BINDING SERVICES
Tel +44 (0)29 2087 4949
Fax.+44 (0)29 2037 1921
E-Mail Bindery@Cardiff.ac.uk

Development of Tangible Acoustic Interfaces for Human Computer Interaction

A thesis submitted to Cardiff University
in candidature for the degree of

Doctor of Philosophy

by

Ze Ji B.Eng, 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

Abstract

Tangible interfaces, such as keyboards, mice, touch pads, and touch screens, are widely used in human computer interaction. A common disadvantage with these devices is the presence of mechanical or electronic devices at the point of interaction with the interface. The aim of this work has been to investigate and develop new tangible interfaces that can be adapted to virtually any surface, by acquiring and studying the acoustic vibrations produced by the interaction of the user's finger on the surface. Various approaches have been investigated in this work, including the popular time difference of arrival (TDOA) method, time-frequency analysis of dispersive velocities, the time reversal method, and continuous object tracking.

The received signal due to a tap at a source position can be considered the impulse response function of the wave propagation between the source and the receiver. With the time reversal theory, the signals induced by impacts from one position contain the unique and consistent information that forms its signature. A pattern matching method, named Location Template Matching (LTM), has been developed to identify the signature of the received signals from different individual positions. Various experiments have been performed for different purposes, such as consistency testing, acquisition configuration, and accuracy of recognition. Eventually, this can be used to implement HCI applications on any arbitrary surfaces, including those of 3D objects and inhomogeneous materials.

The resolution with the LTM method has been studied by different experiments, investigating factors such as optimal sensor configurations and the limi-

tation of materials. On plates of the same material, the thickness is the essential determinant of resolution. With the knowledge of resolution for one material, a simple but faster search method becomes feasible to reduce the computation. Multiple simultaneous impacts are also recognisable in certain cases.

The TDOA method has also been evaluated with two conventional approaches. Taking into account the dispersive properties of the vibration propagation in plates, time-frequency analysis, with continuous wavelet transformation, has been employed for the accurate localising of dispersive signals. In addition, a statistical estimation of maximum likelihood has been developed to improve the accuracy and reliability of acoustic localisation. A method to measure and verify the dispersive velocities has also been introduced.

To enable the commonly required "drag & drop" function in the operation of graphical user interface (GUI) software, the tracking of a finger scratching on a surface needs to be implemented. To minimise the tracking error, a priori knowledge of previous measurements of source locations is needed to linearise the state model that enables prediction of the location of the contact point and the direction of movement. An adaptive Kalman filter has been used for this purpose.

Acknowlegments[1]

I would like to express my deepest gratitude and respect to my supervisor, Prof. D. T. Pham, for his invaluable guidance, advice and support throughout my studies. The help and understanding that he has shown are gratefully acknowledged.

My thanks are also due to all those members of the laboratory who contributed to my work through their valuable help, advice and technical support. In particular, I would like to mention Dr. M. Yang, Mr. M. Al-kutubi and Dr. Z. Wang for all that I learnt from them during our many long technical discussions. Especially, I would like to thank Dr. M. Yang for his help in the electronics hardware design. Their contribution is acknowledged.

I am indebted to all those colleagues, whose friendship and support I valued more than any technical help. I would also like to thank those French visiting students for their contributions to the project.

Thanks to Mr. C. Matthews for his English correction.

I reserve my deepest gratitude for my parents, who have given me continuous support and encouragement throughout my studies. Many thanks also go to my sister. Last, I would like to express my innermost gratitude to my wife, Diandian, for her moral support and tolerance over the past years.

This work is supported by the European FP6 IST Project "Tangible Acoustic Interfaces for Computer-Human Interaction (TAI-CHI)". The Manufacturing Engineering Centre (MEC) is the coordinator of this project.

[1]This thesis was typed using free software: the editor Emacs and the document preparation system LaTeX 2_ε.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

As the development of the computer grows so considerably, the multiplicity and amount of information to be processed expand vastly. However, the variety of Human-Computer Interaction (HCI) means, used to access or communicate the mass of information in computers, remains limited. In general, human-computer interfaces can be categorised into two types, tangible (touchable) and intangible (or non-touchable, such as audio or video). A tangible interface is a user interface in which a person interacts with digital information through the physical environment. The vast majority of input interfaces are tangible. Currently, the most important forms of such tangible interfaces include keyboards, mice, touch pads and touch screens. A common problem with these devices is that they restrict the mobility of users, constraining them to be near certain devices during interaction with the computer [Yan06]. The motivation of this work is to overcome the mobility limitation that the current devices impose.

Various new intangible approaches for interactions between users and computers have been exploited. One typical example is the use of the technology of speech recognition [Red76], freeing users from dependence on the mouse, keyboard and stylus for many applications, such as the famous software *IBM ViaVoice* [IBM]. Moreover, computer vision has been used for new computer interfaces in various

applications, such as autonomous steered cameras used in video conferences with face recognition [HW94] [Goo97] [YKA02] [WYZJ04]. Speech recognition and computer vision techniques have also been combined together in various ways, such as speech recognition with lip reading, gesture with speech for multimodal interpretation, speech with handwriting as flexible input, and face and eye tracking with sound localisation or speech recognition [WVDM96].

Rather than designing new ways to communicate with PCs, the primary interest of this research is to transform the real physical environment into an interactive cyberspace (figure 1.1). More importantly, the final outcome should be easily adapted to existing software. Regarding the tangible interactive surfaces currently available, most of them are embedded in electronics/optoelectronics devices, such as the widely-used resistive and capacitive touch screens. These interactive surfaces are designed and sold specifically as part of the products, causing the mobility limitation. Several projects have tried to overcome such problems. The project *Everywhere Interactive Displays* [Pin01] [KLP03] used computer vision techniques to convert any normal flat surface, such as a wall or desktop, into an interactive touch screen, by recognising predefined finger gestures to give commands. A research pioneer in tangible user interfaces is Ishii in the MIT Media Laboratory [IU97], who has headed several research projects for similar purposes.

Among these novel approaches, acoustic-based sensing technologies are becoming a promising way to bring another of the five human senses to the realm of computers – the sense of touch. The target of this research is to design multimodal tangible acoustic interfaces (TAI) that employ physical objects and the surrounding space as media to bridge the gap between the virtual and physical worlds. The method that will be developed is based on the principle that interacting with a physical object modifies its surface acoustic patterns, due for instance to the generation of acoustic vibrations (passive method) or the absorption of acoustic energy (active method) at the points of contact. By visualising

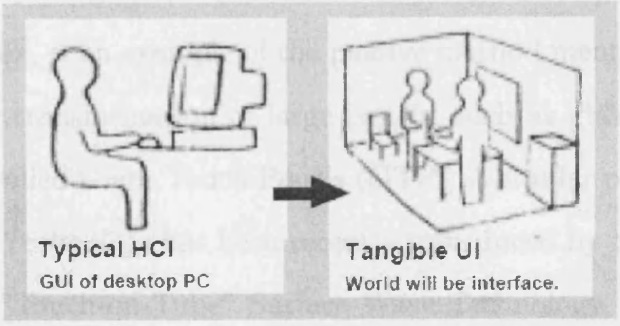


Figure 1.1: From typical HCI to interactive world [IU97].

and characterising such acoustic patterns, it will be possible to transform almost any object (for example, a wall, window, table top, giant screen, or arbitrary 3D object) into an interactive interface, opening up new modes of computer-user interaction for responsive environments. Because of their numerous advantages over other methods, including the spatial freedom they provide to the user, their durability and their ease of installation, acoustics-based interfaces will become a major sensing paradigm in the future, implying enormous potential for the whole computer and information industry. This project has explored how physical objects or flat or complex surfaces, can be transformed into natural, seamless, unrestricted touch interfaces.

A very limited number of commercial products can be regarded as preliminary attempts to develop tangible acoustic interfaces. The first such product, Intelligent Vibration [Vib], is an example of the passive method mentioned earlier. The system allows the transformation of large panels, such as glass plates and shop windows, into so-called Giant Touch Panels (GTP). A similar product using their *Dispersive Signal Technology* has been recently introduced by *3M* [3M]. Another example, *iTouch* "Touch-on-Tube" Surface Wave Technology, was developed by ELO [Tou], a world leader in touch screen technology, and is representative of the active method. Surface waves are directed onto the screen of a cathode ray tube (CRT) faceplate, eliminating the need for any glass overlay on the CRT. This technology is currently only available for screens with glass CRT faceplates up to 21 inches. These first two existing commercial products [Vib][3M] have some similarities with part of this project but are limited to the use of glass, small and flat surfaces, or non-continuous tracking. A very similar research project using acoustic techniques has been developed with preliminary experiments by the MIT media lab [IWO+99] [Che01] [Leo02], but is only capable of single impact detection. The innovation proposed here will allow the project to go well beyond these limitations.

1.2 Research objectives

The main objective of the research was to design and develop acoustics-based remote sensing technologies which can be adapted to virtually any physical object to create tangible interfaces, allowing the user to communicate freely with a computer, an interactive system or the cyber-world. The detailed research objectives were as follows:

- To evaluate different technologies for the characterisation of acoustic propagation patterns and the positioning of acoustic sources in various media;
- To research into acoustic transmission behaviour in solid media;
- To develop novel acoustic transducers for various application scenarios;
- To develop acoustic signal processing algorithms and establish intelligent tangible interfaces for the following two functions:
  - Finger impact localisation;
  - Continuous tracking of finger scratching.
- To develop prototypes of HCI applications.

Eventually, the possibilities for acoustic interactive interfaces will be virtually unlimited. Here are some examples of their potential applications:

- Low cost desktop keyboards and consumer keypads (no more switches needed);
- "Invisible mouse" (the user moves their finger directly on the surface of the table);
- Device-free electronic whiteboards and pointing systems (using conventional pens and fingers);
- Low cost consumer interfaces for PC, 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 (visitors to museums, information centres, and exhibitions);
- Part of the interactive environment of Virtual Reality (VR);
- Sensitive skin for robots and sensitive surfaces in multi-transducer environments;
- Interactive interfaces for security systems (interactive doors, windows, walls and floors);
- Novel musical instruments.

1.3 Outline of the thesis

Chapter 2 gives an overview of the relevant background in various aspects of the work, including the state-of-the-art multimodal tangible interface, acoustic wave propagation in solids, and different wave localisation approaches. Firstly, various techniques of tangible multimodal interface are reviewed. This is followed by an introduction to acoustic localisation techniques using sensor arrays. In the next section, wave propagation in solids will be introduced in detail, focusing on the dispersion properties of plate waves. A review of wave source localisation in solids follows.

Chapter 3 introduces a method, named *Location Template Matching (LTM)*, based on the principle of the time reversal theory. The original method derived directly from the time reversal theory is introduced and experiments are reported. An improvement to this method is then described, and experimental results are given to show the improvement by comparison.

Chapter 4 extends the LTM method with more specific analyses and experiments on the resolution limit with this method. Factors, such as sensor placements, geometrical symmetry, or the dependence on the source location, are tested first, in order to standardise the following experiments. Resolution experiments are then performed on different materials with different thicknesses. Discussion is based on the principle of the time reversal theory and the effect of wave propagation on different materials. With this knowledge, a faster searching technique and the simultaneous localising of multiple impacts at specific resolution become feasible and will be explained.

Chapter 5 introduces techniques of locating wave sources by four sensors at given positions. The conventional Time Difference Of Arrival (TDOA) method is first introduced, and approaches to determine the arrival times of signals, namely Time of Flight (ToF) and Cross-Correlation (CC), are explained and proven by experiment. It is difficult to establish the actual time of arrival of a dispersive wave because of the dependency of its velocity on its frequency. In addition, the ToF method lacks the ability to track continuous signals. To overcome these limitations, two novel approaches for the localisation of an impulsive acoustic source in a solid object are proposed in this chapter. The Wavelet Transform is utilised to extract different frequency components from the recorded acoustic signals for estimation of the group velocities of the various frequencies. Maximum Likelihood Estimation (MLE) is introduced to improve the accuracy and reliability of the localisation. Three localisation methods based on these techniques are introduced and compared. In addition, a novel technique to measure and to verify the dispersive phase velocities is presented.

Chapter 6 describes a novel approach to track a continuous wave source on a flat surface to enable the function of *drag & drop*, which is an everyday operation in Graphical User Interface (GUI) software. First, the Kalman filter and other extended adaptive Kalman filters are reviewed briefly. Then, the implementation of the Kalman filter is explained. Observing the relatively poor performance

for non-linear trajectories, an adaptive method is introduced to enhance the result. Comparative results between the standard Kalman filter and the adaptive Kalman filter on straight and non-linear trajectories are shown with both real and synthetic data.

Finally, chapter 7 summarises the conclusions of the present research and makes suggestions for further work.

A few prototypes of HCI applications developed during this work are outlined in Appendix A. Sensors and hardware development are introduced in Appendix B. The schematics and the specification of the data acquisition card used to communicate with computers are illustrated in Appendix C.

Chapter 2

Related work

This chapter includes four sections, covering background material related to this project. In the first section, an overview of the state-of-the-art of the HCI techniques will be given. The acoustic approach will be discussed and compared with other techniques. In section 2.3, the acoustic characteristics of vibrations in solids is studied from two aspects, namely the excitation of the vibration in solids and the in-solid wave properties. A thorough review of various approaches for acoustic localisation is then given in the last section of this chapter.

2.1 State-of-the-art multimodal tangible interfaces

2.1.1 Overview

Rather than conventional tangible interfaces, this section will only survey and introduce tangible multimodal interfaces. Touch technology can be divided into two broad categories: passive and active. Passive systems do not require a special pointer or energy source to activate the touch surface, only the user's finger or any other pointing device. Active systems, on the other hand, require a special device, which can emit some type of energy (acoustic or infrared) to contact the touch surface.

Passive touch systems: They do not rely on special pointing devices or energy distribution on the touch surface. Instead, they depend on technology in their surface area to register contact. Passive systems include:

Analog resistive (AR): Two sheets coated with a resistive material are separated by an air gap. Pressure applied to the surface brings the sheets together and registers a contact point.

Capacitive: Circuits at the corners of a charge-carrying glass panel register a current when a finger touches the surface. Frequency changes determine the position of the contact.

Acoustic: Several acoustic sensors are installed on the edges or corners of a solid object to determine the position of the acoustic source when a finger touches the surface. Such acoustic passive touch systems are durable and do not require special tools that could be lost or broken, so they are often appropriate for tough environments.

Computer vision: One or more cameras are used to monitor the interactive area, either to locate the finger position or to recognise the pre-defined finger gestures.

Active touch systems: Active systems do not always require a special touch surface, but they do require some kind of special pointing device or energy injection devices. Active touch technology includes:

Electromagnetic: A digitising screen communicates with a battery-powered pen or a pen containing a magnetic coil (a tuned resonant circuit).

Ultrasonic and infrared: Three types of approach as detailed below:

- Infrared sensor and ultrasonic technology track a battery-operated pen;
- An ultrasonic transducer generates acoustic surface waves on the touch screen. These waves are directed across the front surface

of the touch screen by an array of reflectors. Reflectors on the opposite side gather and direct the waves to the receiving transducer, which reconverts them into an electrical signal - a digital map of the touch screen surface. When one touches the screen, a portion of the wave travelling across it is absorbed. The received signal is then compared to the stored digital map. The change is recognised, and a coordinate is calculated.

- Infrared sensors are used to detect changes of the signal from infrared emitters for glass materials. A finger touch can block the path of signal propagation, such that the touch location can be determined.

In the next two sub-sections, a review of some typical existing tangible sensing technologies will be given in two parts: non-acoustic and acoustic approaches.

2.1.2 Computer vision approach

Such an approach can be traced back to the early 90's, when the project "Digital Desk" was introduced by Wellner [Wel93]. A rather complete implementation of this concept has recently been developed in the IBM lab, named "Everywhere Interactive Displays" [Pin01]. These two projects are similar in that both are trying to use a camera to track the finger in the captured images, so that any surface can be eventually used interactively. Figure 2.1 shows a diagram of the hardware setup of the "Everywhere Interactive Displays" project. This work uses an LCD/DLP projector combined with a motorised rotating mirror and a video camera. As the mirror moves, different surfaces become available to be used as displays. The video camera is used to detect hand interaction with a projected image using computer vision techniques. Many other experiments have been performed to track finger motion in the image [OZ97] [CBC95]. Rather than just tracking fingers on a surface to enable an interactive surface, there is also a similar work, which tracks the finger in a 3-D environment for the interaction

between users and computers, such as the project "DigitalEyes" [RK94].

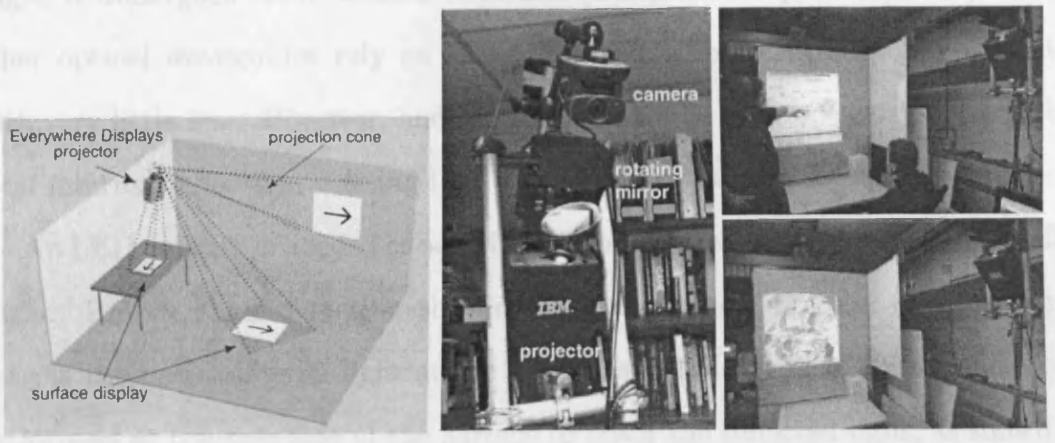
There are several technical and practical difficulties associated with this technique. The first one is the huge space required for the hardware setup. Secondly, any obstacle between the camera and the surface can block the view of the camera. Thirdly, it would be difficult to detect a real touch between the finger and the surface for such a 2-D computer vision technique. In addition, the system can only work with those predefined objects, such as a fingertip, whose shape template has been learnt by the computer beforehand. Finally, thanks to the projectors used in the first two projects, the computer vision technique is suitable in situations of low light. Otherwise, it would be difficult to use it in such low light conditions.

2.1.3 Infrared approach

CarrollTouch [Tou] is an example of using such a technique. Scanning infrared (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 (figure 2.2). 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 detect the absence of IR light and transmit a signal that identifies the X and Y coordinates. One obvious drawback of this approach is the mobility problem, as this technique can only work on glass, and electronic devices must be mounted as part of the product.

2.1.4 Computer vision approach using laser emitter

Based on an interesting phenomenon of light transmission, called *frustrated total internal reflection* (FTIR), this technique uses a combination of computer vision and laser techniques. When light encounters an interface to a medium with a



(a) (b)

Figure 2.1: Using Everywhere Interactive Displays to create displays on different surfaces.

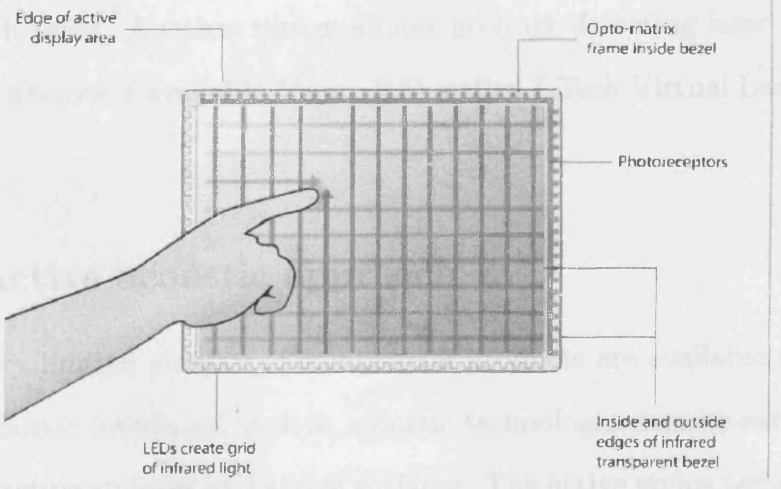


Figure 2.2: CarrollTouch infrared technology [Tou].

lower index of refraction (e.g. glass to air), the light becomes refracted to an extent which depends on its angle of incidence, and beyond a certain critical angle, it undergoes total internal reflection (TIR). Fibre optics, light pipes, and other optical waveguides rely on this phenomenon to transport light efficiently with very little loss. However, another material at the interface can frustrate this total internal reflection, causing light to escape the waveguide at this point.

An LED emitter is needed to send light into the interactive glass from a specific angle. Rather than detecting the finger in the image directly, this technique detects the backscattered light at the locations of finger touch. One camera can be located at the rear side of the surface to track the reflected light, as shown in figure 2.3 [Han05]. A similar system, Digital Vision Touch Technology (DviT), uses four cameras at the corners of the surface that can be adapted on a display screen. Each camera scans the surface and detects the target by calculating its own angle from the affected pixel, as shown in figure 2.4 [dvi03] [Mor05]. A mathematical formula has been developed to automatically record the distance between the two cameras and their viewing angles in relation to each other. With this information, the technology can then triangulate the location of the contact point. This technique allows the precise tracking of multiple fingers only on glass in both cases. Another rather similar product detecting laser reflection for small input devices is available (figure 2.5), called I-Tech Virtual Laser Keyboard [TVLK00].

2.1.5 Active acoustic approach

So far, a very limited number of commercial products are available to be used as tangible acoustic interfaces. Active acoustic technologies can be categorised into two types: active styluses and active surfaces. The active stylus technique locates an active stylus, which emits acoustic signals, normally ultrasound. The active surface technique is different in that the surfaces are fitted with active transducers emitting waves to the surface. A touch by a fingertip on the surface will modify

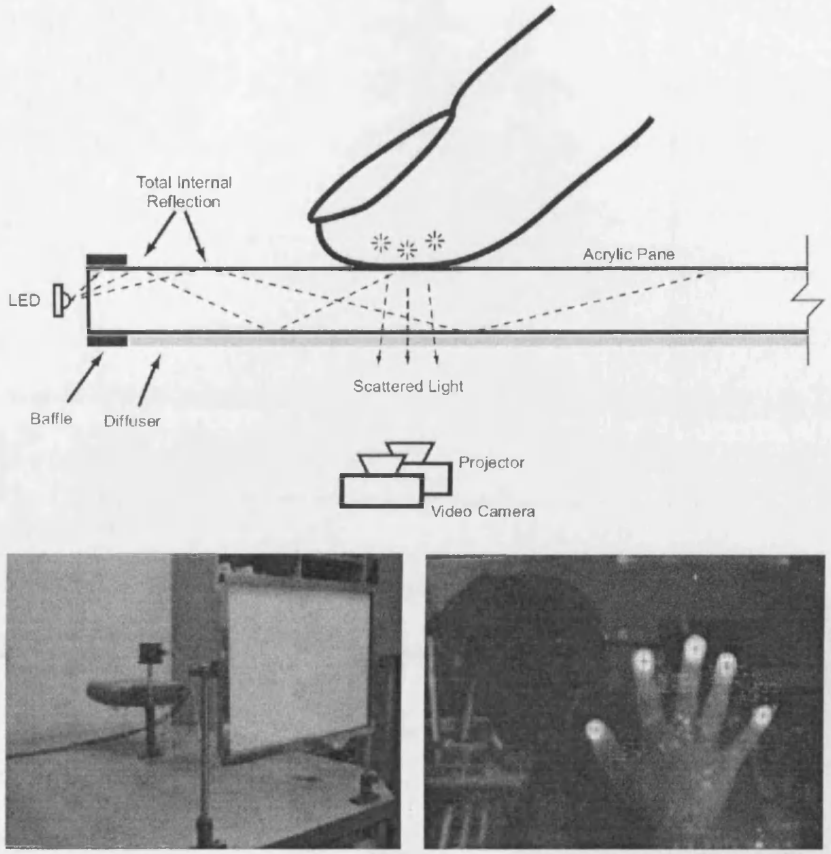
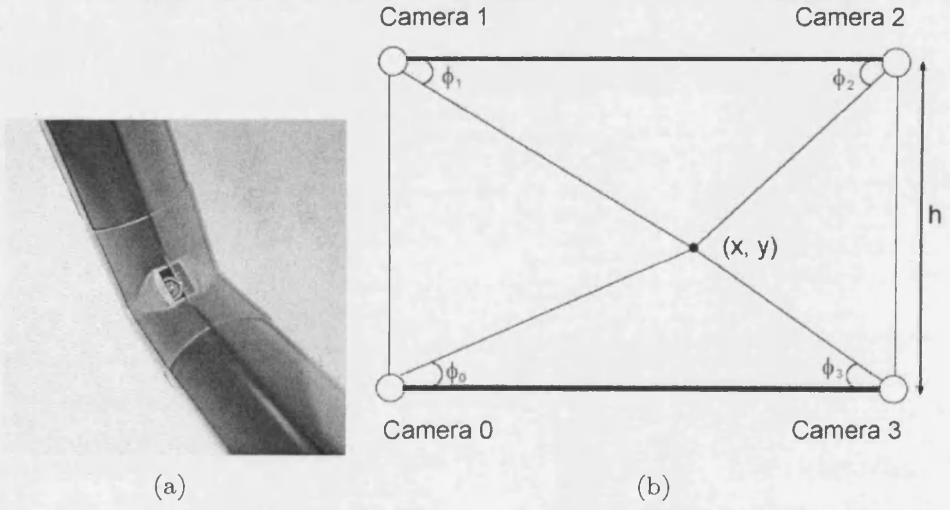
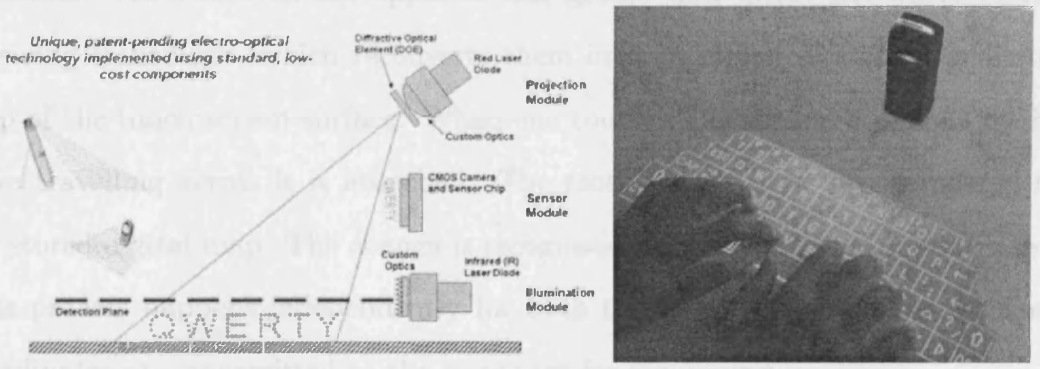


Figure 2.3: FTIR interactive surface [Han05].



(a) (b)

Figure 2.4: DviT Technology [dvi03].



(a) (b)

Figure 2.5: I-Tech Virtual Laser Keyboard [TVLK00].

the original signal map, and in this way the coordinates will be determined.

A company famous for using such active surfaces is ELO [Tou], which has introduced three slightly different techniques, namely iTouch, IntelliTouch, and SecureTouch surface wave touch technology. They are based on similar structures, but on slightly different algorithms. Figure 2.6 depicts an example workflow of the IntelliTouch system. The touch screen is a glass overlay with transmitting and receiving piezoelectric transducers for both the X and Y-axes. The touch screen controller sends a five-megahertz electrical signal to the transmitting transducer, which converts the signal into ultrasonic waves on the surface of the glass. These waves are directed across the front surface of the touch screen by an array of reflectors. Reflectors on the opposite side gather and direct the waves to the receiving transducer, which reconverts them into an electrical signal - a digital map of the touch screen surface. When one touches the screen, a portion of the wave travelling across it is absorbed. The received signal is then compared to the stored digital map. The change is recognised, and a coordinate is calculated. This process happens independently for both the X and Y-axes. The digitised coordinates are transmitted to the computer for processing.

Rather than tracking finger touches, the other technique is to locate a special active device, such as a stylus pen that can emit ultrasound signals when used for writing. Two commercial products are found of this type, namely:

1. The Pegasus PC notes tracker (figure 2.7) [Ame]

2. The MIMIO interactive whiteboard (figure 2.8) [Mac01] [Wor02]

As can be seen from the figures, a receiver is mounted on the surface, which can receive the emitted signals. The surface is required to be flat and limited in size, depending on the output power from the stylus. One advantage of this technique is the mobility feature, that can be used on almost any flat surface.

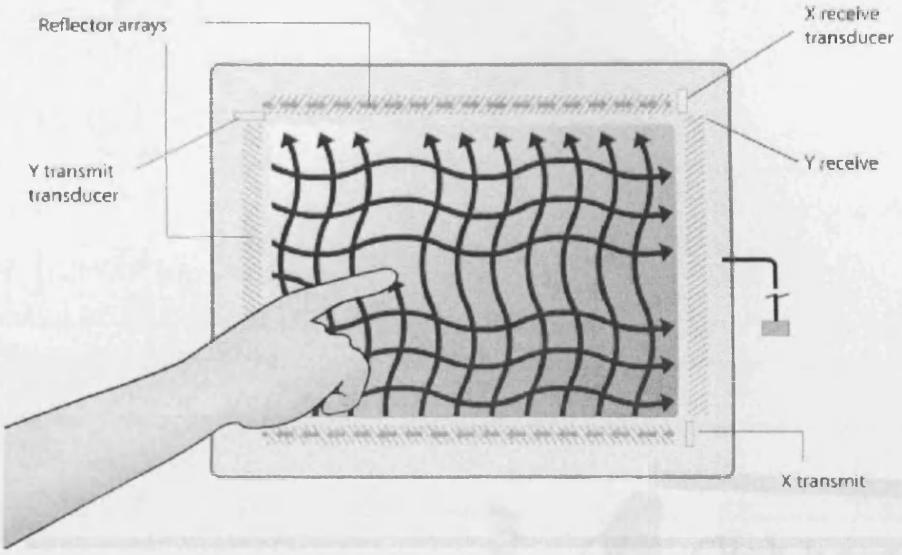
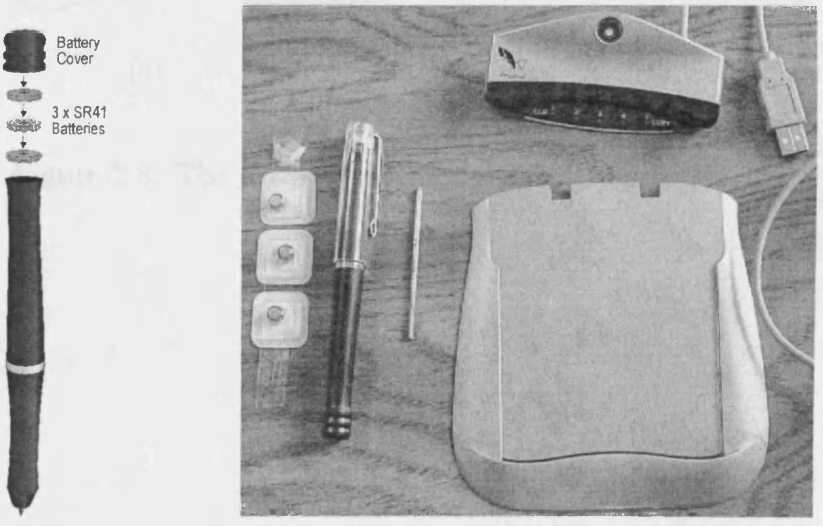
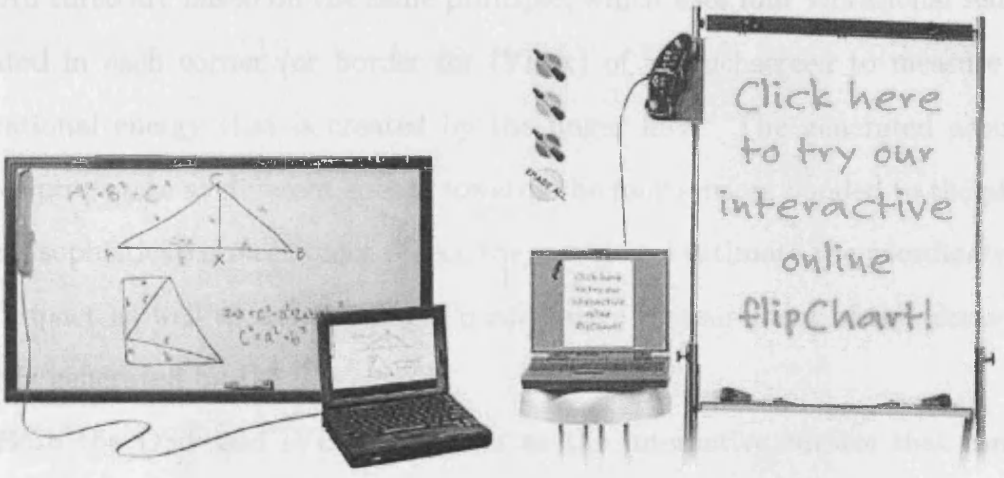


Figure 2.6: ELO iTouch principle [Tou].



(a) (b)

Figure 2.7: The Pegasus PC notes tracker [Ame].



(a) (b)

Figure 2.8: The MIMIO interactive whiteboard [Mac01].

2.1.6 Passive acoustic approach

For the passive techniques, only two commercial products and one research project have been found, as listed below:

1. IVBox by Intelligent Vibration (figure 2.9) [Vib]

2. PingPongPlus project and its extensions by the tangible media group of the MIT media lab (figure 2.10) [IWO+99]

3. Dispersive Signal Technology (DST) by 3M (figure 2.11) [3M]

All three are based on the same principle, which uses four vibrational sensors located in each corner (or border for IVBox) of a touchscreen to measure the vibrational energy that is created by the finger flick. The generated acoustic waves propagate at different speeds towards the four sensors bonded to the plate. Then, sophisticated electronics detect the waves and estimate the coordinates of the impact as well as determining a quantitative measurement of the ultrasonic energy generated by the flick.

Both the DST and IVBox use glass as the interactive surface that can be adapted as a screen, but the MIT media lab experimented on a Ping-Pong table and a lecture whiteboard. One main difference among the three projects is the variety of ways in which the sensors are attached. In the IVBox system, each one of the four sensors is a pair of piezoelectric transducers bonded on opposite faces of the plate. The DST uses one sensor mounted on the backside of the glass. However, the MIT media lab mounted the sensors on the front side. Although they are all based on the same principle, the algorithms used to calculate the coordinates of finger impact are slightly different. The DST technique is a method that tries to compensate for the dispersion caused by both reflections and the materials. In the IVBox system, only the information in the ultrasound bandwidth is taken into account, in order to provide accurate detection of signal arrivals. In the MIT system, two conventional methods, cross-correlation and rising-edge,

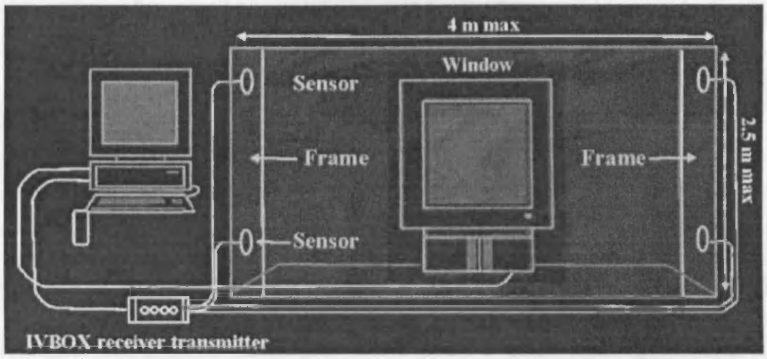
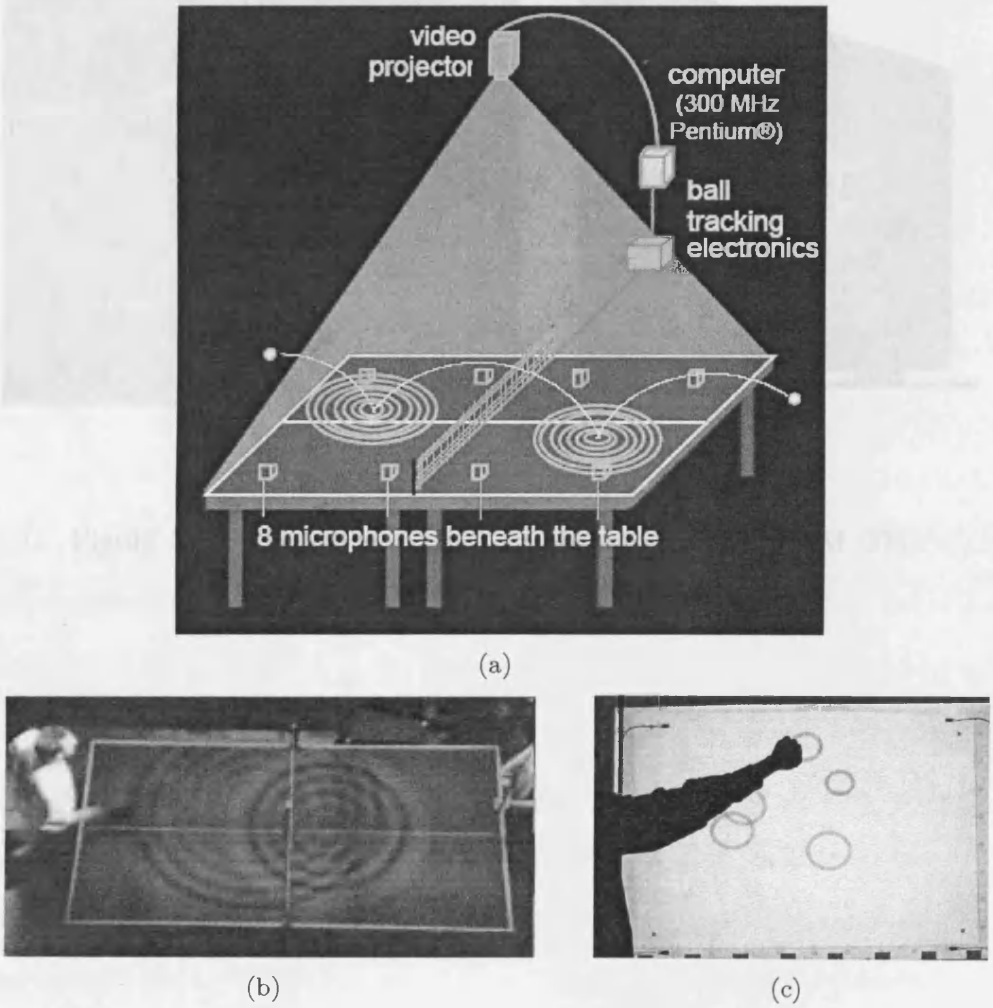


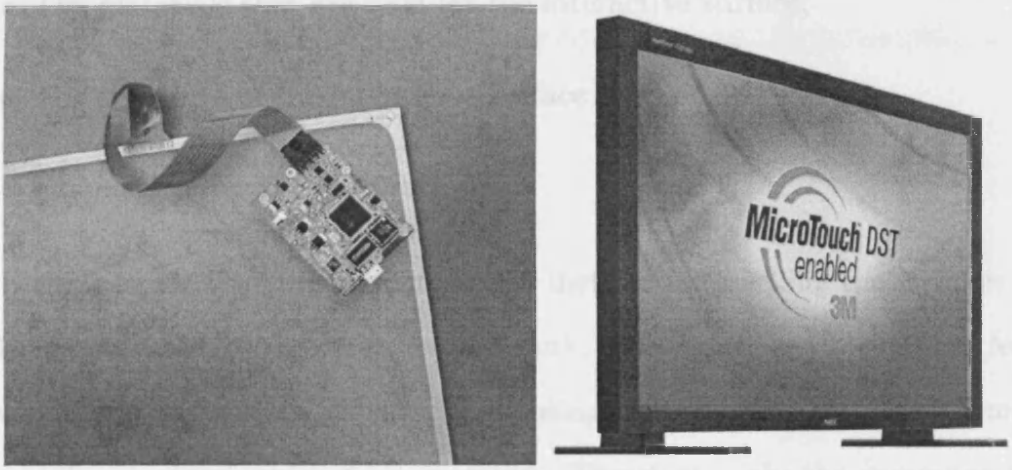
Figure 2.9: IVBox by Intelligent Vibration [Vib].



(a)

(b) (c)

Figure 2.10: PingPongPlus project and its extensions [IWO+99].



(a) (b)

Figure 2.11: Dispersive Signal Technology (DST) by 3M [3M].

were used to detect the time differences of signal arrivals between different pairs of sensors.

2.1.7 Discussion

These different techniques are discussed and compared to the proposed method based on the passive acoustic method in the following aspects (table 2.1):

- Function of tracking continuous movement;
- The type of object to be tracked;
- The materials that are used for the interactive surface;
- The structure of the interactive surface (2-D or 3-D);
- Extra space for the hardware setup.

It can be seen that the acoustic active method for tracking the stylus is obviously not suitable for the target of this work, although it has the desired feature of working on any surface. The methods using IR and active ultrasound emitters work only on glass, which limit mobility. Therefore, only the computer vision method and the acoustic passive method have the advantage of working on any surface. However, as seen in the table, there are several technical difficulties associated with the computer vision method. The object to be tracked with the computer vision method must be clearly defined as a template in the computer memory. That object is normally the fingertip. In addition, the computer vision method has the limitations of requiring huge space for setup, and any obstacle between the camera and the surface can block the view of the camera. Other limitations include the inability of working in a low light environment and the difficulty of judging the physical contact between the finger and surface with a 2-D image. Therefore, it seems that only the acoustic passive method has the advantages over other approaches that it is not limited to the type of objects and surface materials. However, at this stage, all the commercial and research

	Computer vision	IR	Computer vision & IR	Acoustic active surfaces	Acoustic active devices	Acoustic passive method	Proposed method
Continuous tracking	N/A	Yes	Yes	Yes	Yes	No	Yes
Surface materials	Any	Glass	Glass	Glass	Any	Glass and Pingpong table in MIT	Any
Object to track	Fingertip	Fingertip	Fingertip	Fingertip	Special stylus	Any	Any
Structure of surface	2-D	2-D	2-D	2-D	2-D	2-D	3-D
Extra space for setup	Big (steered projector and camera)	No	Big (back projector and camera)	No	Receiver on surface	No	No
Other comment	Not suitable in situation of low light without projector. Computer vision cannot tell the event of finger touch. Obstacles can block the camera and projector.						

Table 2.1: Comparison of existing tangible multimodal interfaces with the propose approach.

work on the acoustic passive method lacks functions of tracking of continuous movement and has mostly been tested on glasses. Moreover, they are all based on the same principle as detailed later. The proposed method will be based on the acoustic passive approach and will investigate more approaches in order to overcome these functional limitations.

Regarding the response time of an interactive system, MacKenzie and Ware [MW93] discussed the fact that the user performance is degraded by the perception of increasing lag in system responsiveness. A lag of up to $75ms$ will not degrade the speed and accuracy of mouse movement. In addition, an optimal range of response time has been determined that is between $25ms$ and $75ms$. For this reason, the proposed method will be evaluated with this criterion in the future.

2.2 Array signal processing

As introduced above, to determine the impact position, the passive acoustic method involves fusing data collected at several sensors in order to carry out a given estimation task (space-time processing). This leads to an interest in reviewing the topic of array signal processing. One notable problem in this research area is the source localisation in radar and sonar using prior information from the data acquisition system (i.e., array geometry, sensor characteristics, etc.). In particular, this section will focus on acoustic localisation. Different approaches will be described in brief and a review of related work in each area will be given.

2.2.1 The beamforming method

Spatial filtering or the beamforming method is the first space-time processing approach to data sampled by an array of sensors. The initial (Bartlett) beamformer was introduced in 1948, as presented in [KV96]. The beamformer is a method used in combination with sensor arrays in order to enhance the acoustic source

of interest present in the recorded mixture of this source and interfering noises. Temporal filtering techniques are useless in the situation where the interfering noises occupy the same frequency band as the desired signal. However, due to the different spatial locations of the different sources, the beamformer is used to separate the signals from the interference spatially. The array of sensors is normally configured in a uniform linear or circular geometry. Figure 2.12 shows an example of a linear array of sensors.

The output of each sensor is properly filtered and will be added up to estimate the power from one direction. A general description in matrix form of this operation is given as

$$\mathbf{y}(t) = \mathbf{w}^H \mathbf{x}(t) \tag{2.1}$$

where $\mathbf{y}(t)$ represents the output at time t, $\mathbf{w}$ and $\mathbf{x}$ are columns vectors of the weights and signal inputs. The array will be steered in all possible directions and the corresponding time delays τ on each of the sensors will be inserted.

$$\tau_i = \frac{(i-1)d}{v} \sin\theta_0 \tag{2.2}$$

where d is the distance between two sequential sensors, v is the signal propagation speed, i represents the index of the sensor and θ_0 is the steered angle. The maximum power indicates the Direction Of Arrival (DOA).

In order to solve the drawbacks of limited spatial resolution capability and susceptibility to interfering sources, modern beamforming approaches have been developed to enhance performance either by assuming or by measuring prior knowledge of the acoustic environment [Tho90]. The effect of unequally spaced array geometry has also been discussed [Har61] [Ish62] [SNS64] [DB88]. The main effort to improve the beamforming performance has been focused on the selection of weight values, which can be categorised into two types: *data independent* and *statistical optimum*, depending on the selection of weight values [VB88]. The

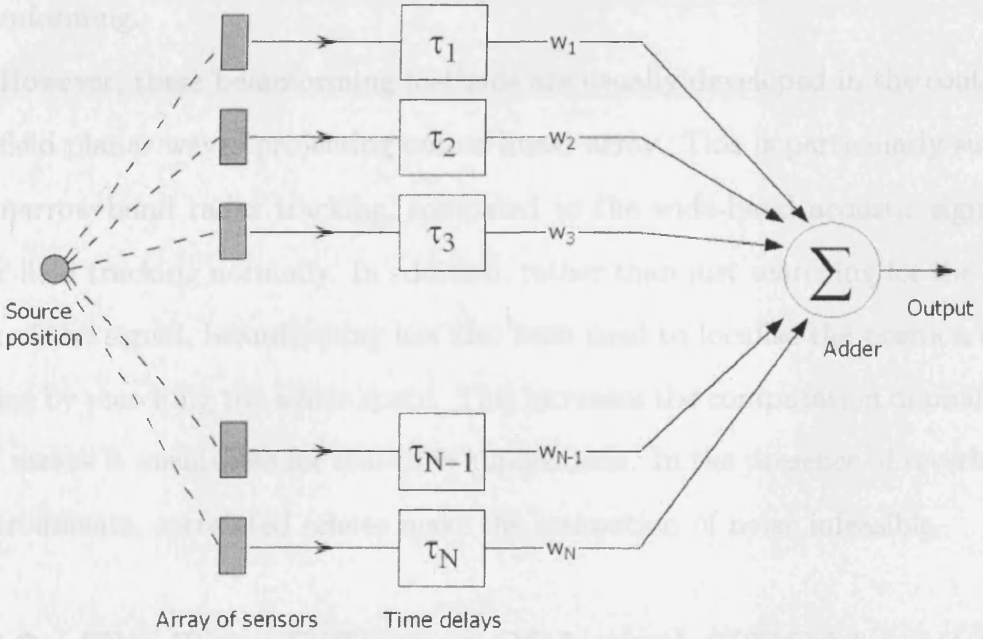


Figure 2.12: Typical beamforming system with an array of sensors.

classical approach of delay and sum beamforming is an example of the data independent type. Further modifications made to this type are always based on the prior knowledge of the desired response, such as the direction of the signals and interference. Statistical optimum beamforming adjusts the weights based on the statistics of the data received at the array. Various criteria have been attempted to minimise the contribution due to noise and interference from other directions than the source. This includes approaches, namely multiple sidelobe cancellers, use of a reference signal, maximisation of signal to noise ratio, linearly constrained minimum variance beamforming, and signal cancellation in statistically optimum beamforming.

However, these beamforming methods are usually developed in the context of far-field planar waves projecting onto a linear array. This is particularly suitable for narrow-band radar tracking, compared to the wide-band acoustic signals in near-field tracking normally. In addition, rather than just searching for the direction of the signal, beamforming has also been used to localise the position of the source by searching the whole space. This increases the computation dramatically and makes it unsuitable for real-time applications. In the presence of reverberant environments, correlated echoes make the estimation of noise infeasible.

2.2.2 The Time Difference Of Arrival (TDOA) method

The Time Difference (sometimes also referred to as Time Delay) Of Arrival (TDOA) method has been a popular research area in recent decades. Rather than directly searching for the location with the received data, the TDOA method tries to derive the time differences between each pair of the signals received by the sensors in the array. These delays result from the differences in distances from an acoustic source to the sensors placed at known geometric positions. The source location can then be calculated with simple triangulation relationships of the sensors. Figure 2.13 depicts the TDOA measurement between two sensors. Knowing the TDOA between 2 sensors, the locus where the excitation point lies

is a branch of the hyperbola (this comes from the definition of a hyperbola). It follows that to know the source point in a plane, 3 sensors are necessary, and the solution is the intersection between the 2 branches of the hyperbola obtained considering 2 sensors at a time.

To estimate the time difference between the received signals at two sensors, various approaches have been investigated. Using the simple signal model introduced before where reverberations and impulse response functions are not considered, two signals received at two sensors can be described as:

$$\begin{aligned} x_1 &= s(t) + n_1(t) \\ x_2 &= \alpha s(t-\tau) + n_2(t) \end{aligned} \tag{2.3}$$

where $s(t)$ represents the source signal, n_1 and n_2 represent the noise for the two channels, α is the factor of attenuation at the second sensor, and τ is the time delay between the two channels. With the simple assumption of independent noises, cross-correlation is the most widely used method to determine the time difference

$$R_{12}(\tau) = E[x_1(t)x_2(t+\tau)] = \alpha E[s(t)s(t+\tau)] \tag{2.4}$$

which is the autocorrelation function of s with its peak value at time τ. The distance difference $\Delta d = d_1 - d_2$ can be formulated as $\Delta d = \tau/v$, where v is the wave propagation velocity. This defines the hyperbola as shown in figure 2.13(b).

The 1980's was the peak period for this research area. Extensive theoretical analysis for time delay estimation was studied in one special issue in 1981 [tdo81]. Most popular methods are based on the above conventional cross-correlation calculation by applying various pre-filters, known as the Generalized Cross Correlation (GCC) [KC76]. There are five different GCC pre-filters discussed in [KC76], which are the Roth processor, the smoothed coherence transform (SCOT), the

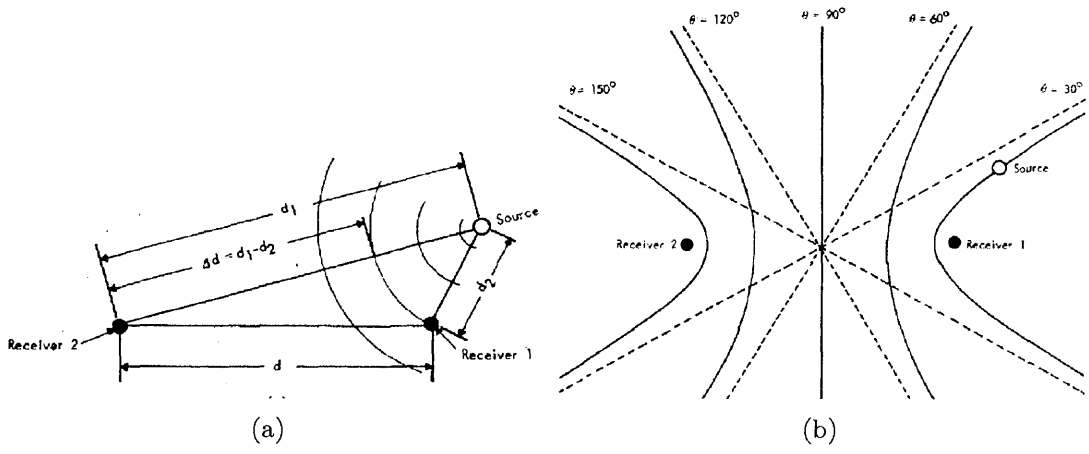


Figure 2.13: Source location (direction) is measured based on the distance difference between two sensors [Pie81].

phase transform (PHAT), the Eckart processor, and the HT filter respectively. The PHAT is the most effective and prevalent one, defined as:

$$\psi^g(f) = \frac{1}{|G_{x_1x_2}(f)|} \tag{2.5}$$

where $G_{x_1x_2}$ is the cross power spectrum of the two signals x_1 and x_2.

Derivation of the time difference estimation has also been carried out by minimising the spectral phase domain using a weighted least squares criterion [CHP78] [Pie81]. However, all these methods to determine the TDOA in either the time domain or the spectral domain suffer the same problem from noise and reverberation. In the spectral domain, the phase information is only usable when the source signal is dominant in power compared to other noises, in the selected frequencies. With correlated reverberations, these methods become even more vulnerable.

Statistical methods using the Maximum Likelihood Estimate (MLE) have been used [BG02] [MA04] to overcome this shortcoming by searching for the most likely source position. This method can effectively lessen the errors introduced by the ill measurement. Despite of the heavier computation introduced by MLE, a considerable advantage of TDOA is that there is less computational demand than the beamforming method [Bra95], so that it can be employed in real-time applications. Another limitation of the direct TDOA estimate is the inability to locate multiple sources. However, the MLE method inherited from the steered beamforming method can look for all the source positions if the number of sources is known beforehand.

There has also been some related work carried out which identifies the onset of pitch periods and then cross-correlates these within a certain temporal window [FC84] [Goo97] [RYPD05]. However, this method has a limitation that it must take a very long segment of signals, which contain enough information, to estimate the time delay accurately. Also, the long signal means that the source must be

relatively more stationary spatially than with other methods that need shorter signals.

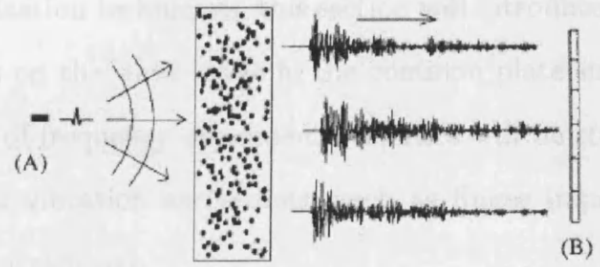
Another simple but efficient method is to measure the Time of Flight (ToF) to obtain the arrival times of signals [MW01] [BK92]. In brief, the ToF method determines the arrival time of a signal by a threshold at a predefined level. The arrival time is obtained when the signal exceeds the threshold. Quite often, the ToF method is used to locate objects with active sonar sources underwater. Four different techniques, including optimum correlation detection, have been compared experimentally in [BA98]. The optimum correlation detection technique outperforms all the other three methods, which are based on the threshold principle, in terms of accuracy. However, the simple ToF method has the advantage of a low computation demand suitable for real-time application. Moreover, one big problem with these methods is that prior knowledge of the signal model is required. Time-frequency analysis has also been applied in music labelling applications to recognise the onset of different music instruments at different spectral bands [BBP04].

2.2.3 The time reversal method

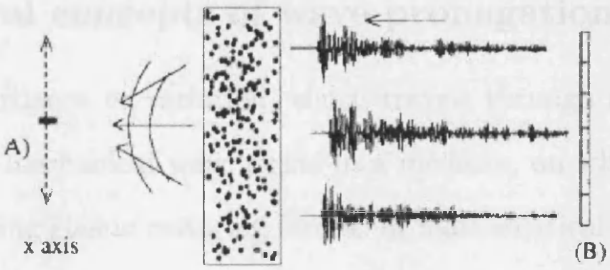
Compared to conventional approaches, the time reversal method has shown a significant importance in the area of acoustic source localisation in difficult media. In the time reversal experiment, there are two steps as shown in figure 2.14. At the first step, an array of transducers, laid out as a cavity or mirror, are used to record a transient signal emitted from a point like position. The recorded signals in the memory will be reversed in time (Last In First Out) and re-emitted from the transducers to the same environment or medium. The re-emitted signals will be propagating backwards as the time goes back and focusing at the original source location with a high peak of energy. This parallel procedure allows the conversion of a divergent incident wave, issued from a point-like source, into a convergent reflected wave, focusing on the source [FCD+00]. Theoretically, a

time reversal experiment, during traceback to the wave source location, is able to discover all paths of the wave propagation, including the direct, reflected, refracted, or scattered waves. The field of displacement can be considered as a superposition of all these waves. This introduces the primary advantage of this time reversal process. The distortions by an inhomogeneous medium located between the object and the transducers are compensated for by the time reversal.

The Time-Reversal Cavity (TRC) is an array of reversible transducers that completely surround the source position in 3-D space in order to capture all the signals passing through multiple paths. Thus, the re-emitted signals will go backwards through all the paths that they propagated from the source. The Time-Reversal Mirror (TRM) is made of an array (1D or 2D) of reversible transducers. As the TRC is difficult to realise in practice, TRM was introduced and is commonly used in time reversal experiments [FPWC89] [FP01] [Fin92]. TRM was designed to be an array of transducers as a mirror that can capture the signals within a specific angular sector. TRM can re-emit signals to limited paths resulting in a lower quality of refocusing. However, it has the significant advantage of easier implementation since the TRC is in fact impossible to implement, especially in a large open environment. TRM does, however, have the limitation of losing information about signal paths. In addition to the single point localisation, iterative experiments with TRM have been applied to find the target of interest in the situation of existing multiple targets. By moving the TRM to an appropriate position, it can capture the most significant feature of paths and will ignore signals from other sources. This is normally used in the so-called passive time reversal method, where the signal is emitted from one of the transducers and reflected by target obstacles. In contrast, the active time reversal method is designed to localise the active acoustic source that sends out the signal.



(first step)



(second step)

Figure 2.14: Time reversal experimental setup. In the first step the source (A) transmits a short pulse that propagates through the rods. The scattered waves are recorded on a 128-element array (B). In the second step, the 128 elements retransmit the time-reversed signals through the rods. The piezoelectric element (A) is now used as a detector. It can be translated along the x axis while the same time-reversed signals are transmitted by B, in order to measure the directivity pattern. [FCD+00]

2.3 Wave propagation in solids

However, as mentioned in the last section, the signal model normally used is assumed to be non-dispersive and so the time difference between signals arriving at two sensors can be formulated with the distance difference and a constant velocity, as $\tau = \Delta d/v$. However, wave propagation in solids is much more complicated than the simple non-dispersive assumption for sound in air. In order to better utilise source localisation techniques, this section will introduce wave propagation in solids and focus on the wave mode in the common plate structure. In particular, the property of frequency dependent velocities will be studied. Finally, the physical process of vibration excitations, such as finger impacts or continuous scratches, will be introduced.

