction of travel. In solids, waves can be characterised by oscillatory patterns that are capable of maintaining their shape and propagating in a stable manner and molecules that can support vibrations in other directions. A number of different propagation modes therefore exist. There are four principle modes in solids based on the way the particles oscillate; longitudinal wave, transverse (or shear) wave, surface wave and plate wave. In contrast with wave propagation in seismological science [47], there are two main types of waves; body wave and surface wave. A body wave is one traveling through the interior of the medium and is of two kinds; longitudinal, designated P (primary) wave and transverse, designated as S (secondary) wave. A surface wave travels only through the surface of the medium and is of two kinds; a Love wave and a Rayleigh wave. Body waves arrive before the surface waves and they are of a higher frequency.

Minutes **1 – 60**	**Cyclones and Anticyclones** high and low pressure systems; tropical cyclones including hurricanes, typhoons; circulation patterns caused by vertical effects **Global Cycles** atmospheric tides, cloud cycles, sea/land breezes, hill/valley flows	**Trapped Waves** within the Earth's liquid core **Tsunamis** (tidal waves) caused by seismic activity beneath the ocean floor	**Thermal Cycles** expansion and contraction of solids due to fluctuations in temperature	**Circadian Rhythms** of plants and animals	**Electroacoustic Waves** caused by local turbulence within the solar wind
Seconds **1 – 60**	**Cyclonic Storm Cycles** tornadoes, waterspouts **Cyclonic Minicycles** caused by vertical effects, wind, and turbulence **Caustics** caused by boundary interactions	**Surface Waves** 'ring' waves, 'breaking' waves, storm waves, 'rollers', seichs **Tsunamis** (tidal waves) **Tidal Bores** **Internal Waves** **Trapped Waves**	**Surface Waves** snow, sand, and dust waves **Seismic Waves** generated on the surface crust of the Earth or other terrestrial bodies **Seismic Waves** generated within the interior of the Earth or other terrestrial bodies	**Surface Waves** waves of grain and grass, etc. **Biomechanical Cycles** heartbeat, respiration, metabolic and other biomechanical rhythms internal to humans and other species	**Electroacoustic Waves** caused by local turbulence within the solar wind
Normal sound range ***(include the range of human hearing)*** **1 – 100 k cps**	**Internal Waves** caused by turbulence **Radial Waves** **Shock Waves** **Wind and Whistle Tones** **Internal Waves** caused by spontaneous boundary interactions **Acoustic Whistlers** **Cylindrical and Conical Waves** **Normal Sound Waves** in air and other gas	**Caustics** caused by complex boundary interactions **Deep Channel Waves** **Internal Waves** caused by acoustic cavitation **Internal Waves** caused by spontaneous boundary interactions **Normal Sound Waves** in water and other liquids	**Surface Waves** circular waves, eliptical waves, cymatic patterns generated on the surface of a solid body **Mechanical Oscillations** generated on or within a solid body **Normal Sound Waves** which are transmitted through a solid	**Biochemical Oscillations** spontaneously generated within humans and other living systems **Traveling Waves** spontaneously generated within living systems **Internal Waves** caused by acoustic cavitation **Normal Sound Waves** which are transmitted through organic substances	**Electroacoustic Waves** caused by local turbulence within the solar wind
Microacoustic 100 k – millions cps	**Microthermal Waves** trapped within normal sound waves **Opticoacoustic Waves** **Microcycles** the oscillation of particles and subparticles	**Microthermal Waves** trapped within normal sound waves in liquids and fluids **Microcycles** the oscillation of particles and subparticles	**Stress Waves** in crystalline structures **Microcycles** the oscillation of particles and subparticles	**Microthermal Waves** trapped within normal sound waves in organic substances **Ultrasound Waves** **Microcycles** the oscillation of particles and subparticles	**Electroacoustic Waves** caused by local turbulence within the solar wind **Magnetoacoustic Waves** caused by the interaction of the solar wind with the magnetic fields of various planets and satellites **Microcycles** the oscillation of particles and subparticles
Frequency in cycles	Air & other gas	Liquids & fluids	Solids	Organic substances	Interplanetary Space

Figure 2.12. Acoustic wave spectrum

As shown in figure 2.13, the oscillations occur in longitudinal waves in the longitudinal direction or the direction of wave propagation. It can be generated in liquids as well as solids because the energy travels through the atomic structure by a series of compression and expansion movements. In the transverse or shear wave, the particles oscillate at right angles, or transverse, to the direction of propagation. Shear waves require an acoustically solid material for effective propagation, and therefore are not effectively propagated in materials such as liquids or gasses. Shear waves are relatively weak when compared to longitudinal waves. In fact, shear waves are usually generated in materials using some of the energy from longitudinal waves. Longitudinal and shear waves are the two modes of propagation most widely used in active non destructive testing using ultrasonics [48].

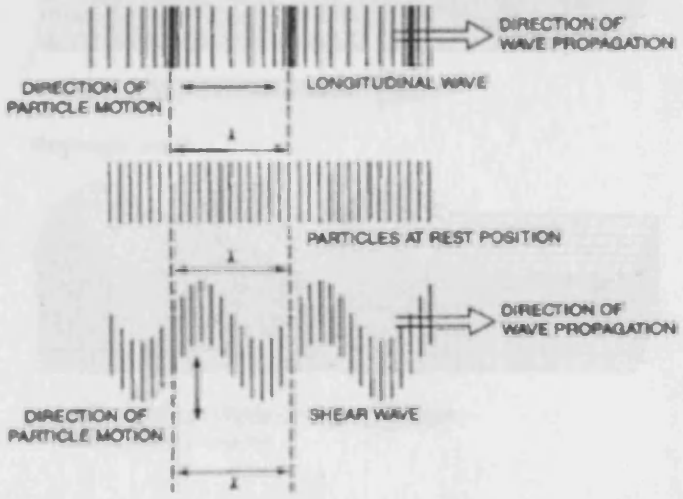


Figure 2.13 Longitudinal and shear body waves.

As illustrated in figure 2.14, Love surface waves produce an entirely horizontal motion with respect to the direction of propagation. Rayleigh surface waves travel on the surface of a relatively thick solid material, penetrating to a depth of one wavelength. The particle movement has an elliptical orbit. Rayleigh waves are useful in non destructive testing because they are very sensitive to surface defects and they follow the surface around curves. They can, therefore, be used to inspect areas that other waves might have difficulty reaching. Plate waves can be propagated only in very thin metals.

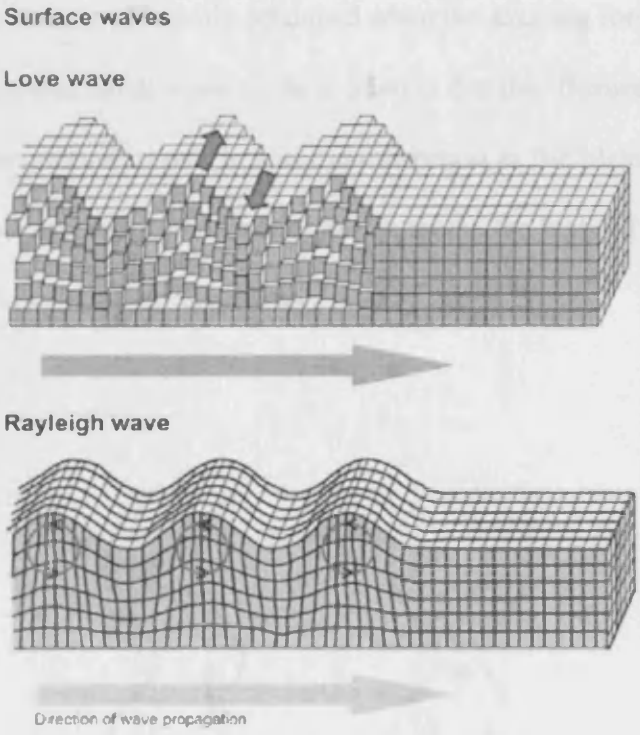


Figure 2.14. Love and Rayleigh surface waves.

Lamb waves are the most commonly used plate waves in non destructive testing. Lamb waves are complex vibration waves that travel through the entire thickness of a material. Propagation of Lamb waves depends on the density and the elastic material properties of a medium. They are also influenced a great deal by the test frequency and material thickness. With Lamb waves, a number of modes of particle vibration are possible, but the two most common are symmetrical and asymmetrical as shown in figure 2.15. The complex motion of the particles is similar to the elliptical orbits for surface waves. The symmetrical Lamb waves mode, also called the 'extensional mode' move in a symmetrical fashion about the median plane of the plate. Wave motion in the symmetrical mode is most efficiently produced when the exciting force is parallel to the plate. The asymmetrical Lamb wave mode is often called the 'flexural mode' because a large portion of the motion moves in a normal direction to the plate, and little motion occurs in the direction parallel to the plate. In this mode, the body of the plate bends as the two surfaces move in the same direction.

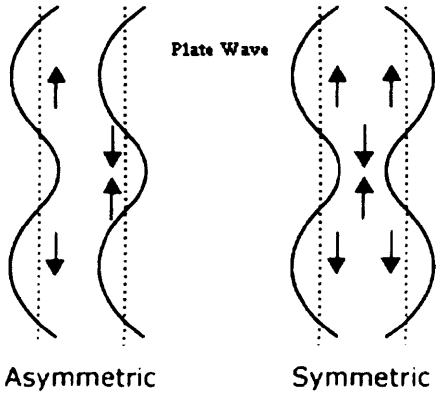


Figure 2.15 Modes of Lamb wave in plate

Wave velocity is essential for the location calculations. For an acoustic source, the propagation speed in air is a known constant of approximately 345 m/s. In [49], experimentation was carried out for measuring the ultrasonic wave velocity in Medium Density Fiber. It was proposed to measure the phase velocity based on the travel time of a specific phase point within the waveform and the group velocity based on the time of the waveform centroid, defined as the energy center of the wave as shown in figure 2.16.

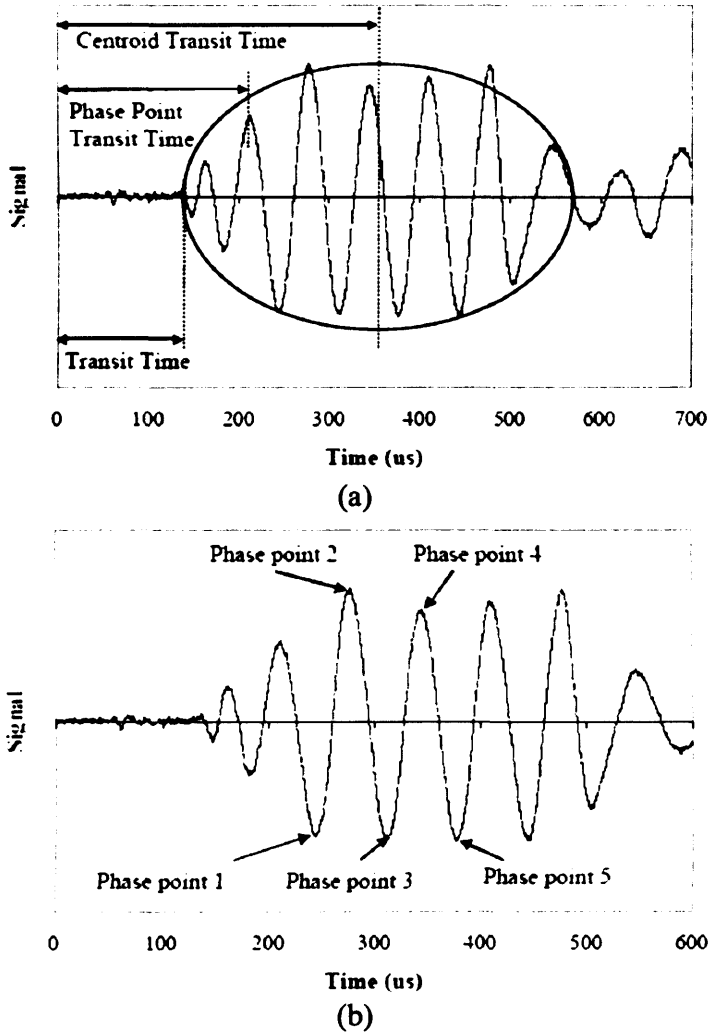


Figure 2.16. Transit time measurement for (a) group velocity and (b) phase velocity using ultrasound signal [49].

Another approach for the analysis of transient waves propagating in composite laminates is presented in [50]. A wavelet transform approach is proposed here for the time-frequency analysis of a dispersive plate wave. Figure 2.17 shows the results of the measured and the theoretical dispersion in unidirectional laminate.

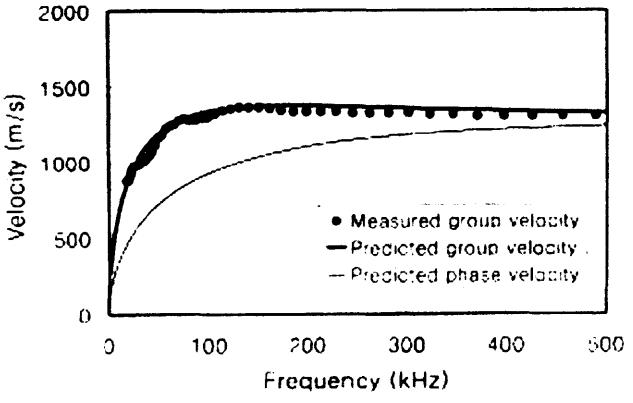


Figure 2.17 [50]. Flexural dispersion in unidirectional laminate.

CHAPTER 3

Enhanced Pattern Matching for Impact Localisation

The use of pattern recognition for source localisation is a new approach based on matching the pattern of the signal's features with a template associated with a predefined location rather than by calculating the actual coordinates of the source. The approach is named here, therefore, Location Pattern Matching (LPM). Although LPM requires a learning process for each location, it has the unique feature of being able to work in heterogeneous medium of any shape or material using one or two sensors. This feature overcomes the limitations of the more widely used approaches based on time difference. The use of cross correlation to match the time series pattern of the received signals has been verified experimentally in [51]. Maximum Likelihood estimation is proposed for pattern matching in [52] to localise an object in acoustically activated room. In [53], an ultrasound source is localised in a room with one receiver based on using Vector-distance Metrics to compare measurements with the simulated signature of location obtained from the room acoustic channel model. Cross correlation is more popular as a matching technique in other applications such as in the medical field, where it is used to compare test ECG signals with the database of known diagnoses to find the most similar waveform and hence the related cardiac information can be found [54]. In image processing, pointing finger is tracked for gesture interface by matching a searched image from a video camera with a finger template using two-dimensional cross correlation [55]. The method of pattern matching for localisation, although it is a

simple method, it has not been explored intensely in literature from an engineering point of view. In this chapter, pattern matching for in-solid localisation is thoroughly investigated and novel algorithms are proposed to solve the common problems of resolution and reliability.

3.1. Hypothesis of Pattern-based Localisation

The hypothesis behind identifying a location from the received signal pattern is that a signal propagating from source to destination inherits a specific signature in its pattern associated with it's source location as a result of scattering. In this section, the hypothesis is illustrated in a mathematical model.

When a driving force is applied to a medium, a travelling wave is generated transporting energy away from the source of disturbance. In a closed system, the waves propagate until they meet the boundaries and are reflected or absorbed. The reflections cause reverberation to the received signal acquired by a transducer at a certain location in the medium. To comprehend the effect of the boundary on the magnitude and phase of a received signal, a plane wave is assumed propagating in the x direction, and the acoustic wave equation is given by,

$$\frac{\partial^2 p}{\partial x^2} = \frac{1}{v_p{}^2}\frac{\partial^2 p}{\partial t^2} \tag{3.1}$$

where p is the acoustic pressure as a function of time t and distance x and v_p is the phase speed. The solution of the differential equation (3.1) is the travelling wave equation given by [56],

$$p = Ae^{j(\omega t - \beta x)} \tag{3.2}$$

where A is the amplitude constant and β is the wave number. With this wave equation, it can be shown how a transmission medium with one reflection affects the received signal. Assume a simple model of two signal paths from the transmitter to the receiver. One direct path with unity gain and a delay t_d, and the other is reflected with attenuation α and a delay $t_d + \Delta_t$ resulting from the path length difference. The overall transfer function of such a transmission medium $H(\omega)$ can be expressed by,

$$\begin{aligned} H(\omega) &= e^{-j\omega t_d} + \alpha e^{-j\omega(t_d+\Delta_t)} \\ &= \overbrace{\sqrt{1+\alpha^2+2\alpha\cos\omega\Delta_t}}^{|H(\omega)|} e^{\overbrace{-j(\omega t_d + \tan^{-1}\frac{\alpha\sin\omega\Delta_t}{1+\alpha\cos\omega\Delta_t})}^{\theta_h(\omega)}} \end{aligned} \tag{3.3}$$

Therefore, multi-path causes distortion in the magnitude $|H(\omega)|$ and the phase $\theta_h(\omega)$ characteristics of the transmission medium. Since the time delay is the product of the path length difference and the wave-number, it is clear from (3.1) that the magnitude and phase of the received signal at a certain location will vary as the source location changes. In a random medium the received signal is a combination of the direct wave and multiple delayed scattered waves that have gone through reflections and refractions

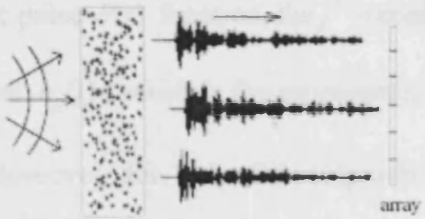
plus the effect of non-isotropic material, and therefore the received signal in reality has a much more complicated relation to its source location than that given in equation (3.1). With this hypothesis, signals received from different locations will have a distinctive feature that can be used to localise the source origin if there is some knowledge about this feature. Practically, the locations features are obtained from the received signal in the training stage.

3.1.1. Focusing in Time-reversal Theory

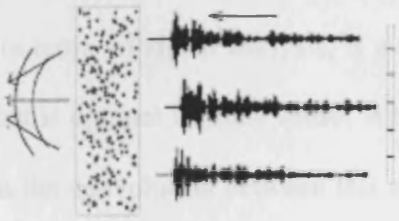
In time-reversal acoustics, a source is applied to a medium at a certain location and a received signal is recorded by an array of transducers as shown in figure 3.1. The received signals are reversed in time and then re-transmitted into the medium. The re-transmitted signal propagates back through the same medium and goes through all the multiple scattering, reflections and refraction that it underwent in the forward direction and refocuses on the source location. If only an aperture of limited area, called the time-reversal mirror, is performed in the time reversal operation, a small part of the field radiated by the source is captured and time reversed, thus limiting focusing quality [57].

In a bounded medium, multiple reflections along its boundaries significantly increase the apparent aperture of the time reversal mirror and effectively the transducers are replaced by reflecting boundaries that redirect part of the incident wave towards the aperture. Thus, spatial information is converted into the time domain and the reversal quality depends crucially on the duration of the time-reversal window, i.e. the length of

the recording to be reversed. The heterogeneity of the medium or the boundaries produces multi-paths and this contributes to having an aperture that is much larger than its physical size. It has been shown experimentally that in a cavity with a specific geometrical property, focusing with time-reversal can be obtained using one transducer [58].



(a)



(b)

Figure 3.1. Sketch of time reversal focusing in random medium. a. impulse transmission and reception. b. time-reversed transmission and impulse localisation.

The time-reversal approach is clearly connected to the inverse source problem. They both deal with propagation of a time-reversed field, but the propagation is real in the time-reversal experiment and simulated in the inverse problem. Moreover, the most important distinction is that time-reversal doesn't need knowledge of the propagating medium while the inverse problem method does.

As for any linear and time-invariant process, wave propagation through a multiple scattering medium may be described as a linear system with a certain impulse response. If the source sends a Dirac pulse $\delta(t)$ function, the j^{th} transducer of the Time-Reversal-Mirror will receive a signal $h_j(t)$, which is the propagation impulse response from the source to transducer j. Moreover, due to spatial reciprocity, $h_j(t)$ is also the impulse response describing the propagation of a pulse from the j^{th} transducer to the source. Thus, if the transducer is able to record and time-reverse the whole impulse response as $h_j(-t)$, the signal generated at the source is given by the convolution $h_j(t) * h_j(-t)$. This convolution product, in terms of signal analysis, is a typical matched filter which is a linear filter whose output is optimal in some sense. Whatever the impulse response $h_j(t)$, the temporal result is the convolution between this response and its time-reverse version $h_j(t) * h_j(-t)$ which is maximal at time $t=0$. This maximum is always positive and equals $\int h_j^2(t)dt$, i.e. the energy conveyed by $h_j(t)$. This has an important consequence. Indeed, with an N-element array, the signal recreated on the source can be written as,

$$s(t) = \sum_{j=1}^{N} h_j(t) * h_j(-t) \tag{3.4}$$

Even if $h_j(t)$ are completely random and apparently uncorrelated signals, each term in this sum reaches its maximum at time $t=0$. So, all contributions add constructively around $t=0$, whereas at earlier or later times uncorrelated contributions tend to interfere destructively one with another. Thus the recreation of a sharp peak after time-reversing on an N-element array can be viewed as an interference process between N outputs of N matched filters.

3.1.2. Realisation of Source Localisation from Time-reversal Focusing

As described in the previous section, it is possible with time-reversal theory to reconstruct an acoustic signal in its original location in a scattering medium by recording the received signals and sending back the time-reversed version of these signals through the medium. This implies that the received signal carries its source location signature as a result of scattering in the transmission medium and reflections from complex boundaries. With the same assumption of Dirac delta source excitation, the response term $h_j(t)$ of the temporal correlation as given by 3.1 can be interpreted in LPM as a template obtained in the learning stage. The $h_j(-t)$ is the applied test signal that with a negative sign turns the convolution into a cross correlation operation. Therefore, cross correlation is a focusing process in time-reversal but a similarity measure in template matching localisation.

3.1.3 Deduced Rules

Although source reconstruction in time-reversal is a transmission process or an active one, it is comparable to passive source localisation with template matching by the operation of cross correlation as illustrated in the previous section. However, time-reversal can help to provide an explanation and physical limitations of the source localisation problem.

Ideally, source reconstruction is achieved with an array of sensors surrounding the source origin with element spacing of at least half a wavelength. Practically, a limited aperture area is used at the cost of focusing resolution. The smaller the array, the larger the focal spot. As a result of wave diffraction, the waves will refocus to a spot not smaller than the shortest wavelength [58]. Accordingly, the achievable localisation resolution using template matching can be increased with more sensors but still limited to the smallest wavelength, and since wavelength is inversely related to frequency, higher accuracy can be obtained with interactions that generate higher frequency signals, such as using nail clicks or a metallic object, than those which generate lower frequencies such as a finger tap or damping material.

3.2. Criterion of Pattern Matching

Localisation by pattern matching encounters two stages, learning and recognition. In the learning stage, signals received from impacts are tagged to their known location zones. In the recognition stage, a received signal is localised by finding the best

matched feature in the template. Pattern matching is performed using one of the similarity measure criterion, maximum likelihood estimation, vector-distance metrics or cross correlation coefficient. The LPM system is depicted in figure 3.2.

Suppose the location feature is represented by the column vector g_i of the template signal samples for the i^{th} location and s is the column test signal that needs to be matched to a template signal. The matching criterion is to find location i that maximises the matching process. One possible matching criterion is the maximum likelihood decision used in [52] to localise the object in a room with active sources. Assuming the vector s consists of zero mean independent Gaussian random variables with standard deviation of σ, the likelihood function can be modelled as:

$$l_{s|\eta_i}(s \mid \eta_i) = e^{-\frac{1}{2\sigma^2}(s-g_i)^T(s-g_i)} \tag{3.5}$$

where η_i is the i^{th} hypothesis. The maximum likelihood decision can be stated by maximising (3.5) and the matching criteria can be simplified to;

$$i = \arg\max_i \{f_{s/g_i}(s \mid g_i)\} = \arg\min_i (|s - g_i|^2) \tag{3.6}$$

It is clear that equation (3.6) is effectively a least square estimator.

The approach used in Dijk PhD work [53] for locating an ultrasonic source is by comparing the measured signal with reference signals obtained by simulation that takes

into account the 3-D geometrical reflections in the room. Vector-Distance metrics are proposed for comparing the two time series of length N with a tuning factor ,q, for the required resolution as given by,

$$L(s,g) = (\sum_{n=1}^{N} |s(n) - g_i(n)|^{q})^{1/q} \tag{3.7}$$

The same concept has been used in [59] to localise an active source underwater using single hydrophone.

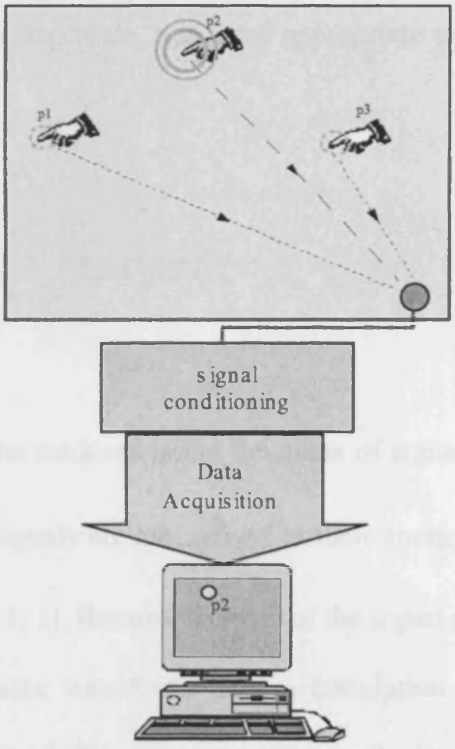


Figure 3.2 LPM system layout.

Another matching approach is cross correlation, a familiar process used in signal processing to compare the similarity between two patterns, where the peak value indicates the degree of linear correlation and the peak argument indicates the time lag between the two signals. Mathematically, the cross correlation between two time signals *s(t)* and *g(t)* is given by;

$$R_{sg}(\tau) = \int_{-\infty}^{\infty} s(t)g(t+\tau)dt \tag{3.8}$$

Because R_{sg} varies as the signal energy varies with time, and the correlation range is dependant on the signal's amplitude, it is more appropriate to use the cross correlation coefficient expressed by;

$$\Gamma_{sg}(\tau) = \frac{R_{sg}(\tau)}{\sqrt{R_{ss}(0)R_{gg}(0)}}, \qquad -1 \le \Gamma_{sg} \le 1 \tag{3.9}$$

where R_{ss} and R_{gg} are the autocorrelation functions of signals s and g respectively at time lag zero. Here, the signals are normalised to their energy and the range of Γ_{sg} is bounded by the interval [-1, 1]. Because the sign of the signal doesn't affect the pattern, for example sine and cosine waveforms have a correlation of -1, both positive and negative peaks are considered. The criteria to estimate the location i of the test signal s from the best match with the template signals g_i can be expressed by,

$$i = \arg\max_i \left| \Gamma_{sg_i}(\tau) \right| \tag{3.10}$$

There are good reasons to choose cross correlation as a preferred matching algorithm. One is that from physics point of view cross correlation can be interpreted as a focusing in time reversal theory. The other reason is that practical tests have shown much better results with cross correlation than the other two methods. Very importantly, the existence of the Fourier transform relationship given by the famous Wiener-Khintchine theorem allows for efficient operations in the frequency domain. The location feature used in the matching process above is the pattern of the time series signals s(t) and g(t), but it is not limited to time series. Any other quantity that carries a location feature can be used instead. The extraction of more specific location features will be considered later. A practical example for identifying the location of a finger tapping on a fibre board is shown in figure 3.3. The result of equation (3.9) is shown in a 3-D map for four individual sensors where the highest correlation peak corresponds with the source location while the matching with other location templates results in lower peaks.

3.3 Empirical Analysis

Since the LPM method conquered the complexity of defining the location signature by recording real signals rather than by simulation or by using an analytical solution, it is more appropriate to analyse the performance of LPM system empirically. The performance of LPM localisation can be characterised by the resolution and the reliability.

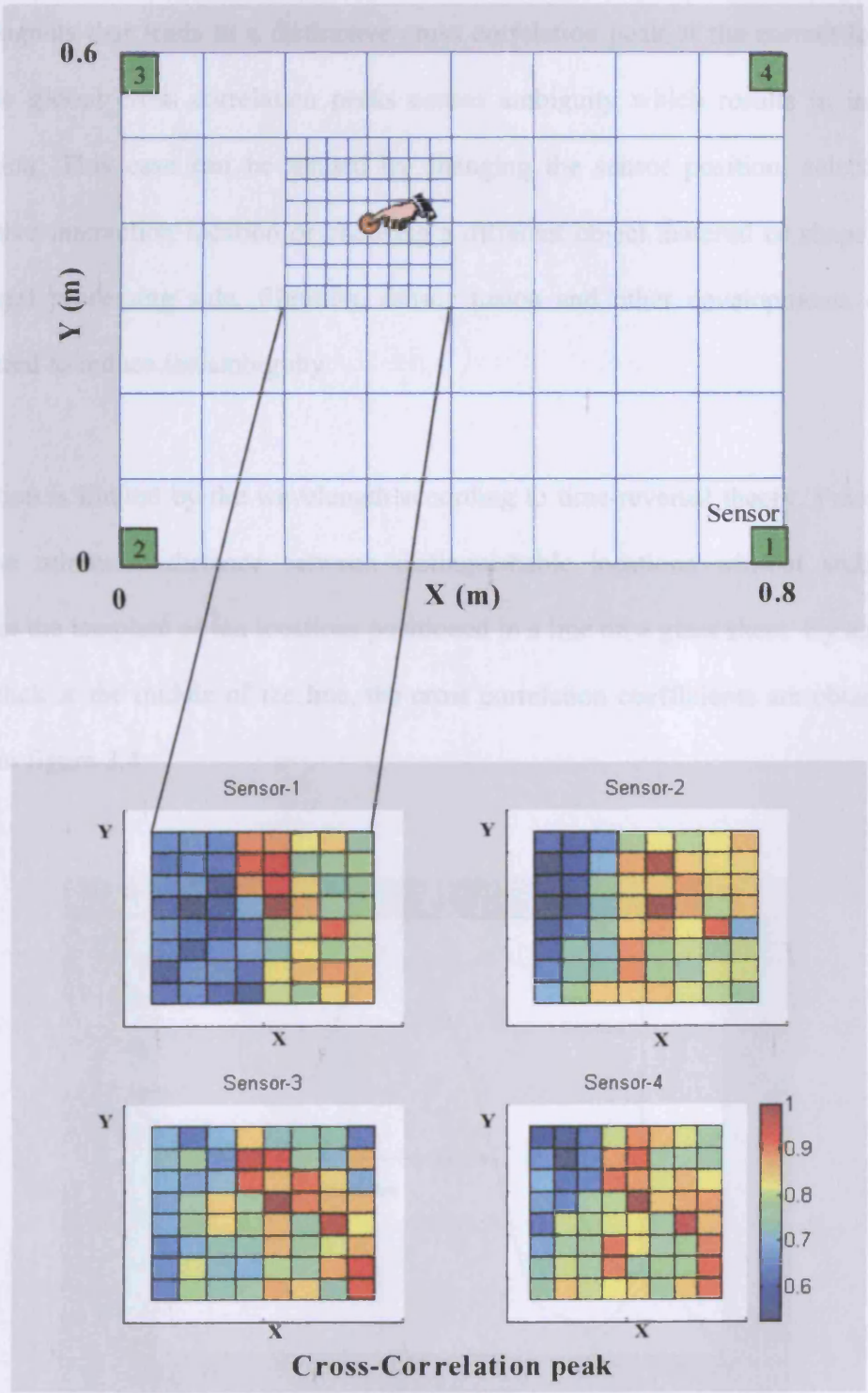


Figure 3.3. Spatial cross correlation peaks of impact detected by four sensors.

