Interventional Ultrasound of the Breast
## Contents

List of contributors vii  
Foreword *Erika Denton* ix  
Preface xi  

1. A history of breast diagnosis  
   *Michael Michell*  
   1  
2. Anatomy of the breast  
   *Menos Lagopoulos*  
   11  
3. Technology for surgeons  
   *Mike Halliwell*  
   23  
4. Training and accreditation  
   *Sasiresh Govindarajulu and Simon Cawthorn*  
   35  
5. Ultrasound appearance of benign and malignant breast lesions  
   *Narasimhaiah Srinivasaiah and Anne Hubbard*  
   43  
6. Preoperative diagnosis and biopsy guns  
   *Karl J Sweeney and Michael J Kerin*  
   63  
7. Techniques of ultrasound biopsy  
   *Michael Michell*  
   79  
8. Techniques of ultrasound-guided vacuum-assisted biopsy  
   *Mike Shere*  
   89  
9. Future developments  
   *Obi Iwuchukwu and Philip Drew*  
   109  
10. Future developments: focused ultrasound ablation  
    *Obi Iwuchukwu and Philip Drew*  
    115  

Index 121
Contributors

Simon Cawthorn
Consultant Breast Surgeon
Frenchay Hospital
Bristol
UK

Philip Drew
Professor of Tissue Engineering and Repair
Hull York Medical School
University of Hull
Hull
UK

Sasirekha Govindarajulu
Staff Grade Surgeon
Breast Care Centre
Frenchay Hospital
Bristol
UK

Mike Halliwell
Physicist
Bristol General Hospital
Bristol
and
Medical Physics
Bristol General Hospital
Bristol
UK

Anne Hubbard
Consultant Radiologist
Castle Hill Hospital
Cottingham
UK

Obi Iwuchukwu
Consultant Oncoplastic and General Surgeon
Breast Unit
Department of Surgery
Sunderland Royal Hospital
Sunderland
UK

Michael J Kerin
Department of Surgery
Clinical Science Institute
University College Hospital
Galway
Ireland

Menos Lagopoulos
Senior Lecturer in Anatomy
Hull York Medical School
The University of Hull
Hull
UK

Michael Michell
Clinical Director
Breast Screening Training Centre
King’s College Hospital
London
UK
List of contributors

**Mike Shere**
Associate Specialist Breast Clinician
North Bristol Healthcare Trust
Frenchay Hospital
Bristol
UK

**Narasimhaiah Srinivasaiah**
Research Fellow
Academic Surgical Unit
University of Hull
Cottingham
UK

**Karl J Sweeney**
Consultant Surgeon Breastcheck
University College Hospital
Galway
Ireland
Foreword

It is a great pleasure to write a foreword to this new text on interventional ultrasound of the breast. Radiological examination of the breast is pivotal in the screening, diagnosis and management of both benign and malignant lesions of the breast. Progressively over the last decade interventional radiology techniques have replaced surgical and clinically guided diagnostic procedures to such an extent, that the vast majority of patients now have a non surgical diagnosis of a breast problem. Ultrasound guided diagnostic breast biopsy is now a fundamental component of the assessment of symptomatic and screen detected abnormalities. The equipment used for both ultrasound imaging and breast interventional procedures has developed with alarming speed and this text will bring the reader entirely up to date.

In recent years ultrasound interventional techniques have progressed to include therapeutic options, and the publication of this book is timely in addressing the practicalities of these procedures.

At a time when professional boundaries are changing and practitioners, other than those traditionally using ultrasound intervention, are delivering these services, this book comprehensively addresses the issues that will inevitably arise. The editors have brought together leaders in their individual fields to great effect.

Erika Denton MBBS MRCP FRCR
Consultant Radiologist
Medical Director, PACS Programme,
Connecting for Health, National
Clinical Lead for Diagnostic Imaging,
Department of Health
Norfolk and Norwich University Hospital
Norwich
UK
Preface

All patients referred to breast clinics should be examined by specialists who are trained to perform the ‘triple assessment’ of the patients breasts. The competencies to perform the three aspects of the triple assessment have been identified, and the Royal College of Radiologists have done the same for breast sonography and U/S guided biopsies. There is, therefore, no reason why the patient cannot have the triple assessment done by the same practitioner. The background of that individual can be either doctor, surgeon or radiologist, or a nurse or radiographer.

This book covers all you need to know about how to examine the breast, biopsy and remove lesions within it assisted by Ultra-sound. Becoming accredited in advanced breast U/S can then be achieved by attendance on a recognised course with assessment of competence once the training has been completed.

Simon Cawthorn
A history of breast diagnosis

Michael Michell

INTRODUCTION

Until the beginning of the 20th century, the diagnosis of breast cancer relied solely on clinical findings. Thus many cancers were large at the time of diagnosis with a high frequency of metastatic spread, the treatment was limited to surgery and the prognosis was often poor. However, the limitations of surgery alone in offering a cure were recognised – DH Agnew commented in 1883 ‘I do not doubt that cancer will one day be curable, but I do not believe that this will be procured through the surgeons scalpel’.1

The development of high quality mammography and ultrasound imaging of the breast during the 20th century have enabled both benign and malignant breast disease to be diagnosed with accuracy and often before clinical signs are apparent. The diagnosis of breast cancers when the tumours are small or preinvasive, together with improvements in treatment, has dramatically improved the prognosis from the disease.

X-RAY MAMMOGRAPHY: EARLY DEVELOPMENTS

In 1913, Dr A Salomon, a surgeon working at the University of Berlin, studied mastectomy specimens using X-ray and described the principal mammographic features of malignancy including microcalcification.2 He also recognised the potential value of mammography for demonstration of clinically occult breast cancer. In the late 1920s, Stafford Warren carried out the first series of in vivo mammograms using a technique adapted from chest radiography.3 He recognised the importance of high quality images and correctly diagnosed 54 of 58 malignant lesions in his first clinical results published in 1930.

Over the following 20–30 years, surgeons, radiologists and pathologists gradually improved knowledge of the relationship between pathological processes occurring in the breast and mammographic appearances. Important work was published by Leborgne in South America on the appearances of benign and malignant calcifications.4 In 1960, Gershon-Cohon and his pathology colleague Helen Ingleby published numerous studies demonstrating mammographic appearances of normal tissue and changes seen in both benign and malignant disease including the classic treatise ‘Comparative Anatomy, Pathology and Roentgenology of the Breast’.5–10
X-ray mammography was not, however, used widely in clinical practice before the 1960s because of variable image quality and a lack of enthusiasm from many radiologists and clinicians.

**MODERN MAMMOGRAPHY**

In 1960, Robert Egan, working at MD Anderson Hospital, University of Texas, reported on a low kV/high mAs technique which produced images with improved definition of soft tissue structures in the breast. Following publication of his experience of 1000 studies in 634 women, in which 228/240 malignant lesions were correctly reported, there was recognition by both clinicians and radiologists of the potential use of mammography in both the diagnosis of breast disease and the detection of breast cancer. The work of Egan was also crucial in demonstrating the importance of both specialised X-ray equipment including X-ray tube, film and processing, and radiographic technique in order to produce images with sufficient detail to demonstrate the mammographic features of breast pathology.

Following Egan’s work the principal technical requirements for the consistent production of high quality X-ray images of the breast were established. A low kV energy X-ray beam with a particular spectrum of X-ray photons is required because of the low inherent contrast of the soft tissues of the breast. The breast tissue needs to be compressed in order to reduce movement unsharpness and decrease scattered radiation. Dedicated highly sensitive film screen combinations and a small focal spot are required for the demonstration of the fine detail of structures in the normal breast and in disease. Finally, specialist training is required for technologists to ensure that correct positioning and radiographic technique is consistently used and for radiologists to ensure correct interpretation of the images.

The first modern dedicated mammography unit, the Senograph was developed by Companies General de Radiologic in France in collaboration with Dr Charles Gros and incorporated a molybdenum rather than a tungsten anode and a compression device independent of the X-ray tube. At the same time, film companies such as Dupont, Kodak and Xerox, realising the potential importance and likely future growth of X-ray mammography, invested in the development of dedicated highly sensitive film screen systems capable of recording the fine detail of breast structures and minimising the radiation dose. Although the principles of modern mammographic technique were established by the work of Egan and colleagues, numerous further technical improvements have been made to both the X-ray unit, film screen combinations and film processing over the past 40 years. These developments have resulted in the very high quality, high resolution mammograms that are now in widespread use in both diagnostic clinics and in population breast cancer screening programmes.

The use of such images together with the development of magnification and localised compression techniques using fine focal spot has enabled the most detailed correlation to be carried out between the mammographic features and the pathology
of benign, preinvasive and different types of invasive breast disease. Tabar and colleagues have linked imaging features, histopathology and natural history of breast tumours through the study of cases detected by population screening.

**DIGITAL MAMMOGRAPHY AND BEYOND**

Full field digital mammography has now become commercially available following the development of modern digital X-ray technology and detectors for breast imaging. Although the spatial resolution of digital mammography is not as high as for the best analogue systems, this is compensated for by the improved contrast resolution and other advantages of digital imaging. Recent large scale trials in Europe and North America have demonstrated that digital mammography achieves the same sensitivity for cancer detection as analogue systems and may be more effective for screening women in their forties and those with dense breasts. The next decade will see the replacement of analogue systems by digital equipment and the full integration of digital imaging with picture archiving and communication systems (PACS).

Digital mammography systems enable the development of new techniques for breast imaging which will further increase the diagnostic information obtained. Techniques which are currently undergoing evaluation include tomosynthesis, contrast enhanced mammography, duel energy subtraction contrast mammography, computed tomography (CT) mammography and computer aided detection. In tomosynthesis, a three-dimensional image of the breast is obtained by acquiring a number of low radiation projection images while the X-ray tube moves in an arc above the breast. The individual images are then reconstructed into a series of high resolution slices which are viewed subsequently. Lesions which are invisible on conventional imaging due to overlying dense breast tissue may be demonstrated using this technique.

Contrast enhanced mammography demonstrates malignant lesions due to the increased uptake of contrast related to tumour angiogenesis. Sequential imaging, dual energy and subtraction techniques allow dynamic time intensity curves to be produced.

Current experimental work is being undertaken to develop breast CT. With cone beam CT using flat panel detectors, three-dimensional images of the breast can be produced without the discomfort of breast compression and without unwanted exposure of the thoracic cavity. In developmental models, breast CT has been shown to detect objects 2 mm diameter with a mean glandular dose of less than conventional two view mammography.

**BREAST ULTRASOUND**

The first breast ultrasound examinations were carried out on palpable masses by Wild and Reid using A-mode apparatus in 1951 and demonstrated the potential for the differentiation of cystic from solid lesions. Further work on the clinical application of breast ultrasound was carried out in conjunction with technical developments
Interventional Ultrasound of the Breast

including B-mode scanning and use of dedicated automated water path scanners. The availability over the past 10 years of small hand-held transducers with a high frequency range of 7.5–12 mHz has enabled rapid real-time examination of the whole breast and axilla to be carried out. Improved resolution and image quality has allowed workers to visualise many non-palpable as well as palpable breast lesions and to define the criteria used to distinguish benign from malignant lesions. Early work was carried out in the 1970s by Kobayashi in Japan, Jellins and colleagues in Australia, and Cole-Beuglet and associates in the US. The most recent detailed analysis of the ultrasound features of both benign and malignant breast lesions has been carried out by Stavros and colleagues. Current and future developments in the use of breast ultrasound include evaluating the potential clinical application of advanced techniques such as Doppler and contrast examination.

MAGNETIC RESONANCE IMAGING

High quality images of the breast are obtained using dedicated surface coils. The demonstration of tumours depends on the increased signal produced following uptake of intravenous gadolinium-diethylene-triamine-pentaacetic acid (DTPA) in areas of tumour angiogenesis. Numerous studies over the past 20 years have shown that MR has a high sensitivity for breast cancer detection but a low specificity. Current studies in clinical practice are aimed at defining which patient groups will benefit from MR in initial tumour staging and treatment planning. MR will have a role in monitoring response to neoadjuvant chemotherapy, predicting response before changes in tumour volume have occurred, and in differentiating between residual tumour and fibrosis. MR will also be used for screening of young high risk patients, including BRCA-1 and BRCA-2 carriers.

EARLY DIAGNOSIS AND SCREENING FOR BREAST CANCER

The pioneers of X-ray mammography recognised that this technique had the potential to detect cancers before any clinical signs of disease were apparent but, until the technical breakthroughs of the 1960s, the images were not of sufficient quality for the technique to be applied systematically in population screening. The first trial to test the efficacy of screening for breast cancer was set up by Shapiro and colleagues in New York State in 1963 to determine whether periodic breast screening with mammography and clinical examination of the breast was effective in lowering mortality from breast cancer. In all, 62,000 women aged 40–64 years receiving their medical care from the Health Insurance Plan of New York State were randomised either to be invited to four annual screens or to form a control group. The trial ended in 1986 and showed that screen detected cancers were less likely to have involved lymph nodes and that there was a reduction of breast cancer deaths of 30% in the study group. Further large scale prospective randomised controlled trials were conducted.
in Europe to measure the effect of screening by mammography on breast cancer mortality.

In the Swedish two counties study, women aged 40–74 years were randomised to either invitation to mammography or a control group – there was no clinical examination as part of the screen. The combined results of the Swedish trials showed a risk reduction for death due to breast cancer of approximately 29% for women aged 50–69 years. In the UK, a working group chaired by Patrick Forrest examined all the available evidence on the benefits and costs of screening and recommended the provision of regular mammography screening for women aged 50 years and older. The National Health Service Breast Screening Programme (NHS BSP) began in 1989 and by 1995 a network of some 90 programmes had been established to provide 3-yearly screening by invitation to women aged 50–64 years using a single mediolateral oblique view of each breast. Subsequently, the screening technique has been improved with two view mammography used at each screen and the invitation range extended up to 70 years. Similar mammography screening programmes have now been established in other European countries, North America, Australia and New Zealand.

**ASSESSMENT AND DIAGNOSIS**

The widespread use of mammography in clinical practice together with the introduction of large population breast cancer screening programmes has led to the detection not only of many small cancers but also many ‘indeterminate’ lesions without specific mammographic signs of benign or malignant disease. A multidisciplinary approach using the triple diagnosis method, consisting of clinical examination, imaging work-up and needle biopsy, has evolved in order to achieve an accurate diagnosis and plan appropriate management. The importance of high quality special views including magnification mammography using a fine focal spot and localised compression techniques for the demonstration of fine detail mammographic features, particularly microcalcification and parenchymal distortion, has been emphasised by Tabar and colleagues in Sweden. This multidisciplinary approach has now been applied both to the assessment of women with screening detected lesions and to those presenting with symptomatic breast disease. Specialist diagnostic breast clinics run jointly by breast surgeons, clinicians, radiologists, radiographers and specialist breast care nurses have been established. This form of specialist clinic has been essential to meet the needs of increasing numbers of women presenting to the health service with minimal signs and symptoms of breast disease, encouraged by health promotion campaigns and publicity emphasising the importance of early diagnosis and treatment of cancer.