2.3.1 General concepts of wave propagation

A wave is a disturbance or variation, which travels through space, often transferring energy. A mechanical wave exists in a medium, on which deformation is capable of producing elastic restoring forces. In mathematical terms, a travelling wave in one dimension is defined by an expression of the type $f(x - ct)$, where the function f represents a disturbance advancing in the positive x-direction with a phase velocity c and t is the travelling time. The function f is called a simple wave function, and the argument $(x - ct)$ is the phase of the wave function. For mechanical waves f generally denotes a displacement, a particle velocity or a stress component.

The phase velocity of a wave is the rate at which the phase of the wave propagates in space. This is the velocity at which the phase of any one frequency component of the wave will propagate. Wave disturbances in nature rarely occur at a single frequency; however, it is mathematically and conceptually more convenient to study single-frequency characteristics. The phase velocity can be given in terms of the wave frequency ω and its wave number k_ω by

$$c_p = \frac{\omega}{k_\omega} \tag{2.6}$$

A non-dispersive system is defined as one where the phase velocity c_p does not depend on the wavelength (or frequency). In general cases, acoustic waves, such as sounds in air, are considered non-dispersive. Only if the relationship between k and ω is linear will an arbitrary spatial form of disturbance not be subject to change as it propagates: such change is known as *dispersion.* Dispersion is an important phenomenon because it governs the shape of a pulse as it propagates through a dispersive medium.

Group velocity c_g is another important concept, describing the propagation speed at which energy is conveyed along a wave. The group velocity of a wave is the velocity with which the variations in the shape of the wave amplitude (known as the modulation or envelope of the wave) propagate through space. It can be formulated as:

$$c_g = \frac{\partial \omega}{\partial k} \tag{2.7}$$

In most cases the group velocity can be considered the signal velocity of the waveform. However, it is worth mentioning that the accuracy of this relationship may be significantly weakened for a medium with high attenuation.

2.3.2 Wave propagation in solids

This section gives a brief overview of wave motion in solids, with the focus on the plate (Lamb) wave in the last sub-section. More detailed information can be found from many in-depth books [Gra75] [KD76].

In solids, sound waves can propagate in four principle modes that are based on the way the particles oscillate, as summarised in table 2.2. Sound can propagate as longitudinal waves, shear waves, surface waves, and in thin materials as plate waves. Longitudinal and shear waves are the two modes of propagation most

widely used in ultrasonic testing.

Media type	**Wave type**	**Particle behaviours**
Unbounded	P-wave (Longitudinal, compressional, or dilatational wave)	Parallel to the direction of wave propagation
Unbounded	S-wave (Transverse, shear, or distortional wave)	Perpendicular to the direction of the wave propagation
Bounded (semi-infinite)	Surface wave (Rayleigh wave and love wave)	*Rayleigh wave:* Elliptical orbit. *Love wave:* Parallel to plane layer, perpendicular to wave direction
Bounded (plate)	Plate Wave (Rayleigh-Lamb wave)	Similar behaviour as Rayleigh wave, but bounded by the two surfaces

Table 2.2: Wave types in solids.

Waves in unbounded media

As can be seen in the table 2.2, in unbounded media, waves are categorised into two modes: primary wave (P-wave) (figure 2.15(a)) and secondary wave (S-wave) (figure 2.15(b)).

P-wave: In the P-wave, the oscillations occur in the longitudinal direction or the direction of wave propagation. In general, this wave mode is normally known as the longitude wave. Since compressional and dilational forces are active in these waves, they are also called pressure or compressional waves. They are also sometimes called density waves because their particle density fluctuates as they move.

S-wave: In the S-wave, the particles oscillate at a right angle or transverse to the direction of propagation. This wave mode is normally referred to as the transverse or shear wave. S-waves require an acoustically solid material for effective propagation and, therefore, are not propagated effectively in

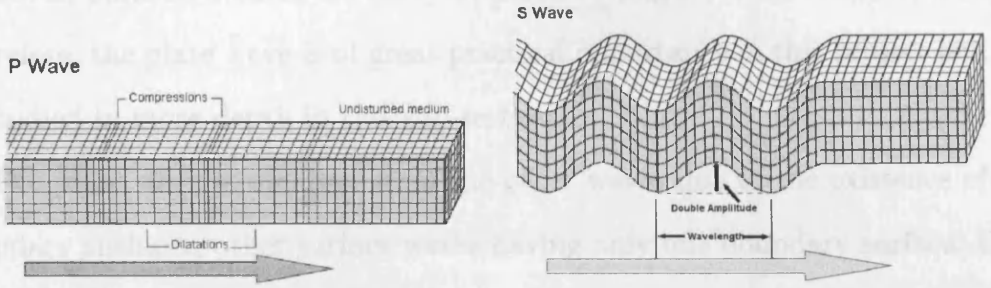
materials such as liquids or gasses. The S-wave is relatively weak when compared to the longitudinal wave. In fact, S-waves are usually generated in materials using some of the energy from longitudinal waves.

Surface wave

The first type of surface wave is called the Love wave, named after A.E.H. Love, a British mathematician who worked out the mathematical model for this kind of wave in 1911. It's the fastest surface wave and moves the ground from side-to-side, as shown in figure 2.16(a).

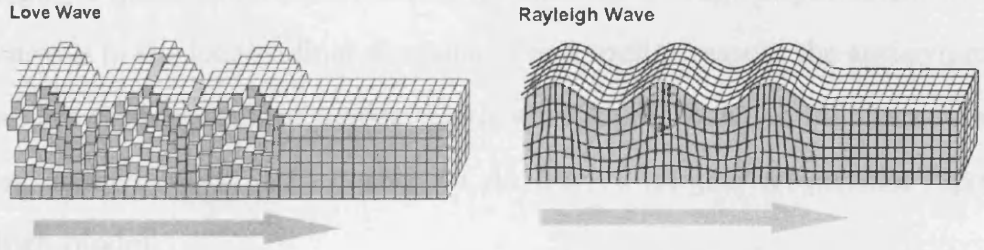
The other kind of surface wave is the Rayleigh wave, named for John William Strutt, Lord Rayleigh, who mathematically predicted the existence of this kind of wave in 1885. Most of the shaking felt from an earthquake is due to the Rayleigh wave, which can be much larger than the other waves.

Rayleigh 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 as shown in figure 2.16(b). The particles in a solid, through which a Rayleigh surface wave passes, move in elliptical paths, with the major axis of the ellipse perpendicular to the surface of the solid. As the depth into the solid increases, the "width" of the elliptical path decreases. Rayleigh waves are useful because they are very sensitive to surface defects and since they follow the surface around, they can also be used to inspect areas that other waves might have difficulty reaching. When defects of any sort (cracks, scratches, irregularities, etc.) are present on the surface over which a Rayleigh wave is propagating, the wave is scattered, as are longitudinal and transverse waves propagating into the depth of the medium. The initial Rayleigh wave is attenuated due to these scattering effects. This effect is very important in application to ultrasonic surface-flaw detection.



(a) Particle movements in P-wave mode. (b) Particle movements in S-wave mode.

Figure 2.15: Wave movements in unbounded media.



(a) Particle movements in Love wave mode. (b) Particle movements in Rayleigh wave mode.

Figure 2.16: Wave movements in semi-infinite bounded media.

Plate (Lamb) wave

The plate structure is the most common geometry structure that is available in everyday life, such as the surface of a desktop, a window, or a blackboard. Moreover, plate structures are easily vibrated by fingers, due to their thin depth. Therefore, the plate wave is of great practical importance in this project and will be studied in more depth in this sub-section.

The plate wave is different from the other waves due to the existence of two boundary surfaces, other surface waves having only one boundary surface. Plate wave propagation occurs when the wavelength, λ, is much greater than the thickness, $2b$ of the plate ($\lambda \gg 2b$) (see Figure 2.17). Some authors recommend the wavelength be ten times greater than the thickness ($\lambda > 20b$), while others propose less stringent wavelength requirements [Tuc01]. The remaining dimensions (length, l and width, w) of the plate must be distinctly greater than the wavelength.

In a traction-free infinite plate, the plate wave propagation is normally governed by the Rayleigh-Lamb theory. The modes of wave propagation in the elastic layer may be split up into two systems of symmetric mode and anti-symmetric mode, as shown in figure 2.18. The symmetric modes are usually termed the extensional or quasi-longitudinal modes because the average displacement over the thickness is in the longitudinal direction. For a similar reason, the anti-symmetric modes are termed flexural modes. Both wave modes have an infinite number of modes, named $S0, S1, S2, \ldots$ and $A0, A1, S2, \ldots$. Higher frequencies represent the high modes.

Plate waves are dispersive by nature, meaning that different frequencies travel at different speeds (phase velocities). The phase velocity c_p is the fundamental characteristic of the Lamb wave and once it is known, the wave number ξ can be determined and the stresses and displacements can be calculated at any point of the plate.

First consider the case of symmetric waves. The Rayleigh-Lamb frequency

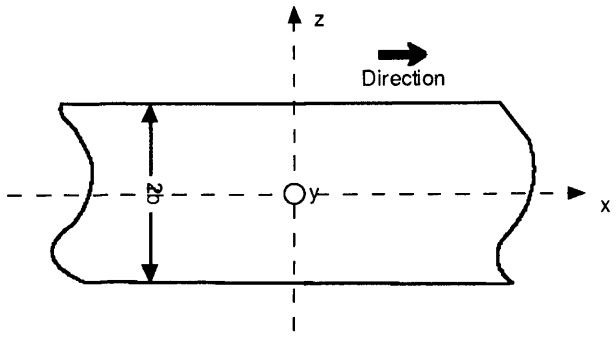


Figure 2.17: Plate with free boundary.

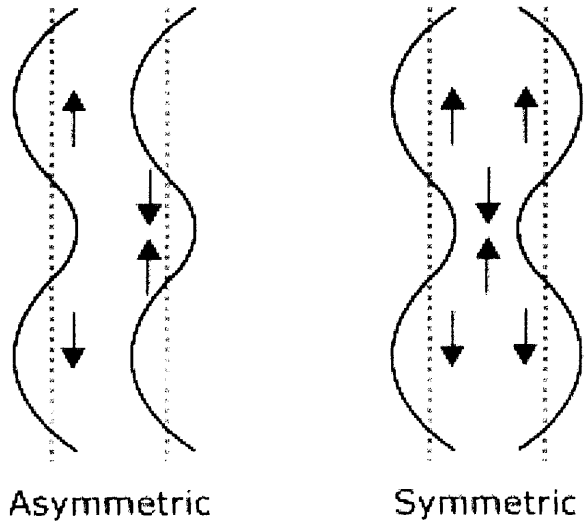


Figure 2.18: Plate wave modes: symmetric and anti-symmetric.

equation for the propagation of symmetric waves in a plate is given by

$$\frac{\tan \beta b}{\tan \alpha b} = -\frac{4\alpha\beta\xi^2}{(\xi^2 - \beta^2)^2}, \tag{2.8}$$

and in a similar form, the frequency equation for the anti-symmetric waves in a plate is formulated as

$$\frac{\tan \beta b}{\tan \alpha b} = -\frac{(\xi^2 - \beta^2)^2}{4\alpha\beta\xi^2}, \tag{2.9}$$

where

$$\alpha^2 = \frac{\omega^2}{c_1^2} - \xi^2,$$
$$\beta^2 = \frac{\omega^2}{c_2^2} - \xi^2,$$

ω is the angular frequency,

$$c_1 = \sqrt{\frac{\lambda + 2\mu}{\rho}}$$

is the p-wave velocity,

$$c_2 = \sqrt{\frac{\mu}{\rho}}$$

is the s-wave velocity. λ and μ are the Lamé constants of the solid material. Following the standard notation as in [Gra75], here λ is used to denote both the wavelength and the Lamé constants, depending on the context. The Lamé constants are related to the elastic modulus E and Poission's ratio v by the relations

$$\lambda = \frac{vE}{(1+v)(1-2v)}, and$$
$$\mu = \frac{E}{2(1+v)}.$$

With numerical methods, the dispersive curves of phase velocities can be calculated. Given the frequency ω, the wave numbers satisfying the Rayleigh-Lamb equation can be determined. From the relationship $\omega = \xi c_p$, the propagation velocity of the waves can be established. Figure 2.19 shows the result for an aluminium plate, using the default configurations (thickness: $10mm$, density: $2700kg/m^3$) of a GPL licensed program [Con] in Python language. In chapter 5, numerical calculation of the Flexural mode $A0$ on an experimented material will be carried out, in order to verify the accuracy of the measured velocity.

In the case when the transverse wavelength with respect to the thickness is quite large, so that $2\pi/\beta, 2\pi/\alpha \gg b$, only the lowest modes, $S0$ and $A0$, of the Lamb wave are available [Gra75]. For the anti-symmetric case, the phase velocity of the flexural wave $c_{p,f}$ becomes

$$c_{p,f} = \xi b \left\{ \frac{E}{3\rho(1-v^2)} \right\}^{\frac{1}{2}}, \tag{2.10}$$

and the relationship between phase velocity $c_{p,f}$ and frequency ω can be deduced to

$$c_{p,f} = \left\{ \frac{E}{3\rho(1-v^2)} \right\}^{\frac{1}{4}} (\omega b)^{\frac{1}{2}} \tag{2.11}$$

For the symmetric mode, the phase velocity of the extensional wave $c_{p,e}$ becomes

$$c_{p,e} = \left\{ \frac{E}{\rho(1-v^2)} \right\}^{\frac{1}{2}} \tag{2.12}$$

that is not dependent on the frequency.

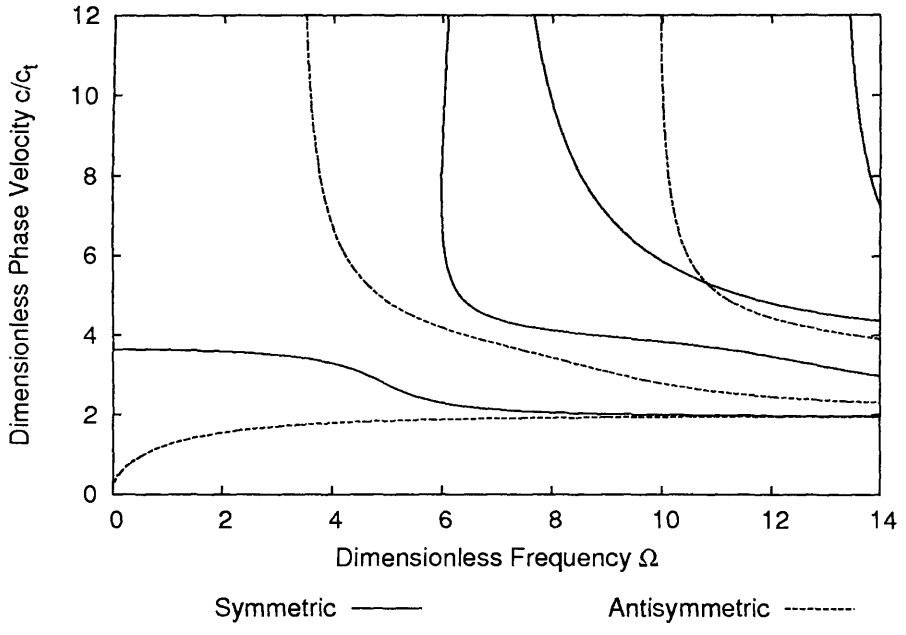


Figure 2.19: Dispersion of plate wave.

This result represents only the lowest vibrational mode in a limited frequency range in the overall frequency spectrum. In the following experiments, it will be demonstrated that these wave modes are the only available part and of the most practical importance. Moreover, these two relationships are also found to agree with the derivation from the classical plate theory (CPT), which has been widely accepted in research [Gor91] [PG94] [PG92].

A brief summary of the plate wave movements can be given from the above equations and descriptions. The flexural wave is a frequency dispersive wave, where the higher frequency components propagate faster than lower frequency components. On the other hand, it can be seen that the extensional wave velocity does not depend on the frequency, which means there is no dispersion existing in this wave mode. An extensional wave travels much faster than a flexural wave, but carries only a little energy, that can be neglected, and is attenuated more quickly than lower frequency components. There are also some other in-plane wave modes, which are not of interest here.

2.3.3 The excitation of vibrations on solid surfaces

The authors of [Leo02] and [Che01] have mentioned the phenomena that for different impacts, in terms of the frequency spectrums, hard impacts (e.g. metal or nail) may produce higher frequencies or broader spectrum bandwidth than soft impacts (e.g. fingertip). Due to the requirement of a constant for the wave velocity in the TDOA method, it is obvious that the dispersion of plate waves may impair the accuracy of the result. In [Leo02], a rather practical solution was employed by using two different measured velocities for two impacts, where a simple zero-crossing method was used to judge the type of impact. Other localisation approaches will also be affected by different impacts, and this will be introduced in detail later. Therefore, it is clearly of practical importance to understand the physical process of the contact between an object and a surface.

In fact, there are many more different ways of exciting vibration that have

been employed in vibration researches. They may influence the characteristics of the vibrations in various ways. In this section, only the two most common excitation techniques that are useful for the HCI application will be explored: impulse impacts and frictional (continuous) contacts.

Dynamic impulse impacts

In theory, the transient excitation is referred to as a transverse force applied to the surface of an elastic structure and mathematically represented by a Dirac delta function δ, as shown in figure 2.20.

However, in the real world, it is impossible to produce a perfect delta signal, in which the contact time and the contact area are both assumed to be infinitely short. Basic mathematical forms are sometimes used to approximate the excitation source for the purpose of theoretical research, such as the derivative of a Gaussian distribution. However, they are not enough to model the real physical process.

The tactile contact between a finger and an elastic plate surface has not been discussed by any of the HCI projects. However, the literature has shown the interest of contact modelling in the area of mechanical vibration testing [McC95]. Other research areas, such as sound simulation of physical collisions [RF03] and modelling of contact for simulation in robotic systems [MO99], have also performed similar studies. The following is a summary of the dynamic system model, according to the description given by [McC95].

A simple model of an impulse impact by a finger on a surface (figure 2.21(a)) can be considered as a collision between two single Degree Of Freedom (DOF) systems (figure 2.21(b)). The base structure is represented by *system 1*, which consists of mass m_1, damping constant c_1, and spring constant k_1 (stiffness). Similarly, *system 2* consists of mass m_2, damping constant c_2, and spring constant k_2, and represents the impacting object.

Each system has its own natural frequency that is given by

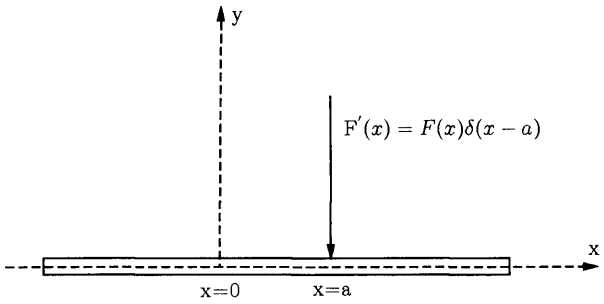


Figure 2.20: Point force excitation at position $x = a$ represented by a Dirac delta function $\delta(x - a)$.

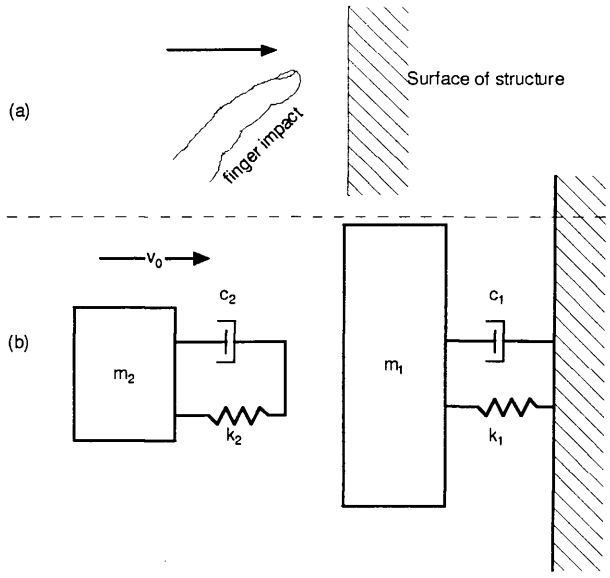


Figure 2.21: Impact by the fingertip on the structure (a) is modelled by two single degree of freedom systems (b).

$$\omega_{11} = \sqrt{\frac{k_1}{m_1}} \tag{2.13}$$

and

$$\omega_{22} = \sqrt{\frac{k_2}{m_2}} \tag{2.14}$$

and it is assumed that $\omega_{22} \gg \omega_{11}$. System 2 is assumed to have an initial velocity of v_0 before it impacts system 1, which is initially at rest. Spring k_2 and damper c_2 constitute the interface between the two systems during impact, and spring k_2 and damper c_2 are assumed to only support compressive loads. Furthermore, assume that m_1 is very large compared to m_2, then the impact force can be given by

$$F_t = \begin{cases} \frac{m_2 v_0 \omega_n}{\sqrt{1-\xi^2}} e^{-\xi\omega_n t} \sin(\omega_d t + \varphi), & \text{for } 0 < t < T/2 \\ 0, & \text{for } t > T/2 \end{cases} \tag{2.15}$$

where

$$\omega_n = \omega_{22}$$

$$\xi = \frac{c_2}{2\sqrt{k_2 m_2}}$$

$$\omega_d = \omega_n \sqrt{1-\xi^2},$$

also,

$$\tan\varphi = \frac{2\xi\sqrt{1-\xi^2}}{1-2\xi^2}.$$

For very light damping, equation 2.15 reduces to be nearly the same as the well known undamped half sine impulse that is given by

$$F_t = \begin{cases} m_2 v_0 \omega_n \sin(\omega_n t), & \text{for } 0 < t < T/2 \\ 0, & \text{for } t > T/2 \end{cases} \tag{2.16}$$

Equation 2.15 shows that the peak force F_0 is controlled by the initial momentum of $m_2 v_0$ and the impacting structure's natural frequency ω_n. This relationship works well as long as there is a large inertia mismatch between the impacting structure and the impacted structure and $\omega_{11} \ll \omega_{22}$, so that impact duration is short compared to the structure's natural period, usually less than $0.1T$.

Consider a traditional linear impulse momentum analysis where the effects of k_1 and c_1 are ignored. Let V_1 be the velocity of mass m_1 after the impact, V_2 be the velocity of the impacting structure after impact, and e be the traditional coefficient of restitution. Also assume that V_1 and V_2 are in the same direction as v_0. Then the velocities of masses m_1 and m_2 after impact can be written as:

$$V_1 = \left(\frac{m_2(1+e)}{m_1+m_2}\right) v_0 = \left(\frac{M(1+e)}{M+1}\right) v_0 \tag{2.17}$$

$$V_2 = \left(\frac{m_2}{m_1+m_2} - \frac{m_1 e}{m_1+m_2}\right) v_0 = \left(\frac{M-e}{1+M}\right) v_0 \tag{2.18}$$

where the system mass ratio M is given by

$$M = m_2/m_1$$

When mass ratio M is small, it can be seen, from equation 2.17, that structure 1 has only a small percentage of the impact velocity v_0 transmitted to it, while structure 2 is seen to rebound with a velocity that is essentially $-ev_0$. However, the described phenomenon of double strikes will occur whenever $V_2 \geq 0$, since structure 2 remains in the vicinity of structure 1, so that structure 2 is impacted

a second time by the rebounding structure 1. In fact, when mass ratio M is equal to the coefficient of restitution, equation 2.17 indicates that the impacting structure is at rest and will remain so until struck by system 1 during its return stroke one half natural period later.

Thus, it is clear that the impact force in the linear system is represented by a half sine function. The spring constant, which represents the stiffness of the impacting object, is related to the frequency of the half sine, as a higher value of k_2 indicates a higher frequency. In other words, the contact time for soft impacts is longer than for hard impacts. On the other hand, the half sine may become clipped when the initial velocity increases, due to nonlinear spring behaviour. However, this clipping only changes the input frequency spectrum.

Frictional (continuous) touch

Figure 2.22 shows the continuous frictional contact between a fingertip and a flat surface.

Friction is defined as the force that opposes the relative motion, or tendency of such motion, of two surfaces in contact. The physical friction model is normally referred to as two simple static and kinetic models. The friction force is given by the product of the friction coefficient and the normal force,

$$F = \mu N \tag{2.19}$$

where μ represents the coefficient of kinetic friction μ_k or static friction μ_s, depending on the motion of the object. Physically these reaction forces are the results of many different mechanisms, which depend on contact geometry and topology, properties of the bulk and surface materials of the bodies, displacement and relative velocity of the bodies and presence of lubrication [OAW+98].

However, this relationship is not enough to explain if vibration can be induced by friction.

There is a new class of model of friction that can explain how friction induces

vibration. In [OAW+98], different friction models are discussed. In this kind of model, small displacements and velocities are involved. To better illustrate the induction of vibration by friction between fingertip and surface, the Bristles and the LuGre models [OAW+98] give a good representation for such phenomenon, as shown in figure 2.23.

Due to irregularities in the surfaces, the number of contact points and their location are random. Each point of contact is thought of as a bond between flexible bristles. As the surfaces move relative to each other the strain in the bond increases and the bristles act as springs, giving rise to a friction force. The force is then given by

$$F = \sum_{i=i}^{N} \sigma_0 (x_i - b_i) \tag{2.20}$$

where N is the number of bristles, σ_0 is the stiffness of the bristles, x_i is the relative position of the bristles, and b_i is the location where the bond was formed. As $|x_i - b_i|$ equals δ_s, the bond snaps, and a new one is formed at a random location relative to the previous location.

The complexity of the model increases with N. The stiffness of the bristles, σ_0, normally can be made velocity dependent. An interesting property of the model is that it captures the random nature of friction. The randomness depends on the number of bristles. The model is inefficient in simulations due to its complexity.

Since it is almost impossible to model the real behaviour of the friction and it is not of practical importance in the scope of this project, the detailed mathematical model will not be presented here. However, this model gives a view of this phenomenon and shows that the friction force is related to the moving speed and other parameters.

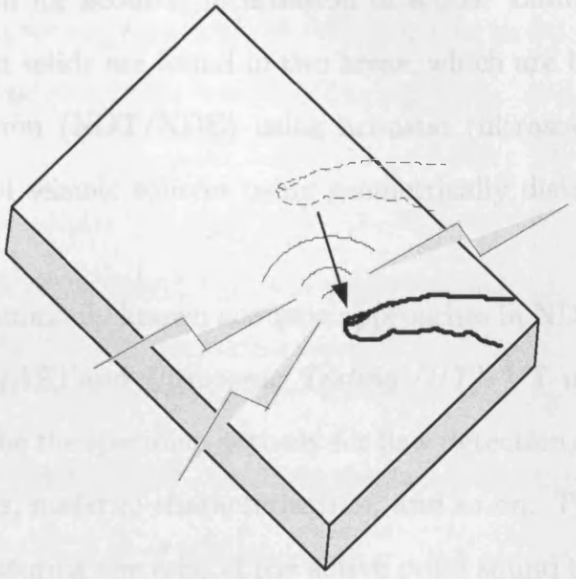


Figure 2.22: Friction between the fingertip and the surface.

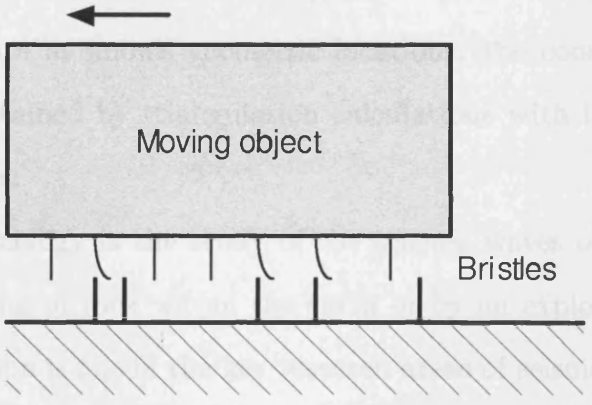


Figure 2.23: Bristles model of friction.

2.4 Acoustic localisation in solids

Due to the complexity of wave propagation in different solid materials, there is no arbitrary solution for acoustic localisation in solids. Limited applications of sound localisation in solids are found in two areas, which are the non-destructive testing and evaluation (NDT/NDE) using acoustic (ultrasound) emission and in the localisation of seismic sources using geometrically distributed monitoring seismographs.

There are two commonly known acoustic approaches in NDT/NDE, which are *Acoustic Emission (AE)* and *Ultrasonic Testing (UT).* UT uses high frequency sound energy to probe the specimen actively for flaw detection/evaluation, dimensional measurements, material characterisation, and so on. Typically, inspection is conducted by measuring the echo of the active pulse sound at the discontinuity in the material (such as a crack). Rather than actively probing the material, AE is a method of passively monitoring the emissions from active defects. AE is a class of phenomena whereby a transient ultrasonic wave (ranged from $20KHz$ to $1MHz$) is generated by rapid local stress release upon damage formation in a loaded material. The energy release can propagate through the solid and be measured on the surface of the material by vibrational sensors, which convert the mechanical wave into an electrical signal. With several sensors mounted on the material surface at known geometric locations, the coordinates of possible sources can be obtained by triangulation calculations with the measurement of signal arrival times.

Similarly, seismology is the study of the seismic waves of energy caused by the sudden breaking of rock within the earth or by an explosion. The problem of source localisation is one of the key research areas of seismology. Wave modes involved in the seismogram include the P-wave, S-wave, and the two types of surface waves. Frequencies of these waves are usually below human hearing. Various methods have been employed for locating wave sources. One conventional method of calculating the source distance is measuring the arrival time difference

between the p-wave and s-wave modes within one signal measured at the seismograph. Another approach is using a network of seismographs to estimate the TDOA of signals. This is similar to the previously introduced TDOA method, but different in the way it extracts the arrival times or time differences.

One common difficulty in these two applications is the complicated wave propagation, which is reflected, dispersed, scattered, and more. Rather than detailing the two applications completely, this section will cover two selected categories that are related and are of interest to this research work. The two methods are the Time of Flight (ToF) and the time-frequency analysis.

2.4.1 The ToF method

The simplest method of determining the arrival time of a signal is to detect its rising-edge, also referred as Time of Flight (ToF). As introduced before, the estimated ToF, $\hat{\tau}_{thres}$ is the moment at which the waveform crosses a threshold level T. Some applications require cheap solutions that are suitable for direct implementation using dedicated hardware, such as instrumental electronics [Hof72]. Commercial products usually take the first threshold crossing for the arrival time. However, this conventional method is affected significantly by the level of threshold voltage and the average value of velocity measured several times.

In recent years, various approaches have been attempted to improve the reliability of timing signal arrivals [ES94] [SE99] [SHP02] [SHRP04] [KGR05]. Many of these have been performed with an automatic P-phase onset picker approach in a seismogram. A short review of methods of picking the first arriving p-wave is given in [Ald04]. The most popular method was initially introduced by Allen in 1978 and modified by Baer and Kradolfer in 1987. The ratio of a short-term average (STA) and a long-term average (LTA) is used as the criterion to identify the arrival of a signal. This algorithm involves four processes, briefly: enveloping the seismogram, the so-called STA/LTA filter to obtain the ratio between the short window and the long window, smoothing the ratio function, and checking

the threshold crossing of the smoothed ratio function [ES94]. This algorithm checks the ratio between the amplitudes of the short window and long window through the whole seismogram. The drawback of this method is that it encounters difficulties when attempting to identify the change in frequency domain.

In [SHP02], other wave picker approaches have been mentioned, including neural networks, fuzzy logic, wavelet transform, and higher-order statistics. One of the problems with these methods is the inadequacy for real-time implementation due to the computation complexity. This paper also introduced a novel and promising method based on higher-order statistics that overcomes the subjectivity of human intervention and eliminates the noise factor in automatic phase picker algorithms. Briefly, the use of the third and the fourth order statistics moment has provided a more robust arrival detector than the STA/LTA algorithm. However, it has the same problem of identifying the change in spectral domain as the STA/LTA method. Similar work has been performed [YI97], that is based on the so-called bicoherence correlation in the third order domain to pick arrival times in a seismic data set.

Usually, P-wave is predominantly used as the wave mode for many NDT and seismic applications. However, in plates, the dominant wave mode, the Lamb wave, has dispersive velocities, such that the ToF is unsuitable. This problem also arises for the onset detection for a seismogram when Rayleigh waves dominate.

Gorman proposed an alternative method for AE source location in dispersive media[ZG91]. In this approach, a single frequency cosine wave modulated by a Gaussian pulse is prepared first. This single frequency wave is cross-correlated with the transducer outputs to locate the corresponding phase points at that single frequency. The arrival times are then determined by the correlation coefficients. This is of practical importance for locating sources when only dispersive flexural waves are measurable. In contrast, the threshold crossing method and other ToF methods are more suitable for the dispersionless extensional wave, which is usually more attenuated than the low frequency flexural wave.

2.4.2 Time-frequency analyses of dispersive signals

Time-frequency representations, such as the Short Time Fourier Transform (STFT) and the Wavelet Transform (WT), have been used for analysing dispersive signals to good effect [NJQJ01] [KK01]. In this subsection, an overview of time-frequency analysis techniques will be given, followed by a review of current applications of these techniques in the fields of dispersion analysis and localisation.

Basics of time-frequency analysis

The primary purpose of time-frequency analysis is to investigate and understand the changes of frequencies in time. Rather than transforming a temporal signal into the frequency domain with the versatile Fourier transform, time-frequency analysis maps the signals from the form of 1-D in time to a 2-D representation in both time and frequency domains. Figure 5.19 shows an example of time-frequency representation of a signal.

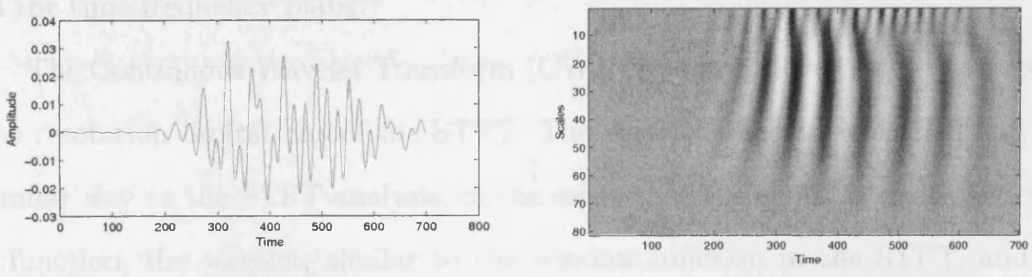
Unlike Fourier-based analyses that use global (nonlocal) sine and cosine functions as bases, time-frequency analysis uses bases that are localised in time and frequency to effectively represent non-stationary signals, which are commonly used in AE and seismic applications. The Short Time Fourier Transform (STFT) and Wavelet Transform (WT) are two popular techniques used for time-frequency analysis in AE and seismography applications.

The STFT of a function f is a Fourier-related transform used to determine the sinusoidal frequency and phase content of local segments of a signal as it changes over time. The mathematical definition can be written as:

$$Sf(u,\xi) = \int_{-\infty}^{\infty} f(t) g^*_{u,\xi}(t) dt = \int_{-\infty}^{\infty} f(t) g(t-\tau) e^{-i\xi t} dt \qquad (2.21)$$

where the real and symmetric window function $g(t) = g(-t)$ is translated by u and modulated by the frequency ξ:

$$g_{u,\xi}(t) = e^{i\xi t} g(t-u) \qquad (2.22)$$



(a) (b)

Figure 2.24: The time-frequency representation (b) of the signal (a).

and $*$ denotes the complex conjugate.

One significant limitation of the STFT technique is its fixed resolution. For a given window $g(t)$ with the time spread σ_t and frequency spread σ_ξ, the time and frequency spreads of the modulated translated window $g_{u,\xi}(t)$ are independent of (u, ξ). The width of the windowing function relates to how the signal is represented - it determines whether there is good frequency resolution or good time resolution. A wider window gives better frequency resolution but poorer time resolution, and a narrower window gives better time resolution but poorer frequency resolution. Usually, the resolution covered by σ_t and σ_ξ is denoted by a rectangle, which is referred to as a Heisenberg box. Figure 2.25 represents the two examples with better frequency resolution (a) and better time resolution (b) in the time-frequency plane.

The Continuous Wavelet Transform (CWT) was introduced later to overcome the resolution limitations of the STFT. The wavelet analysis is carried out in a similar way to the STFT analysis, in the sense that the signal is multiplied with a function, the wavelet, similar to the window function in the STFT, and the transform is computed separately for different segments of the time-domain signal. However, the most significant characteristic is that the width of the window is changed as the transform is computed for every single spectral component. It can give good time resolution for high-frequency events, and good frequency resolution for low-frequency events. This is the best suited type of analysis for many dispersive signals. The CWT of $f(t)$ is defined as:

$$\begin{aligned} Wf(u,s) &= \int_{-\infty}^{\infty} f(t)\psi^*_{u,s}(t)dt \\ &= \frac{1}{\sqrt{s}}\int_{-\infty}^{\infty} f(t)\psi^*(\frac{t-u}{s})dt \end{aligned} \tag{2.23}$$

where u is the translation parameter representing the shift parameter of the wavelet $\psi(t)$ in the time domain, and s is the dilation or scale parameter, which

is related to the frequency. The relationship between the scale variable s and frequency ξ is $\xi = \xi_0/a$, where ξ_0 is the central frequency of the mother wavelet $\psi(t)$. The normalisation factor $\frac{1}{\sqrt{s}}$ is included so that $\| \psi_{u,s} \| = \| \psi \|$.

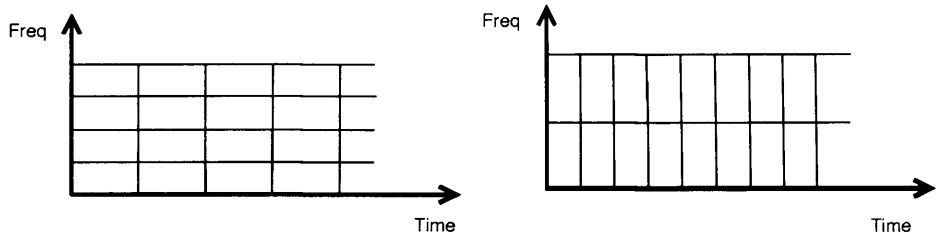
Increasing the scale (decreasing frequency) results in increased window width, so that the transform starts picking up the lower frequency components, and vice versa, as depicted in figure 2.26.

The algorithm of the CWT is heavily computational and places redundant information on the time-frequency plane. By taking into consideration the reconstruction of the original signal, the Discrete Wavelet Transform (DWT) provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in computation time. The DWT is a special case of the CWT that provides a compact representation of a signal in time and frequency that can be computed efficiently. In brief, rather than computing all scales in the CWT, the DWT does not choose continuous scales and translations, but rather chooses discrete steps. Detailed coverage of this topic is not of significant interest here, and can be found in a number of books and articles on this topic [GC99].

Applications of time-frequency representation in dispersive wave analyses and localisation

Applications of dispersion analysis with time-frequency representation are numerous from dispersion analysis to source localisation.

Various techniques for time-frequency representation of wave dispersion have been tested and discussed. The efficiency of the CWT in the analysis of impact induced transient wave propagation along a beam was demonstrated in [OH94]. In addition, the STFT and the CWT have been compared in representing the dispersion of an elastic flexural wave generated by an impact in a solid circular cylinder [KK01]. The experimental results have confirmed the theoretical analysis, that the CWT outperforms the STFT in representing clearer instantaneous frequencies with the ridge analysis. Two other time-frequency analysis tech-



(a) (b)

Figure 2.25: Comparison of STFT resolution. (a) has better frequency resolution, and (b) has better time resolution.

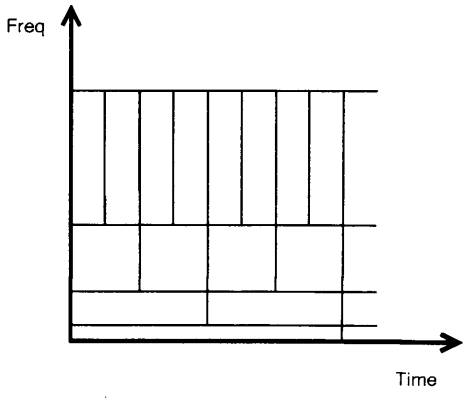


Figure 2.26: CWT resolution.

niques, called the smoothed pseudo-Wigner distribution and the Hilbert-Huang transform, were also compared with the CWT in [ADH+06] for signals induced by impacts on beams. In conclusion, the CWT still appeared to be the best choice for its ease of implementation, smaller computational burden, and finer resolution in identifying the ridges.

The relationship between the group velocities of different wave modes and the CWT time-frequency plane has been discussed in a number of papers, stating that the maximum at each scale indicates the arrival time at the group velocity of the corresponding frequency. With such knowledge, the authors of [HKWS00] have carried out work to determine the distance between the source and the receiver on a plate by matching the inverse ridge curves in the CWT domain. The work precluded automatic matching due to its complexity, and was therefore conducted manually. This paper emphasised the contribution of using a single signal only for the source detector separation determination. Another work has also shown the possibility of detecting the distance between the source and receiver on a plate with one channel [JHW+04]. The principle is simple, that two distinct peaks in the CWT domain may indicate the arrival times of the extensional and flexural waves, which propagate at different velocities. With knowledge of the velocities, the distance between the source and the receiver can be determined. It should be noted that both works are only useful when the signal has a very broad frequency band, which contains higher modes of Lamb waves.

Inverse modelling of dispersive seismic wave propagation has been conducted in the CWT domain in [KHD+05]. A mathematical model was established to link two seismic signals received at two stations with the knowledge of wave dispersion and attenuation. To estimate the propagated signals with good accuracy from one prior known signal was shown to be possible.

Returning to the source localisation problem, some have performed experiments on localising impacts on plates with the CWT [GH97] [JJ00] [HOG02] [WTAT05] with a few (usually 4) transducers at known positions. The basic

principle is to determine the arrival times at the corresponding group velocities for different frequencies (the scales) in the CWT domain. The impact location can then be estimated by the TDOA method. In [GH97] [JJ00] [HOG02], the maximum in the CWT plane at a scale (frequency) with a time shift indicates the arrival time at the corresponding group velocity. By contrast, the authors of [WTAT05] used the first distinct maximum in the CWT domain as the arrival time. In [HOG02], two sets of simulated data on big and small specimens were examined and compared, concluding that the big plate gave higher accuracy than the smaller plate, due to the considerable edge reflections in the smaller specimen.

Another technique called the wavelet packet transform has been used to localise the AE source position in an arbitrary structure, and compared with the conventional ToF and cross-correlation methods [DRS04]. DWT has also been applied for time difference estimates with two approaches in [BT03]. One is a simple use of the DWT as the pre-filter before applying the GCC, and the other one is to calculate the time difference via the discrete wavelet domain inner product. The result has shown the advantages of higher accuracy for the simulated highly transient signals.

2.5 Summary

This chapter has covered background material relevant to this research. In particular, the chapter has reviewed the state-of-the-art in the area of multimodal tangible interfaces, the acoustic characteristics of vibrations in solids, and different approaches for acoustic localisation.

Chapter 3

Location Template Matching method

Although most of the source localisation techniques are based on the TDOA or other similar principles, as reviewed in chapter 2, time-reversal as an alternative has shown its great importance in related research areas. Location Template Matching (LTM), derived from the time-reversal theory, is a method of locating impact source positions by means of only one or more sensors. One advantageous feature of this approach is that the sensor locations need not be known beforehand.

This chapter will first review the time reversal theory. The LTM method will then be described, linked between the time reversal and the cross-correlation function. The second section discusses experiments with the LTM method. The next section introduces a technique to improve the performance of the initial LTM approach with an additional pre-filter to the cross-correlation function. Finally, further improvement of the LTM method is proposed.

3.1 Time reversal review and Location Template Matching

3.1.1 Acoustic time reversal

Acoustic time reversal theory has, in recent years, shown significant importance in the area of acoustic source localisation [FCD+00]. The time reversal theory is based on the reciprocity feature of the wave equation, stating that if one has a solution to the wave equation, then the time reversal of that solution is also a solution. The theory behind this statement can be understood from the standard wave equation that contains only even order derivatives. In addition, the medium is assumed to be approximately linear in the time reversal theory.

In the time reversal experiment, an array of transducers configured as a mirror or cavity is used to record a signal at the first stage. The transducer array is called a time reversal mirror (TRM) and a time reversal cavity (TRC), depending on the layout. The emitted source signal is usually a short pulse-like signal generated from a point position. The received signals are first recorded in the memory of computers, and will be time-reversed and re-emitted back from the transducers to the same environment or medium. The re-emitted signals will be propagating backwards as the time goes back and focusing at the original source location with a high peak. Theoretically, a time reversal experiment is able to discover the multipath of wave propagation, by tracing back the wave source position, which could have been reflected, refracted, and scattered. The field of displacement can be considered as a superposition of all these waves.

In the simplest configuration of the time reversal experiment, assume that there is only one receiver, which is located at position r. A signal s, emitted from the source position r_0, can be received by the sensor at time t. Moreover, when the source signal is assumed as a simple impulsive excitation $v(t) = \delta(t)$, the obtained signal at position r can be considered as a linear function of the impulse response of the propagation $h(r, t)$.

Function G, known as Green's function, is usually used to represent the wave field at any other position and at any subsequent time. The wave function can then be expressed as the integral of Green's function weighted by the source distribution. With the source position r_0 and the time t_0, Green's function is denoted by $G(r_0, t_0|r, t)$ expressing the disturbance at the position r and time t. As this disturbance will only depend on the elapsed time $t - t_0$, this expression can also be formulated as

$$G\left(r_0, t_0|r, t\right) = G\left(r_0, 0|r, t - t_0\right) \tag{3.1}$$

Usually the transducer geometries need to be considered for analytic solutions [Fin92]. However, this would be very complicated to obtain in practice and, in fact, is not necessary for this research in order to calculate any impulse response in any media. Here, it is only necessary to apply one of the properties of time reversal, that the impulse responses in the transmit mode and in the receive mode are identical in almost any lossless media, which could be inhomogeneous or homogeneous.

Ignoring the transducer geometry effect, the wave function $h(r, t)$ at position r corresponding to the initial source signal, which is assumed as a Dirac Delta function $\delta\left(t\right)$, can be expressed as the integral of the Green's function

$$h(r, t) = \int G\left(r_0, t_0|r, t\right) dr \tag{3.2}$$

The reciprocity theorem is an important theory stating that the source position and the receiving position can be exchanged, so that the same signal s can be obtained at the new receiving position, which is originally the source position. Thus, the Green's function has the following relation $G\left(r_0, t_0|r, t\right) = G\left(r, t_0|r_0, t\right)$.

On the other hand, the function $h\left(r_0, t\right)$ denoting the wave from the receiver position r to the source position r_0 can be formulated as an integral of Green's

function $G(r, t_0|r_0, t)$:

$$h^t(t) = \int G(r, t_0|r_0, t)\, dr \tag{3.3}$$

This function implies that the propagations of an impulsive signal between two locations are identical, that $h^t(t) = h(t)$.

Since it is impossible to really inverse the signal in time, a finite length of signals is used instead, denoted with time of sample T. Thus, the time-reversed signal is denoted as $h(T-t)$. In general, the length T can be neglected, and the process of time-reversal becomes $h(t) * h(T-t) = h(t) * h(-t)$, which is exactly the convolution of the response function.

In both of the TRM and TRC experiments, there are a number of transducers used in an array. With the linearity assumption of the system, the superposition principle can be applied. Assuming the number of transducers is N, the refocused signal s at the source position r_0 is the summation of all the emitted signals from the N transducers

$$s(r_0, t) = \sum_{i=1}^{N} h_i(t) * h_i(-t) \tag{3.4}$$

To be more rigorous, the acoustic-electric responses of the transducers need to be taken into account. In the transmit mode, $h^{tae}(t)$ relates the electrical input to the normal velocity output, and, by contrast, $h^{rae}(t)$ represents the response of the acoustic force input to the electrical output. In general, these two responses are identical, and can be expressed as $h^{ae}(t)$. The complete representation of the above formula can then be given as

$$s(r_0, t) = \sum_{i=1}^{N} h_i^{ae}(t)(t) * h_i(t) * h_i^{ae}(-t) * h_i(-t) \tag{3.5}$$

It can then be briefly summarised that, in a lossless environment, the refocused

signal is in fact the convolution of the impulse response function and its time-reversed version.

3.1.2 Location Template Matching

According to the time reversal theory described above, it is possible 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 Location Template Matching (LTM) approach, the unique feature of the received acoustic signal as a function of its source location, which is based on the time reversal theory, is employed to localise an impact on an object. This is achieved by creating a template of received signals generated from taps mapped to their known corresponding locations [Yan06] [IQCF05] [PAKY+06]. As mentioned before, the time reversal method is a process of convolution between signal and its time reversed version. This is identical to the cross-correlation function. Here, the template signals are used as the filters, and the process of recognising signals is therefore a process of finding the most suitable filter that can maximise the similarity. In order to compare the similarity between signals, cross-correlation needs to be normalised in terms of energy to ensure the correlation coefficient is between -1 and 1. Therefore, the maximum of the correlation coefficients can be used as the measurement of similarity.

Cross-correlation is a technique widely used to measure similarity between two signals. For example, cross-correlation has been applied to match unknown test ECG signals with template ECG signals in a database of known diagnoses designed to determine cardiac information [BK00]. In image processing applications, a finger has been tracked for real-time gesture interfaces by two-dimensional cross-correlation matches between the target image and the template images [OZ97].

However, this has not been applied in the application of human computer interaction with the method of acoustic pattern matching.

Considering a template acoustic signal $s_0(t)$ and an acoustic signal to be recognised $s_k(t)$, the correlation coefficient of signals s_0 and s_k can be written as

$$r_{s_0 s_k} = \frac{cov(s_0, s_k)}{\sigma_{s_0}\sigma_{s_k}}, k = 1, 2, 3 \ldots \tag{3.6}$$

where the $cov(s_0, s_k)$ is the covariance between signal s_0 and s_k, and σ_{s_0} and σ_{s_k} are the standard deviations of signals s_0 and s_k.

The hardware setup is illustrated in figure 3.1, comprising a piece of hardboard, one sensor located at the bottom right of the board, signal conditioning hardware, a PCI data acquisition card and a PC. A knock on the hardboard is detected by the sensor. The received signal is amplified and filtered by the signal conditioning hardware, digitised by the PCI data acquisition card and processed by the PC, using the cross-correlation analysis.

The flowchart of the LTM method is illustrated in figure 3.2.

From the figure, it can be seen that the LTM method involves two phases: the learning phase and the testing phase. In the learning step, a group of signals, generated by impacts with a point-like object at several pre-defined positions, are recorded as a template. In the testing step, one can tap at one of those defined positions. By comparing with the recorded reference signals, the most similar referenced signal can be found, thus the corresponding position can be determined.

Taking into account the external noises or signals from unmapped positions, a threshold is needed as the criterion for this purpose. Any correlation coefficient below the threshold will be considered as a false trigger. The value of the threshold will be decided by experiments and by the resolution required for the application.

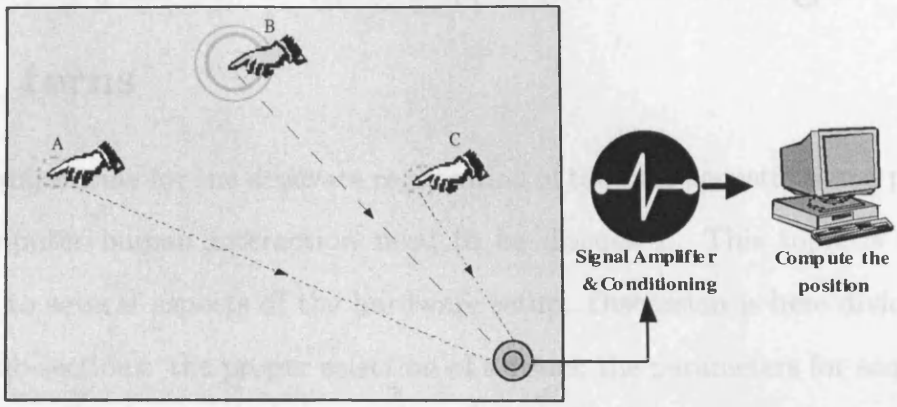


Figure 3.1: Hardware setup of LTM method

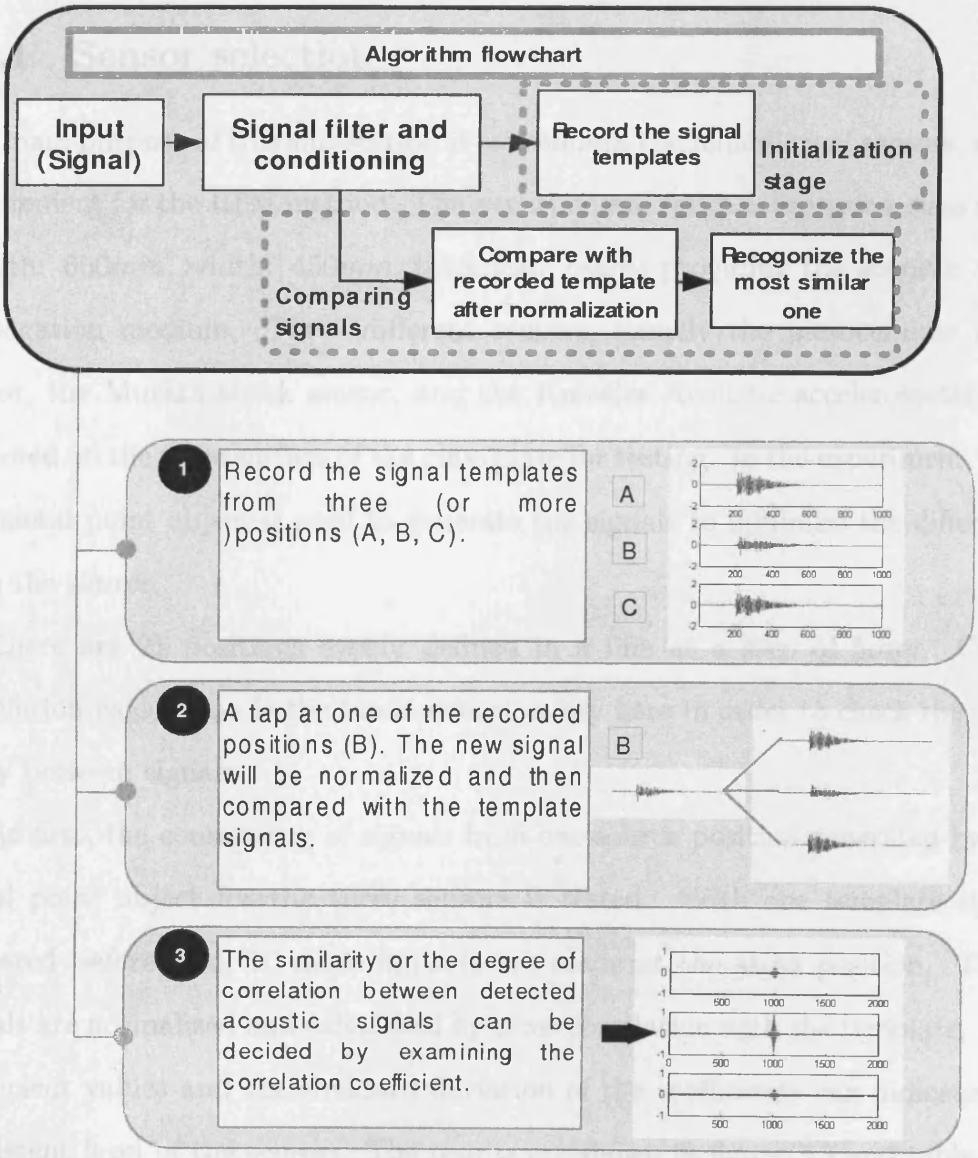


Figure 3.2: LTM method flowchart

3.2 Acquisition and registration of signal patterns

The requirements for the accurate registration of tangible acoustic signal patterns for computer human interaction need to be discussed. This topic is directly related to several aspects of the hardware setup. Discussion is here divided into three sub-sections: the proper selection of sensors; the parameters for acquisition and registration of signal patterns; and signal consistency testing for different impacts.

3.2.1 Sensor selection

The main purpose of this sub-section is to evaluate the reliability of sensors, a key requirement for the LTM method. The experimental setup comprises a glass plate (length: $650mm$, width: $450mm$, thickness: $5mm$) providing the acoustic wave propagation medium. Three different sensors, namely the piezoceramic plate sensor, the Murata shock sensor, and the Knowles Acoustic accelerometer, are mounted on the same surface of the glass plate for testing. In the experiment, only one metal point object is used to generate the signals to minimise the difference from the source.

There are 21 positions evenly defined in a line at a step of $5mm$. Cross-correlation calculation is the fundamental utility here in order to check the similarity between signals.

At first, the consistency of signals from one source position generated by the metal point object for the three sensors is tested. With one template signal prepared beforehand, 10 more impacts are made at the same position. These signals are normalised and calculated by cross-correlation with the template. The coefficient values and the standard deviation of the coefficients can indicate the consistent level of the sensors. The results are shown in figure 3.3 and table 3.1.

Next, the template signal from the centre position is used as the reference

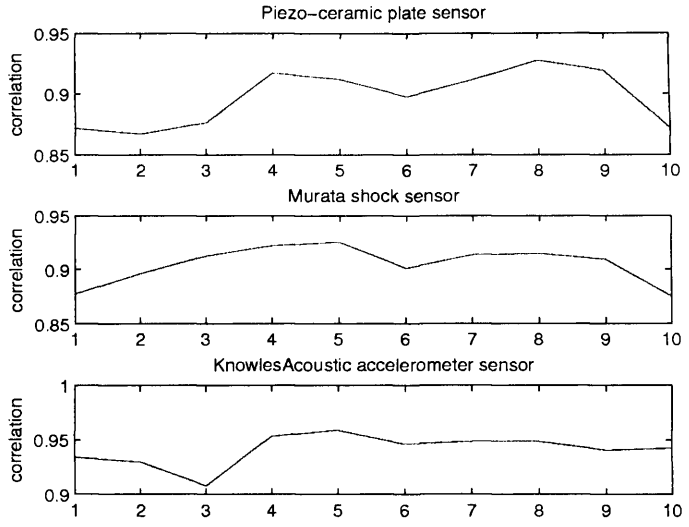


Figure 3.3: Correlation coefficients of signals from one position.

Sensor	Piezoceramic	Murata	Accelerometer
std	0.0039	0.0040	0.0100

Table 3.1: Standard deviation of the correlation coefficients for three sensors.

signal. The cross-correlation between this template and all other signals is calculated.

Figures 3.4, 3.5, 3.6, 3.7, 3.8, and 3.9 show the results from 8 of the 21 positions, located $\pm 50mm$, $\pm 30mm$, $\pm 10mm$, and $\pm 5mm$ from the centre position. At each position, 11 samples were prepared. The corresponding correlation coefficients are shown in the figures.