Practically, there is no guarantee that for each location there is always a unique feature in the signals that leads to a distinctive cross correlation peak at the correct location. Multiple global cross correlation peaks causes ambiguity which results in incorrect estimation. This case can be treated by changing the sensor position, selecting an alternative interaction location or choosing a different object material or shape. From the signal processing side, filtration, sensor fusion and other developments can be considered to reduce the ambiguity.

Resolution is limited by the wavelength according to time-reversal theory. Practically, it is the minimum distance between distinguishable locations without ambiguity. Consider the template of ten locations positioned in a line on a glass sheet. By applying a nail click at the middle of the line, the cross correlation coefficients are obtained as shown in figure 3.4.

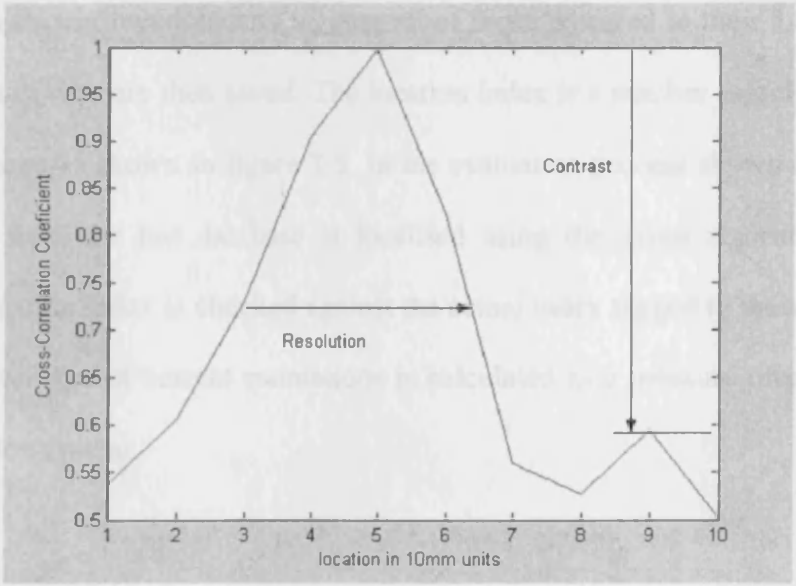


Figure 3.4. Resolution in LPM system.

The resolution can be shown in this figure, as arbitrary defined in [51] by the half power beam width of the spatial cross correlation peaks and the ambiguity level by the contrast.

On the other hand, reliability is indicated by how well the LPM system is correctly responding to different types of interactions, which is restricted by the sensitivity of the matching algorithm to the type of interactions used in the learning stage.

3.3.1 Evaluation Procedure

Obviously for a proper evaluation or comparison of LPM systems, given data collected from measurements has to be used for evaluating different algorithms and parameters since interactions cannot be reproduced exactly the same. For a quantitative evaluation of large data, a simple procedure is used by generating both template signals and test signals from known locations and tagging all of those received to their location index. The two sets of data are then saved. The location index is a number associated with the interaction zone as shown in figure 3.5. In the evaluation process shown in figure 3.6, each signal from the test database is localised using the given algorithm, then the estimated location index is checked against the actual index tagged to the signal. In this way the percentage of correct estimations is calculated as a measure of confidence in the localisation system.

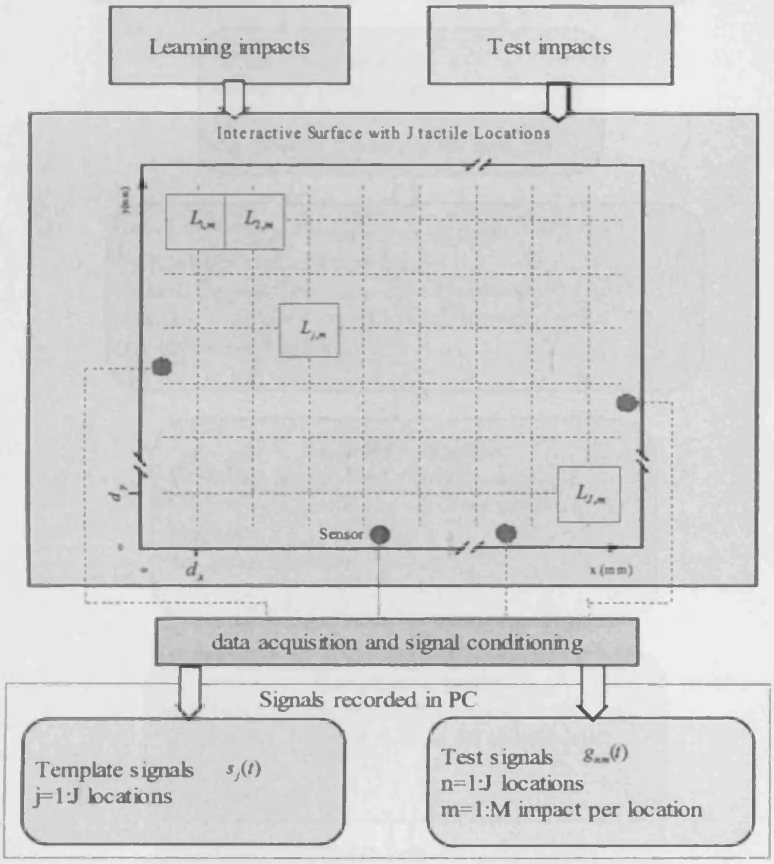


Figure 3.5. Data recording for evaluation.

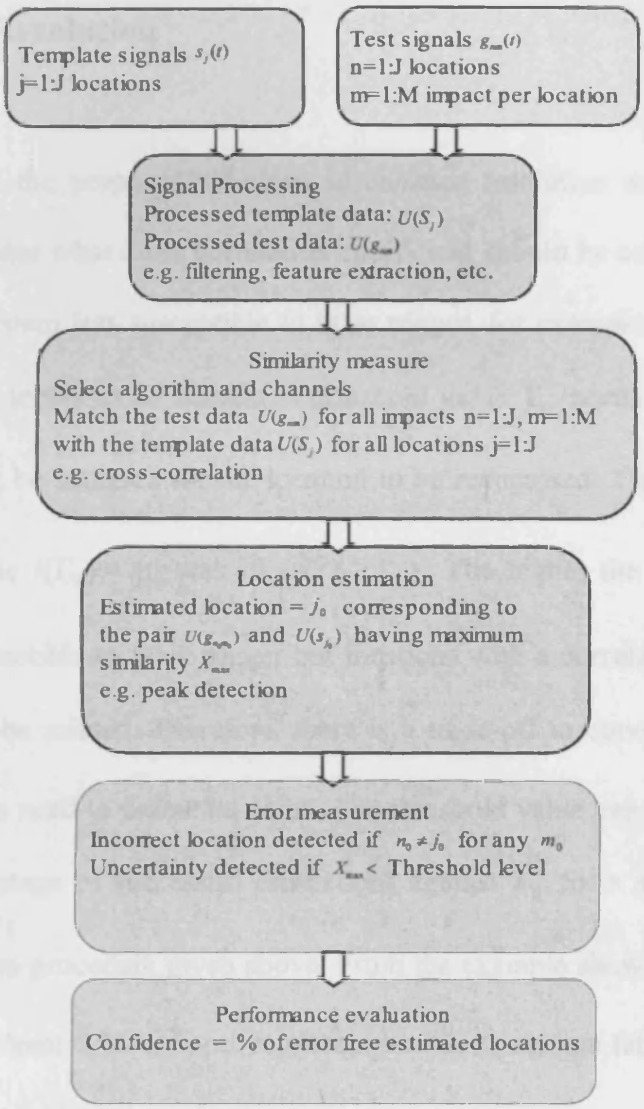


Figure 3.6. Evaluation procedure for the data collected as in figure 3.5.

3.4 Enhanced Resolution

Before looking at the proposed solutions to enhance resolution and reliability, it is important to examine what cross correlation coefficient should be accepted. In order to make the LPM system less susceptible to false trigger, for example to ignore impacts outside the tactile locations of interest, a threshold value Γ_0 needs to be assigned so that $\Gamma_{sg} \geq \Gamma_0$ must be satisfied for the location to be recognised. Thus equation (3.10) can be rewritten as $i(\Gamma_0) = \arg\max_i(|\Gamma_{sg_i}(\tau)| \geq \Gamma_0)$. The higher the threshold the less the system is susceptible to false trigger but locations with a correlation coefficient of less than Γ_0 will be missed. Therefore, there is a trade-off in choosing the threshold level and there is a need to define its value. The threshold value can be investigated by plotting the percentage of successful estimations against Γ_0 for a given test database using the evaluation procedure given above. From the example shown in figure 3.7, Γ_0 should be set to at least 0.95 to improve system immunity against false trigger as far as possible without degrading its performance.

3.4.1 Multi-dimensional signal

It was shown that focusing with time-reversal can be performed in an open system with an aperture of transducers at the cost of resolution. However, in a closed system, one transducer can be sufficient depending on the object geometry. Based on this fact, spatial diversity using multiple sensors can improve the resolution. Two options of

sensor fusion can be considered. One basic method is by calculating the statistical average of the matching peaks detected by K channels as,

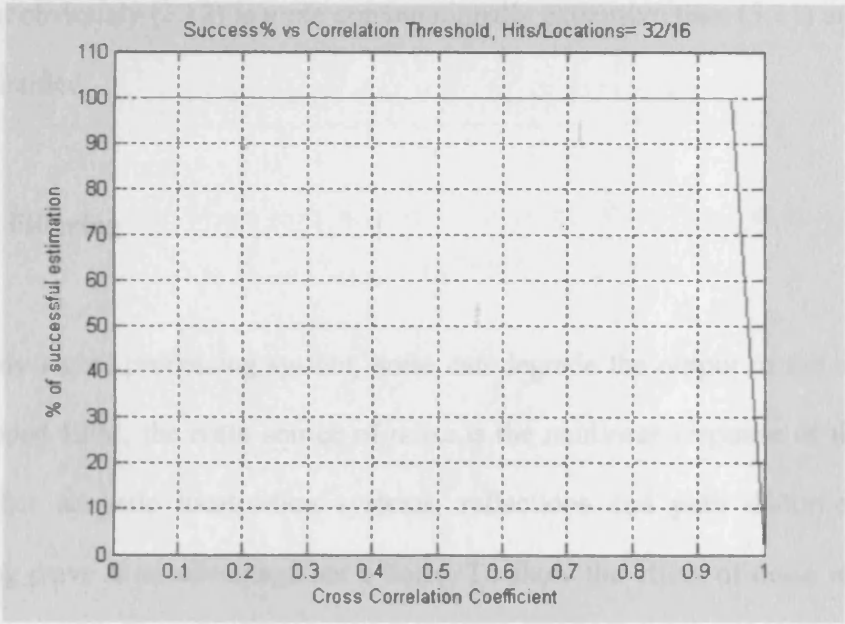


Figure 3.7. Threshold of cross correlation coefficient.

$$i = \arg\max_i \sum_{1}^{K} \max(\Gamma_{sg_i,k}(\tau))/K \tag{3.11}$$

Note that maximum likelihood estimator can also be used as in (3.11) but with minimisation of the matching kernel rather than maximisation. The other option is to adopt the signal reconstruction formula from time-reversal theory given by equation (3.4) to combine the outputs of the cross correlation prior to peak detection as follows,

$$i = \arg\max_i \left| \sum_1^K \Gamma_{sg_i,k}(\tau) \right| \qquad (3.12)$$

The evaluation process shown in figure 3.6 has shown comparable results for both criteria but obviously (3.12) is more computationally expensive than (3.11) and thus its use is unjustified.

3.4.2 Post Filtering

As with any signal processing system, noise can degrade the output of the system. In the developed LPM, the main source of noise is the nonlinear response of the sensor. Unlike other acoustic localisation systems, reflections and path distortion of the propagating wave is an advantage not a noise. To show the effect of noise removal as well as to demonstrate the relationship of wavelength and resolution, the following experiment was conducted. Sensors with nonlinear properties above 10 KHz were used and therefore a low pass filter was required. A template was created from nail clicks at 2x4 grid of eight locations spaced by 100 mm on a fibre board and four channels were used to acquire the signals at a sampling rate of 100 K samples per second. A Chebyshev digital filter of the frequency response shown in figure 3.8 was used with a cutoff frequency of 8 KHz, then modified to 4 KHz. By applying ten impacts at each location, the resulting database was evaluated using raw signals, filtered signals with 8KHz low pass filter and then with a 4KHz low pass filter. The evaluation results shown in table 3.1 indicate the improvement with noise filtering for both filters as well

as the contribution of higher frequencies in the location features. The minor improvement of higher frequencies is due to the major power content within the lower frequency band.

Filtering	Sensor a	Sensor b	Sensor c	Sensor d	Average from 320 test signals
None	91.00	98.00	95.00	93.00	94.25
LPF 8 KHz	100.00	100.00	100.00	100.00	100.00
LPF 4 KHz	99.00	99.00	100.00	98.75	99.19

Table 3.1. The percentage of correct estimations from four channels.

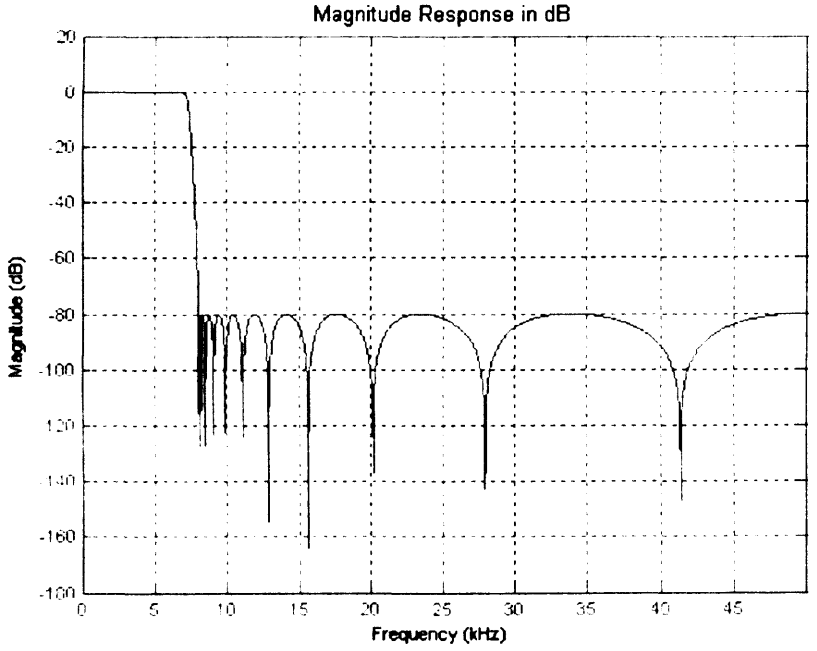


Figure 3.8 Low pass filter

3.4.3 Coherence Filtering

As seen with the previous filtering example, the removal of some frequency components from row signals improved the results of cross correlation, but that was only for known noise sources and unnecessary filtration of higher frequencies will degrade the resolution. Other frequency components which can negatively affect the cross correlation may still exist in the signals. Therefore it is desirable to select only the frequency components which are correlated. This can be achieved using the coherence function.

Coherence is one of the techniques used in acoustics for signal analysis [60]. The coherence function quantifies the linear relationship between two signals in the frequency domain at each frequency ω. The magnitude squared coherence between signals $s(t)$ and $g(t)$ is given by,

$$\gamma_{sg}^2(\omega) = \frac{\left|P_{sg}(\omega)\right|^2}{P_{ss}(\omega)P_{gg}(\omega)} \tag{3.13}$$

where $P_{sg}(\omega)$ is the cross spectral density and $P_{ss}(\omega)$ and $P_{gg}(\omega)$ are the auto spectral density functions of $s(t)$ and $g(t)$ respectively. Equation (3.13) produces a real number between 0 and 1 that represents the degree of matching at frequency ω. Rather than detecting the peak of equation (3.13), the mean value of γ^2 denoted by $\bar{\gamma}$ is computed and used as a matching criterion given by,

$$\bar{\gamma} = \frac{1}{\Delta_\omega} \int_{\Delta_\omega} \gamma_{sg}^2(\omega) d\omega \qquad (3.14)$$

where Δ_ω is the range of frequencies for which the power spectrum is above a threshold level. To eliminate the coherence produced as a result of division by nearly zero quantity, the magnitude squared coherence is multiplied by the cross spectral magnitude.

	Sensor a	**Sensor b**	**Sensor c**	**Sensor d**	**Average**
Algorithm	correct ----- wrong	correct ----- wrong	correct ----- wrong	correct ----- wrong	% of correct
Cross Correlation	824 ----- 256	710 ----- 370	810 ----- 270	803 ----- 277	72.8
Coherence	936 ----- 144	895 ----- 185	939 ----- 141	926 ----- 154	85.6

(a)

	sensor a	**sensor b**	**sensor c**	**sensor d**	**Average**
Algorithm	correct ----- wrong	correct ----- wrong	correct ----- wrong	correct ----- wrong	% of correct
Cross Correlation	891 ----- 189	823 ----- 257	879 ----- 201	803 ----- 277	78.6
Coherence	1036 ----- 44	978 ----- 102	1031 ----- 49	1025 ----- 55	94.2

(b)

Table 3.2. Results of localising 1080 impacts at 12x9 locations at 20mm resolution for four channels using (a) finger tap and (b) nail click impacts on a glass sheet.

A typical example of using the coherence function is shown in Figure 3.9.

This algorithm has been compared with the cross correlation algorithm using the evaluation procedure in section 3.3.1 for nail clicks and finger tapping on a glass sheet with four low-frequency piezoelectric sounders positioned arbitrarily near the edges. The percentage of successful localisations is shown in table 3.2, where a 14.2% improvement on average has been achieved for coherence compared to cross correlation with single sensor.

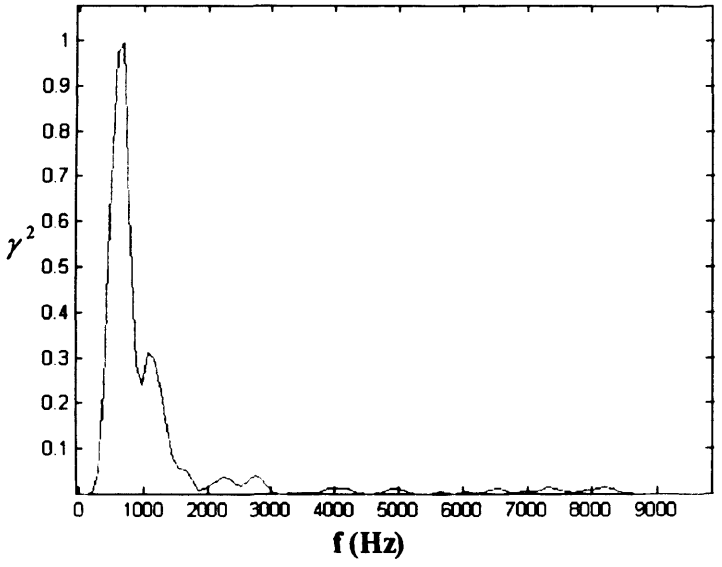


Figure 3.9. Magnitude squared coherence versus frequency of two nail click signals at the same location on a fibre board.

3.5 Location Feature Extraction for Enhanced Reliability

Ideally, to comply with the time-reversal theory, an impulsive source is needed for both learning and recognition stages. Practically, it is found that cross correlation is sensitive to the template type, which means that similar sorts of impacts should be used in both stages, which is expected since the variation in the signal will appear as a variation in the location signature. One option to make the system more reliable is to use multiple templates for different types of interactions, but this is impractical as the calibration work will be intensified. Signal filtering improves the resolution but is found to be not effective in improving reliability because it filters the frequency components and not the location features. Therefore, an attractive novel solution is proposed here to solve the reliability problem by extracting specific features from the signal that is associated with the source location rather than the source information.

Let an unknown source signal given by $s(t)$ emitted from location i on the surface of a tangible object as shown in figure 3.10, and two sensors are receiving the signals $g_1(t)$ and $g_2(t)$. The propagation path from the given source to sensor-1 and sensor-2 can be expressed by specific transfer functions denoted by $h_1(t)$ and $h_2(t)$ respectively. The transfer function is characterised by the complex propagation path and is independent of the source signal information. Accordingly, the transfer function for a specific source to receiver path represents the actual location signature. Treating the transmission medium as a black box of a single input/multiple output system, the output signal received by the i^{th} sensor, as any time invariant system, can be expressed analytically by the convolution integral given by,

$$g_i(t) = \int_{-\infty}^{\infty} h_i(t)s(t-\tau)d\tau \qquad (3.15)$$

For instance, consider the output from one sensor only. It is possible to measure the transfer function for a given location by applying an impulse $\delta(t)$ at that location and measuring the output. In that case, the received signal is the transfer function, as one can tell from equation (3.15) which in turn can be used as location signature in the template. If the test signal is also an impulse, then the matching process is comparing location signatures and therefore high accuracy is anticipated. Otherwise, if the signal used for test, or for the template, is not an impulse, the resulting received signal will include source signal information plus location information, which accordingly results in estimation error. This is why LPM is very sensitive to template signals and works better with impulsive types of impact. The task now is to develop a technique to extract the source information from any type of interaction by employing two sensors.

Let the input in the system shown in figure 3.11 is a stationary random signal. The input/output relationship in the frequency domain is given by a Fourier transform of equation (3.16) as,

$$G_i(f) = S(f)H_i(f) \qquad (3.16)$$

The hypothesis of extracting the location signature involves utilising a measurable quantity that doesn't require any knowledge about the input excitation or medium

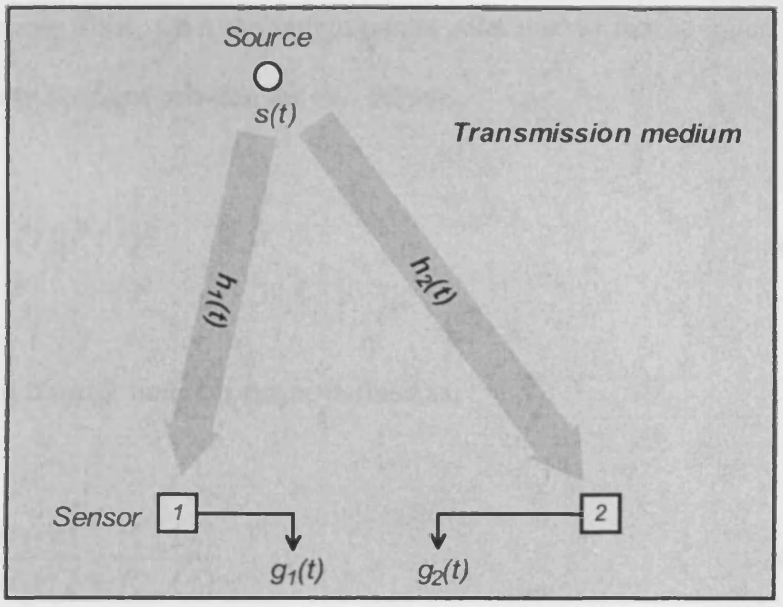


Figure 3.10. Received signals from two different paths.

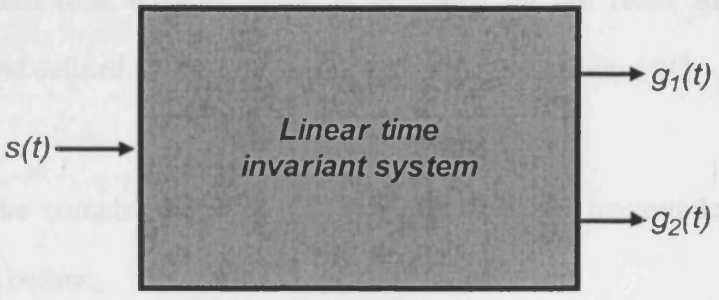


Figure3.11. Black box LPM model.

transfer function. Thus, from the output/output relationship that is given by the cross spectral density function between the two outputs,

$$P_{g_1g_2}(f) = G_1(f)G_2^*(f) \tag{3.17}$$

a hypothetical transfer function can be defined as,

$$\tilde{H}_{g_1g_2}(f) = \frac{H_2(f)}{H_1(f)} = \frac{P_{g_1g_2}(f)}{P_{g_1g_1}(f)} \tag{3.18}$$

where $P_{g_1g_1}$ is the autocorrelation function of $g_1(t)$. It can be seen from the above three equations that $\tilde{H}_{g_1g_2}$ is a function of the two transfer functions $h_1(t)$ and $h_2(t)$ which still represents an independent location signature. A related subject in literature is the binaural localisation in humans which is simulated by the Head Related Transfer Function cue and defined by the ratio of the two output spectrums [61].

By rewriting the complex equation (3.18) in the form of magnitude and phase as equation (3.19) bellow,

$$\tilde{H}_{g_1g_2}(f) = \mathrm{V}_{g_1g_2}(f)\angle\varphi_{g_1g_2} \tag{3.19}$$

either the magnitude or phase patterns can be used as location signature information. To use both amplitude and phase information, the pattern given by,

$$\tilde{h}_{g_1 g_2}(t) = \int_{-\infty}^{\infty} \psi(f) e^{-j2\pi ft} P_{g_1 g_2}(f) / P_{g_1 g_1}(f) df \qquad (3.20)$$

can be used as signature, where $\psi(f)$ is a filter introduced to adjust the bandwidth variation which differs depending on the interaction with different materials. With $\psi(f) = V_{g_1 g_2}(f)$, only phase information is extracted. Obviously, utilising only the phase or the magnitude pattern is computationally faster than using (3.20) since it is not necessary to convert them back into the time domain.

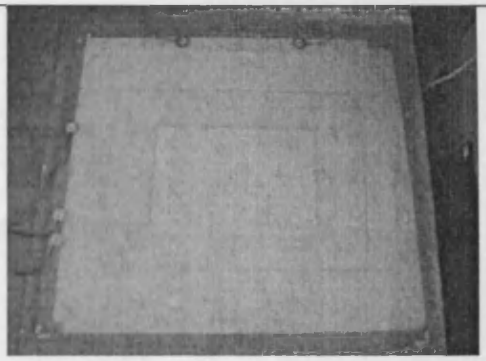
An experimental result was carried out by registering a template from impulsive impacts at defined locations generated by pen tip hits on a glass sheet. The test database consists of different interaction types such as pen tip hits, nail clicks and finger tapping. Then, with the evaluation procedure, it is found that the highest percentage of correct estimations are obtained using (3.20), then using the phase information only and lastly when only magnitude information is used.

3.6 Experimentation

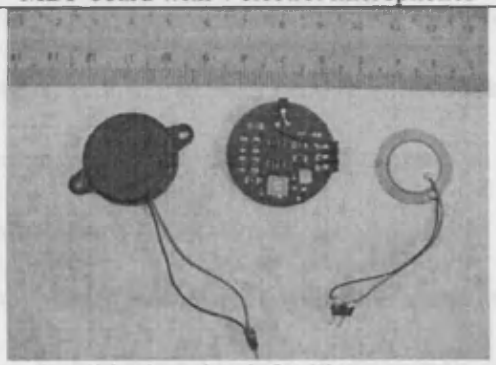
The experimental setup consists of the interactive object, sensors, signal amplifier, data acquisition card and a PC to process the signals. Different object materials have been tested including metal, glass, plastic and fibre boards. The suitable sensors were the piezoelectric discs, electret microphones and accelerometers. For data acquisition, a four channel PCI card is used for evaluation and a two channel sound card used for demonstrations, such as the portable USB sound card and the wireless audio

transmitter. These equipments are pictured in figure 3.12. The LPM system is found to be working on a variety of materials. The piezo-ceramic sounders and electret microphones are the cheapest but can only pick up low frequencies when firmly attached to the surface. The piezoelectric microphone is very sensitive with wide bandwidth response but the most expensive. The piezoelectric shock sensor from Murata is the best sensor with a sufficient frequency response and a very reasonable price. A drawback of the LPM system is that it may require re-calibration of all the interaction points whenever a physical change is imposed on the tangible object.

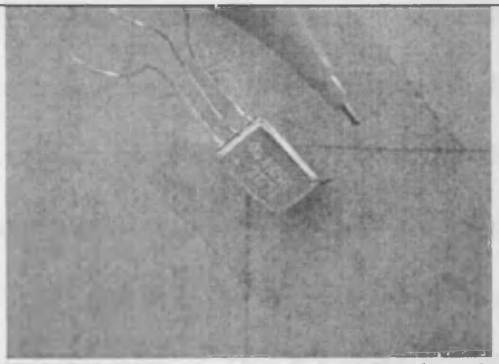
MDF board with 4 electret microphones



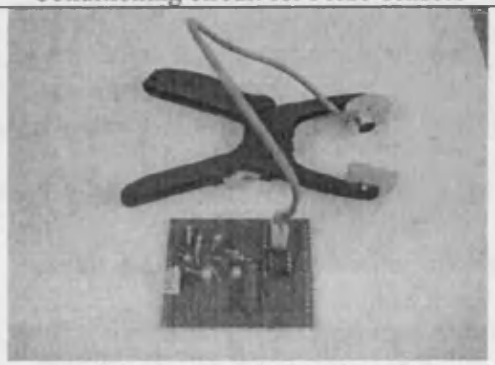
Glass with four Piezo-ceramic sounders



Conditioning circuit for Piezo-sensors



Piezoelectric microphone on MDF board



Conditioning circuit for microphone

FM Audio transmitter/receiver

Plastic tray with shock sensors and USB sound card

Metal tray with shock sensors and USB sound card

Figure 3.12. Various experimental equipment for LPM localisation.

Chapter 4

An Investigation into Localisation Approaches for TAI

The LPM approach proposed in the previous chapter utilises the uniqueness of the transmission channel property between source and sensors. It benefits from the propagation complexity which results from object geometry, boundaries and the structure of material and thus is suitable for impact localisation for any object material and shape. However, the disadvantage of the LPM method is the learning requirement for each individual point of interaction on the tangible surface. This is a problem when it is preferred to interact with arbitrary locations or to work on large surface with a large number of interaction points or if it is not desired or not possible to carry out the learning stage. Another problem with the LPM approach is its unsuitability for tracking a moving source. The promising approach that doesn't require learning, with good potential for tracking a continuous moving source, is the one based on measuring the time differences of signal arrivals to multiple spatially separated sensors which has been a hot area in the development of passive source localisation for modern applications.

Since various algorithms have been developed for the passive source localisation problem in the literature, as seen in section 2.3, it is important now to identify the fundamental features of these algorithms and to match it with the essential

requirements of TAI. These approaches are based on a known wave propagation velocity v, the measured time difference τ and sensor geometry with no learning required. In the research by MIT for developing a large tangible window, only heuristic methods were considered, with little attention paid to other methods. The purpose of this chapter is to investigate the main theoretical localisation methods and identify the problem of source localisation in the TAI model, which in turn leads to a narrowing down of the possible options and justifies the methods used in the development of the algorithms in the following chapters.