**IMAGE-GUIDED BREAST BIOPSY**

A histological diagnosis for non-palpable lesions detected by mammography was initially obtained by preoperative marking and surgical excision. Marking was first
Interventional Ultrasound of the Breast

carried out in the 1960s using a standard needle and was rapidly replaced by the development of techniques using specialist hook wires for better fixation.42

The disadvantages of diagnostic surgery for such lesions were unnecessary operations for women with benign lesions and the possibility of additional surgical procedures to obtain clear margins and stage the axilla for women with carcinoma. During the past 30 years, there has been steady development and improvement in image-guided needle sampling techniques which have made possible a definite preoperative diagnosis of malignancy for more than 90% of women with carcinoma, and has allowed most women with benign lesions to be spared surgery.

Image guidance

Although stereoscopic mammography was reported by Warren in 1930, it was not until the 1970s that stereotactic mammography was introduced into clinical practice, first for the placement of localising wires prior to surgery and subsequently for directing a sampling needle.43 The first stereotactic units were attached to existing mammography equipment and the procedure performed with the patient in the sitting position. Subsequently, the equipment has been developed so that the patient may be sitting, lying prone or in the decubitus position, the biopsy device may be introduced either in the line of the X-ray beam or perpendicular to it, and digital imaging has increased accuracy and reduced the procedure time. Using modern equipment, virtually all non-palpable lesions can be accessed for sampling.

Improvements in the quality and resolution of breast ultrasound mean that most soft tissue lesions and some florid microcalcification clusters are visible. For these lesions, ultrasound has become the method of choice for biopsy guidance – it is most comfortable for the patient, and the accuracy of the procedure is verified by real-time visualisation of the needle traversing the lesion.

With the increasing use of breast MR, particularly for screening of young women with a high risk of breast cancer, lesions which are initially only visible on MR are being detected. Some of these are visible on a ‘second look’ directed ultrasound examination and can therefore be sampled under ultrasound guidance. For lesions which remain only visible on MR, MR-guided biopsy systems have been developed.

Sampling

Stereotactic breast biopsy was initially carried out using fine needle aspiration cytology.44 Although in some centres very good results were obtained, many series reported problems with low sensitivity, high inadequate rates, equivocal results and occasional false positive results. Problems with cytology are related to the need for a specialist breast cytologist, the difficulty of obtaining cellular samples adequate for diagnosis from some benign and malignant lesions, and equivocal
interpretations resulting from cellular changes in some benign lesions and epithelial hyperplasia.

Stereotactic core biopsy, followed by ultrasound-guided core biopsy were introduced in the US by Parker and colleagues in the early 1990s and have since been incorporated into routine practice in breast clinics throughout the world. Image-guided core biopsy has been shown to be associated with higher sensitivity and specificity, and lower inadequate and equivocal rates compared with fine needle aspiration cytology. Tumour tissue obtained by core biopsy also provides information on tumour type and grade, presence of invasion and receptor status which are necessary for planning management.

For some types of lesion, however, multiple 14 G core samples may not provide a definite diagnosis. This may be because of difficulties in accurately targeting very small lesions only a few millimetres in diameter or may be related to the underlying pathology – this applies particularly to some clusters of indeterminate microcalcification where there may be a spectrum of different degrees of epithelial hyperplasia with or without atypia. In these cases, more tissue is required to enable the histopathologist to make a diagnosis. Vacuum-assisted core biopsy using 11 G or 8 G devices allows larger volumes of tissue to be obtained both because of the diameter of the device and the effect of the negative pressure which pulls tissue into the biopsy port. Use of these devices with X-ray, ultrasound or MR guidance for diagnostic procedures has become widespread over the past 5 years. The devices are now also being used under ultrasound guidance for the therapeutic removal of some benign breast lesions such as fibroadenoma or papilloma.

CONCLUSION

Radiologists working in the first half of the 20th century recognised the potential impact of imaging not only on the diagnosis and management of breast disease, but also, through screening, on the natural history and prognosis for women with breast cancer. Rapid developments in diagnostic techniques, and the implementation of large scale population screening programmes have led to the detection, diagnosis and treatment of breast cancers often when the disease is at an early stage before spread has occurred and the prognosis is very good. Improvements in image-guided biopsy techniques over the past 30 years mean that surgery is now very rarely required for diagnosis either of benign or malignant lesions.

Over the coming decades we can look forward to the full implementation of digital mammography and clinical evaluation of new techniques including contrast and CT mammography, improved sensitivity for cancer detection in young patients and high risk groups, and more accurate staging of disease both in the breast and in the axilla. We will also develop a better understanding of the natural history and appropriate management of ‘borderline’ and precancerous lesions diagnosed particularly by screening programmes, and further develop tests for monitoring the effectiveness of different treatments.
REFERENCES

Anatomy of the breast

Menos Lagopoulos

DEVELOPMENT

The breasts are modified and specialised sweat glands.

The mammary line (crest, ridge) is an ectodermal (epidermal) thickening that appears during the 4th–5th week of development. It extends from axilla to groin, on each side of the body (Figure 2.1). Only a small portion of the line persists in the thoracic region.

Invasion of the underlying mesenchyme (dermis) in the 6th week gives rise to the mammary buds. These lengthen, branch and are canalised to form the lactiferous ducts. The lactiferous ducts come together in a depression on the surface of the skin called the mammary pit. Shortly after birth the pit is converted to the nipple (Figure 2.2).

Persistence of remnants of the mammary line may give rise to accessory nipples (polythelia). They are found along the line of the mammary line and are commonly mistaken for moles. An extra breast develops if a remnant of the mammary line completely develops into a breast (polymastia) (Figure 2.3). Amastia is the congenital absence of the breast. It can be either unilateral or bilateral and is very rare. In amastia there is absence of breast tissue but the nipple is formed.1–5

ANATOMY

The adult (female) breast lies on the anterior thoracic wall. Its base extends from the 2nd to the 6th rib. It lies from the edge of the sternum to almost the mid-axillary line. Part of the superior lateral quadrant is sometimes extended towards the axilla. This is the axillary tail of the breast.

The superficial fascia splits to contain the breast. The deep layer of the superficial fascia overlies the chest muscles, separated from them by the retromammary space. The superficial (or subcutaneous) layer lies deep to the dermis. Cords of connective tissue connect the dermis to the ducts of the gland and to the deep layer of the superficial fascia – the suspensory ligaments of Astley Cooper. Contraction of these cords leads to indentation of the skin associated with some tumours (Figure 2.4).
Muscles

The breast lies over the muscles of the anterior thoracic wall. Also, there are muscles associated with the axillary region. Knowledge of these muscles and their blood and nerve supply is important to the surgeon.
The muscles of the anterior and lateral chest wall include pectoralis major and minor, serratus anterior, external oblique abdominis and rectus abdominis. The latissimus dorsi lies posteriorly. The subclavius lies under the clavicle whilst the deltoid and subscapularis are in the axillary region. The pectoralis major muscle is important

Figure 2.3  Congenital abnormalities associated with the breast. Reproduced from Netter, Volume 2. Reproductive System: The Netter Collection of Medical Illustrations, 1997 with permission from Saunders.

The muscles of the anterior and lateral chest wall include pectoralis major and minor, serratus anterior, external oblique abdominis and rectus abdominis. The latissimus dorsi lies posteriorly. The subclavius lies under the clavicle whilst the deltoid and subscapularis are in the axillary region. The pectoralis major muscle is important
Interventional Ultrasound of the Breast

in reconstructive breast surgery. The serratus anterior receives its nerve supply from the long thoracic nerve. The nerve can be damaged during dissection of the axillary lymph nodes (Figure 2.5).

Blood supply of the breast

The main vessels are the internal thoracic artery, the axillary artery and intercostal arteries. The lateral thoracic artery supplies the upper and lateral borders of the breast. The internal thoracic artery sends branches through the 1st to 4th intercostal spaces. The 2nd and 3rd branches are the largest. They supply the medial aspect of the breast. The posterior intercostal arteries also send small branches. There are variations in the distributions of these vessels.

The veins form a superficial plexus (around the nipple) and a deep plexus. From there, blood drains into deep veins that run with the arteries. It should be noted that the posterior intercostal veins can communicate with veins that drain the bony spine (Figure 2.6).

Nerve supply

The sensory supply of the breast is from branches of the 4th, 5th and 6th intercostal nerves. These nerves also carry afferent sympathetic fibres. The secretory activity of the breast is mainly controlled by the ovarian and pituitary hormones.4,8,9

HISTOLOGY

The tissue of the breast is composed of about 10–20 lobes separated by connective and adipose tissue. Each lobe opens independently at the nipple. A lobe is made of several
lobules. A lobule consists of clusters of milk secreting sacs, the alveoli. Myoepithelial cells lie around the alveoli. Their contraction helps the release of milk. When milk is produced it passes from the alveoli into a complex system of tubules and eventually reaches the intralobular duct. Outside the lobule the intralobular duct becomes the extralobular duct (Figure 2.7).

The lactiferous (mammary) duct drains each lobe. Near the nipple it dilates to form the lactiferous sinus. The ducts drain at the nipple, near the tip. The nipple is a raised pigmented area. The areola surrounds the nipple (Figures 2.8–2.10). Near the surface, the lactiferous ducts are lined with squamous stratified epithelium.
A gradual epithelial transition is seen, from stratified epithelium in the lactiferous ducts to a double layer of cuboidal cells in the lactiferous sinus to a single layer of columnar or cuboidal cells for the rest of the duct system. Changes in the epithelium of the duct system of the breast may give rise to breast cancer.\textsuperscript{9,10,12}

**LYMPHATIC DRAINAGE**

The quadrants of the breast, o’clock positions and codes are shown in Figure 2.11. The breast is divided into four quadrants. Upper inner (superior medial), upper outer...
In addition, the position of a lump in the breast can be described as the position on a clock (Table 2.1).

**The axilla**

The axilla is pyramidal in shape. It lies between the arm and the thorax, and communicates with the posterior triangle of the neck. It contains vessels, nerves and lymph
nodes. It has an apex and a base (floor), and four walls, anterior, posterior, medial and lateral. The axillary fascia forms the floor. The anterior wall consists of three muscles, pectoralis major and minor, and subclavius. The fascia extends between the clavicle and pectoralis minor muscle. It is pierced by lymphatics, the cephalic vein, the lateral pectoral nerve and branches of the thoraco-acromial axis (a branch of the axillary artery). The posterior wall is formed by the subscapularis and teres major muscles, and the tendon of latissimus dorsi. The medial wall is the chest wall with the upper portion of serratus anterior. The lateral wall is the humerus. It contains the axillary artery and vein, the brachial plexus and lymph nodes (Figures 2.12 and 2.13).

**The axillary lymph nodes**

There are usually 20–30 nodes scattered in the region. The anterior (pectoral) group lies in the medial wall of the axilla, anteriorly, close to the lateral thoracic artery at the

---

Figure 2.10 The lactiferous ducts opening at the nipple. Reproduced from Stevens and Lowe, Human Histology, 2005: 388–94 with permission from Elsevier.

Figure 2.11 The quadrants of the breast.
### Table 2.1 Staging codes for the breast

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C50.0</td>
<td>Nipple</td>
</tr>
<tr>
<td>C50.1</td>
<td>Central (subareolar) portion extending 1 cm around the areola</td>
</tr>
<tr>
<td>C50.2</td>
<td>Upper inner quadrant</td>
</tr>
<tr>
<td>C50.3</td>
<td>Lower inner quadrant</td>
</tr>
<tr>
<td>C50.4</td>
<td>Upper outer quadrant</td>
</tr>
<tr>
<td>C50.5</td>
<td>Lower outer quadrant</td>
</tr>
<tr>
<td>C50.6</td>
<td>Axillary tail</td>
</tr>
<tr>
<td>C50.8</td>
<td>In case there is a tumour overlapping two or more subsites, and the subsite of the origin of the tumour is not known</td>
</tr>
<tr>
<td>C50.9</td>
<td>In case there are multiple tumours in at least two quadrants of the breast</td>
</tr>
</tbody>
</table>

This information is used when describing breast pathology or operative reports, during physical examination and when interpreting a mammogram or ultrasound investigation.

![Intertubular groove of humerus](image)

**Figure 2.12** The boundaries of the axilla. Reproduced from Moore and Agur, Essential Clinical Anatomy, 2nd edn, 2002 with permission from Lippincott, Williams & Wilkins.

![The boundaries and contents of the axilla](image)

**Figure 2.13** The boundaries and contents of the axilla. Reproduced from Standring, Gray’s Anatomy the Anatomical Basis of Clinical Practice, 2005: 8400 with permission from Elsevier.
lower border of pectoralis major. The posterior (scapular, subscapular) group lies in
the medial wall, posteriorly, close to the subscapular artery. The lateral group is
found near the medial side of the axillary vein. The central group of nodes lies in the
fat in the middle of the axilla. These nodes are the most easily felt in the axilla. The
apical group lies at the apex of the axilla and receives from all the other groups. Lymph
from the apical nodes drains into the thoracic (on the left) or the right lymph trunk.

The axillary lymph nodes are also described according to their position in relation
to pectoralis minor muscle. Level I nodes are found lateral to the lower border of the
muscle. Level II nodes lie posterior to the muscle. Level III nodes are located medial
to the upper border of pectoralis minor muscle (Figure 2.14).

**Internal thoracic nodes**

There are also lymphatics lying on the medial aspect of the medial side of the breast.
These go with branches of the anterior intercostal vessels and eventually reach the
internal thoracic nodes, alongside the internal thoracic artery. They also receive affer-
ents from the liver and diaphragm, the rectus abdominis muscles and its sheath. The
internal thoracic nodes drain into the thoracic (on the left) or the right lymph trunk.

---

**Figure 2.14** The lymphatic drainage of the breast. PM, pectoralis minor; PMJ, pectoralis
major. Reproduced from Moore and Agur, Essential Clinical Anatomy, 2nd edn, 2002 with
permission from Lippincott, Williams & Wilkins.
Other groups of nodes described include the interpectoral nodes (Rotter’s), between pectoralis major and minor (Figure 2.14).

**Sentinel lymph node**

The sentinel lymph node is the very first node to receive drainage from a tumour of the breast.

**Lymph drainage**

Afferent lymphatics drain into large sinuses lying under the capsule of the node. These cortical (subcapsular) sinuses run towards the medulla of the gland. In the medulla, the network of interconnecting sinuses will get together at the hilum, where the large efferent lymphatic vessel arises. The afferent vessel can operate in the node in two ways. The node will receive the lymph through the afferent vessel, filter it and release it in the efferent vessel. Sometimes, however, the afferent vessel goes through the node with no filtration taking place. This explains why the first node receiving lymph from an organ is not always the one to be involved in metastases. About 75% of the drainage of the breast goes to the axillary nodes, mostly the anterior group. Direct drainage to the apical or central groups is also possible. It is believed that involvement of these nodes is a contraindication for surgery. Some vessels may also drain directly into the subscapular (posterior) group. The medial part of the breast usually drains to parasternal nodes and from there to nodes along the internal thoracic artery. Some lymphatics can also accompany the intercostal arteries to posterior intercostal nodes. Superficial lymphatics may communicate with the opposite breast and the anterior abdominal wall. Also, less frequently, there may be direct drainage to supraclavicular (deep cervical) nodes. Involvement of the supraclavicular nodes of the opposite side and of the supraclavicular nodes implies advanced disease.