It can be seen that the coefficients of the positions that are $50mm$ and $30mm$ from the centre position are mostly under 0.5, which is in contrast with the values of figure 3.3. The correlation coefficients are increasing significantly as the positions get closer to the target centre position.

The stability of the sensors can be seen with low standard derivations of the coefficients. Figure 3.10 depicted the results, showing that there are no significant differences in receiving repeatable and reliable signals.

Figure 3.11 shows the coefficients with all tested signals around the centre template. The centre position is located at the position of $0mm$. At every $5mm$ from the centre, 10 samples are prepared and calculated for the correlation analysis. Averaged coefficients are taken as the final value.

As above, all three sensors performed similarly well, with coefficients decreasing as the position gets further. For the accelerometer, the coefficient decreases more sharply than the other two sensors, although there are two local peaks at about $-40mm$ and $40mm$, which might be ambiguous in terms of recognising the target template. However, compared to the centre position, the coefficient values at the two positions are still low enough. The sharper decrease of the coefficients is in fact partially because of the wider frequency response of the sensor and the wavelength of the signals. This contrast is directly related to the highest resolution that can be achieved and will be explained in the next chapter.

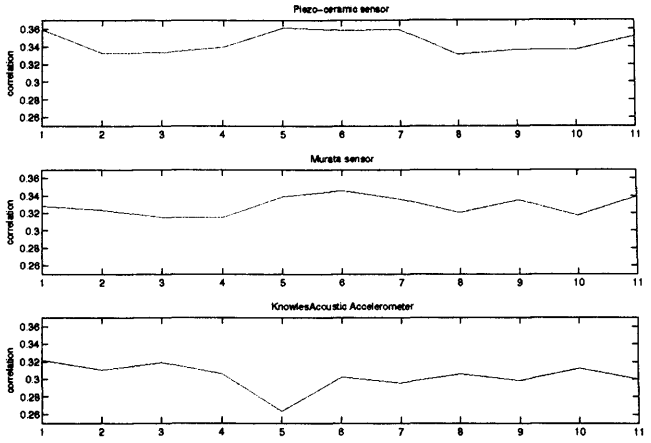


Figure 3.4: Correlation between template at the middle position and signals from the position 50 mm to the left.

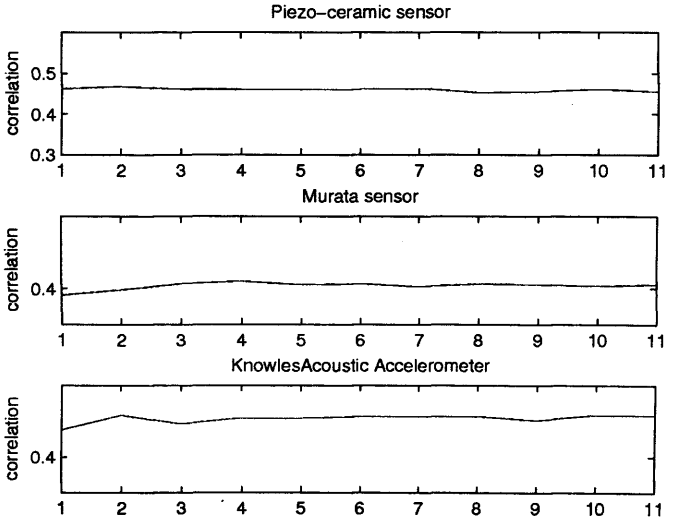


Figure 3.5: Correlation between template at the centre position and signals from the position 50 mm to the right.

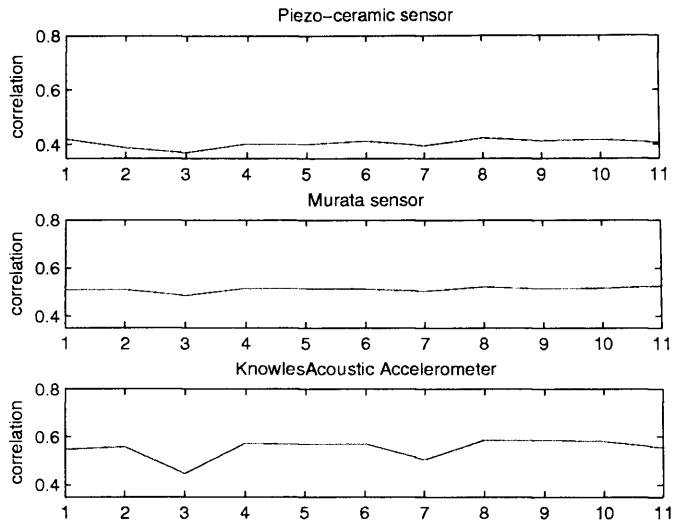


Figure 3.6: Correlation between template at the centre position and signals from the position 30 mm to the left.

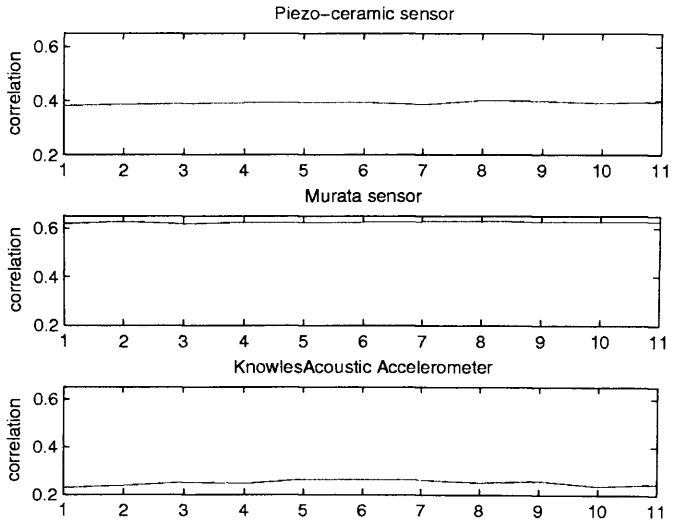


Figure 3.7: Correlation between template at the centre position and signals from the position 30 mm to the right.

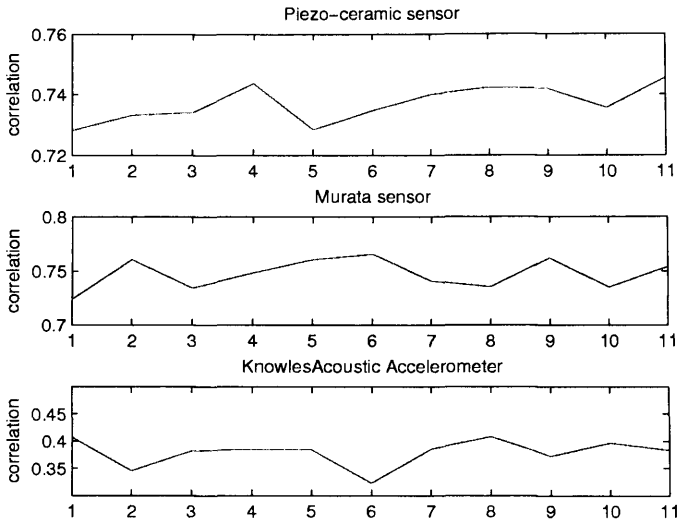


Figure 3.8: Correlation between template at the centre position and signals from the position 10 mm to the left.

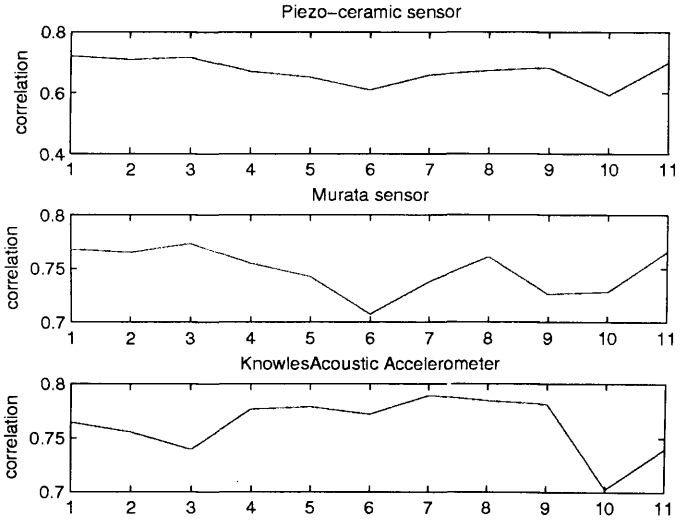


Figure 3.9: Correlation between template at the centre position and signals from the position 10 mm to the right.

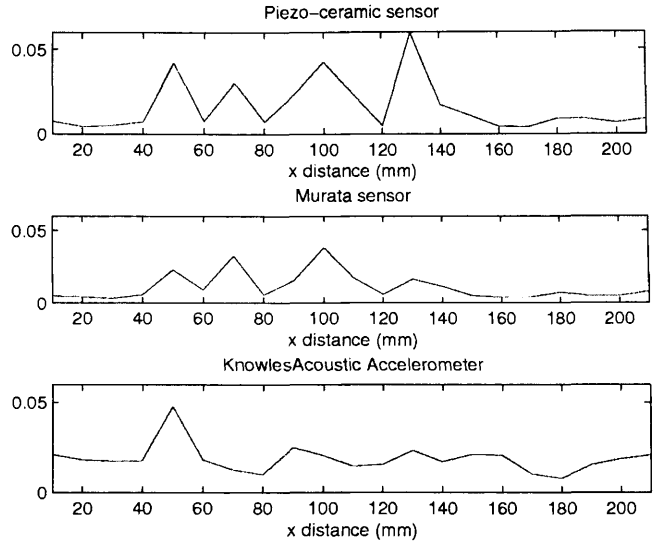


Figure 3.10: Standard derivations of correlation coefficients at all positions.

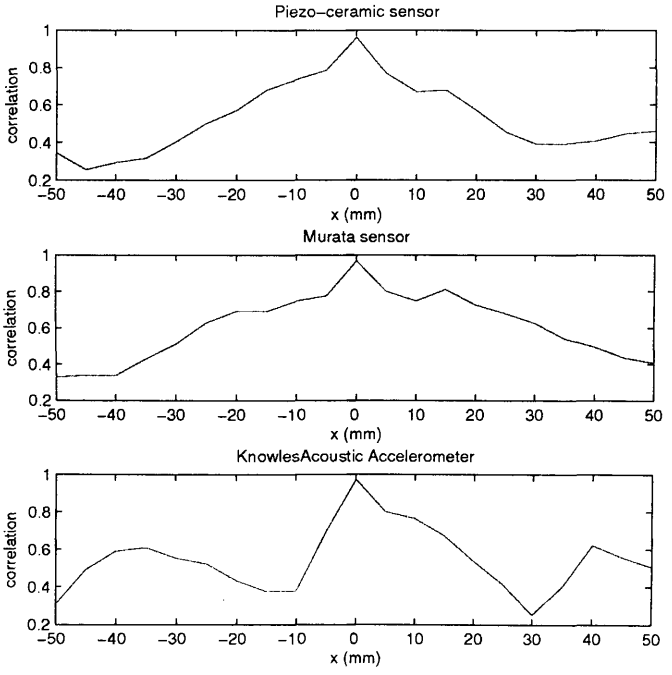


Figure 3.11: Averaged coefficients at all positions with the template at the centre position.

3.2.2 Signal acquisition parameters

Parameters such as signal sampling rates and recorded signal lengths and the use of filters before computing correlation coefficients are all important for signal pattern registration accuracy [PWJ+05]. The number of features in a pattern is directly proportional to its sampling rate and recorded signal length, and the signal amplitude accuracy is confined by its digitisation resolution levels or bits. In a particular application, the registration accuracy can be considered together with the sampling rate, signal recorded length and the object properties.

In order to precisely express impact patterns, high resolution and high speed analogue to digital converters are required for the registration of acoustic signal patterns. However, considering the registration accuracy and real-time processing requirements in a particular application, a proper selection of data registration formats is necessary. These aspects will now be discussed along with experimental results.

In this experiment, a Knowles Acoustic accelerometer was used as an acoustic sensor, due to its high sensitivity and wider bandwidth. Sampling rates between $1KHz$ and $500KHz$, amplitude digitisation resolution levels between 8 and 16 bits, and recorded signal lengths between $5ms$ and $60ms$ were used to acquire the acoustic signals. The correlation coefficients and their deviations were obtained with 50 signal samples.

Figure 3.12 shows three acoustic signal patterns recorded with sampling rates of 500 KHz, 31 KHz and 1 KHz. In figure 3.12(c), it can be seen clearly that the high frequency components were lost when the low sampling rate was used. The amplitude digitisation resolution employed in the experiments was 16 bits. The recorded signal length was 60 ms. Figure 3.13(a) shows the correlation coefficient distribution when the sampling rate was 1 KHz, and figure 3.13(b) depicts the standard deviations of correlation coefficients with sampling rates from 1 KHz to 500 KHz. It is clear that the standard deviation increases with a decrease in the sampling rate.

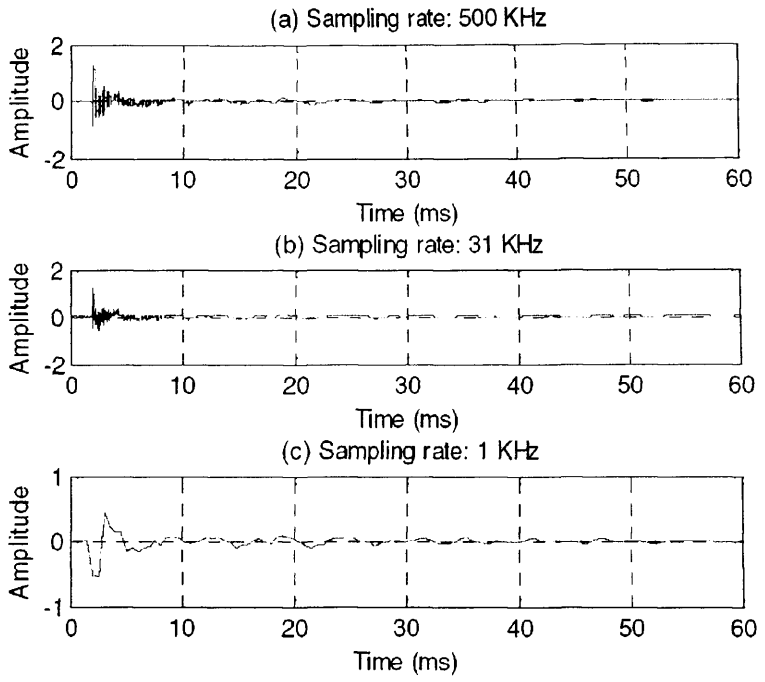


Figure 3.12: Three recorded acoustic signal patterns with sampling rates of 500 KHz, 31 KHz and 1 KHz.

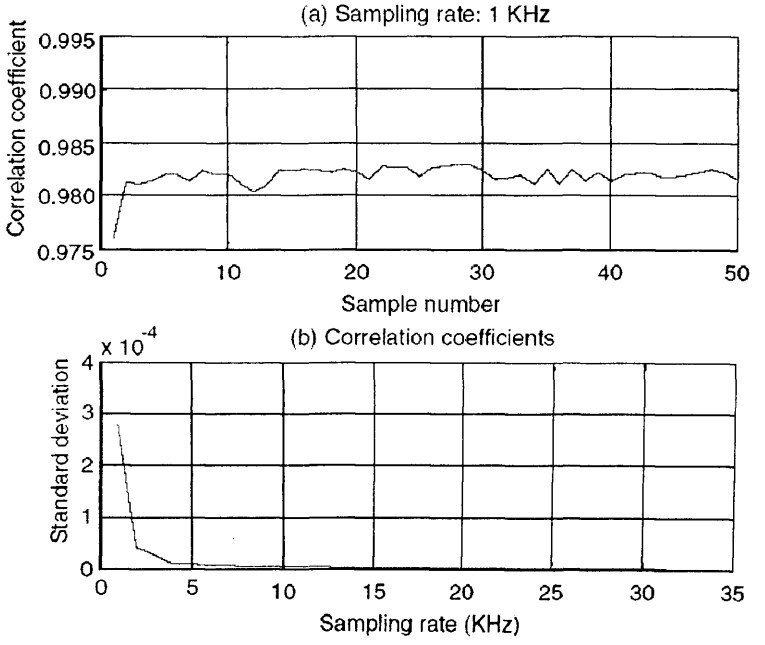


Figure 3.13: (a) correlation coefficient distribution when sampling rate = 1 KHz, and (b) standard deviations of correlation coefficients with sampling rates from 1 KHz to 31 KHz.

Figure 3.14 shows three recorded acoustic signal patterns with digitisation resolution levels of 16, 12 and 8 bits. The sampling rate used in the experiment was 500 KHz. The recorded signal length was 60 ms. Figure 3.15(a) shows the correlation coefficient distribution when the amplitude digitisation resolution was 8 bits, and figure 3.15(b) plots the standard deviations of correlation coefficients with digitisation resolution levels from 8 to 16 bits. It can be seen that the standard deviation increases with a decrease in the amplitude digitisation resolution.

Figure 3.16 shows three recorded acoustic signal patterns with signal lengths of $60ms$, $30ms$ and $5ms$. The sampling rate used in the experiment was $500KHz$. The amplitude digitisation resolution was 16 bits. Figure 3.17(a) shows the correlation coefficient distribution when sample length was $5ms$, and figure 3.17(b) shows the standard deviation of correlation coefficients with sample lengths from $5ms$ to $60ms$. It can be seen from figure 3.17(b) that the longer recorded signal length gave higher accuracy.

This research has shown that high accuracy of the registration of acoustic signal patterns can be achieved through correct selection of the sampling rates, amplitude digitisation resolution levels and recorded signal lengths. The experimental results have confirmed that the standard deviation of the correlation coefficient can be reduced by increasing the sampling rate, amplitude digitisation resolution and recorded signal length. Correct selection of data registration parameters can be made based on the registration accuracy and real-time processing requirements for a particular application.

3.2.3 Signal consistency

The above experiments have demonstrated the high consistency of signals induced by one object from one source position. However, it is in fact not possible to reproduce the same signal with different objects, such as a fingertip, nail, or pen point, used in the interaction with computers. Thus, how to extract the key feature that is common to all the signals from one source position now becomes

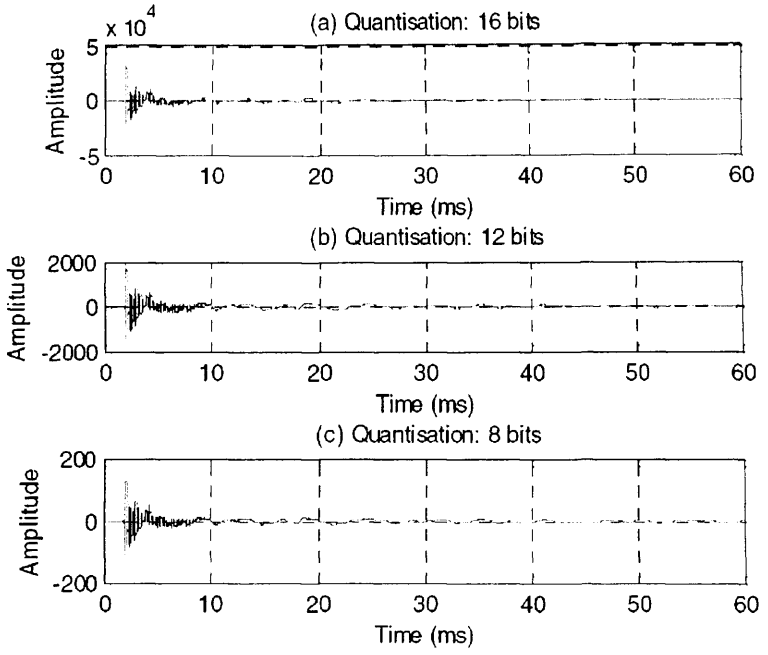


Figure 3.14: Three recorded acoustic signal patterns with digitisation resolution levels 16, 12 and 8 bits.

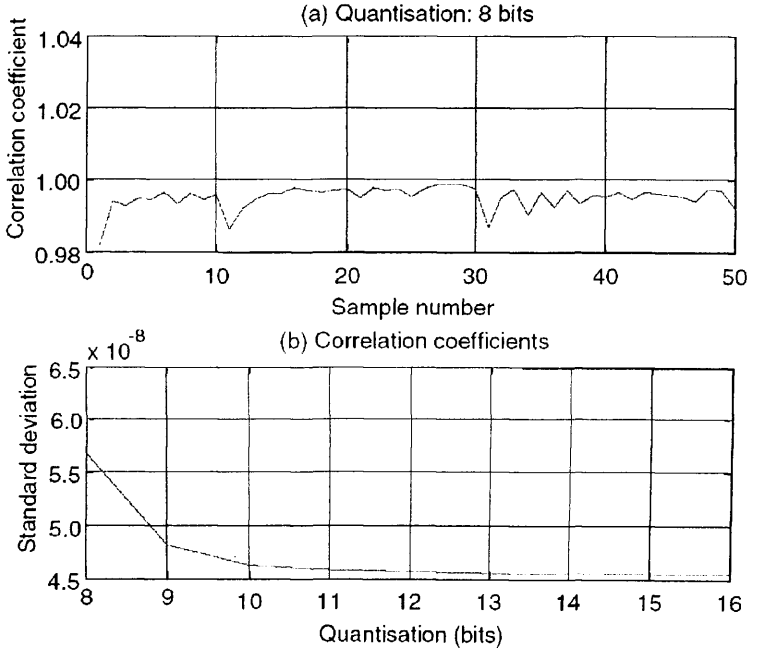


Figure 3.15: (a) correlation coefficient distribution when digitisation resolution = 8 bits, and (b) standard deviation of correlation coefficients with digitisation resolution levels between 8 and 16 bits.

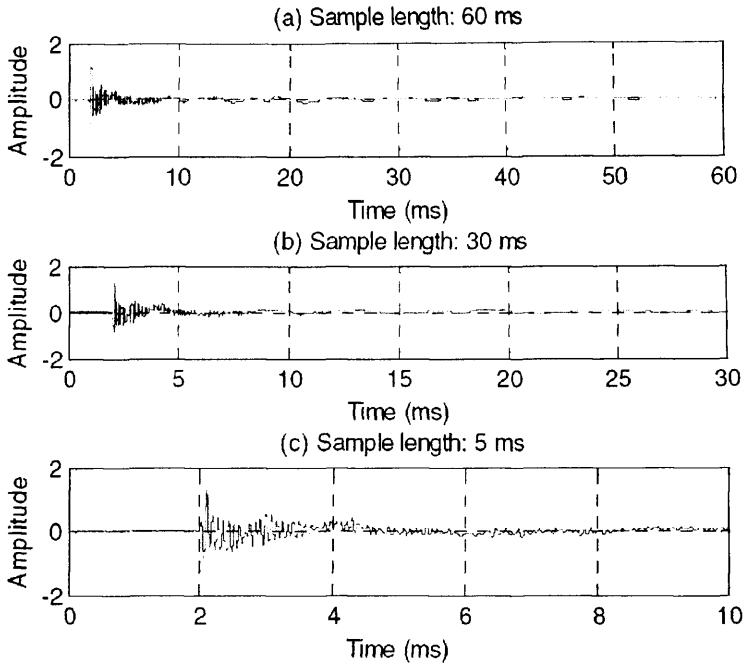


Figure 3.16: Three recorded acoustic signal patterns with different signal lengths.

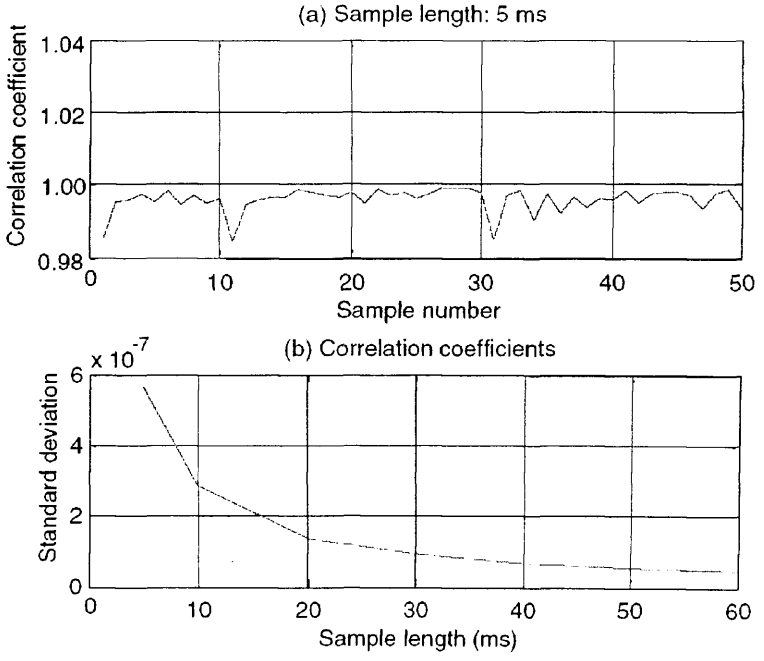


Figure 3.17: (a) correlation coefficient distribution when sample length = 5 ms, and (b) standard deviation of correlation coefficients with sample lengths from 5 ms to 60 ms.

important. A few conditions to be satisfied for such a purpose can be summarised as below [BdVB80]:

1. The generated signal must contain enough energy, such that high signal noise ratios occur at all frequency components of interest;

2. The source must produce a short, simple, pulse-shaped signal (i.e., no significant pressure oscillations may occur outside the main lobe);

3. The source must be repeatable along time and space, i.e.,

   (a) The generated time signal must be repeatable.

   (b) The directivity characteristic must be repeatable (preferably omnidirectional).

The sensor tests in appendix B have shown the clear ratio between the energy of the signal and the background noise. The source signals, generated by tapping with nails and fingertips, are assumed to be short enough. In order to measure or estimate the direct force that is applied on the surface by the nail or finger tapping, some use specially designed hammers. A force sensor is attached at the end, so that forces of impact can be measured directly. However, this is not suitable for this project, which considers different types of impact, generated by nails, pen points, or fingertips. Due to the inconvenience of measuring the repeatability of the tapping at the source position, here the experiment is designed to check the consistency of the received signals between the source position and the receiver.

Experiment design: In addition to the glass used before, another glass plate with similar dimensions($570 \times 510 \times 4mm$) is tested for the purpose of comparison. Here, the first plate is named *plate 1* and the new glass plate as *plate 2*.

First, a signal, generated by an impact with a pen point, from one mapped position is used as a template. The template is obtained using a pen point because

of the small rigid contact area. This is supposed to be able to produce a neater and shorter impulse, so that the received signal is expected to be closer to the real propagation transfer function between the sensor and the source position. Also, a shorter signal, closer to the Delta function $\delta(t)$, will contain wider frequency bandwidth. That means a pen point is better for producing a signal template with more complete information than that generated by nails or finger tips. Although it is possible to create separate templates for different objects, it is not practical to cover all types. This would need a lot more computation and is, therefore, too time-consuming for real-time applications.

Three groups of signals, induced by pen points, nails, and finger tips respectively, are then prepared for testing. Each group contains 40 signal samples. These signals are normalised and calculated by cross-correlation with the template signal for comparison. The standard deviation of the maximum coefficients can indicate the consistency level of the source signals and the sensor characteristic statistically.

Results: The results are shown in figures 3.18, 3.19 and 3.20. The solid line represents the maximum correlation coefficients for *Plate 1*, and dashed line for *Plate 2*. For both materials, the maximum correlation coefficients for those nail and finger tip samples are lower than the pen samples. The finger tip samples are more uncorrelated with the template. In addition, as shown in the graph, there are more fluctuations in the coefficients distribution for *Plate 2*. Moreover, the correlation coefficients for *Plate 1* are higher than for *Plate 2*.

Table 3.2 shows the standard deviations of the coefficient values in the figures. *Plate 1* has lower standard deviations for all three different types of tapping. The fingertip signals are more unreliable due to their high standard deviations for both materials.

It is obvious that the cross-correlation method is not always reliable when different objects are applied to generate the vibrations. On the other hand, material properties and support structures can also make a difference to the

results. All these factors contribute to the unreliability of the cross-correlation method.

Discussion on the effect of noise: The unreliability caused by external noises on the cross-correlation function can be quantified as below. Two measured signals x_1 and x_2, in the presence of the background noise, can be modelled as:

$$x_1(t) = s_1(t) + n_1(t) \tag{3.7}$$

and

$$x_2(t) = s_2(t) + n_2(t) \tag{3.8}$$

where $x_1(t)$ and $x_2(t)$ are the received signals, which include the real signals $s_1(t)$ and $s_2(t)$ and the external uncorrelated noises $n_1(t)$ and $n_2(t)$. As the noise at two measurements are assumed uncorrelated with each other and with the signals, the cross-correlation function of $x_1(t)$ and $x_2(t)$ can be approximated as:

$$r_{x_1x_2}(\tau) = r_{s_1s_2}(\tau) \tag{3.9}$$

Therefore, the cross-correlation coefficient $\rho_{x_1x_2}(\tau)$ can be given as:

$$\begin{aligned} \rho_{x_1x_2}(\tau) &= \frac{r_{x_1x_2}(\tau)}{\sqrt{r_{x_1x_1}(0)}\sqrt{r_{x_2x_2}(0)}} \\ &= \frac{\rho_{s_1s_2}(\tau)}{\sqrt{(1+\sigma_{n_1}^2/\sigma_{s_1}^2)(1+\sigma_{n_2}^2/\sigma_{s_2}^2)}} \end{aligned} \tag{3.10}$$

where $\rho_{s_1s_2}(\tau)$ represents the correlation coefficient of signals $s_1(t)$ and $s_2(t)$, and $\sigma_{n_1}^2$, $\sigma_{s_1}^2$, $\sigma_{n_2}^2$ and $\sigma_{s_2}^2$ are the variances of $n_1(t)$, $s_1(t)$, $n_2(t)$ and $s_2(t)$ respec-

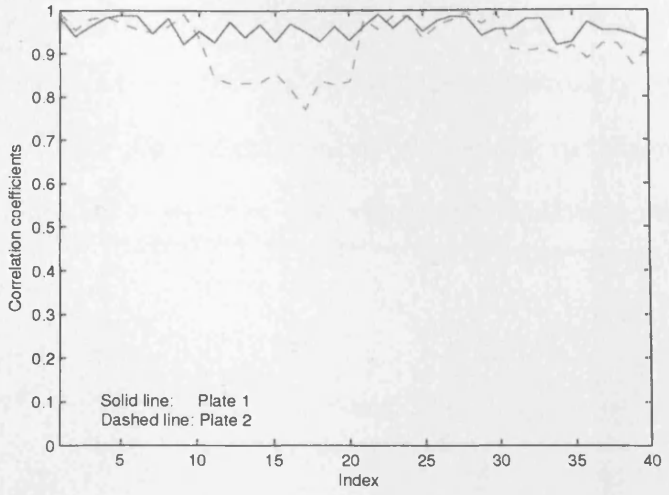


Figure 3.18: Correlation coefficients between a template and 40 signal samples generated with a pen point. The solid line shows the result of *Plate 1*, and the dashed line for *Plate 2*.

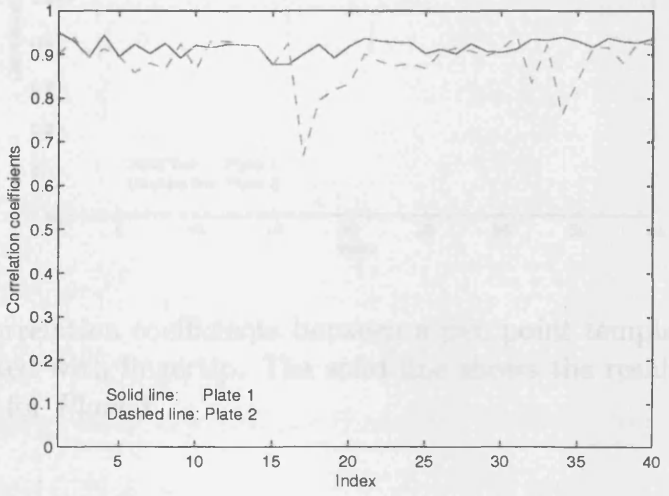


Figure 3.19: Correlation coefficients between a pen point template and 40 signal samples generated with nail. The solid line shows the result of *Plate 1*, and the dashed line for *Plate 2*.

	Pen points	Nails	Finger tips
Plate 1	0.0220	0.0176	0.0545
Plate 2	0.0621	0.0531	0.1420

Table 3.2: Standard deviations of correlation coefficients for two glass plates with three different types of tapping.

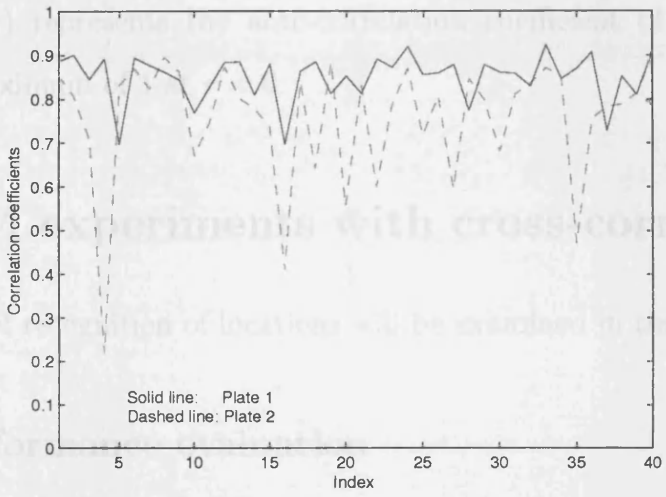


Figure 3.20: Correlation coefficients between a pen point template and 40 signal samples generated with fingertip. The solid line shows the result of *Plate 1*, and the dashed line for *Plate 2*.

tively.

For the experiments using the pen point, the source impact can be assumed to be the same as $s_1(t) = s_2(t) = s(t)$, and the unreliability can, therefore, be considered as the direct effect of external noise. Assume the noise levels at the two signals are the same $\sigma_1 = \sigma_2 = \sigma$, the correlation coefficient can be formulated as:

$$\rho_{x_1x_2}(\tau) = \frac{\rho_{ss}(\tau)}{1 + \sigma_n^2/\sigma_s^2} \tag{3.11}$$

where $\rho_{ss}(\tau)$ represents the auto-correlation coefficient of signal s, which achieves its maximum of 1 at $\tau = 0$.

3.3 LTM experiments with cross-correlation

The accuracy of recognition of locations will be examined in this section.

3.3.1 Performance evaluation

The aim of the following experiments is statistically to examine the accuracy and robustness of the LTM method.

As shown above, with experiments that involve generating signals by applying natural impacts, such as finger tips, nails or pen points, it is not possible to reproduce the same data with different objects. Figure 3.21 shows an example of correlation coefficients for the three different impacts at the same location. As above, the template was prepared using a pen point impact. As expected, the three objects produced very different signals, and directly degraded the correlation coefficients, especially for the finger tip tap.

The experiments were performed as follows. With the template signals prepared by a pen point, the data for testing was prepared by applying impacts at each mapped location. In this experiment, the mapped locations were in the cen-

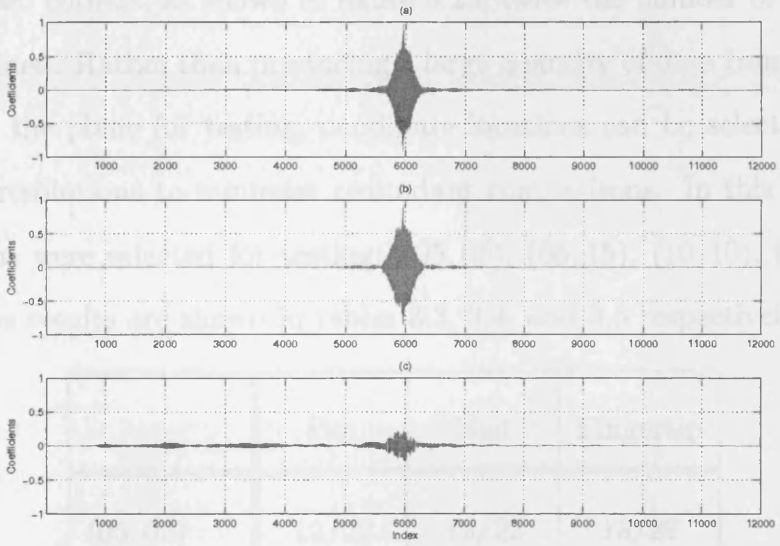


Figure 3.21: LTM coefficient distributions for three different impacts using pen points(a), nails(b), and finger tips(c).

tre of the glass plate with a mesh of 20×20 grids, as shown in figure 3.22. Each grid has a dimension $10 \times 10mm^2$. Different impacts, using finger tips, nails, and pen points, were conducted individually to create their own databases.

The percentage of correct recognition of the mapped locations was used to estimate the reliability of the method. The reliability is closely related to the resolution for the mapped locations. To investigate the relationship between reliability and resolution, experiments were carried out at different resolution levels, $10 \times 10mm^2$, $20 \times 20mm^2$, and $40 \times 40mm^2$.

Impacts were applied by tapping 11 times at each location. With sensors located at two corners, as shown in figure 3.23, twice the number of the samples can be obtained. Rather than producing a large quantity of data from all mapped locations in the plane for testing, candidate locations can be selected at different spatial resolutions to minimise redundant comparisons. In this experiment, five locations were selected for testing: $(05, 05)$, $(05, 15)$, $(10, 10)$, $(15, 05)$, and $(15, 15)$. The results are shown in tables 3.3, 3.4, and 3.5 respectively.

	Pen	Nail	Fingertip
$(05, 05)$	12/22	12/22	13/22
$(05, 15)$	22/22	15/22	09/22
$(10, 10)$	22/22	21/22	12/22
$(15, 05)$	14/22	17/22	18/22
$(15, 15)$	20/22	19/22	12/22
Percentage	81.8%	76.4%	58.2%

Table 3.3: Result of successful recognition with LTM method for three different impacts on 5 predefined positions at the resolution $10 \times 10mm^2$.

It is shown that the pen point has the highest reliability, which is reasonable

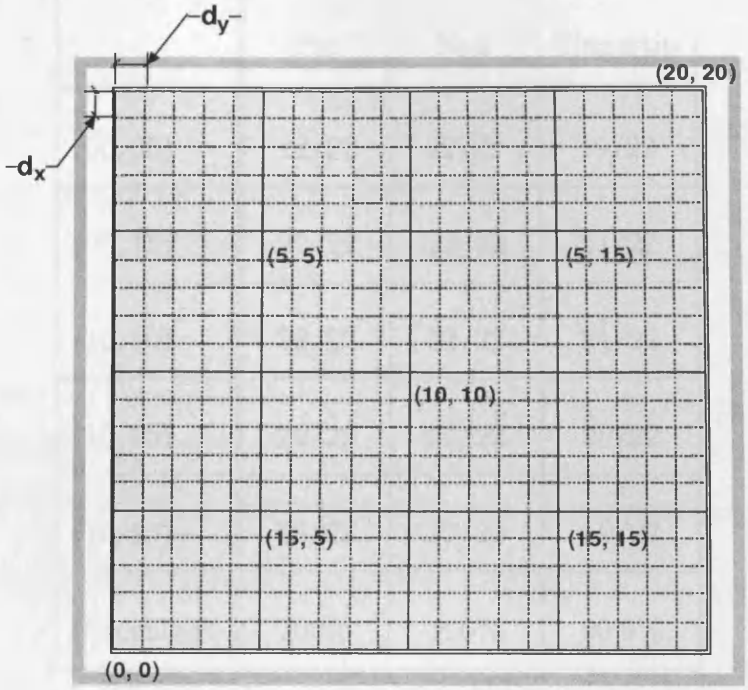


Figure 3.22: Layout of the LTM experiment.

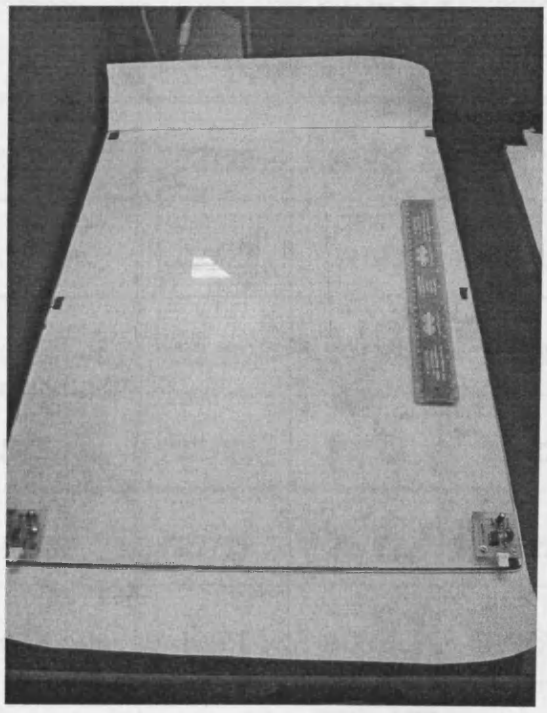


Figure 3.23: Hardware of the LTM experiment.

	Pen	Nail	Fingertip
(05, 05)	22/22	22/22	19/22
(05, 15)	22/22	22/22	21/22
(10, 10)	22/22	22/22	21/22
(15, 05)	22/22	22/22	20/22
(15, 15)	22/22	22/22	19/22
Percentage	100%	100%	90.9%

Table 3.4: Result of successful recognition with LTM method for three different impacts on 5 predefined positions at the resolution $20 \times 20mm^2$.

	Pen	Nail	Fingertip
(05, 05)	22/22	22/22	19/22
(05, 15)	22/22	22/22	21/22
(10, 10)	22/22	22/22	21/22
(15, 05)	22/22	22/22	21/22
(15, 15)	22/22	22/22	19/22
Percentage	100%	100%	91.8%

Table 3.5: Result of successful recognition with LTM method for three different impacts on 5 predefined positions at the resolution $40 \times 40mm^2$.

as the template was also prepared with the pen point. However, it can be seen that, with the resolution of $10 \times 10mm^2$ at two positions $(5,5)$ and $(15,5)$, the accuracy of recognition is very low at just 12/22 and 14/22. Although this is very low compared to other positions, it should be noted that incorrect samples were recognised at two very close locations of $(5,6)$ and $(15,6)$. This indicates that this method is not able to produce high reliability at such a high resolution. The nail can also produce signals at very high correlation levels, but with more errors.

In contrast, the fingertip has produced considerably more false triggers. With the resolution of $10 \times 10mm^2$, the percentage of correct recognition is just 58.2%, which rises to 91.8% at the lowest resolution of $40 \times 40mm^2$.

3.4 LTM improvement

3.4.1 Acquisition of impulse response

According to the experimental results above, the cross-correlation function is not reliable enough to recognise the location template for different impacts. Assuming the medium impulse transfer function $h(t)$ of source signal $s(t)$, together with external noise $n(t)$, the received signal $x(t)$ can be modelled as:

$$x(t) = s(t) * h(t) + n(t) \tag{3.12}$$

where the signal $s(t) * h(t)$ is assumed to be uncorrelated with $n(t)$. Thus, to acquire a quality impulse response between the source and the receiver, the source signal $s(t)$ and the noise $n(t)$ play a very important role.

Berkhout [BdVB80] introduced a simple method for acquiring impulse responses in concert halls, in order to improve the impulse response in the following two aspects:

- Repeatable impulse response;
- High signal-to-noise ratio (SNR) and high resolution.

Considering the existence of noise, the optimal solution of obtaining a clean impulse response should follow the least squares criterion of $E[h(t) * s(t) - u(t)]^2$ to be minimum, where $u(t)$ is the desired impulse response and E denotes the expectation. Also, the desired impulse response should follow the condition of:

$$E\{u(t)u(t+\tau)\} = \delta(t) \tag{3.13}$$

With the assumption of the uncorrelated noise:

$$E\{u(t)n(t)\} = 0 \tag{3.14}$$

the satisfying optimal filter was obtained as:

$$f(t) * [r_s(t) + r_n(t)] = s(-t) \tag{3.15}$$

where $r_x(t)$ denotes the autocorrelation of x. This can be represented in frequency domain as:

$$F(f) = |S(f)|/[|S(f)|^2 + R_n(f)]e^{-j\phi(f)} \tag{3.16}$$

From the frequency domain representation, two properties of this filter can be summarised as:

1. The phase calculation with $e^{-j\phi(f)}$ results with zero phase that the energy is concentrated at the arrival time τ. This is identical to the cross correlation in frequency domain;

2. The magnitude calculation has the function of suppressing the frequencies, where the noise effect is large enough to make errors more likely.

3.4.2 Improvement with phase cross-correlation

The above technique is not ideal for this application. However, towards achieving the desired target of equation 3.13, a simple modification to the above solution can be made to suit different types of signal. Similarly, an optimal filter is applied before the process of the cross-correlation. This is the so-called phase cross-correlation[1]. With a similar style as above, it can be described as:

$$F(f) = \frac{X_1(f)}{[|X_1(f)X_2^*(f)|]} \approx \frac{|S_1(f)|}{[|S_1(f)S_2^*(f)|]} e^{-j\phi_1(f)} \tag{3.17}$$

where X_1 and X_2 are the Fourier transforms of the two received signals, S_1 and S_2 represent the Fourier transforms of the two real signals, $*$ denotes the conjugation, and ϕ_1 is the angle of S_1. Assuming the noise effect is small enough to be neglected, this inverse filter is actually normalising all frequency components. Only the phase spectrums are kept and processed with a convolution calculation. The filtered output $O(f)$ will be:

$$O(f) = e^{-j\phi_1(f)} e^{j\phi_2(f)} \tag{3.18}$$

When the two signals have the same phase spectrums, they are mutually eliminated as a series of unit constants. Its inverse Fourier transform clearly is a Delta function at time t $\delta(t)$. This also indicates the principle of phase conjugate mirrors, which is identical to the time reversal in the frequency domain [FCD+00].

As analysed before, the uniqueness of the signals is caused by the reflections

[1]The phase correlation idea described in section 3.4.2 was originally suggested by one Tai-Chi partner, LOA, during the period of Tai-Chi project. The practical experiemnts and analyses were carried out independently.

from the surface's boundaries and the material's dispersion. Therefore, the only difference is the source signals, which contain different frequency components for different impacts. This has been explained in section 2.3.3. By treating all frequencies equally, the difference in frequency amplitudes can be minimised. This also implies that the signature of a location is embedded in the phase spectrum of the received signals.

This is actually identical to the PHAT type of the Generalised Cross-Correlation (GCC), which has been applied and discussed as an optimal solution to time difference estimation. In its original form, the PHAT method requires the estimation of cross spectrum between signals. However, this would be very computational if there are a large number of templates to be calculated. In this research, the cross spectrum estimation is not required for each calculation. Instead, as indicated in equation 3.17, only the normalised phase spectrum is required. The templates are replaced with the normalised phase spectrum instead of the original signal in the time domain. Thus, the comparison between signals can be carried out in three steps:

1. Calculate the phase spectrum of candidate signal to be compared;
2. Apply dot product between this new phase spectrum and the template, which is also a phase spectrum;
3. Inverse Fourier transform of the result from last step to obtain the correlation coefficients in the time domain.

The above method is based on the assumption of zero noise. However, as this pre-filter treats all individual frequencies equally, noises have as important an effect as the main components of the signals after the normalisation. As it is not necessary to use the complete spectrum, here only the part of signal with a relatively high SNR is considered. A threshold is applied in order to check if any individual frequency component is more likely to be a noise or a signal. Specifically, for the template, which is prepared under good conditions, only the

range with higher energy in the frequency spectrum is taken into account. A proper window is applied to keep the frequencies of interest and to eliminate the effect of noise.

Compared to the results of using the general cross-correlation (figure 3.21), figure 3.24 shows the coefficients with the new technique for the three different impacts. In contrast, this phase correlation has a significantly different result, as shown in the graph. For all the three impacts, at the approximate time difference position, the coefficients increase very sharply to a peak value. At other positions, the coefficients decline to a stationary or low level. This is clearly the desired expectation that can reduce the ambiguities during the process of recognition among the templates.

3.4.3 Performance evaluation

Comparing the previous LTM method with the cross-correlation function, tables 3.6, 3.7, and 3.8 below show the results of the same experiment using the same data, except that the new method is applied instead. A significant improvement for the finger tap can be observed for all three resolutions. At the resolution of $40 \times 40mm^2$, the percentage increased to 100%. However, at the resolutions of $10 \times 10mm^2$ and $20 \times 20mm^2$, it improved from 58.2% to 65.4% and from 90.9% to 91.8% respectively.

For the impacts with both nail and pen point, all of the test signals were successfully recognised at the two lower resolutions. At the highest resolution of $10 \times 10mm^2$, it can be seen that the average percentage of successful recognition at only two of the positions $(05, 05)$ and $(15, 05)$ are degraded significantly. However, the others maintain a very high accuracy and the average performance has been improved.

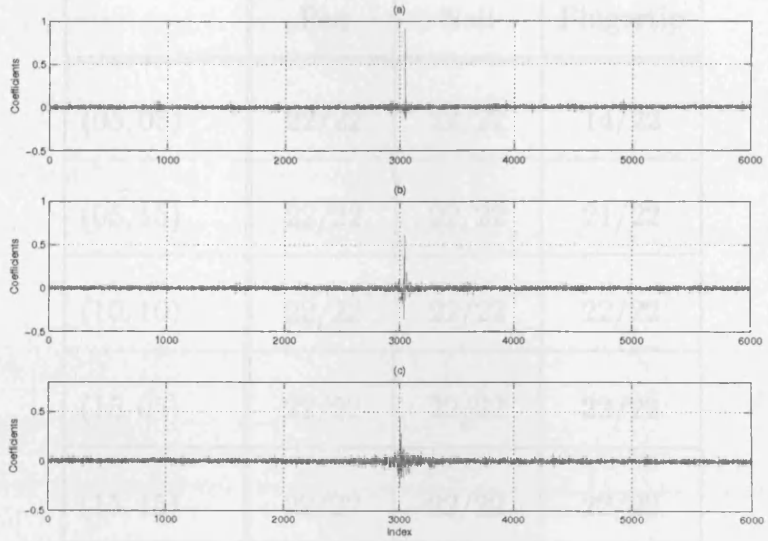


Figure 3.24: Phase correlation coefficients for three different impacts using pen points, nails, and fingertips.

	Pen	Nail	Fingertip
(05, 05)	12/22	11/22	08/22
(05, 15)	22/22	21/22	11/22
(10, 10)	22/22	21/22	18/22
(15, 05)	21/22	10/22	15/22
(15, 15)	22/22	22/22	20/22
Percentage	90%	77.3%	65.4%

Table 3.6: Result of successful recognition with the phase correlation for three different impacts on 5 predefined positions at the resolution $10 \times 10mm^2$.

	Pen	Nail	Fingertip
(05, 05)	22/22	22/22	14/22
(05, 15)	22/22	22/22	21/22
(10, 10)	22/22	22/22	22/22
(15, 05)	22/22	22/22	22/22
(15, 15)	22/22	22/22	22/22
Percentage	100%	100%	91.8%

Table 3.7: Result of successful recognition with the phase correlation for three different impacts on 5 predefined positions at the resolution $20 \times 20mm^2$.

	Pen	Nail	Fingertip
(05, 05)	22/22	22/22	22/22
(05, 15)	22/22	22/22	22/22
(10, 10)	22/22	22/22	22/22
(15, 05)	22/22	22/22	22/22
(15, 15)	22/22	22/22	22/22
Percentage	100%	100%	100%

Table 3.8: Result of successful recognition with the phase correlation for three different impacts on 5 predefined positions at the resolution $40 \times 40mm^2$.

3.5 Improvement by using two sensors

3.5.1 Improvement of successful recognition

A second sensor can be applied in addition to increase the recognition probability. With the previous recorded data, two sensors were used to record signals simultaneously. The second sensor location was chosen randomly at one of the corners of the plate. The summation of the coefficients from both signals is used. Using the new phase cross correlation function, it was found that a second sensor can help to improve the accuracy so that it is almost perfect. With the same recorded data, all of the three different impacts can be recognised successfully for all three resolutions. Although false recognition happens occasionally at the high resolution, the percentage of false recognitions is still too low to be counted. With the original cross-correlation function, the recognition accuracy has also been improved; however it is still less accurate than the phase-correlation function.

3.5.2 Enhancement of the correlation coefficient

In order to give a clearer view of how the two methods behave, the coefficient distribution in the plane of $200 \times 200mm^2$ is presented in this sub-section. The square area is divided into 20×20 grids. Figures 3.25 and 3.26 illustrate the typical results of the pen impact at position $(10, 10)$ for both cases, one-sensor and two-sensor approaches, with the cross-correlation and the phase-correlation methods separately.

It can clearly be seen that the second sensor is very effective in enhancing the contrast of the coefficients between the source position and other positions, for both methods. The coefficient at the source position is almost doubled in value, but in contrast the coefficients at other positions do not seem to be enhanced as significantly as the source position. For the phase-correlation method, in particular, the source position can be recognised confidently.

The results for the nail-generated signals and the fingertip-generated signals

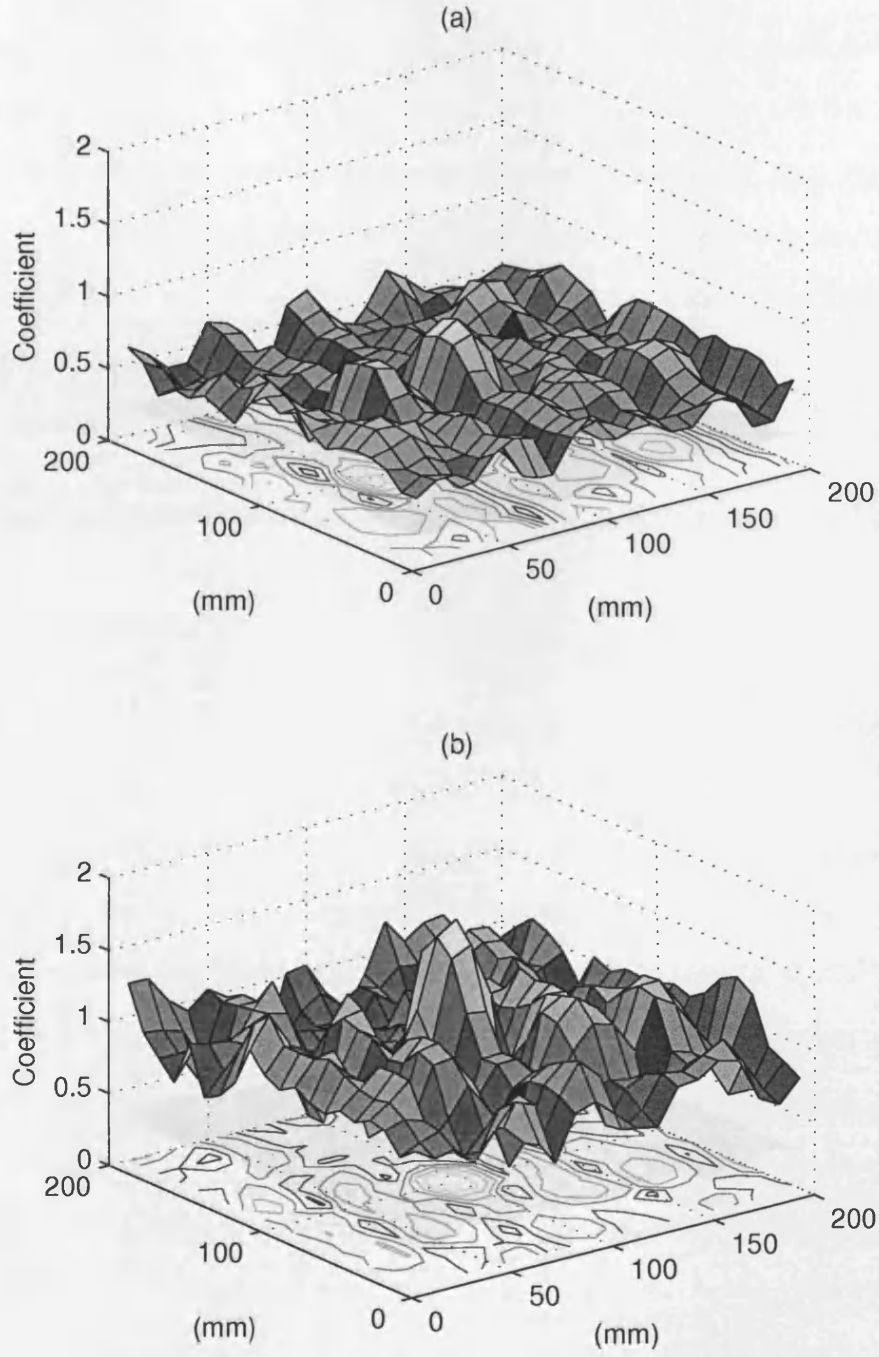


Figure 3.25: Cross-correlation coefficients in a square area ($200 \times 200(mm)$) for the impact using a pen point, using one sensor (a) and two sensors (b).