4.1 Active and Passive Sources

In an active localisation system, a deterministic signal is transmitted from the source to work as an embedded time stamp. The Time of Arrival (TOA) from source to sensor can be known from the signals acquired by each sensor, using synchronisation with the emitter clock. Since each TOA forms a circle of possible source location centred at the relevant sensor, the source location ambiguity can be resolved from the intersection of three circles using a minimum of three sensors as shown in figure 4.1. Such an active method is common in mobile phones and in ultrasound applications [62].

In a passive localisation system, the time of arrival is unknown. But the Time Difference of Arrival (TDOA) between a couple of sensors can be measured and used as the fundamental unit to calculate the source location. In comparison with the TOA circles in an active system, a given TDOA value between two sensors forms a hyperbolic curve for the possible source location as illustrated in figure 4.2. Therefore,

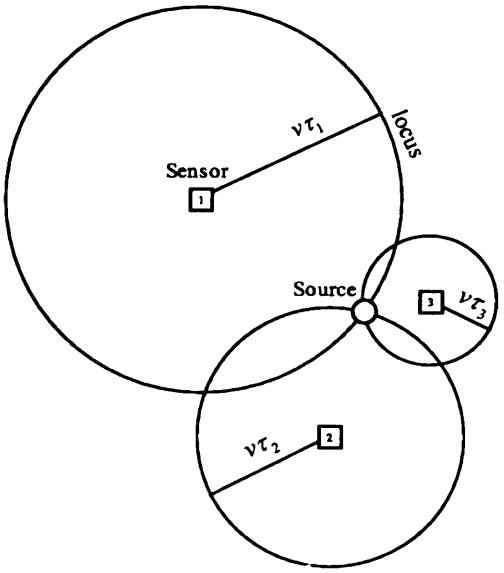


Figure 4.1. Active source localisation from time TOA circles' intersection

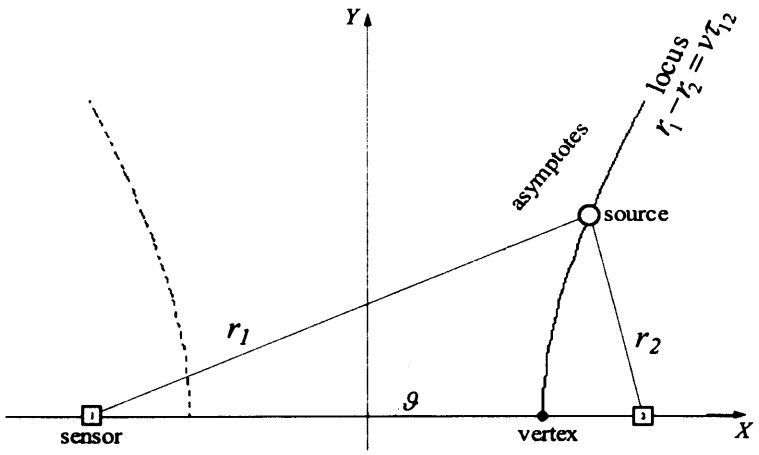


Figure 4.2. TDOA hyperbola for passive system

an additional sensor is required to find the location from the intersection of two hyperbolas.

4.2 Bearing Estimation

The best known source localisation approach for in-air and underwater application is the bearing estimation or Angle of Arrival (AOA) using a linear phased array. With reference to figure 4.3, assume a plane wave is incident on an array of two sensors spaced a distance d apart. From the geometry, it can be shown that the angle of arrival ϑ can be found from,

$$\cos(\vartheta) = \frac{v\tau_{12}}{d} \tag{4.1}$$

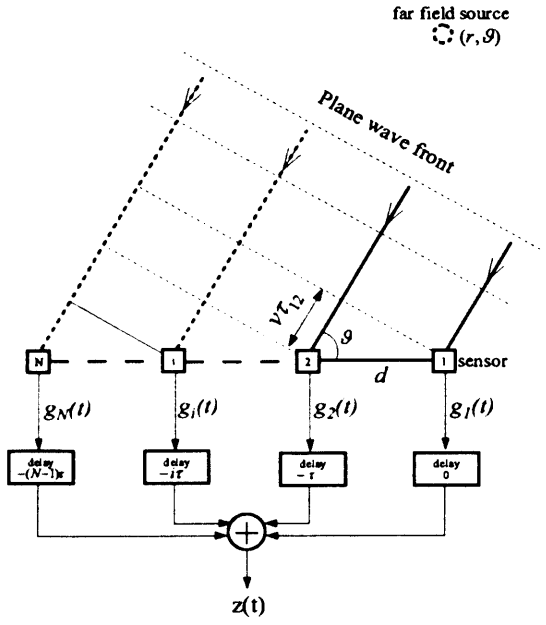


Figure 4.3. Far-field plane wave front and beamforming

The condition for assuming a plane wave that leads to equation (4.1) is that the source is located in the far-field zone specified by $r >> d$ [63]. Another condition to resolve ambiguity resulting from spatial aliasing is that the maximum distance between sensors must satisfy $d \leq \frac{\lambda}{2}$, where λ is the wavelength [64]. The direction of arrival of a far-field source can also be approximated by the asymptotic line of a hyperbola curve as depicted in figure 4.2.

The source coordinates can be obtained in the far-field from the triangulations of intersecting two angles of arrival using two arrays as illustrated in figure 4.4 as given by,

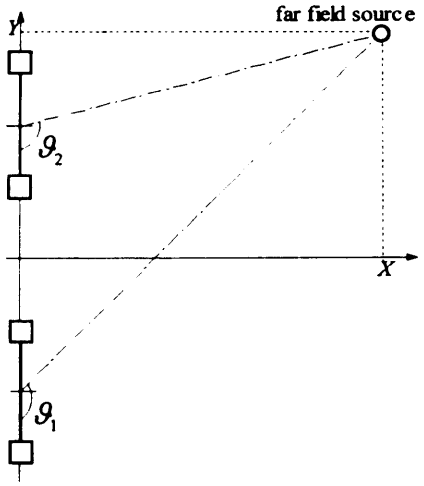


Figure 4.4. Far-field source location (x,y) from two AOA's

$$x = d\frac{\sin(\vartheta_1)\sin(\vartheta_2)}{\sin(\vartheta_1 - \vartheta_2)}$$
$$y = -d\frac{\sin(\vartheta_1)\cos(\vartheta_2)}{\sin(\vartheta_1 - \vartheta_2)} \qquad (4.2)$$

This approach has been proposed for sound localisation in service robots [35].

Due to practical and physical limitations including signal strength, accessibility, wavelength variation and the need to distribute the sensors apart for better performance, the working area of a reasonable size object, i.e. reachable by hands such as a whiteboard and a shop window, is in the order of the distance between sensors. In this case, the source is located in the near zone where the wave front is spherical and the time difference is a function of the radial difference, as illustrated in figure 4.5. Thus the far-field formula is not a proper approximation and near-field localisation algorithms should be considered. The source location in the near-field can be calculated from the hyperbolic intersection geometry.

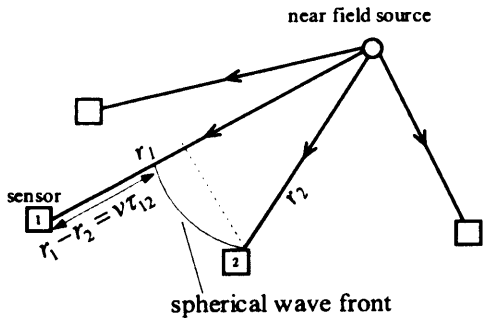


Figure 4.5. Near-field TDOA assuming spherical wave front.

4.3 Hyperbolic Localisation Geometry

The hyperbola is defined by the set of points that are an equal difference in distance from two focal points. Consider the scenario of locating a source *s*, produced by an impact on a solid object, using the acoustic signal picked up by sensors 1, 2 and 3 as depicted in figure 4.6. The focal points here are the sensors. For a given TDOA measurement between sensors 1 and 2, the source *s* can be located anywhere along the red hyperbola curve. With additional measurement from sensors 1 and 3, for instance, the source location (x,y) can be resolved from the intersection of the two hyperbolas, the red and the green curves. Mathematically, let sensors 1, 2 and 3 be at (x_1,y_1), (x_2,y_2) and (x_3,y_3) respectively. The emitted energy from source *s* arrives at sensors 1 and 2 with a delay difference of τ_{12} and at sensors 1 and 3 with a delay difference of τ_{13}. Thus two hyperbolas can be formed,

$$\sqrt{(x-x_1)^2+(y-y_1)^2}-\sqrt{(x-x_2)^2+(y-y_2)^2}=v\tau_{12}$$

$$\sqrt{(x-x_1)^2+(y-y_1)^2}-\sqrt{(x-x_3)^2+(y-y_3)^2}=v\tau_{13} \qquad (4.3)$$

These are two equations with two unknowns, x and y. However, because of the root square, solving fourth-order equations in x and y is not promising and therefore an alternative solution is sought.

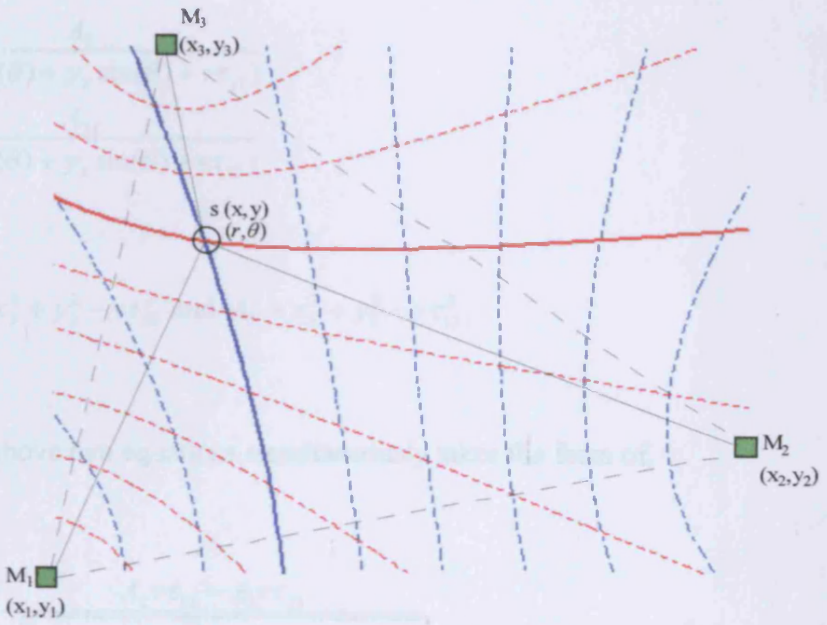


Figure 4.6. Hyperbolic intersection geometry

4.3.1 Analytic Solution

The analytic solution for the hyperbolic intersection problem can be simplified using polar coordinates as proposed in [65] for detecting acoustic emission in non destructive testing. Assuming the source is located at (r, ϑ) from the reference sensor, g_1, in the origin (0,0) and applying the cosine rule for triangles M_1,M_2,s and M_1,M_3,s yield after algebraic simplification,

$$r = \frac{A_1}{2(x_2\cos(\theta) + y_2\sin(\theta) + v\tau_{12})}$$
$$r = \frac{A_2}{2(x_3\cos(\theta) + y_3\sin(\theta) + v\tau_{13})} \qquad (4.4)$$

where $A_1 = x_2^2 + y_2^2 - v\tau_{12}^2$ and $A_2 = x_3^2 + y_3^2 - v\tau_{13}^2$

Solving the above two equations simultaneously takes the form of,

$$\cos(\vartheta - \phi) = \frac{A_2 v\tau_{12} - A_1 v\tau_{13}}{\sqrt{(A_1 x_3 - A_2 x_2)^2 - (A_1 y_3 - A_2 y_2)^2}} \qquad (4.5)$$

where $\tan(\phi) = \frac{A_1 y_3 - A_2 y_2}{A_1 x_3 - A_2 x_2}$.

Now ϑ can be obtained from equation (4.5) and substituted in equation (4.4) for r.

Other compact solutions have been developed for special cases of array geometries. In passive sonar acoustics [66], a compact solution is developed using a linear array of three sensors positioned as ($x_1=0, y_1=0$), ($x_2=-L_1, y_2=0$) and ($x_3= L_2, y_3=0$). The source location (x,y) is found using the time differences of arrival between each pair of sensors ($\tau_{12}=r_{12}/v$, $\tau_{13}=r_{13}/v$ and $\tau_{23}=r_{23}/v$) in a closed form by simultaneously solving two cosine law equations of the two adjacent triangles formed by four points defined by the three sensors and the source as,

$$x = \frac{r_{12}L_2^2 - r_{13}L_1^2 - r_{12}r_{13}r_{23}}{2(r_{12}L_2 + r_{13}L_1)} \tag{4.6}$$

and y is obtained from $\sqrt{r^2 - x^2}$ where,

$$r = \frac{L_1[1-(\frac{r_{12}}{L_1})^2] + L_2[1-(\frac{r_{13}}{L_2})^2]}{2(\frac{r_{13}}{L_2} + \frac{r_{12}}{L1})}$$

Similarly, another compact solution has been formulated for the special case of having three sensors in a right angle geometry [67] for the development of a golf simulator. Although these analytic solutions are attained in a closed form, they are very sensitive to errors in time delay values and they do not use redundant information from additional sensors to improve accuracy. For these reasons, alternative algorithms need to be considered.

4.3.2 Iterative solution

A numeric solution to the hyperbolic intersection equations that can handle the error in TDOA is obtained by defining the error in equation (4.3) as,

$$\begin{aligned} e_{12} &= \sqrt{(x-x_1)^2 + (y-y_1)^2} - \sqrt{(x-x_2)^2 + (y-y_2)^2} - v\tau_{12} \\ e_{13} &= \sqrt{(x-x_1)^2 + (y-y_1)^2} - \sqrt{(x-x_3)^2 + (y-y_3)^2} - v\tau_{13} \end{aligned} \tag{4.7}$$

Then, x and y that satisfies both hyperbolas with least mean error are found by minimising the term,

$$e(x,y) = e_{12}^2(x,y) + e_{13}^2(x,y) \tag{4.8}$$

This numeric algorithm has been proposed for the tangible acoustic interface application in [16].

4.4 Time difference estimation

In the two stages localisation approach, time difference of arrival is the key element used in the positioning algorithm. One basic method with limited capability for estimating the time delay is based on the time when the signal amplitude passes a threshold value within a certain region of the signal. Other sophisticated methods are based on the cross correlation operation. In [15], the first maximum in the signal is used as an index to measure the time difference between two sensors. The x and y coordinates are then found from a polynomial function of time delays with coefficients that have been previously determined from calibration impacts at known locations. An improved routine then followed by [16] based on raising edge detection. This routine spots where the signal first surpassed a quarter of the peak of the signal, then it backtracked specific steps before proceeding forward to find the first spot that rose above some lower predefined threshold that should be above noise level. Another heuristic method in [16] is the peak of the cross correlation performed on a selected part of the signals. Because the used direct cross correlation is not reliable enough, the

final time delay is taken from the average value of these two methods, then substituted in equation (4.7) for the location coordinates using an iterative solution.

The approaches in [15, 16] rely on the signal magnitude regardless of the whole signal shape and phase. It is, therefore, sensitive to factors that affect the signal magnitude such as the object's damping property, homogeneity, noise and is severely affected by dispersion. Also, this method is only applicable to individual impacts but is not suitable for continuously tracking a moving source. The theoretical development in most of the algorithms for estimating the time difference in the applications of sonar, radar, speech and acoustic signal processing is based on cross correlation as described below.

4.4.1 Cross Correlation based TDOA

If the source produces signal $s(t)$, then the received signals $g_i(t)$ and $g_j(t)$ acquired by spatially separated sensors M_i and M_j can be modelled by,

$$\begin{aligned} g_i(t) &= h_i(t) * s(t) + n_i(t) \\ g_j(t) &= h_j(t) * s(t-\tau) + n_j(t) \end{aligned} \tag{4.9}$$

where τ represents the time difference of arrival to be determined, * signifies the convolution operation, $h_i(t)$ is the channel impulse response between the source and the i^{th} sensor and $n_i(t)$ is an additive white Gaussian noise assumed uncorrelated with $s(t)$ and $n_j(t)$. In the ideal propagation, $h(t)$ is the Dirac delta function, therefore

$g_i(t) = s(t) + n_i(t)$ and $g_j(t) = s(t-\tau) + n_j(t)$. For such an assumption, τ is the time lag corresponding to the global maximum of the cross correlation function given by,

$$R_{ij}(\tau) = \int_{-\infty}^{\infty} g_i(t) g_j(t+\tau) dt \qquad (4.10)$$

The noise term *n(t)* is eliminated by the cross correlation process since it is assumed uncorrelated. If the given assumptions are not valid, the peak detection of (4.10) will encounter error. However, this can be compensated for by introducing a weighting function in the frequency domain in a process called generalised cross correlation, a closer approximation to the real environment.

4.4.1.1 Generalised Cross Correlation

The most popular cross correlation method for time delay estimation is the Generalised Cross Correlation (GCC) [68]. The advantage of the GCC is that it encompasses a weighting function in the frequency domain to improve the TDOA estimation accuracy in a real reverberant environment to some extent.

From the well known Wiener–Khinchin theorem, the cross spectral density of a wide-sense-stationary random process is related to the cross correlation function by the Fourier transform relationship. The GCC technique introduces a weighting function $\Psi(f)$ as a filtering process within the cross correlation operation as given below,

$$\hat{R}_{ij}(\tau) = \int_{-\infty}^{\infty} \Psi(f) G_i G_j^* e^{j2\pi f\tau} df \qquad (4.11)$$

where G_i is the Fourier transform of $g_i(t)$ in equation (4.9). The estimated TDOA can then be found from

$$\hat{\tau} = \arg_{\tau} \max \hat{R}_{ij}(\tau) \qquad (4.12)$$

The solution of (4.12) is simply obtained by numerical search. Mathematically, detecting the maximum of the cross correlation is equivalent to detecting the zero crossing of its derivative when the second derivative is negative. This solution has several advantages over the numerical search, particularly for hardware implementation, since shift register can be replaced with logic counter and XOR gates [69].

4.4.1.2 Criteria of the GCC Filtering Process

The choice of the filter $\Psi(f)$ is important in practice. If $\Psi(f)=1$, no weighting is introduced to compensate for the effect of propagation in a real environment and thus the classical cross correlation is obtained which is equivalent to equation (4.10). Because of noise and reverberation in a real environment, some criteria for the weighting functions have been developed particularly for in-air acoustics. Among the common criteria are the Phase Transform (PHAT), Smooth Coherence Transform (SCOT) and Maximum Likelihood Filter (MLF). The classical PHAT filter is given by

$$\Psi_{PHAT} = \frac{1}{|X_1(f)||X_2(f)|} \tag{4.13}$$

This PHAT processor can perform well in a moderately reverberant room. It has been used extensively in the literature for the localisation of a speaker in a room [70] and in robotics applications [71]. If the noise spectrum N is known, then MLF expressed by

$$\Psi_{MLF} = \frac{|G_i(f)||G_j(f)|}{\frac{|N_i(f)|^2}{|G_i(f)|^2} + \frac{|N_j(f)|^2}{|G_j(f)|^2}} \tag{4.14}$$

can reduce the effect of noise but does not perform well with reverberation. The SCOT filter is given by

$$\Psi_{SCOT} = \frac{1}{\sqrt{G_i(f)G_i^*(f)}\sqrt{G_j(f)G_j^*(f)}} \tag{4.15}$$

which effectively leads to the coherence function when substituted in equation (4.11).

These GCC filtering processes, in particular PHAT, are the most popular approach for in-air applications for the treatment of the dominant problems of reverberation and noise. This is also required for TAI application although these problems can be physically reduced. The reflection effects, even if less in damping material, can be decreased by placing the sensors away from the edges or by fitting an absorbing

material on the edges. The ambient noise can be significantly reduced by affixing sensors firmly to the object with directional isolation plus a good choice of electronics to transfer the signal at low impedance. Other problems arise with wave propagation in solids which are minor in the air. The velocity of sound waves in the air is well known and precisely modelled as a function of temperature with no substantial dispersion phenomenon. Also, air can be assumed uniform where the inverse square law applies and therefore energy based localisation is an option [72] and a cross correlation based approach with some filtering can perform well. While wave propagation in solids is far more complicated, it has different modes and may experience dispersion and amplitude distortion. The GCC filtering technique is, therefore, one option for the development of TAI application but further improvement is required by alternative or supportive techniques.

4.4.2 Beamforming

Beamforming, or spatial filtering, is another approach to source direction estimation based on bearing estimation, where the reception pattern of an array of sensors is steered virtually to the direction where the signal energy is maximised. This is a primary approach in radar and sonar applications.

In the uniform array shown in figure 4.3, the signal $g_i(t)$ received by the i^{th} sensor is a time delayed version of the signal from sensor i-1 by τ given in equation (4.1). The beamformer delays and sums the received signals and the energy of the output signal during the time interval [-T T] can be found from

$$E_{BF}(\tau) = \int_{-T}^{T}\left[\sum_{i=1}^{N} g_i(t+(i-1)\tau)\right]^2 dt \qquad (4.16)$$

Accordingly, the time delay τ or the corresponding direction of arrival from equation (4.1) is the value that maximises $E_{BF}(\tau)$. Various improvement techniques can be applied by performing the beamforming process in the frequency domain, such as signal weighting and filtering, which has been an area of research [73]. This approach has the potential to detect multiple targets by searching for multiple peaks in equation (4.16). However, side loops and local maxima are issues that have received high attention in research. Beamforming is a known technique in radar but has also received attention in modern applications such as source localisation to enhance speech recognition as in [74].

4.5 Signal Analysis

The Fourier transform is a very powerful tool for analysing the frequency content *G(f)* of the entire time signal *g(t)*. If it is desired to provide information about the time intervals when certain frequencies occur, such as searching for voiced intervals in speech, then the frequency content in a finite time window of length *b* can be found from a logical extension of the Fourier transform known as the Short Time Fourier Transform (STFT) given by

$$STFT_g(b,f) = \int_{-\infty}^{\infty} g(t)h(t-b)e^{-j2\pi ft}dt \qquad (4.17)$$

For a time limited signal, the time window width Δ_t is related to the frequency window width Δ_f by the uncertainty principle [75] given by the condition of

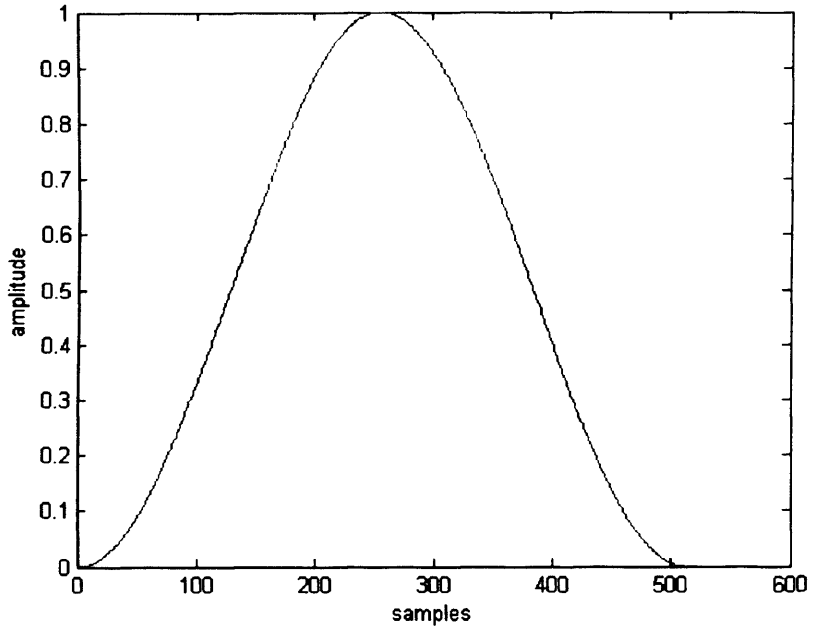
$$\Delta_t \Delta_f \geq 0.5 \qquad (4.18)$$

Relation (4.18) shows that the size of a time-frequency frame cannot be made arbitrarily small and that a perfect time-frequency resolution cannot be achieved, i.e., the higher the frequency resolution, the lower the time resolution and vice versa. For example, if the time frame is chosen as 0.1 s, then $\Delta_f \geq 5$ Hz, which is high resolution and more than required. However, a duration of 0.1 s at a sampling rate of 100 k sample/s results in processing large amounts of data, which may not be useful, for example in the case of an impact, the signal duration lasts for about 20 ms only. For the case of tracking a continuous moving source, the details of the location information can be lost within 0.1 second. A compromised time frame would be 20 ms, giving $\Delta_f \geq 25$ Hz frequency resolution, which is less than the 50 Hz noise considered appropriate. Various forms can be used for the windowing function *h* to reduce spectral leakage, such as the Hanning window shown in figure 4.7(a). Upon choosing the window function, the time-frequency resolution is fixed over the entire analysis plan. If the signal is highly dynamic, i.e. more to non-stationary, then a Wavelet transform can provide more details than STFT. The continuous Wavelet transform (WT) of a signal *g(t)* at scale *a* is given by [76]

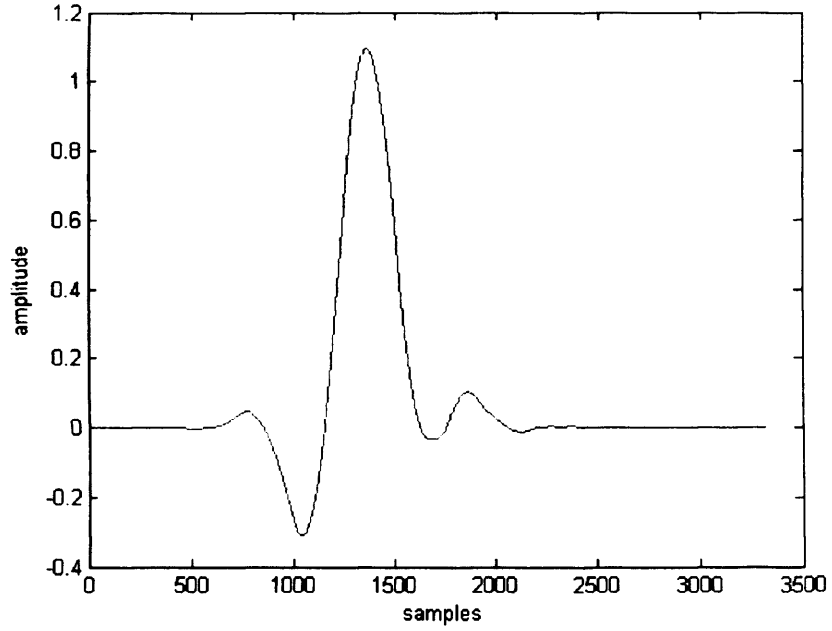
$$WT(b,a) = \int_{-\infty}^{\infty} g(t)\frac{1}{\sqrt{a}}\Psi^{*}(\frac{t-b}{a})dt \qquad (4.19)$$

where a is the scale parameter, b is the position parameter and $\Psi(t)$ is the mother wavelet. An example of the mother wavelet is the one shown in figure 4.7(b) from the Symlet family of wavelets. The trouble with wavelet analysis is that it requires attention on how to choose the decomposition level and how to choose the proper mother wavelet from various families of wavelets to match the signal characteristics and to consider the computation cost.

To demonstrate these time-frequency analysis tools, a typical signal obtained from a nail scratch on an MDF board is shown in figure 4.8 with its power spectrum. The resulting STFT analysis using the window function in figure 4.7(a) and the wavelet analysis using the mother wavelet in figure 4.7(b) are shown in figure 4.9. Both techniques indicate no significant fluctuation of frequency with time.

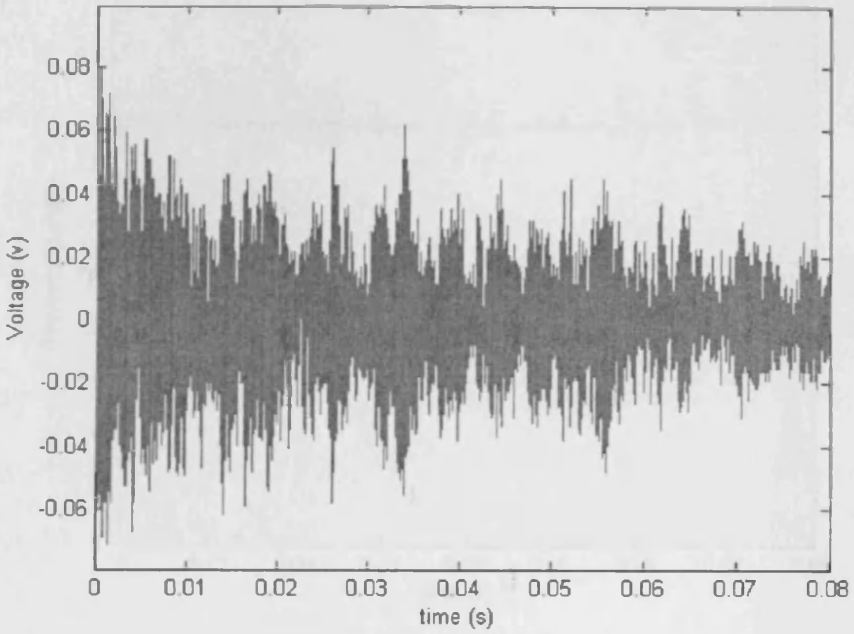


(a)

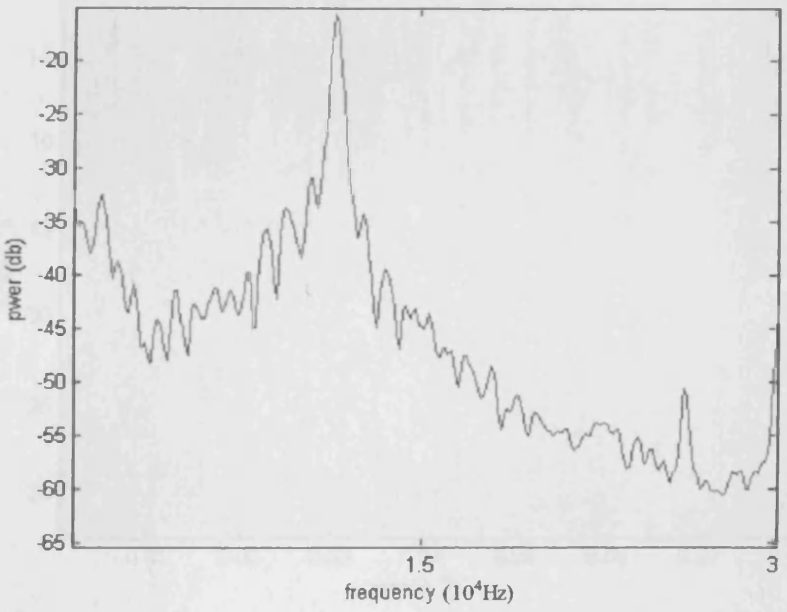


(b)

Figure 4.7. (a) Hann window function (b) mother wavelet from the Symlet family

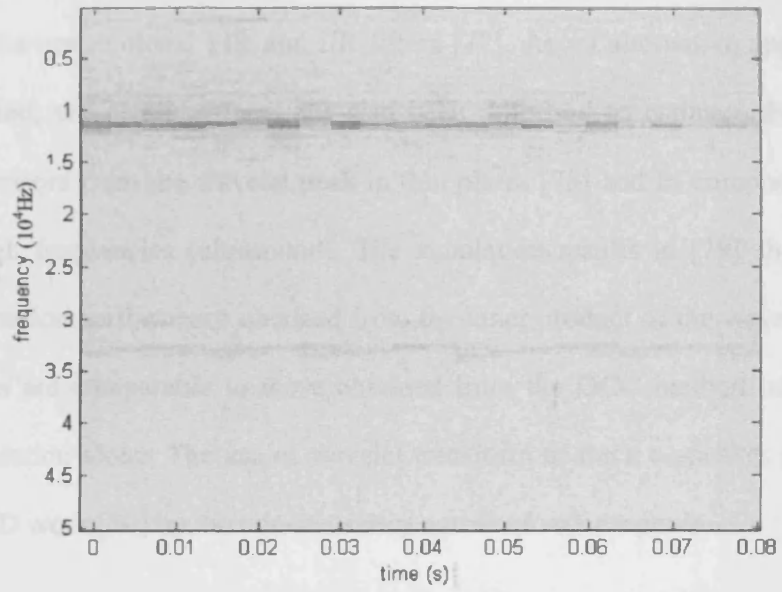


(a)

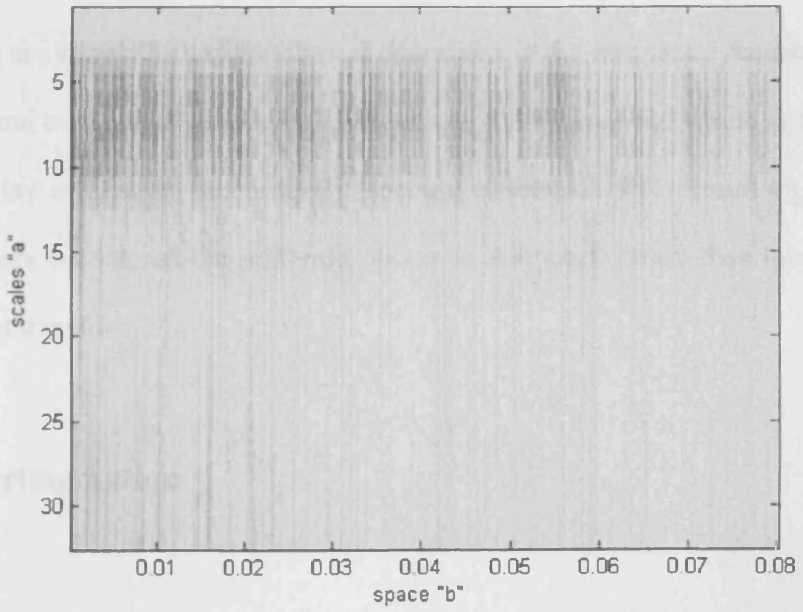


(b)

Figure 4.8. (a) Time signal example from nail scratch on MDF board (b) power spectrum of the signal.