**Metastatic spread**

Although the initial spread of breast cancer is through the lymphatic system, bony metastases occur by way of the bloodstream. Commonly affected are the lumbar vertebrae, the femur, the thoracic vertebrae, the ribs and skull. Metastases can also occur in the liver, lung and brain. The spread of cancer to regional lymph nodes and the spread to distant sites suggest two independent but also related events.4,8,10,13,14

**REFERENCES**

3

Technology for surgeons

Mike Halliwell

INTRODUCTION

Why a chapter on technology? After all ‘Ultrasound scanners are easy to use. You turn them on, put gel on the patient, put the probe on the patient and, hey presto, you have a picture.’ But then it gets a bit trickier. Sometimes there are unaccountable things in the image or else stuff, that you jolly well know is in the breast somewhere, is invisible. That is where this chapter is designed to help. In an entirely non-mathematical way it describes the very basic functions of the scanner, how ultrasound interacts with tissues to make images and how the operator can adjust the machine’s controls to give the most useful image, or to retrieve the situation when the image looks completely hopeless.

ULTRASOUND

Ultrasound is simply sound which is too high in pitch for humans to hear. Sound waves are a variation in pressure, they can be generated, for example, by plucking a guitar string. You can almost see the string vibrating and it is these vibrations which are transferred to the surrounding air as pressure waves, alternately squashing and releasing the air. These waves of pressure travel away from the string. If they reach our eardrum then the pressure changes cause it to move backwards and forwards and we hear sound. Vibrations at more than about 20 000 a second (20 kHz) are inaudible to humans and thus named ‘ultrasound’.

For breast imaging it is usual to use vibrations at frequencies of 5–15 million cycles a second (5–15 MHz). (The term ‘frequency’ is just a measure of how many vibrations per second are associated with the sound wave, it is not related to the energy or intensity of the sound wave. A high frequency wave is not necessarily more powerful than a low frequency one. The difference is that it gives a better quality image but of a smaller depth of tissue. To have more energy (to be more powerful or have greater intensity) it would have to have been created by a larger voltage.)

There are just three particular properties of ultrasound which make it useful in medical imaging.
1. It travels in a narrow beam;
2. It travels at constant speed;
3. It is reflected at boundaries.

These properties mean that it is possible, from the skin surface of a patient, to tell the position, depth and type of objects within the body.

The simplest kind of ultrasound device contains a single transducer which when struck (excited) by a voltage spike produces a click (pulse) of sound waves (Figure 3.1). The click travels away from the transducer, like light from a torch, in a well defined cylindrical beam. When it meets a boundary between two different tissues part of the click is reflected and the rest travels on. This is analogous to light being reflected from a pane of glass, a small part of the light is reflected but the bulk passes through the glass to illuminate objects beyond it. The part of the ultrasound click which is reflected, the echo, returns to the transducer where it is converted into a voltage and displayed on a computer screen.

The display screen shows a horizontal line with a blip at the left hand side corresponding to the transducer and others from the targets, somewhere to the right. The distance between the blips shows how far away is the target. The depth of the target is found by measuring the time delay between launching the pulse and receiving the echo. The depth is then given by multiplying the time delay by the speed of sound, and dividing by two to account for the ultrasound round trip being to the target and back again to the probe (Figure 3.2). (Thankfully the operator does not have to do any of this calculation, the machine does it all.)

This A-scan (amplitude scan) does not show the location of the target in the patient as there is no information available about the direction of the beam (apart from the operator squinting down the side of the transducer holder). This is an example of the basic pulse–echo technique, used by bats and dolphins. However, these animals are able to add the directions of their targets to their ranges because their ears are used to provide stereo (directional) reception. In medical applications this directional knowledge is achieved by steering the beam from the probe electronically in known directions.

![Figure 3.1](image.png) The ultrasound pulse. The pulse of ultrasound is about 2–5 cycles in duration, about 1 mm long, about 3 mm wide and occupies a three-dimensional onion shaped volume.
For breast scanning the best kind of probe is a flat faced one, the kind known as ‘a linear array’. When pushed against the skin surface, the flat face of the probe can be used to squash the breast against the chest wall. This helps to straighten out boundaries between the fat and glandular components. Flattening these boundaries reduces image distortion (caused by refraction at curved surfaces). The flattening process also improves the visibility of rigid lesions as boundaries in their vicinity do not flatten out but remain curved around them and easily visible on the image.

Linear arrays consist of a flat faced collection of tiny transducer elements arranged side-by-side (an array), like the teeth of a comb. There are typically between 100 and 200 elements in the probe housing. Each ultrasound beam is made by exciting a group of 10–15 of these at a time. The location of the beam emerging from the transducer face depends on which group is excited, this is controlled by scanner electronics (the beamformer). For the most basic scan, the first beam is produced from element group 1 to 10, the second beam from 2 to 11, the third from 3 to 12, the fourth from 4 to 13 and so on. For a probe with 100 elements this would result in approximately 100 beams in the image (actually 92 because the first and last four would not be the centre elements of the beam forming groups, in this simplified example) (Figure 3.3).

The display screen is arranged to show lines in directions and positions corresponding with each ultrasound beam. Usually there is a horizontal line drawn across the top of the display corresponding with the face of the probe. Lines corresponding to each ultrasound beam are then drawn vertically down the screen from the appropriate part of this probe face line. These beam lines are invisible until echoes from the pulse are received. Each echo is displayed as a bright dot on the beam line at the appropriate depth from the probe face. As the beams sweep across the face of the probe, the echo dots on the display draw out the outlines and texture of the tissues beneath the probe.
Interventional Ultrasound of the Breast

It is possible to sweep the beams across the face of the probe in about one 20th of a second, consequently the display is refreshed 20 times a second (the frame rate is 20 frames per second, fps). The image appears in ‘real time’, any movement of the tissue – cardiac, respiratory or probe induced – are shown as they happen. Of significance in breast imaging is that the placement of biopsy needles under ultrasound control is visualised ‘live’. Accurate needle placement is relatively easy to achieve. As far as the operator is concerned the beam sweeping occurs automatically. As soon as the scanner is turned on the beam starts sweeping to and fro.

Echo size

The image on the screen is the result of reflections of the pulse of ultrasound. Bright regions tend to correspond to fibrous or calcific areas which are denser than their surroundings, and blank regions (no echoes) correspond with liquid volumes such as cysts which contain no reflectors. The brightness of a particular part of the image cannot be rigorously related to the tissues there but general trends can be identified (Figure 3.4).

Reflections

The reason behind this lack of direct correspondence between echo size and tissue or target type is to do with the way in which ultrasound is reflected. As well as being dependent on the relative acoustic densities (acoustic impedance) at tissue boundaries the echo strength also depends on the size and shape of the boundary. The two extreme shapes encountered in medical scanning are the large flat plane (e.g. pectoralis major boundary) and the small point target (e.g. microcalcification).

A large flat plane acts like a mirror. The echo is best seen if the beam strikes the surface of the plane perpendicularly. At an oblique angle the returning echo will be reflected away from the probe and seen only faintly if at all.
A tiny target acts like a finger wiggled in bath water – the reflections come away in all directions. The size of the echo is spread wide (over the surface of a sphere) so that the portion that gets back to the transducer is quite weak but it will be picked up irrespective of the orientation of the target.

In real tissues most targets are intermediate in size and shape, and display both directivity (flat surfaces) and universality (point targets). For example Cooper’s ligaments are best seen when perpendicular to the beam but it is possible to follow them on the image as they curve away. The echo brightness lessens as the curvature increases (less of the main echo is directed back to the probe) but continuity with the brightest parts is maintained (Figure 3.5).

**Beam size**

Another factor relating echo brightness and target type is the size of the beam. It is helpful to realise that the ultrasound pulse occupies a volume in space – it is a bit like an onion in shape, a petit pois in size (Figure 3.1). As the pulse travels through the tissues it will generate echoes from every target within its volume. The smaller the volume the more precise will be the echo patterns as a representation of what is really there. If the ‘onion’ is large then fine detail will be lost and the result will be an average of target types within that volume. For example, if there is a 3 mm diameter cyst within the breast and the ultrasound onion is 6 mm across, then the ‘no echoes’ from the content of the cyst will have added to them the ‘tissue echoes’ from the 1.5 mm of stuff on either side of the cyst too. On the image these echoes will overlay the cyst and the result will be a slightly less bright patch but not a completely echo free region. This is one of the reasons for focusing the beam (Figure 3.6).
Focusing reduces the beam width, and improves the volume of the onion, and hence the ability to find small regions of different reflectivity. This is fairly easy to see in the case of a small cyst but it also applies to small solid lesions.

**Figure 3.5** Types of reflection from typical breast tissues.

**Figure 3.6** The effect of beam size on contrast resolution. (a) Very thin beam – good focus (thin slice). Only echoes from the mass are picked up (here the mass is a cyst, no internal echoes). (b) Very wide beam – poor focus (thick slice). Echoes from the mass and from the tissues around mass are picked up (display shows both at once).

**Focusing**

Focusing reduces the beam width, and improves the volume of the onion, and hence the ability to find small regions of different reflectivity. This is fairly easy to see in the case of a small cyst but it also applies to small solid lesions.
Resolution

Another reason to reduce the beam dimensions by focusing is embodied in the concept of lateral resolution. If there are two small targets at the same depth and side by side, then they will only be seen as two targets if the onion is narrower than their separation. If they sit inside the onion, then they will be drawn on the image as a single extended object. This problem is most commonly related to the example given above, i.e. to resolving two closely spaced objects, but the lack of clarity caused by the volume of the onion also applies to single targets. Any single target will be drawn as an elongated line on the display as the beam sweeps from side to side. The first indication of the presence of the target will be a dim dot as the edge of the beam just touches it. The target will reappear on each subsequent scan line until the beam has been moved to the position where the opposite edge just misses it. The length of the line denoting the target corresponds with the width of the beam. Focusing reduces the beam width and reduces the length of the target line (Figure 3.7).

Depth compensation

Echo brightness also depends on the weakening of the pulse as it travels through tissue. There are two mechanisms leading to a reduction in pulse size: energy is lost from the pulse as heat; and energy is lost as reflections and scattering occur. The significance of this is that identical targets at different depths will result in different sized echoes at the probe. This phenomenon is experienced in everyday life too, if you shout at your teenage son from a distance of 100 metres he will ignore you because your pressure waves have been reduced by attenuation and he can’t hear you. If you shout from a distance of 1 metre he will certainly be able to hear you because your pressure waves have not been significantly attenuated, but he will still ignore you because he is a teenage son.

In medical imaging it is possible to compensate for this effect (attenuation). The compensation can only be done in the scanner after the echoes have been received.
The method of compensation is a process of increasing the amplification applied to echoes according to their depth of origin. Deep echoes will be amplified much more than shallow ones. The process has a variety of names: depth gain compensation, time gain compensation and swept gain are frequently used. The operator has some control over this (Figure 3.8).

**Operator controls**

The components of an ultrasound scanner and operator controls are shown in Figure 3.9.

**Focus**

Focusing the beam improves image clarity (spatial resolution) and lesion conspicuity (contrast resolution), and is under operator control. For best image quality as many focal points as possible should be selected; unfortunately the trade off is that the frame rate is reduced with each new focal point. The frame rate reflects the rate at which each complete image is produced. If an image takes one 100th of a second to complete
then there will be 100 frames per second, if it takes half a second the frame rate is two frames per second. Often, if all possible focal points are selected, the frame rate is so low that the image becomes very jerky. In this case the operator has to move the probe very slowly over the patient to obtain readable images. An acceptable compromise is often reached with four or five focal points selected. The depth of these foci should be such that they straddle the region of interest. (A useful tip is that if the operator has only one focus to play with then it is best to locate it slightly deeper than the region of interest, the beam shape is such that it gets worse more quickly deep to the focus than it does before the focus.)

**Depth gain compensation**

Most machines have a built-in average compensation setting so that the initial scans should be nearly correct. The operator can ‘fine tune’ echo levels on the display to achieve a balanced result using controls on the scanner’s front panel. These controls are usually in the form of about five sliders. Each slider controls echo levels over about one-fifth of the depth on the display. Judicious use of the sliders can improve an irregularly attenuated image; however, injudicious use can completely ruin it.

**Overall gain**

This control changes the overall brightness of the display by changing amplification equally to all echoes. The operator uses this, frequently, to adjust the brightness of the mid-level echoes to be mid-grey before fine tuning with the depth compensation sliders.

**Freeze**

The freeze control stops the scan process and causes the last image to be displayed as a still image on the screen. This image can be measured, annotated, printed, recorded or stored. Frequently scanners have a cine loop facility. This feature allows the last few seconds of scans to be replayed as either a real-time cine loop or a sequence of still images. This is useful in breast scanning when often the best scan is produced just before the operator freezes the image. (The action replay brings back the best scan which can then be measured.)

**Depth**

This changes the size of the image according to the maximum range that the operator needs to be visible. For large patients it may be necessary to have a depth of 8 cm, smaller individuals may only need 2 cm. In general the minimum depth necessary to visualise the region of interest should be employed.
Zoom

Zoom makes it possible to magnify portions of the scan. This is particularly useful in large patients where the depth is necessarily set to a high value making lesions appear to be relatively small on the screen. A box cursor is used, positioned and sized using the scanner’s trackball mouse, to define the region of interest around the lesion. The zoom control then magnifies this part of the scan. The zoomed scan can be frozen and then processed in the usual ways.

The use of ultrasound in needle guidance

Ultrasound is ideal for needle guidance.

1. It is real time – the needle can be visualised as it moves through the tissues;
2. The probe is hand held – the location of the needle can be checked frequently in relation to anatomical landmarks and the target lesion;
3. Ultrasound is strongly reflected from metallic objects – the needle is easily identifiable.

Later chapters will deal with the practical clinical techniques involved; here it may be helpful to touch on a few of the more basic principles of needle guidance.

How is the needle visualised?

Type of boundary

The needle is metallic. This means that its acoustic properties are significantly different from those of soft tissue, an ultrasound pulse will be strongly reflected at its surface.

Shape of boundary

The needle is flat (as far as ultrasound is concerned) in one direction and cylindrical in the other direction. Ultrasound will undergo mirror-like reflection in the flat direction and scattering in the cylindrical direction. In practice this means that the needle is seen best if the shaft is parallel with the probe face (perpendicular to the beams). Because of its cylindrical nature the needle will be equally well visualised even if it is rotated, around its main axis, provided the parallel orientation is maintained (Figure 3.10).

The scan volume

The ultrasound beam is a three-dimensional entity; when scanned along the face of the probe it traces out a three-dimensional volume, the scan or slice volume. If the
needle lies within this volume it will be seen in its entirety; if it lies obliquely across the volume it will be seen as a cigar shaped object. It is easy to mistake this for the whole of the needle, the error results in the tip of the needle not being visualised. If this happens then sampling can occur from tissues other than the intended target. It is important to check, by rocking the probe to and fro, that the tip is clearly seen. A few moments practice with the needle in a shallow beaker of water helps elucidate this particular problem.