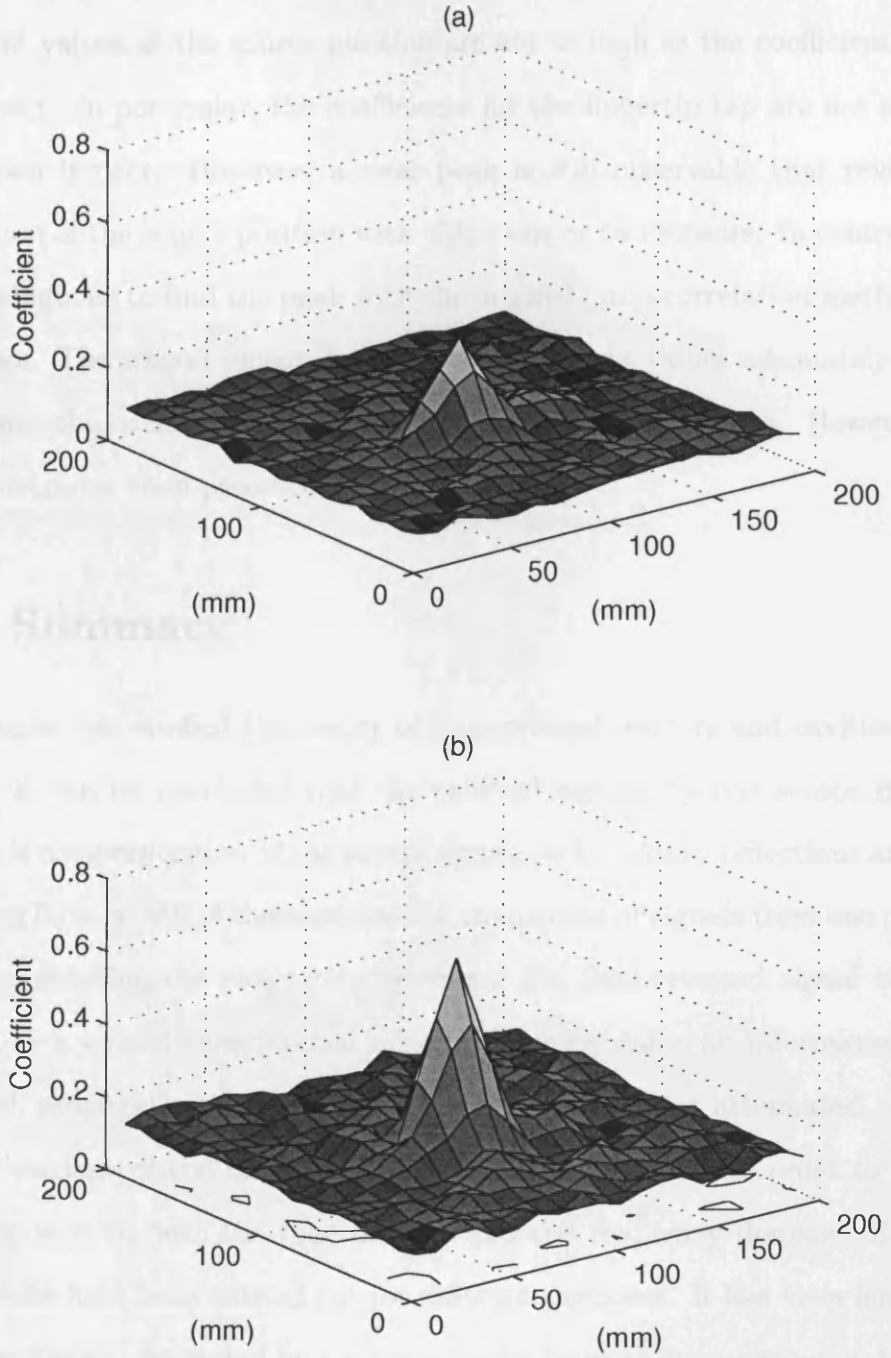


Figure 3.26: Phase correlation coefficients in a 2-D square area ($200 \times 200(mm)$) for the impact using a pen point, using one sensor (a) and two sensors (b).

are shown in figures 3.27, 3.28, and 3.29, 3.30 respectively. Similarly, the second sensor has increased the contrast between the coefficients at the source position and other positions in all the cases. With the phase-correlation method, the coefficient values at the source position are not as high as the coefficient for the pen impact. In particular, the coefficients for the fingertip tap are not as sharp as the pen impact. However, a clear peak is still observable that reveals the recognition of the source position with either one or two sensors. In contrast, it is more ambiguous to find the peak with the original cross-correlation method with one sensor. The second sensor increased the coefficient values adequately enough to enhance the contrast of the coefficient at the source position. However, it is more ambiguous than phase-correlation.

3.6 Summary

This chapter has studied the theory of time-reversal mirrors and cavities. Based on this, it can be concluded that the received signals by one sensor from one position is a superposition of the source signal, its boundary reflections and other scattering factors. All of these ensure the uniqueness of signals from one position. Inversely, recalling the reciprocity principle, the time-reversed signal is in fact working as a virtual time-reversal mirror, which contains all information of the multipath propagation if the signal is long enough and not attenuated.

Two versions of the LTM method have been developed, in order to find the similarity level in both the time domain and the frequency domain. Extensive experiments have been carried out for different purposes. It has been found that vibration signals, generated by a point impulse force at one position, hold strong signature information of the propagation between the source and the receiver position. The phase spectrum of signals is found to be correlated at a high level among those signals from one position for different impacts.

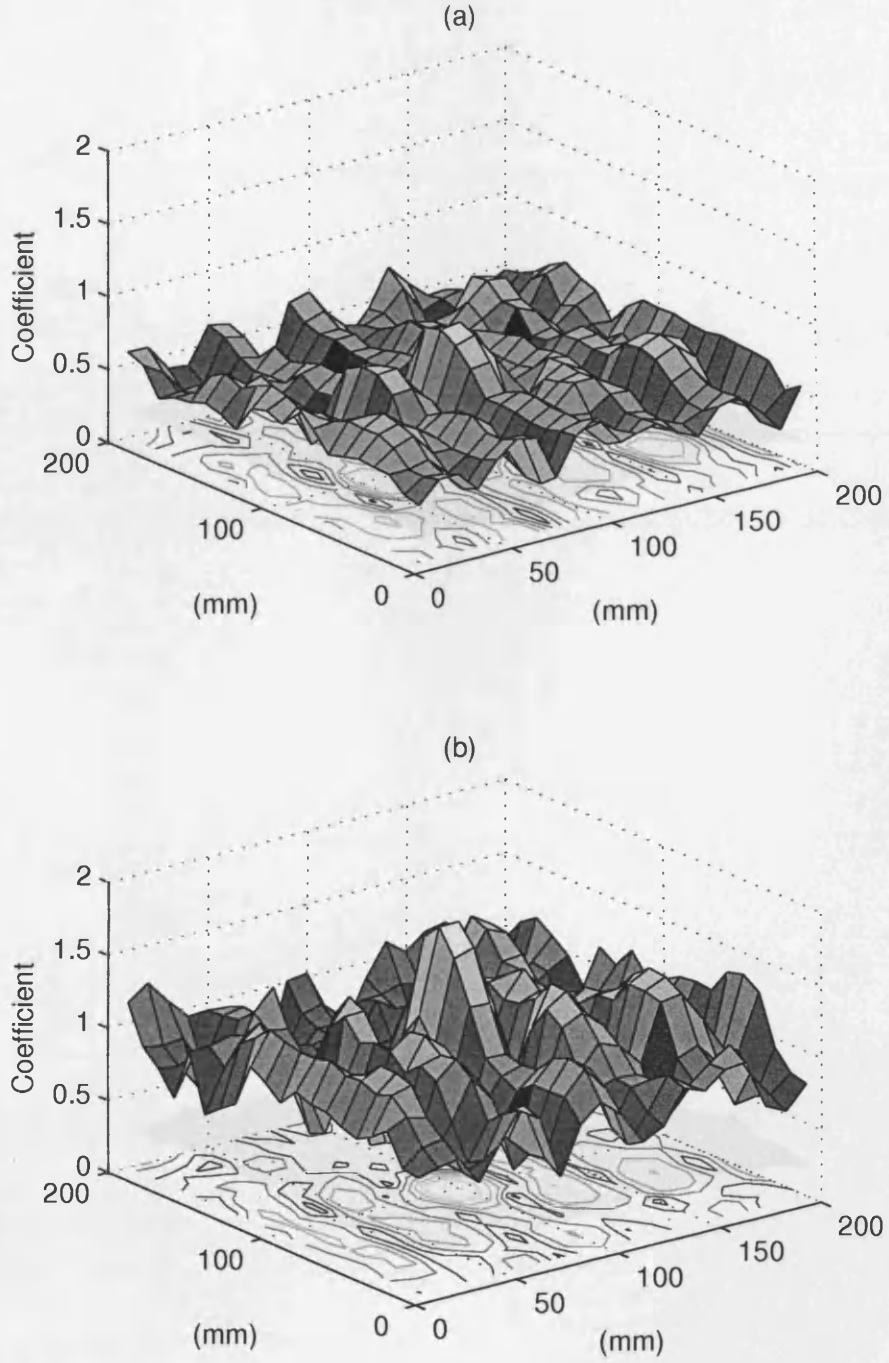


Figure 3.27: Cross-correlation coefficients in a 2-D square area ($200 \times 200(mm)$) for the impact with a nail, using one (a) and two sensors (b).

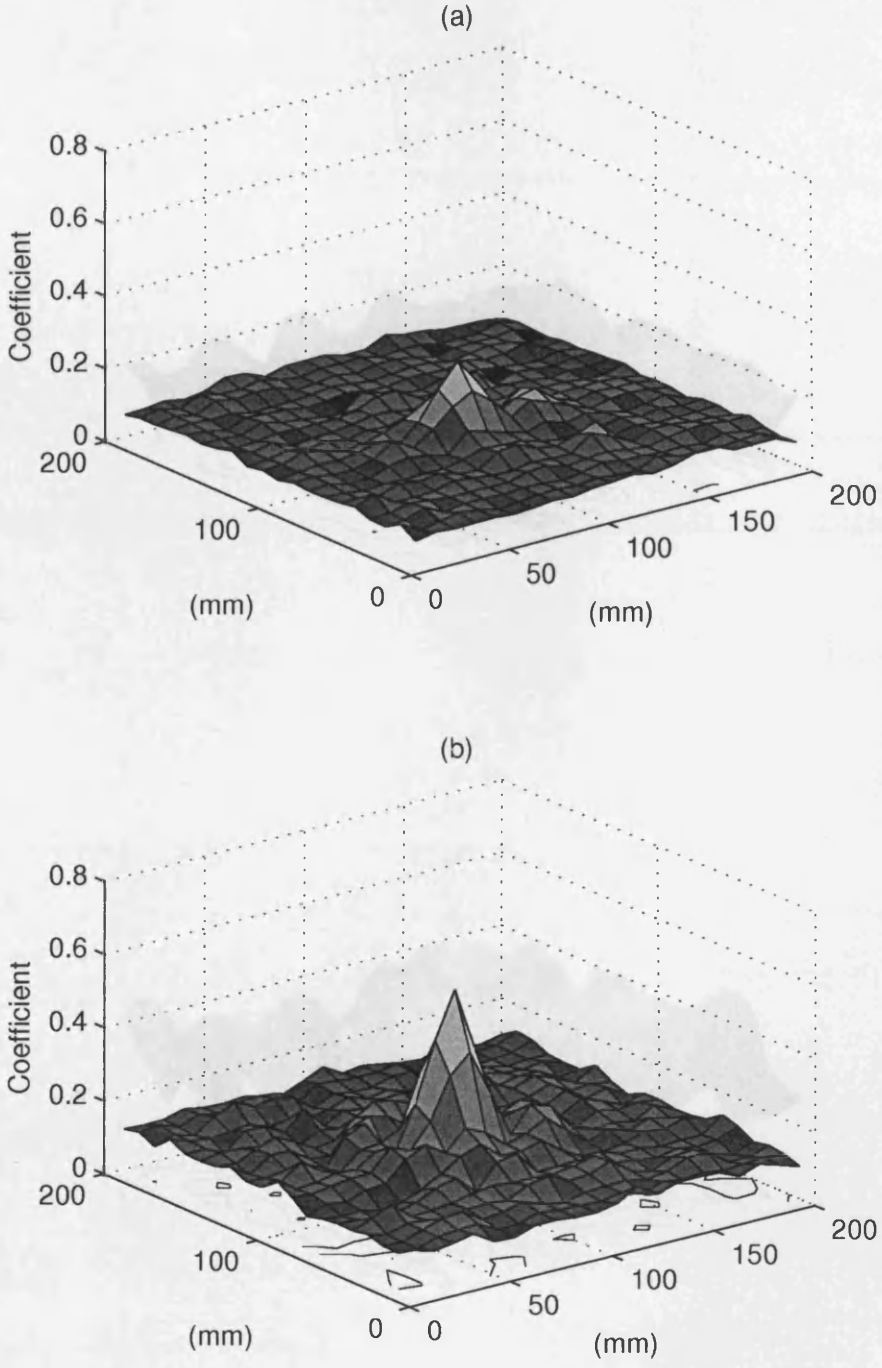


Figure 3.28: Phase-correlation coefficients in a 2-D square area ($200 \times 200(mm)$) for the impact with a nail, using one (a) and two sensors (b).

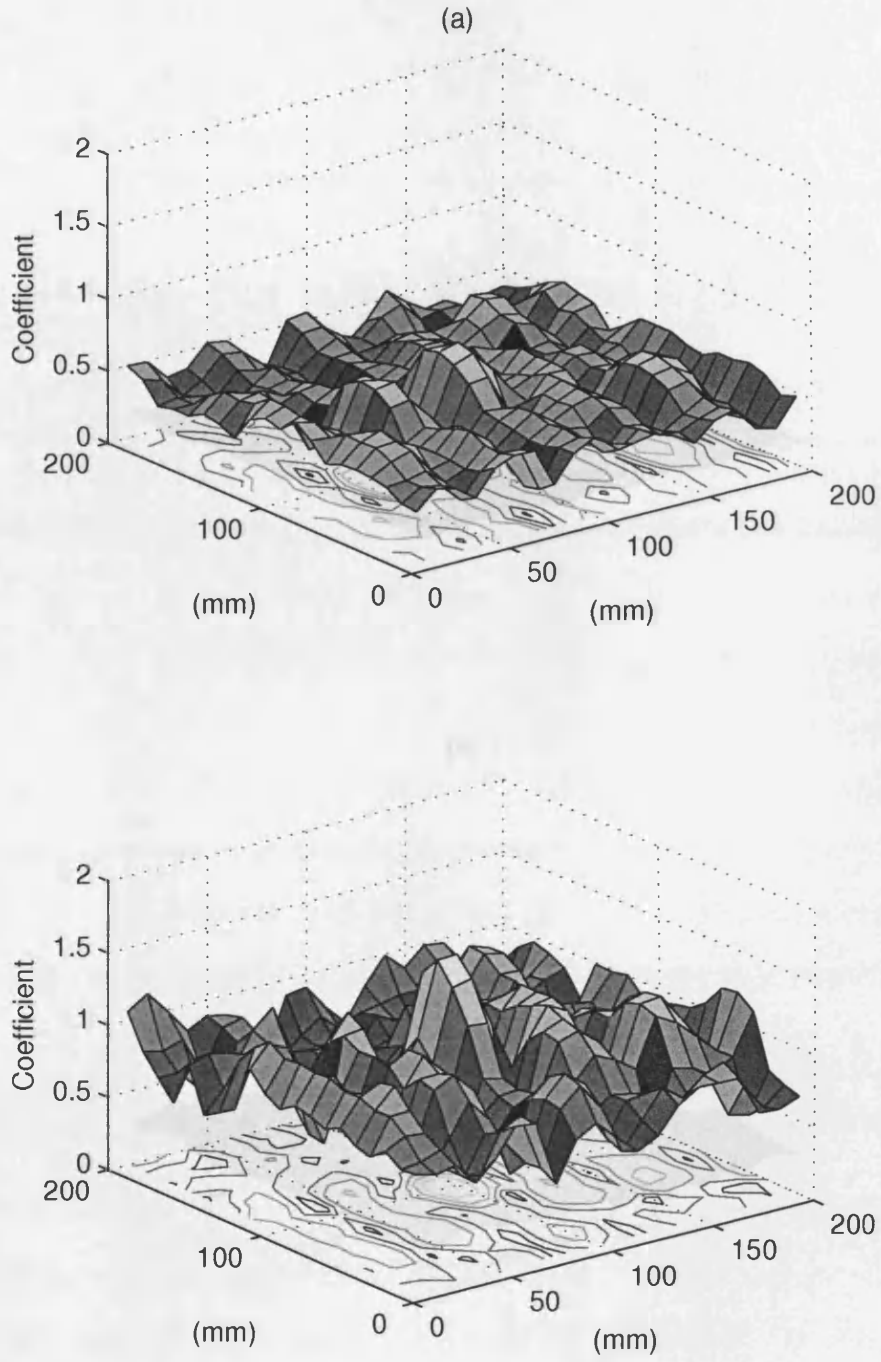


Figure 3.29: Cross-correlation coefficients in a 2-D square area ($200 \times 200(mm)$) for the impact with a finger tip, using one (a) and two sensors (b).

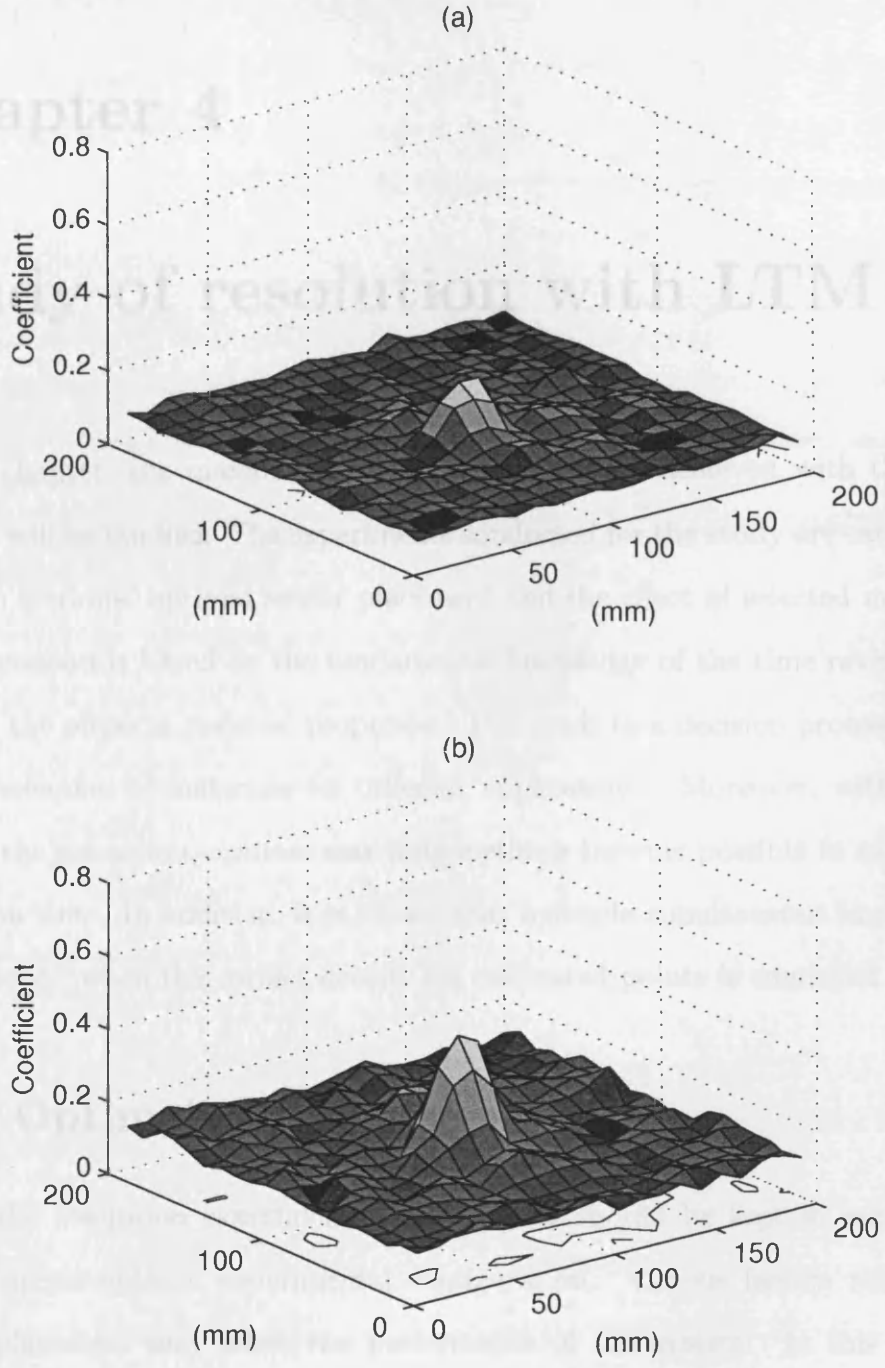


Figure 3.30: Phase-correlation coefficients in a 2-D square area ($200 \times 200(mm)$) for the impact with a finger tip, using one (a) and two sensors (b).

Chapter 4

Study of resolution with LTM

In this chapter, the maximum resolution that can be achieved with the LTM method will be studied. The experiments conducted for the study are categorised into two sections: optimal sensor placement and the effect of selected materials. The discussion is based on the fundamental knowledge of the time reversal theory and the physical material properties. This leads to a decision process for the proper selection of materials for different applications. Moreover, with knowledge of the resolution, optimal searching methods become possible to reduce the execution time. In addition, it is shown that multiple simultaneous impacts are recognisable, when the correct density for calibrated points is employed.

4.1 Optimal sensor placement

Before the resolution experiments, a few points should be kept in mind when setting up an optimal experimental configuration. Various factors relating to sensor placement may affect the performance of the system. In this section, the problem will be discussed and experiments conducted in three sub-sections, including the effect of the geometrical symmetry relative to the sensor location, the effect of the number of sensors on the resolution, and the principle of the optimal location of sensors.

Geometrical symmetry

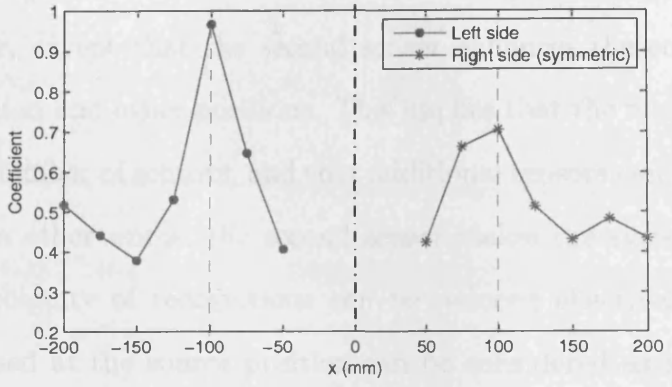
Geometrical symmetry of two or more locations relative to the sensor will cause ambiguities for the LTM method, because of the identical propagation path from all these source positions to the sensor. On a regular shaped surface, when a sensor is located in the centre of the surface, symmetric positions around the sensor can be found at the two sides of the sensor. This was tested on a rectangular Plexiglass plate with dimensions of $1000 \times 600 \times 6mm^3$. The sensor was located at the centre of the plate $(500, 300)$.

The experiments were carried out on the axis of $y = 300mm$. 14 points were calibrated beforehand, located symmetrically at the two sides of the sensor with distances of ± 200, ± 175, ± 150, ± 125, ± 100, ± 75, and $\pm 50(mm)$ respectively. Each position was calibrated with 5 samples, and the average coefficient was used. Two experiments were prepared to test the effect of the geometrical symmetry, centred at $100(mm)$ and $150(mm)$ to the left side to the sensor. Figure 4.1 shows the results of the two centre positions at $100(mm)$ and $150(mm)$.

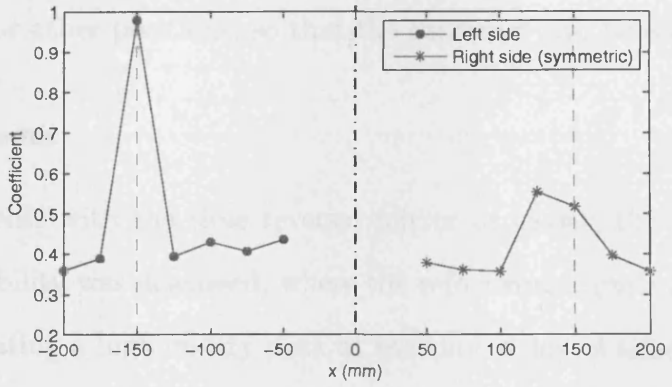
The coefficients at the mirrored positions are considerably higher than the average of other positions. The high similarity between the two symmetric positions will cause ambiguities in recognising the location template. Therefore, this problem should be avoided when designing the application in practice.

Number of sensors

In the time reversal experiments, the effect of using different numbers of sensors has been discussed for various media [YTF03] [FP01]. The improvement with more sensors is that the refocusing amplitude at the source position is significantly enhanced. However, the resolution remains the same. Fink has experimented with TRM in random media, concluding that the resolution of the system is independent of the array aperture. Two experiments were performed using only 1 and 122 transducers in the same medium. The resolution remained the same for the two experiments, except the significant enhancement of the contrast with



(a)



(b)

Figure 4.1: Coefficients at symmetrical positions, at position $100mm$ (a) and $150mm$ (b).

more sensors.

Similarly in this experiment, as in the last chapter, the second sensor is very helpful to enhance the contrast of the coefficients between the source position and other positions. With the same experiment as reported in the last chapter (figure 3.26), the typical correlation coefficients along the line at $x = 100mm$ are represented in figure 4.2. The coefficients are normalised here so that it can be more clearly compared. It can be seen that basically the shapes in both cases are very similar, except that the second sensor enhances the contrast between the source position and other positions. This implies that the resolution does not depend on the number of sensors, and that additional sensors can help to increase the contrast. In other words, the second sensor makes the system more robust so that the ambiguity of recognitions can be reduced effectively. The energy amplitude focused at the source position can be considered as a summation or superposition of signals from all sources. However, the summation of all sources does not hold for other positions, so that the constrast can be enhanced.

Sensor placement

In the experiments with the time reversal mirror or cavity, the principle of the refocusing capability was discussed, where the refocusing signals are propagating back and generating a high energy peak at the time delay of time reversed duration. In addition, regarding the resolution that can be achieved, the aperture of the TRM and TRC are important. However, with one sensor, it is theoretically impossible to recreate the high focusing energy in an open environment. In a bounded environment, the wave does not only propagate through a direct path from the source to the sensor, but is also reflected by the boundaries of the object. The reflections make the received signal more informative. Consequently, the reflections can construct a lot more virtual sensors [IF98], which can even further construct infinitely more virtual sensors, as shown in figure 4.3. Thus, one sensor in a bounded plate can eventually work as a TRM that includes the

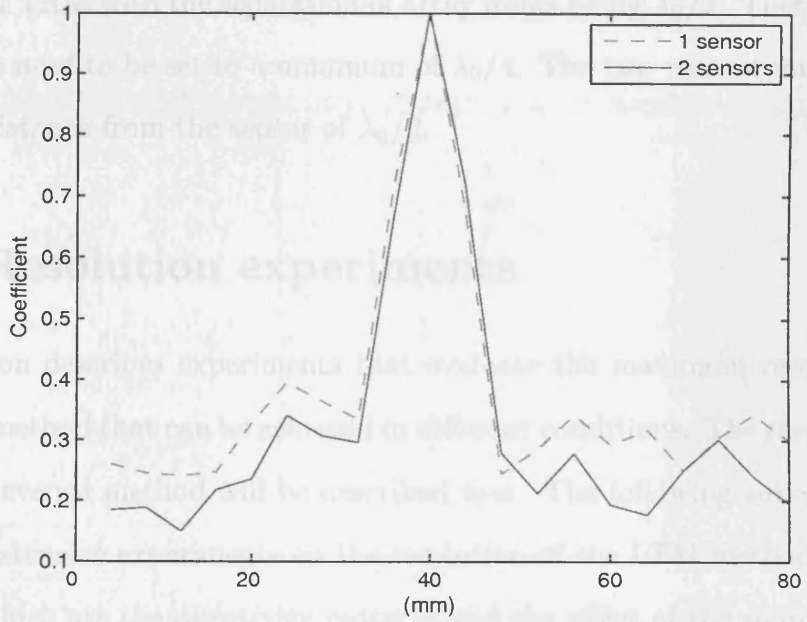


Figure 4.2: A typical example of contrast enhancement with a second sensor.

sensor and its virtual reflection sensors. As discussed in [FP01], the aperture of the TRM determines the quality of refocusing. Therefore, the optimal placement of sensors may effectively enhance the performance of LTM methods in certain conditions.

In the TRM method, the separation $h/2$ of sensors in an array is determined as $\lambda_0/2$, where λ_0 is the wavelength of the central frequency or the carrier frequency of the pulse [BPT03]. This can keep interference at a minimum. The aperture of the sensor array is thus $a = Nh$, where $2N + 1$ represents the number of sensors. Here, the same principle can be employed directly, in order to construct the virtual TRM with the separation of array items being $\lambda_0/2$. Therefore, values of x and y need to be set to a minimum of $\lambda_0/4$. The two nearest virtual sensors are at a distance from the sensor of $\lambda_0/2$.

4.2 Resolution experiments

This section describes experiments that evaluate the maximum resolution with the LTM method that can be achieved in different conditions. The resolution with the time reversal method will be described first. The following sub-sections will describe extensive experiments on the resolution of the LTM method from three aspects, which are the directivity patterns and the effect of the source locations, the effect of the thickness of the plates, and the dimensions of the plates. It should be noted that, in the following experiments, only a pen point is used for the generation of signals to ensure a consistent excitation source.

4.2.1 TRM resolution analysis and wavelength estimation

In the TRM experiments, the resolution is limited by the wavelength of the central frequency. Figure 4.4 shows the simulated refocusing process of time-reversed signals. It can be observed that a few side lobes are spreading around the source position with a distance of about a full wavelength of the central

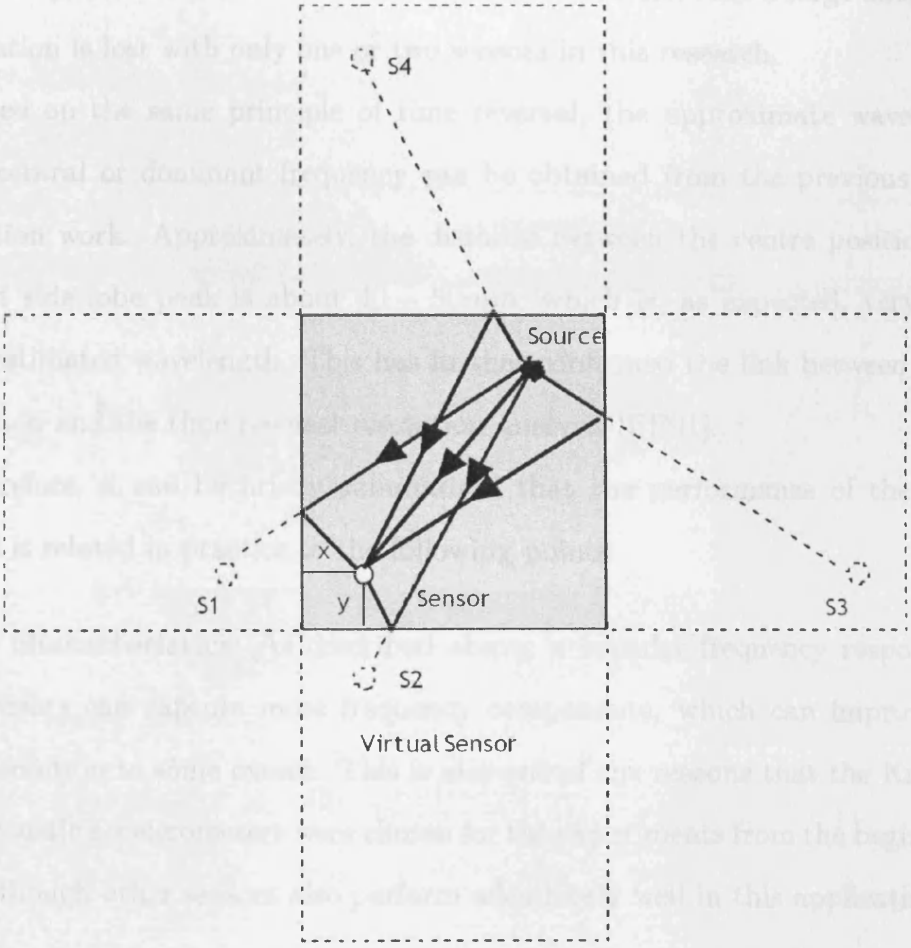


Figure 4.3: Virtual sensors constructed by the boundary reflections.

frequency. It is interesting to note from figure 3.25 in chapter 3 that the cross-correlation coefficients between the template and the new signal have a similar trend and that the centre is highly correlated and that a few local peaks or lobes are at an almost constant distance spreading from the centre outwards. This trend is not as strong as the numerical time reversal result, as the correlation is just an assumption of the time reversal process in an ideal lossless medium. Importantly, the number of sensors also affects the result, that a large amount of information is lost with only one or two sensors in this research.

Based on the same principle of time reversal, the approximate wavelength of the central or dominant frequency can be obtained from the previous cross-correlation work. Approximately, the distance between the centre position and the first side lobe peak is about $40 - 50mm$, which is, as expected, very close to the estimated wavelength. This has further confirmed the link between cross-correlation and the time reversal resolution analysis [FP01].

Therefore, it can be briefly summarised that the performance of the LTM method is related in practice to the following points:

Sensor characteristics: As described above, a broader frequency response of sensors can capture more frequency components, which can improve the resolution to some extent. This is also one of the reasons that the Knowles acoustic accelerometers were chosen for the experiments from the beginning, although other sensors also perform adequately well in this application.

Impact type: This determines the bandwidth of the source signal. Rigid objects generate higher frequency and broader bandwidth signals than soft objects, such as finger tips, which generate relatively lower frequency and narrower bandwidth signals.

Interactive surface materials: The wave velocity in a material is closely related to the wavelength. Also, different materials may have different frequency responses, that will eventually affect the resolution of the LTM

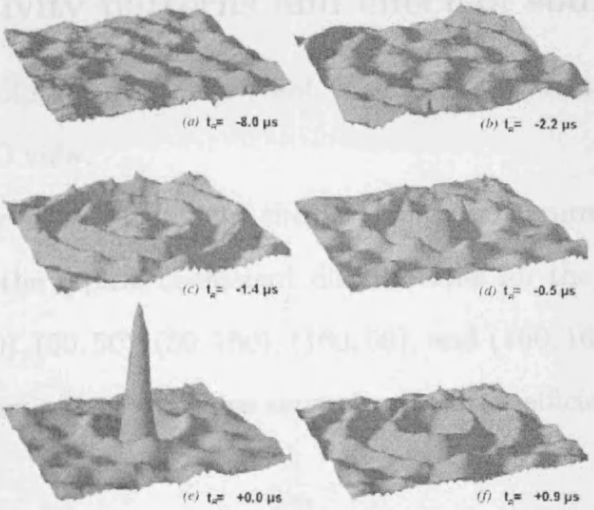


Figure 4.4: Time-reversed wave field observed at different times around the central point on a square of $15 \times 15mm^2$ ([FP01]).

method.

Geometry of surfaces: At the boundaries, the signals can be reflected or refracted. The more complicated geometry the surface has, the more reflections and mode conversions there are. Presumably, this complicated wave propagation multipath has an effect similar to more virtual sensors. Due to the significant enhancement with more sensors as reported above, the complicated geometry of the surface can effectively increase the reliability, but not the resolution.

4.2.2 Directivity patterns and effect of source locations

Regarding the resolution, it is important to check the directivity pattern of the coefficients in a 2D view.

With the same data as used in the last chapter, figures 4.5, 4.6, 4.7, 4.8, and 4.9 illustrate the typical coefficient distributions for the impacts at the five positions $(100, 100)$, $(50, 50)$, $(50, 160)$, $(160, 50)$, and $(160, 160)$ respectively. The experiment was carried out with two sensors and the coefficients were divided by 2.

From the figures, it is a common feature among all the tested five positions that the focal spots have a similar sized diameter of about $40 - 50(mm)$. This is a very close agreement with the value of the approximate wavelength. It can then be concluded that the LTM method resolution is limited by the wavelength of the dominant frequency, which is determined by the physical properties of the selected material. In addition, due to the elastic isotropy of the selected glass plate, the wavelength remains the same along both the x and y axes. On the other hand, this experiment has revealed the independence of the focal resolution from the source locations. This means that, for the same material, the resolution at different positions, where the elastic properties are the same, remains almost unchanged.

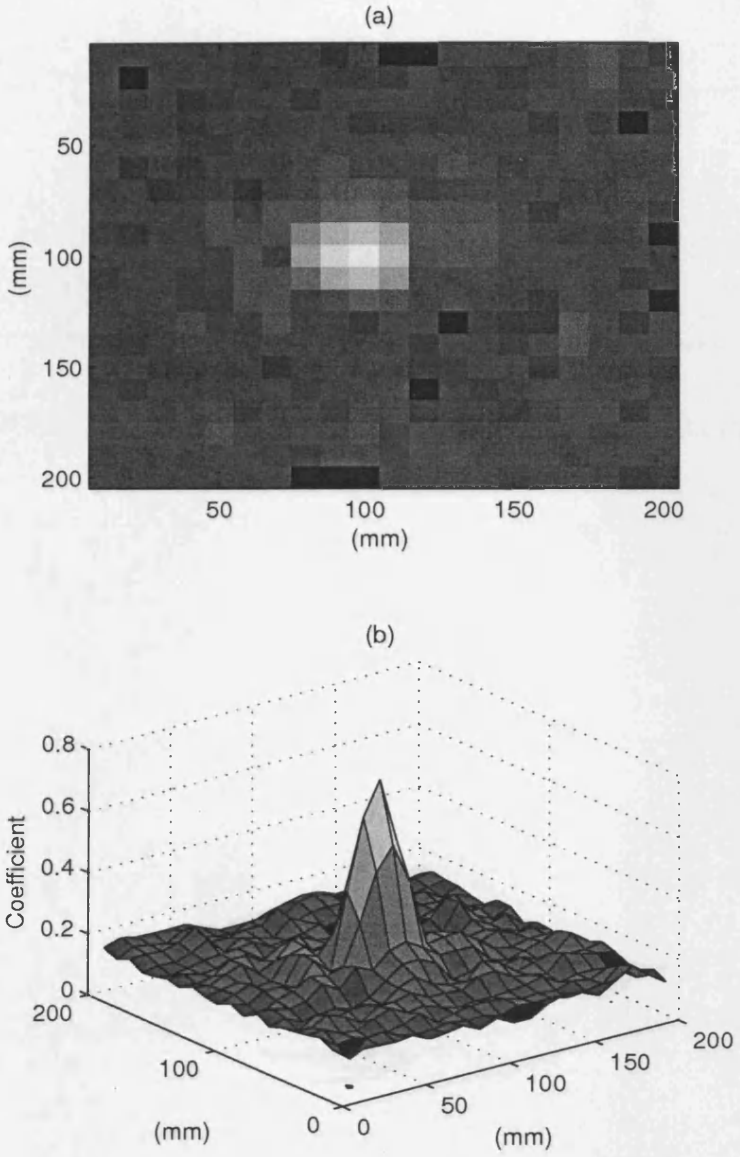


Figure 4.5: Coefficient distribution in a 2D area. The impact was conducted at position $(100, 100)$.

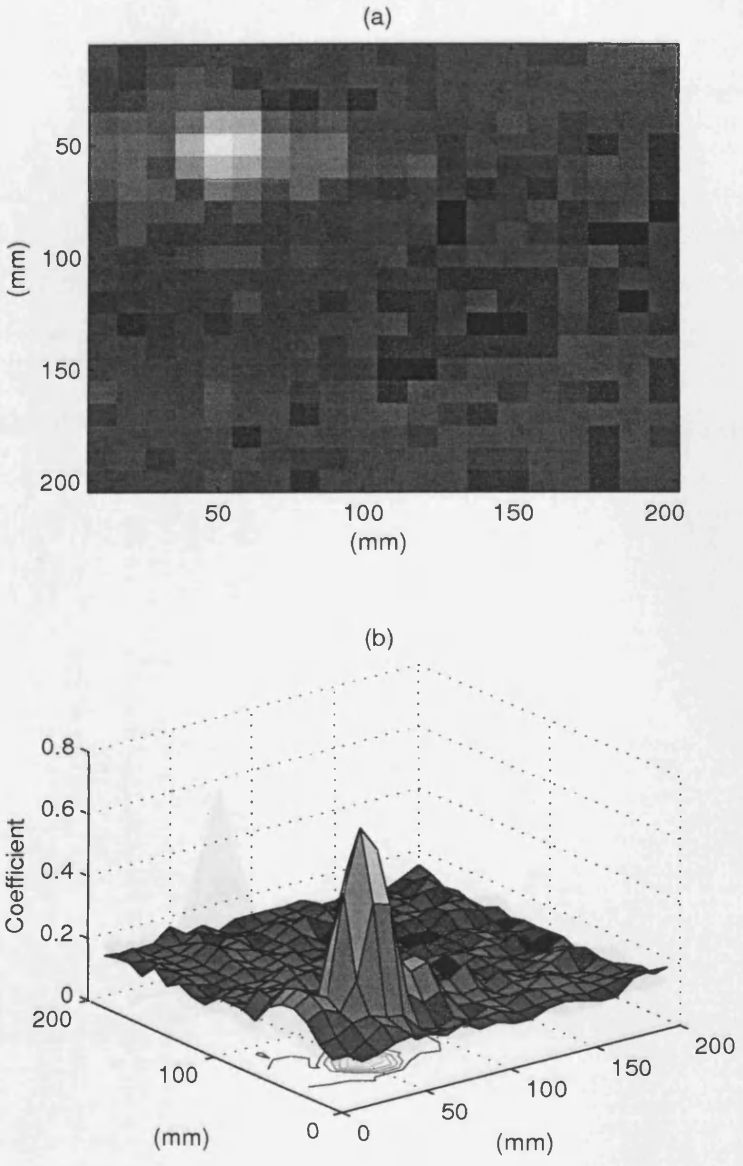


Figure 4.6: Coefficient distribution in a 2D area. The impact was conducted at position $(50, 50)$.

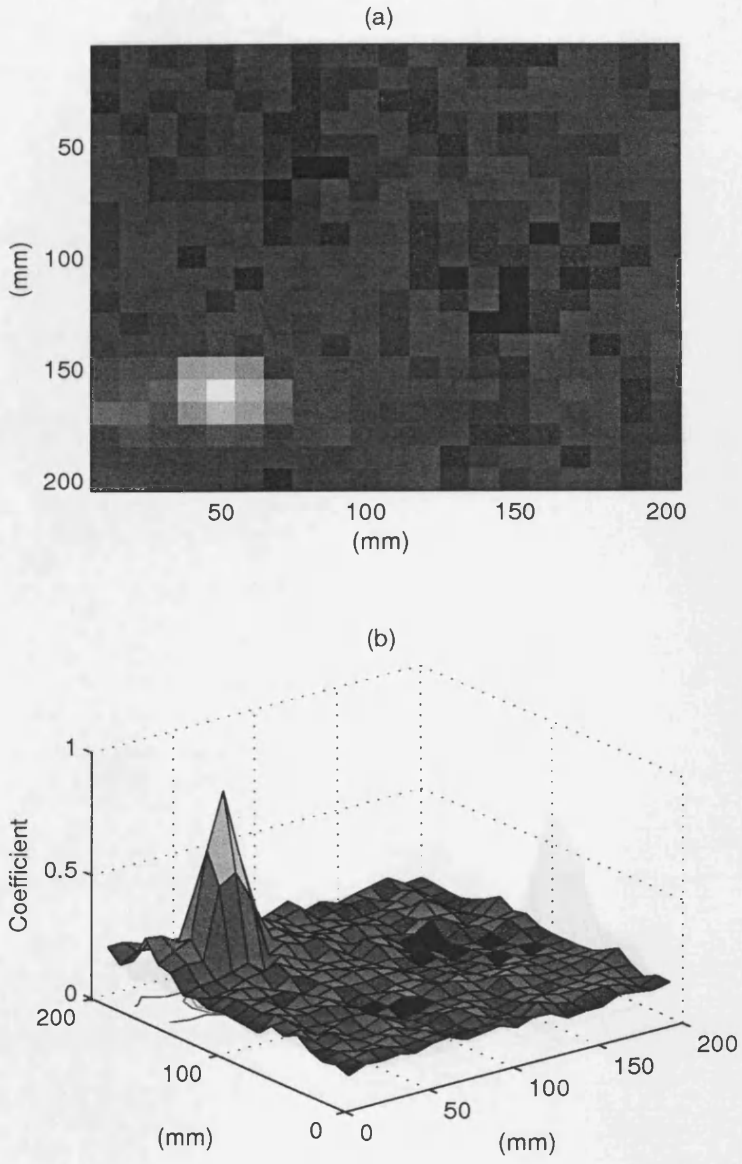


Figure 4.7: Coefficient distribution in a 2D area. The impact was conducted at position $(50, 160)$.

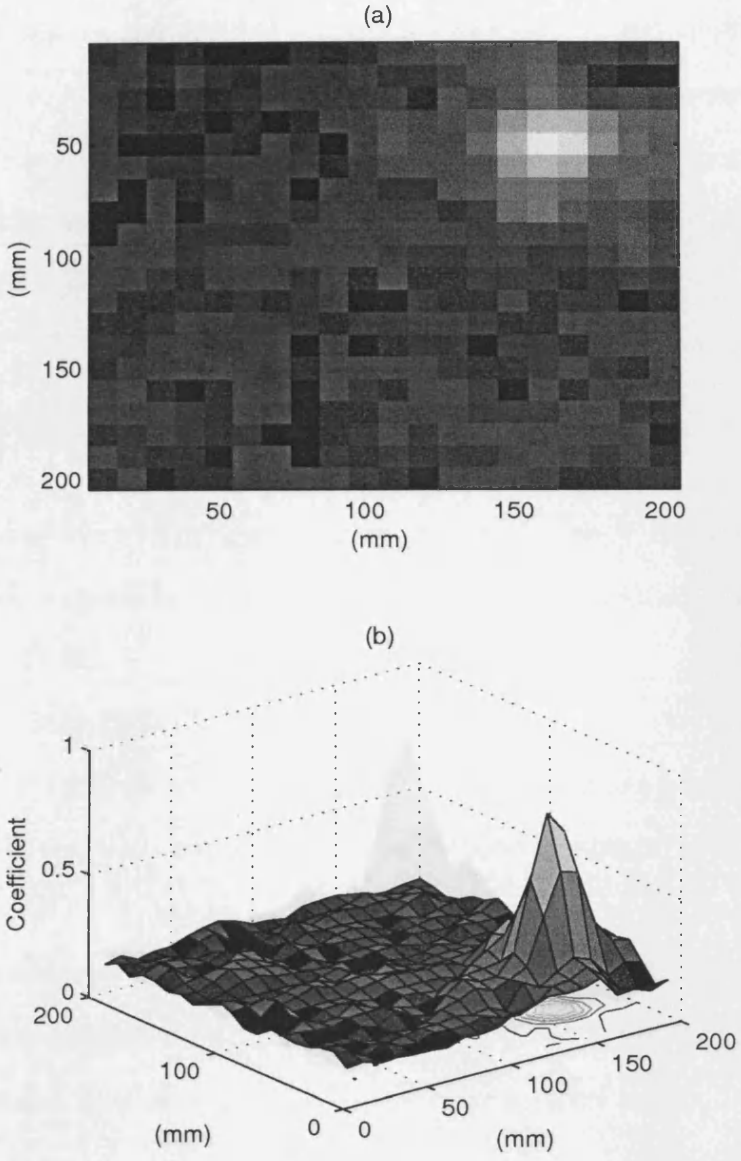


Figure 4.8: Coefficient distribution in a 2D area. The impact was conducted at position $(160, 50)$.

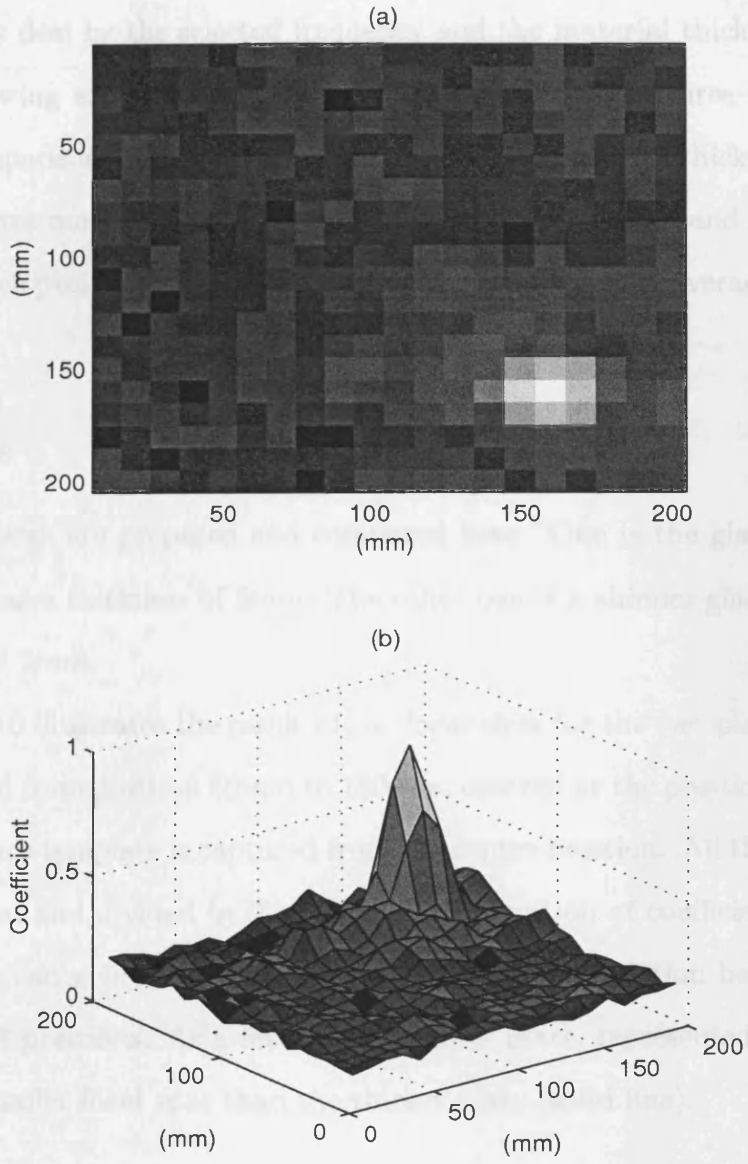


Figure 4.9: Coefficient distribution in a 2D area. The impact was conducted at position (160, 160).

4.2.3 Thickness of plates

As stated above, Lamb waves are complex vibrational waves that travel through the entire thickness of a plate. Propagation of Lamb waves is dependent on the density, elasticity, and material properties of the tested specimen, which are influenced a great deal by the selected frequency and the material thickness [Fah85].

The following experiments are, therefore, performed on three different materials. Comparisons are made in terms of the difference of thickness for each material. Three materials are selected for testing: glass, wood, and metal (steel). As before, each position is calibrated with 10 samples, and the average coefficients are used.

Glass plates

Two glass plates are prepared and compared here. One is the glass plate used before that has a thickness of $5mm$. The other one is a thinner glass plate, with a thickness of $2mm$.

Figure 4.10 illustrates the result of the focal sizes for the two plates. Samples were prepared from position $50mm$ to $150mm$, centred at the position $100mm$. A global reference template is captured from the centre position. All the coefficients are normalised and divided by 2 as this is a summation of coefficients from two sensors. This can give a better understanding of the correlation between signals from different positions. As a result, the thinner plate, represented by a dashed line, has a smaller focal spot than the thicker plate (solid line).

Wood boards

There are many different wood materials, in which the properties, such as density and elasticity parameters, vary significantly from each other. Medium Density Fibreboard (MDF) is dense, flat, stiff, and has no knots. It is widely used in everyday life, for applications such as cabinets, wall-panels, and storage units. Taking into consideration the applications, MDF is used here for its popularity.

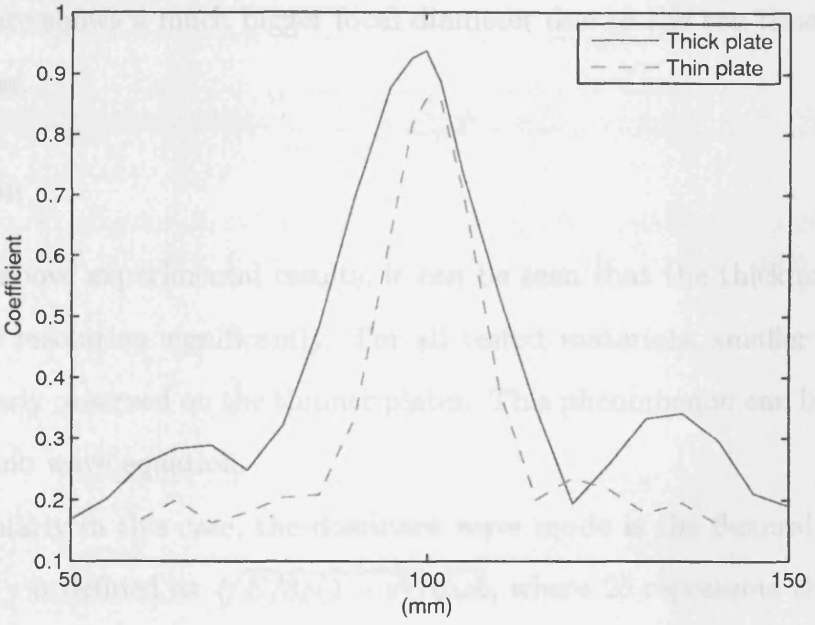


Figure 4.10: Coefficient distribution for two glass plates with different thicknesses in 1D view. The template position was captured at the centre position $100mm$.

Two samples with different thicknesses of $12mm$ and $6mm$ are prepared and compared. Figure 4.11 illustrates the comparison of the two boards. Similarly, the thinner board shows a higher resolution than the thicker board. In addition, the MDF boards show much larger focal spots than do the glass plates.

Metal (Steel)

Similarly, for the metal, two steel plates were selected for this experiment. Their thicknesses are $1mm$ and $10mm$ respectively. Figure 4.12 shows the comparison of the two plates. As before, the thinner plate gives very high resolution. The thicker plate shows a much bigger focal diameter due to the ten times difference in thickness.

Discussion

From the above experimental results, it can be seen that the thickness of plates affects the resolution significantly. For all tested materials, smaller focal spots can be clearly observed on the thinner plates. This phenomenon can be explained by the Lamb wave equation.

Particularly in this case, the dominant wave mode is the flexural wave. The velocity $c_{p,f}$ is defined as $\sqrt[4]{E/3\rho(1-v^2)}\sqrt{\omega b}$, where $2b$ represents the thickness of the plate, ω represents the angular speed and ρ represents the density. For each selected material, the elastic properties and the density can be approximately considered to be the same for different thicknesses. This can be understood as the signals on these materials containing a similar frequency spectrum. Therefore, the only factor that affects the phase velocities is the thickness $2b$. The thicker the plate is, the more dispersive the vibration will be. Then, within the same frequency spectrum, the higher velocity means larger wavelength, and hence lower resolution.

For example, for the experiment on the two glass plates, the approximate focal spots are $45mm$ and $30mm$ for the plates with a thickness of $5mm$ and

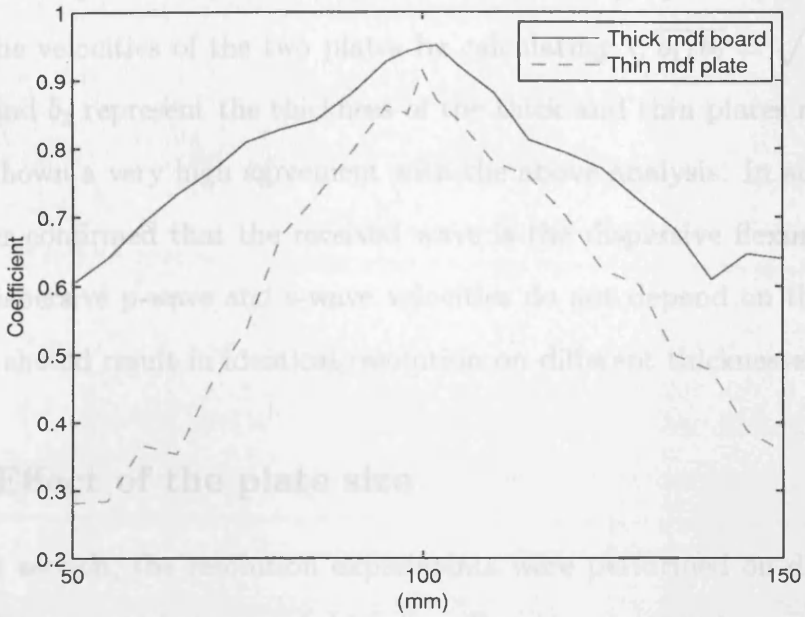


Figure 4.11: Coefficient distribution for two metal plates with different thicknesses in 1D view. The template position was captured at the centre position $100mm$.

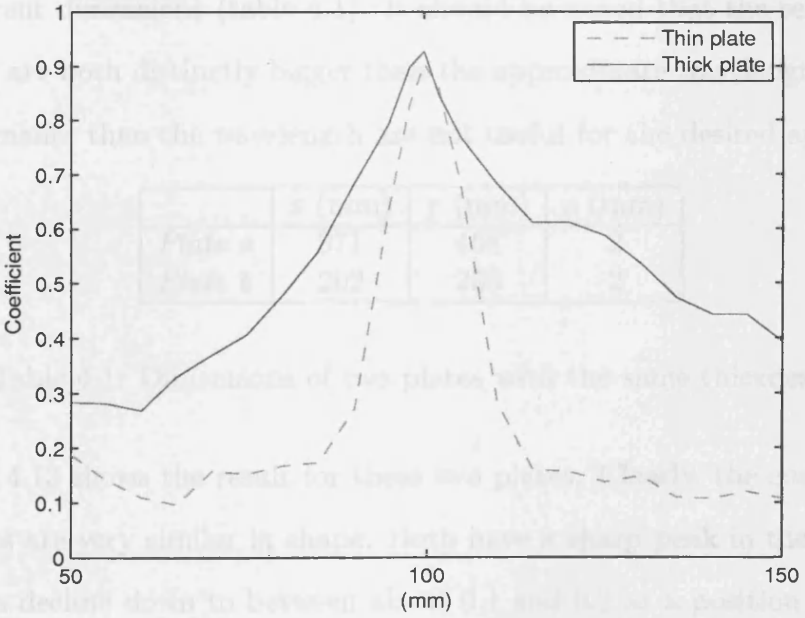


Figure 4.12: Coefficient distribution for two metal plates with different thicknesses in 1D view. The template position was captured at the centre position $100mm$.

$2mm$ respectively. The ratio between the wavelengths of the two plates can be obtained directly with $45/30 = 1.5$, which should be proportional to the ratio between the velocities of the two plates by calculating $\sqrt{b_1/b_2} = \sqrt{5/2} \approx 1.58$, where b_1 and b_2 represent the thickness of the thick and thin plates respectively. This has shown a very high agreement with the above analysis. In addition, this has further confirmed that the received wave is the dispersive flexural wave, as the non-dispersive p-wave and s-wave velocities do not depend on the thickness at all and should result in identical resolution on different thicknesses.

4.2.4 Effect of the plate size

In the last section, the resolution experiments were performed on different materials and compared in terms of thickness. However, as the sizes or dimensions of these plates are also different, it is worth investigating the influence on the resolution of the different dimensions of the surfaces. The purpose of this section is also to prove the independence of resolution from the geometry of the surface.

Two glass plates with the same thickness ($2mm$) are prepared for testing, but with different dimensions (table 4.1). It should be noted that the selected plate specimens are both distinctly bigger than the approximate wavelength. Surfaces that are smaller than the wavelength are not useful for the desired applications.

	x (mm)	y (mm)	z (mm)
Plate a	371	468	2
Plate b	202	255	2

Table 4.1: Dimensions of two plates with the same thickness.

Figure 4.13 shows the result for these two plates. Clearly, the coefficients for both plates are very similar in shape. Both have a sharp peak in the centre and coefficients decline down to between about 0.1 and 0.2 at a position $15mm$ out. This indicates the independence of resolution from different plate sizes.