(a)



(b)

Figure 4.9. Time-frequency representation for the signal in figure (4.8) using (a) STFT and (b) wavelet transform.

Wavelet transform has been used to de-noise signals before applying GCC as another option to the conventional FIR and IIR filters [77]. As an alternative approach to the GCC method, wavelet transform has also been proposed to estimate the time delay between sensors from the wavelet peak in thin plates [78] and in composite laminates [50] at high frequencies (ultrasound). The simulation results in [79] show that time delay estimation performance obtained from the inner product of the wavelet transform coefficients are comparable to those obtained from the GCC method but better than cross correlation alone. The use of wavelet transform to track a speaker is justified in Griebel PhD work [80] by the non-stationary nature of voice signals.

In the TAI application, the nature of the observed signals is stationary to some extent and the use of the STFT tool is sufficient and effective. Furthermore, Fourier techniques are known for their analytical operations in the frequency domain, providing powerful and convenient tools such as cross spectral estimation, which is fundamental in time delay estimation. Accordingly, spectral estimation tools based on the Fourier transform are considered the preferred option in this work rather than those based on the wavelet transform.

4.6 Experimentation

Characterising the features of the source signal, particularly wave velocity and spectrum, as well as choosing the signal analysis tool as seen in the previous section, is part of the investigation to provide essential information for the algorithms' development in the following chapters. Therefore, it is chosen to include the

experimentation here in a compact form rather than having it in a separate chapter. The experimental set up used in this work consists of tangible object, sensors, signal conditioning circuit, data acquisition card, a PC with Matlab software and a display unit with the following description,

- **Interactive Object**

The chosen interactive object is a 6mm thickness Medium Density Fibre (MDF) board of size 1.5 x 1.2 m^2 and a smaller board of size 0.8 x 0.9 m^2 used for exhibitions as shown in figure 4.10. This board is made from composite material. It is available on the shelf and has wide domestic uses such as for wall partitions, furniture and decoration.

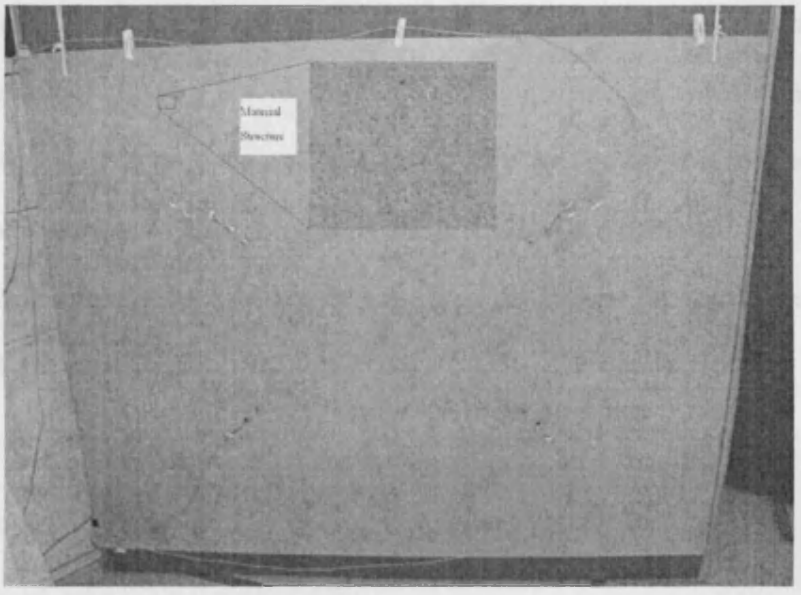


Figure 4.10. MDF interactive board (1500mm x 1200mm).

- **Sensors**

The choice of an appropriate sensor is vital for development success since it is the core element responsible for obtaining the important information within the acquired signal. Therefore the required specification for a good sensor is to be of high sensitivity to vibration, of wide frequency response, of low noise and preferably of small compact size. After intensive experimental tests of various types of sensors including microphones, VDF piezoelectric film, Ceramic piezoelectric sensor, sounder and accelerometer as pictured in figure 4.11, the accelerometer model BU-1771 from 'Knowles Acoustics' was found to be the best for its specifications as well as because of its reasonable price.

This sensor has a built in JFET transistor, which is a great feature used to convert the very high impedance of the piezoelectric element into low impedance. This has the practical benefit of allowing the transformation of signals through wires without having the amplifier placed close to the sensor. The two-wire configuration is used to feed the sensor with power and transfer the signal back to the data acquisition board. The frequency response of the sensor is shown in figure 4.12 where the resonance occurs at around 10 KHz.

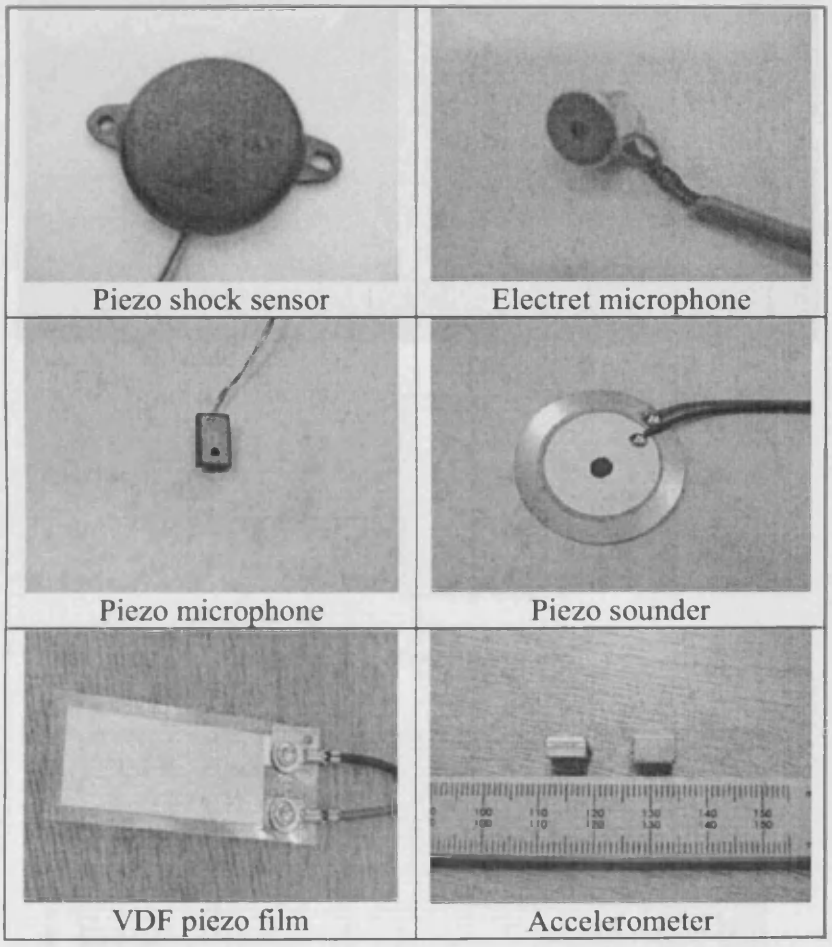


Figure 4.11. Different types of tested sensors.

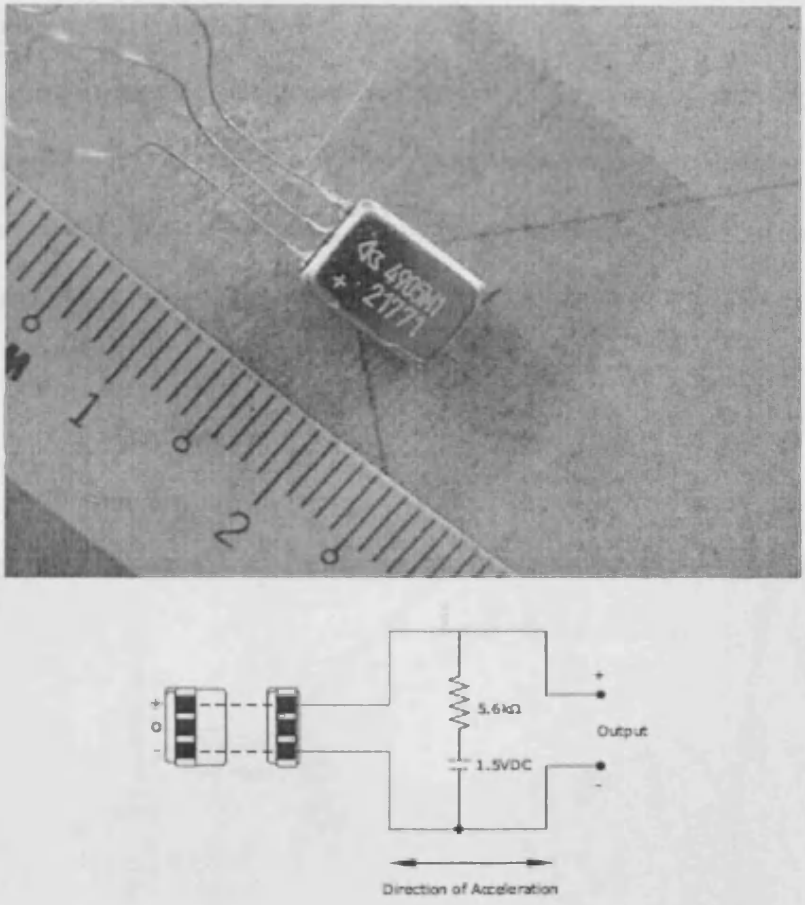


Figure 4.12. BU-1771 sensor, connection diagram and frequency response.

➢ **Signal Conditioning Circuit**

The signal conditioning board shown in figure 4.13 consists of signal preamplifier model Kemo M040, used to boost the signal received from sensors plus other components for basic filtering and for providing phantom power to sensors. The preamplifier gain is 25 at supply voltage of 9V.

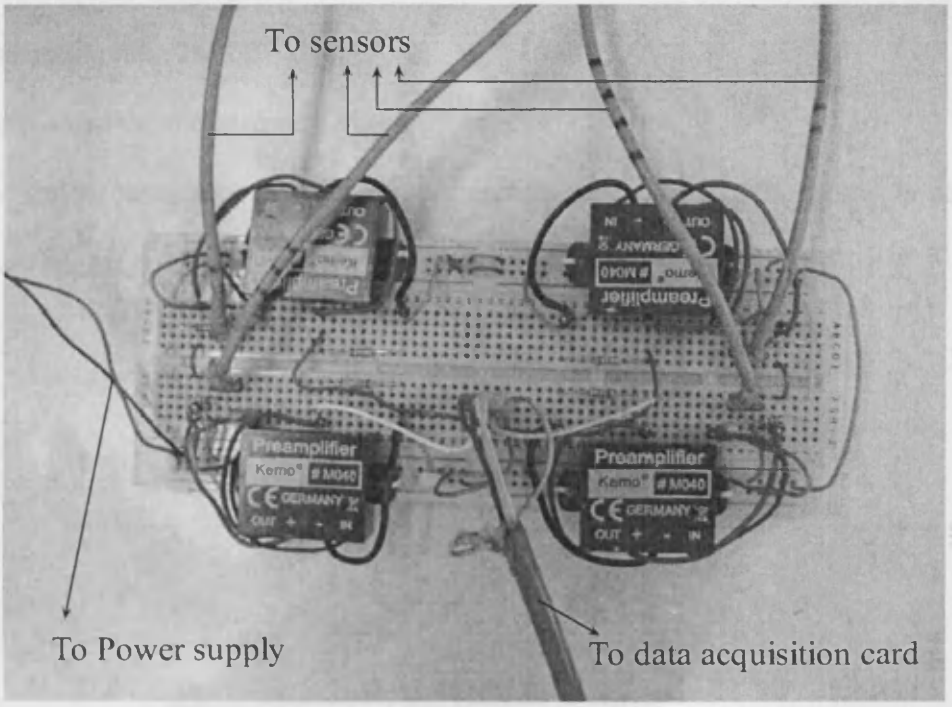


Figure 4.13. Signal conditioning board.

- **Data Acquisition Card**

The theoretical localisation accuracy is limited by the sampling frequency since the sampling time must be grater than the shortest time difference. Therefore, it is better to choose a data acquisition card with a sampling frequency as high as possible. The data acquisition board used is a 32-bit PCI-Bus architecture model NuDAQ-2010, which is pictured in figure 4.14 and has the following main specifications:

4-channel simultaneous analogue inputs

4-bit ADC with sampling rate up to 2MHz

Max sampling rate: 2MS/s

Supports software and hardware trigger

Single ended connection is used for the measurements input. The board is set for software trigger and the highest sampling rate of 100 Kbps is used.

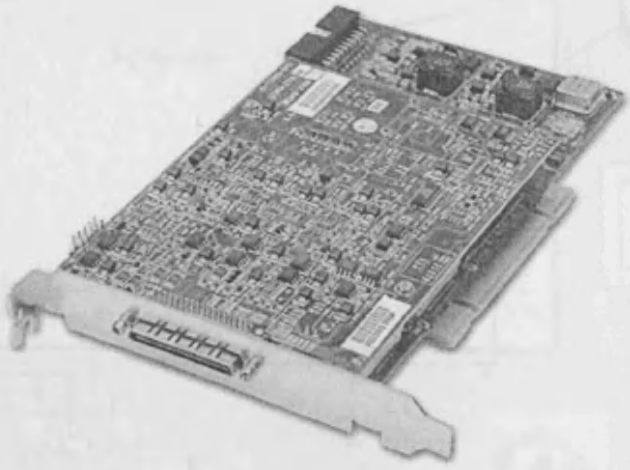


Figure 4.14. DAQ-2010 data acquisition board.

➢ **Signal Processing**

All the developed algorithms are written in Matlab code version 6.1 and the data acquisition card is operated by Matlab driver. Mouse movements and clicks are controlled from Matlab via Java Classes which allows interaction with any Windows application such as Microsoft Paint.

The complete TAI system including the above list of equipments is depicted in figure 4.15. Human interaction with the tangible object can be performed passively with a finger nail or other solid object such as a plastic brush, wooden stick or metal bar. The result of processing the interaction can be displayed directly onto a monitor or more intuitively by projecting the screen onto the interactive object.

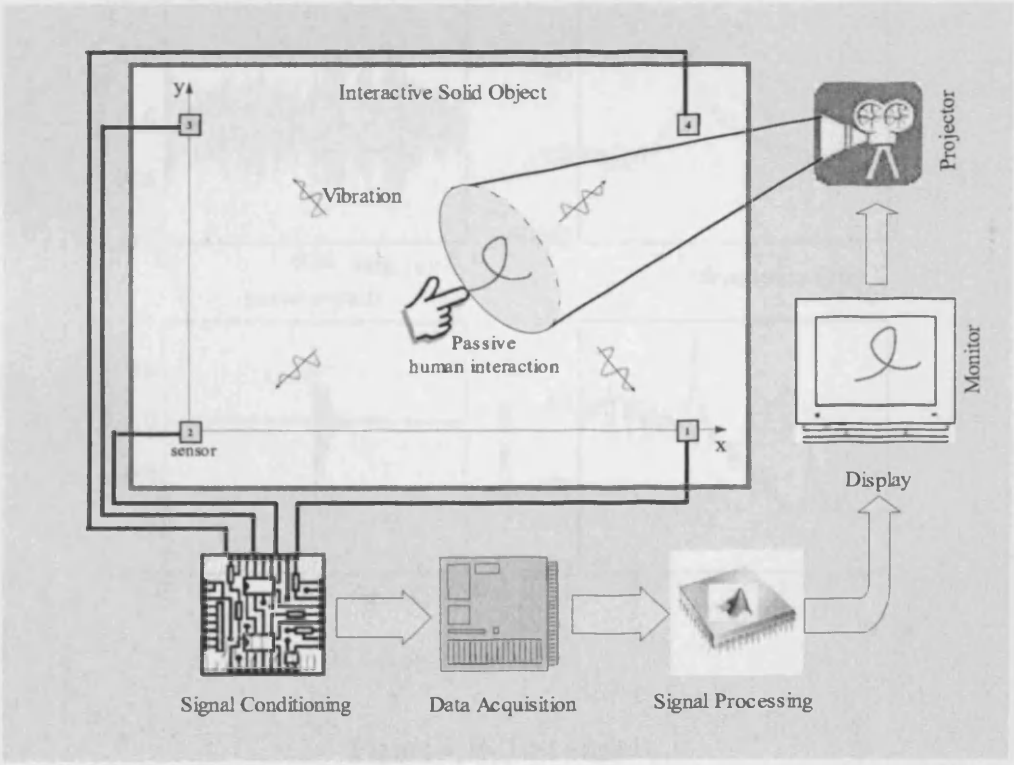


Figure 4.15. TAI model diagram

The application used in this system is Microsoft Paint as a means of creating graphics from a finger touch on a dead object.

Test signals of nail click, fingertip rubbing and metal spoon impact on the surface of the MDFF tangible object are shown in figure 4.16 together with their corresponding

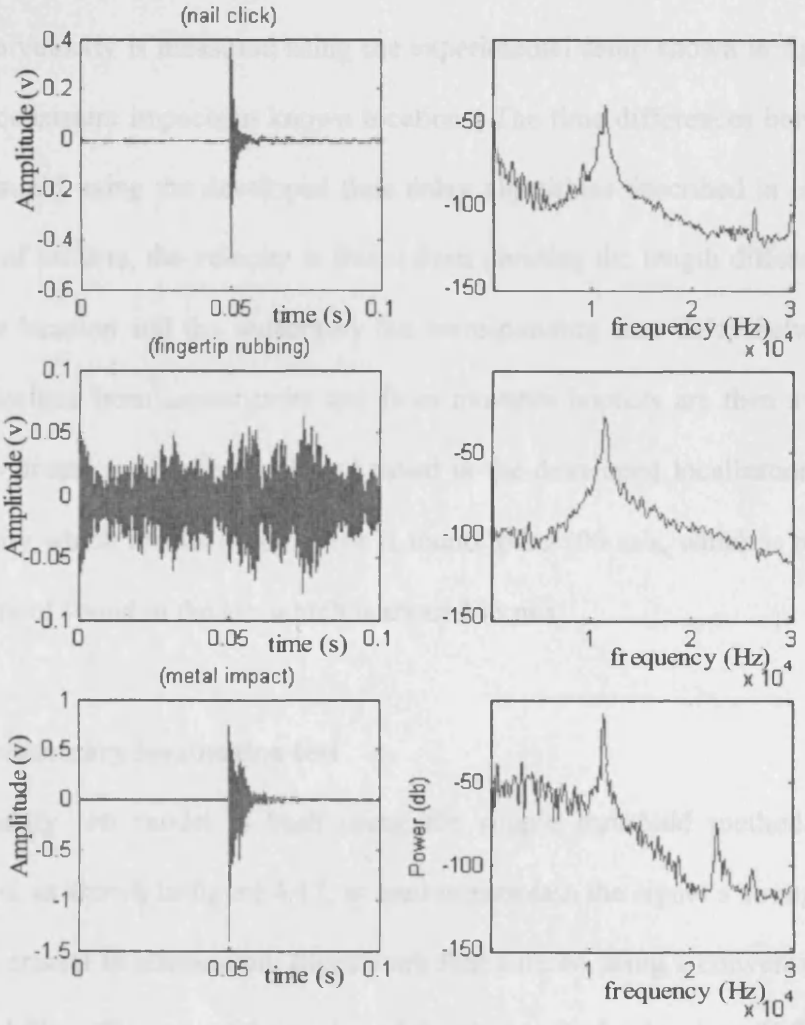


Figure 4.16 Test signals.

power spectrum. It is observable from the power spectrum of these signals that they occupy most of the audible frequency range and there is a nonlinear part caused by the sensor resonance. Also, the presence of considerable power is observable in the continuous signal at the higher frequencies beyond the resonance region which can also be utilised for tracking.

- **Wave Velocity measurement**

The group velocity is measured using the experimental setup shown in figure 4.15, by applying consistent impacts at known locations. The time differences between sensors were measured using the developed time delay algorithms described in chapter 5. For each pair of sensors, the velocity is found from dividing the length difference between the source location and the sensors by the corresponding time delay between sensors. These velocities from sensor pairs and from multiple impacts are then averaged. The obtained velocity is checked back and tuned in the developed localisation algorithms. The velocity which results in less error is found to be 700 m/s, which is roughly twice the velocity of sound in the air, which is about 345 m/s.

- **Preliminary localisation test**

A preliminary test model is built using the simple threshold method. A metallic whiteboard, as shown in figure 4.17, is used to maintain the signal's strength since this method is crucial to attenuation. Signals are first filtered using a conventional IIR low pass digital filter. The cut-off frequency of the chosen Chebyshev type II filter is 3KHz, with the frequency response shown in figure 4.18. Then the time difference is determined from the threshold level exceeding a threshold value of 0.3 v after

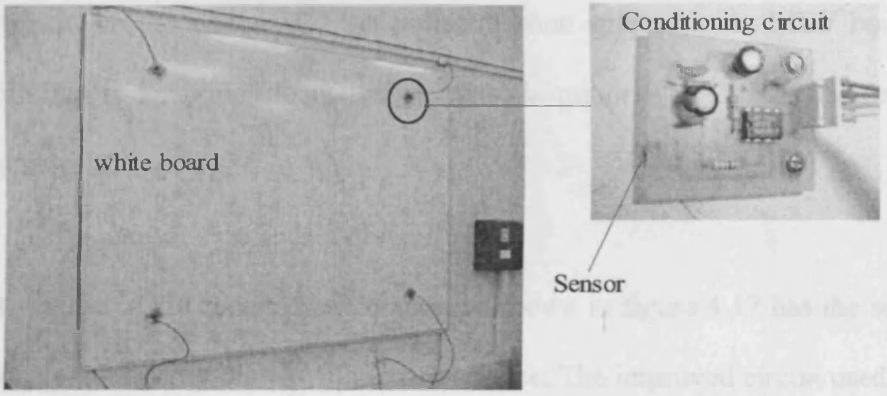


Figure 4.17 Whiteboard interactive object and conditioning circuit.

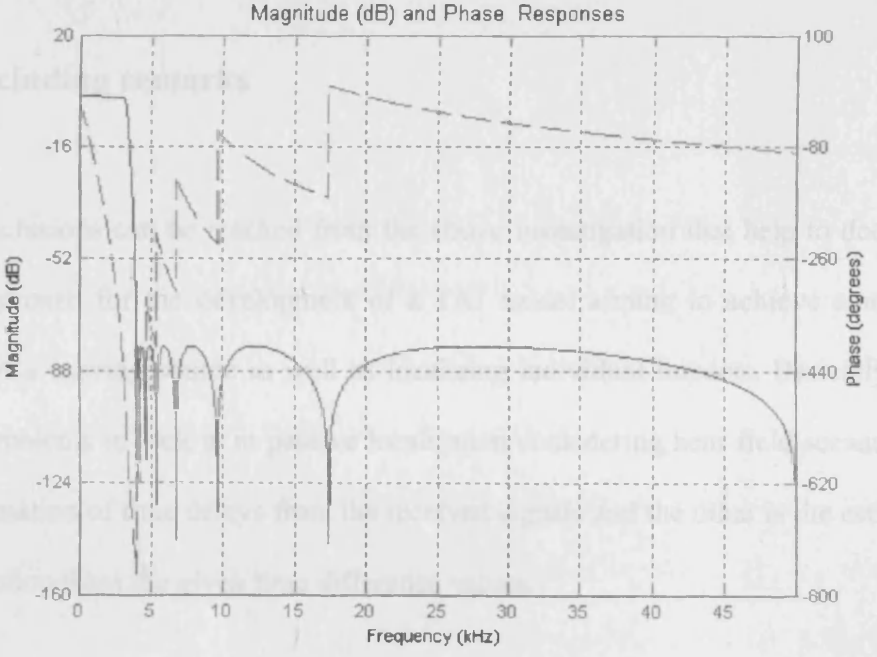


Figure 4.18. Frequency response of the IIR Low pass digital filter.

normalising each signal to its first lobe peak detected above noise level. Good results were achieved on the whiteboard but it failed when applied to an MDF board. The reason for this is that the attenuation in metal is minor while it is considerable in damping material.

The first version of the conditioning circuit, as shown in figure 4.17 has the sensor on board and placed directly on the whiteboard surface. The improved circuit used later on the MDF board takes advantage of the FET transistor built into the sensor to convert the impedance so that the signal is less susceptible to noise and the conditioning circuit can be placed away from the sensor, making sensor attachment easier and more practical.

4.7 Concluding remarks

Some conclusions can be reached from the above investigation that help to decide the proper approach for the development of a TAI model aiming to achieve continuous tracking of a moving source as well as localising individual impacts. Basically, there are two problems to look at in passive localisation considering near field scenario; one is the estimation of time delays from the received signals and the other is the estimation of the location from the given time difference values.

Due to the adverse effect of noise in a real environment caused by propagation, sensors and ambient factors, time delay estimation becomes problematic. The localisation in a TAI model involves the estimation of time delays from acoustic signals propagating in

a solid object and hence is subject to amplitude distortion, particularly in composite or non-uniform material, as well as dispersion. The method of raising edge passing a threshold level relys completely on the amplitude and therefore it can easily fail with considerable or uneven attenuation and dispersion. In any event, it is not suitable for tracking. Direct cross correlation is not sufficient on its own because of the amplitude dependence. PHAT filtering processes in the GCC approach are a promising choice to handle this problem because, as can be seen from equation 4.11, they treat the reverberation and noise by normalising the amplitude for all frequencies and therefore the operation becomes less sensitive to amplitude distortion and more suitable for wideband signals. The development of time delay estimation algorithms that imply optimisation and are phase dependant rather than magnitude dependant, with attention to dispersion, is crucial for TAI application.

In the TAI model, the intended resolution of the location estimation is comparable to the finger tip size, where a point source is assumed. This is in contrast to a speaker localisation application in the air, where the resolution is comparable to the head size. Also, tracking the trajectory of a continuously moving source requires good consistency of time delay estimation and the computation cost becomes a significant factor for real time implementation. To achieve such estimation accuracy from time delays which are corrupted with error, sensor fusion is required. The option of the beamforming approach is an example of sensor fusion and optimisation. Although it is accurate, it requires intensive computation and is based on bearing estimation assuming a far-field scenario, which is not an appropriate approximation for the TAI model.

The Maximum Likelihood approach is an error-based method that minimises the error between the measured time differences and the theoretical ones, searched over a hyperbolic grid area. This method is a common technique used for talker localisation and tracking in a room [81, 82]. Another probabilistic approach known as accumulated correlation [28] or spatial likelihood [29] has shown promising results in room acoustics by optimising the location based on cross correlation information from multi-sensor pairs and, importantly, uses spatial mapping rather than a far-field assumption. Therefore, these approaches are strongly nominated for the TAI approach with a good potential for success, although they encounter numerical optimisation. Other solutions [42] involving linearisation of the hyperbolic equations with sensor fusion are also of interest, particularly for tracking. These solutions will be considered in the following chapters.

Chapter 5

Enhanced Acoustic Source Localisation

In theory, two pieces of time difference of arrival information from three sensors are adequate to resolve the source location ambiguity problem. However, to achieve the required resolution for TAI application in practice, sensor fusion with nonlinear optimisation are proposed in this chapter using two different methods where the information from each pair of sensors including redundant sensors are utilised to optimise the source location. The first method finds the location where the probability of possible time differences is maximised based on the spatial likelihood while the second method determines the location where the error in the estimated time differences is minimised based on least squared error.

For further localisation enhancement, the dispersion problem has been considered. This is accomplished by introducing the detection of the cross correlation envelop via a Hilbert transform, which can be regarded as a temporal smoothing filter. Moreover, a criterion is developed for each method to detect outlier estimations.

Real signals from four sensors are used throughout this chapter to illustrate the functionality of sensor fusion and the proposed algorithms, but without losing the generality of employing additional sensors. That means the developed algorithms can handle an optional number of sensors. The experimental TAI model used in this work

consists of an MDF board of 6mm thickness with four sensors located at the corners of a 600 mm x 400 mm rectangle as illustrated in figure 5.1. The signals, $g_i(t)$, acquired by the i^{th} sensors in response to source signal $s(t)$ excited by nail click and scratch at location (200,300) mm on the board surface are shown in figure 5.2. These signals will be used as an example in the following theoretical development.