Another problem occurs if the slice thickness happens to be wider than the lesion. In this case it is possible for the lesion and the needle to be visualised simultaneously but with the needle lying alongside rather than inside the lesion (Figure 3.11).

**Figure 3.10** Effect of angle on the visualisation of a biopsy needle.

**Figure 3.11** The importance of containing the needle within the scan plane.
CONCLUSION

What kind of machine should you buy? Almost any machine on the market will allow real-time needle guidance. But you do get what you pay for and it is helpful to seek advice from other users before deciding. Do not forget that it is more complex than a television and applications support from the manufacturer is almost essential, certainly in the learning period with a new machine.

FURTHER READING

The incidence of breast cancer continues to increase worldwide and it is the most common cancer in women in the UK. The rise in the number of breast cancer patients has increased breast awareness, and the heightened awareness has resulted in an increasing demand for both diagnostic and interventional breast ultrasound procedures. Traditionally ultrasound was performed only by radiologists but there has been a shortage of radiologists, and inability to cope with the demand to reduce waiting times and provide a more efficient service for all patients. Central to this demand is the need to have the highest quality imaging of the breast. With the multidisciplinary approach adopted in most breast units the training in breast ultrasound can be aimed at breast surgeons, clinicians, oncologists, nurse practitioners and health professionals, and tailored to their individual needs.

Breast ultrasound is now seen as an essential part of modern clinical examination of the breast, both in the imaging and in needle biopsy. In one-stop clinics there is a need for a clinical diagnosis at a ‘one-stop’ visit. Ultrasound allows almost all impalpable cancers to be biopsied and improves the accuracy of biopsying palpable cancers. With ultrasound it is possible to follow the movement of the biopsy needle as it takes place.

Sonography of the breast, like many other aspects of ultrasound, remains a technically challenging procedure, as breast tissue is heterogeneous. It is essential that the sonographer possesses skills in the technique of performing and interpreting the scans. Acceptable levels of performance can only be achieved by ‘hands on experience’ in routine practice.

In order to improve practice and deliver high standards of care, training in ultrasound is essential. With appropriate training this can be delivered by a non-radiologist with the appropriate core skills training and, of course, assessment and accreditation when these are successfully acquired. However, the standard of training for the non-radiologist performing the ultrasound would be expected to be the same as that for the radiologist.

**TRAINING REQUIREMENTS**

It is well known that ultrasound diagnosis is highly operator dependent and there is a potential for diagnostic error. The European Federation of Societies for Ultrasound
in Medicine and Biology (EFSUMB)\(^1\) has proposed and published minimal training requirements for the practice of medical ultrasound in Europe in their web site. These are supported by the Royal College of Radiologists and the British Medical Ultrasound Society (BMUS). These organisations recommend theoretical knowledge (core knowledge) and practical skills.

The theoretical training of breast ultrasound starts with understanding of the physics of ultrasound and knowledge of the instrumentation, which enables the practitioner to apply the knowledge of the physical principles of ultrasound and the technical processes of the ultrasound scanner to obtain diagnostically useful images and avoid misdiagnosis.

The WHO Study Group on training in diagnostic ultrasonography\(^2\) developed a curriculum for basic physics and instrumentation. The curriculum stresses that the following should be mastered to achieve maximum benefit of the ultrasound technology:

- the basic physics of ultrasound, its interaction with tissues and its bioeffects
- the knowledge of transducer construction
- the effects of frequency of sound waves on the depth of penetration and the image quality
- methods of focusing, issues on resolution and trade-off between the resolution and tissue penetration
- artefacts and their effects on clinical practice
- safety and limitations in use of ultrasound
- ultrasound instrumentation with emphasis on real time and Doppler
- colour flow and power Doppler
- use of the ultrasound controls with emphasis on operator controlled variables which would enhance image quality
- image recording
- knowledge of statistics and computer science.

Understanding the anatomy of the breast is essential. Ultrasound scanning should be comprehensive. The scan field should include all the anatomical layers of the breast. To be able to understand the anatomy in relation to ultrasound a basic knowledge of anatomy of the normal breast (both male and female), and the physiological changes with age (puberty, pregnancy, lactation and menopause) is essential. Knowledge of changes in the breast due to hormone replacement and other medications is also necessary. To understand the ultrasound appearances a comprehensive knowledge of various pathologies of the breast (benign, indeterminate and malignant) is essential.

The WHO document\(^2\) stresses that even established, well advanced teaching centres should be encouraged to invite external lecturers to ensure fostering of new ideas and to constantly improve the quality of the curriculum.

Practical training with hands on experience is essential after acquiring the required core knowledge by attending the theoretical course. Training should include the different techniques of performing a systematic examination of the breast and the advantages
and disadvantages of each. Ultrasound can be targeted or focused on the symptomatic lesion but this limits the use of ultrasound and its applications, and fails to exploit the advances in technology of modern scanners. There is also a potential to miss lesions that are impalpable. Guidelines for three levels of minimum training requirements have been proposed by the Royal College of Radiologists (Table 4.1).3

The training should be undertaken in a multidisciplinary environment where there is an adequate input from all specialties. Diagnostic and interventional ultrasound procedures should be correlated with clinical findings, other imaging and histopathological results including surgical resections.

The Royal College of Radiologists3 recommends that a competency assessment form should be completed and signed by the training supervisor. Regular appraisal should take place during the training period. During the training period a log book should be maintained. The log book is a formal record of the experience and reflections on undertaking the required number of ultrasound scans.

### Table 4.1 Minimum training requirement for Levels 1–3 proposed by the Royal College of Radiologists

<table>
<thead>
<tr>
<th>Level</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level 1</td>
<td>Perform ultrasound of breast and axilla safely and accurately. Recognise normal anatomy. Identify the pathology (benign, intermediate and malignant). Understand the importance and indications of ultrasound in the context of triple assessment. Correlate ultrasound findings with clinical, imaging and pathological findings. Give a detailed report and grading of ultrasound finding. Offer differential diagnoses and recommend further management. One ultrasound session a week for a period of 6 months to a year and a minimum of 100 scans should be undertaken. A log of 50 cases should be kept.</td>
</tr>
<tr>
<td>Level 2</td>
<td>Accept and manage referrals from Level 1. Be able to identify and diagnose almost all breast pathology. Be able to perform all ultrasound-guided interventional procedures. Initial observation, later performance under supervision and when competence achieved perform independently with support readily available. Should conduct research in ultrasound. One ultrasound session a week with a minimum of ten scans per session for at least 3 months.</td>
</tr>
<tr>
<td>Level 3</td>
<td>Accept referrals from Levels 1 and 2. Able to identify and diagnose all breast pathology. Be familiar with ultrasound-guided vacuum-assisted biopsy. Able to teach and supervise Level 1 and 2 practitioners. Should conduct research in ultrasound. Be aware of and pursue the developments in ultrasound. One ultrasound session a week with a minimum of 100 ultrasound scans per year.</td>
</tr>
</tbody>
</table>

ASSESSMENT AND ACCREDITATION OF THE INDIVIDUAL

On completion of theoretical and practical training, training programmes should have sufficiently rigorous mechanisms in place to evaluate theoretical knowledge and practical scanning and interpretive skills.

Log books should be assessed as pass or fail based on the range of conditions, description of ultrasound finding with grading, interventional procedures undertaken,
results of the interventional procedures, attendance to multidisciplinary meeting, reference to literature and correlation with clinical, other radiological and histological findings. Reflections in the log book should also be evaluated.

One examination that trainees must take covers the physical principles of ultrasound and the instrumentation that makes ultrasound equipment work. This material is technically complex, and it gets more sophisticated every year, as scientists discover new ways of using ultrasound to get better images. Of the marks 30–40% should be allocated to a written examination with short answers/multiple choice questions and the remaining 60–70% of marks should be allocated to Objective Structured Clinical Examination (OSCE).

In the OSCE the trainee should be able to demonstrate the ability to use the operator controlled variables in the machine to obtain good quality images and also be able to describe and grade ultrasound findings on hard copy pictures. The trainee should be able to identify the hidden lesion (olive), and demonstrate the ability to perform a core biopsy or any interventional technique. Strategies relating to a failed assessment should be clearly documented.

Sonographers or other health-care professionals without a medical degree who wish to train in breast ultrasound should undergo comprehensive training programmes for sonographers and these programmes require high standards of core knowledge and practical scanning skills. These training programmes should be strictly regulated with established schemes for assessment and accreditation of the trainees.

A minimum amount of breast ultrasound scanning is to be performed to maintain ongoing experience after achieving the required training. It is the responsibility of the person performing the scan to ensure that he/she maintains practical skills by hands on experience in regular ultrasound sessions. It is also essential to scan a range of pathology and practice in a multidisciplinary environment.

To achieve accuracy and confidence the International Breast Ultrasound School (IBUS) recommends performance of a minimum of 500 examinations in a multidisciplinary environment with at least 300 cytology or histology correlation cases, and performance of at least 50 interventional procedures with appropriate follow-up.

**Continuing medical education and professional development**

The practitioner should maintain skills by performing a minimum amount of ultrasound scans. To attain high standards of practice, audit of one’s own practice is encouraged and is of educational value. One should keep up to date with recent advances in ultrasound and attempt to participate in research. The individual should include elements of ultrasound in the continuing medical education or continuing professional development. There are potential advantages of using distant learning and multimedia approaches in achieving this.

Teachers should be highly qualified to teach the practical skills of ultrasound and should regularly attend ultrasound courses and conferences to ensure that they are aware
and pursue the developments of this fast growing field. To deliver the above curricu-

lum, there is a need for physicists/engineers who not only must possess the required
professional background, but who are also able to teach a non-engineering audience.

WHO\textsuperscript{2} recommends intensive training programmes for the teachers at major
training centres. Jefferson Research and Education Institute, Philadelphia, Pennsylvania
is one institution which offers such programmes. The teacher receives an update
of the clinical and technical knowledge, and also instruction in the methods and
techniques of teaching.

\section*{Teaching aids}

ATS model 551 (ATS Laboratories, Bridgeport, CT, USA) is a tissue-mimicking
phantom used to monitor performance changes which may occur during normal
operation of an ultrasound imaging system and can also be used in training for verti-
cal and horizontal measurement calibration, to understand the focal zone, axial and
lateral resolution.

Model BB-1 the ATS Breast Biopsy phantom mimics the appearance, touch and
acoustic properties of the human breast. Combination of 5 and 10 mm diameter, cystic
and solid-like tissue-mimicking target structures have been embedded randomly
throughout the phantom. This can be used for practising the biopsy techniques.
Home made chicken breast/turkey breast with olives or fruit can also be used and are
cheap.

\section*{INSTRUMENTATION, QUALITY ASSURANCE AND
PATIENT SAFETY}

The IBUS\textsuperscript{4} and the American Institute of Ultrasound in Medicine (AIUM)\textsuperscript{5} have
published guidelines for equipment requirements for performing breast ultrasound.
Breast ultrasound should be performed with a high quality high resolution near field
real-time and linear array scanner operating at a centre frequency of 7 MHz. Equipment
permitting electronic adjustment of focal zone is recommended. Adequate penetra-
tion to the depth of the breast to display all anatomical layers of the breast is essential.
A penetration depth of at least 4 cm is required and a field of view of greater than 4 cm
is preferable in larger breasts. Coupling devices for the transducer such as stand off
pads are recommended for evaluation of superficial lesions.

Ultrasound scanning should be reproducible and any ultrasonographic finding
should be identified clearly on the image. Knowledge of methods of annotation
(laterality, quadrant of the breast, position on the clock face, distance from the nipple
and the depth of the lesion) of the ultrasonic finding is essential. Image recording,
storing and filing should also be an integral part of the training. Knowledge of the
different methods of storage of still pictures, cine loops, video loops and retrieval of
the stored images and clips would facilitate education, teaching and research.
Like any high-tech equipment, an ultrasound machine needs to be calibrated regularly against some recognized benchmark, like a standardized plastic ‘phantom’ that mimics the acoustic properties of human tissue. Calibrated test structures are embedded within the tissue-mimicking environment. Phantoms are essential to detect performance changes that occur through normal ageing and deterioration of system components. This ensures that the ultrasound images from each procedure will give the most accurate and complete clinical information. The trainee should be aware of the quality assurance guidelines.

Formal training programmes should include appropriate teaching material on the safe use of ultrasound, the potential for bioeffects and the rationale and means for limiting output. The AIUM has approved the following statements regarding diagnostic ultrasound.

1. ‘Based on the epidemiologic evidence to date and on current knowledge of interactive mechanisms, there is insufficient justification to warrant a conclusion that there is a causal relationship between diagnostic ultrasound and adverse effects.’6
2. ‘Although there are no confirmed biological effects on patients caused by exposures from present diagnostic ultrasound instruments, the possibility exists that such biological effects may be identified in the future. Thus, ultrasound should be used in a prudent manner to provide medical benefit to the patient.’7
3. ‘Ultrasound monitors must display the Mechanical Index (MI) for B mode scans and Thermal Index (TI) for M mode and Doppler.’8 The trainer should make sure the trainee is aware of the displayed safety indices, their meaning, and their function in the management of safety.
4. It has been recommended that users follow the principles of ALARA (as low as reasonably achievable).

**Breast imaging pearls**

Training programmes should include appropriate teaching material on how to safeguard patients.

1. All personnel performing ultrasound scans must strictly adhere to professional ethics and behaviour ensuring patient confidentiality.
2. Always be respectful of patients, their families and medical colleagues.
3. In the event of an accident or occurrence of a complication a policy or a procedure must exist for responding and reporting.
4. At the end of each examination, transducers used on the skin surface should be cleaned with wipes as directed by the manufacturer (see operating manual). Additional cleansing may be necessary as per guidelines of the unit or the manufacturer in cases of blood or contamination with tissue fluid.
It is possible that the practice of ultrasound by non-radiologists will increase and will be introduced into more breast units over the next few years which will lead to more training centres. It is therefore important that consideration be given not only to the setting up of suitable training programmes but also given to accreditation of the providers of such training programmes. This accreditation should encourage such providers of training programmes to meet and exceed nationally recognised standards. Through accreditation of the training programmes the providers have an opportunity to audit their training methods and initiate changes to improve the standards of training. Seeking of such accreditation proves the providers commitment to the highest quality patient care.

Ultrasonography may end up as a victim of its own success. The use of ultrasound has increased due to the fact that it is one of the fastest growing diagnostic imaging modalities, and magnified by the ease of availability of smaller, cheaper more advanced scanners. To date, no harmful effects of ultrasound have been reported. The absence of ionising radiation, the inexpensive and non-invasive nature of the investigation could lead to more and more users. It is therefore important that the training programmes should be of high standard in order to ensure that the breast sonographers are properly trained for their job.

REFERENCES

5

Ultrasound appearance of benign and malignant breast lesions

Narasimhaiah Srinivaiah and Anne Hubbard

INTRODUCTION

Ultrasound is the examination of choice to evaluate any palpable breast mass, being quick, painless and cheap, and having none of the risks of ionising radiation that limit the use of mammography in younger women. Ultrasound should be used to guide needle sampling of all palpable lesions, as it provides a lower rate of inadequate sample collection than clinically guided biopsy and has a much smaller risk of pneumothorax.