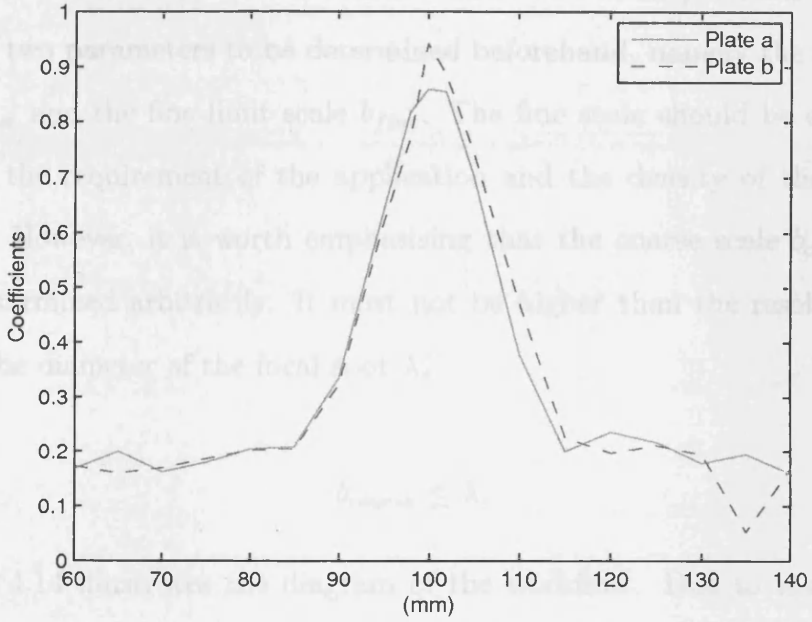


Figure 4.13: Coefficient distribution in 1D view. The impact position was at the centre position $100mm$.

4.3 Coarse-to-fine search

Despite the rapid progress of comtemporary computer development, to search across a large number of calibrated positions would still be very computationally demanding. To optimise performance, a coarse-to-fine searching technique can be employed.

This technique arises from the observation of the above resolution experiments. As can be seen, the resolution limit is dependent on the properties of the material. To search all positions is clearly not optimal. Thus, it is proposed to search from coarse to fine scales, rather than to search all positions one by one. There are two parameters to be determined beforehand, namely the coarse limit scale b_{coarse} and the fine limit scale b_{fine}. The fine scale should be easily determined by the requirement of the application and the density of the calibrated positions. However, it is worth emphasising that the coarse scale b_{coarse} should not be determined arbitrarily. It must not be higher than the resolution limit, which is the diameter of the focal spot λ,

$$b_{coarse} \leq \lambda.$$

Figure 4.14 illustrates the diagram of the workflow. Due to the occasional ambiguities of the local maximums and the global maximum, a simple two-step searching method is used, rather than the more widely known gradient search. In the first step, a coarse search is performed. The highest coefficient at the scale of a wavelength in the area of interest can be obtained. Then a finer scaled search will be carried out at the possible positions that are located within the limit of one wavelength around that observed target. In this way, the number of points to be compared will be reduced from $m \times n$ to $m + n - 1$, where m is the number of points to be searched at the coarse scale, and n is the number of points at the fine scale within one coarse scale. For example, on the glass plate with a resolution of about $40(mm)$, the coarse scale will be set at $40(mm)$. Suppose the

application requires the resolution of $10(mm)$ as the fine scale. For the area of $200 \times 200(mm^2)$, the number of points to be considered will be reduced from 400 to only $5 \times 5 + 4 \times 4 - 1 = 40$.

Practically, this algorithm was implementated using the C++ programming language on the Linux operating system. A laptop computer, with an Intel Core Duo T2400 processor, was used. The execution time measurement was carried out with different numbers of templates. With 800 templates, the latency is about $154ms$, and with 400 templates, the latency is about $67ms$. The signals were sampled at a sampling rate of $44100Hz$ for a length of 1024 samples. This is fast enough for real-time applications for such a large number of templates. Using the coarse-to-fine technique introduced above, this method becomes feasible for large surface applications.

4.4 Multiple impacts evaluation

Apart from the recognition of a single position, it has been observed that this method is capable of recognising multiple positions simultaneously. In this section, experiments for multiple impacts are performed, showing the potential and limitations of this method. When multiple impacts are conducted on the same surface, the received signal is actually the summation of all individual signals from all source positions with relative time shifts.

As shown in figure 4.15, it is found that, when two impacts were conducted simultaneously, the coefficients for both impact positions are still noticeable. The coefficient values are degraded to some extent due to the normalisation of correlation.

However, due to the resolution limitation caused by the material properties, the setup of a multiple impact application should be carefully designed. Every two pre-defined positions should be located with a minimum distance apart of a full wavelength. When two source positions are separated with a distance that is smaller than the wavelength, the two peaks adjoin, resulting in ambiguous local

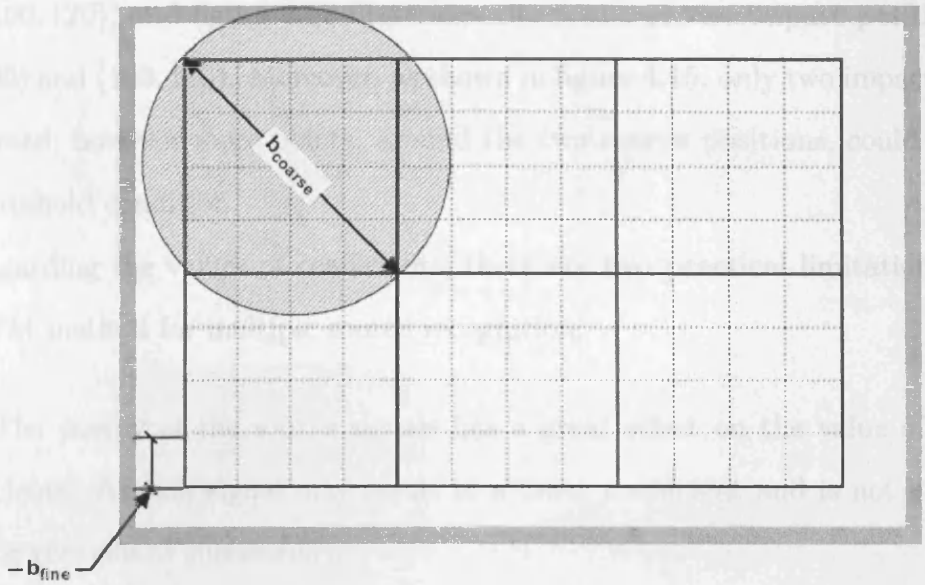


Figure 4.14: Diagram of coarse-to-fine search.

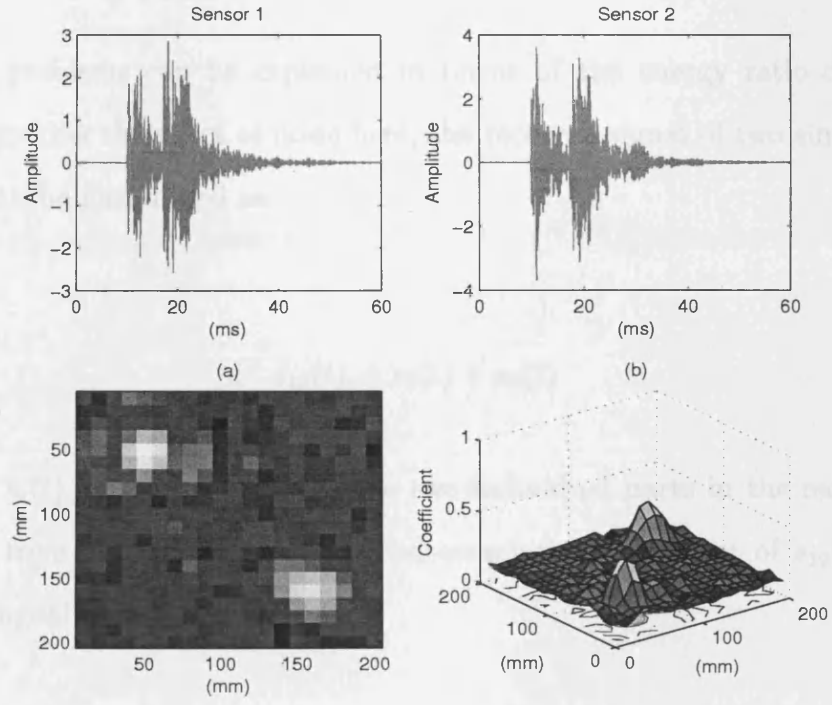


Figure 4.15: Coefficient distribution when two impacts are conducted simultaneously.

peaks. Figure 4.16 shows an example of two source positions located at $(100, 80)$ and $(100, 120)$, and figure 4.17 illustrates the result of two impact positions at $(100, 90)$ and $(100, 110)$. Moreover, as shown in figure 4.15, only two impacts were conducted; however more points, around the two source positions, could satisfy the threshold condition.

Regarding the values of coefficients, there are two practical limitations with the LTM method for multiple source recognition:

- The energy of the source signals has a great effect on the value of coefficients. A weak signal may result in a lower coefficient and is not likely to be recognised successfully.
- The other factor is the sample duration of the signals. The sample duration also affects the reliability. When the second signal is truncated in the sample captured by the sensor, the incomplete information of this signal will be difficult to recognise.

These problems can be explained in terms of the energy ratio of the two signals. Ignoring the effect of noise here, the received signal of two simultaneous impacts can be formulated as:

$$s_{12}(t) = s_1(t) + s_2(t) \tag{4.1}$$

where $s_1(t)$ and $s_2(t)$ represent the two individual parts in the received signal $s_{12}(t)$ from two positions. The cross-correlation coefficient of $s_{12}(t)$ with a template signal $s_T(t)$ is given by:

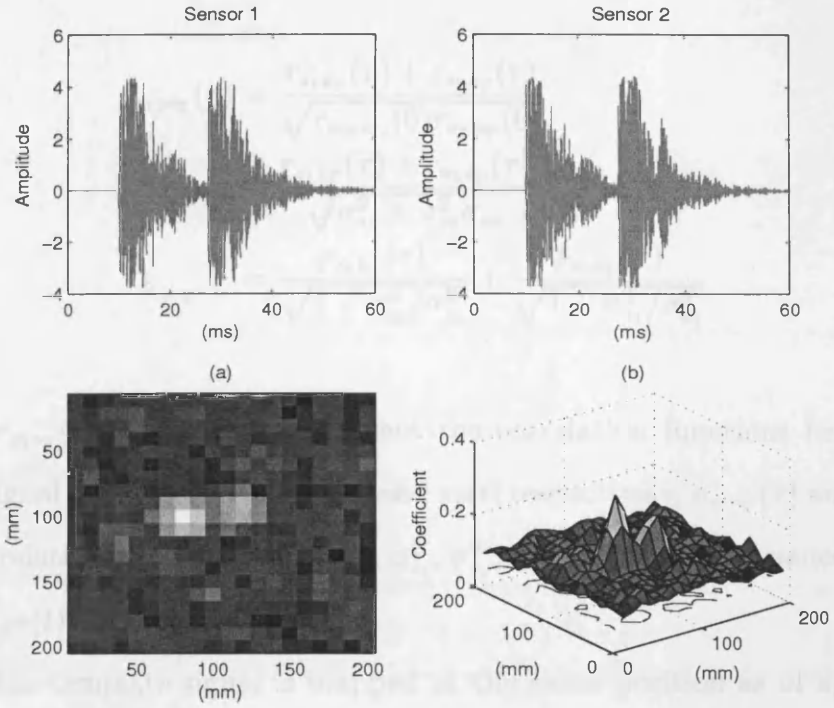


Figure 4.16: Example of two simultaneous impacts at positions of $(100, 80)$ and $(100, 120)$.

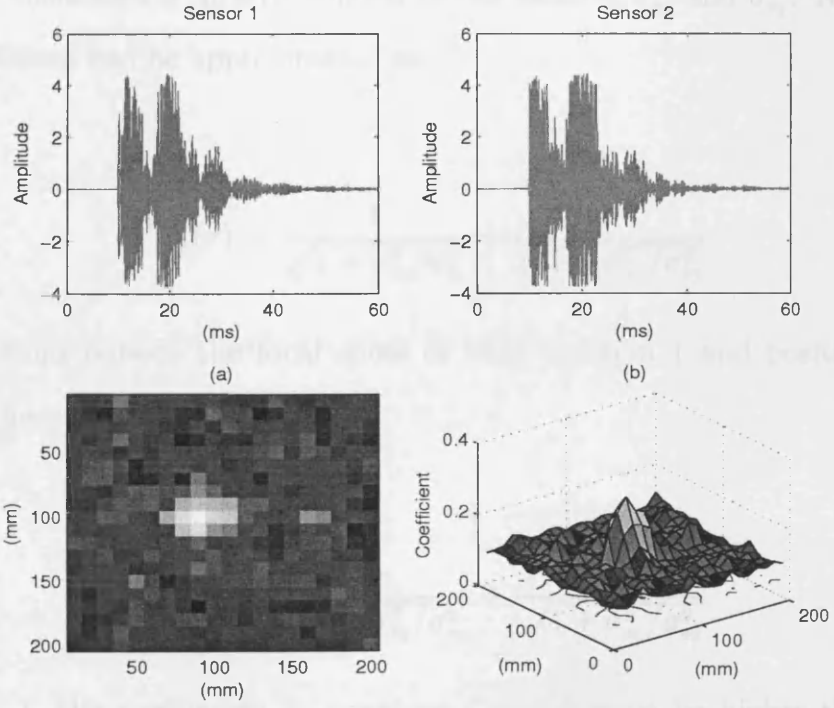


Figure 4.17: Example of two simultaneous impacts at positions of $(100, 90)$ and $(100, 110)$.

$$\begin{aligned}\rho_{s_{12}s_T}(\tau) &= \frac{r_{s_1s_T}(\tau) + r_{s_2s_T}(\tau)}{\sqrt{r_{s_{12}s_{12}}(0) r_{s_Ts_T}(0)}} \\ &\approx \frac{r_{s_1s_T}(\tau) + r_{s_2s_T}(\tau)}{\sqrt{\sigma_{s_1}^2 + \sigma_{s_2}^2}\sigma_{s_T}} \\ &= \frac{\rho_{s_1s_T}(\tau)}{\sqrt{1 + \sigma_{s_2}^2/\sigma_{s_1}^2}} + \frac{\rho_{s_2s_T}(\tau)}{\sqrt{1 + \sigma_{s_1}^2/\sigma_{s_2}^2}}\end{aligned} \tag{4.2}$$

where $r_{s_1s_T}(\tau)$ and $r_{s_2s_T}(\tau)$ represent the correlation functions between the template signal $s_T(\tau)$ and signals $s_1(t)$ and $s_2(t)$ respectively, $\rho_{s_1s_T}(\tau)$ and $\rho_{s_2s_T}(\tau)$ are the correlation coefficients of them, $\sigma_{s_1}^2$, $\sigma_{s_2}^2$, and $\sigma_{s_T}^2$ are the variances of $s_1(t)$, $s_2(t)$, and $s_T(t)$.

When the template signal is mapped at the same position as of s_1, $\rho_{s_1s_T}(\tau)$ reaches its maximum of about 1 at the corresponding time shift. When position 2 is far from position 1 (outside the spatial focal spot of position 1), $\rho_{s_2s_T}(\tau)$ can be considered as a constant c, which is lower than 1. The value of the correlation coefficient is directly related to the ratio of $\sigma_{s_1}^2$ and $\sigma_{s_2}^2$. At position 1, the coefficient can be approximated as:

$$\rho_{s_{12}s_T}(\tau) = \frac{1}{\sqrt{1 + \sigma_{s_2}^2/\sigma_{s_1}^2}} + \frac{c}{\sqrt{1 + \sigma_{s_1}^2/\sigma_{s_2}^2}} \tag{4.3}$$

At positions outside the focal spots of both position 1 and position 2, the coefficient becomes:

$$\rho_{s_{12}s_T}(\tau) = \frac{c}{\sqrt{1 + \sigma_{s_2}^2/\sigma_{s_1}^2}} + \frac{c}{\sqrt{1 + \sigma_{s_1}^2/\sigma_{s_2}^2}} \tag{4.4}$$

As $c < 1$, the coefficients at positions 1 and 2 must be higher than other positions. This indicates the potential of the multiple impacts problem. However, the contrast of the coefficients is degraded to $\frac{1-c}{\sqrt{1+\sigma_{s_2}^2/\sigma_{s_1}^2}}$, that affects its reliability

·significantly.

In the case of $\sigma_{s_1} = \sigma_{s_2}$, the correlation coefficient can be given by:

$$\rho_{s_{12}s_T}(\tau) = \frac{\rho_{s_1s_T}(\tau)}{\sqrt{2}} + \frac{c}{\sqrt{2}} \tag{4.5}$$

At positions 1 and 2, the coefficient can be approximated as:

$$\rho_{s_{12}s_T}(\tau) = \frac{1+c}{\sqrt{2}} \tag{4.6}$$

At positions outside the focal spots of both position 1 and position 2, the coefficient is:

$$\rho_{s_{12}s_T}(\tau) = \frac{c}{\sqrt{2}} + \frac{c}{\sqrt{2}} = \sqrt{2}c \tag{4.7}$$

Considering the extreme case that $\sigma_{s_1} \gg \sigma_{s_2}$, the coefficient $\rho_{s_{12}s_T}(\tau)$ becomes:

$$\rho_{s_{12}s_T}(\tau) = \begin{cases} 1, & \text{at position 1} \\ \rho_{s_1s_T}(\tau), & \text{at other positions} \end{cases} \tag{4.8}$$

Similarly, for the extreme case that $\sigma_{s_2} \gg \sigma_{s_1}$, the coefficient $\rho_{s_{12}s_T}(\tau)$ becomes:

$$\rho_{s_{12}s_T}(\tau) = \begin{cases} 1, & \text{at position 2} \\ \rho_{s_2s_T}(\tau), & \text{at other positions} \end{cases} \tag{4.9}$$

In such cases, the part of signal with a small portion will be totally ignored, and will not be recognised.

In fact, these problems are unavoidable in practice, as one can not always

tap at two positions simultaneously with similar strength. However, this method has shown its potential for the multiple impacts problem. Although most of the HCI applications nowadays do not demand such functionality, it is still worth investigating for improvement in the future.

4.5 Summary

In this chapter, the resolution of the proposed LTM method has been analysed. First, three aspects of the sensor placement, including the symmetrical locations, the number of sensors, and the optimal sensor location, have been discussed. Next, experiments were performed on different materials and different thicknesses. The achievable resolution was found to be related to some factors, including the frequency bandwidth of the received signals, the dispersive properties of the materials, the type of impacts, the sensor characteristics, and so on. In addition, the coarse-to-fine searching technique has been applied to reduce the computation time. Finally, multiple simultaneous impacts have been evaluated, so that two impacts can be localised or identified when the two signals are both acquired by the computer adequately. When the acquired signal does not include enough energy from one of the impacts, it is less reliable. An understanding of the resolutions for different materials is useful for designing appropriate applications.

Chapter 5

TDOA methods and the Wavelet transform with maximum likelihood estimation

In this chapter, time difference methods will be discussed. First, two conventional approaches, the Time-of-Flight (ToF) and cross-correlation methods, will be introduced in sections 5.1 - 5.3. Section 5.4 presents a method to measure and validate the phase velocities of flexural waves, which are used to calculate the group velocities. Section 5.5 deals with the problem of the localisation of dispersive signals using the Continuous Wavelet Transform (CWT). Then, statistical estimation of maximum likelihood is proposed to improve the accuracy and reliability of acoustic localisation.

5.1 Source location in a plane

The principle of the TDOA method is to use the measurement of time differences between arrivals of signals to estimate the source location [PAKJ+05][PJP+06]. These time arrival delays result from the different distances from the source to the sensors, which are placed at known geometric positions.

Figure 5.1 presents a typical TDOA system comprising a plate, four sensors

installed at the four corners, signal conditioning hardware, a data acquisition card and a PC. A knock on the plate can be detected by the sensors. The sensor outputs are amplified and filtered, then digitised and finally processed by the PC to obtain the coordinates of the position of the knock.

Theoretically, three sensors are sufficient for localisation in such an application. When four sensors are used for the localisation of an impact, the additional sensor is used to reduce the error. The source location (x, y) has the following relationship with the time differences of the sensor received signals:

$$\sqrt{(x-x_1)^2+(y-y_1)^2}-\sqrt{(x-x_3)^2+(y-y_3)^2}=c\Delta t_{13} \tag{5.1}$$

and

$$\sqrt{(x-x_2)^2+(y-y_2)^2}-\sqrt{(x-x_4)^2+(y-y_4)^2}=c\Delta t_{24} \tag{5.2}$$

where (x_1, y_1), (x_2, y_2), (x_3, y_3) and (x_4, y_4) are the coordinates of the four sensor positions respectively, c is the acoustic wave propagation velocity in the plate, Δt_{13} is the true time difference for the sensor locations (x_1, y_1) and (x_3, y_3), and Δt_{24} is the time difference for the sensor locations (x_2, y_2) and (x_4, y_4).

The above equations can be expressed in the general form

$$\Delta d_{i,j}=\sqrt{(x-x_i)^2+(y-y_i)^2}-\sqrt{(x-x_j)^2+(y-y_j)^2} \tag{5.3}$$

where i and j represent the sensor indices. The source location can therefore be calculated by minimising the following expression:

$$E=\sum_{i=1}^{4}\sum_{j=2}^{4}(c\Delta t_{i,j}-\Delta d_{i,j})^2, \quad j>i \tag{5.4}$$

where $\Delta t_{i,j}$ represents the time difference between signal arrivals at the ith

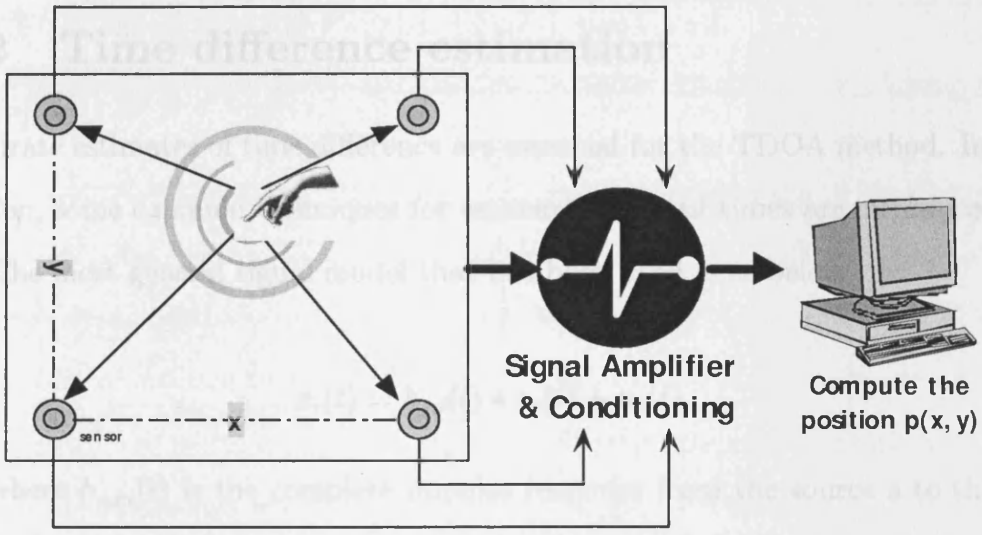


Figure 5.1: Typical TDOA system with four sensors.

sensor and the jth sensor for all the four sensors.

Taking into account external noise, it is proposed to introduce a threshold to determine the confidence of the signal quality. A confident estimation of source location can be obtained by a very low value of E, which must be non-negative. The threshold needs to be decided, and any E that is higher than the threshold will be considered a false trigger.

It should be pointed out that the determination of the source location in a plane is based on the hypothesis of acoustic waves travelling in the plate at a constant velocity.

5.2 Time difference estimation

Accurate estimates of time difference are essential for the TDOA method. In this section, some common techniques for estimating arrival times are introduced.

The most general signal model that has been used is as below:

$$x_i(t) = h_{s,i}(t) * s_s(t) + n_i(t), \tag{5.5}$$

where $h_{s,m}(t)$ is the complete impulse response from the source s to the ith sensor, $s_s(t)$ represents the source signal, and $n_i(t)$ is the additive white noise, which is uncorrelated with the source signals and noises from other sensors. The symbol $*$, as used previously, denotes the temporal convolution. Normally, researchers are interested in sound source localisation problems in reverberant environments. In addition, attenuation factors should also be considered in practice. The impulse response from the source position to the ith sensor can, therefore, be described as two parts:

$$x_i(t) = a \cdot s_s(t) \cdot h_s(t - \tau) + a \cdot s_s(t) \cdot r_{s,i}(t - \tau) + n(t), \tag{5.6}$$

where the nominal response is represented by $h_s(t - \tau)$ and interfering echoes $r_{s,i}(t - \tau)$ are represented separately. With the assumption that the waveform

·is transmitted at time $t = 0$, τ is the ToF. Equation 5.6 simply states that the observed waveform equals the nominal response $s \cdot h(t)$, but time-shifted by τ and attenuated by a. Such an assumption is correct for a medium like air as the propagation of a waveform through air does not show a significant dispersion within the bandwidth of interest. The attenuation coefficient a depends on many factors, but also on the distance, and thus also on τ. The possible echoes are represented by $a \cdot r(t - \tau)$. They share the same time shift τ because no echo can occur before the arrival of the nominal response. The additional reverberations are also modelled in $r(t)$ implicitly.

5.2.1 The Time-of-Flight (ToF) method

Introduction

The simplest method used to determine the arrival time of a signal is to detect its rising-edge, also referred as the Time-of-Flight (ToF). Some applications require cheap solutions that are suitable for direct implementation using dedicated hardware, such as instrumental electronics. For that reason, a simple thresholding of the observed waveform at a level T has gained great interest. The estimated ToF $\hat{\tau}_{thres}$ is the moment at which the waveform crosses a threshold level, as illustrated in figure 5.3.

Filtering and normalisation

Figure 5.2 shows an example of typical signals acquired at two sensors. Clearly, the signal after propagation over different distances is dispersed variously. The velocity dispersion can be observed by the shape change of the waveforms. Attenuation is another factor that may impair the accuracy of the ToF determination by an arbitrary selected threshold.

In practice, the value of the threshold should be tuned experimentally in order to achieve an accurate localisation result. The threshold level T should be normally chosen just above the maximum possible noise level. However, due

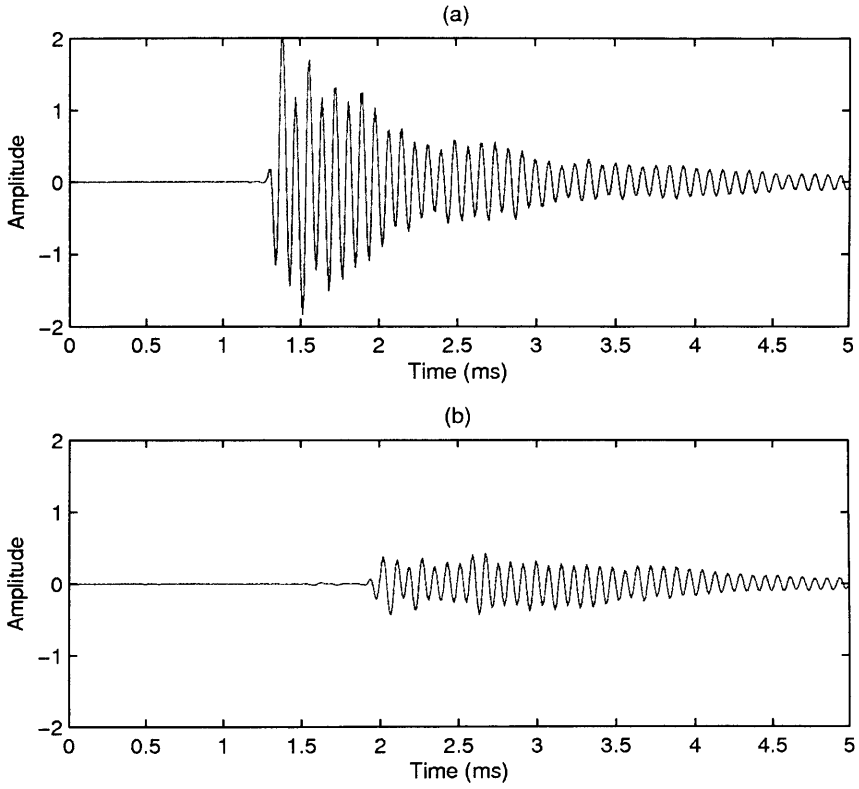


Figure 5.2: Typical signals acquired by two sensors.

to the slow rising of the nominal response, sometimes the moment $\hat{\tau}_{thres}$ of level crossing appears just after the true τ and that may create a bias. Therefore, it was found that a simple threshold does not work well, due to the lack of the information about signal magnitude and shape for each channel. Hence, a threshold is normally determined as a value related to the maximum of the waveform, at $T = \alpha \max(x(t))$. α is a constant set, for instance, to 30%, which is identical to the process of normalisation.

In addition to designing a proper threshold, the accuracy and reliability of the localisation can also be improved by proper filtering of the signals.

Envelope extraction

On the other hand, a simple narrowband signal can also be modelled as $g(t)\cos(2\pi ft + \varphi) + n(t)$, which contains the envelope $g(t)$ and the carrier $\cos(2\pi ft + \varphi)$. To extract the ToF of a signal, the carrier part of the waveform may cause a resolution error equal to $1/f$. Therefore, rather than applying the threshold operation directly, the envelope is applied instead. The simplest method of extracting the envelope is to apply a very low-pass filter to the signal. Another solution could be using techniques of interpolation to minimise the error. However, the optimal method is to use quadrature filtering. First, the waveform is filtered to reduce the noise. Next, the filter signal is phase-shifted over 90° to obtain the quadrature component $q(t) = g(t)\sin(2\pi ft + \varphi) + n_q(t)$. Finally, the envelope is estimated using $\hat{g}(t) = \sqrt{w^2_{band-filtered}(t) + q^2(t)}$. The Hilbert transform provides a simple solution to obtain the quadrature component, as below:

$$\hat{s}(t) = \mathcal{H}\{s\}(t) = h(t) * s(t) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{s(\tau)}{t-\tau} d\tau. \tag{5.7}$$

where $h(t) = \frac{1}{\pi t}$, representing the system impulse response, and $\mathcal{H}$ denotes the Hilbert transform.

Figure 5.3 provides an example of the envelope extracted from a real vibrational signal.

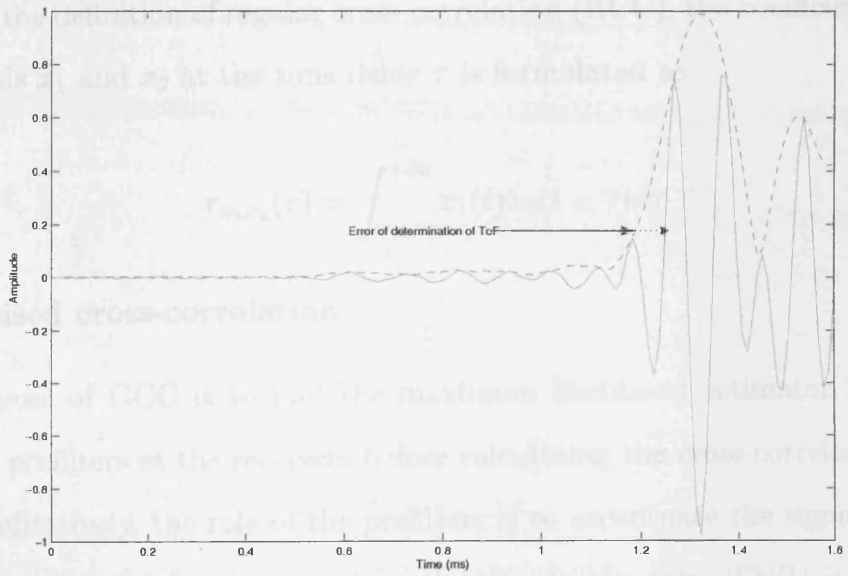


Figure 5.3: Example of envelope extraction from a real signal.

5.2.2 Cross-correlation and Generalised cross-correlation

The cross-correlation function is the most popular method to determine the time difference between two signals received at two spatially separated sensors in the presence of uncorrelated noise. The maximum of the cross-correlation coefficients of two signals indicates the lag of time between them. Usually, cross-correlation (CC) is considered in two types: regular cross-correlation (RCC) and generalised cross-correlation (GCC).

Regular cross-correlation

To recall the definition of regular cross-correlation (RCC), the coefficient between two signals x_1 and x_2 at the time delay τ is formulated as

$$r_{x_1x_2}(\tau) = \int_{-\infty}^{+\infty} x_1(t)x_2(t+\tau)d\tau \tag{5.8}$$

Generalised cross-correlation

The purpose of GCC is to find the maximum likelihood estimator [KC76], by applying prefilters at the receivers before calculating the cross correlation (figure 5.4). Qualitatively, the role of the prefilters is to accentuate the signal passed to the correlator at the frequencies where the signal noise ratio (SNR) is the highest and to suppress the noise power.

The generalised correlation between $x_1(t)$ and $x_2(t)$ is described as

$$r^g_{x_1x_2}(\tau) = \int_{-\infty}^{\infty} \psi_g(f)G_{x_1x_2}(f)e^{j2\pi f\tau}df, \tag{5.9}$$

where

$$\psi_g(f) = H_1(f)H_2^*(f) \tag{5.10}$$

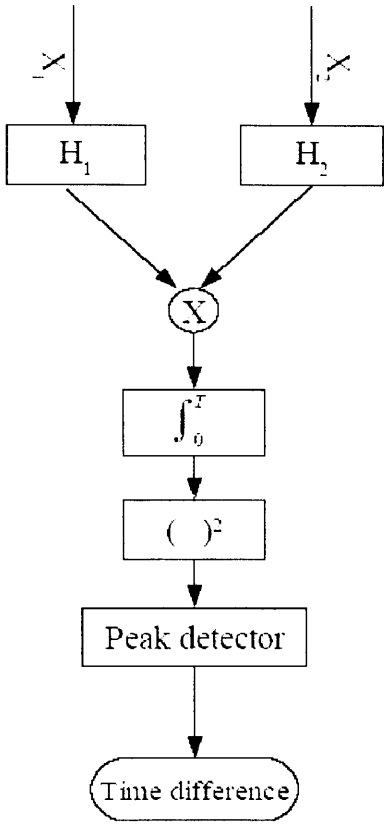


Figure 5.4: Flowchart of the GCC method. Prefilters H_1 and H_2 are used before applying to cross-correlator.

and $G_{x_1x_2(f)}$ is the cross power spectral density function of $x_1(t)$ and $x_2(t)$. Symbol $*$ denotes the complex conjugate.

In general, there are five different GCC pre-filters [KC76], namely Roth, SCOT, PHAT, Eckart, and HT. Each of them has its own advantages and disadvantages for different conditions [KC76]. PHAT is the most popular one and has been widely used in many source localisation applications, defined as:

$$\psi^g(f) = \frac{1}{|G_{x_1x_2}(f)|} \tag{5.11}$$

PHAT is in fact identical to the method for the template matching problem used in the last two chapters. It has been shown that, in plate structures, the signal is distorted significantly by the dispersion and edge reflections. The template matching method provides a very low coefficient for two signals from two different locations. This is an indication of the unsuitability of PHAT for time difference estimation here, due to the reciprocity between the source and receivers.

The drawbacks of the CC methods can be briefly summarised in the two points below:

- The true peak corresponding to the source location is not always represented as a global maximum, due to the strong effect of the multipath propagation and the frequency dispersion caused by the nature of plate wave mode. As a result, ambiguous peaks will appear in the correlation coefficients causing abnormal results.

- In the presence of uncorrelated white noise, a peak is always observed in the middle of the correlation coefficients, resulting in a source location calculated to be in the centre of the plate. This type of noise is particularly common from electronics devices.

5.3 Experimentation and discussion

This section first gives a description of the hardware setup. Then the above two methods are analysed experimentally. Their advantages and disadvantages will be discussed.

5.3.1 Experimentation hardware

The experiments were carried out on a thin Medium Density Fibreboard (MDF) board with dimension of $1200 \times 900 \times 6(mm^3)$, as shown in figure 5.5. The sensors were located at the four corners of the central experimental area of $600 \times 400(mm^2)$.

5.3.2 Resolution limit caused by sampling rate

The sampling rate f_s of the analog-to-digital conversion is one of the factors that can affect the resolution limit of the TDOA method. A resolution limit arises due to the finite sampling frequency of a practical system of which the estimated time difference τ is quantised. This time difference τ is then described as a multiple of the time interval of $T_s = \frac{1}{f_s}$. The digitised time delay N_s is then expressed as:

$$N_s = \frac{L}{cT_s} = \frac{f_s L}{c} \tag{5.12}$$

where L is the distance between the two sensors. The implication of this relation is that f_s and L are proportional to the time step N_s. Accordingly, a higher sampling rate f_s can have a bigger value of N_s, so that the system's resolution is improved, as shown in figure 5.6. In the experiment, the maximum distance for all pairs of sensors is on the diagonals, that is $\sqrt{600^2 + 400^2} = 721mm$. With the sampling rate of $100kHz$ and an approximate velocity $\approx 600m/s$, the highest achievable resolution is $6mm$ and $N_s = 120$. This is high enough for this application.

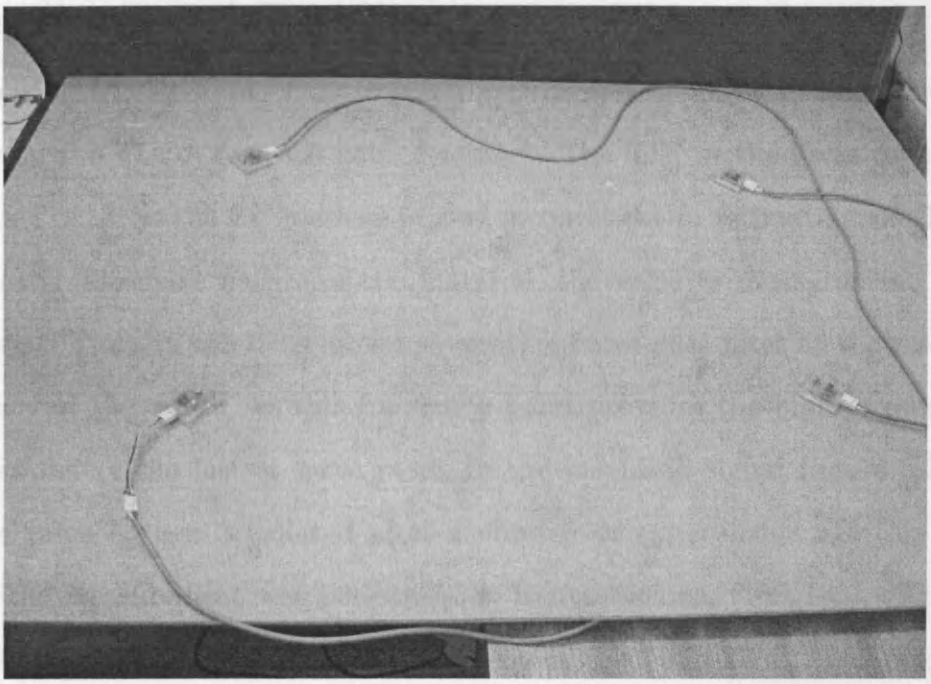


Figure 5.5: TDOA experimental setup using an MDF plate and four accelerometers.

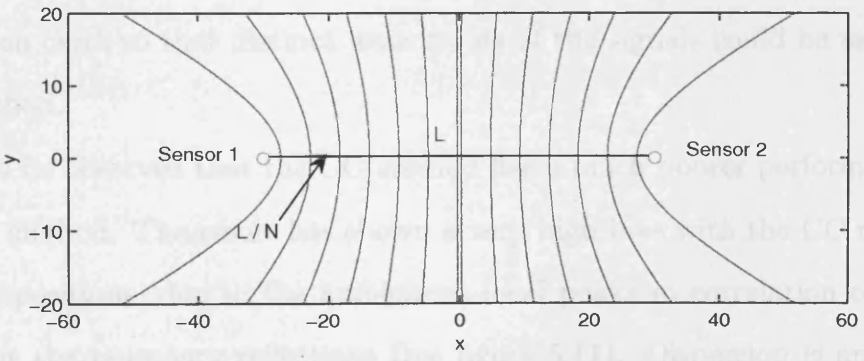


Figure 5.6: Example of resolution limit caused by sampling rate in one-dimension.

5.3.3 Comparison of the ToF and CC methods

In this sub-section, performance of both the ToF and the CC methods are compared and discussed experimentally.

Knowing the location of the tap, an approximate speed can be deduced by measuring the TDOA for each pair of sensors. The ToF method was used to decide the TDOA, as the CC method proved inconsistent in estimating the TDOA. Due to the dispersive nature of the material, the velocity measurement is very influential. Thus, it was determined to apply a band-pass filter at the resonance frequency of the sensor, as this frequency band provides the highest sensitivity that can detect the fastest wave mode in the incoming signal (figure 5.7). An average value is then calculated after a number of experiments. In this experiment, the measurement was performed at four positions, $(100, 100)$, $(500, 100)$, $(100, 300)$, and $(500, 300)$ for two pairs of sensors, $1-3$ and $2-4$. An approximate velocity was obtained as $650m/s$.

As before, experiments were carried out on data collected from impacts of nails and fingers respectively. Each set of signal data included 25 samples at three positions: $(200, 100)$, $(400, 300)$, and the centre position $(300, 200)$. Figure 5.8, 5.9 and 5.10 show the results of the two methods respectively. Tables 5.1, 5.2, and 5.3 show the standard deviations and the mean values of the results for both methods. All the signals were sampled in the pre-trigger mode of the data acquisition card, so that distinct wave fronts of the signals could be used by the ToF method.

It can be observed that the CC method has a much poorer performance than the ToF method. The result has shown a very high bias with the CC method at all three positions, due to the ambiguous local peaks in correlation coefficients caused by the boundary reflections (see figure 5.11). Dispersion is another important factor that impairs the performance. In contrast, it can be seen that the ToF method performed much better. However, it too is affected by the material dispersion. The signals received by different sensors at different distances

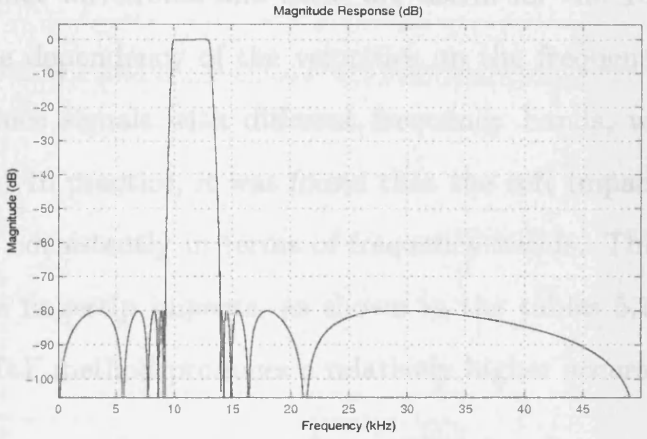


Figure 5.7: Band-pass digital filtering.

Method (impact)	Mean (x)	Mean (y)	Std (x)	Std (y)
ToF (nail)	204.800	92.000	5.099	5.773
ToF (finger)	203.600	101.200	11.136	14.236
CC (nail)	124.400	117.200	87.038	106.99
CC (finger)	104.800	194.800	85.348	89.634

Table 5.1: Statistical result of the ToF and the CC methods at position $(200, 100)$.

must contain different levels of dispersion and attenuation. For this reason, the ToF method performed slightly less accurately at the positions $(200, 100)$ and $(400, 300)$ than at the centre position $(300, 200)$.

When comparing the two different impacts, the signals made by the nails have more distinct wavefronts and these are useful for the ToF method. Moreover, due to the dependency of the velocities on the frequencies, the two types of impact produce signals with different frequency bands, which propagate at different speeds. In practice, it was found that the soft impacts (fingers) do not always perform consistently in terms of frequency bands. This results in higher variance for the fingertip impacts, as shown in the tables 5.1, 5.2, and 5.3. In summary, the ToF method produces a relatively higher accuracy with nails than with fingers.

It should be noted that the ToF method is only suitable when the wave fronts of signals are available. Clearly, one disadvantage is the lack of the capability to locate signals without a wave front, such as signals generated by continuous scratching. Another disadvantage of the ToF method is its vulnerability in noisy environments.

5.4 Dispersion estimation by measuring phase velocity

The results above have shown that conventional cross-correlation (and also GCC) is not suitable for this application, due to the strongly dispersed waveform. As

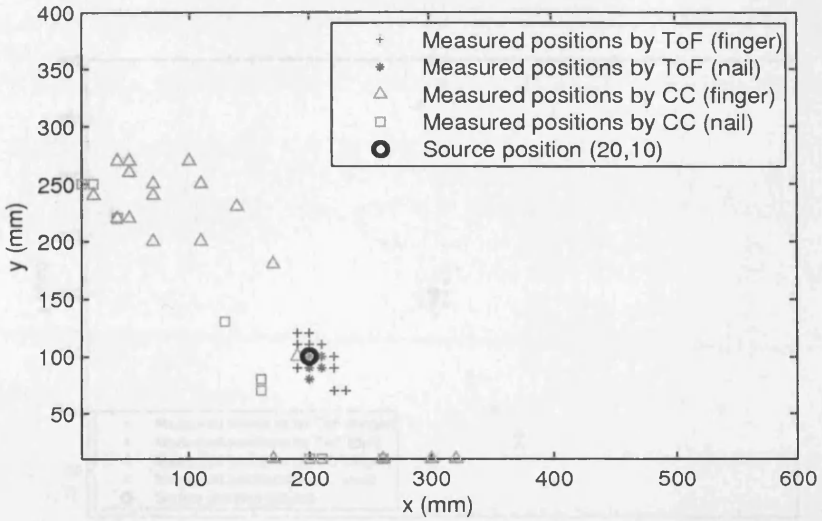


Figure 5.8: Result of the ToF and the CC methods at position $(200, 100)$.

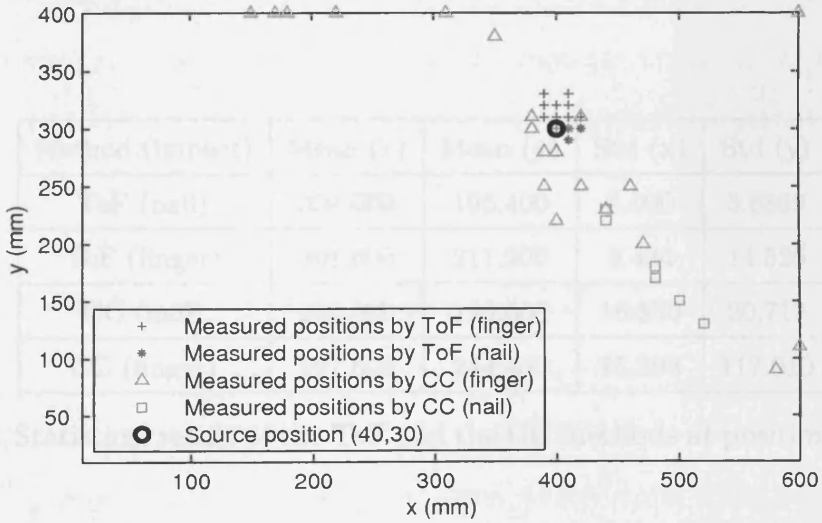


Figure 5.9: Result of the ToF and the CC methods at position $(400, 300)$.

Method (impact)	Mean (x)	Mean (y)	Std (x)	Std (y)
ToF (nail)	412.800	299.600	4.5826	2.0000
ToF (finger)	407.600	310.400	8.3066	10.198
CC (nail)	475.200	182.400	28.449	36.661
CC (finger)	372.800	298.800	128.830	91.210

Table 5.2: Statistical result of the ToF and the CC methods at position $(400, 300)$.

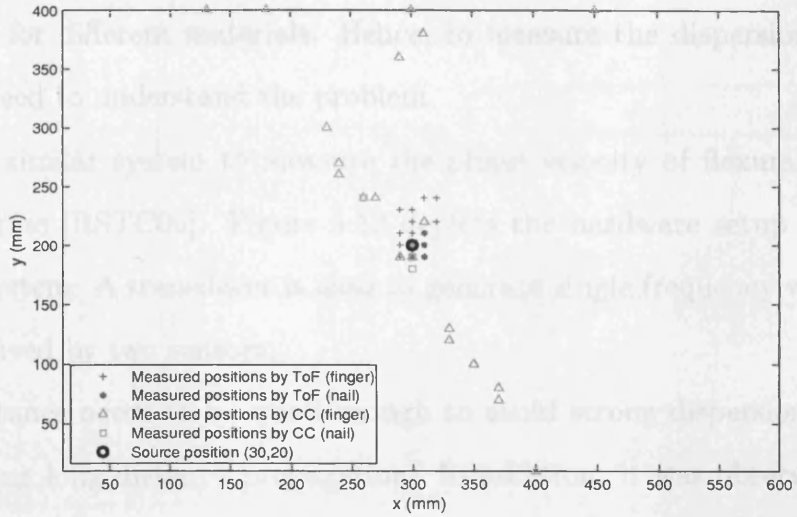


Figure 5.10: Result of the ToF and the CC methods at position (300, 200).

Method (impact)	Mean (x)	Mean (y)	Std (x)	Std (y)
ToF (nail)	304.000	196.400	5.000	5.6859
ToF (finger)	301.600	211.200	9.434	14.526
CC (nail)	292.000	199.600	16.330	20.712
CC (finger)	297.600	244.400	75.293	117.010

Table 5.3: Statistical result of the ToF and the CC methods at position (300, 200).

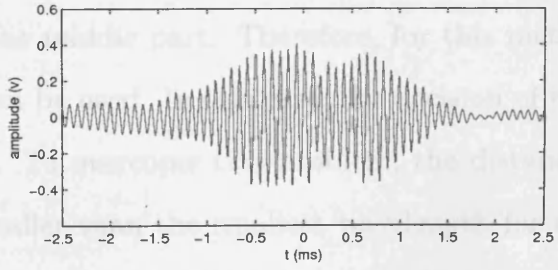


Figure 5.11: Cross-correlation coefficients with ambiguous peaks.

described above, the dispersion of phase velocities can be calculated with prior knowledge of the material properties. However, it is not practical to obtain all the parameters for different materials. Hence, to measure the dispersion properties there is a need to understand the problem.

Here, a similar system to measure the phase velocity of flexural waves was employed from [RSTC06]. Figure 5.12 depicts the hardware setup of the measurement system. A transducer is used to generate single frequency waves, which can be received by two sensors.

The distance needs to be small enough to avoid strong dispersion and attenuation during long distance propagation. In addition, it was observed that the quality of the received signals is strongly affected by the resonance frequency of the accelerometer. Figure 5.13 shows a received signal and its Fourier transform, generated by the transducer located $150(mm)$ from the sensor (50 cycles of sinusoid at frequency of $4000Hz$). It can be seen that the transmitted frequency is quite weak compared to the resonance frequency. A smaller distance can reduce this difference.

Sinusoidal waves in the frequency range from $1000Hz$ to $5000Hz$ were produced by the transducer for every $250Hz$. Two methods have been employed to determine the phase shift between two received signals.

The first is to measure the time at zero crossing in the time domain. The idea is to obtain the phase point of a sinusoid wave, which starts from zero. When the start point is observable, the time difference between the two sinusoid waves can then be calculated. However, as seen above, the start point of the received signal is not as clean as the middle part. Therefore, for this method, only the middle part of the signal can be used. In this case, the decision of the phase point would become ambiguous. To overcome this problem, the distance d between the two sensors must be smaller than the smallest wavelength for all frequencies in this experiment. This will guarantee that the phase difference is smaller than a full cycle 2π.

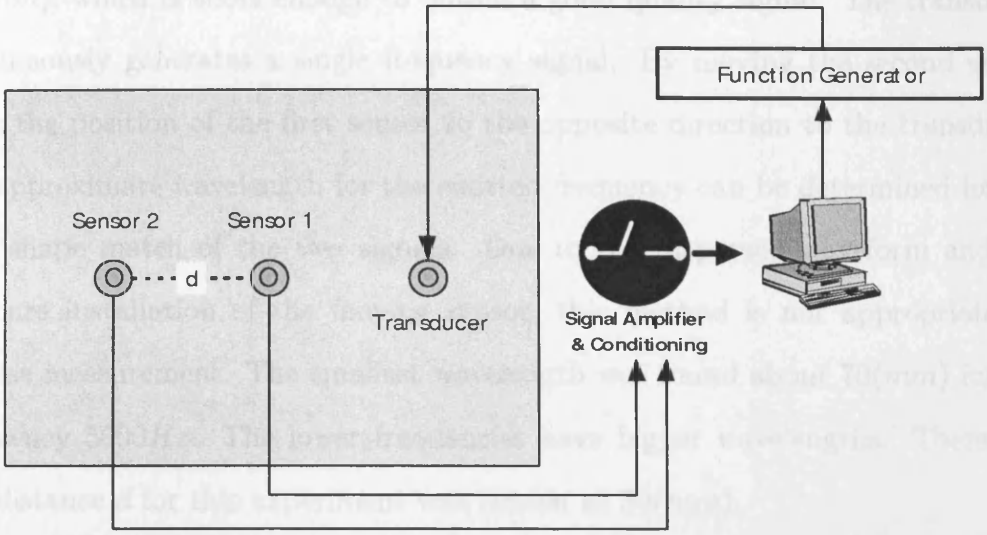
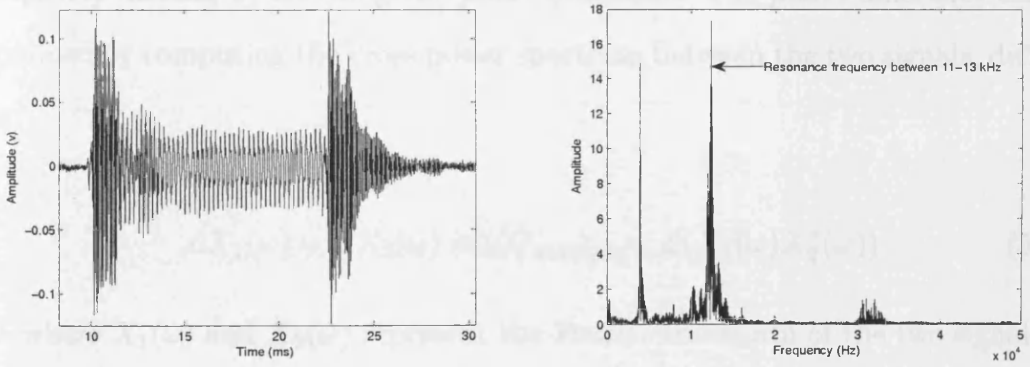


Figure 5.12: Hardware setup for phase velocity measurement.



(a) (b)

Figure 5.13: The signal (a) received at the sensor generated by the transducer at the frequency of $4000Hz$. The Fourier transform (b) shows the strong amplitude at the resonance frequency of $11kHz - 13kHz$.

To determine the smallest wavelength, a simple technique was employed for this purpose. The first sensor and the transducer are fixed with a distance of $50(mm)$, which is short enough to obtain a good quality signal. The transducer continuously generates a single frequency signal. By moving the second sensor from the position of the first sensor to the opposite direction to the transducer, the approximate wavelength for the emitted frequency can be determined for the first shape match of the two signals. Due to the dispersed waveform and the insecure installation of the moving sensor, this method is not appropriate for precise measurement. The smallest wavelength was found about $70(mm)$ for the frequency $5000Hz$. The lower frequencies have bigger wavelengths. Therefore, the distance d for this experiment was chosen as $50(mm)$.

As expected, the signal at the second sensor is more dispersed and attenuated compared to the other one (figure 5.14). A cleaner signal can be obtained by applying proper digital filtration. Then the time difference can be determined by the two closest zero crossing points for the two signals. Figure 5.15 illustrates an example of the phase difference of two normalised signals at $4000Hz$.

The second method of measuring the phase velocity is carried out in the frequency domain by checking the phase difference. The phase difference can be obtained by computing the cross power spectrum between the two signals, defined as:

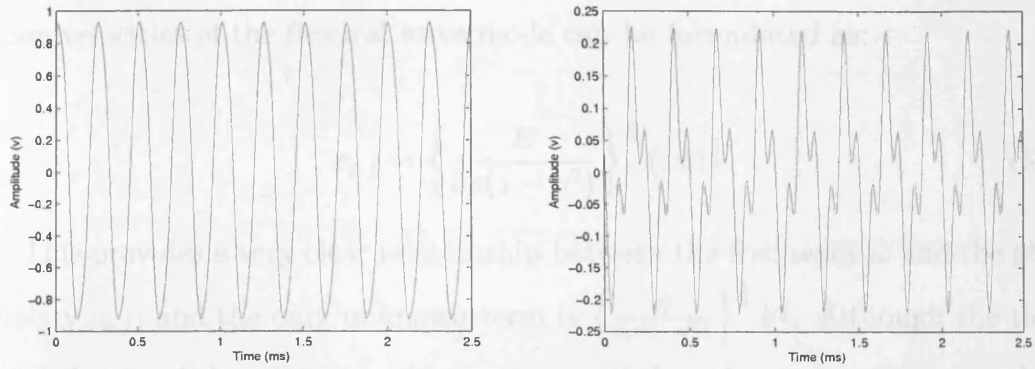
$$\angle X_1(\omega) - \angle X_2(\omega) = \angle G_{x_1x_2(\omega)} = \angle(X_1(\omega)X_2^*(\omega)) \tag{5.13}$$

where $X_1(\omega)$ and $X_2(\omega)$ represent the Fourier transform of the two signals, $\angle$ denotes the angle, and $*$ denotes the conjugation.

The time difference can be obtained with the following relationship:

$$c_p(\omega) = \frac{\omega d}{\angle G_{x_1x_2(\omega)}} \tag{5.14}$$

The experiments were carried out a number of times with both methods. The



(a) (b)

Figure 5.14: The signal received by the first sensor (a) and second sensor (b), located $50mm$ apart from each other.

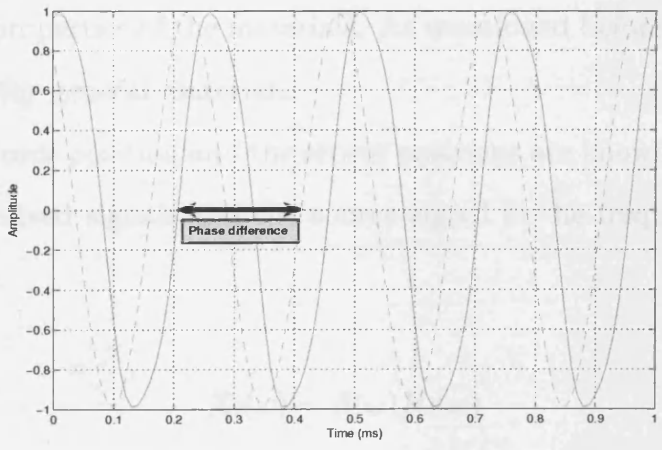


Figure 5.15: Phase difference is determined by measuring the two closest zero crossing points.

average values of these experiments are used as the final optimum results. The measured phase velocities are shown in figure 5.16.