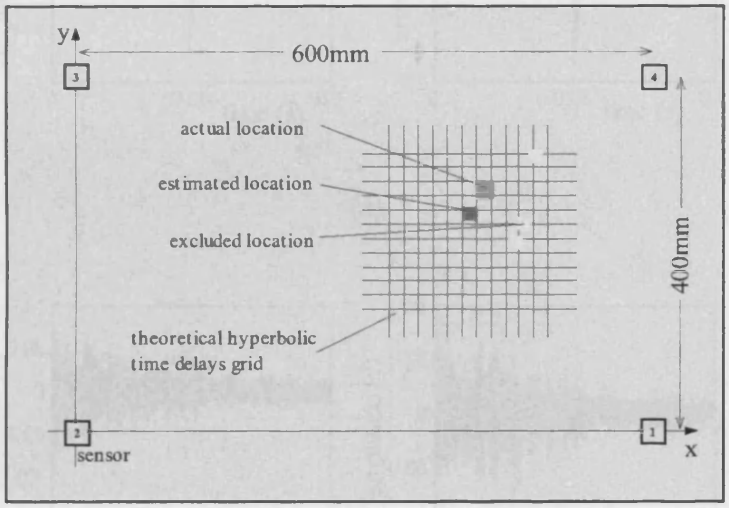
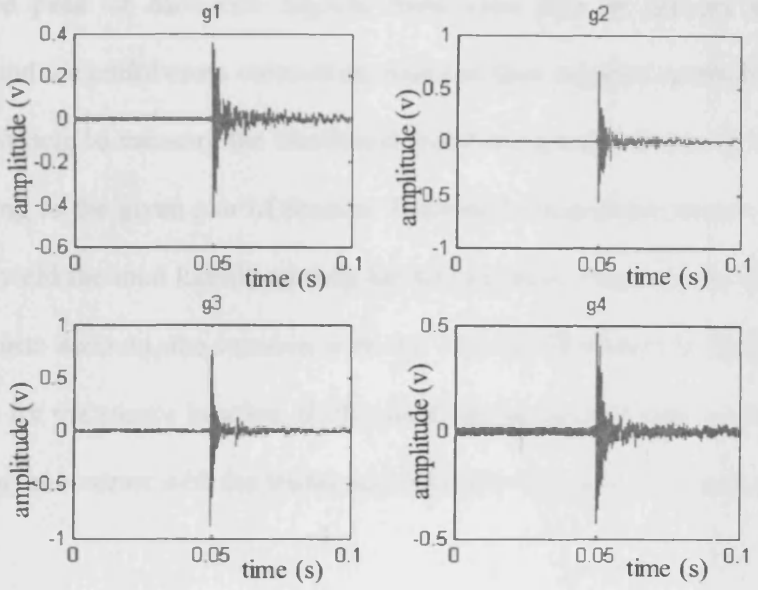


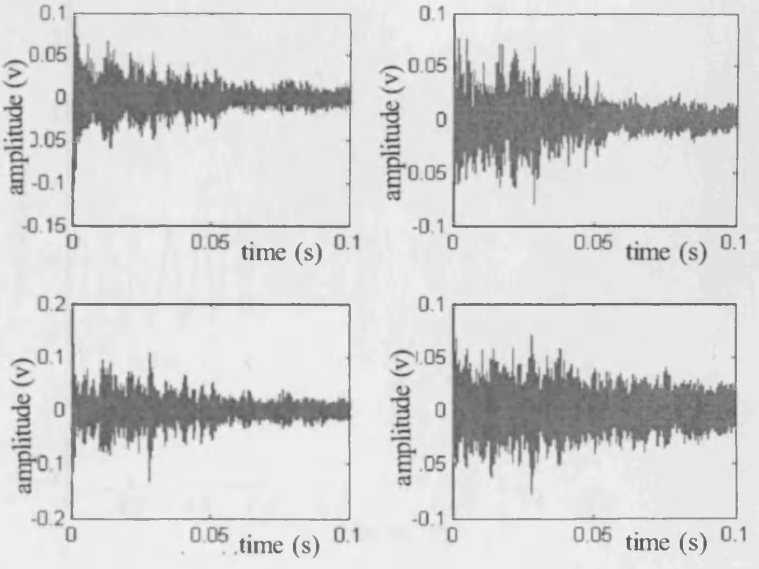
Figure 5.1 TAI model with four sensors

5.1 Spatial Likelihood-based Localisation

With real signals, cross correlation is not a perfect technique since it usually contains multiple peaks as shown in figure 5.3 and therefore there is no guarantee that the peak will occur at the correct time difference. This problem of noisy cross correlation can be handled by retaining the entire cross correlation vector of each sensor pair rather than



(a)



(b)

Figure 5.2. Real signals, g_i, from the i^{th} sensor generated by (a) nail click (impact) and (b) nail scratch on the MDF board.

selecting the peak of each one. Signals from each pair of sensors are first cross correlated, and the entire cross correlation vector is then mapped spatially to a common coordinate system to measure the likelihood that the source is at any of the hyperbolas corresponding to the given pair of sensors. Vectors from multiple sensor pairs are then summed to yield the total likelihood map for that location. After all the information has been taken into account, the location with the highest likelihood is finally selected as the estimate for the source location. Such a localisation process can be looked at as 3-D beamforming in contrast with the traditional beamforming used in phased array.

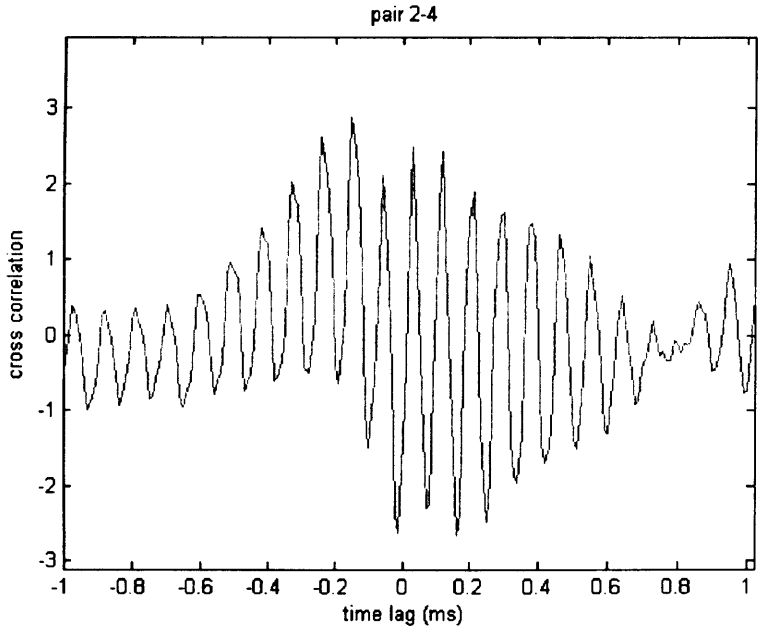


Figure 5.3. Cross correlation of impact signals showing multiple peaks

This localisation algorithm has been developed for room acoustics by two authors individually using different mathematical formulations. In [83] the processing is performed in the time domain and is known as accumulated correlation, and in [84] the processing is performed in the frequency domain and is known as the spatial likelihood function. In this work the algorithm is enhanced for TAI development and referred to as Enhanced Likelihood Mapping (ELM).

5.1.1 Theory

By rewriting equation (4.9) as $g_i(t) = h_i(t) * s(t-\tau_i) + n_i(t)$, assuming n_i independent zero mean white Gaussian noise with variance σ_i^2, the theoretical proof of the algorithm in [83] is based on treating $g_i(t)$ as an estimator for τ_i and $s(t-\tau_i)$ where $\tau_i = \|q-u_i\|/v$ given u_i the i^{th} sensor location. Using Bayes' Rule stating: *posterior* = *likelihood* * *prior* / *marginal likelihood*, the posterior probability that the source is located at q is

$$p = P(q,s \mid g_1,..,g_N) = \frac{P(g_1,..,g_N \mid q,s)P(q,s)}{P(g_1,..,g_N)} \tag{5.1}$$

Ignoring the denominator, which is a normalisation constant and not a function of q, and assuming the prior $P(q,s)$ uniform, then maximising (5.1) reduced to maximising the likelihood $P(g_1,..,g_N|q,s)$. With g_i considered as an independent random variable, it can be shown that,

$$p' = \prod_{i=1}^{N} P(g_i \mid q,s) = \prod_{i=1}^{N} e^{-\int_{-T}^{T} \frac{[g_i(t+\tau_i)-s(t)]^2}{2\sigma_i^2} dt} \tag{5.2}$$

Substituting s with its maximum likelihood estimate given by $\hat{s} = \frac{1}{N}\sum_{i=1}^{N} g_i(t+\tau_i)$ in (5.2), and assuming equal σ_i for all sensors, then taking the logarithm yields,

$$\log p' = -\sum_{i=1}^{N} \int_{-T}^{T} [g_i(t+\tau_i) - \hat{s}(t)]^2 dt = \frac{2}{N} V_{AC} - \frac{N-1}{N} V_E \tag{5.3}$$

where

$$V_{AC} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} \int_{-T}^{T} g_i(t+\tau_i) g_j(t+\tau_j) dt \tag{5.4}$$

is the accumulated correlation and

$$V_E = \sum_{i=1}^{N} \int_{-T}^{T} g_i^2(t+\tau_i) dt \tag{5.5}$$

is a constant representing the combined energy of the signals. Accordingly, the estimated location can be found from the maximum of V_{AC} given in (5.4).

In [84], the theory derivation of the algorithm is based on treating the cross correlation as an observational estimate of $P(g_1,..,g_N \mid q,s)$ which is related to the posterior estimate as given in (5.1) under the same assumption for the prior probability $P(q,s)$. Thus, by substituting the time difference $\tau_{ij}(q)$ between sensors i and j given as a function of source location q in the GCC formula (4.11), the spatial likelihood function (SLF) in the frequency domain for a pair of sensors is obtained as,

$$SLF_{ij}(q) = \int_{-\infty}^{\infty} \Psi(f) G_i G_j^* e^{j2\pi f \tau_{ij}(q)} df \tag{5.6}$$

The advantage of (5.6) is that it allows for the filtering processes $\Psi(f)$ as discussed in section 4.4.1.1 to be performed inclusively in the frequency domain.

5.1.2 ELM Algorithm

Given N sensors, the usable number of time differences is given by M none-repeated combinations of sensor pairs given by

$$M = C_2^N = \frac{N!}{2!(n-2)!} \tag{5.7}$$

With three sensors, three time differences are available, providing one redundant time difference. By adding a fourth sensor, the time delay information from equation (5.7)

is doubled to six. The ELM algorithm allows for sensor fusion by utilising M time difference information to improve the accuracy and robustness of the estimated location.

From the above theory and by introducing a Hilbert envelope detection operator Θ, to be expressed later, the proposed ELM algorithm for TAI can be formulated in compact form for both of time domain and frequency domain filtering processes as follows,

$$ELM_t(x,y) = \sum_{\substack{i=1,j=2 \\ i\neq j \\ i,j\neq j,i}}^{N-1,N} \Theta[\int_{-\infty}^{\infty} g_i^{BPF}(t) g_j^{BPF}(t+\tau_{ij}(x,y))dt] \tag{5.8}$$

$$ELM_f(x,y) = \sum_{\substack{i=1,j=2 \\ i\neq j \\ i,j\neq j,i}}^{N-1,N} \int_{-\infty}^{\infty} \Psi(f) G_i G_j^* e^{j2\pi f \tau_{ij}(x,y)} df \tag{5.9}$$

given,

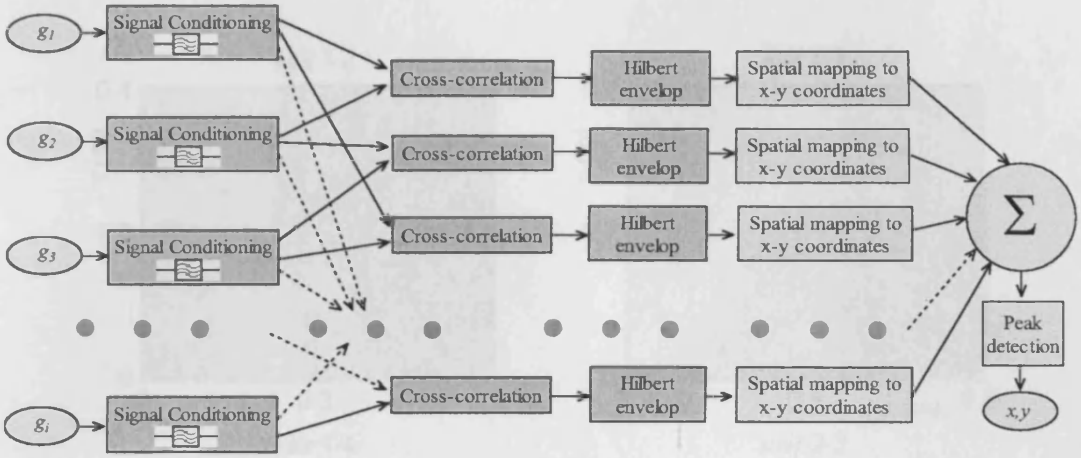
$$\tau_{ij}(x,y) = (\sqrt{(x-x_i)^2+(y-y_i)^2} - \sqrt{(x-x_j)^2+(y-y_j)^2})/v \tag{5.10}$$

where $g^{BPF} = h_{BPF}(t) \otimes g(t)$ is the band pass filtered signal resulting from the convolution (denoted by $\otimes$) of the signal g and the impulse response of the band pass filter $h_{BBF}(t)$. $\Psi(f)$ is the weighting filter. The estimated locations $\hat{x}, \hat{y}$ can then be found by locating the maximum of equation (5.9) or (5.10) as,

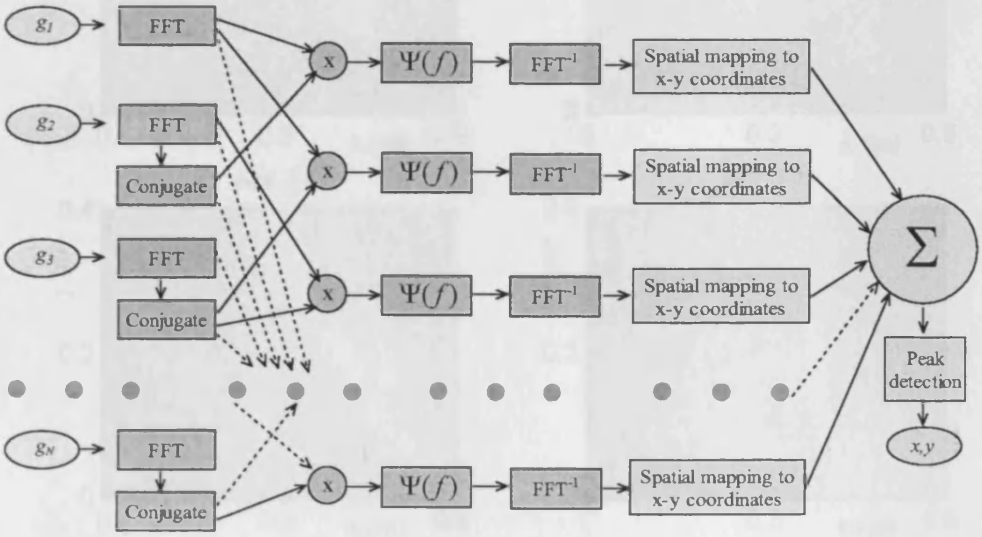
$$(\hat{x},\hat{y}) = \arg_{x,y} \max(ELM(x,y)) \tag{5.11}$$

Attention should be paid to the search time required to maximise the 3-D equation of (5.11). This ELM algorithm is illustrated in the diagram shown in figure 5.4.

Let for instance evaluate the core algorithm using raw signals by letting $\Theta([.]) = [.]$ and $h_{BPF}(t) = \delta(t)$. Considering the example of having four sensors located on the surface of the tangible object as in figure 5.1, the theoretical time difference map for all sensor pairs can be computed numerically with a 1 mm step from the hyperbolic equation (5.10) and the resulting time difference maps are shown in figure 5.5. Given the signals in figure 5.2, the spatial likelihood map for each pair of sensors obtained from equation (5.8) before performing the summation is shown in figure 5.6. The darker hyperbola curve observed in each image represents the locus of the possible source location sited by the corresponding sensor's pair. By summing the spatial likelihood from all pairs, the source location can be obtained from the maximum of the absolute value of the final likelihood map. The result of equation (5.8) is shown in figure 5.7 for nail click. The estimated location from equation (5.11) is (220,380) mm.



(a)



(b)

Figure 5.4. Algorithm diagram for (a) ELM_t and (b) ELM_f

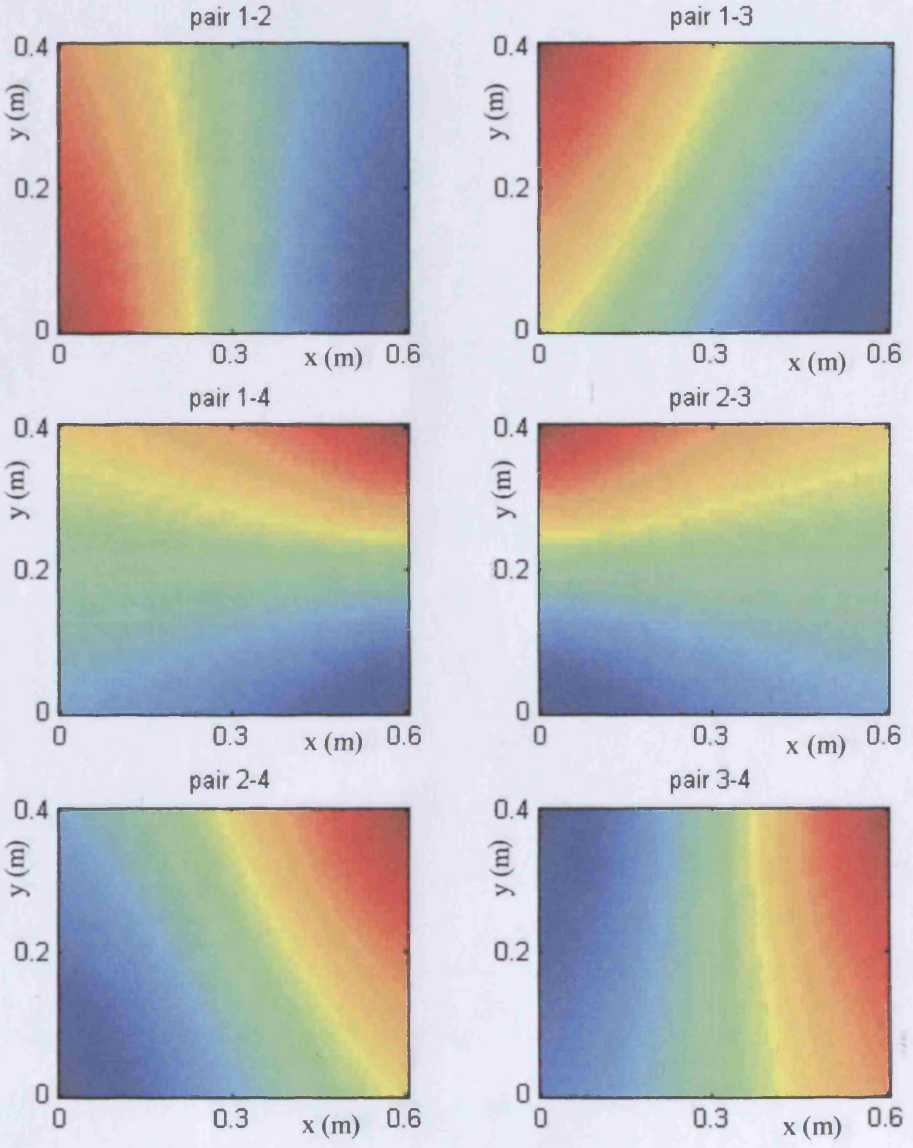


Figure 5.5. Theoretical hyperbolic map of time differences

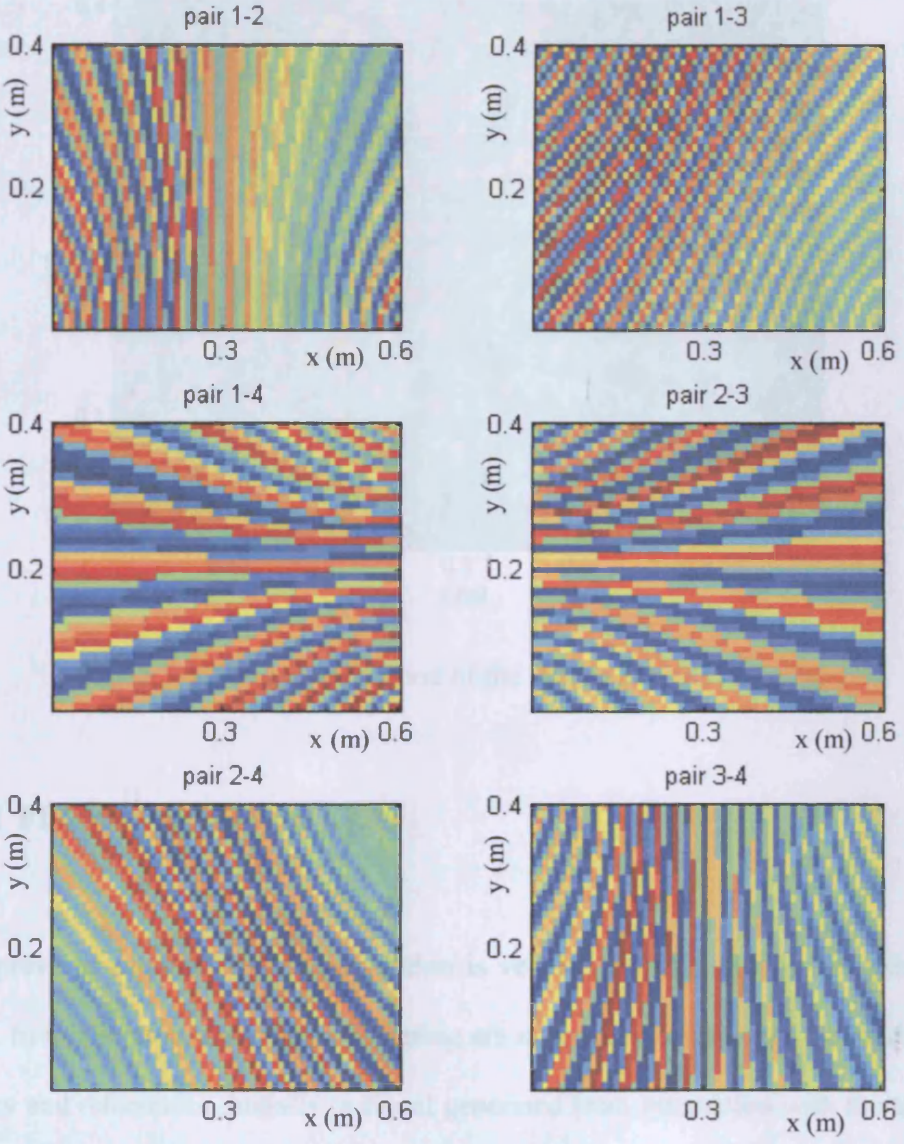


Figure 5.6. Spatial likelihood for the source at (200,300)mm for each pair of sensors

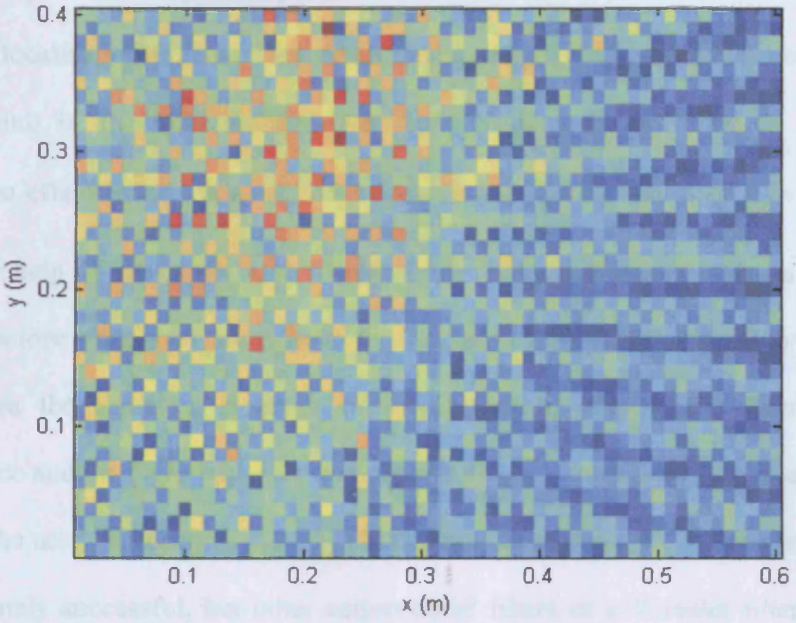


Figure 5.7. Spatial Likelihood of the source at (200,300)mm.

5.1.2.1 Filtering Process in ELM

In the previous section the ELM algorithm is verified for TAI application using raw signals. In this section, two types of filtering are introduced to improve the localisation accuracy and robustness. Initially, a signal generated from interaction with the tangible object is distinguished from the background noise using a simple threshold technique. This noise can be seen when there is no interaction activity with the board. Practically, the noise is lower than 5 mV, which is very low compared to 50 mV produced by weak signals generated from rubbing the surface depending on the signal amplifier used. Hence, only signals above the threshold of 50 mV are processed and localised.

The first expected attempt to improve the localisation is to condition the input signals prior the localisation process. The conditioning stage comprises signal normalisation and filtering in the time domain. The normalisation is performed to reduce the attenuation effect by dividing each signal by its standard deviation, $g_i'(t) = g_i(t)/\sigma_{g_i}$. This will help to make the signal waveforms from multiple sensors as similar as possible before cross correlating them. The purpose of the pre-filtering is to remove the noise from the signal as a result of low frequency components from electrical interference and the high frequency components from the nonlinear response around 10 KHz for the used sensor model BU-1771. The use of the popular IIR filter was found to be adequately successful, but other conventional filters or a Wavelet filter is also an option. The designed digital filter is a 10^{th} order band-pass Elliptic filter with a lower cut off frequency of 500 HZ and an upper cut off frequency of 8 KHz. The frequency response of the designed filter is shown in figure 5.8. Expanding the upper or lower bandwidth of the filter was found to degrade the results.

While the design of the IIR filter involves defining the stop band cut off frequencies where the noise is dominant and setting the attenuation level, in comparison, the Wavelet filter design requires decomposing the signal into levels, identifying which components contain the noise, and then reconstructing the signal without those components. This method may result in losing sharp features of the signal and therefore an alternative thresholding technique can be used which involves discarding only the portion of the details that exceeds a certain limit.

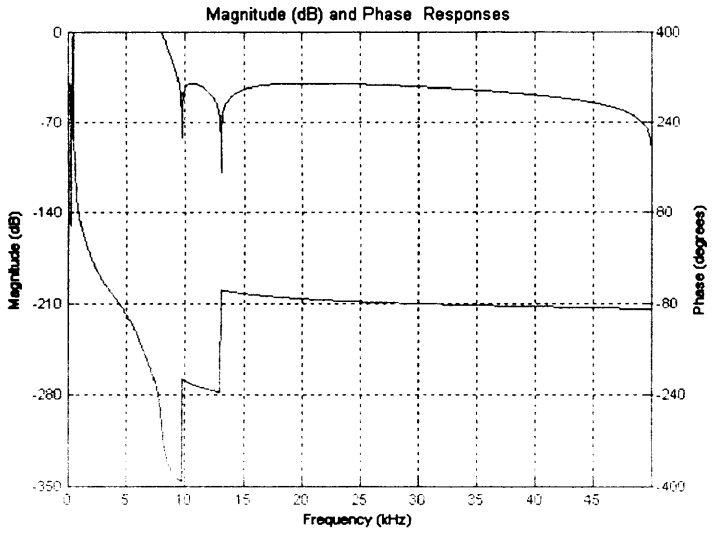


Figure 5.8 Elliptic band-pass filter design

The generation of the spatial likelihood maps are repeated for the same test impact signals after applying the proposed filter. The results are shown in figure 5.9 for the individual pair of sensors and the summation of these maps is the final map given by the ELM_t equation as shown in figure 5.10. It is clear that pre-filtering has significantly improved the reliability of the estimation, as can be seen from the smoothness achieved in the likelihood map in figure 5.10, where the local maximum becomes distinctive compared to the multiple peaks in figure 5.7 using raw signals.

The second filtering type employed here is the PHAT process discussed in section 4.4.1.1. This is achieved by substituting the filtering process $\Psi_{PHAT}(f)$ given in equation (4.13) into equation (5.9). The resulting map of ELM_f is calculated for the same signals used in the above example as shown in figure 5.11. It is clear that a sharper peak is obtained compared to the pre-filtering method in the time domain.

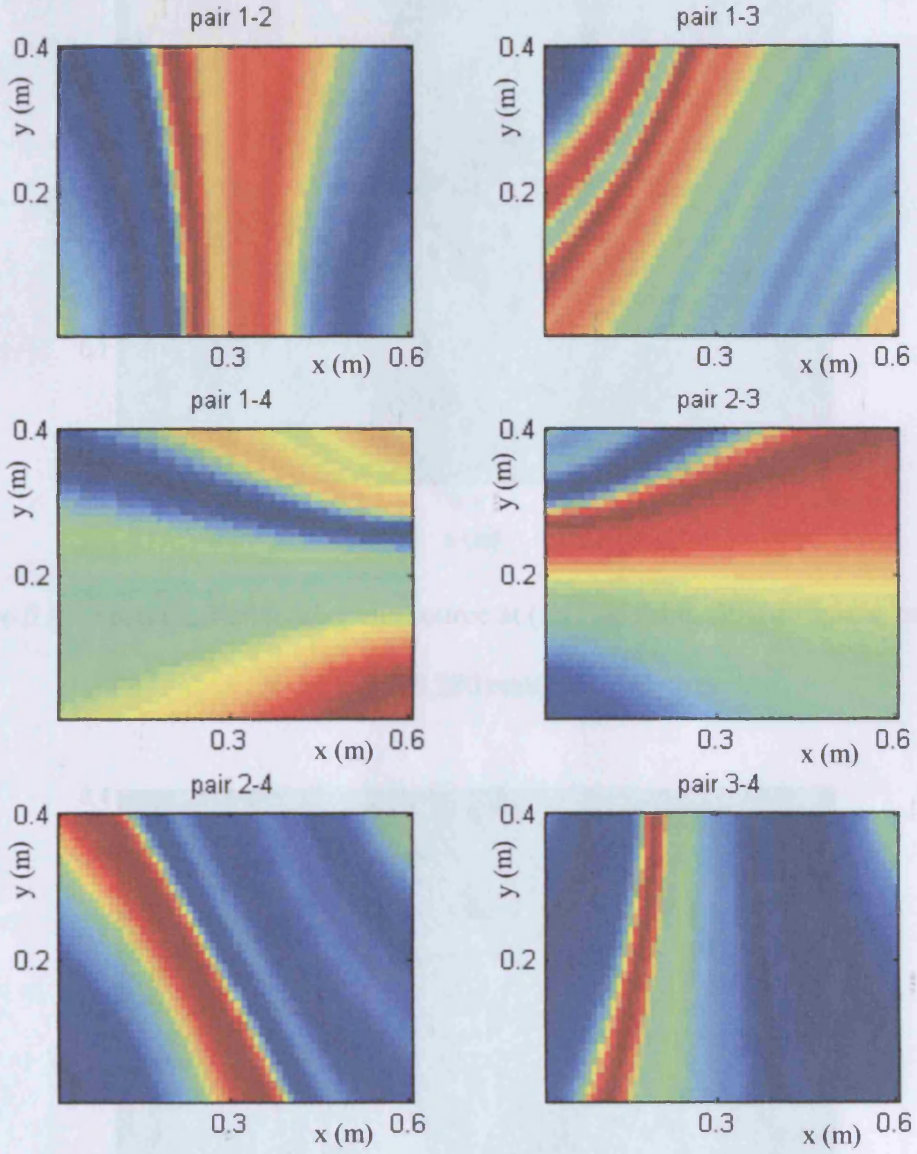


Figure 5.9 Spatial likelihood for the source at (200,300)mm for each pair of sensors after filtering the signals.