The sensitivity and specificity of ultrasound vary enormously depending on the experience of the operator and the equipment used. Ultrasound is not infallible and 1–2% of palpable lesions will not be identified. Approximately 30% of cancers of less than 10 mm in size found at screening are not visualised on ultrasound, and need stereotactic biopsy for diagnosis. The risk is greatest for lobular carcinomas, and cancers hidden amongst benign breast disease. Small lesions close to the chest wall are also more difficult to perceive, as backscatter from the underlying lung degrades the resolution of the image, this effect being greater in obese women. Those with a large amount of sound-absorbing fibrous tissue amongst their glandular tissue also present a problem, as acoustic shadowing from fibrous tissue can mask small cancers. There remains a place for fine needle aspiration (FNA) or core biopsy in women with significant asymmetrical nodularity or thickening with a normal ultrasound. This should preferably be performed by clinicians trained in both clinical examination and ultrasound.

Ultrasound is not a substitute for mammography, and should not be used as a cancer screening tool except under exceptional circumstances, such as in those who require screening due to family history or past radiotherapy and are too young for mammography but cannot tolerate or have contraindications to magnetic resonance imaging (MRI).

Getting the best image of a lesion for diagnostic evaluation is based on a few simple principles.

1. Ensuring that the breast presets are set up correctly. The complexities of this, in modern computer software enhanced equipment, are beyond the scope of this chapter; however, if images are disappointing, a session with an application
Interventional Ultrasound of the Breast

specialist with breast experience can be invaluable. A good explanation of colour Doppler and power Doppler can be found in Breast Ultrasound by Sohn and colleagues.4

2. Always using the highest ultrasound frequency possible, as higher frequencies give improved fine detail. However, there is always a trade-off with depth penetration. With slim patients a frequency of 13–15 MHz should be aimed for, but to scan obese patients with fibrous or dense glandular breast tissue, it may be necessary reduce this to 7.5 MHz, or even to 5 MHz. To achieve this full frequency range usually means the purchase of two probes.

3. Using the focal zones appropriately. Large numbers of active focal zones are satisfactory for still images, but cause the image to blur with movement as there is a trade-off with frame rate. One or two focal zones are best and should be positioned at the area of interest or just below it.

4. Ensuring that the probe is always vertical to the skin surface. Angulation will cause loss of the typical bright acoustic enhancement deep to cysts and the acoustic shadow behind breast cancers, and make it impossible to follow the track of the core biopsy or FNA needle.

5. Using an appropriate amount of breast compression. The amount of compression should be varied to see whether the lesion will change in shape, or slip out from under the probe as is typical of fibroadenomas.

6. Remembering that the breast can be felt at the same time as it is scanned, so that the texture of the tissues being scanned, and the fixation of lesions relative to the skin and deep fascia can be assessed.

7. Always remembering to confirm that the lesion identified on ultrasound is the same lesion that is palpable and/or visible on mammography. This is usually straightforward in the symptomatic clinic, but can be challenging in breast screening practice. The injection of a very small amount of radiopaque contrast, 0.2 ml is sufficient, or an air bubble and a repeat film are often needed.

For non-radiologists it can be difficult to access and validate breast ultrasound training. The Royal College of Radiologists recommends practical training, working within the requirements of national occupational standards,5 of at least one ultrasound session per week over a period of 6 months to 1 year, with the mentorship of an experienced consultant radiologist to reach level 1 competency.6 For average unaided symptomatic clinic practice level 2 is required, with the ability to perform immediate guided FNA or core biopsy of any lesion seen. In the hands of an experienced practitioner, breast ultrasound looks easy; however, the time needed to acquire these skills should not be underestimated.

The report of a breast ultrasound examination should always be graded. This enables pathways for audit of departmental and individual performance. The simple system of U1 normal, U2 benign, U3 uncertain, U4 possibly malignant and U5 malignant is recommended for easy comparison with subsequent histological data.
THE NORMAL BREAST

The appearance of the normal breast evolves with age, the most important change being the proportion of fat to glandular tissue. In the young breast fat may be almost absent, with bright, reflective glandular tissue from the chest wall to the skin. Small poorly reflective lesions are easily identified. In adult life the glandular tissue forms the filling in a sandwich, with a poorly reflective fatty layer superficially beneath the skin, and deeply between the glandular tissue and the chest wall. The amount of fibrous tissue is very variable. When present in excess, corresponding to a prominent duct pattern on mammography, or in areas of fibrocystic disease, fibrous tissue absorbs and reflects sound, severely decreasing the depth of penetration and producing confusing acoustic shadows. The postmenopausal breast gradually becomes entirely fat, with reflective areas persisting in areas of fibrosis. Ultrasound has a higher predictive value in the postmenopausal breast due to the decrease in benign disease.

BENIGN LESIONS

General principles

Benign lesions have a wholly smooth or undulating lobulated surface. They are oval, with the long axis parallel to the skin. Blood vessels, when visible on Doppler, branch regularly, in a tree-like fashion. Growth by expansion pushes aside the structures of the breast, so that Cooper’s ligaments are bent around the lesion, not disappearing at the edges as though eaten through.

When adjacent to the skin, the bright lower layer of the dermis is intact. Acoustic shadowing may occur in any lesion containing excess fibrous tissue, and does not necessarily imply a malignant lesion. Fine edge shadows may be present with benign lesions, created by the incident beam on the smooth edges. A lesion should always be judged by its worst area, and a well defined lesion must be well defined over all of its circumference. Any trace of a reflective corona moves a lesion into a suspicious group. It is safe to assume that any lesion definitely containing fat is benign.

Common benign lesions

Fibroadenomas

Fibroadenomas are characterised by an oval or smoothly lobulated solid mass, usually less than 3 cm in the longest axis, which lies parallel to the skin. The typical appearance evolves with age. In young women fibroadenomas may be of the same echotexture as fat, distinguishable only by shape, correspondence to the palpable mass, and an increase in vascularity, with characteristic regular vessels that sweep around the edge before entering the lesion. With increasing age the fibroadenomas become echo poor...
compared with fat, the outline becomes more lobulated and the lesion less vascular. Calcifications, initially microscopic calcifications, gradually enlarge to form bright echoes with acoustic shadow. Old fibroadenomas, usually in postmenopausal women, may hyalinise, and contract slightly, with a coarsely lobulated outline and patchy acoustic shadowing, resembling a malignant lesion. Vascularity decreases with age.

In young women very large, rapidly growing ‘giant’ or juvenile fibroadenomas occur. Apart from their greater size, the appearance is typical. They are often very vascular. During lactation, pre-existing fibroadenomas may enlarge rapidly and develop small cystic spaces, with a characteristic appearance. They involute equally rapidly with the end of lactation (see Figures 5.1 and 5.2).

**Figure 5.1** Young fibroadenoma – Smooth, oval, long axis parallel to the skin. Fine homogenous internal echoes. Slips out from under probe with compression.

**Figure 5.2** Lactating fibroadenoma – Fluid filled clefts cause rapid enlargement with the onset of lactation.
Cysts

Cysts are found in pre- or perimenopausal breasts, or in older women on hormone replacement therapy (HRT), but rarely after the age of 60. With frequencies over 10 MHz cyst fluid is seldom echo-free, as the cyst fluid accumulates debris, dead or degenerated apocrine cells, and cholesterol crystals which reflect sound. There is a characteristic bright shadows, acoustic enhancement, and there may be fine edge shadows at the borders of the lesion. A cyst must always be avascular on Doppler. There may be a prominent debris layer, accumulated along the cyst wall with gravity, in the direction of the woman’s feet. In old cysts the fluid may become so thick that the lesion is indistinguishable from a solid mass acquiring a pasty texture that is difficult to aspirate, and core biopsy is necessary to establish a diagnosis. Intracystic papillomas may be difficult to distinguish from debris, but are usually hummock shaped, and occur anywhere on the cyst wall. Some vessels are usually apparent within the papilloma on Doppler. Core biopsy of these lesions is a challenge, as the needle must be directed through the base of the papilloma, and visualisation is often lost after the first pass. If the biopsy is inadequate, a review in 3–6 weeks is needed in order to give the cyst time to refill, as small papillomas will be invisible against a background of normal tissue plus oedema and haematoma.

A cyst should have no discernable wall. High-grade carcinomas may occasionally become centrally necrotic, and appear as a thick-walled cyst. Vascularity will be apparent within the wall on Doppler, and there will usually be reactive, inflammatory, increases in echoes in the surrounding fat. Simple aspiration reveals serosanguinous fluid, not the typical dark green or ‘squid ink’ fluid found in an old cyst, or the thick pus expected from an abscess. Cytology is not successful from the necrotic cells in the fluid. If any doubt remains a core biopsy should be performed. During lactation milk-filled cysts, galactoceles, are occasionally found, with a typical layered appearance. (See Figures 5.3, 5.4 and 5.5.)
Phyllodes tumours may appear identical to fibroadenomas when small, but grow to a much larger size. They become less even in texture when large and may show prominent bright strands parallel to the skin surface. They tend to occur in older women and may appear postmenopausally.
Fat necrosis

Following severe blunt trauma as haematoma resolves, the characteristic features of crushed fat appear, with brightly reflective oedematous fat, containing superficial cyst-like structures, oil cysts, which when aspirated yield fluid with the appearance of sunflower oil. Fibrous irregular masses may also form as haematoma resolves, more commonly with surgical injury, these may be difficult to distinguish from malignancy.

Lipomas

A lipoma is an oval area of breast fat, in small lesions frequently showing a diffuse increase in fine uniform echoes compared with normal subcutaneous fat, although large lipomas are distinguishable only by the presence of a fine outer capsule and their correspondence to a palpable mass. Small lesions may be difficult to distinguish from the rarer neurofibroma, although these are usually more vascular and may be tender. Large lipomas may arise within the chest wall and extend through the pectoral muscle, causing gradually worsening breast asymmetry. (See Figure 5.6.)

Sebaceous cysts

Sebaceous cysts are frequently referred to breast clinics, and occasionally presenting as an apparent mass on mammography, ultrasound is only needed if the clinical appearance is not characteristic. They appear as sharply defined, usually brightly reflective, occasionally layered masses within or immediately beneath the skin.

Figure 5.6  Lipoma – Faint fibrous capsule surrounds fat which is slightly more reflective than the surrounding normal fatty lobules.
When infected the surrounding fat shows bright oedema and the overlying skin is thickened.

**Less common benign lesions**

**Radial scars/complex sclerosing lesion**

Radial scars/complex sclerosing lesions are usually found during screening. They are often subtle on ultrasound, showing a small irregular mass, with a faint acoustic shadow. They cannot be reliably distinguished from small carcinomas, although with the typical mammographic appearances a complex sclerosing lesion can be strongly suspected.

**Diabetic fibrous mastopathy**

Diabetic fibrous mastopathy produces a poorly reflective mass with marked acoustic shadowing and irregular margins, indistinguishable from malignancy except by core biopsy. It may involve large areas of the breast. A similar condition may occur in non-diabetic women, of unknown aetiology, usually referred to as sclerosing lymphocytic lobulitis.

**BREAST CANCER**

The typical appearance of breast cancer is of an ill-defined echo-poor mass. An acoustic shadow is often present, and is more apparent in low-grade lesions. A corona, a ring of featureless brightly reflective tissue, varying from 1 to 6 mm is frequently present. Normal breast structures appear sharply cut off at the periphery of the mass. Changing the angle of the scan to align the probe with a Cooper’s ligament will help to demonstrate this characteristic feature. Spiculation, with the local Cooper’s ligaments drawn into the mass, is sometimes apparent – this is much more clearly seen if three-dimensional scanning is used. This technique produces beautiful images, but is time consuming, and adds little to the diagnostic usefulness of breast ultrasound. Microcalcifications are frequently seen as bright specks. Large masses often have central necrotic areas, visible as irregular areas of increased echoes. These areas should be avoided when performing an ultrasound guided biopsy; always aim at the edge of a large lesion. The presence of skin invasion, with loss of the bright line of the lower surface of the dermis should be noted. Expanded ducts passing from the edge of the main mass should be looked for carefully, these frequently represent ductal carcinoma in situ (DCIS) and should be included in the total measurement. The area surrounding the mass, and between the mass and the nipple should be examined for satellite lesions. (See Figures 5.7 and 5.8.)
Less common appearances of breast cancer

An apparently well-defined mass

Breast cancer can appear as an apparently well-defined mass which is round or oval in any direction with fine internal echoes, and may have a bright shadow of
Interventional Ultrasound of the Breast

acoustic enhancement. Up to 2% of cancers have this appearance. The appearance may be confused with a cyst, until vessels are seen on colour Doppler, or aspiration fails. Lesions are very rarely entirely well defined over their whole circumference – a mass should always be judged by its worst area. This appearance may occur in high-grade rapidly growing cancers where closely packed cells provide few interfaces to reflect sound and expansion is so rapid as to mask the signs of invasion, so that the Cooper’s ligaments appear to bend around the lesion rather than to go through it. (See Figure 5.9.)

A round or finely irregular mass

The echo texture of a round or finely irregular mass may be almost equal to fat, but with a fine speckled appearance, often described as ‘pepper and salt’ and a thin bright corona. This appearance usually represents a mucinous carcinoma, and is commoner in women over the age of 60.

Cavitating lesion resembling a breast abscess

Any apparent breast abscess, peripherally placed in a non-lactating breast must be regarded as a carcinoma until proven otherwise. While the red oedematous appearance of the breast clinically resembles infection, cancers are much less painful. Aspiration will yield serosanguinous fluid, not typical pus. The fluid is unhelpful for cytology, as any cells will be necrotic, and a core biopsy of the wall is required for histology.
Papillary lesions

Papillary lesions can be seen as intracystic or intraductal solid vascular tissue, which varies from a small mound to tissue which fills the space so that the original cyst is no longer visible. These usually represent a form of DCIS, but the appearance of an adjacent extraluminal mass suggests invasion.

Diffuse oedema

In diffuse oedema the breast shows diffuse inflammatory change with bright fat, thickened skin and markedly dilated subcutaneous lymphatics. This appearance corresponds to inflammatory carcinoma clinically. A similar picture is seen with severe oedema from other causes, usually heart failure, reaction to radiotherapy, or hypoalbuminaemia. No discrete mass may be found. If histology cannot be obtained from biopsy of axillary lymphadenopathy, skin punch biopsy is more likely to be successful than blind core biopsy as the dilated subcutaneous lymphatics are often full of tumour cells.

An almost entirely bright mass

Rarely an almost entirely bright mass can be found in which the bright corona is much more prominent than the dark central echo-poor mass. This is a very rare appearance. (See Figure 5.10.)

Figure 5.10  Unusual cancer – Appears as a mass predominantly brighter than the surrounding fat, with an irregular poorly reflective core.
**Enlarged axillary lymph nodes with no apparent breast abnormality**

Approximately 1–2% of all breast cancers present as enlarged axillary lymph nodes with no apparent breast abnormality. A ‘second look’ ultrasound to biopsy any lesion found on MRI offers the best chance of diagnosis. Any abnormality however slight should have a core biopsy.

**Benign lesions that may mimic cancers**

**Areas of fibrosis**

Fibrous areas may occur alone or with cysts as fibrocystic disease. Irregular poorly reflective areas with acoustic shadowing are seen which may resemble carcinoma. There will be no architectural change or destruction of Cooper’s ligaments. Compression will reduce or remove the acoustic shadow.