Recalling the dispersion relation introduced in chapter 2, the frequencies and phase velocities of the flexural wave mode can be formulated as:

$$c_{p,f} = \left\{ \frac{E}{3\rho(1-v^2)} \right\}^{\frac{1}{4}} (\omega b)^{\frac{1}{2}} \tag{5.15}$$

This provides a very clear relationship between the frequency ω and the phase velocity $c_{p,f}$, and the only unknown term is $\left\{ \frac{E}{3\rho(1-v^2)} \right\}^{\frac{1}{4}} b^{\frac{1}{2}}$. Although the thickness is known, it is still impossible to obtain all the other parameters directly. A simple curve fitting analysis of the measurement with this formula is then carried out with the least square error criterion. The dispersive phase velocities are plotted against the respective frequency, as shown in figure 5.16.

5.4.1 Blind verification of measured velocities

In order to validate the accuracy of the measured velocity, the most straightforward method should be to compare the estimated values against the theoretical calculation of the dispersion curve. However, this requires the precise knowledge of the physical properties of the materials. As mentioned before, this information is not available for general materials.

When the source position and the sensor positions are known, the relationship between the received signals and the source signal in the frequency domain can be described as:

$$X_i(\omega) = S(\omega) H_i(\omega) \tag{5.16}$$

where $H_i(\omega)$ represents the propagation response of the source $S(\omega)$ and the received signal $X_i(\omega)$ at the ith sensor for different frequencies in the frequency domain. The attenuation factors are assumed to be negligible here. When reflections are not considered, $H(\omega)$ can be simply described with the different time

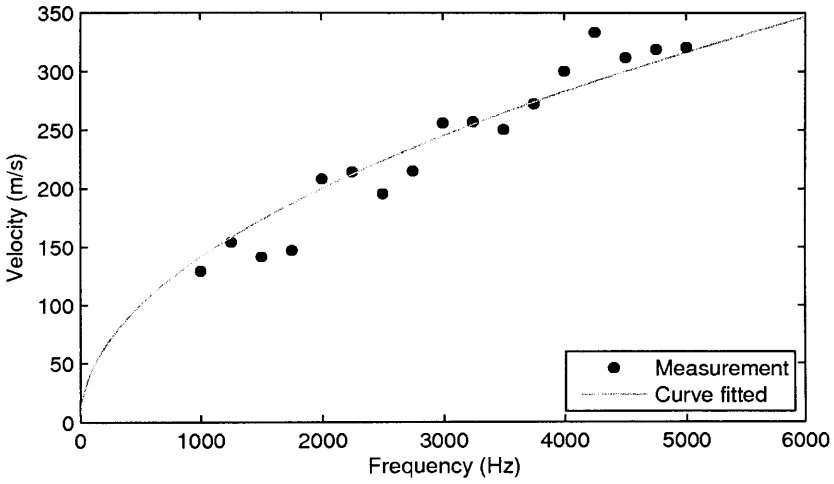


Figure 5.16: Phase velocities measured on MDF plate.

shifts of the corresponding frequencies, as:

$$H(\omega) = e^{-i\omega\tau_\omega} \tag{5.17}$$

where τ_ω represents the time shift at the velocity of frequency ω. Therefore, $S(\omega)$ can be reconstructed by multiplying $1/H(\omega)$ and $X(\omega)$.

The received signals are first band-pass filtered, such that only the useful frequencies are kept. High frequencies are filtered according to the sensor characteristics, which includes the resonant frequencies at about $11 - 13kHz$.

Figure 5.17 shows the result of the back propagation from the receivers to the source position at $(500, 100)$. It can be seen that the signals are dispersed and attenuated significantly at different sensors. To compare the back propagated signals, normalisation was first carried out. It can be seen that the signals were matched very accurately in shape (figure 5.18). From experience, it was also observed that, for positions in the central area, the received signals at different sensors do not vary as significantly as this, due to the shorter distance differences. Therefore, this example with the source position at $(500, 100)$ indicates that the measurement of the phase velocities is very satisfactory.

5.5 Continuous Wavelet Transform method

To overcome the problem of uncertain timing, Wavelet methods have been investigated to exploit their ability to obtain the time-frequency representation of a signal simultaneously. It would be beneficial to know the time intervals between different spectral components travelling in a solid object for the correct and precise measurement of the time differences of acoustic signals.

5.5.1 Continuous Wavelet transform (CWT)

The Continuous Wavelet Transform (CWT) of signal $f(t)$ is defined as

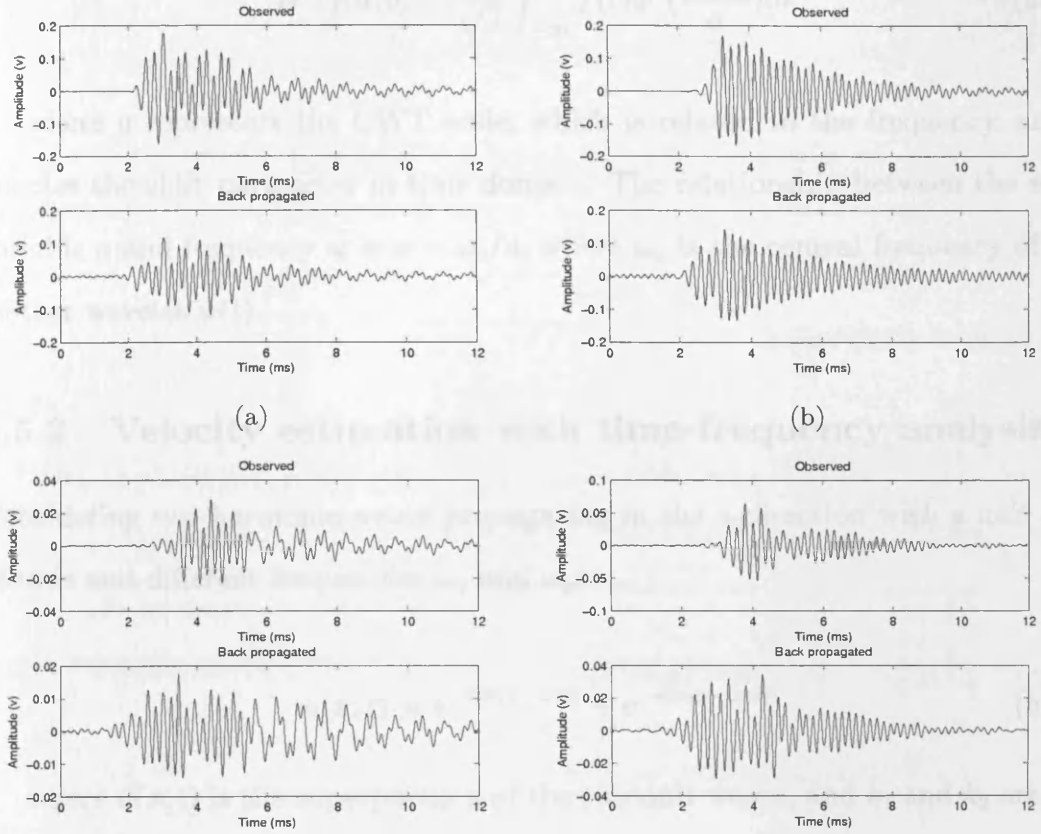


Figure 5.17: The received signals at four channels and their reconstructed signals with the measured phase velocities.

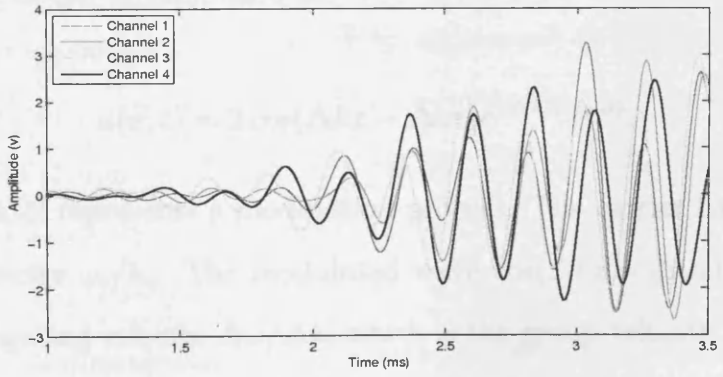


Figure 5.18: The back propagated signals from four channels after normalisation.

$$WT_f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} f(t)\psi^*(\frac{t-b}{a})dt \tag{5.18}$$

where a represents the CWT scale, which is related to the frequency, and b denotes the shift parameter in time domain. The relationship between the scale variable a and frequency ω is $\omega = \omega_0/a$, where ω_0 is the central frequency of the mother wavelet $\psi(t)$.

5.5.2 Velocity estimation with time-frequency analysis

Considering two harmonic waves propagating in the x-direction with a unit amplitude and different frequencies ω_1 and ω_2:

$$u(x,t) = e^{-i(k_1 x - \omega_1 t)} + e^{-i(k_2 x - \omega_2 t)} \tag{5.19}$$

where $u(x,t)$ is the superposition of the two unit waves, and k_1 and k_2 are the wave numbers.

Introducing

$$(k_1 + k_2)/2 = k_c, \quad (k_1 - k_2)/2 = \Delta k, \tag{5.20}$$

$$(\omega_1 + \omega_2)/2 = \omega_c, \quad (\omega_1 - \omega_2)/2 = \Delta\omega \tag{5.21}$$

equation 5.19 can be simplified as

$$u(x,t) = 2\cos(\Delta k x - \Delta\omega t)e^{-i(k_c x - \omega_c t)}. \tag{5.22}$$

Equation 5.22 represents a modulation process. The carrier has frequency ω_c and phase velocity ω_c/k_c. The modulated wave $\cos(\Delta k x - \Delta\omega t)$ has frequency $\Delta\omega$ and propagating velocity $\Delta\omega/\Delta k$, which is the group velocity when $\Delta k \to 0$.

With CWT, the signal $u(t)$ is transformed into the following form [JJ00]:

$$WT(x,a,b) = \sqrt{a}\{e^{-i(k_1 x - \omega_1 b)}\hat{\psi}^*(a\omega_1) + e^{-i(k_2 x - \omega_2 b)}\hat{\psi}^*(a\omega_2)\}. \quad (5.23)$$

The magnitude of the CWT can be obtained as

$$\mid WT(x,a,b) \mid = \sqrt{a}\{\hat{\psi}(a\omega_1)^2 + \hat{\psi}(a\omega_2)^2 + 2\hat{\psi}(a\omega_1)\hat{\psi}(a\omega_2)\cos(2\Delta k x - 2\Delta\omega b)\}^{1/2}. \quad (5.24)$$

The value of $\Delta\omega$ is assumed to be small. Therefore, it can be neglected after CWT giving $\hat{\psi}(a\omega_1) \approx \hat{\psi}(a\omega_2) \approx \hat{\psi}(a\omega_c)$. After simplification, the magnitude of CWT can be written as:

$$\mid WT_u(x,a,b) \mid = \sqrt{2a} \mid \hat{\psi}(a\omega_c) \mid [1 + \cos(2\Delta k x - 2\Delta\omega b)]^{1/2}. \quad (5.25)$$

The right side of equation 5.25 reaches its maximum value at $a = \omega_0/\omega_c$ and $b = (\Delta k/\Delta\omega)x = x/c_g$. So the maximum of CWT of scale a with time shift b indicates the arrival time at the group velocity.

Research work has been performed using the maximum in the CWT plane at a scale (frequency) with a time shift to indicate the arrival time at the corresponding group velocity [GH97] [JJ00] [HOG02]. However, the authors of [WTAT05] used the first distinct maximum in the CWT domain as the arrival time.

In fact, although promising results have been obtained by them, the maximum in the wavelet domain is not the true arrival time. The right side of the formula can be understood as a modulation by a very low frequency $\Delta\omega$ envelope. For very low frequency signals in this research, this introduces a very high bias in obtaining consistent arrival times. A more precise arrival time, therefore, is the maximum of the envelope, which can be extracted with the Hilbert transform. Moreover, instead of using a single frequency in the calculation, a combination of

different scales can be used together in order to increase reliability.

5.5.3 Practical consideration of the Wavelet selection

The CWT can be considered as a group of narrow-band filters with the central frequencies corresponding to the selected scales in the computation. It is usually best to select the optimum mother wavelet depending on the application. Different mother wavelets have been used in previous work, including the Gabor [JJ00] [WTAT05] and Morlet wavelets [OH94]. The Gabor wavelet has often been used because it provides the best resolution in both frequency and time domains. The Gabor wavelet is defined as

$$\psi_g(t) = \frac{1}{\sqrt[4]{\pi}}\sqrt{\frac{\omega_0}{\gamma}}\exp\left[-\frac{(\omega_0/\gamma)^2}{2}t^2\right]e^{j\omega_0 t} \tag{5.26}$$

and its Fourier transform is expressed as:

$$\hat{\psi}_g(\omega) = \frac{\sqrt{2\pi}}{\sqrt[4]{\pi}}\sqrt{\frac{\gamma}{\omega_0}}\exp\left[-\frac{(\gamma/\omega_0)^2}{2}(\omega-\omega_0)^2\right] \tag{5.27}$$

where γ is a positive constant that affects the shape of the wavelet and is usually chosen as $\pi\sqrt{2/\ln 2} \approx 5.3364$. The centre frequency ω_0 can be set to 2π. This function can be considered as a complex sinusoidal function with a Gaussian window centred at $t = 0$ and its Fourier transform centred at $\omega = \omega/a$. Its derivative $\psi_g(a, b)$ is then translated to the centre position $t = b$ and its Fourier transform is dilated with the scale a centred at $\omega = \omega_0/a$.

In general, the most popular practice for mother wavelet selection is based on the coefficient values for different mother wavelets [MZK02]. The highest coefficient value in the time-frequency domain indicates the best pattern matching between the signal and candidate mother wavelet. In this research, some common wavelets have been tested, and found to perform similarly. In addition to Gabor wavelet function mentioned above, *db6 (Daubechies)* has also been used.

5.6 Maximum likelihood estimation for localisation

Preliminary experiments have shown the unreliability of the proposed method by using the Wavelet transform for the TDOA calculation, due to some large errors in determining the arrival time. Maximum Likelihood Estimation (MLE) can overcome problems of ambiguities in time difference estimation, which may happen with techniques such as cross-correlation or CWT peak detection as described above. This method searches for a target position over all possible positions in the workspace instead of just looking for the precise time difference of the arrived signals. The highest likelihood determines the target presence.

5.6.1 Theory description

The likelihood distribution can be described with Bayes' rule that the posterior probability of a source at location L is expressed as [BG02]:

$$P(L, s|x_1, x_2, x_3, x_4) = \frac{P(x_1, x_2, x_3, x_4|L, s)P(L, s)}{P(x_1, x_2, x_3, x_4)} \tag{5.28}$$

In the simplified condition, the signal model at the ith sensor is assumed as:

$$x_i(t) = h_i(t) * s(t - \tau_i) + n_i(t) \tag{5.29}$$

where $*$ denotes the temporal convolution between the impulse response $h_i(t)$ and the source signals $s(t - \tau_i)$ at the ith sensor. $n_i(t)$ and τ_i represent the external noise and the time delay. Therefore, the probability can be expressed with a Gaussian distribution as below:

$$P(x_1, x_2, x_3, x_4 \mid L, s) = \prod_{i=1}^{4} P(x_i \mid L, s) \tag{5.30}$$

$$= \prod_{i=1}^{4} \exp(-\int_{t_0-W}^{t_0+W} \frac{[x_i(t+\tau_i) - s(t)]^2}{2\sigma_i^2} dt). \tag{5.31}$$

With the assumption of non-dispersive signals, the estimated $\hat{s}(t)$ can be obtained using a maximum likelihood estimator as below:

$$\hat{s}(t) = \frac{1}{4} \sum_{i=1}^{4} x_i(t+\tau_i). \tag{5.32}$$

The likelihood is thus yielded [BG02] as:

$$\log P' = \frac{2}{N} V_c - \frac{N-1}{N} V_E \tag{5.33}$$

where $N = 4$ in this work, and

$$V_c = \sum_{i=1}^{3} \sum_{j=i+1}^{4} \int_{t_0-W}^{t_0+W} x_i(t+\tau_i) x_j(t+\tau_j) dt \tag{5.34}$$

is the cross-correlation of all possible pairs of signals with the calculated time delay τ, and

$$V_E = \sum_{i=1}^{4} \int_{t_0-W}^{t_0+W} x_i^2(t+\tau_i) dt \tag{5.35}$$

is a constant value representing the summation of signal energy over all the channels and does not affect the final result.

It should be noted that the MLE method used here is based on the assumption of non-dispersive signals. As discussed above, CWT can provide the time-frequency representation of a signal. At individual scales or frequencies, CWT can be understood as band-pass filters that extract narrowband information from a signal. Due to the frequency-velocity dependency of flexural waves, CWT is effective in dealing with different frequencies separately. Although the arrival

time at a particular group velocity can be extracted at the corresponding scale, it was found to be unreliable because the assumption of two unit harmonic waves does not hold. Nevertheless, any individual scales of CWT can be regarded as narrowband signals with a central frequency ω_0. One main feature of the cross-correlation of narrowband signals is the presence of distinct local peaks with time intervals equal to the period of the central frequency. One of the local peaks can represent the group delay at a specific frequency band. Figure 5.19 illustrates a typical example of CWT of the vibration signals from two channels. The correlation at one scale is shown in the lower part of figure 5.19. When CWT is applied, the calculation of MLE will be carried out for all selected scales (frequencies) in the CWT domain separately. Taking into account the requirement of real-time computation, only 7 scales are used in this work, according to the level of the coefficients in the time-frequency plane.

5.6.2 Implementation

The implementation of MLE consists of a few stages, as described below:

Determine the map of time differences

Knowing the wave velocity c measured during the initialisation stage for various kinds of taps, it is easy to determine the time difference of arrival τ_{ij} expected for each pair of sensors, for every position on the test plate: $\tau_{ij} = \Delta d_{ij}/c$.

Figure 5.20 shows the map of time difference between each pair of sensors on the test plate, where the measured velocity is at $600m/s$.

Fill the cross-correlation coefficients into the map

The next stage consists of filling in these time difference maps with the correlation coefficients of signals from the sensors concerned at the corresponding time difference. As a result, many maps can be obtained, each one corresponding to the estimation of the target position by a couple of sensors, as shown in figure

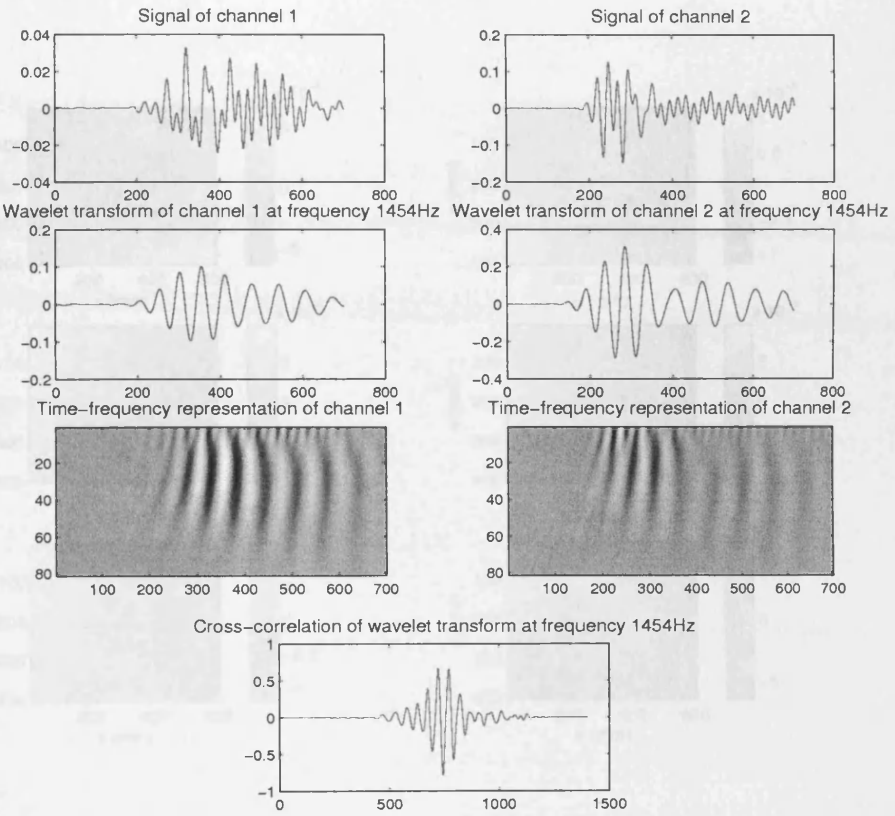


Figure 5.19: Wavelet transform of typical signals from two channels. The graph at the bottom is the cross-correlation of the signals CWT at frequency of 1454Hz.

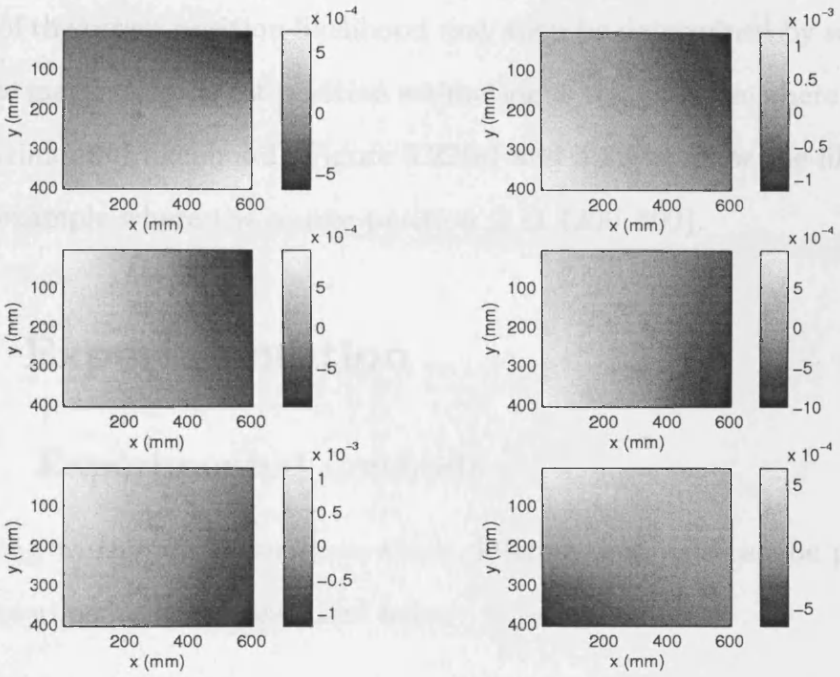


Figure 5.20: Time difference of arrival to each pair of sensors on the test plate, for a wave velocity of 600m/s.

5.21. The number m of maps depends of the number n of sensors used, and is equal to the number of combinations of two different sensors among all the sensors:

$$m = \frac{n!}{2 \times (n-2)!}$$

Calculate the maximum likelihood

To calculate the likelihood for each position in the map, as shown in equation 5.33, a map of the target position likelihood may then be determined by superimposing all these maps. The target position estimation is the position where this map has the maximum of likelihood. Figure 5.22(a) and 5.22(b) show the likelihood map for an example where the source position is at $(200, 300)$.

5.7 Experimentation

5.7.1 Experimented methods

According to the above analysis, three different methods can be proposed and used experimentally, as described below:

MLE with cross-correlation

This method (defined in equation 5.33) is suitable for non-dispersive waves. In practice, a digital high-pass filter is applied before the cross-correlation.

MLE with cross-correlation in CWT domain

This is similar to the above approach, but the cross-correlation function is computed in the CWT domain on a number of specific scales. As described above, narrowband signals tend to have cross-correlation coefficients with the same frequency. Thus one of the peaks at any individual scale should represent the arrival time at the corresponding group velocity.

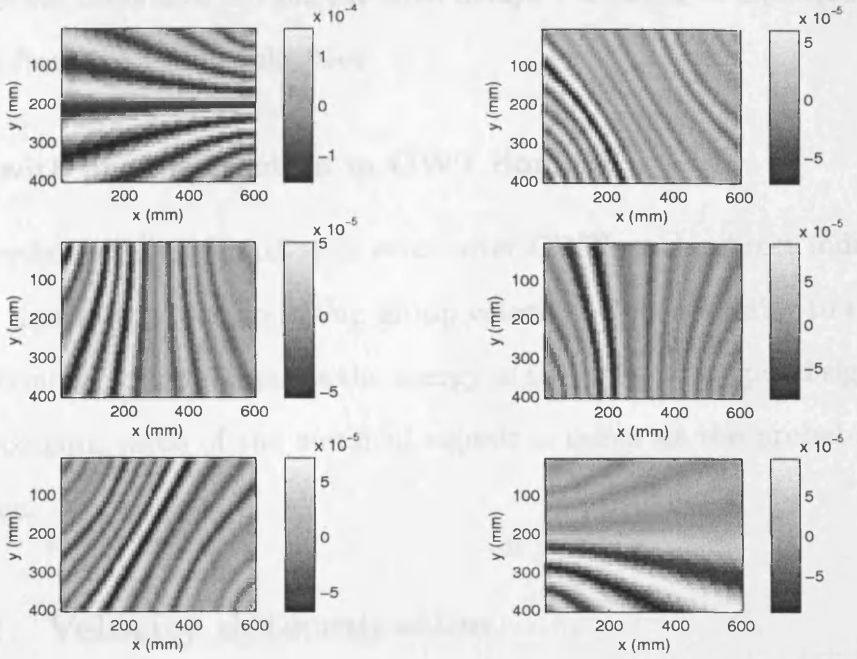
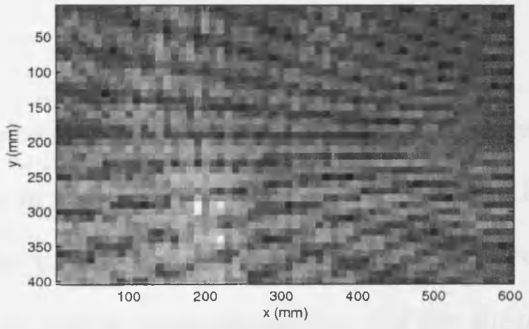
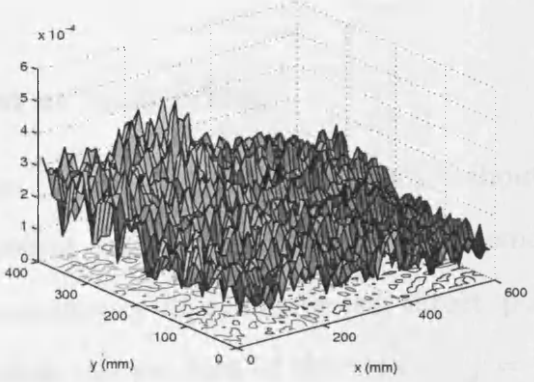


Figure 5.21: Maps of target position estimation for every couple of sensors (for a tap at position $(200, 300)$).



(a)



(b)

Figure 5.22: Target location estimation (for a tap position $(200, 300)$).

Another difference is that the time delays τ defined in equation 5.32 are extracted from the group velocities.

MLE with peak extraction in CWT domain

This method uses the maximum value after CWT as the direct indication of the arrival time at the corresponding group velocity. This is similar to the method of beamforming, which computes the energy of the back-propagated signals; instead, the maximum value of the summed signals is taken as the probability of target presence.

5.7.2 Velocity determination

Knowing the measured phase velocities, it is then possible to compute the group velocities with the following relationship:

$$c_g = \frac{\partial \omega}{\partial k} \tag{5.36}$$

where ω represents the angular velocity and k represents the wave number. Figure 5.23 shows the computed group velocity with the previously measured phase velocity. Due to the inconsistency of signals produced every time, the group velocity is obtained by this calculation, rather than by direct measurement using the CWT method mentioned above.

5.7.3 False alarm removing

In noisy environments (e.g. doors slamming, people shouting), the system may detect unwanted ambient signals. It is possible to remove these false alarms by looking at the *consistency* of the detected target position. The position-consistency is defined as the product of the ratios $r_{i,j} = \frac{\max(C_{ij})}{C_{ij}(\tau_{ij})}$ for the pair of sensors (i, j), where $\max(C_{ij})$ is the maximum of the cross-correlation function C_{ij} and $C_{ij}(\tau_{ij})$ is the value of this function for the time lag τ_{ij} corresponding to

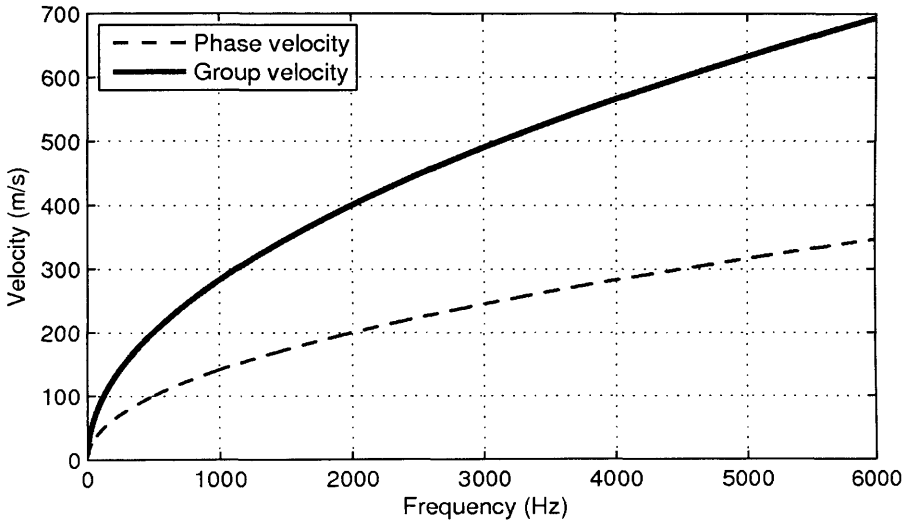


Figure 5.23: Computed group velocity with the measured phase velocity.

the estimated target position (x, y). Therefore, position-consistency $= \prod_{i,j} r_{i,j}$.

A large value of this function indicates a high probability that a received signal is a false alarm. Thus, a simple cut-off threshold can be applied to classify the consistency to identify a real tap.

5.7.4 Comparative results

The experiments were carried out with the source positioned at $(40, 30)$. Figures 5.24, 5.25, and 5.26 show typical results for the three methods. A brighter region indicates a higher probability of the target presence. The experiments have shown that the first method using cross-correlation (figure 5.24.) and the second method using cross-correlation in CWT (figure 5.25) have similarly good performance. However, the third method produced a more blurred image (Figure 5.26). The first method can be further improved with a high-pass filter. Cross-correlation in the CWT domain was found sensitive to specific frequencies when different types of object were used to produce impacts. It is also a method of high computational demand. However, the main advantage of the second and the third methods is their ability to separate frequencies of dispersive waves with CWT for a precise localisation. Compared with the third method, the second method of MLE with cross-correlation in the CWT domain has shown greater reliability.

In order to evaluate the performance quantitatively, similar experiments were carried out as for the ToF and CC methods above. The experimental results, using the same data at the three positions $(400, 300)$, $(200, 100)$, and $(300, 200)$, are depicted in figures 5.27, 5.28 and 5.29 respectively. As before, statistical analyses of the results are shown in tables 5.4, 5.5, and 5.6.

It can be seen that MLE with cross correlation in the CWT domain outperformed the other two methods in both accuracy and reliability. Directly applying the MLE method has shown a higher reliability for the impacts using nails rather than fingertips by observing the standard deviations. This is because of the strong ambiguities of local maximums in the likelihood map caused by the inconsistency

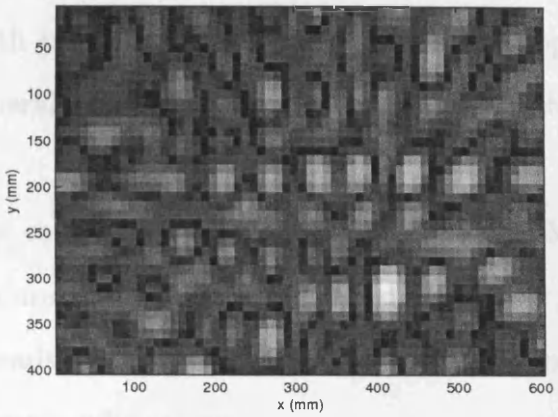


Figure 5.24: Example of MLE with cross-correlation. The source position is at (400,300).

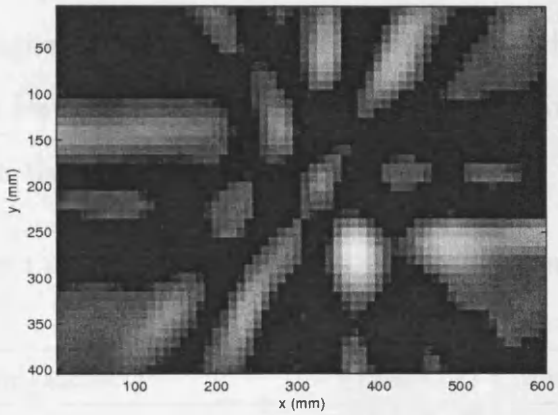


Figure 5.25: Example of MLE with cross-correlation in CWT. The source position is at (400,300).

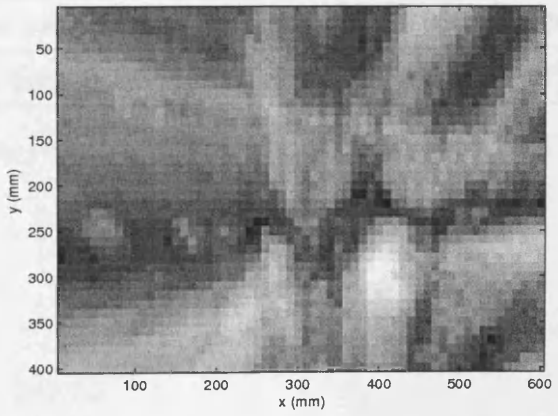


Figure 5.26: Example of MLE with peak extraction in CWT. The source position is at (400,300).

of taps made by soft fingertips. A very similar performance is observed for the method of MLE with peak extraction of the CWT envelopes, that the impacts induced by nails generally have a higher accuracy and reliability than those using fingertips.

Compared to the conventional ToF and CC methods, MLE in the CWT domain is not more accurate than the ToF method. Moreover, as mentioned before, it has the further disadvantage of higher computational demand. However, it has the following three main advantages:

- Much higher accuracy than the CC method, which directly uses cross correlation for the time lag determination.
- Suitable for signals without distinct wave fronts, which the ToF method requires. This feature is particularly useful for tracking continuous signals, as described in the next chapter.
- Provides a more robust false alarm removing criterion.

Method (impact)	Mean (x)	Mean (y)	Std (x)	Std (y)
MLE in CWT (nail)	380.000	320.400	0.000	2.000
MLE in CWT (finger)	372.000	281.600	4.0829	3.7417
MLE (nail)	387.200	302.800	17.205	14.866
MLE (finger)	405.200	349.200	66.340	40.303
MLE with envelope peak extraction (nail)	386.000	326.000	28.137	9.1285
MLE with envelope peak extraction (finger)	355.600	317.600	142.420	68.634

Table 5.4: Statistical result of the methods using the CWT and MLE at position $(400, 300)$.

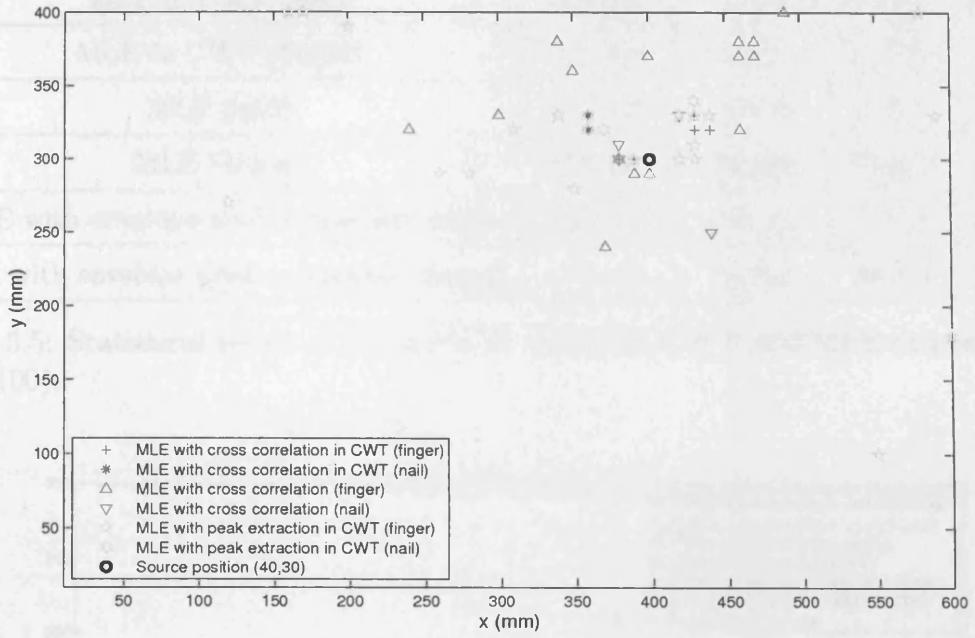


Figure 5.27: Result of the methods using the CWT and MLE at position $(400, 300)$.

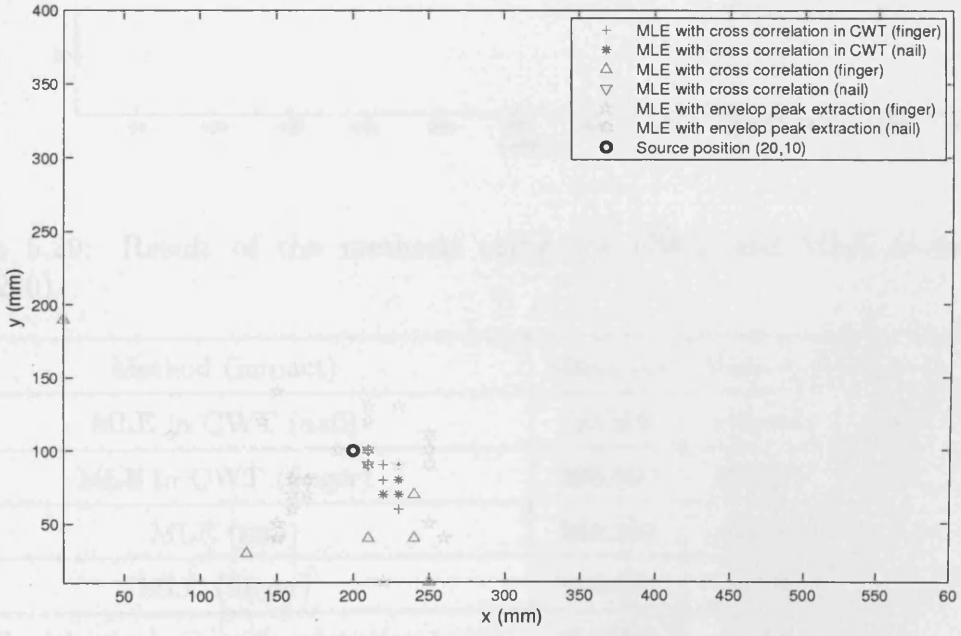


Figure 5.28: Result of the methods using the CWT and MLE at position $(200, 100)$.

Method (impact)	Mean (x)	Mean (y)	Std (x)	Std (y)
MLE in CWT (nail)	229.200	71.600	2.7695	3.7417
MLE in CWT (finger)	219.200	82.000	8.1240	11.180
MLE (nail)	210.000	95.200	0.000	5.099
MLE (finger)	218.400	58.400	49.890	37.157
MLE with envelope peak extraction (nail)	250.000	100.000	0.000	6.4552
MLE with envelope peak extraction (finger)	177.600	84.800	39.611	35.838

Table 5.5: Statistical result of the methods using the CWT and MLE at position (200, 100).

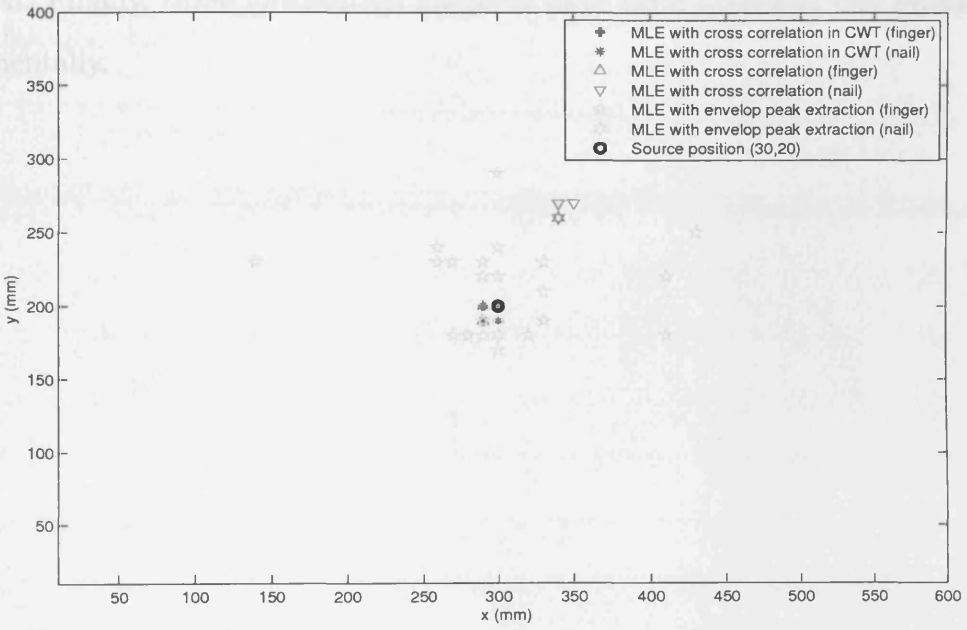


Figure 5.29: Result of the methods using the CWT and MLE at position (300, 200).

Method (impact)	Mean (x)	Mean (y)	Std (x)	Std (y)
MLE in CWT (nail)	293.600	198.400	4.899	3.7417
MLE in CWT (finger)	296.000	197.200	5.000	4.5826
MLE (nail)	338.000	256.000	11.903	20.207
MLE (finger)	300.400	199.600	8.8882	13.379
MLE with envelope peak extraction (nail)	300.000	187.200	0.000	4.5826
MLE with envelope peak extraction (finger)	300.400	216.400	58.201	28.562

Table 5.6: Statistical result of the methods using the CWT and MLE at position (300, 200).

5.8 Summary

This chapter has discussed two conventional TDOA-based methods for the localisation of an impulsive source. By studying the dispersion properties of solid materials, a technique of measuring phase velocities and verifying this measurement, has been proposed. Taking into consideration the frequency-velocity dispersive property of Lamb waves, CWT has been applied to deal with narrower band signals in different frequency bands. Maximum likelihood estimation was used to overcome the problems posed by the ambiguity of precise time lag determination. Finally, three localisation methods have been proposed and evaluated experimentally.

Chapter 6

Continuous tracking with the Kalman filter

In the general operations of a computer Graphical User Interface (GUI), drag-and-drop is an everyday action of clicking on a virtual object and dragging it to a different location or onto another virtual object. In general, it can be used to invoke many kinds of action, or create various types of associations between two abstract objects. In order to fulfil this drag-and-drop function, the capability to continuously track a finger's motion is required. In this chapter, continuous tracking based on a passive acoustic localisation method will be introduced with experimental results.

First, a brief review of existing technologies for continuously tracking an object will be given, with a focus on the Kalman filter. Second, several techniques to improve the standard Kalman filter will be reviewed. In the following section, implementation of the Kalman filter for this application will be introduced. A simple improvement using the fading memory filter is implemented with both experimental and synthetic data to quantify the errors. Finally, the complete tracking system design will be presented.

6.1 The Kalman filter and other techniques in a tracking system

6.1.1 Prediction in tracking systems

As presented in chapter 2, the scratching of a finger on a surface can generate micro-impacts and hence generate very complicated signals. With the MLE-based method introduced in the last chapter, it was found to be possible to measure the locations of continuous sources with reasonable errors.

The discretisation of the continuous signals will produce losses of information during the process of tracking. For each step of measurement, the computation takes a short time to acquire the signals and to compute their localisation. The former is a fixed length of time. The latter may differ for various computers or other kinds of processors, such as Digital Signal Processors (DSP). Faster computers can minimise the time break between two scans.

Several approaches have been applied for tracking with discrete-time sensor scans, including techniques using radar (aerospace), vision, and sonar. Tracking with a passive in-air acoustic localisation method has also been discussed by a small number of researchers, for both on-line (real-time) tracking and off-line tracking [KGM06] [Leh04]. However, no previous research has been reported on tracking acoustic sources on solid objects for any purpose, except for the little information found for tracking active sources in 1-D with preliminary experimental results [GKM98].

Tracking an object, which is moving through the view of the sensor's detection, is a common application of sequential estimation of its state space. The concept of state space is fundamental to this problem. Figure 6.1 shows a linear dynamical system, which is operating in the state domain. The state is normally represented as a vector, which is a minimum but sufficient data set, which can describe the system's current behaviour. In this diagram, the state vector is represented as $\mathbf{x}_k$, where k is the time in the discrete form. Past data is typically needed to predict

the future state of the process. It should be noted that the process function (prediction) $F(\cdot)$ and the measurement function $H(\cdot)$ are required prior to the operation. In addition, the initial state $\mathbf{x}_0$ is required and noise properties should be well defined.

Conventional approaches include the Wiener filter, $g-h$ or $g-h-k$ filter, Kalman filter (including Extended Kalman filter or others) and Particle filter. The Wiener filter was specifically designed for stationary processes, and, hence, limited by this assumed condition. The $g-h$ or $g-h-k$ filters are simple filters designed to deal with the process that can be formulated into a linear equation. A $g-h$ filter is used for a process model with a constant velocity, and a $g-h-k$ filter is used for a model with a constant acceleration. The constant $g-h$ or $g-h-k$ filter meets the accuracy requirements and at the same time requires very little computer time. However, it lacks the capability to compute the accuracy of the predictions. Another problem with using the constant $g-h$ filter is that it does not provide good performance when the target has a non-linear trajectory.

The Kalman filter is an optimum solution to the problem of discrete data linear filtering. It outperforms the approaches introduced above. The Extended Kalman filter (EKF) is essentially the same as the Kalman filter, except the state transition and observation models need not be linear functions of the state, but may instead be differentiable functions. The Particle Filter (PF) is a sophisticated model estimation technique for nonlinear state model prediction. The basic idea is that any probability density function (pdf) can be represented as a set of samples or particles. An advantage, it has, over the EKF is that the Bayesian optimal estimate can be achieved with sufficient samples. However, it would be more time consuming than the EKF.

In this work, the Kalman filter can be directly applied, as the state model is a simple linear model where the coordinates of the source can be directly measured.

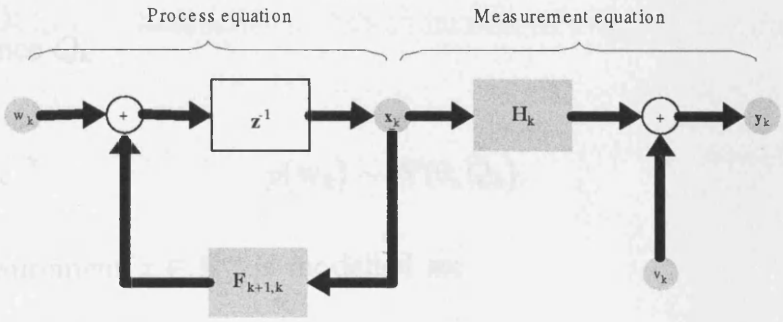


Figure 6.1: Workflow of a dynamic system [Hay01].

6.1.2 The Kalman filter

The Kalman filter is an optimal approach that can be used for both stationary and non-stationary processes [WB04]. Given a sequence of noisy observations, the Kalman filter is used to estimate the state $\mathbf{x} \in \Re^n$ of a discrete-time controlled process that is governed by the following linear stochastic difference equation

$$\mathbf{x}_k = F_k \mathbf{x}_{k-1} + B\mathbf{u}_{k-1} + \mathbf{w}_{k-1}, \tag{6.1}$$

where F_k is the state transition model which is applied to the previous state $\mathbf{x}_{k-1}$, B_k is the control-input model which is applied to the control vector u_{k-1}, w_{k-1} is the process noise which is assumed to be zero mean normal distribution with covariance Q_k

$$p(\mathbf{w}_k) \sim N(0, Q_k). \tag{6.2}$$

The measurement $z \in \Re^m$ is modelled as:

$$z_k = H_k \mathbf{x}_k + \mathbf{v}_k. \tag{6.3}$$

where H_k is the measurement function and v_k is the noise of the observation (measurement) with the same assumption of zero mean normal distribution with covariance R_k:

$$p(\mathbf{v}_k) \sim N(0, R_k). \tag{6.4}$$

In general, the noise covariances Q_k and R_k are assumed to be constant. The subscript k is used to indicate that the noise covariances may change with each step of measurement.

With the given knowledge of the process prior to step k, the a priori state

estimate $\hat{x}_k^- \in \Re^n$ can be defined. Similarly, $\hat{x}_k \in \Re^n$ is defined as the a posteriori state estimate at step k with the measurement z_k. The goal of the Kalman filter is to compute the a posteriori state estimate by a linear combination of the a priori estimate $\hat{x}_k^-$ and a weighted difference between an actual measurement z_k and a measurement prediction $H\hat{x}_k^-$ as shown in equation 6.5:

$$\hat{x}_k = \hat{x}_k^- + K(z_k - H\hat{x}_k^-) \tag{6.5}$$

where the difference $(z_k - H\hat{x}_k^-)$ is called the *measurement innovation*, or the *residual.* The residual reflects the discrepancy between the predicted measurement $H\hat{x}_k^-$ and the actual measurement z_k. A residual of zero means that the two are in complete agreement.

The $n \times m$ matrix K in 6.5 is chosen to be the gain or blending factor that minimises the a posteriori error covariance. One form of the resulting K is given by:

$$\begin{aligned} K_k &= P_k^- H^T (H P_k^- H_T + R)^{-1} \\ &= \frac{P_k^- H^T}{H P_k^- H^T + R}, \end{aligned} \tag{6.6}$$

where $P_k^- = E[e_k^- e_k^{-T}]$ represents the a priori estimate error covariance, and the a posteriori estimate error covariance is represented by P_k, that $P_k = E[e_k e_k^T]$. The definition of the a priori and a posteriori estimate errors are:

$$e_k^- \equiv x_k - \hat{x}_k, and \tag{6.7}$$

$$e_k \equiv x_k - \hat{x}_k \tag{6.8}$$

The weight K is adjusted in accordance with error covariances R and P. It can

be seen that the gain weight K is higher when the measurement error covariance R approaches zero.

$$\lim_{R_k \to 0} K_k = H^{-1} \tag{6.9}$$

In contrast, the gain weight K becomes zero when the a priori estimate error covariance P_k^- approaches zero.

$$\lim_{P_k^- \to 0} K_k = 0. \tag{6.10}$$

Table 6.1 [Hay01] is a detailed summary of the Kalman filter. The properties of the Kalman filter can be found from many sources. One advantage of the Kalman filter is the efficient computation. Only the state at the previous step and the new observation is required for the new estimation.

6.1.3 Adaptive filtering approaches for non-linear trajectories

The Kalman filter is constructed based on the assumed knowledge of the state variables propagation, which are the positions and velocities in time for the tracking system in the form of a linear dynamic model. This linear dynamic model is also known as a target motion model, that can be modelled as a movement at a constant velocity or acceleration. However, due to the lack of knowledge of the dynamic motions in some applications, it may degrade the performance of the Kalman filter. In the tracking system, the random dynamic motion can be explained in some cases, such as during non-linear trajectories.

In order to deal with this problem, it is often necessary to apply adaptive techniques to characterise the target models and then predict and compensate for the uncertainties [FBQL91]. An overview of some of these techniques is given

State-space model
$\mathbf{x}_{k+1} = F_{k+1,k}\mathbf{x}_k + \mathbf{w}_k$,
$\mathbf{y}_k = H_k\mathbf{x}_k + \mathbf{v}_k$,
where $\mathbf{w}_k$ and $\mathbf{v}_k$ are independent, zero-mean, Gaussian noise processes of covariance matrices Q_k and R_k, respectively.

Initialisation: For $k = 0$, set
$\hat{\mathbf{x}}_0 = E[\mathbf{x}_0]$,
$\mathbf{P}_0 = E[(\mathbf{x}_0 - E[\mathbf{x}_0])(\mathbf{x}_0 - E[\mathbf{x}_0])^T]$.

Computation: For $k = 1, 2, \ldots$, compute:
State estimate propagation
$\hat{\mathbf{x}}_k^- = F_{k,k-1}\hat{\mathbf{x}}_{k-1}^-$,

Error covariance propagation
$\mathbf{P}_k^- = F_{k,k-1}\mathbf{P}_{k-1}F_{k,k-1}^T + Q_{k-1}$,

Kalman gain matrix
$K_k = \mathbf{P}_k^- H_k^T[H_k\mathbf{P}_k^- H_k^T + R_k]^{-1}$,

State estimate update
$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_k^- + K_k(\mathbf{z}_k - H_k\hat{\mathbf{x}}_k^-)$,

Error covariance update
$\mathbf{P}_k = (\mathbf{I} - K_kH_k)\mathbf{P}_k^-$.

Table 6.1: Summary of the Kalman filter [Hay01]

below.

Identification of the process noise variance

This is probably the most popular approach for compensating for the inadequacies in the presumed state dynamic model by means of identifying the process noise variance to determine the model state. The simplest way to achieve this is to establish a detection scheme. The process noise variance, and thus the Kalman gain, will then be changed to adapt to the new state model. The filter residual sequence is monitored and various techniques have been employed to calculate the process change. The sudden change of motion can be observed by a large value of:

$$\alpha = (\tilde{y}_k^T \tilde{y}_k)/(\frac{1}{N}\sum_{i=1}^{N} \tilde{y}_{k-i}^T \tilde{y}_{k-i}) \tag{6.11}$$

where $\tilde{y}_k$ is the measurement residual, defined as $\tilde{y}_k = z_k - H\hat{x}_k^-$.

In practice, the summation of α over a window in the time domain is applied to determine the state of the current motion. When this value exceeds a pre-defined threshold, it can then be considered as a sudden model change.

When the non-linear motion is detected, the system noise covariance Q needs to be modified to suit the new state:

$$Q_k \rightarrow \Gamma Q_k \Gamma^T = Q_k^* \tag{6.12}$$

where

$$||Q_k^*|| > ||Q_k|| \tag{6.13}$$

The multiple model and interacting multiple model approaches

As the names indicate, Multiple Model (MM) and Interacting Multiple Model (IMM) approaches use multiple state models to describe the process system, in

order to overcome the difficulty of motion uncertainty. The principle of this approach is that several pre-defined motion models are assumed to contain the true target displacement model at one time [LA93]. Each model will be weighted with its corresponding probability.

The MM and IMM methods are very similar to each other, except for the definition of the probability. The IMM uses the presumed a priori Markovian transition probability $p_{i,k}$, which represents the probability from the ith model at the scan time k. Given a parameter vector $\theta \in \theta_1, \theta_2, \ldots, \theta_n$ denoting the different models, the probability is given as

$$p_{i,k} = P(\theta_k) = P(\theta_{i,k}|Z_k) \tag{6.14}$$

where $Z_k = z_{k_0}, \ldots, z_k$. This set of parameter vectors $\theta_1, \theta_2, \ldots, \theta_n$ could correspond to different noise levels, different non-linear models, or others. In brief, both methods share the same formulation to describe the final outcome

$$\hat{x}_k = \sum_{i=1}^{n} \hat{x}_{i,k} p_{i,k} \tag{6.15}$$

where $\hat{x}_{i,k}$ represents the estimate produced by the ith sub-filter.

Figure 6.2 is a diagram of the IMM filter. The measurement input z and the previous states, after the step of interaction, are used as the inputs for the sub-filters. The final result is obtained by calculating the equation 6.15, where the required parameters are obtained from the sub-filters and the step of weighting.

Obviously, the result can be very accurate when an adequate number of sub-models is used. It can be seen that the computational demand with such an approach is very high as all the models must be calculated. In addition, the performance relies on the proper estimation of the transition probability. The result will only be optimal when one of the models corresponds to the correct displacement motion.

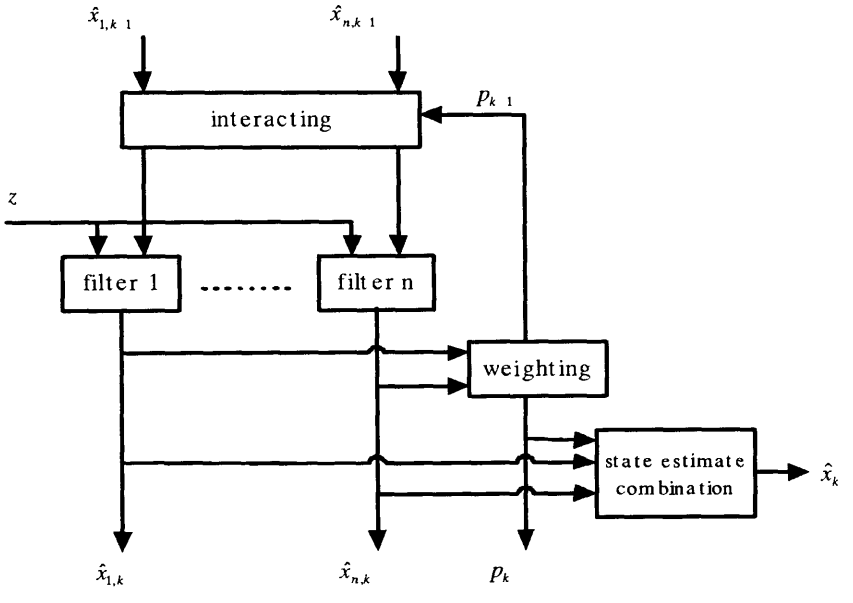


Figure 6.2: The IMM filter diagram.