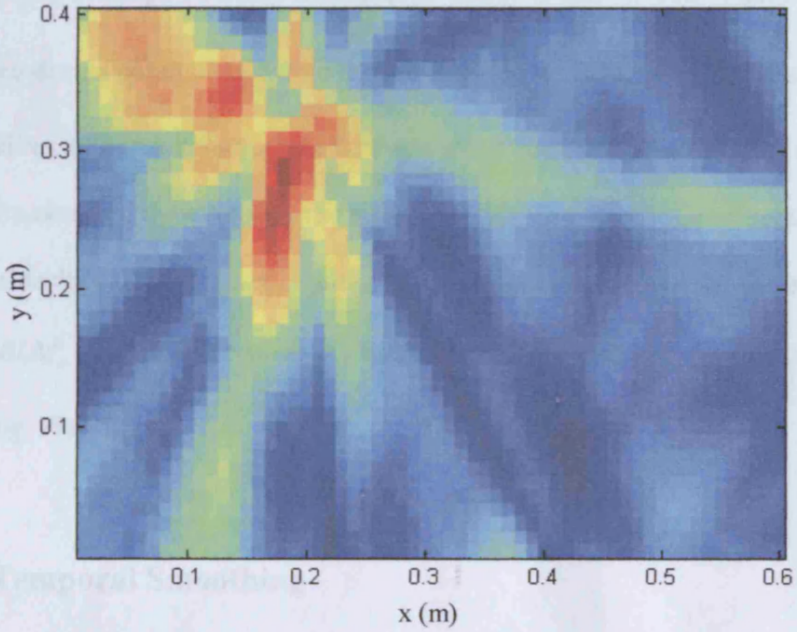


Figure 5.10. Spatial Likelihood of the source at (.2,.3)m using filterd signals, peak at (180,280)mm.

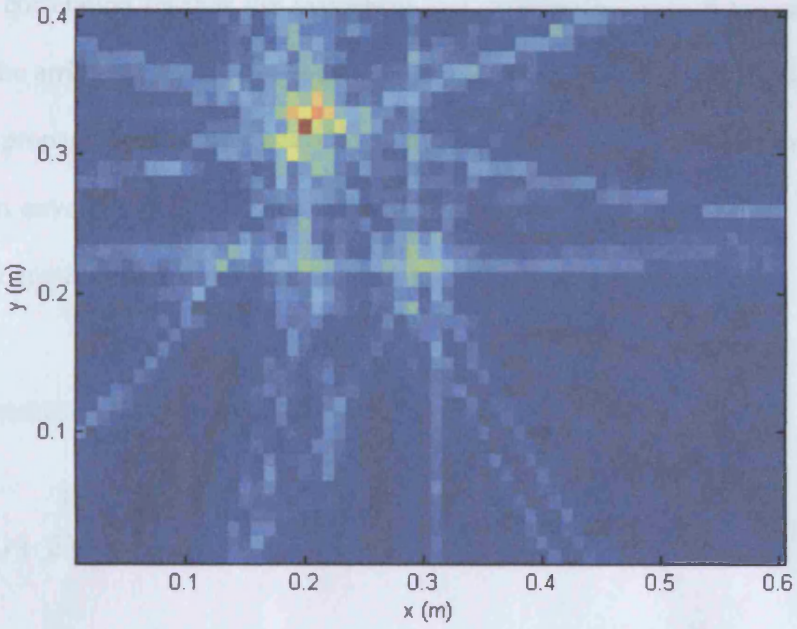


Figure 5.11 Spatial likelihood map using PHAT process

A significant advantage of using ELM_f over ELM_t is that the PHAT process does not require any design parameters, while the pre-filtering in ELM_t, whether it is IIR, FIR or Wavelet filter, requires knowledge of the dominant signal components and noise which is normally obtained by analysing the signals. That means that if these parameters have been considerably changed as a result in changing the object material for example, the filter of ELM_t (IIR, FIR or wavelet) has to be redesigned but this is not necessary when using PHAT.

5.1.2.2 Temporal Smoothing

Further enhancement in the ELM algorithm is achievable by treating the dispersion effect in solids. Theoretically, in non dispersive multiple output systems, the output of the cross correlation reaches the maximum at a time lag equal to the time difference between the arrival of the input signals. On the other hand, in dispersive systems, where the wave propagation velocity is a function of frequency, the output peak of the cross correlation envelope occurs as the time lag equals the group delay of the wave. This fact can be interpreted in practice using a Hilbert transform [85].

The analytical signal of a given function $z(t)$ is defined by

$$Z(t) = z(t) + j\tilde{z}(t) \tag{5.12}$$

where the imaginary part in (5.12) is the Hilbert transform of $z(t)$ given by,

$$\tilde{z}(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{z(\zeta)}{t-\zeta} d\zeta \tag{5.13}$$

the envelope function of $z(t)$ is found from the magnitude of $Z(t)$ as

$$\Theta(t) = [z^2(t) + \hat{z}^2(t)]^{1/2} \tag{5.14}$$

Equation (5.14) can be used through the operator Θ in equations (5.8) or (5.9) to reduce the error in cross correlation caused by dispersion.

To visualise the difference between algorithms and the effect of dispersion treatment, the ELM algorithm is applied to the impact signals shown in figure 5.2(a) and the produced spatial likelihood maps are shown in figure 5.12. With raw signals, the ELM algorithm produces multiple peaks in the likelihood map with several sharp local maxima comparable to the global maximum as shown in figure 5.12(a). By conditioning the input signals, the local maxima are reduced, as shown in figure 5.12(b). When a Hilbert envelope is applied, it is observable that the peak is enhanced by shifted local maxima towards the global peak and the overall ELM surface is smoothed as shown in figure 5.12(c). It is clear from figure 5.12(d) that the PHAT process produces sharper peaks with lower side-loops. For the purpose of evaluating the functionality of the ELM algorithm for tracking in the next chapter, the above results are repeated in figure 5.13 but for scratch signals as shown in figure 5.2(b). Although scratch signals produce more local maxima than the impact signals, the proposed ELM

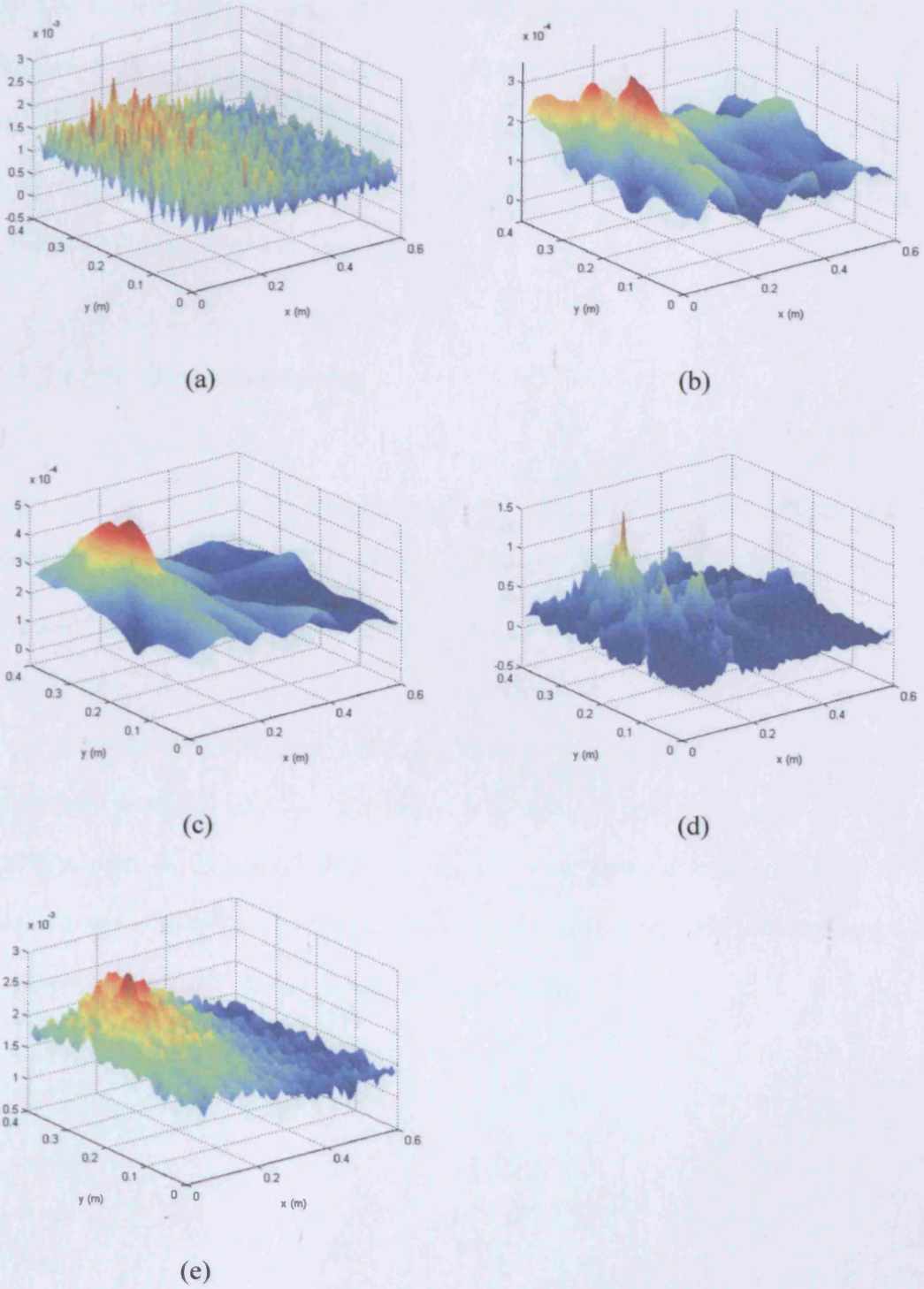


Figure 5.12. ELM of impact signals using (a) raw signals, (b) conditioned signals, (c) as in (b) with Hilbert envelope, (d) PHAT and (e) with Hilbert envelope only.

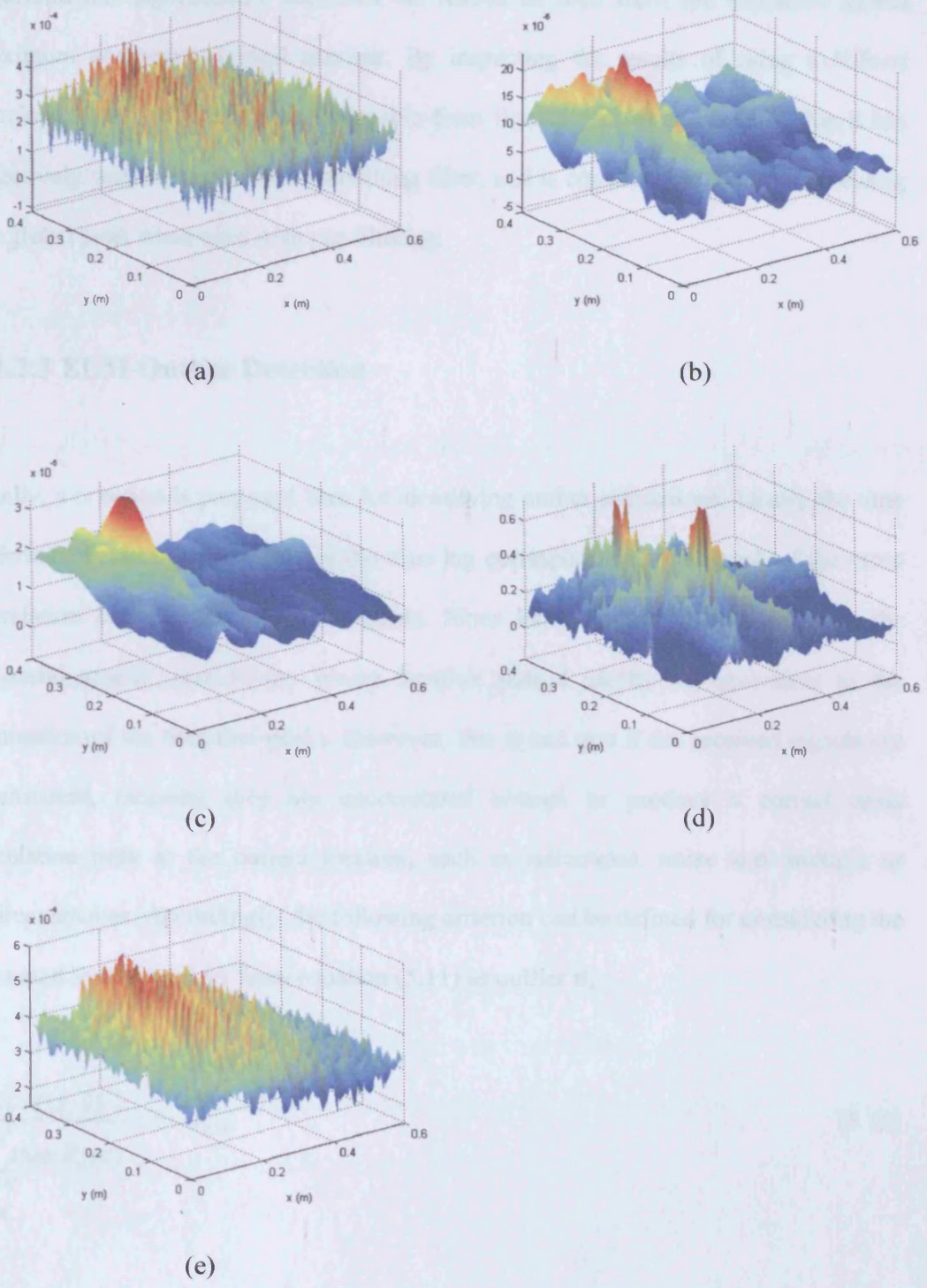


Figure 5.13. ELM of scratch signals using (a) raw signals, (b) conditioned signals, (c) as in (b) with Hilbert envelope, (d) PHAT and (e) Hilbert envelope only.

algorithm has significantly improved the results as seen from the enhanced global maximum and reduced local maxima. By inspecting the results of using a Hilbert envelope as the only filer, it is noticeable from figures 5.12(e) and 5.13(e) that it has effectively worked as temporal smoothing filter, and is considerably better at revealing the global peak when used with pre-filtering.

5.1.2.3 ELM Outlier Detection

Finally, a criterion is proposed here for identifying outlier estimations. Ideally the time difference between two sensors is the time lag corresponding to the peak of the cross correlation between the received signals. Since ELM applies spatial mapping, the produced spatial peak at the source location should ideally be equivalent to the summation of the temporal peaks. However, this is not true if the received signals are inconsistent, meaning they are uncorrelated enough to produce a correct cross correlation peak at the correct location, such as reflections, noise and multiple or indirect sources. Accordingly, the following criterion can be defined for considering the estimated location $(\hat{x}, \hat{y})$ from equation (5.11) as outlier if,

$$\frac{ELM(\hat{x},\hat{y})}{\sum_{\substack{i=1,j=2 \\ i\neq j \\ i,j\neq ji}}^{N-1,N} \max R_{ij}(\tau)} > 1 + \varepsilon_{ELM} \tag{5.15}$$

where $0 < \varepsilon_{SLM} < 1$ is a threshold value that can be adjusted to the required sensitivity.

Practically, with ε_{ELM} set to 0.1, the criterion in (5.15) has successfully identified estimated locations resulting from a handclap or from dropping an object on the floor as an outlier by marking them with a different colour that those obtained from tidy impacts on the board.

5.2 Time Difference-based Localisation

The unreliable time delay estimation using the cross correlation method was conquered in the ELM algorithm by maximising the likelihood of the source location rather than finding the time delays from each cross correlation.

Another option is to estimate the time differences first, corrupted with error of course, and then the sensor fusion algorithm is used to minimise the error while positioning the source. For this purpose, Maximum Likelihood (ML) estimators are proposed. This option is more flexible than ELM because it allows the choice of the required positioning algorithm. However, the accuracy of the positioning algorithm here vastly depends on the level of error in the time difference values. Therefore, it becomes crucial to develop a reliable algorithm to estimate time differences with as little error as possible, taking into account the processing speed. An efficient algorithm is developed for estimating time differences based on spectral estimation.

5.2.1 Linear Cross Spectral Phase

The classical time difference estimation is defined by equation (5.10) as the argument of the cross correlation peak. In the time domain, this can be improved by pre-filtering the signals or applying the most popular PHAT process using GCC, which involves filtering in the frequency domain, then returning to the time domain to extract the time difference from the peak as given by equations (5.9) and (5.10). An alternative method for estimating the time difference is to compute the Linear Cross Spectral Phase (LCSP). The proposed LCSP algorithm for TAI in this section estimates the time difference entirely in the frequency domain, making the estimation process more efficient and robust than the time domain algorithms and therefore particularly important for tracking a continuous source.

Let the received signal $g(t)$ be assumed a wide sense stationary process. Although some real signals are not so, this assumption usually holds for a signal within a time frame of small size, which is a practical necessity. The cross spectral density of signals $g_i(t)$ and $g_j(t)$ can be found from

$$P_{ij}(f) = G_i(f)G_j^*(f) \tag{5.16}$$

where $G(f)$ is the Fourier transform of $g(t)$. Since $g_j(t)$ is time delayed from $g_i(t)$ by τ as given in the signal model of (4.9), then in terms of the Autospectral density $A_{ii}(f)$ of $g_j(t)$, (5.16) can be expressed by,

$$P_{ij}(f) = A_{ij}(f)e^{-j2\pi f\tau} = |P_{ij}(f)|\angle\phi(f) \tag{5.17}$$

The time difference τ appears only in the phase angle ϕ of (5.17) as a linear function of the frequency f. Since the group velocity is used to compute the length difference, the group delay, not the phase delay, must be extracted from the phase function in (5.17) as given by [86],

$$\tau = -\frac{d\phi(f)}{df} \tag{5.18}$$

The cross spectrum and auto spectrum functions can be effectively estimated based on using STFT as given in chapter 4, then equation (5.18) can be computed numerically for the quantities given in samples, i, using linear regression of the form $\sum(f_i - f)\phi_i / \sum(f_i - f)f_i$ [87, 88].

5.2.2 ML Positioning

Given M pair of sensors, the Maximum Likelihood algorithm (ML) proposed here for TAI can handle the error in the time difference values by minimising the error between the given time difference $\hat{\tau}_m$ of the m^{th} pair and the ideal time difference $\tau_m(q)$ associated with the searched location q. If the estimated time differences are modelled by the random variable $\hat{\tau}_m = \tau_m + e_m$, where e_m is zero-mean additive white Gaussian noise with known standard deviation σ_m, then by assuming that the time differences

from each pair of sensors are statistically independent, the likelihood function can be expressed by the conditional probability density function given by [89],

$$p(\hat{\tau}_1,...,\hat{\tau}_M \mid q) = \prod_{m=1}^{M} \frac{1}{\sqrt{2\pi\sigma_m^2}} e^{\frac{-[\hat{\tau}_m - \tau_m(q)]^2}{2\sigma_m^2}} \tag{5.19}$$

taking the log of both sides of (5.19) yield,

$$\ln(p(\hat{\tau}_1,...,\hat{\tau}_M \mid q)) = -\frac{1}{2}\sum_{m=1}^{M} \ln(2\pi\sigma_m^2) - \sum_{m=1}^{M} \frac{(\hat{\tau}_m - \tau_m(q))^2}{2\sigma_m^2} \tag{5.20}$$

The ML estimation of location q is the position that maximises the likelihood function (5.20) or equivalently that minimises the second term, since the first term is not a function of q, which results in the following localisation criterion

$$J_{ML}(q) = \arg\min_q \left(\sum_{m=1}^{M} \frac{[\hat{\tau}_m - \tau_m(q)]^2}{\sigma_m^2}\right) \tag{5.21}$$

It is clear that J_{ML} is a weighted least error estimator. If no statistics are considered or σ_m is the same for all sensor pairs, then the denominator is constant and (5.21) is reduced to the following formula,

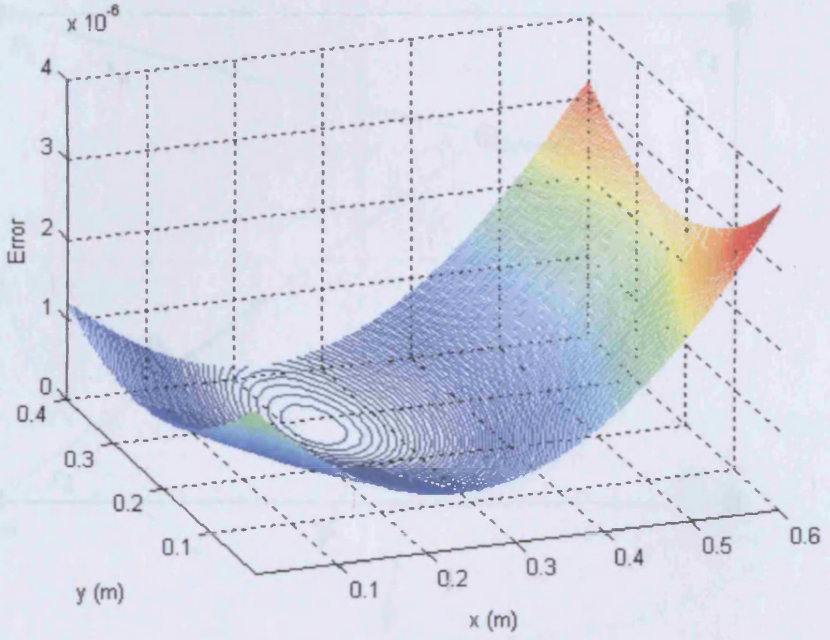
$$\hat{J}_{ML}(q) = \arg\min_q \left(\sum_{m=1}^{M} [\hat{\tau}_m - \tau_m(q)]^2\right) \tag{5.22}$$

Figure 5.14 shows the location of the impact and scratch test signals used in the above example computed from (5.22) using the input from (5.10). A significant difference between ML and ELM can be observed by comparing figure 5.13 with figure 5.14 where it is seen that ML does not suffer from side loops but at the cost of sharpness, which means that the ML algorithm provides more stability while ELM algorithm provides higher accuracy. The similarity between the ML maps of the two different signals is due to the dependence of the ML algorithm on the time differences already estimated, not on the signals themselves.

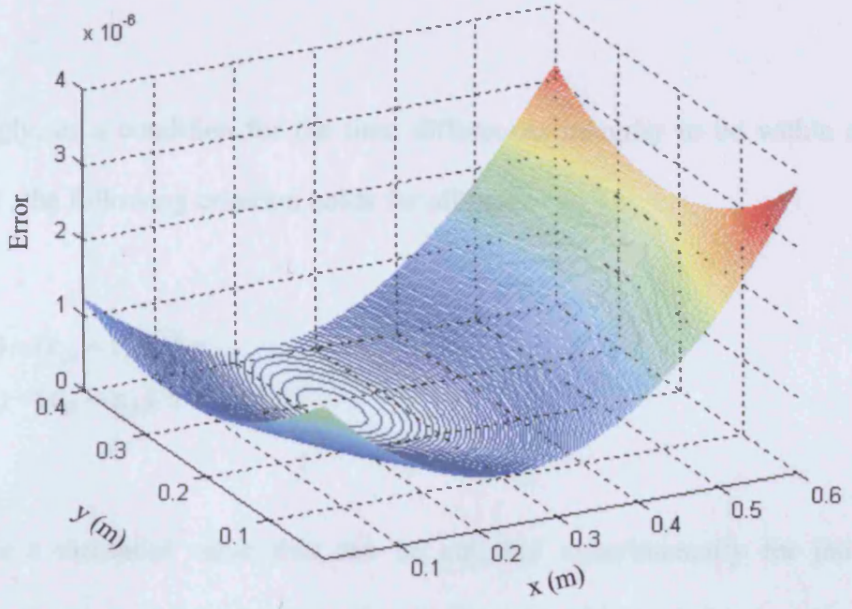
5.2.3 Time Difference Outlier Detection

A large error in any time difference measurement will result in outlier estimation of the source location. To reduce this error a validation criterion is proposed here by checking the integrity of time differences from all pairs of sensors. This criterion can be used as a prior check of time differences before proceeding to estimate the location.

With reference to figure 5.15, let the source be located at distances r_1, r_2, r_3 and r_4 from sensors m_1, m_2, m_3 and m_4 respectively and the associated time difference of arrival between each pair of sensors is given by $\tau_{ij} = (r_i - r_j)/v$, where i,j = 1:4, $i \neq j$. From the geometry of the time differences, both segments $r_a = (r_3 - r_4) - (r_3 - r_1)$ and $r_b = (r_2 - r_4) - (r_2 - r_1)$ resulting from the intersection of the two circles of radius r_1 and r_4 with lines r_3 and r_2 respectively must have the same length of $\tau_{14} v$ since both circles are centered at the same origin.



(a)



(b)

Figure 5.14. Contour plot of ML algorithm for (a) impact signal and (b) scratch signal

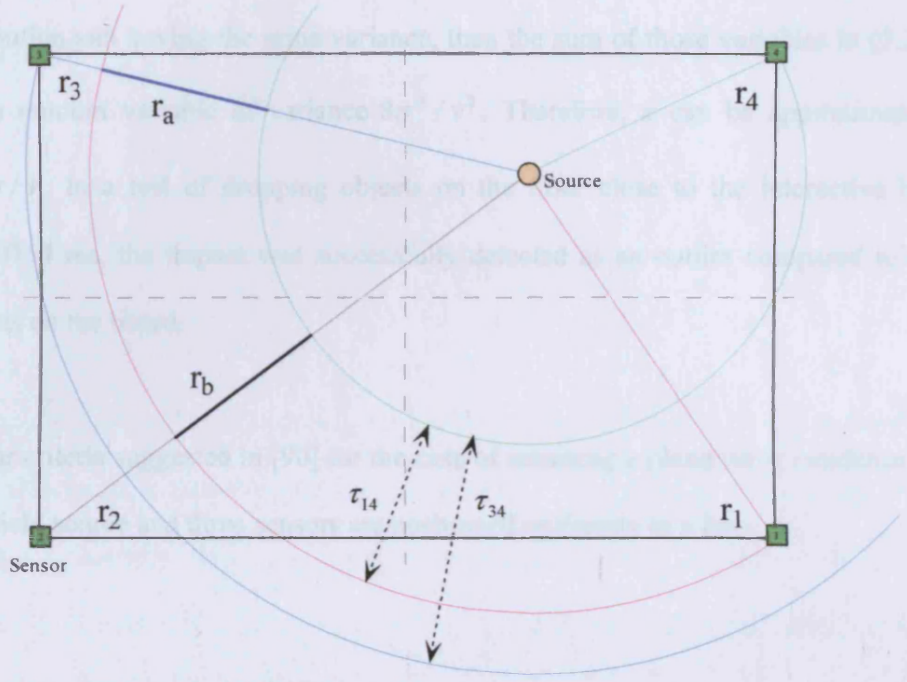


Figure 5.15. Geometry of time differences relationship

Accordingly, as a condition for the time differences integrity to be within an certain error of ε, the following criterion holds for all pairs,

$$\begin{aligned} \left|(\tau_{34}+\tau_{13})-(\tau_{24}+\tau_{12})\right| &\le \varepsilon \\ \left|(\tau_{24}-\tau_{23})-(\tau_{14}-\tau_{13})\right| &\le \varepsilon \end{aligned} \qquad (5.23)$$

Here ε is a threshold value that can be adjusted experimentally for the required sensitivity. For an initial assumption of ε, let the acceptable deviation in each of x and y directions from the source location be σ, then the associate variance of the time difference is $2\sigma^2 / v^2$. Suppose τ_{ij} are independent random variables of Gaussian

distribution and having the same variance, then the sum of those variables in (5.23), is also a random variable of variance $8\sigma^2 / v^2$. Therefore, ε can be approximated by $2\sqrt{2}\sigma / v$. In a test of dropping objects on the floor close to the interactive board, ε =0.0754 ms, the impact was successfully detected as an outlier compared to sharp impacts on the board.

Similar criteria suggested in [90] for the case of assuming a plane wave incidence from a far-field source and three sensors are positioned uniformly in a line.

Chapter 6

Enhanced Continuous Tracking

The localisation approaches developed in chapter 5 satisfy the TAI requirements and the positioning of a scratch type of signal as a snapshot of continuous tracking has been verified successfully in addition to the impact type of signals. Therefore, both ELM and ML techniques are qualified for developing a continuous tracking algorithm but the use of a Hilbert envelope with ELM increases the computation cost.

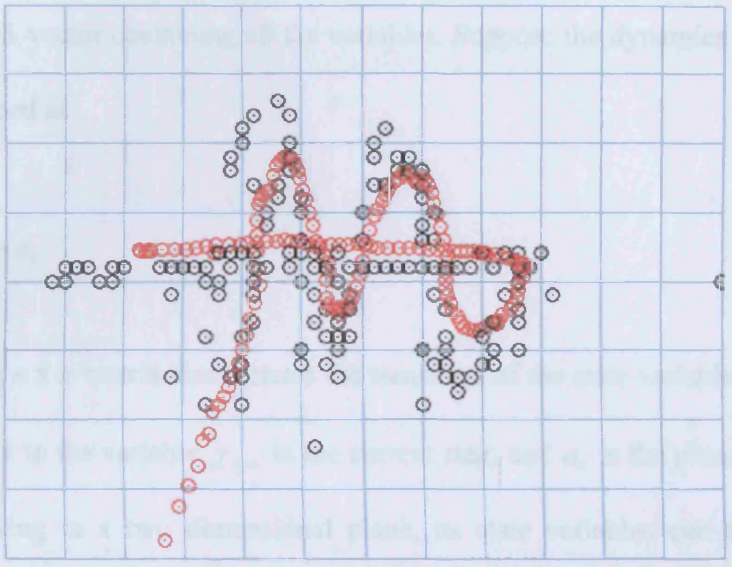
Initially, source tracking can be achieved by iterating the location estimation of consequent sets of time frames acquired by the sensors, but the error produced by the positioning algorithm appears as noise in the estimated trajectory. The computation cost also becomes a significant factor to consider. Because the amplitude of the signals produced by continuous contact movement varies rapidly, power threshold is used instead of amplitude threshold as previously used with an impact signal. The use of power threshold is important for skipping localising noise or the very weak signals that usually appear within the continuous tracking. However, this has the consequence of omitting some events. There are three issues to resolve with continuous tracking: filtering of noise in the trajectory estimation, also called spatial filtering; improving the accuracy and robustness of positioning; making it fast enough to cope with a large

interactive area. The task in this chapter is to develop a continuous tracking algorithm that can handle these three issues efficiently with attention to practical viability.