**Scars and fat necrosis**

Diagnosis of scars and fat necrosis depends on an adequate history. To produce fat necrosis blunt trauma must be sufficient to produce a noticeable bruise. The commonest non-surgical cause is seatbelt crush injury from road traffic accidents. Irregular, poorly reflective areas with acoustic shadowing and distortion of architecture may produce areas indistinguishable from carcinoma. These will be avascular on Doppler, but this is not a secure means of differentiation from carcinoma. Core biopsy or MRI is needed.

**Old fibroadenomas**

Usually in postmenopausal women fibroadenomas may hyalinise or calcify. As they shrink the outline may become irregular. They are avascular on Doppler and there is no architectural distortion. The mammographic appearance often remains typically benign.

**Pseudoangiomatous stromal hyperplasia**

Pseudoangiomatous stromal hyperplasia is a relatively common pathological finding with a variety of appearances. It may appear as a relatively well circumscribed mass, or a diffuse area of patchy reduced reflectivity.

**Sarcoidosis**

Sarcoidosis is rare in the breast, and produces distorted, poorly reflective areas indistinguishable from carcinoma. It is usually multifocal and bilateral.
DUCTAL CARCINOMA IN SITU

Very variable levels of visualisation of DCIS have been reported in the literature from clinical trials. Microcalcification is clearly seen on ultrasound only when lying in poorly reflective tissue, and DCIS tends to be visible only when ducts are grossly distended or there is cancerisation of lobules histologically. The typical appearance is of small irregular poorly reflective areas, with bright specks of calcification trailing off at the edges. Fibrous tissue reaction and inflammation may occur around areas of DCIS, and moderate acoustic shadowing does not necessarily indicate invasion. It is worthwhile to ultrasound all areas of microcalcification. Small masses may be invasive foci; full preoperative diagnosis prevents the need for a second operation for axillary staging. (See Figure 5.11.)

OTHER MALIGNANCIES

Sarcomas

Any sarcoma may occur in breast tissue, but the commonest forms are liposarcoma and angiosarcoma, which occur in or around the edges of areas previously irradiated. Liposarcoma appears as a rapidly enlarging usually well defined mass, with a variable internal structure depending on the grade and proportion of fatty differentiation. Angiosarcoma appears similar to any high-grade vascular breast carcinoma and is often mistaken initially for a second primary or local recurrence. The diagnosis may be suspected when the lesion continues to bleed copiously after core biopsy.

Figure 5.11 DCIS – Patchy, nodular, poorly reflective areas with bright specks of calcification, plus small areas of acoustic shadowing.
Lymphomas

Lymphomas may present either as malignant nodes in the axilla or in non-Hodgkin’s lymphoma as a diffuse irregular poorly reflective mass without acoustic shadow indistinguishable from a breast primary. It is usually very vascular.

Metastases

Metastases may resemble a breast primary, or appear as well defined oval or round masses, without an acoustic shadow, which may be mistaken for fibroadenomas, but tend to be of finer, less reflective internal texture. The commonest lesions to present in a patient with no known malignancy are melanoma and small cell lung cancer. While many other primaries may metastasise to the breast this is usually at a late stage in the disease, rather than the presenting feature.

THE AXILLA

The preoperative identification of involved axillary lymph nodes has assumed greater importance with the technique of sentinel node biopsy, in order to avoid a second operation in node-positive cases. Ultrasound is only moderately sensitive, with up to 15% of axillae which appear to contain only normal nodes having occult metastatic deposits found in surgical sampling or clearance. Ultrasound cannot hope to identify the tiny deposits found histologically. Progressive changes occur as lymph nodes become replaced by metastatic cancer. Once deposits reach 3 mm they may be identified as focal echo-poor thickenings in the lymph node cortex. As they progressively enlarge they distort the shape of the node, displacing the fatty hilus. In a normal node, blood vessels enter through the hilus. Perforating vessels appear as deposits enlarge, although tumour deposits are frequently less vascular than normal nodal tissue. As further enlargement occurs, the fatty hilus is completely lost, and the node loses its normal oval shape becoming round or irregular. Tumour may extend through the capsule into the axillary fat, producing a mass indistinguishable from a second primary tumour.

Lymph nodes may enlarge due to reactive changes, but retain their normal shape, as the cortex thickens it remains even and the fatty hilus is preserved remaining central (see Figures 5.12, 5.13 and 5.14). As there is considerable overlap, FNA or core biopsy should be attempted to establish a diagnosis. A suggested classification based on appearance is given in Figure 5.13.

INFLAMMATORY DISEASE

Abscess

Simple breast abscess occurs almost exclusively in the lactating breast. There will be a bright area of oedema surrounding an irregular cavity, but no discernable wall.
Figure 5.12  R3, indeterminate lymph node – Thickened cortex, uniform thickness, central fatty hilus not displaced.

Figure 5.13  Suggested classification of axillary lymph node appearance. (a) Normal lymph node. Thin rim of cortex central fat: U1. (b) Benign appearance. Some cortical thickening, less than 3 mm: U2. (c) Indeterminate appearance: Uneven cortex, some areas more than 3 mm: U3. (d) Suspicious appearance. Thick cortex with focal poorly reflective areas: U4. (e) Malignant appearance. Greatly thickened cortex, displaced Hilum: U5. (f) Node replaced by tumour. Penetrating vessels, no hilum visible: U5.
All non-lactating abscesses should be regarded as cavitating carcinomas unless definite pus is aspirated. Several aspirations may be necessary for resolution.

**Periductal mastitis**

Central inflammation, involving the ducts immediately deep to the nipple areolar complex, accompanied in severe cases with skin thickening plus dilated irregular ducts and bright oedema of the surrounding fat. There may be deeper extension into a true abscess cavity. Ultrasound has little role in the diagnosis, which is usually obvious clinically. Aspiration of pus is more difficult than in a simple bacterial abscess, as it is usually too thick, and less helpful in encouraging resolution. (See Figure 5.15.)

**Tuberculosis**

Tuberculosis, now very uncommon in Northern Europe, presents as a chronic complex ramifying abscess, with skin thickening and sinuses.

**GYNAECOMASTIA**

Development of the male breast produces a characteristic cone-shaped area of vascular, poorly reflective vascular glandular tissue underlying the nipple, trailing out, rather raggedly, into a peripheral area of more typical, mature brightly reflective glandular tissue. The growing glandular tissue may by asymmetrical, giving rise to
suspicion of malignancy, but a concave peripheral contour is reassuring. Small ducts are often visible. Any of the benign lesions found in the female breast may occur. Any focal lesion should be subject to FNA or core biopsy. Male breast carcinomas usually arise peripherally, and have an identical appearance to those found in the female breast, but skin involvement occurs early due to the small breast size. Cancers may arise without pre-existing gynaecomastia.

**THE POSTOPERATIVE BREAST**

Seroma in the wound or axilla is a frequent immediate postoperative complication, appearing as echo-free fluid if not contaminated with blood, and is easily aspirated under ultrasound guidance.

Haematoma contains more echoes and organises rapidly, with fibrous strands making aspiration difficult.

Infection may present with reflective areas of oedema and overlying skin thickening. Oedema may break down into collections of pus, which should be aspirated.

Local recurrence can be difficult to distinguish from scarring unless there is previous imaging with definite evidence of no progressive change. Both may appear as a poorly reflective irregular mass with a prominent acoustic shadow. The absence of blood flow on Doppler in a mature scar does not definitely exclude local recurrence. MRI is the most accurate method of diagnosis; core biopsy may be needed if this is not readily available. Recurrence remote from the scar is more easily evaluated, and usually has an appearance resembling a new primary. Skin recurrence appears first as small poorly reflective areas within thickened skin. A skin punch biopsy is the easiest route to a histological diagnosis. (See Figure 5.16.)
Interventional Ultrasound of the Breast

FUTURE DEVELOPMENTS

Harmonic imaging and microbubble contrast agents have made little impact on the routine practice of breast ultrasound to date, as the depth of tissue in most breasts is insufficient for harmonics to produce much improvement in image quality. Microbubble agents may improve the identification of small deposits in axillary lymph nodes. Development of systems for delivery of chemotherapy direct to a tumour site, by drug encapsulation within stable microbubbles has been successful in animal models. These drug containing microbubbles are injected intravenously, but remain intact in the general circulation. The microbubbles are then ruptured by the energy deposited when the tumour is scanned, releasing the drug locally within the tumour vessels.

Quantification of the change in shape of a breast lesion with compression, elastography, presented as a change in colour of the lesion, shows promise in improving differentiation between benign and malignant lesions. While this is currently a research tool, equipment with this functionality is about to appear on the market.

REFERENCES

5. www.skillsforhealth.org.uk
Preoperative diagnosis and biopsy guns

Karl J Sweeney and Michael J Kerin

HISTORY

The development of preoperative diagnosis using triple assessment techniques allows one stage resectional surgery for breast cancer. The use of core biopsy techniques for preoperative histological diagnosis is now the norm in many breast units throughout the world. However, it has not always been so.

Historically, pathological confirmation of a breast cancer occurred by either open biopsy or resectional surgery. Mastectomy was performed for all clinically diagnosed breast cancers (a small number of which were done for benign disease) until the seminal NSABP B-06 trial confirmed the efficacy of lumpectomy. The introduction of intraoperative frozen section allowed surgeons to confirm malignancy during the operation but prior to mastectomy, thus sparing some women from unnecessary mastectomy. These unfortunate women consented to open biopsy or mastectomy and the psychological morbidity of this approach identified the need for preoperative diagnosis. Furthermore, improvements in the understanding of breast cancer biology led to the increased use of breast conserving surgery mandating efficient preoperative planning. The era of mastectomy as the definitive treatment for breast cancer has passed and today preoperative diagnosis by minimally invasive core biopsy is expected even for impalpable disease.

The introduction of fine needle aspiration in the 1970s and core biopsy in the 1980s meant that a preoperative cancer diagnosis could be achieved in the majority of patients with symptomatic breast cancer. Fine needle aspiration cytology (FNAC) can give an immediate diagnosis in the outpatient clinic. It is relatively cheap and easy to perform. In contrast, core biopsy requires a skin incision, is more time-consuming, is more expensive and will not give an immediate diagnosis.

THE NEED FOR PREOPERATIVE DIAGNOSIS

The rationale and necessity for preoperative diagnosis is readily understandable. The principle of preoperative assessment of breast cancer is to facilitate rapid referral of patients for one-stage resectional surgery. Definitive non-operative diagnosis of benign conditions is useful for the rapid discharge of patients from the breast clinic. The traditional ‘triple assessment’ of clinical, radiological and pathological evaluation
Interventional Ultrasound of the Breast

provides the highest level of diagnostic accuracy. In isolation, each element may not come to the final true diagnosis; however, concordance between these three modalities provides an accuracy of more than 99% in defining the breast lesion. Discordance mandates further evaluation, usually in the form of radiologically guided biopsy or surgical open biopsy.

The triple assessment clinic evaluates patients with symptomatic breast disease, whereas the mammographic screening clinic evaluates asymptomatic patients with a mammographic abnormality identified as part of a population breast cancer screening programme. Both types of clinic utilise the triple assessment template. The results of all these processes should be discussed at a multidisciplinary meeting which includes all those who assessed the patient and should also include other specialties who may be involved in the adjuvant setting such as medical and radiation oncology. A multidisciplinary management plan is agreed upon and the results and proposed management plan is discussed with the patient.

Clinical assessment is usually the first step in the diagnosis of breast pathology of symptomatic patients. A careful history and examination can establish risk factors for breast cancer, duration of symptoms, comorbidities and often differentiate benign from malignant lesions. Clinical evaluation can identify patients unsuitable for breast conserving surgery despite radiological appearances. Most patients presenting to the clinic have benign conditions, primarily mastalgia or nodularity of a benign nature. A proportion of patients will present with a true abnormality such as a lump or nipple discharge which is clinically benign. A smaller proportion will present with an abnormality which is suspicious but not conclusively malignant, and the most uncommon group of patients attending the breast clinic have a definite clinical cancer.

Ideally, suspicious clinical abnormalities should be discussed with the radiologist at the clinic to ensure that a complete radiological evaluation is carried out in the absence of any mammographic findings. Often either a line drawing indicating the clinical abnormality or even a demonstration of the abnormality to the radiologist is useful for this purpose. Discussion in clinic is particularly important when performing radiologically guided biopsy of lesions to ensure that the radiological findings concur with the clinical findings. Core biopsy should not be performed prior to full radiological assessment as the subsequent tissue distortion may reduce mammographic efficacy.

FINE NEEDLE ASPIRATION CYTOLOGY

The initial design of the triple assessment clinic included the presence of a cytologist at the clinic and a rapid cytological diagnosis during the patient’s outpatient visit. The cytology result is classified as C1: inadequate; C2: benign; C3: uncertain, probably benign; C4: uncertain, probably malignant; and C5: malignant (Figure 6.1). Histological evaluation by core biopsy has replaced FNAC primarily because of a lack of cytological expertise and because it can differentiate in situ from invasive disease. Frozen section evaluation of a core biopsy is not adequate for
preoperative assessment of breast lesions and it often takes a number of days to get a result from a core biopsy.

Patients should be encouraged to bring someone with them when the results of their tests are being discussed at the next outpatient visit. A diagnosis of cancer should be given by the surgeon with a breast-care nurse in attendance. The follow-up arrangements should be clear to the patient, surgeon and breast-care nurse, and the patient must have a contact telephone number for the breast-care nurse for the inevitable postconsultation questions.

The success of preoperative diagnostic testing depends on the experience of the operator, correct equipment and procedural room setup, skilled assistance and careful tissue preparation. The number of staff performing FNAC or core biopsy in a breast centre should be minimised and performance audited regularly. Although the equipment required to perform aspiration or core biopsy is minimal, the ease of performing the procedure and the procedural time can be improved by a skilled assistant who sets up the equipment and provides clinical assistance during the biopsy. When performing FNAC it is recommended that an assistant trained in specimen preparation is present. If a pathologist is available, the adequacy of the aspirate may be assessed immediately thus reducing the recall rate for further FNAC.

The accuracy of FNAC depends on an adequate and representative sample, correct processing of the aspirate and expert interpretation of the cytological material. The ability to transport the cells to the laboratory in liquid medium prior to cytopsin and fixation has also been shown to have a high diagnostic yield. The rate of inadequate or equivocal samples can be as high as 30%, especially in paucicellular lesions such as sclerosing fibroadenomas and this can limit the effectiveness of this technique.

Fine needle aspiration is performed in the outpatient setting. The patient is informed of the procedure and signs a consent form. The patient is usually lying down and the skin overlying the palpable abnormality is prepped with an alcohol solution. Local anaesthetic is usually not required. The lesion is fixed with the non-dominant hand
and a 22 or 23 G (blue or grey) needle attached to a 10 ml syringe is passed into the lesion. Suction may be more easily applied by using a syringe holder. The lesion is aspirated and a number of passes through are made without withdrawing the needle from the skin. If a large amount of blood is obtained it is advisable to discontinue the procedure, apply pressure to the area and repeat the aspiration from a different angle.