Coloured noise model

The coloured noise model is an approach to characterise the motion of non-linear trajectory. The principle of it is that by checking the correlation of the noise process, it is possible to determine the onset of manoeuvre. It is assumed that in practice the action of a manoeuvre will last for a period of time, rather than be completed in a transient time. Therefore, the observed process noise will be correlated. For the system model:

$$x_{k+1} = Fx_k + Bu_k \tag{6.16}$$

the autocorrelation function of the process noise $u(k)$ is assumed to be of the form:

$$r_{uu}(\ell) = E[u_k u_{k+\ell}] = \sigma_u^2 e^{-a|\ell|} \tag{6.17}$$

where a represents a scalar constant for this model. The noise process u_k can be given as [FBQL91]:

$$u_{k+1} = -au_k + \sigma_u\sqrt{2a}w_k \tag{6.18}$$

where w_k represents a white noise process. In order to apply the Kalman filter, the system model needs to be transformed to the linear relationship. Given $\bar{x}_k = [x_k^T u_k^T]^T$, the system can be derived as:

$$\overline{x}_{k+1} = \begin{bmatrix} F & \vdots & B \\ \cdots & \cdots & \cdots \\ 0 & \vdots & -a \end{bmatrix} \overline{x}_k + \begin{bmatrix} 0 \\ \cdots \\ \sigma_u\sqrt{2a} \end{bmatrix} w_k \tag{6.19}$$

and

$$z_k = \begin{bmatrix} H & \vdots & 0 \end{bmatrix} \overline{x}_k + v_k \tag{6.20}$$

In this case, the system is constructed with the independent white noises w and v, which are required by the Kalman filter. However, the performance of this technique is closely related to the selection of the scalar a, as it determines the duration of the manoeuvring window in time. This technique will only work properly when the manoeuvre motion is known a priori.

The variable dimensional filter

Similar to other adaptive Kalman filtering techniques, when the non-linear motion is observed, the process model will be modified to adapt to the new motion. In this technique, an acceleration constant is introduced to the original constant velocity state.

$$\underline{x} = \begin{bmatrix} x \\ \dot{x} \end{bmatrix} \rightarrow \underline{x}_m = \begin{bmatrix} x \\ \dot{x} \\ \ddot{x} \end{bmatrix} \tag{6.21}$$

For such a method, it is suggested that the normalised residual sequence $\tilde{y}$ in a fading memory manner can be monitored as:

$$\varrho_k = \lambda \varrho_{k-1} + \tilde{y}_k \qquad , \lambda < 1 \tag{6.22}$$

$$= \tilde{y}_k + \lambda \tilde{y}_{k-1} + \lambda^2 \tilde{y}_{k-2} + \ldots \tag{6.23}$$

where λ is the scalar value that determines the effective window in time. When ϱ_k exceeds a pre-defined threshold, it will be considered as an event of non-linear trajectory.

The performance of this technique is closely related to the selection of the constant acceleration. On the other hand, this method is effective in correcting the past estimates and eventually has the effect of smoothing the trajectory, rather than of providing a better estimation.

Feedback of estimation error

This is a simple and novel technique for the implementation of the Kalman filter. Different from those traditional adaptive Kalman filters, the position estimation error of the last state w_k is added to the current measurement model. This is based on the assumption that the position estimation errors are the same for the two sequential measurements [PN03]. Figure 6.3 depicts the measurement model of the Kalman filter. The true position of the target at time $k-1$ is denoted by x_{k-1}.

As presented above, the last estimation error w_k is assumed to be

$$w_k = x_{k-1} - \hat{x}_{k-1} \tag{6.24}$$

where x_{k-1} is the true position of the target at time $k-1$, and $\hat{x}_{k-1}$ is the output estimation position at time $k-1$. The measurement used by the new Kalman filter will be described as:

$$z_k = x_k - w_k = x_k - (x_{k-1} - \hat{x}_{k-1}) \tag{6.25}$$

In this case, it is assumed that the state model at time $k-1$ provides the true position of target x_{k-1}. As mentioned before, the position estimation error w in closed state is assumed to be equal, $(x_{k-1} - \hat{x}_{k-1} \approx x_k - \hat{x}_k)$. Therefore, the new constructed measurement will be more correct than the original measurement.

This method has proved to be efficient in computation, and able to provide a closer performance to the IMM method in some situations. As can be seen, the estimation error is compensated for with the last measured state, which is assumed to be the true position. However, due to the independence between the last measurement and the current measurement, it is not a practically useful methodology for general purposes. This method may only be useful when the measurement error is affected by some possible physical factors that will take effect for a period, rather than by general white noise.

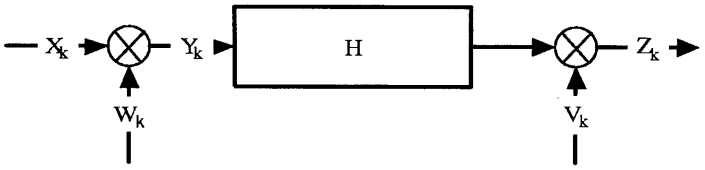


Figure 6.3: The measurement model using the feedback of the estimation error.

The fading memory filter

This is a technique that adaptively changes the Kalman gain through the use of a scale factor. The new introduced scale level factor is determined based on the available information. As the name indicates, this scale factor has the effect of fading the memory of the Kalman filter, which originally uses an unbounded memory to calculate the Kalman gain. By applying a factor to the predicted covariance matrix to increase the variance of the predicted state vector, more weight will be given to the measurements. Therefore, the scale factor level increases during non-linear periods to provide improved tracking.

The fading Kalman filter can be described in the same form as the standard Kalman filter, except the error covariance equation is inserted with a fading scale factor.

$$P_{k+1|k} = S_{k+1} F_k P_k F_k^T + Q_k \tag{6.26}$$

where S_{k+1} is the fading scale factor at time $k+1$. The fading memory effect must be satisfied with the condition $S \geq 1$. When $S = 1$, it becomes the standard Kalman filter.

This method has the significant advantage of very low computational demands. Performance is dependent on the selection of the scale factor. Compared to other adaptive Kalman filter techniques, this method achieves the same target of a proper Kalman gain, but with a comparably low computational demand. In addition, this method has been compared with the IMM method, and has been proved to give better results in general cases and a similar performance when the knowledge of the non-linear motion is known for IMM [EO99].

6.2 Implementation of the Kalman filter

6.2.1 State model for tracking in a 2-D plane

With the method described in chapter 5, the source location can be obtained as a Cartesian coordinate. It means that the Kalman filter fits the state model as a linear problem. The state model thus can be described as below:

$$\chi = \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} \tag{6.27}$$

where x and y represent the coordinate of the source position, and $\dot{x}$ and $\dot{y}$ represent the moving velocity in two directions. The system equations can be described with parameters below:

$$F = \begin{bmatrix} 1 & 0 & dt & 0 \\ 0 & 1 & 0 & dt \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{6.28}$$

where dt represents the update time, and is not a constant, and

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}. \tag{6.29}$$

6.2.2 Initialisation and reinitialisation

Before applying the Kalman filter, the first task is that of initialisation. As indicated in the equations of the Kalman filter, there are some parameters to be determined beforehand.

Initial position and reinitialisation

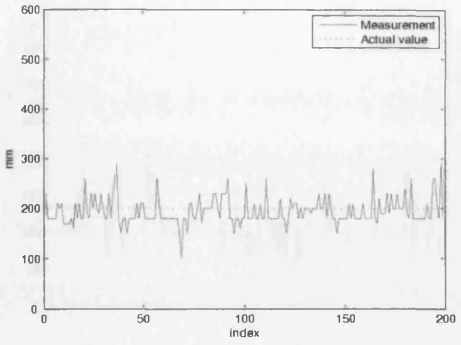
The initial position is required for the calculation of the Kalman filter. However, without any prior knowledge, it is impossible to designate a unique value for it. A compromised solution is to use the first measured position, which is not required to be very accurate, as it will converge to the real trajectory with the Kalman filter in the following scans. At present, the accuracy of the first measurement is considered acceptable.

Moreover, one cannot operate on the surface at one stroke only. A criterion needs to be defined to judge if it is a new start or the continuation of the previous measurement. Therefore, it is necessary to define the time gap, during which no signals, or no good signals, have been detected. In practice, $0.3 - 0.5\ s$ was found to work quite reliably.

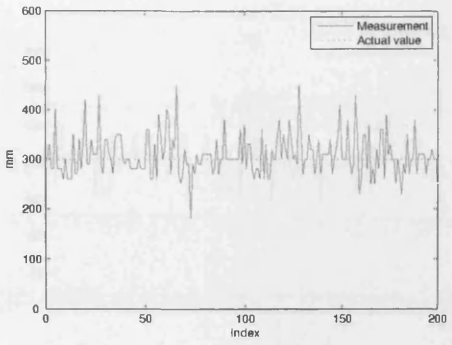
Determination of noise variance

The noise covariance for the measurement needs to be obtained and approximated prior to the operation of the filter. This can be done with some off-line sample measurements. Taking into consideration the real-time requirement, the CWT calculation at a number of scales was replaced with a band pass filter to minimise the computational demand.

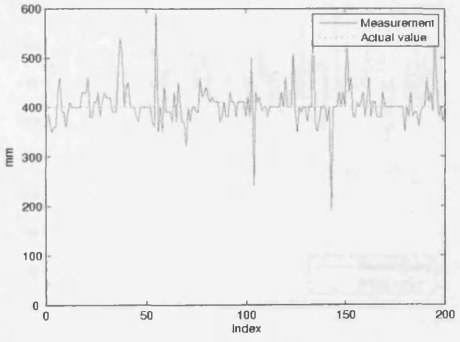
First, the measurement of localising continuous sources must be performed. Experiments were proposed to treat two directions x and y separately. The finger was moving vertically on three straight lines, where $x = 200, 300, 400(mm)$ respectively. The same experiments were performed on three horizontal straight lines, where $y = 100, 200, 300(mm)$. The results are shown in figures 6.4 and 6.5 respectively. It can be observed that the measurement for continuous tracking is less accurate than the case of single taps. This may be because of many factors, such as unwanted noises, weaker signals generated by scratching the surface, and imprecise operation by hand. However, the mean values of the measured results are still close to the actual value, as illustrated in the figures.



(a) At the line $x = 200mm$. Mean value of the measurement is $194.3mm$.

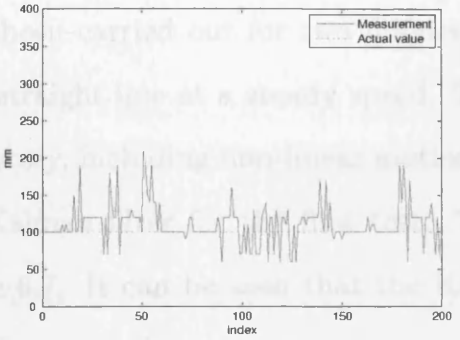


(b) At the line $x = 300mm$. Mean value of the measurement is $308.1mm$.

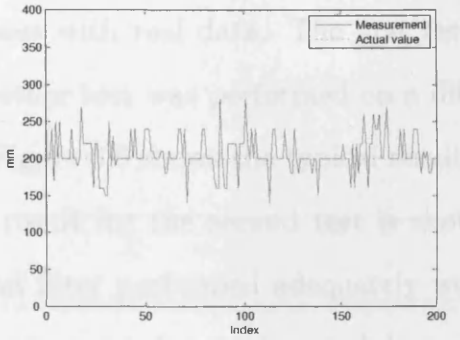


(c) At the line $x = 400mm$. Mean value of the measurement is $403.0mm$.

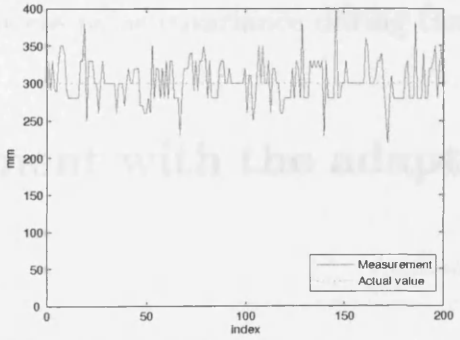
Figure 6.4: Continuous vertical moving at the line $x = 200, 300, 400mm$.



(a) At the line $y = 100mm$. Mean value of the measurement is $108.8mm$.



(b) At the line $y = 200mm$. Mean value of the measurement is $206.0mm$.



(c) At the line $y = 300mm$. Mean value of the measurement is $301.1mm$.

Figure 6.5: Continuous horizontal moving at the lines $y = 100, 200, 300mm$.

On the other hand, the noise covariance for the prediction process cannot be obtained directly in general. This may vary for different motions of finger movement, or even vary from person to person. Thus, an arbitrary value is used, according to the experimental results.

6.2.3 Result and discussion

In order to evaluate the performance of the Kalman filter, several experiments have been carried out for two different cases with real data. The first test was on a straight line at a steady speed. The other test was performed on a difficult trajectory, including non-linear motions. Figure 6.6 shows the typical result with the Kalman filter for the first test. The result for the second test is shown in figure 6.7. It can be seen that the Kalman filter performed adequately well for the simple straight trajectory. However, the second experiment did not show equivalent performance, and behaved as undershooting. This is mainly because of the unpredictable process noise covariance during the non-linear motion.

6.3 Improvement with the adaptive Kalman filter

However, it is occasionally the case that the measurement error (in particular) does not remain constant. For example, different objects or different users may make the system perform at different noise levels. External vibration noises can also affect the accuracy sometimes.

Moreover, as mentioned before, the process noise Q is sometimes changed dynamically during filter operation – becoming Q_k – in order to adjust to different dynamics. For example, the magnitude of Q_k can be reduced if the user moves steadily, and increased if the dynamics start changing rapidly. In such cases Q_k might be chosen to account for uncertainty about the user's intentions and uncertainty in the model.

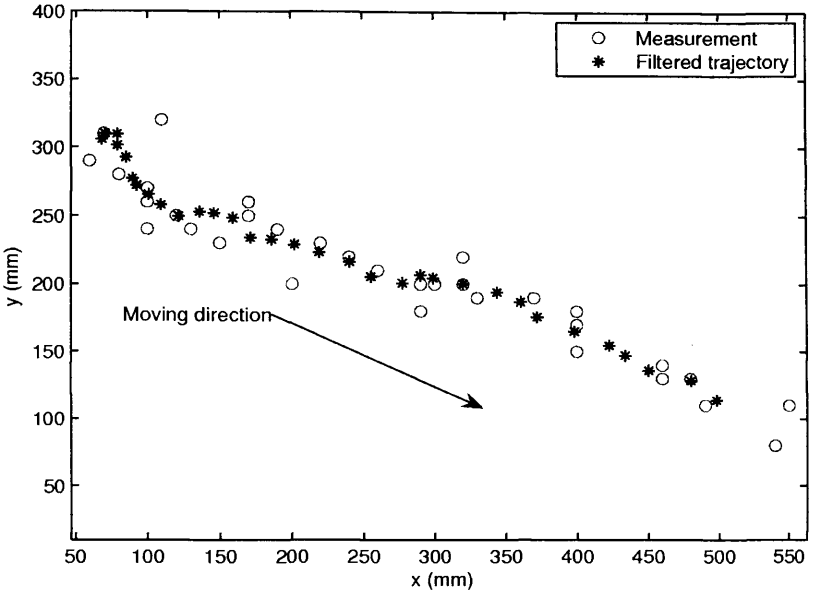


Figure 6.6: A typical result of the Kalman filter.

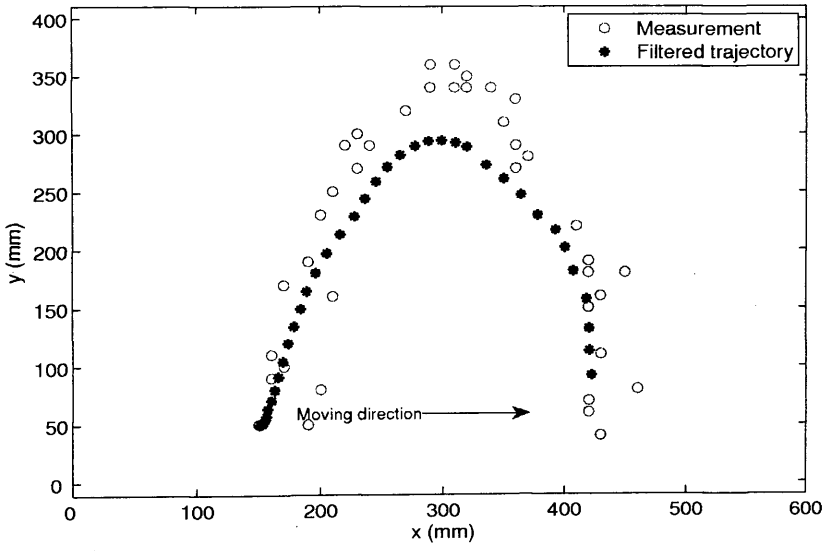


Figure 6.7: A typical result of the Kalman filter.

When tracking the motion of finger movement, changing direction or sudden accelerations/decelerations cannot be predicted by the system. Therefore, constant noise variance cannot be appropriately designed before applying to the Kalman filter. Overshooting and undershooting are common in these cases, as mentioned before. That is because the dynamic model keeps the position estimation of the previous state while the source is actually suddenly changing direction or speed.

In this section, an adaptive Kalman filter using the fading memory approach is introduced and implemented. In addition, due to the occasional high measurement error, a simple likelihood map of the state of movement is studied and applied, in order to determine if the detected target is noise, which will be effectively ignored.

6.3.1 The adaptive fading memory

In the Kalman filter, the weight of gain for the adjustment between the new measurement (observation model) and the predicted state vector is determined on the basis of the dynamic model and all previous measurements. The principle of fading memory is simply to apply a scale factor $S > 1$ to the predicted covariance matrix in order to increase the variance of the predicted state vector (equation 6.26), if the motion is being changed. More weight will be given to the measurements. Otherwise, with the standard configuration, the dynamic model will be given more weight, and will ignore the information from measurements. This will cause the divergence of the filtering process.

6.3.2 Extra state likelihood mapping

However, the approach introduced has only considered the factor of the noise variance due to the non-linear motion. Empirical data has shown that this method does not perform adequately well in some situations. For example, a big measurement error might be considered as a sudden change of direction or accel-

eration/deceleration, and the fading Kalman filter will put more weight on the measurement. Therefore, this will cause a big divergence from the true trajectory. In the next epoch, the prediction will be even farther from the measurement again. This may cause instability of the tracking system.

Big errors from the measurement are normally caused by external noise or boundary reflections, appearing as multiple local maximums in the initial likelihood map. Specifically, when the effect of the external noise is significant, a local maximum will normally appear at the centre position on the plate. These ambiguities will definitely make it difficult to judge the correct source position. In this case, the measurement, and hence the estimate, is not useful.

In order to avoid the above problems, an extra likelihood map of the moving state is introduced for the following purposes:

- The detection of the presence of a non-linear trajectory could be based either on the last residual error or some function of the last N residual errors. Once it is detected, a new configuration of the Kalman filter will be applied in order to suit the real motion;

- A large change of residual errors does not just imply a big change of the target state, but also noise, as described before. Therefore, it would be advantageous to be able to recognise which is which. When the measurement is affected by external noise too significantly, the estimate will not be useful, and hence should be ignored.

An extra likelihood map is designed to be able to cope with different possible situations: invariability (prediction is right), sudden change (changing directions or accelerating/decelerating), and noises to be ignored. Figure 6.8 illustrates a schematic of this. Three non-overlapping windows are presented for the three possible states mentioned above.

The inner window, known as the prediction window based on the previous state, is regarded as a steady state. The manoeuvring window covers the possi-

ble area that the target will move into after non-linear trajectories or acceleration/deceleration. Finally, the area outside these two windows is considered as the noise area and will not be taken into account.

The determination of the size of the first two windows is related to the scale factor for the fading memory and will be introduced first. A simple approach is to use a constant value for the scale factor that is greater than 1 during the non-linear trajecoty. Due to the variety of motions, a constant scale factor is obviously not adequate. A scale factor that is adaptively changed, according to the previous residual errors, is more suitable. As presented above, there are different adaptive approaches to improve the Kalman filter. They are mostly based on the last residual error or some functions of the last N residual errors, for example a weighted sum of the residual errors from the previous observations. Recall the residual error equation 6.5:

$$\tilde{y}_k = z_k - H\hat{x}_k^- \tag{6.30}$$

By calculating the previous N residual errors, a criterion is defined for judging the non-linear motion, as:

$$\alpha = (\tilde{y}_k^T \tilde{y}_k)/(\frac{1}{N}\sum_{i=1}^{N} \tilde{y}_{k-i}^T \tilde{y}_{k-i}) \tag{6.31}$$

The value of α indicates the relationship between the current residual error and the previous N residual errors. When $\alpha > 1$, the current error covariance is higher than the mean of the previous N residual errors. Assuming there is no acceleration or non-linear motion in the previous N states, the mean of the previous N residual errors could indicate the correct error covariance in the steady motion. Therefore, by introducing a threshold $TH1$, satisfying $TH1 \geq 1$, to α, the non-linear state can be determined. Equally, this threshold determines the inner window of the likelihood map.

Unwanted external noise is unavoidable. Therefore, it is necessary to set up

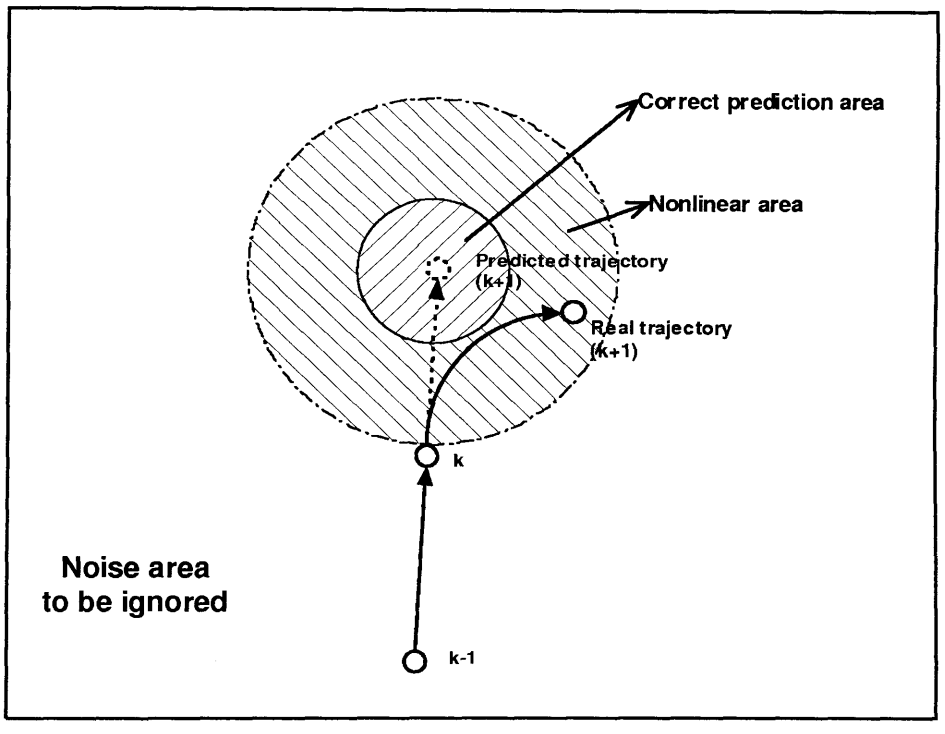


Figure 6.8: Extra likelihood map for Kalman filter prediction.

a criterion to recognise the event of noise. In addition to the false alarm removal criterion introduced in the last chapter, at this prediction stage, a more restrictive criterion is applied by introducing another threshold TH_2 as the constraint of the maximum possible acceleration. This threshold also has the same effect as the outer window in the extra likelihood map. On the other hand, this will limit the speed of finger movement and will not be suitable in some applications.

Here, the scale factor can be decided as below:

$$S = \begin{cases} \sqrt{\alpha}, & TH_2 > \alpha > TH_1 \\ 1, & \alpha \leq TH1 \\ n/a, & \alpha \geq TH2 \quad \textit{Noises. Result should be ignored} \end{cases} \tag{6.32}$$

It can be seen that $\sqrt{\alpha}$ is used as the scale factor, rather than α, to avoid high fluctuation in the filtered trajectory.

6.3.3 Result and discussion

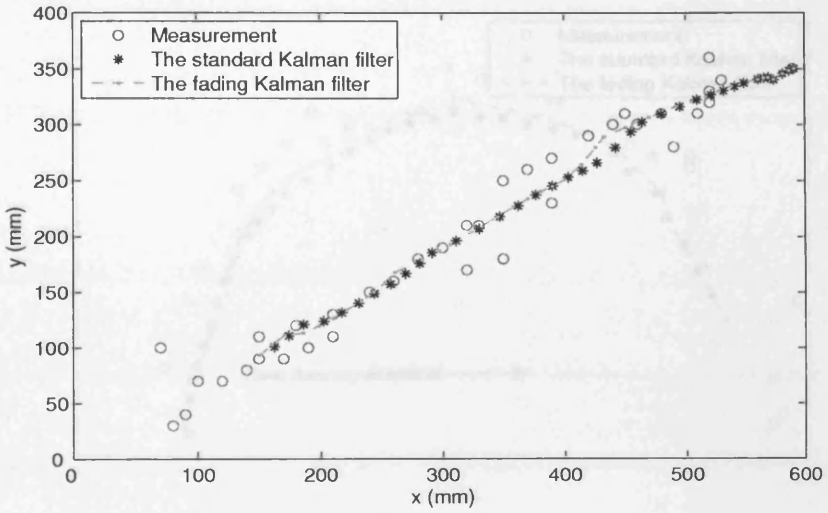
Experiments were carried out with different conditions. Figure 6.9 shows the results of two straight trajectories at relatively steady moving speeds. The blue circles are the initial measurements. The output from the standard Kalman filter is represented by the black asterisks. The adaptive fading memory technique is also applied and its result is displayed by the red dashed line. Clearly, it can be seen that the original measurement shows big variances, and the Kalman filter is effective in reducing errors. In these two cases, the fading Kalman filter did not produce a high discrepancy from the standard Kalman filter. As the fading Kalman filter was introduced to cope with non-linear trajectories and acceleration/deceleration, this result is as expected. However, one drawback of the fading Kalman filter can be observed from the figures, namely that some big errors from the measurement have been considered as sudden motion change by

the fading Kalman filter. This caused estimation closer to the measurement and occasionally that is not always desired. This indicates the inability of the fading memory filter to recognise the real motion when the measurement falls in the non-linear window, with a high measurement error. In contrast, this problem does not affect the standard Kalman filter as significantly as it does the fading Kalman filter.

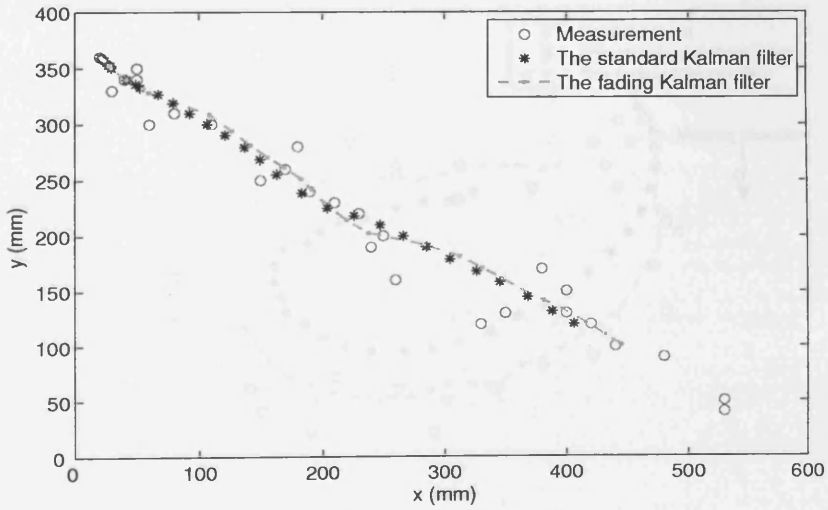
Figures 6.10(a) and 6.10(b) depict the results of two experiments on two non-linear trajectories. The first one was carried out on a relatively steady speed motion with a smaller turning rate. In contrast, the second one was carried out at a higher turning rate and the movement was deliberately accelerated. Similar to the previous experiments on the two straight lines, the first experiment performed adequately well with both filters. The fading Kalman filter was more effective in pulling the estimate closer to the measurement in the non-linear trajectory. For the second experiment, the fading Kalman filter has a more significant effect in correcting the estimate properly during the non-linear trajectory and acceleration. As mentioned before, the standard Kalman filter encountered the problem of undershooting during acceleration and non-linear trajectories. The fading Kalman filter outperformed the standard approach considerably here.

Considering the extreme case of large measurement errors, the performance of the standard Kalman filter is greatly degraded, as shown in figure 6.11. The fading Kalman filter shows its advantage in effectively ignoring the wrong measurement, and hence in providing more accurate estimates.

In order to quantify the tracking performance, synthetic data is prepared, such that the errors against the real trajectories can be measured and compared. Taking into consideration the performance of both the standard Kalman filter and the fading Kalman filter, the same model as for equation 6.1 was used to simulate different trajectories with constant speed and two accelerations on both x and y axes. The term Bu was used to designate the accelerated motion. Additional noises were added to the simulated data. These are Gaussian noises generated

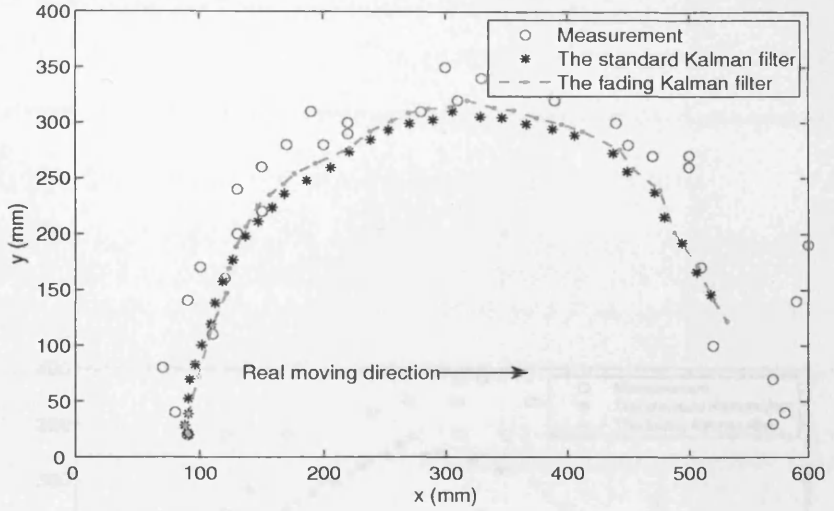


(a)

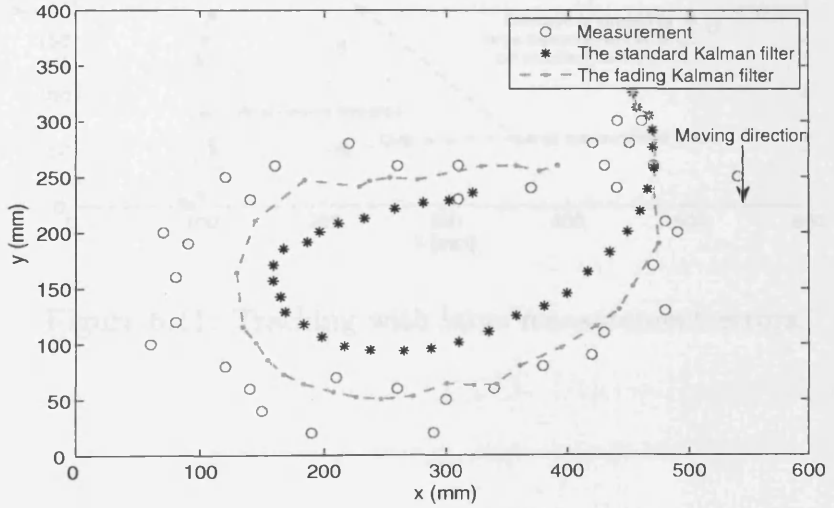


(b)

Figure 6.9: Test on two straight trajectories with Kalman filter and fading Kalman filter.



(a)



(b)

Figure 6.10: Tests on two non-linear trajectories with Kalman filter and fading Kalman filter.

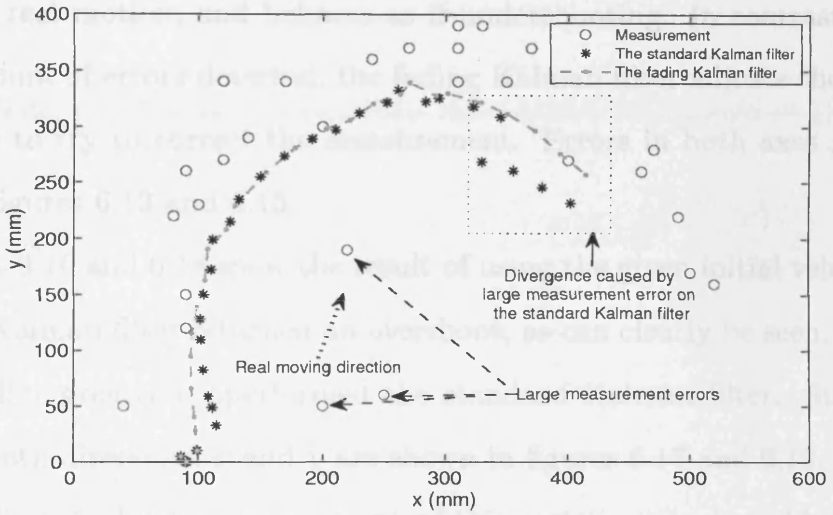


Figure 6.11: Tracking with large measurement errors.

with a given variance and mean. The scan interval was assumed to be constant at $0.05s$.

Two non-linear examples with different turning rates were prepared. Experiments were also carried out using both sets of data with two different initial velocities. One was initialised with zero speed, and the other was given the real velocities used for the modelling. The purpose of the second configuration was to simulate the tracking during the movement, assuming that the Kalman filter was able to predict the correct movement at a stable state.

Figures 6.12 and 6.14 depict the result of the first experiment with initialisation of zero speed. It can be observed that the standard Kalman filter cannot follow the real motion, and behaves as if undershooting. In contrast, according to the amount of errors detected, the fading Kalman filter adjusts the estimation adaptively to try to correct the measurement. Errors in both axes x and y are shown in figures 6.13 and 6.15.

Figures 6.16 and 6.18 show the result of using the given initial velocities. The standard Kalman filter returned an overshoot, as can clearly be seen. The fading Kalman filter greatly outperformed the standard Kalman filter. Similarly, the errors in both directions x and y are shown in figures 6.17 and 6.19.

Regarding the latency requirement of this system, it is desirable to have the lowest latency possible. The upper bound on the latency of this system is the sum of the travel time of the signal, the delay from buffering the signals, and the delay introduced by the computation. The travel time is merely the range between source and sensor divided by the speed of wave propagation. As the maximum distance in this system is the diagonal line, which is about $0.72m$, the delay produced by the travel time is, then, $1.2ms$. The buffer length used in this system was 1200 samples at the sampling rate of $100000Hz$. Therefore, this introduces a delay of $12ms$. The total execution time, including the buffering and the computation delays, is measurable and the averaged execution time was obtained as about $46ms$.

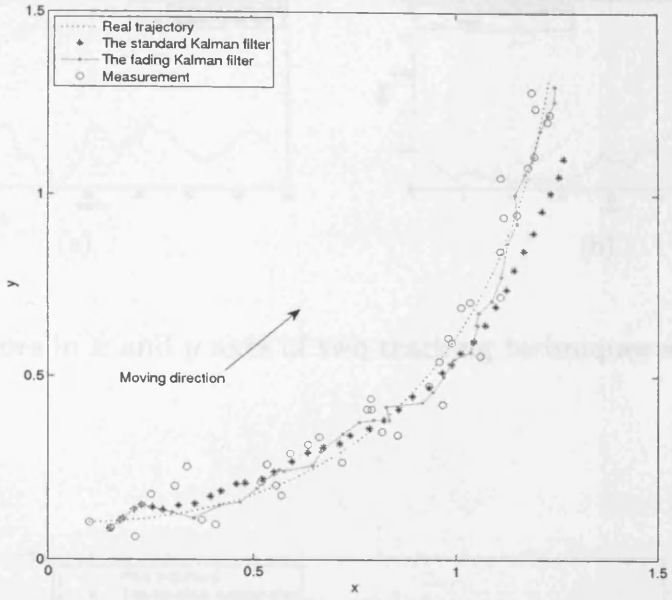
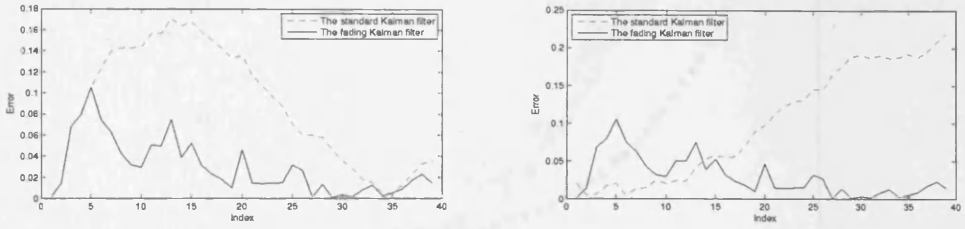


Figure 6.12: Experiment on synthetic data with initialisation of zero speed.



(a) (b)

Figure 6.13: Errors in x and y axes of two tracking techniques with initialisation of zero speed.

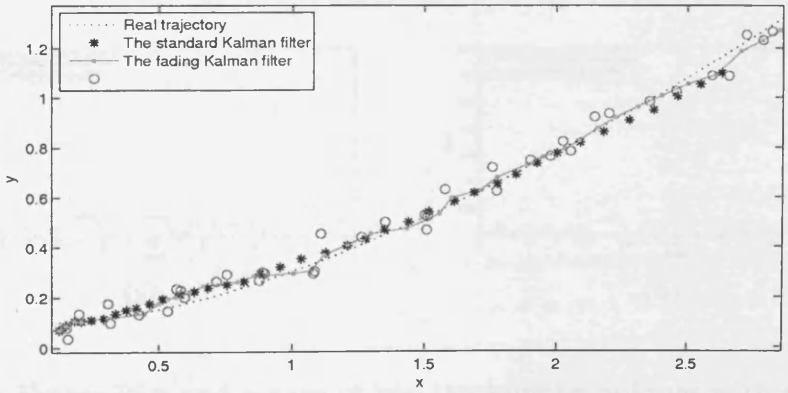
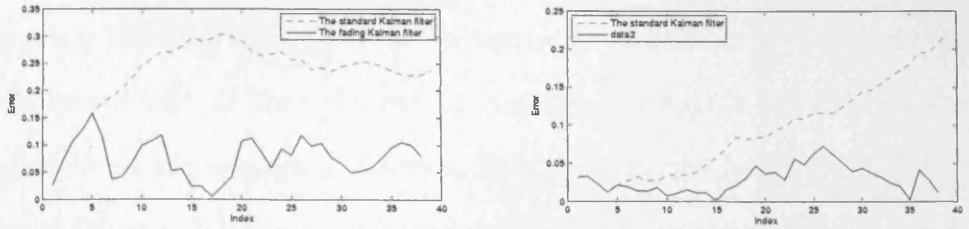


Figure 6.14: Experiment on synthetic data with initialisation of zero speed.



(a) (b)

Figure 6.15: Errors in x and y axes of two tracking techniques with initialisation of zero speed.

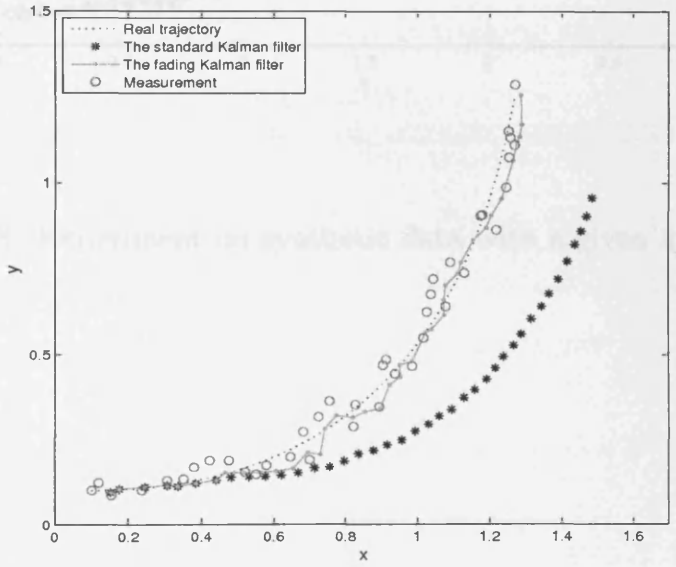
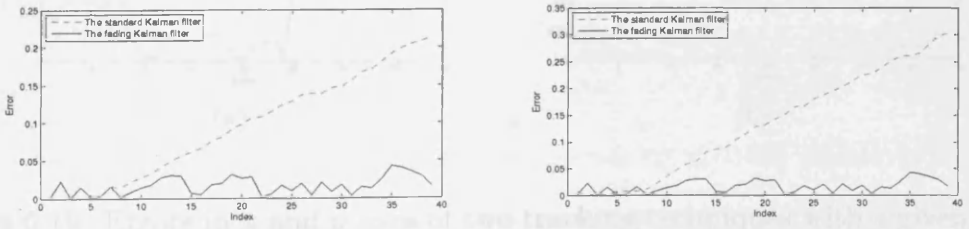


Figure 6.16: Experiment on synthetic data with a given initial speed.



(a) (b)

Figure 6.17: Errors in x and y axes of two tracking techniques with a given initial speed.

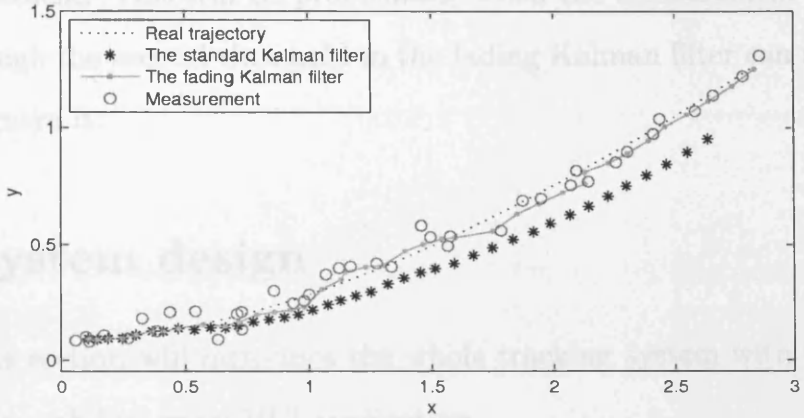
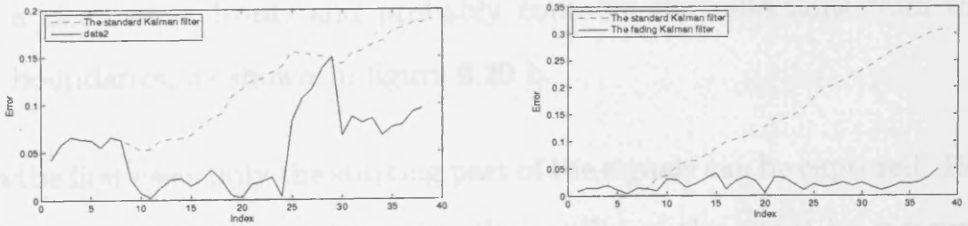


Figure 6.18: Experiment on synthetic data with a given initial speed.



(a) (b)

Figure 6.19: Errors in x and y axes of two tracking techniques with a given initial speed.

In summary, the main advantage of using the fading Kalman filter is the significantly reduced error by its self adaptivity. On the other hand, its disadvantage, in both cases, is that the fading Kalman filter does not provide the result as smoothly as the standard Kalman filter, due to the heavier weights given to the measurement. This will be problematic when the measurement error is very high, although the second threshold in the fading Kalman filter can recognise the case and ignore it.

6.4 System design

Finally, this section will introduce the whole tracking system with the consideration of the usability as an HCI application.

For the continuous tracking functionality, the data acquisition card should no longer work in the pre-trigger mode. Instead, signals will need to be continuously captured for higher efficiency. This can make it more difficult than the pre-trigger mode. Two common difficult cases are summarised below:

- Signals may not be captured over adequately long periods, such that part of the signals will be truncated (figure 6.20 a);
- Signals may only contain the other truncated part, which does not have a clear wave front, and probably contains the reflections from the edge boundaries, as shown in figure 6.20 b.

In the first case, only the starting part of the signals can be captured. However, the first step is usually to determine the quality of the signal by measuring the energy, such that this will be easily recognised as a weak signal or noise. In addition, a very short signal is not informative enough for the cross-correlation function, due to the lost information during the time-shift process. Therefore, the ToF method is the best method for such a situation. The slope of the curve can be used to determine whether the signals are generated by impacts, or from

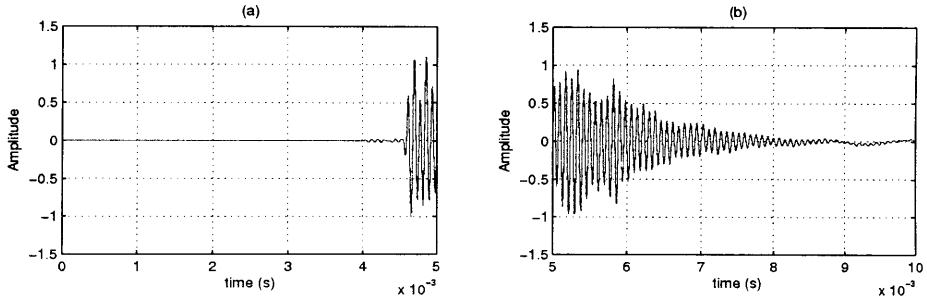


Figure 6.20: Example of truncated signal captured with continuous acquisition mode.

external noises. In the second case, the signals are captured long enough, but without a distinct wave front. In this case, the MLE method will be used.

Figure 6.21 shows the flowchart for the procedure to decide on the measurement method.

Figure 6.22 shows the flowchart of the Kalman filter with the steps of smoothing and fading-memory techniques. This is an extension of the process in figure 6.21 with the addition of a Kalman filter.

6.5 Summary

In this chapter, the Kalman filter has been studied and applied for the continuous tracking of finger movement in real-time. The fading memory technique with an extra state likelihood map was proposed and implemented to fulfil an adaptive Kalman filter, which can effectively overcome the problem of overshooting or undershooting during non-linear trajectories or acceleration/deceleration. Finally, the work flow for a complete tangible system, combining localisation techniques and tracking, was discussed and illustrated.

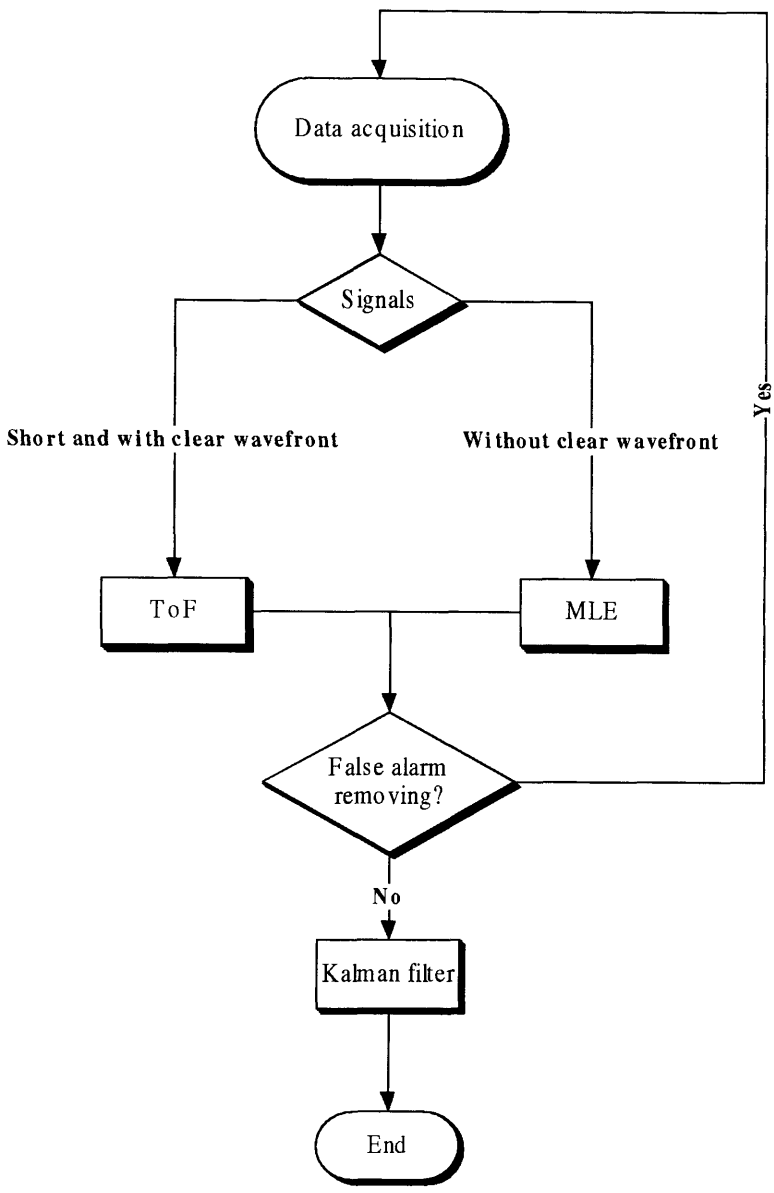


Figure 6.21: Flowchart for the procedure to decide on the measurement method.

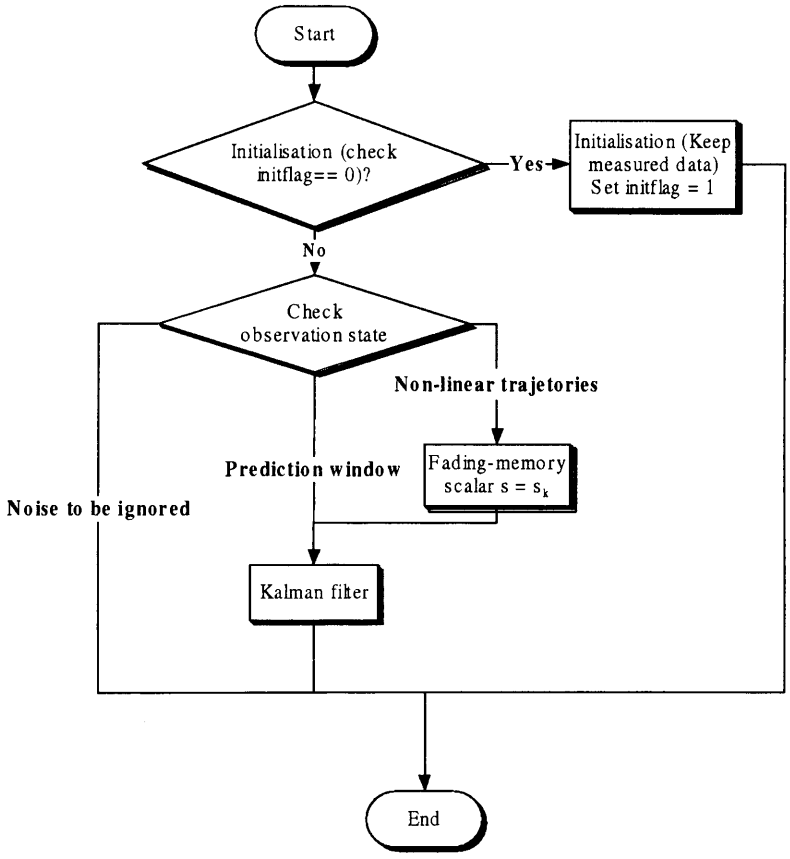


Figure 6.22: Flowchart of the Kalman filter with fading-memory and smoothing.

Chapter 7

Conclusions and future work

This chapter presents the conclusions of this work, outlines the main contributions of the research and makes recommendations for further work.

7.1 Conclusions

The aim of the research reported in this thesis was to investigate and develop acoustics-based sensing technologies, which can be adapted to virtually any physical object to create tangible interfaces, allowing the user to communicate freely with a computer, an interactive system or the cyber-world.

In chapter 2, various multimodal interactive techniques, including computer vision, IR, and active and passive acoustic, were reviewed and discussed. In the rest of the chapter, different technologies for the characterisation of acoustic propagation patterns and the positioning of acoustic sources in various media were evaluated. Array signal processing techniques were summarised from the conventional beamforming and TDOA methods to the alternative time reversal method. Acoustic transmission behaviour in solid materials was also introduced in respect of the desired applications in this work. Related approaches for localisation in solids were reviewed from two areas: NDT/NDE and seismology.

The LTM method was investigated with extensive experiments from various aspects. Chapter 3 first tested the standard cross-correlation function, and pre-

sented an improvement using an additional pre-filter. Experiments were carried out on a glass plate with a large amount of sampled data, resulting in a very accurate recognition of impacts to a resolution of $10mm$ by using two sensors together. Chapter 4 focused on the effect on the resolution limitation in different conditions, including the sensor placement and the physical properties of the materials. In summary, the resolution is limited by the wavelength of the dominant frequency, so that various factors can affect the performance. Most importantly, different thicknesses of the plates have significantly different velocities of wave propagation and eventually this will result in different resolutions. For the same type of material, the thinner plate has the higher resolution.

Chapter 5 discussed various TDOA methods. Two conventional approaches, ToF and cross-correlation, were introduced and experiments on each were performed. The ToF method with a proper band-pass filtration and envelope extraction is much more accurate than the cross-correlation method, which is vulnerable to boundary reverberations and the dispersion of plate waves. Studying the dispersive velocities is very important in estimating the time difference between two signals. Therefore, a method was proposed to measure the velocities at different frequencies and a curve fitting analysis was carried out to obtain the formulated velocities. In addition, a method was proposed to verify the measurement by inverse propagation of the received signals at the four sensors. The CWT was applied for time difference estimation at different scales (frequencies). The group velocities can be obtained with the measured phase velocities, so that the source location can be calculated with the group velocities at the corresponding frequencies. To overcome the ambiguities in measuring the precise time difference, MLE was applied.

Compared to the methods based on the TDOA principle, LTM has the advantages of simpler configuration and higher reliability. The sampling rate for the LTM method, in practice, can be reduced to $20kHz$ without compromising performance. In contrast, the sampling rate is critical to the TDOA method, as

discussed in chapter 5, as the resolution is closely related to the sampling rate. The TDOA has the advantage of easy installation on big surfaces, without calibrating all points as with LTM. Moreover, the computational demand is much lower with the TDOA.

Continuous tracking of finger scratching on a plate were implemented with the Kalman filter presented in chapter 6. Different adaptive techniques were reviewed and the fading memory technique was applied in addition to the standard Kalman filter to improve the performance during non-linear trajectories and acceleration/deceleration.

7.2 Contributions

The contributions of this research are summarised below:

1. Various multimodal technologies have been reviewed and compared with the proposed acoustic method. Advantages and disadvantages have been summarised in terms of different aspects, such as functionality, geometry dimension of interactive surface, portability, and installation.

2. The performace of the LTM method was tested with various experiments, including sensor selection, signal acquisition parameter selection, signal consistency testing, and reliability testing. Effect of applying an additional sensor has also been shown with experiments. The resolution of LTM has been studied based on the principle of time reversal. The experimental result has shown that the resolution is limited by the wavelength of the dominant frequency. The sensor placement has been found useful in optimising the reliability of the system, but not for the resolution. Moreover, experiments have shown that resolution is dependent on the wavelength by comparing different thicknesses of three materials and showing, that thin plates have higher resolution than thicker plates, due to the dispersive velocities of the Lamb wave. Limitations of energy ratio between signals from

multiple impacts have been discussed.

3. The conventional approaches of ToF and cross-correlation for time difference estimation have been compared, and cross-correlation has been shown to be significantly affected by the dispersion and reverberation. In addition, a simple technique was introduced to verify the measured phase velocities by inverse propagation of the received signals to the known source position in the frequency domain for any arbitrary material without physical knowledge. This also provides the means to obtain the group velocities, which are useful for the CWT method for time difference estimation. Another contribution is the combination of CWT and MLE to overcome both the problems of dispersion and reverberation effectively.

4. A successful attempt at applying the Kalman filter to the continuous tracking of finger movement on a plate. The Kalman filter was applied with an additional step of a fading memory technique to improve the performance, which has been demonstrated with both real measured data and synthetic data simulating non-linear trajectories and acceleration/deceleration.

7.3 Future work

There are a number of potential ways to improve on this work in the future:

1. Although the LTM method can provide high reliability and resolution on properly selected materials, it is still tedious to calibrate a large number of points manually if the desired application requires it. The development of a machine for automatic template preparation could save time and preclude manual operation errors. On the other hand, it would be interesting to investigate an intelligent method to reduce the number of points for calibration.

2. As indicated in chapter 4, the question of multiple impact identification

remains open for future work. The main difficulty is the lack of the knowledge of the number of simultaneous impacts. Also, as mentioned before, the question of the practical limitation of measuring exactly simultaneous impacts with the same strength remains open and worth investigating.

3. Another practical consideration of the performance of LTM is the effect of the condition change on the propagation behaviour, such as the temperature change. For this research, most of the work was carried in the laboratory at indoor temperatures. It is worth investigating the physical properties in various conditions, and quantifying the formulation to improve its robustness.

4. The measured phase velocities have been verified satisfactorily. There is a potential method of estimating the source location in the phase domain with the knowledge of phase velocities. A current difficulty arises because the phase domain is not continuously dominated by the vibrational signals only, but by noise as well. It is worth investigating a method for time difference estimation with phase information and different velocities at different frequencies.

5. Particle filtering is a promising technique for tracking objects with nonlinear behaviour [AMGC02]. When the computation can be reduced down effectively with no compromise in accuracy, it would be very interesting in this application to replace the Kalman filter, and even to track multiple objects.

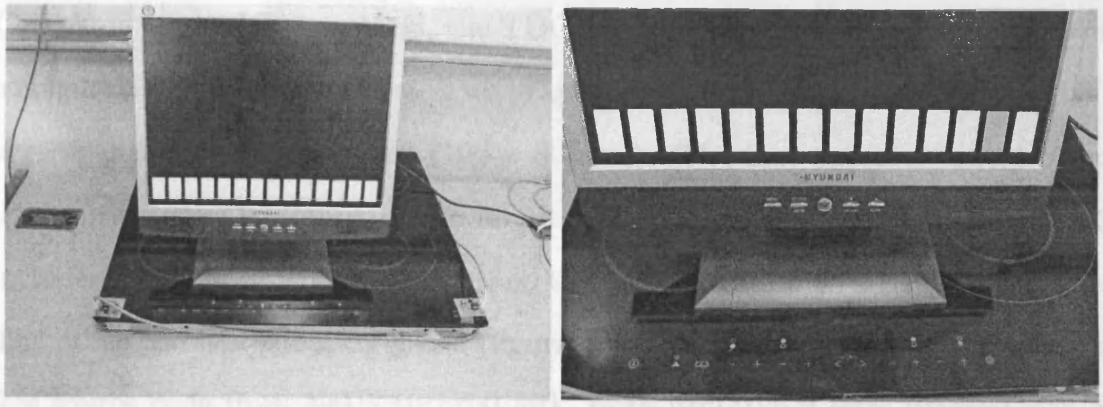
Appendix A

HCI demonstrations

In order to evaluate these developed techniques, some HCI demonstrators have been implemented during this work.