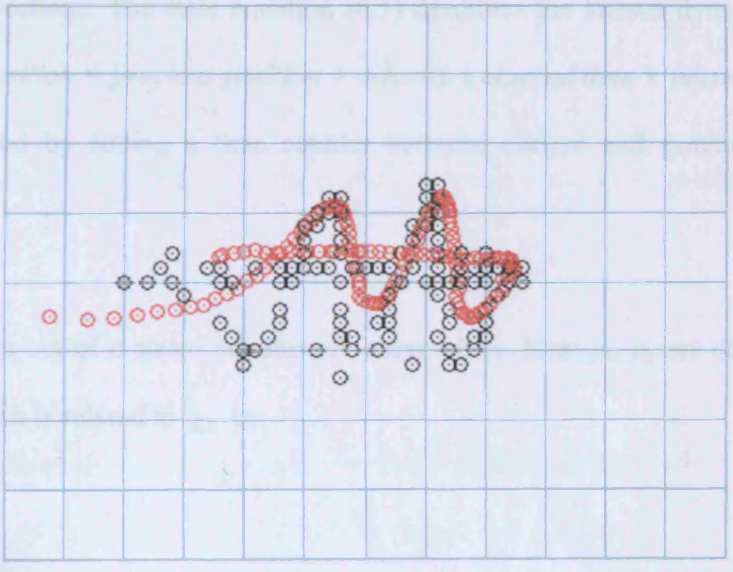
6.1 Trajectory Prediction with Kalman Filter

The Kalman filter is an efficient recursive filter that estimates the state of a dynamic system from a series of incomplete and noisy measurements. Thus it is an ideal option for smoothing the estimated trajectory from noise and missing events.

An example of the noise produced in the trajectory estimation of a continuous moving source using ELM and ML positioning algorithms is shown by the black circles in figure 6.1, where some of the estimated locations are scattered around the actual movement path of a Sine wave and horizontal line created by a nail scratching. Obviously, curve fitting is not an option to smooth the trajectory in real time because only historical information is available and outlier detection is not appropriate since it has a fixed threshold and this results in missing data. The Kalman filter can significantly improve the trajectory estimation by predicting the location based on previous behaviour and current measurement. The theoretical derivation of the Kalman filter is available in textbooks on estimation theory such as [91]. There are various formulas for developing the Kalman filter, such as the g-filter and h-filter presented for tracking from a physics point of view as clearly given in [92]. Here, Kalman filter equations in a compact and efficient matrix form have been implemented for the development of TAI [93].



(a)



(b)

Figure 6.1. Tracking continuous nail scratch using (a) ELM_t and (b) ML algorithms in black. The plot in red is for the same sketches with a Kalman filter applied.

Let the system dynamics of a moving source be described by n system variables and let χ_k be an n x 1 vector containing all the variables. Suppose the dynamics of the system can be described as

$$\chi_{k+1} = \varphi_k \chi_k + a_k \tag{6.1}$$

where φ_k is a n x n matrix that dictates the transition of the state variable χ_k from the previous step k to the variable χ_{k+1} in the current step, and a_k is the process noise. For a source moving in a two dimensional plane, its state variables can be defined as $\chi = (p_x, p_y, v_x, v_y)$, where p_x, p_y are the position coordinates and v_x, v_y are the velocity in x and y directions. The state equation (6.1) describes the known dynamic equation stating *new position = previous position + velocity* x *elapsed time + noise*. The elapsed time is updated by setting a time counter between current and previous measured locations.

Suppose that m out of n state variables are observables. Here p_x, p_y are denoted by the vector z_k which is related to χ_k by

$$z_k = H_k \chi_k + b_k \tag{6.2}$$

where H_k is an m x n matrix that extracts observable variables from the state variables. b_k is the measurement noise. Both a_k and b_k in equations (6.1) and (6.2) are assumed white Gaussian with covariance A_k and B_k respectively. If the initial estimation of the

state variables is $\hat{\chi}_k^-$, the error of this estimate can be found from $e_k^- = \chi_k - \hat{\chi}_k^-$ and the covariance of this error is defined by P_k^-, where '-' means it is a prior estimate.

With Kalman filtering, the measurements of χ can be improved by incorporating z_k in the following relation

$$\hat{\chi}_k = \hat{\chi}_k^- + K_k(z_k - H_k\hat{\chi}_k^-) \tag{6.3}$$

where K_k is an n x m matrix called the Kalman gain yet to be determined from

$$K_k = P_k^- H_k^T (H_k P_k^- H_k^T + B_k)^{-1} \tag{6.4}$$

Having calculated the Kalman gain, the predicted covariance can be obtained from

$$P_k = (I - K_k H_k) P_k^- \tag{6.5}$$

where I is the identity matrix, and both corrected state and corrected error covariance to be projected in the next step as a prior estimate are found from

$$\chi_{k+1}^- = \varphi_k \chi_k \tag{6.6}$$

$$P_{k+1}^- = \varphi_k P_k \varphi_k^T + A_k \tag{6.7}$$

Equations (6.2) to (6.5) represent the Kalman filter. The implementation algorithm of the Kalman filter in this work is illustrated in a flowchart as shown in figure 6.2. The state variables here are the source position coordinates and the measurement inputs are the output of the localisation algorithm. The initial velocity state variable is assumed to be 0.2 m/s, the natural speed of a hand movement, which is reasonable compared to the technical tracking speed of a typical PC mouse. The initial position state variable can be assumed as arbitrary but can be made automatic as given in the following section. It is worth mentioning here that if the source movement velocity (not the wave propagation velocity) can be measured, then it can be updated within equation (6.2). Using the difference between estimated consequent locations divided by the related observable time is an option but is not a real measure of velocity since it includes the computation velocity. Practically, it is noticed that a small variation in the velocity state variable has no effect. At this point, to illustrate how multi-modal sensor fusion can be attained with a Kalman filter, direct velocity measurement from video camera or accelerometer can be used here to provide an actual velocity update by modifying equation (6.2).

It may be observed from figure 6.1 that tracking with ELM results in higher deviation than ML but has better accuracy in following the actual trajectory. This is expected as seen from the analysis in chapter 5. After applying Kalman filtering to the tracking example using ML and ELM, the trajectory is significantly improved as shown by the smooth red circles in figure 6.1.

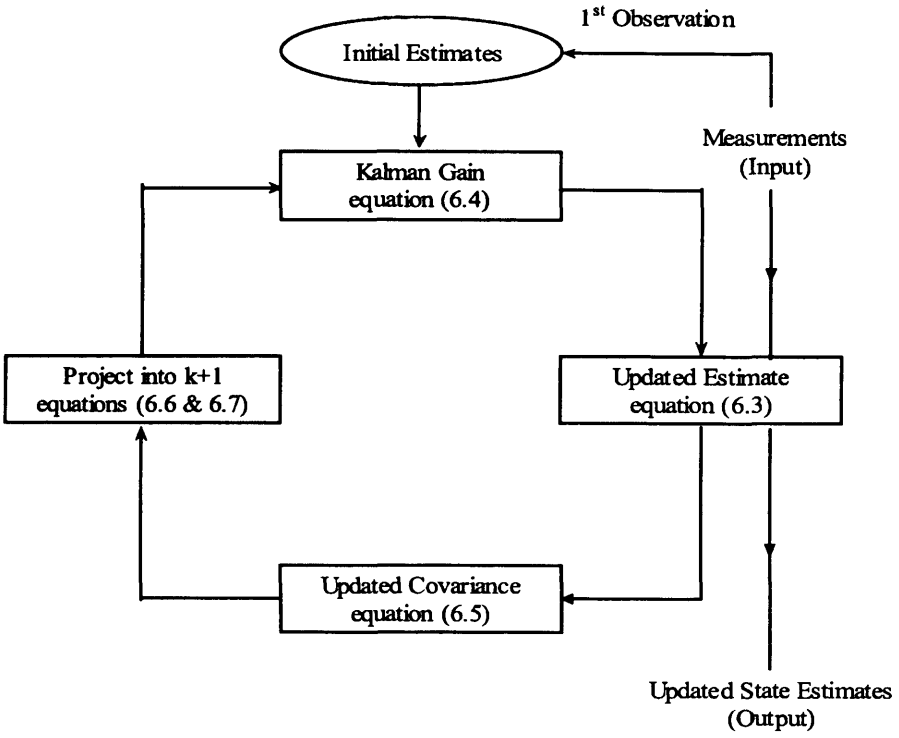


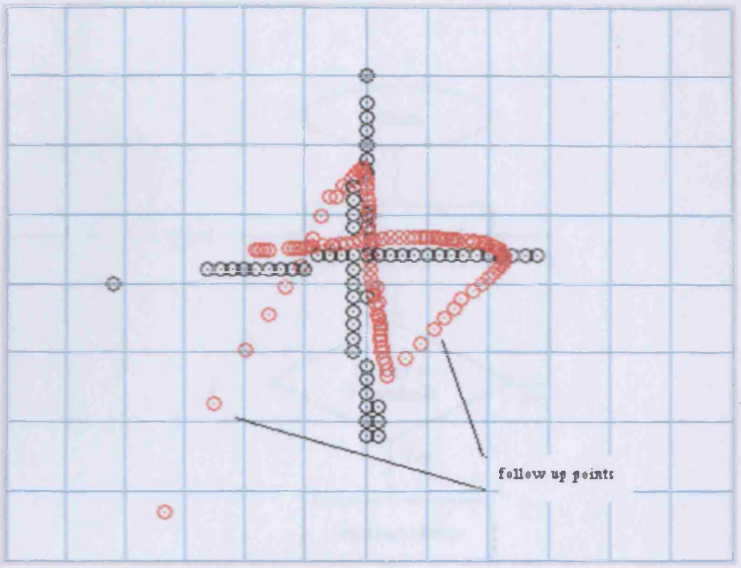
Figure 6.2. Kalman Filter Recursive Algorithm

6.1.1 Automatic Filter Initialisation

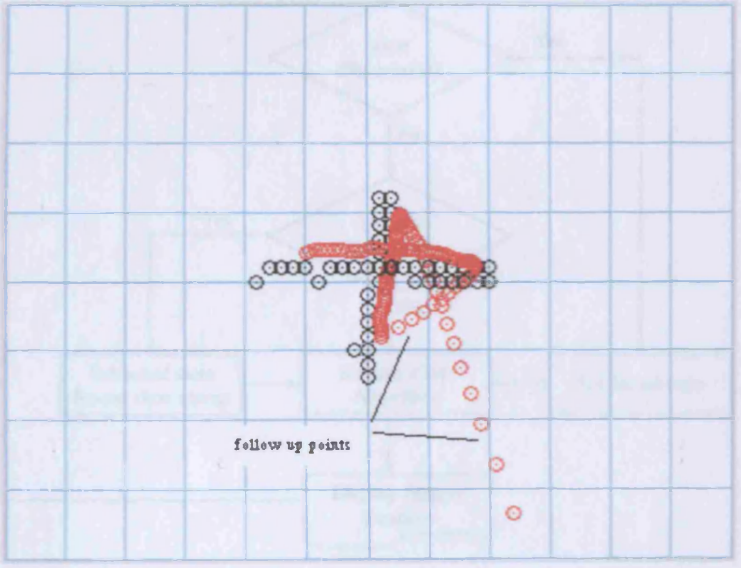
An essential requirement for Kalman filter prediction is to have an initial state value. This causes a practical problem in TAI application which appears at the beginning of each individual trajectory. The prediction of the first point at the beginning of the interaction uses a predefined initial state value, and the last estimated value at the end of a trajectory becomes the initial state of the first point in the next trajectory prediction. A good example can be seen when starting the interaction from a location different from the initial state value such as drawing a cross with two individual lines. The

consequence is the appearance of unwanted points joining the end of the first line with the beginning of the second line, and also at the beginning of the plot as shown in figure 6.3, note that no such follow-up points appear without Kalman filtering.

This problem can be solved by automatically setting the initial state value of the Kalman filter to the first measured value at the beginning of each individual trajectory, but this solution requires a means of distinguishing between discontinued interactions. One option is to oblige the user to start each interaction with a distinctive impact and the tracking algorithm will monitor the input signal amplitude and uses a threshold level to recognise new interaction. However, this method is not reliable and restricts the user to a certain interaction procedure. Another method proposed here which is transparent to the user is by utilising the transition time naturally required between the interactions, which is basically the time taken to move the hand from the last point of a current interaction to the first point of the next interaction. This can be accomplished by measuring the idle time when there is no activity on the board, using a time stamp at the beginning of the data acquisition, and if this time exceeds the transition time, the following interaction is flagged as new. Subsequently a Kalman filter is initialised by the first estimated value obtained from the localisation algorithm and the same for each new interaction and a similar initialisation is also made at the beginning of the first point. This process is illustrated in the flowchart shown in figure 6.4. For example, by setting the transition time at 800 ms, individual sketches could be plotted conveniently using the proposed solution, such as the nail sketching of 'X' shown in figure 6.5. It is clear that the follow-up points in figure 6.3 have disappeared in figure 6.5.



(a)



(b)

Figure 6.3. Undesired estimations appear as a result of Kalman filter initialisation at the beginning and at discontinuity '+' generated with (a) ELM and (b) ML algorithms.

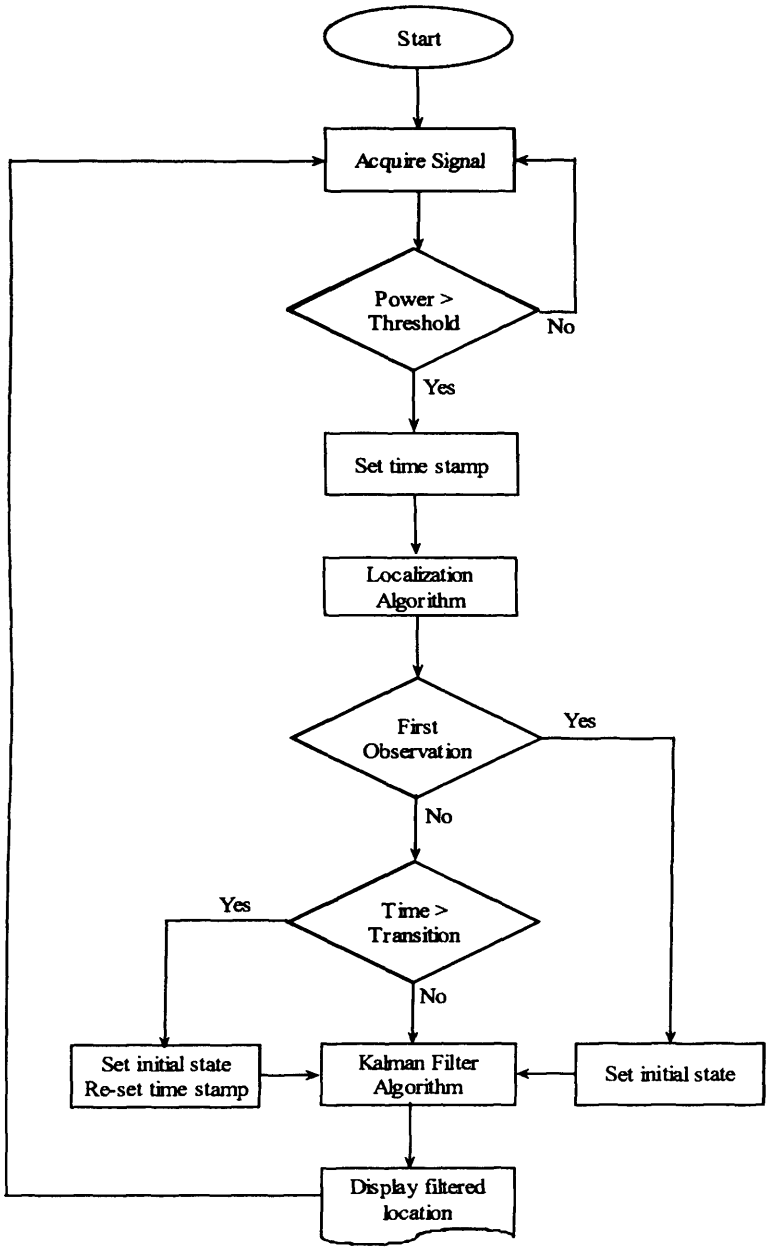
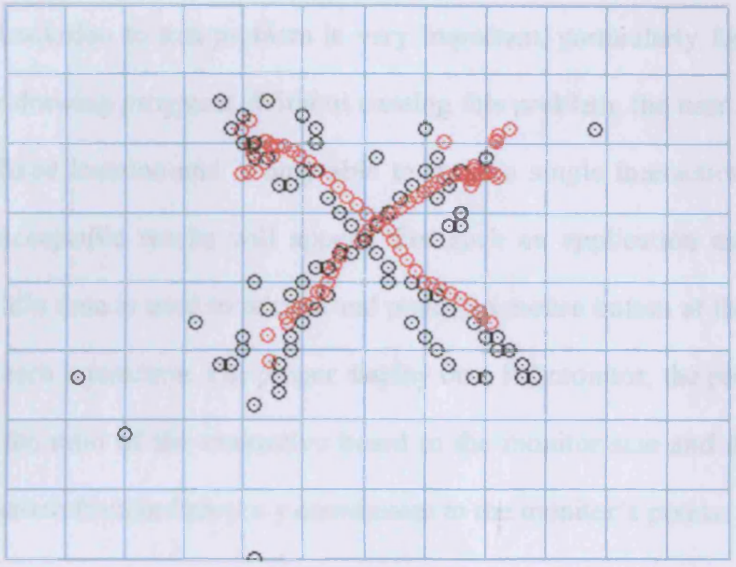
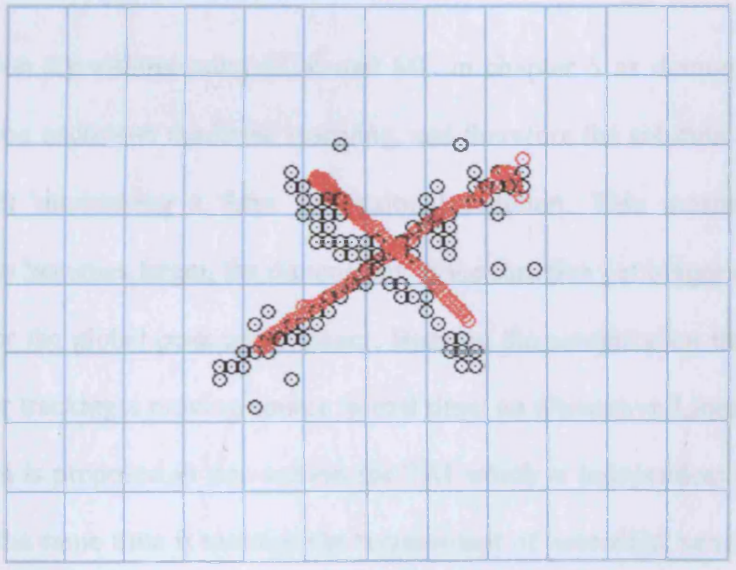


Figure 6.4. Correction flowchart to solve the problem of unwanted estimations produced by a Kalman filter



(a)



(b)

Figure 6.5. Undesired estimations at the beginning and at discontinuity 'X' generated with (a) ELM and (b) ML algorithms disappeared with the proposed correction to a Kalman filter initialisation.

The proposed solution to this problem is very important, particularly for applications like computer drawing programs. Without treating this problem, the user has always to start from a fixed location and is only able to make a single interaction per session, otherwise unacceptable results will appear. For such an application another control based on the idle time is used to release and press the mouse button at the end and the beginning of each interaction. For proper display on a PC monitor, the produced figure is resised by the ratio of the interactive board to the monitor size and the movement steps are converted from ordinary x-y coordinates to the monitor's pixels.

6.2 Linear Positioning Algorithm

The localisation algorithms using ELM and ML in chapter 5 as demonstrated in the previous section encounter nonlinear mapping, and therefore the solution is attained by maximising or minimising a three dimensional function. This means that as the interactive area becomes larger, the dimensions of the function get bigger and hence the search time for the global peak takes longer. Because the computation time is of great importance for tracking a moving source in real time, an alternative Linear Positioning (LP) algorithm is proposed in this section for TAI which is independent of the spatial size while at the same time it satisfies the requirement of near-field, sensor fusion and accuracy.

The proposed algorithm consists of measuring time delays between a number of sensors and a reference sensor which can be obtained using the LCSP algorithm given in chapter 5. Here, with N sensors there are N-1 usable time delays. A closed form

location solution is then found from tangent circles of these time delays instead of from a hyperbolic intersection. The function of sensor fusion is achieved using the Linear Least Square (LLS) technique.

A basic problem with solving the simultaneous hyperbolic equations is that a closed form of algebraic solution is difficult to find, so the solution has to be attained by a numerical search. Another problem with the hyperbola itself is that the point of intersection can move considerably for a relatively small change in eccentricity of one of the hyperbolas. This is unlike circles, defined by a constant distance from each sensor, as an alternative localisation geometry to the traditional hyperbolic intersection as proposed in [94].

Let the source location to be resolved is (x,y), the i^{th} sensor location is (x_i, y_i), the distance between the source and the i^{th} sensor is r_i and the time delay between the i^{th} sensor and the reference sensor is $\tau_{1i} = r_{1i} / v$. Consider sensor 1 as the reference sensor, then with reference to figure 6.6, the circle radius r_1 is given by,

$$(x - x_1)^2 + (y - y_1)^2 = r_1^2 \tag{6.8}$$

and since $r_{1i} = r_i - r_1$, then,

$$\begin{aligned} &(x - x_2)^2 + (y - y_2)^2 = (r_1 + v\tau_{12})^2 \\ &(x - x_3)^2 + (y - y_3)^2 = (r_1 + v\tau_{13})^2 \\ &\cdots \\ &(x - x_N)^2 + (y - y_N)^2 = (r_1 + v\tau_{1N})^2 \end{aligned} \tag{6.9}$$

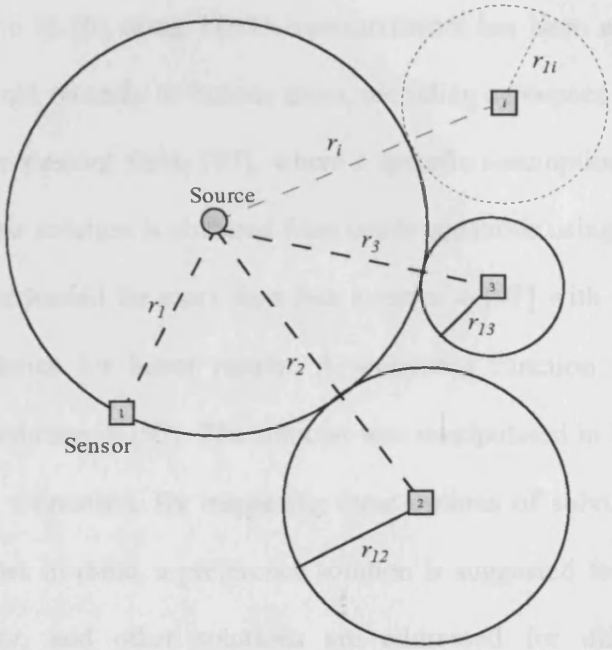


Figure 6.6. TDOA localisation from tangent circles

Substitute 6.8 into 6.9 and rearrange terms to get

$$\begin{aligned}
&(2x_1 - 2x_2)x + (2y_1 - 2y_2)y - 2v\tau_{12}r_1 - v^2\tau_{12}^2 = K_1^2 - K_2^2 \\
&(2x_1 - 2x_3)x + (2y_1 - 2y_3)y - 2v\tau_{13}r_1 - v^2\tau_{13}^2 = K_1^2 - K_3^2 \\
&\cdots \\
&(2x_1 - 2x_N)x + (2y_1 - 2y_N)y - 2v\tau_{1N}r_1 - v^2\tau_{1N}^2 = K_1^2 - K_N^2
\end{aligned} \tag{6.10}$$

where $K_i^2 = x_i^2 + y_i^2$. Note that (6.10) is nonlinear for the unknowns x and y since r_1 appears in the equations. Finding a non-iterative solution for (6.10) has been a challenge for researchers.

Seeking a compact solution for the source localisation problem based on solving similar equations to (6.10) using TDOA measurements has been an area of research since the 1980's until recently in various areas, including aerospace [95], robotics [96] and other signal processing fields [97], where a specific assumption is made for each case. In [94] a linear solution is obtained from circle equations using four sensors only. The solution was extended for more than four sensors in [97] with the introduction of second order statistics for better results. A weighting function for the error was introduced to the solution in [95]. The solution was manipulated in [96] to account for the wave velocity estimation. By inspecting these options of solving (6.10) with the TAI pragmatic sense in mind, a preference solution is suggested for TAI which has a practical feasibility, and other solutions are addressed for different hypotheses depending on the availability of sensors and the method of managing the time delay information to achieve sensor fusion as a goal.

In order to utilise all redundant information with least number of sensors, the terms in equation (6.10) can be rearranged and written in matrix form as,

$$U_1 V_1 = r_1 C_1 + D_1 \tag{6.11}$$

where

$$U_1 = \begin{bmatrix} x_1 - x_2 & y_1 - y_2 \\ x_1 - x_3 & y_1 - y_3 \\ \cdots & \cdots \\ x_1 - x_N & y_1 - y_N \end{bmatrix}, \; V_1 = \begin{bmatrix} x \\ y \end{bmatrix}, \; C_1 = \begin{bmatrix} v\tau_{12} \\ v\tau_{13} \\ \cdots \\ v\tau_{1N} \end{bmatrix}, \; D_1 = \frac{1}{2}\begin{bmatrix} v^2\tau_{12}^2 + K_1^2 - K_2^2 \\ v^2\tau_{13}^2 + K_1^2 - K_3^2 \\ \cdots \\ v^2\tau_{1N}^2 + K_1^2 - K_N^2 \end{bmatrix}$$

With three sensors only, i.e. no redundancy, U_1 is square and (6.11) can be used for an intermediate solution in terms of r_1 as,

$$V_1 = U_1^{-1} r_1 C_1 + U_1^{-1} D_1 \tag{6.12}$$

However, the aim is to achieve sensor fusion from added redundant sensors. In this case (6.12) is not valid since U_1 becomes an over-determined matrix. Note that the right hand side of (6.11) includes the source of error from the noisy TDOA measurements. This is a typical data fitting problem which can be solved using the Linear Least Square optimisation. Assume for instance that r_1 is known, then the problem of the form of $U_1\hat{V}_1 \approx E_1$ can be solved by minimising the sum of squares of $U_1\hat{V}_1 - E_1$, which can be performed by setting the first derivative of the vector inner product to zero as,

$$\frac{\partial}{\partial \hat{V}_1}([U_1\hat{V}_1 - E_1]^T [U_1\hat{V}_1 - E_1]) = 0 \tag{6.13}$$

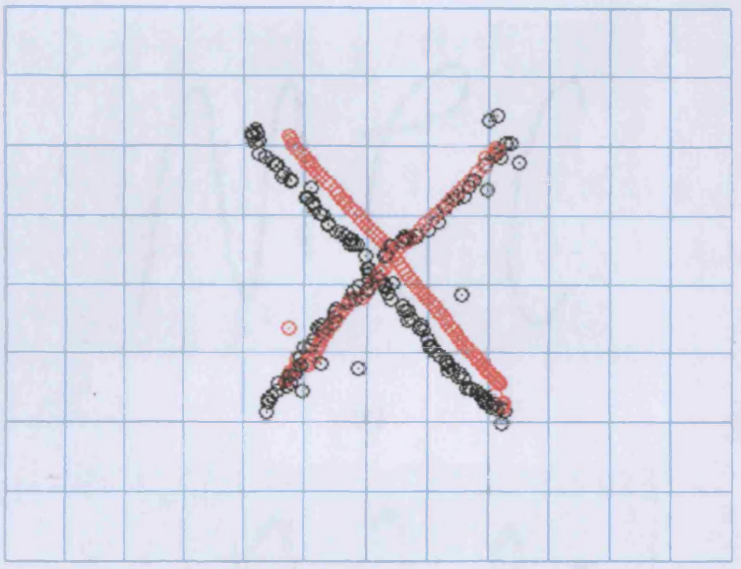
yielding,

$$\hat{V}_1 = (U_1^T U_1)^{-1} U_1^T (r_1 C_1 + D_1) \tag{6.14}$$

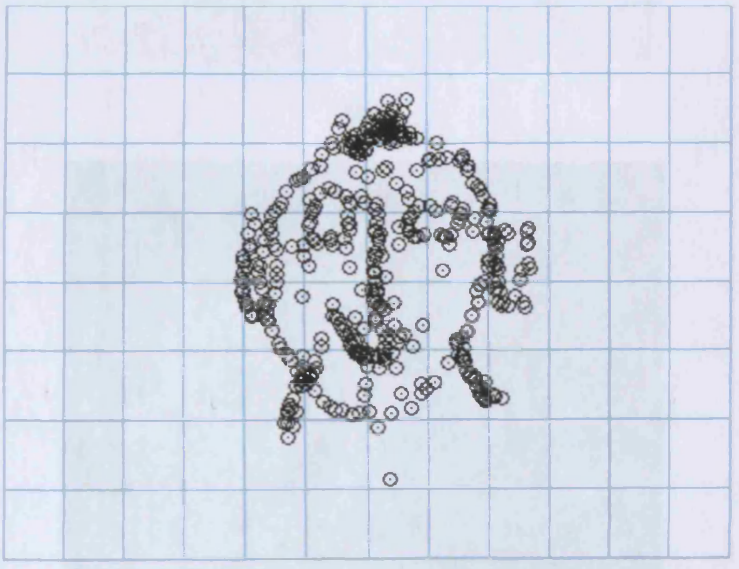
Substituting this intermediate solution (6.14), into equation (6.8), a standard quadratic equation in r_1 is obtained. Then, substituting the positive root back into the above equation yields the final solution of $\hat{V}_1$.

The outcome of this positioning algorithm when used together with the LCSP algorithm for TDOA estimations is impressive. As can be seen in figure 6.7, The sketch of 'X' produced from a nail scratch is better than those shown in figure 6.5 and also shown is a precision sketch of face that is difficult to achieve with ELM and ML algorithms without Kalman filtering. More importantly, the LLS process is independent of the spatial size and thus applicable on large interactive surfaces without additional computation costs, unlike the ML and ELM algorithms. Integrating this algorithm with the Kalman filtering solution given in figure 6.4 and interfacing with the PC mouse to use Microsoft Paint program as an application example, produces the desired tangible acoustic interface for continuous tracking. The result of this final algorithm is demonstrated by the free hand writing of 'mec' with a plastic brush and the Sine wave plot shown in figure 6.8.

A practical test of this proposed continuous tracking algorithm has shown its accuracy and robustness, with satisfying results achieved by interacting with the board using finger tip rubbing, a nail scratch, metal spoon and plastic brush without the need to change the settings or configurations of the system.