Breast lesions are often deeper than initially thought and if there is any doubt a longer needle should be used for the sampling. If there is no resistance and the lesion is not felt to be a lipoma, then it is likely to have been missed and reaspiration should be performed. Large fat droplets or heavily blood-stained aspirates are more likely to be unrepresentative of the lesion.

The resultant tissue is usually visible in the hub of the syringe and this is prepared for cytology. The cytology specimen may be either fixed on a glass slide or placed in a transport medium to be processed in the cytology laboratory. When preparing a slide it is important to spread the aspirated material as a thin layer on the slide before fixing it, either with a fixative solution or by drying it rapidly by waving it in the air.

**CORE BIOPSY**

The spectrum of histological abnormalities of the breast is large and not fully understood. Furthermore, the histological classification of core biopsies is designed to aid the assessment process but not to give a definitive diagnosis (although it is possible in most cases). Thus, a small proportion of biopsies (<10%) cannot be categorised as normal, benign or malignant. The reporting classification is:

- B1 – normal tissue
- B2 – benign lesions
- B3 – uncertain malignant potential
- B4 – suspicious
- B5 – malignant.

A B1 diagnosis should always raise the possibility of a missed lesion, although the histological features of a hamartoma or lipoma may fall into this category. Benign lesions (B2) include fibroadenomas, fibrocystic changes, sclerosing adenosis, duct ectasia and abscess or fat necrosis (Figure 6.2).

Lesions of uncertain malignant potential are often associated with malignancy on excision biopsy. Atypical intraductal epithelial proliferations range in severity from the mildly atypical to having all the features of ductal carcinoma in situ (DCIS) (B5) but an insufficient tissue sample to fulfil all the criteria for DCIS. The definition of atypical ductal hyperplasia (ADH) relies on a combination of histological, morphological and size extent criteria and, therefore, accurate diagnosis is only possible on excision biopsy. In fact, it has been shown that over 50% of excision biopsy specimens, subsequent to core biopsy demonstrating probable ADH, contained in situ carcinoma. Lobular neoplasia (atypical lobular hyperplasia or lobular carcinoma in situ)
is often a coincidental finding in a core biopsy from a screen detected lesion. It does not have the same implications as ductal carcinoma in situ and does not require therapeutic excision. However, recent publications have confirmed the higher incidence of breast cancer in the index quadrant associated with these lesions. Phyllodes tumour may be difficult to distinguish histologically from a fibroadenoma (Figure 6.3). Furthermore, the radiological appearances and clinical findings of the two are identical. The risk of phyllodes increases significantly with the size of the lesion and for this
reason all fibroadenomas greater than 3 cm should be excised locally. An intraductal papilloma is usually benign; however, they may show considerable heterogeneity and in situ carcinoma may be missed on core biopsy. These lesions should be excised. Radial scar (complex sclerosing lesion) should be excised because of the risk of associated DCIS or invasive carcinoma.

The category of B4 (suspicious) should not be an indication for therapeutic surgery. Lesions in this category are usually poorly prepared or insufficient tissue cores that contain probable carcinoma. This histological category mandates further tissue diagnosis either by further non-operative biopsy or by open biopsy.

Malignant cores (B5) are divided into invasive or in situ carcinoma where possible. The presence of in situ disease only on core biopsy does not preclude invasive disease in the final resected tumour. In fact, 20% of tumours diagnosed as in situ carcinoma on preoperative core biopsy will have an invasive component in the final resected specimen. This rate probably increases with increasing size and grade of the DCIS, and for this reason sentinel lymph node biopsy may be recommended as part of the definitive management of larger, higher grade DCIS tumours as the risk of invasion and the need for lymph node assessment increases (Figures 6.4 and 6.5). Further information available from core biopsy includes the grade of the tumour (which is concordant with the final histological grade in 75% of cases), histological type (lobular vs ductal), oestrogen and progesterone receptor status, and human epidermal growth factor receptor-2 (HER2) status. Hormone receptor status should be repeated on the final resected tumour if negative on the core biopsy.

Cytology provides information on the appearance of individual cells and can differentiate malignant from benign cells. It cannot differentiate in situ disease from

Figure 6.4  Histology: ductal carcinoma in situ. Courtesy of Dr Gearoid O’Laoi, Mercy Hospital, Cork.
invasive carcinoma because it does not demonstrate the interaction between the cells and the basement membrane. Core biopsy can provide information, not only on the appearance of the cells themselves, but also on the interaction between cells and the basement membrane, and the appearance of tumour cells as a group (thus differentiating lobular from ductal carcinoma). It may also provide information on oestrogen, progesterone and HER2 receptor status.

The development of automated core biopsy guns has resulted in the rapid and widespread adoption of this technique. The original core biopsy needles were hand fired by pushing an outer cutting cylinder over a solid trochar with a small well at the distal end. These devices were cumbersome, difficult to maintain in position during the biopsy procedure and often failed to obtain sufficient tissue for histological analysis. The introduction of spring loaded devices eliminated much of the difficulty in keeping the needle in position and results in a much greater specimen volume. Although originally introduced as sterile disposable inserts in reusable spring devices, mass produced fully sterile disposable spring loaded devices have become more readily available and cost effective (Figure 6.6).

The sensitivity and specificity of core biopsy are higher than FNAC and there is little difference in patient discomfort between the two sampling methods. In addition, the core can give much more information than an aspiration, particularly with regard to invasiveness of the lesion sampled. A 14 G needle is most commonly used although the optimal number of passes required varies according to the mammographic appearances of the lesions being sampled, with fewer passes required for solid lesions compared with microcalcifications. A minimum of five to six passes are required when sampling microcalcifications to minimise sampling error and specimen radiography is required to ensure that representative calcifications are obtained.

Figure 6.5  Histology: invasive carcinoma. Courtesy of Dr Gearoid O’Laoi, Mercy Hospital, Cork.
There may be false positive and false negative results from core biopsy. False positives are presumed to be either excision of the lesion by the core biopsy or failure to excise the lesion at surgery. The reported false negative rate for malignancy with core biopsy is up to 7% and is more common with microcalcifications and may be due to operator inexperience.

The larger the sample size the higher the accuracy of the biopsy. This size of the biopsy is dictated by the size of the needle, the use of vacuum-assisted biopsy gun and the number of biopsies taken. Mammographic abnormalities such as moderate to low level suspicion microcalcification require a larger volume of tissue for accurate diagnosis, and for these lesions it has been suggested that vacuum-assisted core biopsy demonstrates a lower equivocal sample rate and increased accuracy in the detection of small invasive tumours associated with an area of in situ disease.

Core biopsy is also performed in the outpatient setting; however, it requires local anaesthetic and will leave a small scar as a skin incision is required to facilitate the larger core biopsy needle. Patients who are receiving anticoagulation therapy have an increased risk of haematoma following core biopsy and this procedure should be carried out with care on these patients. Aspirin and clopidrogel (Plavix®) are not contraindications to core biopsy; however, an assistant should apply pressure to the biopsied area between each needle pass.

The consent procedure is similar to that of the FNAC and the procedure is usually performed with the patient in the supine position. The area is prepped with an antiseptic solution and local anaesthetic is applied to the skin and the lesion. Core biopsies should be performed with an automated 14 G needle gun rather than a manual disposable gun as these obtain larger samples. A 1–2 mm incision through the anaesthetised skin and superficial fascia facilitates the biopsy needle. The incision should not be located in the nipple areola complex. The most serious complication of core biopsy is iatrogenic pneumothorax and the risk of this is minimised by keeping the needle as close as possible to parallel with the chest wall when firing. For lesions greater than

Figure 6.6 Disposable biopsy gun.
10 mm in diameter the needle tip should abut the lesion prior to firing, while smaller lesions require placement of the needle short of the lesion. The lesion should be fixed in position during the procedure with the non-dominant hand and the area compressed for at least 5 minutes after biopsy. The incision site is closed with steristrips and covered with a waterproof dressing.

**VACUUM-ASSISTED CORE BIOPSY**

Core biopsy still has the problem of underestimation and the potential for sampling error and this has led to the development of vacuum-assisted biopsy devices. The vacuum-assisted core biopsy system (Mammotome®) uses an 8 G, 11 G or 14 G needle probe. They can be used under either stereotactic or ultrasound guidance. The biopsy probe incorporates a vacuum channel which sucks adjacent breast tissue into the biopsy port. A rotating cutting cylinder passes into the port, separating the specimen from the surrounding tissue and this sample is then delivered by withdrawing the cylinder without removing the main probe. Multiple contiguous specimens may be obtained by this method and the volume of tissue obtained is larger than that of automated core biopsy systems (Figure 6.7).

![Vacuum-assisted core biopsy](image)

*Figure 6.7 Vacuum-assisted core biopsy.*
The specimens obtained by the vacuum-assisted system are also X-rayed to ensure that areas with microcalcifications have been sampled. Because the biopsy specimens are larger, it is possible that smaller radiological abnormalities are wholly or almost wholly removed by this technique. To ensure that the area can be identified in the event of surgical intervention a small metal marker can be deployed down the probe into the biopsy site (Figure 6.8).

The vacuum device has been demonstrated to be superior in the diagnosis of DCIS compared with 14 G core biopsy (6% vs 21% missed invasive carcinoma). Repeat biopsy rates for inadequate sampling of microcalcifications are also significantly lower using vacuum biopsy compared with core biopsy (12% vs 24%).

**COMPPLICATIONS OF CORE BIOPSY**

Complications of preoperative biopsy are fairly uncommon and rarely serious. Although pain is described during fine needle aspiration this usually settles quickly. Pain after core biopsy as the local anaesthetic wears off may be prevented by giving the patient paracetamol at the time of the procedure.

Pain may also be due to haematoma formation which is reduced by applying pressure to the biopsy site at the time of the procedure. Infection is rare if sterility is observed during biopsy. Biopsy sites should be covered with a waterproof dressing and patients should be advised not to remove the dressings for at least 5 days after a core biopsy. A patient re-presenting with a cellulitis at the site of a biopsy should have the extent of cellulitis marked and be placed on oral penicillin. If there is any clinical
suspicion of an underlying infected haematoma or abscess, an ultrasound-guided aspiration should be performed. Seeding of the biopsy needle tract with tumour is theoretically possible and migration of epithelial cells along the tract has been demonstrated previously; however, there is no evidence that breast cancer may spread by this method.

Vasovagal attacks are more common in upright stereotactic procedures and for this reason ultrasound-guided biopsy in the supine position or stereotactic biopsy on a dedicated prone table is recommended. Removal of the lesion by the biopsy probe is more common after vacuum-assisted biopsy and can usually be predicted on radiological evaluation and a marking clip can be placed at the time of biopsy. The risk of pneumothorax is rare and usually occurs in women with small breasts with medial or axillary lesions. Because the distal point of the needle points at the cassette holder, pneumothorax is an unusual occurrence during stereotactic biopsy. Ultrasound guidance is also relatively safe from this complication which occurs most commonly during free-hand biopsies.

**IMAGE GUIDANCE**

Although free-hand biopsy of palpable lesions may be performed in the outpatient setting by the attending surgeon, a large proportion of mammographically detected lesions are impalpable and require image guidance for sampling. Furthermore, image guidance may improve the accuracy of sampling palpable lesions.5

The two principle imaging modalities are ultrasound and X-ray stereotaxis. The use of magnetic resonance imaging is limited by its cost and the availability of equipment and this would not be a routine imaging modality. Ultrasonography is the modality of choice as it is easier to perform, more comfortable for the patient and less time-consuming. The stereotactic technique is used for image-guided biopsies of suspicious microcalcifications (if not visible on ultrasound) or small soft tissue masses that cannot be adequately visualised on ultrasound.

When using ultrasonography to guide FNAC the needle is usually attached by a short tube to the syringe and is held by the assistant. It is introduced into the breast along the line of the long axis of the ultrasound probe and is easily visualised if kept parallel to the surface of the probe. The needle tip is guided into the lesion and an image is taken to record that the needle is in the correct position. The needle is then moved back and forth within the lesion with negative pressure being applied as in the free-hand technique (Figure 6.9).

When using ultrasound to guide core biopsy the needle is again passed along the long axis of the ultrasound probe and advanced to within a few millimetres of the lesion. The gun is fired and an image is taken to record position (Figure 6.10). One or two passes are usually sufficient to obtain diagnostic material.

The patient is usually seated for stereotactic FNAC, unless on a dedicated prone table. After demonstrating the lesion on a straight scout film, paired stereotactic views are obtained with the X-ray tube angled 15° either side of the central tube position.
Interventional Ultrasound of the Breast

The position of the lesion on the stereotactic views is used to determine the position of the needle guide in the x and y axes so that, when a needle of known length is introduced through the guide into the breast, the needle tip will be in the correct position in the lesion. Different areas of the lesion are sampled by up to five aspirates. The first pass is through the centre of the lesion and subsequent passes should form a ‘star’ pattern for stellate lesions.

Setting up for stereotactic core biopsy is similar to FNAC. After checking the position of the needle tip it is withdrawn about 5 mm so that the biopsy port will traverse the lesion on firing (Figure 6.11). It is important to ensure that there is sufficient tissue deep to the needle tip after firing, so that it does not hit the cassette holder. As with stereotactic FNAC the needle is usually placed in a vertical position; however, the lateral approach is possible while the breast is compressed in the craniocaudal

**Figure 6.9** Ultrasound-guided fine needle aspiration cytology. Courtesy of Irene Sweeney, University College Hospital, Galway.

**Figure 6.10** Ultrasound-guided core biopsy of an impalpable lesion. Courtesy of Irene Sweeney, University College Hospital, Galway.
position with some systems (this is particularly useful in lesions in the inferior part of the breast). Most dedicated prone stereotactic biopsy apparatus will include digital mammography. Digital images are displayed on a computer screen for targeting and the biopsy needle holder position is adjusted automatically.

Five or more core samples are obtained using a 14 G needle during stereotactic core biopsy. It is essential that all core biopsies taken of areas of microcalcification are X-rayed to ensure that tissue with microcalcification has been obtained.

Figure 6.11  Stereotactic core biopsy on a prone table. Courtesy of Irene Sweeney, University College Hospital, Galway.

Figure 6.12  Specimen X-ray following core biopsy of microcalcifications. Courtesy of Irene Sweeney, University College Hospital, Galway.
Interventional Ultrasound of the Breast

Stereotactic core biopsy using a 14 G needle is widely accepted to be sensitive and specific in diagnosing breast masses, with rates of 91% and 99%, respectively, compared with 62% and 87%, respectively, for FNAC (Figure 6.12).

**OPERATIVE CHOICES**

An important aspect of patient counselling after a cancer diagnosis is to explain the relative risks and benefits of breast conserving surgery versus mastectomy, the cosmetic result and the reconstructive options available to the patient. The preoperative diagnosis of multifocality of breast cancer by biopsy of each radiological focus is an important element of the decision to proceed to breast conserving therapy. Although the majority of patients who are deemed suitable for breast conserving surgery are successfully treated with one operation, a proportion (between 5% and 20%) will have, on histological evaluation, disease exceeding preoperative radiological and clinical expectation. These patients will require either excision of cavity margins or mastectomy. All women undergoing mastectomy should be considered for immediate reconstructive surgery. Additionally, patients undergoing extensive wide local excisions may benefit from some form of reconstruction (either to replace the excised tissue or to reduce the contralateral breast to achieve symmetry).