A.1 A hob panel and a simple touch screen using LTM

The LTM method can be easily implemented on almost any surface. A hob panel, provided by SIEMENS, was used to evaluate the feasibility of applying this technology on this material. 14 buttons printed on the surface were used, and the demonstration displayed the colour change on the screen when one taps on the corresponding position, as shown in figure A.1. The other prototype was implemented on a small glass plate placed in front of the computer screen, simulating a small touch screen (figure A.2). This is a rather simple prototype with only four points calibrated. Similarly, the colour changed at the tapped position. The small plate can be eventually used as a remote computer interface for some simple applications, such as a media player or a presentation slides controller.



(a) (b)

Figure A.1: Hob panel demonstration with LTM.

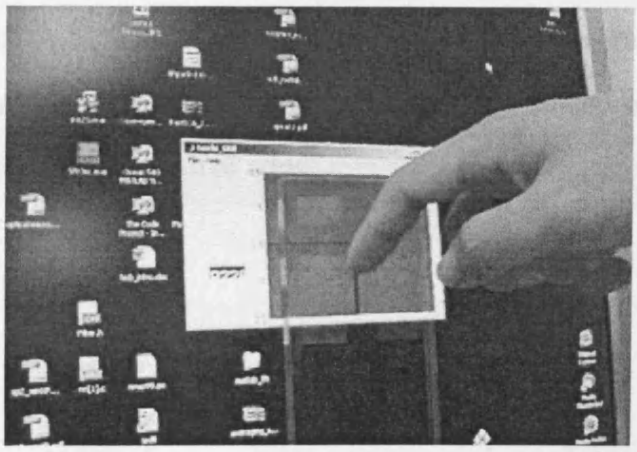


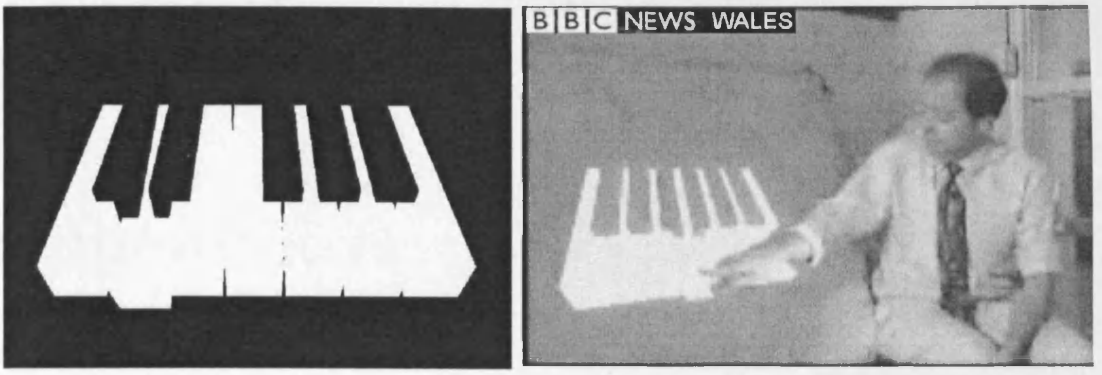
Figure A.2: LTM on a small glass plate simulating a touch screen.

A.2 A virtual piano and a shooting game using TDOA

Compared to the LTM method, the TDOA method has the advantage of simpler configuration on large surfaces. Two graphical applications were implemented on a general purpose whiteboard ($100 \times 80(cm)$). The first one is a virtual piano that converts the whiteboard into an interactive musical instrument. Figure A.3 includes the screenshot of the piano and the demonstration on a BBC programme. The other one is a shooting game (figure A.4), integrating the algorithm with a free source code from NeHe tutorial [ot]. Both prototypes were programmed in C++ and openGL was used for the 3D animation.

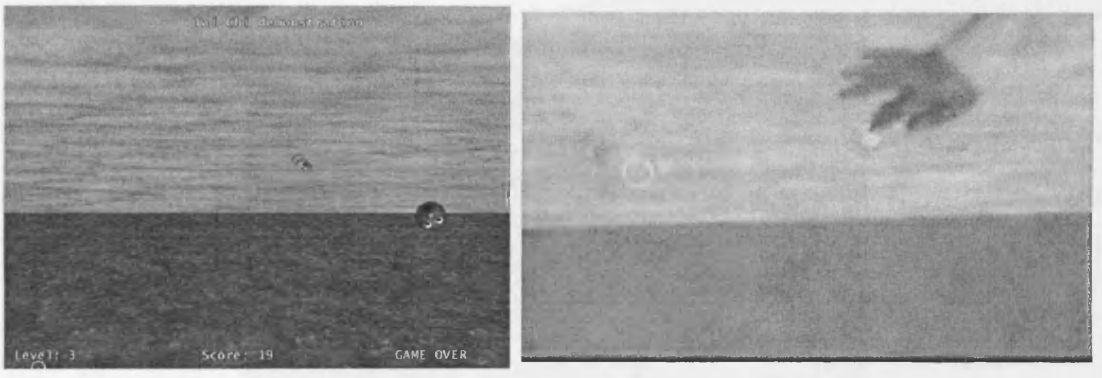
A.3 A drawing pad and an interactive map viewer with continuous tracking functionality

The tracking of the finger is simulated as the mouse event "drag & drop". The control of the mouse cursor by the tracking system allows the using of classic Microsoft Windows applications. The screenshots of drawing with Microsoft Paint are shown in figure A.5. The other demonstration is an interactive map viewer, which is a customisation of the Google map, as shown in figure A.6, so that one can navigate the map in a natural way.



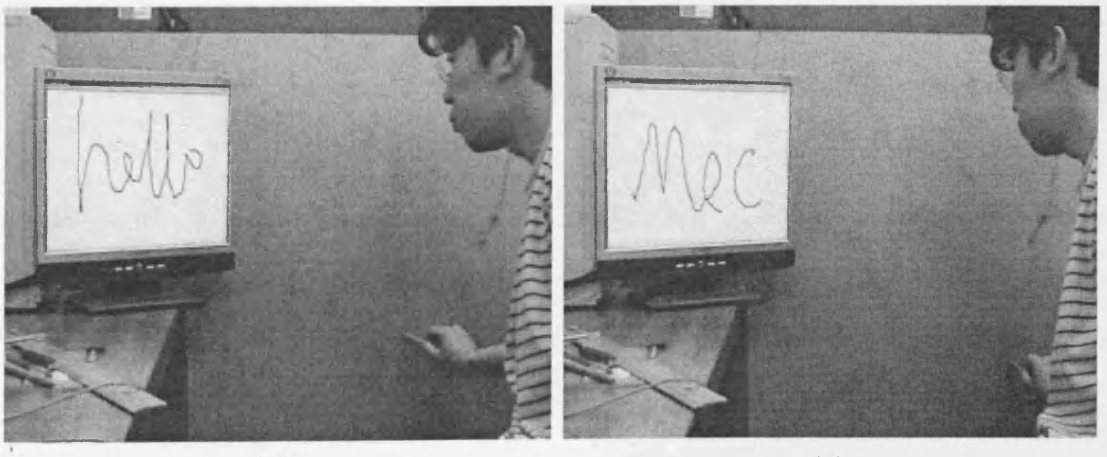
(a) (b)

Figure A.3: Screenshot of the piano (a) and the demonstration (b).



(a) (b)

Figure A.4: Screenshot of the shooting game (a) and the demonstration (b).



(a) (b)

Figure A.5: Screenshots of drawings on Microsoft Paint.

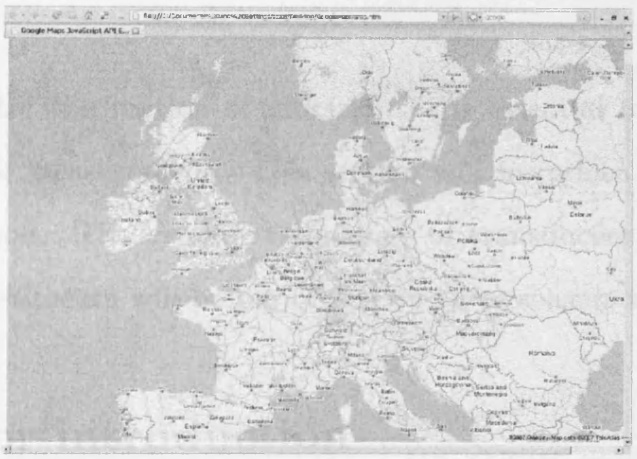


Figure A.6: Screenshot of map viewer based on Google map.

Appendix B

Measurement of acoustic vibrations

In order to monitor the motion of vibration, it is essential to choose a proper transducer for the measurement. This section will introduce the techniques of measuring vibration. The general technology of transducers will be reviewed, and practical experiments will be performed on three sensors with comparisons.

B.1 Sensor introduction

Many different methods have been developed to measure vibratory motions. Piezo-electricity, derived from the Greek for pressure electricity, is the ability of crystals to generate a voltage in response to applied mechanical stresses, and is the most common method used to produce vibration sensing transducers nowadays. The piezo-electric effect is reversible in that piezo-electric crystals, when subjected to an externally applied voltage, can change its shape by a small amount. Certain ceramic materials could be made piezo-electric when subjected to a high polarising voltage, a process analogous to magnetising a ferrous material [Mea99]. By the 1960's, researchers had discovered a weak piezo-electric effect in whale bone and tendon. This began an intensive search for other organic materials that might exhibit piezo-electricity. In 1969, Kawai found very high piezo-activity in

the polarized fluoropolymer, polyvinylidene fluoride (PVDF). While other materials, like nylon and PVC exhibit the effect, none are as highly piezo-electric as PVDF and its copolymers.

Vibration sensors are available in various forms. The accelerometer is the most commonly used vibration measurement transducer due to its small size, wide range of sensitivities, and large usable frequency range [McC95]. An accelerometer attached to a larger object can be modelled as a single degree of freedom (DOF) vibration system excited by a moving base, as shown in figure B.1.

B.2 Sensor testing

Testing of different sensors has been performed. To choose the proper sensor for this application, two primary parameters should be considered, which are the frequency response range and sensitivity. For the vibratory signals generated with a fingertip, the contained frequency spectrum is mainly available in the audible range, rather than the high ultrasonic frequency, which can normally be induced with special piezo-electric transducers [Tuc01].

In this work, four sensors were selected:

Piezo-film (PVDF)

This is the DT series of piezo-film sensor elements from Measurement Specialties Inc [Mea99], as shown in figure B.2(a). The DT series sensors are the simplest form of piezo-film sensors, used primarily as dynamic strain gages and contact microphones for vibration or impact detection. They can be readily adhered to a surface with double-sided tapes or epoxy resin.

This sensor seems well adapted for the application, which requires high sensitivity, wide bandwidth and SNR as high as possible. However, the main drawback is the bad response for low frequency due to the high-pass filter behaviour of this kind of sensor.

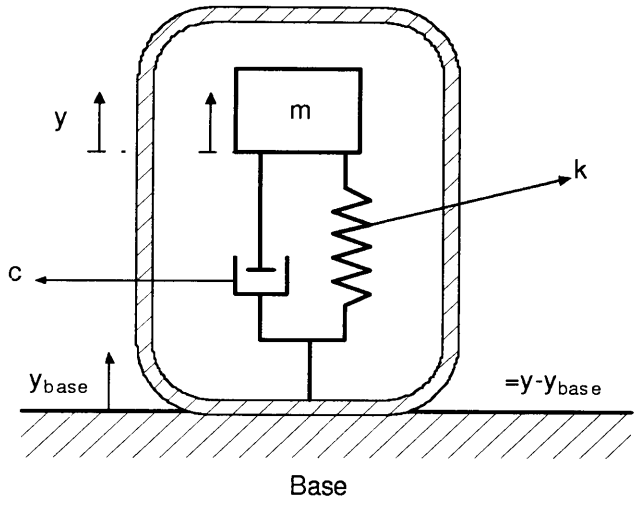


Figure B.1: Accelerometer model.

Piezo-electric ceramic disc

The provided piezo-electric ceramic sensor is bonded to a metallic plate, as shown in figure B.2(b). The resonance frequency of this type of sensor is normally in the range from $1kHz$ to $10kHz$. This is normally used to produce vibration accelerometers or sounders, but can also be used directly as a sensor attached to a flat surface with adhesive or simple double-side tapes.

It has the advantage of very low cost and higher sensitivity to low frequency signals. However, the frequency response is much affected by the method of mounting it, such that the performance might vary on different materials due to the different elastic properties.

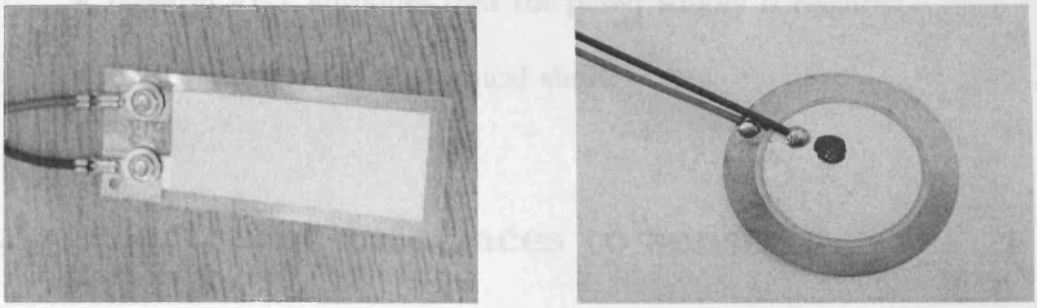
Murata shock sensor

This type of sensor is in fact a piezo-electric ceramic plate formed in a plastic case, with an enhanced sensitivity to the original piezo-electric ceramic sensor. The selected model of this sensor is *pks1-4a1*, which is compact and lightweight, from Murata Manufacturing Company, Ltd. [Mur], as shown in figure B.3(a). The high sensitivity assures it picks up micro level impacts and vibrations. The frequency response for such a sensor is provided in the datasheet as shown in figure B.3(b). A significant resonance frequency is at about $2kHz$ and quite linear frequency response is below $1kHz$.

Knowles acoustic accelerometer (BU-1771)

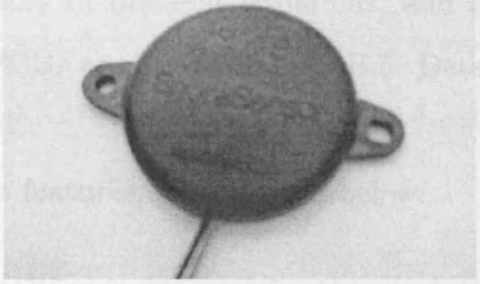
The accelerometer of KU-1771 (figure B.4(a)) from Knowles acoustic [Kno] provides a small sized accelerometer made of piezo-ceramic material. It has the following features:

- High vibration sensitivity (in dB relative to 1.0 Volt/g)
- Small size
- Wider frequency range (figure B.4(b))

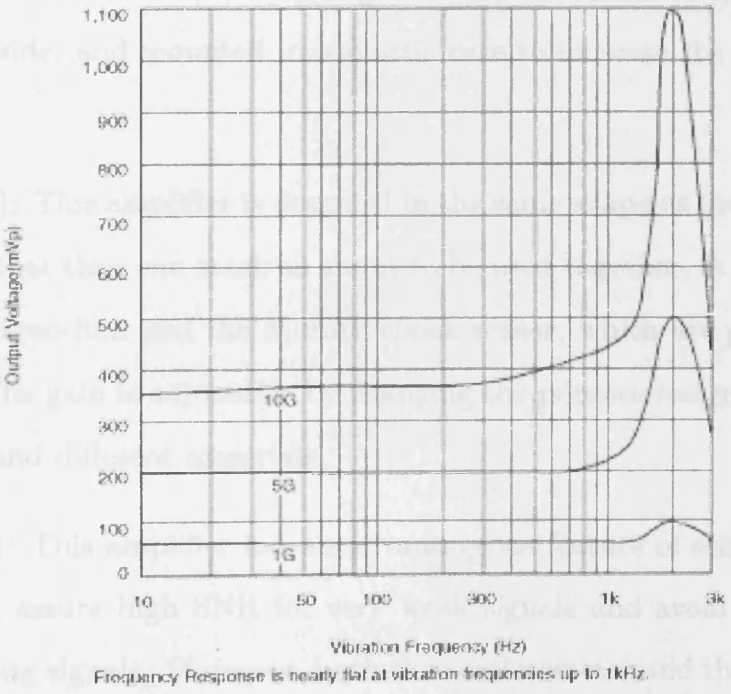


(a) (b)

Figure B.2: DT series piezo-film (PVDF) (a) and piezo-ceramic disc (b).



(a)



(b)

Figure B.3: Murata shock sensor (pks1-4a1) (a) and its frequency response (b).

- Integral FET amplifier that the power supply is required
- High resistance to mechanical shock

B.3 Electronic interfaces to sensors

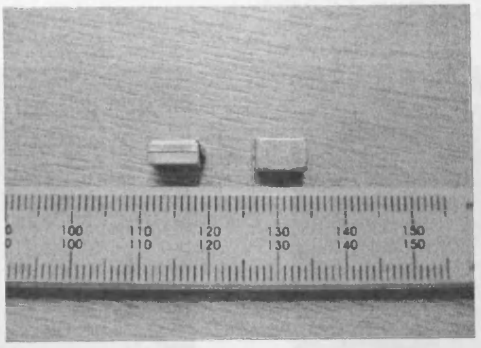
Piezo-electric sensing elements are charge generating devices that require high impedance signal conditioning instruments to interface between them. In order to use the sensors, an amplification circuit is often required that can reduce the impedance and amplify weak signals. Three different amplifiers have been designed to fit the sensors in different situations, and have been produced in printed circuit board (PCB) as shown in figure B.5. Detailed schematics can be found in Appendix C.

Each one has its own features, as described below:

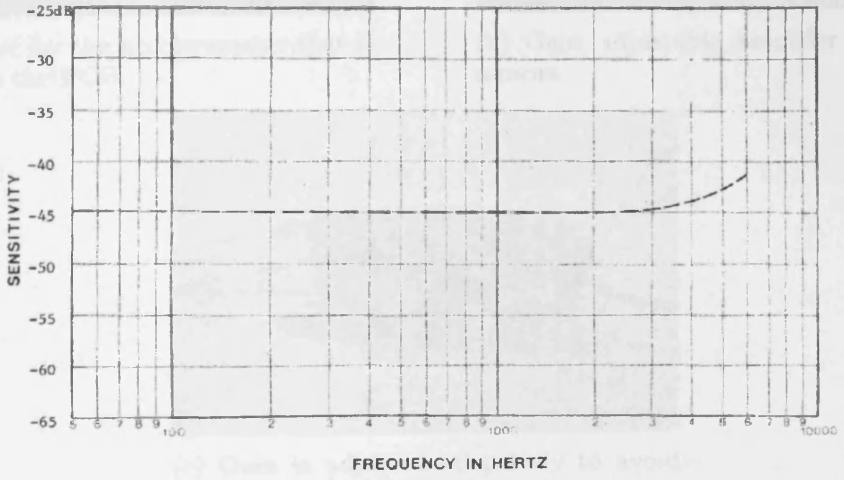
Figure B.5(a): This amplifier is designed for the accelerometer with an amplification gain of ≈ 67. The PCB is glued with the sensor at the corner of the bottom side, and mounted in a plastic case to increase the steady contact area.

Figure B.5(b): This amplifier is designed in the same shape as the piezo-ceramic disc, so that they can stack as a unit to be used together. It is also suitable for the piezo-film and the Murata shock sensor, which are passive components. The gain is adjustable by changing the potentiometer to fit different sensors and different materials.

Figure B.5(c): This amplifier has an advantageous feature of self-adaptive gain that can assure high SNR for very weak signals and avoid saturation for very strong signals. Moreover, both the accelerometer and the piezo-sensors can be used by simply changing the jumpers on the PCB. However, one disadvantage is that it cannot reveal the real signal magnitude.

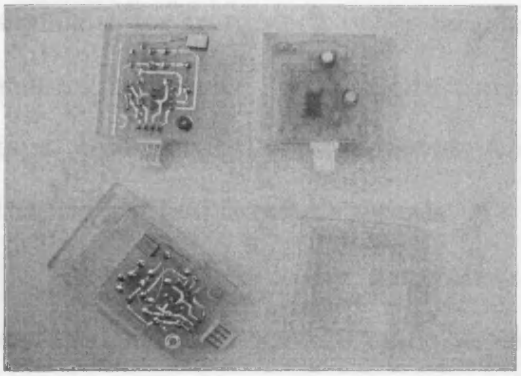


(a)

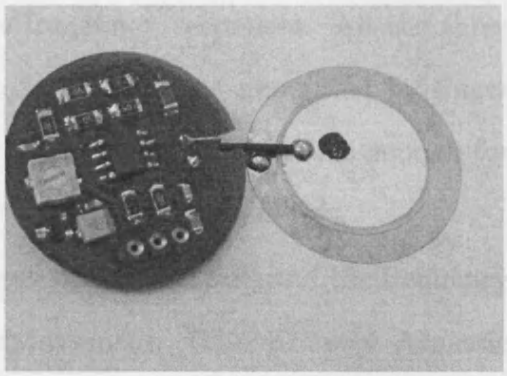


(b)

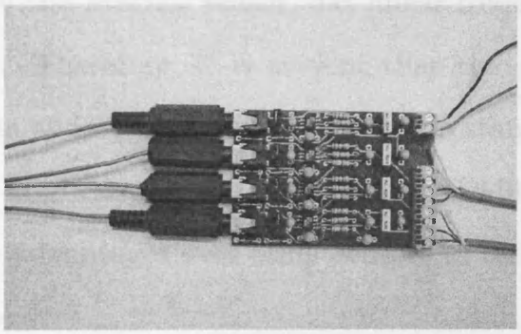
Figure B.4: Knowles acoustic accelerometer model KU-1771 (a) and its frequency response (b).



(a) Amplifier for the accelerometer that is attached on the PCB.



(b) Gain adjustable amplifier for piezo-sensors.



(c) Gain is adjusted adaptively to avoid low SNR and saturation.

Figure B.5: Three amplifier circuits.

The sensitivity of the PVDF strip was found to be very low for the signals induced by finger tapping, and is not useful in practice in this project. Figure B.6 shows the typical signals captured with the other three sensors on a glass plate at the sampling rate of $100kHz$. The data is acquired through the data acquisition card (Adlink PCI DAQ-2005) by the computer (see Appendix C for detailed specifications). The left column shows the signals in the time domain and the right column shows the frequency spectrum of the signals. Clearly, the signals differ from each other largely in the frequency responses. All the three sensors have been found quite sensitive to the vibrations generated by finger taps. However, only the Knowles Acoustic accelerometer is sensitive enough for the friction (continuous) signals.

The piezo-ceramic disc generates the lowest voltage output, and the frequency bandwidth is almost the same as the Murata sensor. The Knowles Acoustic accelerometer has shown to contain much wider spectrum bandwidth. However, recall the data sheet of the Murata sensor, flat linear frequency response is only available below $1kHz$. Therefore, it is evident that the accelerometer has the best performance for its widest bandwidth. In order to standardise the remaining part of this thesis, the Knowles Acoustic accelerometer will be primarily used for experiments due to its advantages over other sensors in terms of both sensitivity and frequency response.

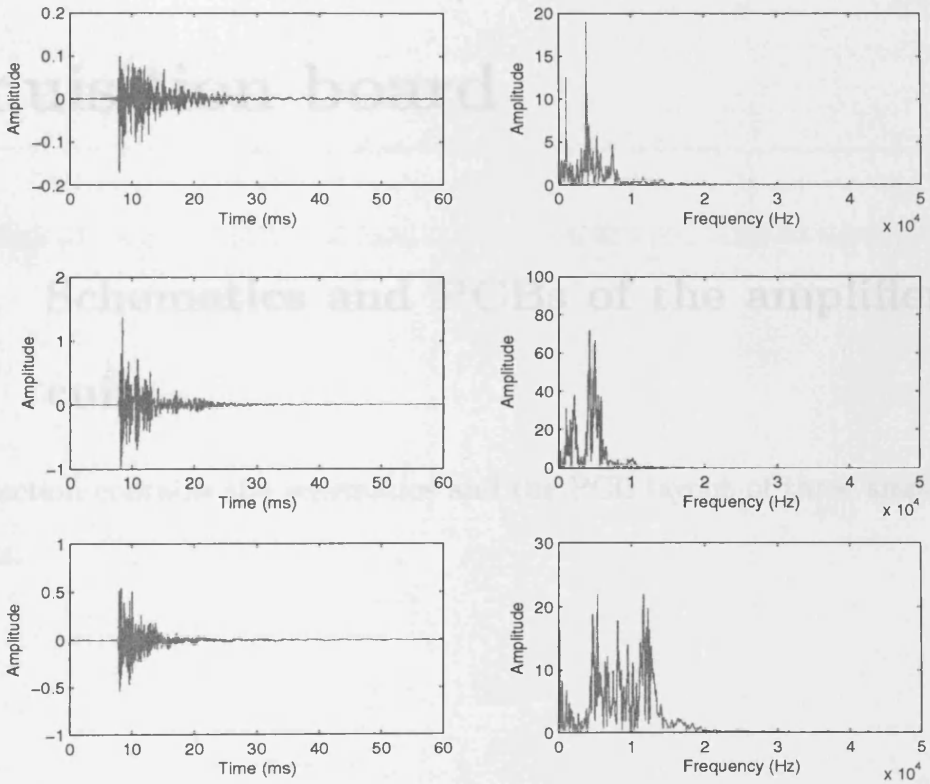


Figure B.6: Signals captured by three sensors on a glass plate.

Appendix C

Amplifier circuits and data acquisition board

C.1 Schematics and PCBs of the amplifier circuits

This section contains the schematics and the PCB layout of three amplification circuits.

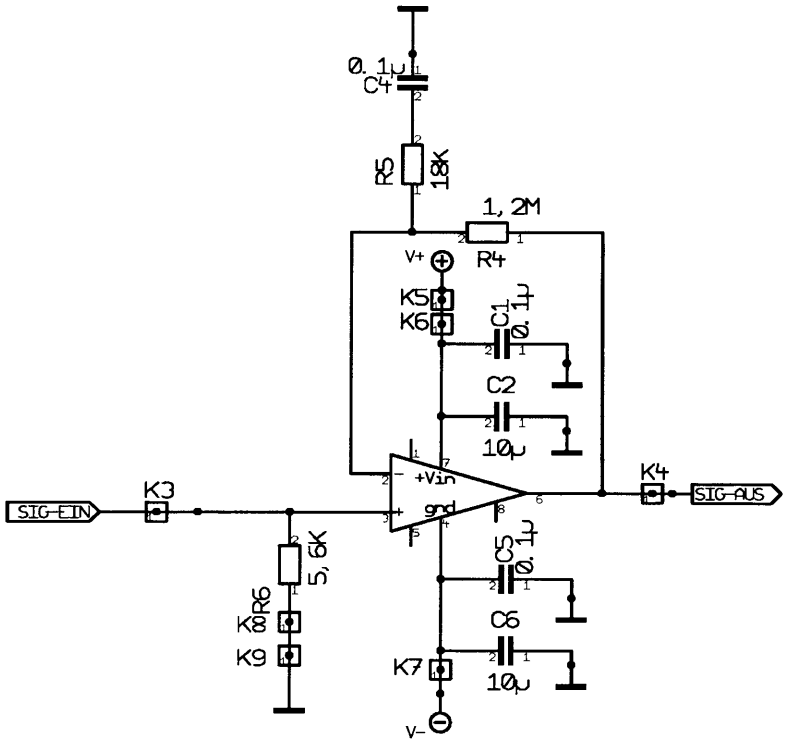


Figure C.1: Schematics of the accelerometer amplifier.

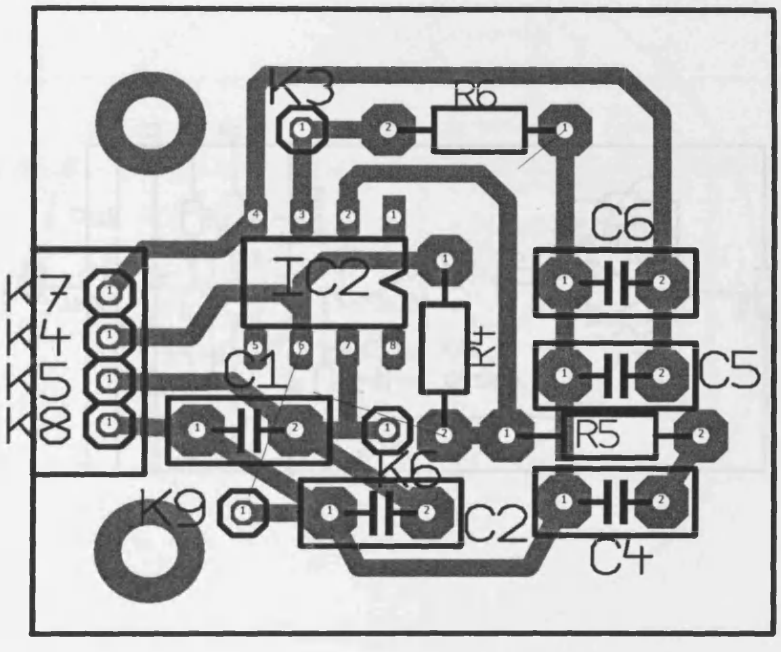


Figure C.2: PCB of the accelerometer amplifier.

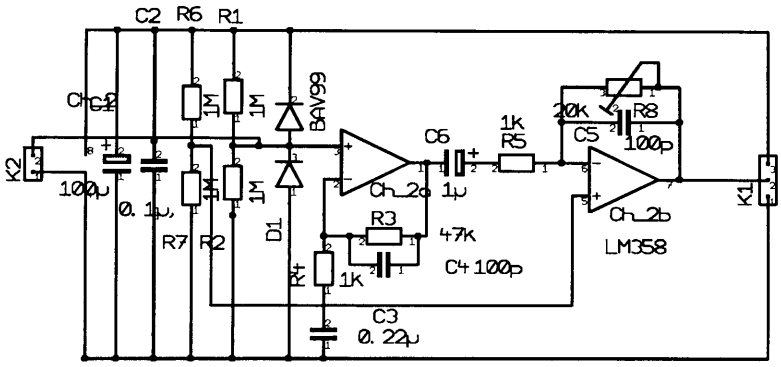


Figure C.3: Schematics of the amplifier circuit for piezo-sensors with adjustable gain.

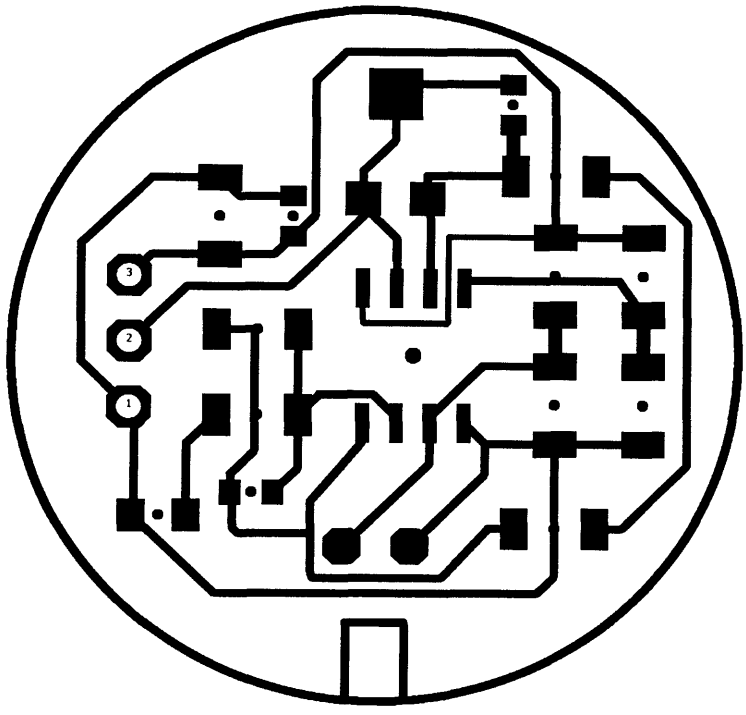


Figure C.4: PCB of the amplifier circuit for piezo-ceramic sensors with adjustable gain.

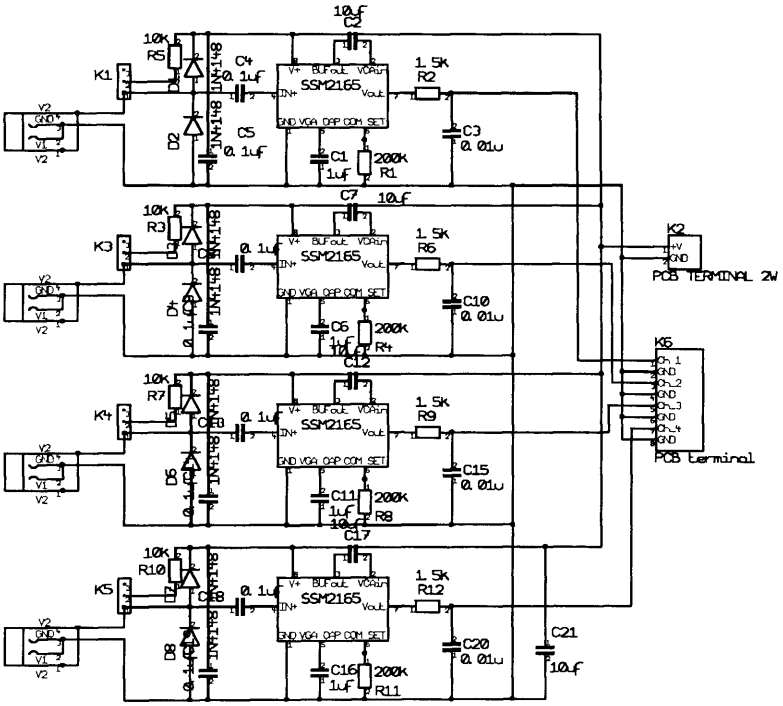


Figure C.5: Schematics of the four channel adaptive amplifier.

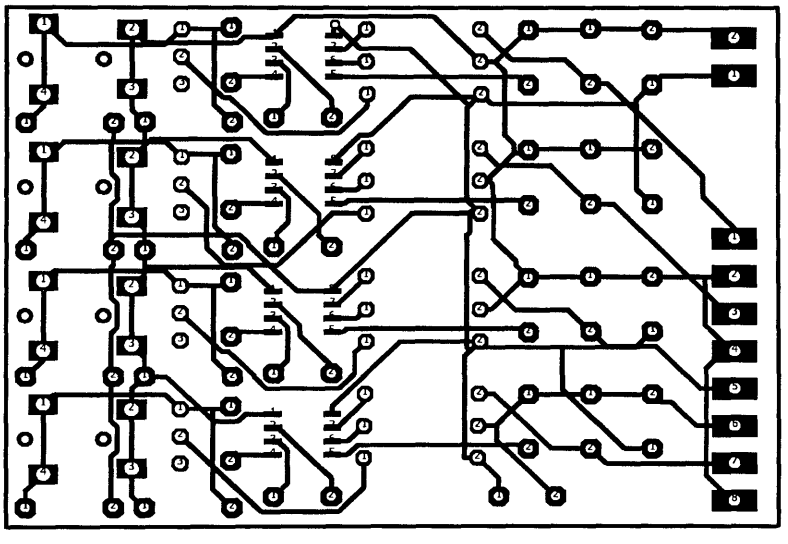


Figure C.6: PCB of the four channel adaptive amplifier.

C.2 Data acquisition board

ADLINK DAQ-2005 is a simultaneous-sampling multifunction DAQ card to meet a wide range of application requirements. The devices can simultaneously sample 4 Analog Input(AI) channels with differential input configuration in order to achieve maximum noise elimination. They also provide 2-CH 12-bit analog outputs with waveform generation capability, which can be performed together with analog input functions. If more analog input or output channels are required, multiple cards can be synchronized through the SSI (system synchronization interface) bus. This makes the DAQ-2000 series ideal for the stimulus/response test. The DAQ-2000 series also feature analog and digital triggering, 24-CH programmable digital I/O lines, and 2- CH 16-bit general-purpose timer/counters. The auto-calibration functions adjust the gain and offset to within specified accuracies such that you do not have to adjust trimpots to calibrate the cards.

Main features

- Supports a 32-bit 3.3 V or 5 V PCI bus
- 4-CH differential analog inputs
- Up to 500 KS/s simultaneous-sampling rate
- 16-bit A/D resolution
- Up to 512 samples A/D FIFO
- Bipolar or unipolar analog input ranges:
  - Bipolar: $\pm 10V$, $\pm 5V$, $\pm 2.5V$, $\pm 1.25V$,
  - Unipolar: $0 - 10V$, $0 - 5V$, $0 - 2.5V$, $0 - 1.25V$
- Programmable gains of $\times 1$, $\times 2$, $\times 4$, $\times 8$
- Versatile trigger sources: software trigger, external digital trigger, analog trigger and trigger from System Synchronization Interface (SSI).

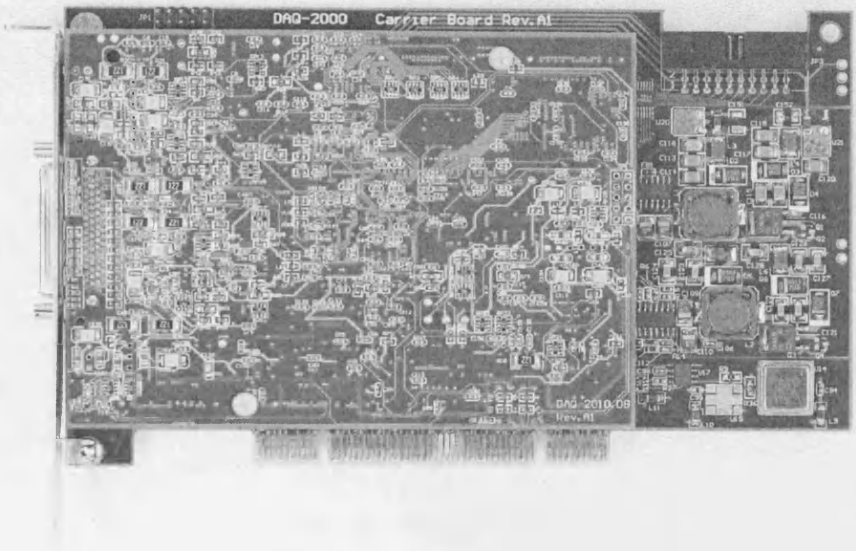


Figure C.7: Adlink DAQ-2005 PCI card.

- A/D Data transfer: software polling & bus-mastering DMA with Scatter/Gather functionality
- Four A/D trigger modes: post-trigger, delay-trigger, pre-trigger and middle-trigger

References

[3M] 3M. Dispersive signal technology. www.3Mtouch.com (last visted 14 May 2007).

[ADH+06] A. Apostoloudia, E. Douka, L.J. Hadjileontiadis, I.T. Rekanos, and A. Trochidis. Time-frequency analysis of transient dispersive waves: A comparative study. *Applied Acoustics*, pages 1–14, 2006.

[Ald04] F. Aldersons. *Toward a three-dimensional crustal structure of the Dead Sea region from local earthquake tomography.* PhD thesis, The Senate of Tel-Aviv University, June 2004.

[Ame] Pagasus America. http://www.pegasusamerica.com (last visited 14 May 2007).

[AMGC02] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2), February 2002.

[BA98] B. Barshan and B. Ayrulu. Performance comparison of four time-of-flight estimation methods for sonar signals. *Electronics Letters*, 34(16):1616–1617, August 1998.

[BBP04] P. Brossier, J. P. Bello, and M. D. Plumbley. Real-time temporal segmentation of note objects in music signals. In *Proceedings of*

the International Computer Music Conference (ICMC 2004), Miami, Florida, USA, November 2004.

[BdVB80] A. J. Berkhout, D. de Vries, and M. M. Boone. A new method to acquire impulse responses in concert halls. *Journal of the Acoustical Society of America*, 68:179–183, 1980.

[BG02] S. T. Birchfield and D. K. Gillmor. Fast bayesian acoustic localization. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 1793–1796, Orlando, FL, USA, May 2002.

[BK92] B. Barshan and R. Kuc. A bat-like sonar system for obstacle localization. *IEEE transactions on Systems, Man, and Cybernetics*, 22(4):636–646, July/August 1992.

[BK00] R. Bousseljot and D. Kreiseler. Waveform recognition with 10,000 ECGs. In *Computers in Cardiology 2000*, pages 331–334, 24-27 Sept. 2000.

[BPT03] L. Borcea, G. Papanicolaou, and C. Tsogka. A resolution study for imaging and time reversal in random media. *Contemporary Mathematics*, 333:63–77, 2003.

[Bra95] M. S. Brandstein. *A Framework for Speech Source Localization Using Sensor Arrays.* PhD thesis, Brown University, May 1995.

[BT03] R.J. Barsanti and M. Tummala. Wavelet-based time delay estimates for transient signals. In *Conference Record of the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 1173–1177, Nov. 2003.

[CBC95] J. Crowley, F. Berard, and J. Coutaz. Finger tracking as an input device for augmented reality. In *International Workshop on Gesture and Face Recognition*, pages 195–200, Zurich, June 1995.

[Che01] Nisha Checka. A system for tracking and characterizing acoustic impacts on large interactive surfaces. Master's thesis, Department of electrical engineering and computer science, Massachusetts institute of technology, May 2001.

[CHP78] Y. Chan, R. Hattin, and J. Plant. The least squares estimation of time delay and its use in signal detection. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(3):217–222, Jun 1978.

[Con] M. Conry. Notes on wave propagation in anisotropic elastic solids. http://www.acronymchile.com/elasticity.html, last visited 14 May 2007.

[DB88] III Doles, J.H. and F.D. Benedict. Broad-band array design using the asymptotic theory of unequally spaced arrays. *IEEE Transactions on Antennas and Propagation*, 36(1):27–33, Jan. 1988.

[DRS04] Y. Ding, R.L. Reuben, and J.A. Steel. A new method for waveform analysis for estimating ae wave arrival times using wavelet decomposition. *NDT&E international*, 37(4):279–290, June 2004.

[dvi03] Digital vision touch (DviT) white report. Technical report, SMART Technologies Inc, Calgary, Canada, 2003.

[EO99] M. Efe and L. Ozbek. Fading Kalman filter for manoeuvring target tracking. *Journal of the Turkish Statistical Association*, 2(3):193–206, 1999.

[ES94] P. S. Earle and P. M. Shearer. Characterization of global seismograms using an automatic picking algorithm. *Bulletin of the Seismological Society of America*, 84(2):366–376, April 1994.

[Fah85] F. J. Fahy. *Sound and structural vibration.* Academic Press, London, 1985.

[FBQL91] M. Farooq, S. Bruder, T. Quach, and S.S. Lim. Adaptive filtering techniques for manoeuvring targets. *Proceedings of the 34th Midwest Symposium on Circuits and Systems*, 1:31–34, May 1991.

[FC84] D. Fischell and C. Coker. A speech direction finder. In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 9, pages 128–131, March 1984.

[FCD+00] M. Fink, D. Cassereau, A. Derode, C. Prada, P. Roux, M. Tanter, J.-L. Thomas, and F. Wu. Time-reversed acoustics. *Report on progress in Physics*, 63:1933–1995, 2000.

[Fin92] M. Fink. Time reversal of ultrasonic fields-part I: basic principles. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 39(5):555, September 1992.

[FP01] M. Fink and C. Prada. Acoustic time-reversal mirrors. *Inverse Problems*, 17(1):R1–R38, February 2001.

[FPWC89] M. Fink, C. Prada, F. Wu, and D. Cassereau. Self focusing in inhomogeneous media with time reversal acoustic mirrors. In *IEEE Proceedings, Ultrasonics Symposium*, volume 2, pages 681–686, Montreal, Que., Canada, 1989.

[GC99] J. C. Goswami and A. K. Chan. *Fundamentals of Wavelets, theory, algorithms, and applications.* Wiley, New York, 1999.

[GH97] L. Gaul and S. Hurlebaus. Identification of the impact location on a plate using wavelets. *Mechanical Systems and Signal Processing*, 12(6):783–795, 1997.

[GKM98] I. Grabec, T. Kosel, and P. Muzic. Location of continuous AE sources by sensory neural networks. *Ultrasonics*, 36:525–530, 1998.

[Goo97] S. G. Goodridge. *Multimedia sensor fusion for intelligent camera control and human-computer interaction.* PhD thesis, Graduate Faculty of North Carolina State University, USA, 1997.

[Gor91] M. R. Gorman. AE source orientation by plate wave analysis. *Journal of Acoustic Emission*, 9(4):283–288, 1991.

[Gra75] K. F. Graff. *Wave motion in elastic solids.* Clarendon press, Oxford, 1975.

[Han05] J. Y. Han. Low-cost multi-touch sensing through frustrated total internal reflection. In *Proceedings of the 18th annual ACM symposium on User interface software and technology*, pages 115 – 118. ACM Press, 2005.

[Har61] R. Harrington. Sidelobe reduction by nonuniform element spacing. *IEEE Transactions on Antennas and Propagation [legacy, pre - 1988]*, 9(2):187–192, Mar 1961.

[Hay01] S. Haykin, editor. *Kalman filtering and neural networks.* Wiley, New York, 2001.

[HKWS00] S. Holland, T. Kosel, R. Weaver, and W. Sachse. Determination of plate source, detector separation from one signal. *Ultrasonics*, 38(1-8):620–623, March 2000.

[Hof72] M. Hoff. Acoustic emission source location using single and multiple transducer arrays. In *Ultrasonics Symposium*, pages 8–12, 1972.

[HOG02] M. A. Hamstad, A. O'Gallagher, and J. Gary. A wavelet transform applied to acoustic emission signals: Part 2: source location. *J. Acoustic Emission*, 20:62–82, 2002.

[HW94] M. Hunke and A. Waibel. Face locating and tracking for human-computer interaction. *In 28th Asilomar Conference on Signals, Sys-*

tems and Computers at Monterey, California Pittsburgh. School of Computer Science, Carnegie Mellon University, 1994.

[IBM] IBM. http://www.ibm.com (last visited 14 may 2007).

[IF98] R. K. Ing and M. Fink. Time-reversed lamb waves. *IEEE Transactions On Ultrasonics, Ferroelectrics, And Frequency Control*, 45(4):1032–1043, July 1998.

[IQCF05] R. K. Ing, N. Quieffin, S. Catheline, and M. Fink. In solid localization of finger impacts using acoustic time-reversal process. *Applied Physics Letters*, 87(204104), 2005.

[Ish62] A. Ishimaru. Theory of unequally-spaced arrays. *IEEE Transactions on Antennas and Propagation [legacy, pre - 1988]*, 10(6):691–702, Nov 1962.

[IU97] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of CHI '97*, pages 234–241. ACM Press, March 1997.

[IWO+99] H. Ishii, C. Wisneski, J. Orbanes, B. Chun, and J. Paradiso. Pingpongplus: design of an athletic-tangible interface for computer-supported cooperative play. In *Proceedings of CHI 99*, pages 394–401. ACM Press, May 1999.

[JHW+04] J. Jiao, C. He, B. Wu, R. Fei, and X. Wang. Application of wavelet transform on modal acoustic emission source location in thin plates with one sensor. *International Journal of Pressure Vessels and Piping*, 81:427–431, 2004.

[JJ00] H. Jeong and Y.-S. Jang. Fracture source location in thin plates using the wavelet transform of dispersive waves. *IEEE Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 47(3):612–619, May 2000.

[KC76] C. H. Knapp and G. G. Carter. The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-24(4):320–327, August 1976.

[KD76] A. Kalnuins and C. L. Dym, editors. *Vibration: beams, plates, and shells.* Dowden, Hutchinson & Ross, 1976.

[KGM06] U. Klee, T. Gehrig, and J. McDonough. Kalman filters for time delay of arrival-based source localization. *EURASIP Journal On Applied Signal Processing*, 2006:1–15, 2006.

[KGR05] J. H. Kurz, C. U. Grosse, and H.-W. Reinhardt. Strategies for reliable automatic onset time picking of acoustic emissions and of ultrasound signals in concrete. *Ultrasonics*, 43:538–546, 2005.

[KHD+05] M. Kulesh, M. Holschneider, M. S. Diallo, Q. Xie, and F. Scherbaum. Modeling of wave dispersion using continuous wavelet transforms. *Pure Applied Geophysics*, 162:843–855, 2005.

[KK01] Y. Y. Kim and E.-H. Kim. Effectiveness of the continuous wavelet transform in the analysis of some dispersive elastic waves. *Journal of the Acoustical Society of America*, 110(1):86–94, July 2001.

[KLP03] R. Kjeldsen, A. Levas, and C. Pinhanez. Dynamically reconfigurable vision-based user interfaces. In *3rd International Conference on Vision Systems (ICVS'03)*, Graz, Austria, 2003.

[Kno] Knowles acoustic, http://www.knowlesacoustics.com/ (last visited 13 March 2007). *Accelerometer datasheet.*

[KV96] H. Krim and M. Viberg. Two decades of array signal processing research: the parametric approach. *Signal Processing Magazine, IEEE*, 13(4):67–94, July 1996.

[LA93] H.-J. Lin and D.P. Atherton. Investigation of IMM tracking algorithm for the maneuvering target tracking. *The First IEEE Regional Conference Proceedings on Aerospace Control Systems*, pages 113–117, May 1993.

[Leh04] E. A. Lehmann. *Particle filtering methods for acoustic source localisation and tracking.* PhD thesis, The Australian National University, July 2004.

[Leo02] C. K. Leo. Contact and free-gesture tracking for large interactive surfaces. Master's thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, May 2002.

[MA04] B. Mungamuru and P. Aarabi. Enhanced sound localization. *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 34(3):1526–1540, June 2004.

[Mac01] MacAddict. Reviews on MIMIO. http://www.macaddict.com, March 2001.

[McC95] K. G. McConnell. *Vibration testing: theory and practice.* Wiley, New York, 1995.

[Mea99] Measurement Specialties, Inc. http://www.msiusa.com, Norristown, US. *Piezo film sensors technical manual*, 1999.

[MO99] D. W. Marhefka and D. E. Orin. A compliant contact model with nonlinear damping for simulation of robotic systems. *IEEE Transactions on Systems, Man, and Cybernetics - part A: Systems and Humans*, 29(6):566–572, November 1999.

[Mor05] G.D. Morrison. A camera-based input device for large interactive displays. *Computer Graphics and Applications, IEEE*, 25(4):52–57, July-Aug. 2005.

[Mur] Murata Manufacturing Company, Ltd., http://www.murata-europe.com/. *Murata product datasheet (last visited 13 March 2007).*

[MW93] I. S. MacKenzie and C. Ware. Lag as a determinant of human performance in interactive systems. In *Proc. Conf. Human Factors in Computing Systems*, pages 448–493, 1993.

[MW01] 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*, 17(1):91–94, February 2001.

[MZK02] X. Ma, C. Zhou, and I.J. Kemp. Automated wavelet selection and thresholding for pd detection. *Electrical Insulation Magazine, IEEE*, 18(2):37–45, March/April 2002.

[NJQJ01] M. Niethammer, L. J. Jacobsa, J. Qu, and J. Jarzynski. Time-frequency representations of lamb waves. *Journal of the Acoustical Society of America*, 109(5):1841–1847, May 2001.

[OAW+98] H. Olsson, K. J. Aström, C. Canudas De Wit, M. Gävert, and P. Lichinsky. Friction models and friction compensation. *European Journal on Control*, 4(3), 1998.

[OH94] T. Önsay and A. G . Haddow. Wavelet transform analysis of transient wave propagation in a dispersive medium. *The Journal of the Acoustical Society of America*, 95(3):1441–1449, 1994.

[ot] NeHe openGL tutorials. http://nehe.gamedev.net/ (last visited 13 may 2007).

[OZ97] R. O'Hagan and A. Zelinsky. Finger track - a robust and real-time gesture interface. In *Proceedings of the 10th Australian Joint Conference on Artificial Intelligence: Advanced Topics in Artificial Intelligence*, pages 475–484. Springer-Verlag, London, UK, 1997.

[PAKJ+05] D. T. Pham, M. Al-Kutubi, Z. Ji, M. Yang, Z. Wang, and S. Catheline. Impact localisation techniques for tangible acoustic interfaces. In Duc Pham, editor, *Intelligent Production Machines And Systems*, pages 497–502. Elsevier, Oxford, 2005.

[PAKY+06] D.T. Pham, M. Al-Kutubi, M. Yang, Z. Wang, and Z. Ji. Pattern matching for tangible acoustic interfaces. In *Intelligent Production Machines And Systems*, pages 556–561. Elsevier, Oxford, 2006.

[PG92] W. H. Prosser and M. R. Gorman. Propagation of flexural mode AE signals in GR/EP composite plates. *Proceedings of the Fourth International Symposium on Acoustic Emission from Composite materials*, pages 418–427, 1992.

[PG94] W. H. Prosser and M. R. Gorman. Plate mode velocities in graphite/epoxy plates. *Journal of the Acoustical Society of America*, 96(2):902–907, June 1994.

[Pie81] A. Piersol. Time delay estimation using phase data. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(3):471–477, Jun 1981.

[Pin01] C. Pinhanez. The everywhere displays projector: A device to create ubiquitous graphical interfaces. In *Proceedings of Ubiquitous Computing*, volume 2201, page 315, Atlanta, Georgia, 2001.

[PJP+06] D.T. Pham, Z. Ji, O. Peyroutet, M. Yang, Z. Wang, and M. Al-Kutubi. Localisation of impacts on solid objects using the wavelet transform and maximum likelihood estimation. In *Intelligent Production Machines And Systems*, pages 541–547. Elsevier, Oxford, 2006.

[PN03] N. Promkajin and S. Noppanakeepong. An improvement to the adaptive kalman filter with the feedback of estimation error. In *Student*

Conference on Research and Development (SCOReD), Malaysia, 2003.

[PWJ+05] D.T. Pham, Z. Wang, Z. Ji, M. Yang, M. Al-Kutubi, and S. Catheline. Acoustic pattern registration for a new type of human-computer interface. In *Intelligent Production Machines And Systems*, pages 481–486. Elsevier, Oxford, 2005.

[Red76] D.R. Reddy. Speech recognition by machine: a review. *Proceedings of the IEEE*, 64(4):501–531, April 1976.

[RF03] D. Rocchesso and F. Fontana, editors. *The Sounding Object.* GNU Free Documentation License, http://www.soundobject.org/SObBook (last visited 9 December 2006), 2003.

[RK94] J. M. Rehg and T. Kanade. Digiteyes: Vision-based hand tracking for human-computer interaction. In *Proc. of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects*, pages 16–22, Austin, Texas, November 1994.

[RSTC06] D. Rovetta, A. Sarti, S. Tubaro, and G. Colombo. Modelling elastic wave propagation in thin plates. In *Intelligent Production Machines And Systems*, pages 548–555, Oxford, 2006. Elsevier.

[RYPD05] V. C. Raykar, B. Yegnanarayana, S. R. M. Prasanna, and R. Duraiswami. Speaker localization using excitation source information in speech. *IEEE Transactions On Speech And Audio Processing*, 13(5):751–761, 2005.

[SE99] R. Sleeman and T. V. Eck. Robust automatic p-phase picking: an online implementation in the analysis of broadband seismogram recordings. *Physics of the Earth and Planetary Interiors*, 113:265–275, 1999.

[SHP02] C. D. Saragiotis, L. J. Hadjileontiadis, and S. M. Panas. Pai-cs/k: A robust automatic seismic p phase arrival identification scheme. *IEEE Transactions on Geoscience and Remote Sensing*, 40(6):1395–1404, 2002.

[SHRP04] C. D. Saragiotis, L. J. Hadjileontiadis, L. T. Rekanos, and S. M. Panas. Automatic p phase picking using maximum kurtosis and κ-statistics criteria. *IEEE Geoscience and Remote Sensing Letters*, 1(3):147–151, 2004.

[SNS64] M. Skolnik, G. Nemhauser, and J. Sherman, III. Dynamic programming applied to unequally spaced arrays. *IEEE Transactions on Antennas and Propagation [legacy, pre - 1988]*, 12(1):35–43, Jan 1964.

[tdo81] Special issue on time delay estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-29(3):1981, 1981.

[Tho90] J.E. Thorner. Approaches to sonar beamforming. In *Southern Tier Technical Conference, 1990., Proceedings of the 1990 IEEE*, pages 69–78, 25 April 1990.

[Tou] Elo TouchSystems. http://www.elotouch.com (last visited 13 may 2007).

[Tuc01] B. J. Tucker. *Ultrasonic plate waves in wood-based composite panels.* PhD thesis, Washington State University, USA, August 2001.

[TVLK00] I Tech Virtual Laser Keyboard. http://www.virtual-laser-keyboard.com (last visited 9 december 2006), 2000.

[VB88] B.D. Van Veen and K.M. Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE ASSP Magazine*, pages 4–24, April 1988.

[Vib] Intelligent Vibrations. http://www.i-vibrations.com (last visited 13 may 2007).

[WB04] G. Welch and G. Bishop. An introduction to the kalman filter. *http://www.cs.unc.edu/ welch, gb, (last visited 14 March 2007)*, 2004.

[Wel93] P. Wellner. Interacting with paper on the digitaldesk. *Communications of the ACM*, 36(7):87–96, 1993.

[Wor02] PC World. Wireless whiteboards free up meetings – new generation of whiteboards takes notes, breaks barriers. *PC World magazine*, February 2002. MIMIO review.

[WTAT05] B. Wang, J. Takatsubo, Y. Akimune, and H. Tsuda. Development of a remote impact damage identification system. *Structural Control and Health Monitoring*, 12:301–314, 2005.

[WVDM96] A. Waibel, M. T. Vo, P. Duchnowski, and S. Manke. Multimodal interfaces. *Artificial Intelligence Review*, 10(3-4):299–319, 1996.

[WYZJ04] X. Wei, L. Yin, Z. Zhu, and Q. Ji. Avatar-mediated face tracking and lip reading for human computer interaction. In *Proceedings of the 12th Annual ACM International Conference on Multimedia*, pages 500–503, New York, USA, 2004. ACM Press.

[Yan06] M. Yang. Tai-chi project report. *http://www.taichi.cf.ac.uk*, 2004–2006.

[YI97] S. K. Yung and L. T. Ikellez. An example of seismic time picking by third-order bicoherence. *Geophysics*, 62(6):1947–1951, November–December 1997.

[YKA02] M.-H. Yang, D.J. Kriegman, and N. Ahuja. Detecting faces in images: a survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34 – 58, Jan. 2002.

[YTF03] S. Yon, M. Tanter, and M. Fink. Sound focusing in rooms. II. the spatio-temporal inverse filter. *The Journal of the Acoustical Society of America*, 114(6):3044–3052, December 2003.

[ZG91] S. M. Ziola and M. R. Gorman. Source location in thin plates using cross-correlation. *Journal of the Acoustical Society of America*, 90:245–251, 1991.