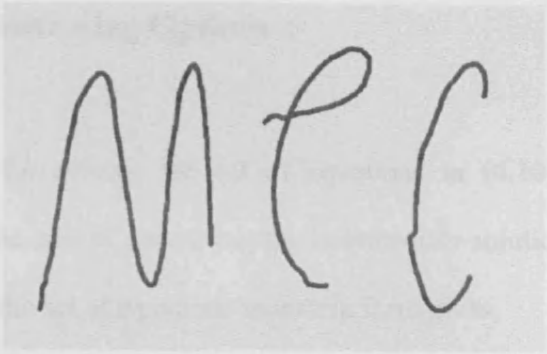


(a)

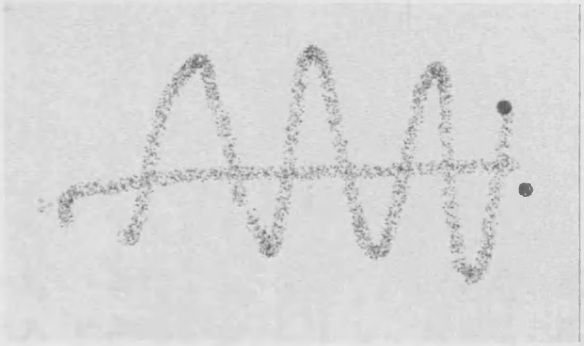


(b)

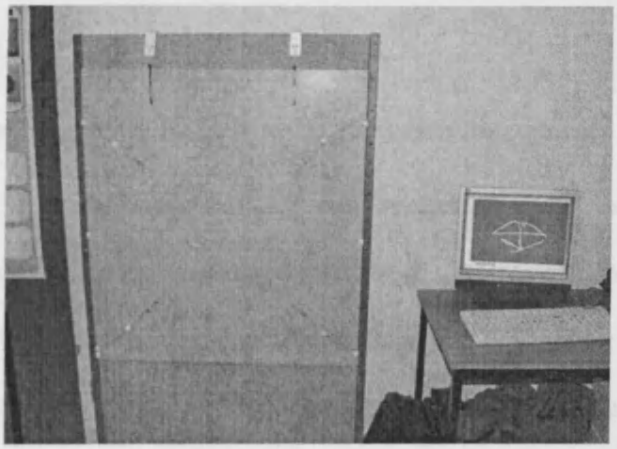
Figure 6.7. Sketching (a) 'X' and (b) 'face' with nail scratch using the proposed algorithm with Kalman filter shown in red for comparison with figure 6.5.



(a)



(b)



(c)

Figure 6.8 Using Paint program as an application interface to the final algorithm. Example of using brush to (a) write mec and (b) draw Sine wave. (c) Live demo in public exhibition.

6.2.1 Extended Positioning Options

Another hypothesis for solving the set of equations in (6.10) is to sacrifice one redundant sensor at the cost of preventing the intermediate solution. Rearranging terms in (6.10) and writing the set of equations in matrix form gives,

$$U_2 V_2 = D_2 \tag{6.15}$$

where

$$U_2 = \begin{bmatrix} x_1 - x_2 & y_1 - y_2 & -2v\tau_{12} \\ x_1 - x_3 & y_1 - y_3 & -2v\tau_{13} \\ \cdots & \cdots & \cdots \\ x_1 - x_N & y_1 - y_N & -2v\tau_{1N} \end{bmatrix}, \; V_2 = \begin{bmatrix} x \\ y \\ r_1 \end{bmatrix}, \; D_2 = \frac{1}{2}\begin{bmatrix} v^2\tau_{12}^2 & K_1^2 - K_2^2 \\ v^2\tau_{13}^2 & K_1^2 - K_3^2 \\ \cdots & \cdots \\ v^2\tau_{1N}^2 & K_1^2 - K_N^2 \end{bmatrix}$$

which results in direct solution from $V_2 = U_2^{-1} D_2$ when four sensors are used. For more than four sensors, the technique of LLS can be implemented as described above. However, although the solution is reached directly, it requires one more sensor than the minimum number required in theory. That means the processing of data from one sensor is not contributing to sensor fusion. On the contrary, a result is duplicated by estimating r_1 and therefore this solution is considered as non efficient.

Another solution of (6.10) is available which is less efficient but with the assumption of having no problem with the availability of sensors and the consequence of handling the

resultant data for the benefit of overcoming unpredictable wave velocity or strong dispersion. This solution can be obtained by rearranging equation (6.10) in matrix form as

$$U_3 V_3 = D_3 \tag{6.16}$$

where

$$U_3 = \begin{bmatrix} x_1 - x_2 & y_1 - y_2 & \tau_{12} & \tau_{12}^2/2 \\ x_1 - x_3 & y_1 - y_3 & \tau_{13} & \tau_{13}^2/2 \\ \cdots & \cdots & \cdots & \cdots \\ x_1 - x_N & y_1 - y_N & \tau_{1N} & \tau_{1N}^2/2 \end{bmatrix}, \quad V_3 = \begin{bmatrix} x \\ y \\ vr_1 \\ v^2 \end{bmatrix}, \quad D_3 = \begin{bmatrix} K_1^2 - K_2^2 \\ K_1^2 - K_3^2 \\ \cdots \\ K_1^2 - K_N^2 \end{bmatrix}$$

Now with a minimum of five sensors, (6.16) can be solved from $V_3 = U_3^{-1} D_3$ but can also be solved for more than five sensors using the LLS technique described. The wave velocity is treated in (6.16) as variable and thus it is estimated within the positioning process. A note on the formulation of (6.16), although it requires a minimum of five sensors without any sensor fusion functionality, it can still be utilised as an efficient tool to measure wave velocity in an object for initial tests. The measured velocity can then be substitutet into the positioning algorithm given in section 6.2, or it can be used to develop a device for measuring wave velocity in materials. Another note regarding the above three formulations for solving (6.10), is that the solution can be extended for estimating source position in three dimensions by adding an additional sensor to the minimum requirement number in each solution. Finally, as a concluding remark, the

solution for the hypothesis of utilising sensor fusion with minimum number of three sensors is the most efficient and most practical compared with the other two hypotheses since most affordable hardware has four channels.

6.2.2 Empirical Error Analysis

The performance of the LLS and ELM localisation algorithms has already been analysed in a 3-D view using real signals in chapter 5 and those have also been compared experimentally with the LLS algorithm for continuous tracking, as shown above. It remains now to provide an empirical comparison between LLS, ELM and ML for single impacts, since a 3-D view for LLS is not valid. This is accomplished by generating M impacts at known locations indexed by the number k on the interactive board. The error in the estimated locations is then analysed for each method, as depicted in figure 6.9.

The actual coordinates of the impact at location k are (x_k, y_k) and the corresponding estimated locations by a certain method from multiple impacts are given by vectors x_{ki} and y_{ki} of size $1xM$. For each location k, the standard deviation of the estimated location from its mean (x_k, y_k) in the x and y directions can be found from,

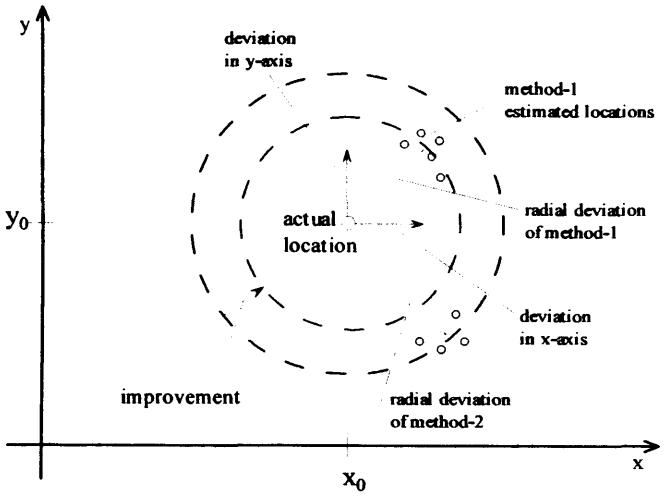


Figure 6.9. Error analysis

$$\sigma_{kx} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(x_{ki} - x_k)^2}$$
$$\sigma_{ky} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(y_{ki} - y_k)^2} \qquad (6.17)$$

Since it is more sensible to consider the radial distance of the estimated location from the actual location as in figure 6.9, the radial standard deviation can be found from equation (6.17) as,

$$\sigma_{kr} = \sqrt{\sigma_{kx}^2 + \sigma_{ky}^2} \qquad (6.18)$$

By applying M=10 consistent and sharp test impacts with nail clicks at each location, the results obtained from equation (6.18) for locations k=1:10 are shown in table 6.1.a. The same test repeated with small nail scratch and the results are shown in figure 6.1.b.

Location k	ELM_t	ELM_t + Hilbert	ELM_f	ML	LP
1	23.022	17.029	3.1623	15.811	3.34
2	18.439	3.1623	6.3246	13.416	2.0992
3	33.015	34.059	31.78	27.019	31.407
4	25.495	7.0711	6.3246	8.9443	3.8629
5	33.5	16.667	3.3333	4.714	21.75
6	31.78	42.544	25.495	31.78	31.202
7	35.355	29.155	28.46	31.78	30.651
8	42.544	40.249	42.544	42.544	42.225
9	36.056	28.284	25.495	28.46	25.901
10	40.139	44.845	26.874	33.333	26.874
Average	**31.9345**	**26.30654**	**19.97928**	**23.78013**	**21.93121**

(a)

Location k	ELM_t	ELM_t + Hilbert	ELM_f	ML	LP
1	9.4868	13.038	69.354	13.038	4.2511
2	11.402	10	42.544	7.0711	7.378
3	34.059	32.249	40.249	25.298	29.426
4	15.811	4.4721	42.544	7.0711	2.7327
5	26.667	24.267	73.106	10	41.092
6	36.878	36.878	72.25	26.077	21.29
7	31.78	33.015	42.544	28.636	26.793
8	56.303	51.865	44.944	49.396	48.292
9	32.558	30	42.544	28.46	24.825
10	118.13	47.14	73.106	51.747	48.796
Average	**37.30748**	**28.29241**	**54.3185**	**24.67942**	**25.48758**

(b)

Table 6.1. Standard deviation of estimated locations for (a) click and (b) scratch test signals

The second and third columns in the table are for ELM_t using pre-filtering and with a Hilbert envelope added. The last three columns are for ELM_f using PHAT, ML and LP algorithms. As the results indicate, the Hilbert envelope improved ELM_t but ELM_f with PHAT has better results than ELM_t for impact type of signals rather than for scratch type. The *LP* algorithm has shown the best results both for impact and scratch signals among the others in addition to the advantage of its computation speed.

Finally, the detection of two simultaneous sources has also been of interest. Theoretically with the ELM algorithm it is possible to detect two impacts by considering the second peak in the outcome of the cross correlation based process. However, the empirical results were not satisfactory. In [98], the second and the third peaks obtained with the GCC method using the PHAT process have been studied for detecting active speakers simultaneously in a room but the results were not sufficiently reliable with successful attempts of about 40%. For the scenario of having two simultaneous sources with different distinctive bandwidths, there might be a good chance to localise them simultaneously by splitting the received signals with two filters before implementing the given algorithms. This case has not been tested but it is an option for active sources in particular. Therefore, for reliable multi-source localisation, there is a need to refine the algorithms and to investigate the feasibility of employing other promising techniques such as Independent Component Analysis as used in [99] to localise multiple sound sources in a room within the proposed algorithms.

CHAPTER 7

Conclusions and Future Work

In this research, tangible acoustic interfaces (TAIs) for human computer interaction have been developed. With the new interfaces it is possible to convert virtually any solid object into an interactive interface in a transparent way by simply attaching sensors to the object's surface to detect the acoustic signals resulting from natural interactions with the object, such as nail clicks. The developed system is capable of responding to two types of interactions, discrete impacts and continuous movement. Various localisation techniques have been thoroughly investigated and the TAI system has been analysed to determine which localisation methods are applicable. Two distinct localisation approaches are proposed. The one that requires configuring each point of interest is not suitable for tracking but can work with one or two sensors with different object materials or shapes. The other method only requires wave velocity and information on the position of the sensors. It is suitable for interacting with arbitrary locations on a flat and reasonably homogeneous surface and suitable for tracking.

The first localisation approach is Location Pattern Matching. The interactions are localised by finding the best matched feature pattern representing a known location. One sensor is adequate for this method, with a higher resolution attained with more sensors, proving the main hypothesis of this research. The reliability problem caused by the sensitivity of the matching process to the pattern of the template has been solved by

the concept of extracting the location signature pattern from received signals using two sensors.

The second localisation approach, widely used for in-air acoustic localisation, is based on measuring the time differences of arrival (TDOA) between spatially separated sensors and uses triangulation to determine the source coordinates. For the efficient operation of TAI with this approach, three requirements have to be met, the assumption of the near-field scenario, the use of an optimisation method and the implementation of sensor fusion. Accordingly, two methods are proposed for impact localisation. The ELM method performs localisation in one step, and two algorithms are proposed. One uses time domain processing with conventional post-filtering and the other employs PHAT filtering in the frequency domain. In the two-step method, TDOA values are found first using either GCC or LCSP. Based on these values, the source is localised using the ML algorithm. The effect of dispersion is treated by introducing Hilbert envelope smoothing. A criterion is proposed for each method to detect outlier estimations that can happen, for example from domestic noise like that made when a door is shut.

Although the above two methods are applicable to tracking, their localisation algorithms are a function of the working area dimensions and the theoretical resolution which in turn can cause significant latency when used for tracking on a large surface. For this reason, an LP localisation algorithm is proposed which is independent of the working area dimensions and uses LCSP for the TDOA inputs. For smooth trajectory prediction of a continuous moving source, Kalman filtering is proposed. A timing-

based algorithm is developed to solve the problem of Kalman filter initialisation, which is important to eliminate the appearance of undesired estimations between consecutive interactions. For a real application, the developed system was interfaced with a mouse pointer. With this system it was possible to draw curved and straight lines on an MDF board with an accuracy of 10 mm using a finger, brush, or wood stick on the Paint program on a PC. All the above algorithms were developed with the sensor fusion concept in mind, which is the trend in sensor applications research. The algorithms were written for an arbitrary number of sensors although in this work four sensors were used due to the limitations of the hardware.

The good results obtained have proved the second hypothesis of this research, namely, that in-air acoustic localisation algorithms can be adopted for use in in-solid applications.

This research has met all the objectives stated in chapter one as follows:

- Objects in different materials (an MDF board and a sheet of glass) were transformed into interactive interfaces.
- Three algorithms have been developed for discrete impact localisation and another algorithm for continuous tracking.
- Filtering techniques were applied to address the problems of wave dispersion and distortion.

For future work, the suggestion is to investigate the integration of LPM with time delay based methods in an attempt to gain the advantages of both approaches. The

localisation of multiple simultaneous sources was not successful and requires more research, possibly with the aid of Independent Component Analysis. Multi-modal sensor fusion is attractive to implement for achieving better accuracy, for example, by introducing velocity information from a dedicated sensor within the parameters of the Kalman filter, or by integrating visual information with the current audio information for more capable applications. Finally, it is worth investigating the extension of these methods for in-air localisation by placing the sensors away from the surface for even freer interaction.

References

[1] www.actkern.info

[2] smarttech.com/DViT

[3] www.elotouch.com/Technologies/ITouch/

[4] www.nextwindow.com/benefits/touchscreen_technology.html

[5] www.pegatech.com

[6] www.mimio.com

[7] www.virtual-laser-keyboard.com

[8] www.sensitiveobject.fr

[9] solutions.3m.com

[10] www.i-vibrations.com

[11] Johanson, B.; Fox, A. and Winograd, T., 'The Interactive Workspaces project: experiences with ubiquitous computing rooms', IEEE Pervasive Computing magazine, Volume 1, Issue 2, April-June 2002, pp 67-74.

[12] Johanson, B.; Winograd, T. and Fox, A., 'Interactive Workspaces', IEEE Computer magazine, Volume 36, Issue 4, April 2003, pp. 99-101.

[13] Vladimir J., Lumelsky, Michael S. and Sigurd Wagner, 'Sensitive Skin', IEEE Sensors Journal, Vol.1, No.1, June 2001, pp. 41-51.

[14] H. Ishii, C. Wisneski, J. Orbanes, B. Chun, and J. Paradiso, 'PingPongPlus: Design of an Athletic-Tangible Interface for Computer-Supported Cooperative Play', Conference on Human Factors in Computing Systems CHI'99, May 1999, pp. 394-401.

[15] N Checka, 'A System of Tracking and Characterizing Acoustic Impacts on Large Interactive Surfaces', Mater thesis, MIT, May 2001.

[16] C. Leo, 'Contact and Free-Gesture Tracking for Large Interactive Surfaces', Master thesis, MIT, June 2002.

[17] J. Paradiso, K. Hsiao, J. Strickon and A. Adler, 'Sensor systems for interactive surfaces', IBM Systems Journal, Vol. 39, Nos. 3&4, 2000, pp. 892-914.

[18] W. Hartmann, 'How we Localize Sound', Physics Today, American Institute of Physics, November 1999, pp.24-29.

[19] Book: S. Haykin, Editor, Array Signal Processing, Prentice-Hall, 1985, ISBN: 013046482-1. tk.5102.5.a7.

[20] H Karim and M Vibreg, 'Two decades of Array signal Processing', IEEE Signal Processing Magazine, Jul. 1996, pp. 67-94.

[21] J.-M. Valin, F. Michaud, B. Hadjou and J. Rouat, 'Localization of Simultaneous Moving Sound Sources for Mobile Robot Using a Frequency-Domain Steered Beamformer Approach'. Proc. IEEE International Conference on Robotics and Automation (ICRA), 2004, pp.1033-1038.

[22] Y. Yoon, Direction-of-arrival Estimation of Wideband Sources Using Sensor Array, PhD thesis, Georgia Institute of Technology, July 2004.

[23] 'Special issue of time-delay estimation', IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-29, No3, June, 1981.

[24] C. Napp et al, 'The Generalized Cross-Correlation for Estimation of time delay', IEEE Trans. On Acoustics, Speech and Signal Processing, Aug. 1976, Vol. 24, No. 2, pp. 320-327.

[25] M. Omologo and P. Savaizer, 'Use of crosspower-spectrum phase in acoustic event location', Correspondence, IEEE Transactions of Speech and Audio Processing, Vol. 5, No. 3, May 1997.

[26] M. Omologo and P. Savaizer, 'Acoustic event localization using crosspower-spectrum phase based technique', Proceedings of the ICASSP, 1994.

[27] J. Benesty, 'Adaptive eigenvalue decomposition algorithm for passive acoustic source localization', Journal of Acoust. Soc. Am, Vol. 107, No.1, Jan. 2000, pp.384-391.

[28] W. Bragg, 'The World of Sound', Dover Publications, 1968.

[29] W. Foy, 'Position-location solutions by Taylor-Serious Estimation', IEEE Transactions on Aerospace and Electronic Systems, Vol. AES-12, No.2, pp. 187-193.

[30] H. Smith and J Abel, 'Closed form least-square source location estimation from range-difference measurements', IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-35, No. 12, Dec. 1987, pp. 1661-1669.

[31] M. Brandstein, J. Adock and H Silverman, 'A closed form method for finding source location from microphone-array time-delay estimates', Proc. ICASSP, 1995, pp. 3019-3022.

[32] M. Brandstein and H Silverman, 'A practical methodology for speech source localization with microphone arrays', Academic Press Ltd, Computer Speech and Language, Vol. 11, 1997, pp. 91-126.

[33] R. Shenoy, 'Design of e-textiles for acoustic applications', Master thesis, Virginia Polytechnic Institute and State University, 2003.

[34] P Julian et al, 'A comparative study of Sound Localization Algorithms for Energy Aware Sensor Network Nodes', IEEE Transactions on Circuits and Systems, Vol. 51, No. 4, Apr. 2004, pp. 640-648.

[35] E. Mumolo et al, 'Algorithms for Acoustic Localization Based on Microphone Array in Service Robotics', Elsevier, Robotics and Autonomous Systems vol. 42, 2003, pp.69-88.

[36] Y Huang et al, 'Passive acoustic source localization for video camera steering', Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2000, Vol.2, pp. II909 - II912.

[37] S Ziola and M Gorman, 'Source location in thin plates using cross-correlation', J. Acoustic Soc. Amer. Vol. 90, No. 5, Nov. 1991, pp 2551-2556.

[38] S. Birchfield and D. Gillmor, 'Acoustic source direction by hemisphere sampling', Proc. Of ICASSP, 2001.

[39] B. Mungamuru and P Aarabi, 'Enhanced sound source localization', IEEE Trans. On Systems, Man and Cybernetics, Vol. 34, No.3, June 2004.

[40] D. Torney and R. Nemzek, 'Least-error localization of discrete acoustic source', Elsevier, Applied acoustics, Vo. 66, 2005, pp. 1262-1277.

[41] A Said, A. Tarighan and N Khajehnouri,'Network-based wireless localization ', IEEE Signal Processing magazine, Jul. 2005, pp. 24-40.

[42] K. Dogancay and A. Hashemi-Sakhtsari, 'Target tracking by time difference of arrival using recursive smoothin', Elsevier, Signal Processing, Vol. 85, 2005, pp. 667-679.

[43] J.-M. Valin, F. Michaud and J. Rouat, 'Robust Localization and Tracking of Simultaneous Moving Sound Sources Using Beamforming and Particle Filtering', Robotics and Autonomous Systems Journal (Elsevier), Vol. 55, No. 3, 2007, pp. 216-228.

[44] D. Ward E. Lehmann and R. Williamson, 'Particle filtering algorithms for tracking an acoustic source in a reverberant environment', IEEE Transactions on Speech and Audio Processing, Volume 11, Issue 6, Nov. 2003, pp. 826 – 836.

[45] Martin Hellebrandt and Rudolf Mathar, 'Location Tracking of Mobiles in Cellular Radio Networks', IEEE Transactions on Vehicular technology, Vol. 48, No.5, Sep. 1999.

[46] J. Holland, 'Acoustic wave spectrum', American Sound Press, 1997, http://www.johnholland.ws/home/acousticwave.

[47] http://www.geo.mtu.edu/

[48] Modes of Sound Wave Propagation, http://www.ndt-ed.org

[49] B. Tucker, 'Ultrasonic plate waves in wood-based composite panels', PhD thesis, Aug. 2001, Washington State University.

[50] H. Jeong and Y. Jang, 'Wavelet analysis of plate wave propagation in composite laminates', Elsevier, Composite Structure, Vol. 49, 2000, pp. 443-450.

[51] R. Ing, N. Quieffin, S. Catheline and M. Fink, 'Tangible interactive interface using acoustic time reversal process', Applied Physics Letters, 2004.

[52] G. Lai and P Aarabi, 'Active object localization using speaker arrays', Proceedings of the IEEE Sixth International Conference on Information Fusion, Vol.1, 2003, pp. 70-73.

[53] E Dijk, 'Indoor ultrasonic position estimation using a single base station', PhD thesis, Eindhoven University of Technology, Netherlands, 2004.

[54] Bousseljot R and Kreiseler D. Waveform Recognition with 10,000 ECGs. IEEE Computers in Cardiology proceedings, 24-27 Sept 2000, Cambridge, MA, 331-334.

[55] O'Hagan R and Zelinsky A. Finger Track – A Robust and Real-Time Gesture Interface. Advanced Topics in Artificial Intelligence, Tenth Australian Joint Conference on Artificial Intelligence Proceedings, 475-484, Dec. 1997.

[56] L. Kinsler, A. Frey, A. Coppens and J. Sanders, 'Fundamentals of Acoustics', 4th ed., 2000, John Wiley & Sons.

[57] M. Fink and C. Prada, 'Acoustic time-reversal mirror', Institute of Physics Publications, Inverse Problems, Vol.17, 2001, pp. R1-R38.

[58] M. Fink, 'Time reversed acoustics', Scientific America, Nov. 1999, pp. 91-95.

[59] S. Jesus, M. Porter, Y. Stéphan, X. Démoulin, O. Rodríguez and E. Coelho, 'Single hydrophone source localization', IEEE Journal of Oceanic Engineering, Vol. 25, No. 3, Jul. 2000, pp. 337-346.

[60] Khong, A W H, Naylor, P A. Stereophonic Acoustic Echo Cancellation Employing Selective-Tap Adaptive Algorithms. IEEE Transactions on Audio, Speech, and Language Processing, Vol. 14, No. 3, May 2006, pp. 785-796

[61] E. Grassi, J. Tulsi and S. Shamma, 'Measurement of Head-related Transfer Function based on the Empirical Transfer Function Estimate', Proceedings of the 2003 International Conference on Auditory Displays, Boston, USA, Jul. 2003, pp. 119-122.

[62] X. Wang, Z. Wang and B. O'Dea, ,'A TOA-based location algorithm due to NLOS propagation', IEEE Trans. Veh. Technol. Vol. 52, No.1, Jan 2003, pp. 112-116.

[63] L. Kinsler, A. Frey, Alan Coppens and J. Sanders, Fundamentals of Acoustics, John Wiley & Sons, 4th edition, 2000.

[64] C. Tan, M. Beach and A. Nix, 'Problems with direction finding using linear array with element spacing more than half wavelength', 1st Annual COST 273 Workshop, Espoo, Finland, 29-30 May 2002.

[65] A. Tobias, 'Acoustic-emission source location in two dimensions by an array of three sensors ', Non-destructive testing, IPC Science and Technology Press, Vol.9, No.1, Feb. 1976, pp. 9-12.

[66] C. Carter, 'Time delay estimation for passive sonar signal processing', IEEE Transactions on Acoustics, Speech and signal processing, Vol. ASSP-29, No.3, Jun. 1981, pp. 463-470.

[67] H. Canistraro and E. Jordan, 'Projectile-impact-location determination: an acoustic triangulation method', Meas. Sci. Technol. Vol. 7, 1996, pp. 1755-1760.

[68] C. Knapp and G. Carter, 'The generalized correlation method for estimation of time delay', IEEE Trans. Acoust. Speech and Signal Processs., Vol. ASSP-24, 1976, pp. 320-327.

[69] P. Julian, A. Andreou, L. Riddle, S. Shamma, D. Goldberg and G. Cauwenberghs, 'A comparative study of sound localization algorithms for energy aware sensor network nodes', IEEE Trans. On Circuits and Systems, Vol. 51, No. 4, Apr. 2004, pp. 640-648.

[70] M. Omologo and P. Svaizer, 'Acoustic localization in noisy and reverberant environment using CSP analysis', Proceedings of the IEEE International

Conference on Acoustics, Speech, and Signal Processing conference, ICASSP-96, Volume 2, Issue , 7-10 May 1996, pp. 921 – 924.

[71] Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS, 27-31 Oct. 2003, Vol. 2, pp. 1228- 1233.

[72] S. Birchfield and R. Gangishetty, 'Acoustic localization by interaural level difference', IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP, Philadelphia, Pennsylvania, 18-23 Mar. 2005, Vol.4, pp.1109-1112.

[73] J. Chen, K. Yao and R. Hudson, 'Acoustic Source Localization and Beamforming: Theory and Practice', EURASIP Journal on Applied Signal Processing, 2003 Hindawi Publishing Corporation, 2003, pp.359–370.

[74] Yamada, S. Nakamura and K. Shikano, 'Robust speech recognition with speaker localization by a microphone array', Proc. of the Int. Conference on Spoken Language Processing, ICSLP, 3-6 Oct. 1996, Philadelphia, PA, USA.

[75] L. Cohen, 'The uncertainty principle in signal analysis', Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, 25-28 Oct 1994, pp. 182 – 185.

[76] S. Mallat, 'A Wavelet tour of signal processing', 2nd ed. Academic press, 1999.

[77] S. Wu, H. So and P. Ching, 'Improvement of TDOA measurement using wavelet denoising with a novel thresholding technique', IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP, 21-24 Apr. 1997, Vol.1, pp. 539-542.

[78] L. Gaul and S. Hurlebaus, 'Identification of the impact location on a plate using wavelets', Mechanical Systems and Signal Processing Vol. 12, No. 6, 1997, pp. 783-795.

[79] R. Barsanti, 'Wavelet-based time delay estimates for transient signals', Thirty-Seventh Asilomar Conference on Signals, Systems and Computers, Vol. 1, 9-12 Nov. 2003, pp. 1173 – 1177.

[80] S. Griebel, ''A microphone array system for speech source localization, denoising and dereverberation', PhD thesis, Harvard University, Massachusetts, Apr. 2002.

[81] M. Brandstein E. Adcock F. Silverman, ,A localization-error-based method for microphone-array design', This paper appears in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP, Atlanta, USA, 7-10 May 1996, Vol. 2, pp.901-904.

[82] V. Duraiswami and Y. Prasanna, 'Tracking a moving speaker using excitation source information', 8th European Conference on Speech Communication and Technology, Geneva, Switzerland, 1-4 Sept. 2003, pp. 69-72.

[83] S. Birchfield and D. Gillmor, 'Fast Bayesian Acoustic Localization', Proceedings of the IEEE International Conference on Speech and Signal Processing, ICASSP, Florida, Vol.2, May 2002, Volume 2, pp. 1793 – 1796.

[84] P. Aarabi and S. Zaky, 'Robust sound localization using multi-source audiovisual information fusion', Elsevier, Information Fusion, Vol.2, 2001, pp. 209-223.

[85] J. Bendat and A. Piersol, 'Random data', John Wiley & Sons, 1986.

[86] A. Mertins, 'Signal analysis', John Wiley & Sons, 1999.

[87] L. Danfeng and S. Levinson, 'A linear phase unwrapping method for binaural sound source localization on a robot', Proceedings of the IEEE Conference on Robotics and Automation, Washington, May. 2002, pp. 19-23.

[88] H. Poor, 'An introduction to signal detection and estimation', Springer, 2nd ed. 1994.

[89] N. Gershenfeld, 'The nature of mathematical modeling', Cambridge University Press, 1999.

[90] J Hietanen, K Koppinen and J Astola, ,'Time-delay selection for robust angle of arrival estimation', Proceedings of the IASTED International Conference on Signal and Image Processing, SIP, Bahamas , Oct. 1999, pp.18-21.

[91] A. Gelb, editor, 'Applied optimal estimation', MIT press, 2nd ed., 1975.

[92] E. Brookner, 'Tracking and Kalman filtering made easy', John Wiley & Sons, 1998.

[93] D. Jwo, 'Remarks on the Kalman filtering simulation and verification', Elsevier, Applied Mathematics and Computation, Vol. 186, 2007, pp. 159-174.

[94] H. Schau and A Robinson, 'Passive source localization employing intersecting spherical surfaces from time-of-arrival differences', IEEE Transactions on Acoustic, Speech, and Signal Processing, Vol. ASSP-35, No.8, Aug. 1987, pp. 1223-1225.

[95] G. Mellen et al, 'Closed-form solution for determining emitter location using time difference of arrival measurements', IEEE Transactions on Aerospace and Electronic Systems, Vol. 39, No.3, Jul. 2003, pp. 1056-1058.

[96] A. Mahajan and M. Walworth, '3-D position sensing using the differences in the time-of-flights from a wave source to various receivers', IEEE Transactions on Robotics and Automation, Vol. 17, No.1, Feb. 2001, pp. 91-94.

[97] Y. Chan and K Ho, 'A simple and effective estimator for hyperbolic location', IEEE Transactions on Signal Processing, Vol.42, No. 8, Aug. 1994, pp. 1905-1915.

[98] D. Bechler and K. Kroschel, 'Considering the second peak in the GCC function for multi-source TDOA estimation with a microphone array', International workshop on Acoustic Echo and Noise Control, IWAENC, Sep. 2003, Tokyo, Japan, pp. 315-318.

[99] H. Buchner, R. Aichner, J. Stenglein, H. Teutsch, W. Kellennann, 'Simultaneous localization of multiple sound sources using blind adaptive MIMO filtering', Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP, 18-23 Mar. 2005, Vol.3, pp. iii/97-iii/100.