Open or excision biopsies are performed less frequently with the improvements in radiological and histological techniques, and the improvements in percutaneous biopsy techniques. Open biopsy is usually performed because: the lesion is inaccessible for radiologically guided core biopsy; the histology of the core biopsy, although benign, suggests associated malignancy not sampled by the core; or the histological subtype is potentially premalignant.

![Histological slide confirming presence of calcium oxalate under polarised light.](https://example.com/histology.jpg)

**Figure 6.13** Histological slide confirming presence of calcium oxalate under polarised light. Courtesy of Dr Gearoid O’Laoi, Mercy Hospital, Cork.
Excision biopsy differs from wide local excision in that the aim of the procedure is not for oncological clearance but for histological confirmation of the presence or absence of disease (Figure 6.13). Thus, this is not a therapeutic procedure and it is imperative that the cosmetic results of the surgery are acceptable. For this reason the amount of tissue excised should be minimal (less than 15 g) and the incision should be placed in as discreet a site as possible. Open biopsy is usually performed under general anaesthetic and is often suitable for day-case surgery.

Biopsy of impalpable breast lesions must be performed with radiological guidance. Lesions close to the chest wall and in other inaccessible areas of the breast may not be amenable to radiologically guided core biopsy. Since these lesions are subclinical, an ultrasound or stereotactically guided wire is placed into the lesion under local anaesthetic on the morning of surgery. The surgeon then performs an excision of the tissue surrounding the tip of the guide wire under general anaesthetic. It is not uncommon that the tissue excised is grossly normal and to confirm excision of the radiological abnormality a specimen mammogram is performed prior to completing the procedure. To facilitate the surgeon, the patient should arrive to theatre with either mammograms following placement of the wire or a diagram drawn by the radiologist indicating the site of the lesion relative to the wire (Figure 6.14). The size of the lesion is recorded prior to the surgery to ensure adequate excision, and the excised specimen should be weighed by both the surgeon and the pathologist.

Adequate oncological management of breast cancer includes pathological assessment of the axillary lymphatic tissue. The presence of malignancy in axillary lymph nodes has a significant impact on disease-free and overall survival. It is also an indication of adjuvant systemic therapy as it is a predictor of systemic disease. The majority of women without palpable axillary disease do not have lymph node involvement.

Figure 6.14 Specimen mammogram following wire-guided excision biopsy. Courtesy of Fidelma Flanagan, Mater Misericordiae Hospital.
The standard of care of the clinically normal axilla is to perform sentinel node biopsy at the time of tumour excision. A proportion of patients will have a positive sentinel node and require completion axillary clearance under further general anaesthetic. Ultrasound-guided FNAC confirmation of clinically or radiologically suspicious axillary lymph nodes avoids a second surgery as the patient can proceed directly to axillary clearance at the time of tumour resection. Although a positive result from axillary fine needle aspiration is useful, the sensitivity and specificity of the technique has not been established to a degree that a negative result obviates the need for sentinel node biopsy. Thus, all patients with a clinically benign axilla and a negative fine needle aspiration of an ultrasonographically abnormal axillary lymph node must have a sentinel node biopsy.

REFERENCES

Techniques of ultrasound biopsy

Michael Michell

Ultrasound-guided needle sampling procedures are now routinely performed as part of the triple diagnostic method on patients presenting with breast lesions to both symptomatic and screening assessment clinics.\cite{1,2} Earlier presentation of symptomatic women with subtle clinical signs and small lumps, and the detection of small lesions which are often non-palpable through the increased use of screening mammography require accurate image-guided sampling in order to obtain a diagnosis. The accuracy of needle biopsy is also improved by ultrasound guidance for palpable lesions and, in modern diagnostic practice, image guidance should be used for all needle biopsy or aspiration procedures.

EQUIPMENT AND ROOM LAYOUT

High quality ultrasound apparatus using a linear probe with a frequency range of 7–12 MHz and good near field imaging is required for adequate visualisation of small breast lesions and the sampling needle. Using modern high frequency equipment, most soft tissue lesions and some microcalcification clusters will be adequately visualised for biopsy. The examination couch should be positioned in the centre of the room giving access to both sides of the couch with the ultrasound machine on the patient’s right side (Figure 7.1). With the radiologist/operator on the patient’s right side, the nurse or assistant on the other side of the couch is ideally positioned both to assist during the procedure and to offer support to the patient.\cite{3}

Ultrasound-guided sampling/biopsy procedures include:

- cyst aspiration
- abscess aspiration
- fine needle aspiration cytology (US FNAC)
- core biopsy (USCB)
- axillary lymph node sampling
- ultrasound-guided preoperative localisation
- vacuum-assisted biopsy
  - diagnostic
  - therapeutic.
All ultrasound-guided biopsy procedures should be carried out with a multidisciplinary team approach to diagnosis and as part of the triple diagnostic method. Clinical assessment and appropriate imaging work-up should be complete before ultrasound-guided biopsy is performed. The diagnostic team should be familiar with locally agreed protocols regarding the indications for biopsy and the selection of sampling technique in accordance with the clinical and ultrasound characteristics of the lesion and patient age. The results of image-guided biopsy should be discussed at a regular prospective multidisciplinary clinical management meeting, in order to decide on patient management.

**EXPLANATION AND CONSENT**

The operator must ensure that the patient fully understands the indication for the procedure and what the procedure involves. This process is helped in many clinics by providing the patient with written information prior to clinic attendance. In most clinics, it is not considered necessary to obtain written consent for ultrasound-guided breast procedures but policy may vary between different institutions or countries.

**LOCAL ANAESTHETIC**

Abscess aspiration, core biopsy procedures and preoperative wire localisation should be carried out using anaesthetic. Cyst aspiration and fine needle cytology are often carried out with no local anaesthetic – local anaesthetic should be offered to such patients, however, for particularly deep-seated lesions.
**TECHNIQUE**

The advantage of using ultrasound guidance is that the needle can be seen in real time traversing the lesion. In order to see the needle clearly, a parallel approach should be used – the needle is introduced and passed through the breast tissue with its long axis parallel to and in line with the surface of the ultrasound transducer. This ensures clear visualisation of the shaft of the needle and the needle tip, and also ensures that there is no danger of the needle tip traversing the chest wall and damaging intrathoracic structures. It is good practice to record ultrasound images of the needle traversing the lesion, ideally including both an image taken along the long axis of the needle, as well as an image at 90° to this in the orthogonal plane.

**CYST AND ABSCESS ASPIRATION**

Simple cyst aspiration is performed for symptomatic relief of a swelling which may be tender. Asymptomatic cysts do not need to be aspirated unless there are imaging features present which suggest that there may be an intracystic lesion present. The needle, attached to the syringe by a connecting tube, is introduced into the cyst and the contents aspirated (Figure 7.2). The cyst aspirate does not need to be sent for cytological examination unless the aspirate is blood stained.

![Figure 7.2](image_url)  
*Figure 7.2* Ultrasound cyst aspiration. (a) Needle, syringe and connecting tube. (b) Aspiration needle parallel to ultrasound probe and chest wall. (c) Aspiration needle within cyst.
Most breast abscesses can be successfully managed by ultrasound-guided aspiration and oral antibiotics. Before aspiration, the skin is cleaned and local anaesthetic is injected into the skin, subcutaneous tissue and down to the wall of the abscess. Aspiration may be carried out using a 19 G needle but more effective drainage is likely with larger 16 G or 14 G needles. Repeated drainage may be required to achieve resolution of larger abscesses. If the abscess is large and complex with multiple septae, percutaneous drainage is unlikely to achieve full resolution and surgical drainage should be considered.

**FINE NEEDLE ASPIRATION CYTOLOGY**

Following skin preparation and local anaesthetic administration, the 21–23 G aspirating needle, attached to a 10 ml syringe with a connecting tube, is introduced into the lesion by the operator. The needle is moved to and fro within the lesion under ultrasound visualisation with simultaneous rotation and with negative pressure applied by the assistant who holds the syringe (Figure 7.3). Aspiration is continued until material is seen within the hub of the needle. The syringe is removed before the needle is withdrawn so that the aspirate is not sucked back into the connecting tube. The aspirate is then delivered onto slides and both dry and wet preparations are made in accordance with guidance from the cytology laboratory. Two to three separate passes may be made, particularly if the aspirate appears scanty. If local circumstances allow, it may be helpful to have the cellularity of the specimens checked by a cytology technician immediately following the procedure.

![Ultrasound fine needle aspiration cytology](image)

**Figure 7.3** Ultrasound fine needle aspiration cytology.

**CORE BIOPSY**

Recent studies have demonstrated that core biopsy has a higher sensitivity and specificity than cytology in most centres and has therefore been increasingly used as the first sampling procedure in the diagnostic work-up. The use of larger needles is not associated
with increased pain or other complications.\textsuperscript{14} For rapid one-stop clinic diagnosis, cytology specimens can be obtained at the same time as core biopsy using imprint cytology techniques – the core specimen is dragged across the slide producing a cellular smear.\textsuperscript{15,16}

The procedure is performed with the patient in the supine position with the ipsilateral arm raised above the head. For lesions which are situated in the lateral aspect of the breast, the patient should be in a supine oblique position with the ipsilateral side raised and supported on a cushion.

The direction of approach should be planned according to the position and depth of the lesion, and the size and texture of the breast. In very fatty breasts with little residual glandular tissue, the biopsy needle passes easily through the tissue. In glandular breasts or in cases where there is benign breast change with fibrosis, considerable resistance may be encountered. In these cases, use of a radial approach in relation to the nipple will make passage of the needle easier because fewer of the fibrous septae of the breast are traversed.

The transducer is placed so that the target lesion is approximately in the centre of the image. For small or subtle lesions, it may be helpful to mark the skin so that the correct position can be quickly found again between needle passes. The operator should use pressure from the probe as well as the hand holding the probe to immobilise the surrounding breast tissue.

The skin is cleaned and the site selected for entry of the biopsy needle is infiltrated with local anaesthetic. The biopsy needle entry point should be approximately 1 cm away from the edge of the ultrasound probe for most lesions. For deep lesions, however, the entry point should be 2–3 cm away from the edge of the probe – this allows a longer needle path to the lesion and enables the parallel relationship between probe and needle to be retained.

Local anaesthetic is injected both superficially and down to the site of the lesion. For a very mobile lesion, local anaesthetic can also be injected around the lesion to decrease movement during sampling. For very posteriorly situated lesions, local anaesthetic may be injected posterior to the lesion in order to displace it anteriorly and achieve a more satisfactory position for access with the sampling needle. Addition of sodium bicarbonate (0.5 ml sodium bicarbonate to 4.5 ml 1% lignocaine) neutralises the acidic pH and eliminates the initial sting of the anaesthetic injection.

A small skin nick is made to allow easy entry of the biopsy needle into the breast. A 14 G needle is most commonly used (Figure 7.4) but a 16 G needle may be easier to position accurately in very glandular breast tissue.

The needle is advanced through the breast tissue under direct ultrasound guidance. It is important to ensure that the needle remains aligned with the long axis of the ultrasound transducer and that, as far as possible, the needle and face of the transducer are parallel – this will ensure that the full length of the needle is clearly visualised allowing accurate positioning for sampling (Figure 7.5). The tip of the needle is positioned at the proximal surface of the lesion and the biopsy gun is fired. An image of the needle through the lesion is stored. The needle is withdrawn and the specimen delivered into formal saline while the assistant maintains manual pressure
Figure 7.4  Automated core biopsy gun.

Figure 7.5  Ultrasound-guided core biopsy. (a) Local anaesthetic is injected down to the site of the lesion. (b) The biopsy needle is parallel to the ultrasound probe and chest wall with the tip at the proximal edge of the lesion. (c) The biopsy needle has traversed the lesion. Postfire ultrasound image of biopsy needle position (arrow).
over the site of the biopsy to prevent haematoma formation. The macroscopic appearance of the sample is examined and the operator should decide on the basis of the ultrasound appearances during biopsy and the macroscopic appearance of the specimen whether an adequate representative sample has been obtained. For solid breast masses, the tissue will generally appear pale cream in contrast to the yellow fatty normal breast tissue. If there is doubt about the adequacy of the specimen, further samples should be obtained. For a discrete solid mass, one or two samples are generally sufficient; for a less well defined area of abnormality, e.g. distortion or microcalcification, further samples should be taken. For microcalcification, core sample X-ray images should be obtained to confirm that the samples contain representative calcium. At the end of the procedure, the assistant should apply pressure over the biopsy site for at least 5 minutes in order to prevent haematoma formation. A sterile dry dressing, e.g. Steri-Strip, is applied to the site of the skin nick with instruction to keep the site dry for 48 hours before removal. The patient should be provided with aftercare advice, including action to be taken should bleeding occur, and should also be given clear information regarding arrangements for obtaining the biopsy results.

**Difficult procedures**

**Deeply situated lesions**

The direction of approach is carefully chosen so that the route allows the parallel relation of the needle to both the transducer surface and the chest wall to be maintained as far as possible (Figure 7.6). Local anaesthetic may be injected behind the lesion in order to displace it anteriorly away from the chest wall. The biopsy needle may be initially introduced in a posterior direction to achieve adequate depth and then rotated so that the parallel position is achieved. Using this technique, the operator must take care to control the movement of the biopsy needle to avoid traversing the chest wall.

**Mobile lesions**

Some small solid lesions, particularly fibroadenomas, may be mobile and present difficulties for sampling. Liberal infiltration of local anaesthetic all around the lesion together with firm pressure using the ultrasound transducer and the operator’s hand help to immobilise both the lesion and surrounding breast tissue. When the biopsy needle is introduced to the ‘prefire’ position, it should be advanced so that the tip is engaged in the proximal edge of the lesion in order to minimise the risk of the needle ‘skating’ along the edge of the lesion and failing to collect an adequate sample. ‘Postfire’ images of the needle in both longitudinal and transverse planes will demonstrate whether the needle has traversed the lesion.
Multiple lesions

If possible, all necessary biopsy procedures should be carried out during the initial clinic attendance. The operator should decide on the biopsy strategy based on both clinical and imaging findings. For young patients with probable multiple fibroadenomas, as long as the imaging features of all the lesions are consistent with benign lesions, it is only necessary to sample one of the masses – usually the mass which caused the clinical presentation.

In patients with a carcinoma, information regarding both the extent and other foci of tumour are important in deciding management. If further possible foci of tumour are seen in the same quadrant as the main mass, these foci can be effectively sampled using ultrasound FNAC following core biopsy of the target mass. Further lesions in a different quadrant of the ipsilateral breast or in the contralateral breast should be sampled by core biopsy.
AXILLARY LYMPH NODE SAMPLING\textsuperscript{17–20}

Staging of the axillary nodes remains a key step in deciding on the surgical and oncological management of patients with invasive breast cancer. Axillary lymph nodes can be visualised in most patients with modern high frequency ultrasound apparatus. Several studies have described abnormal ultrasound features of axillary nodes such as eccentric cortical thickening of $>2$ mm and loss of the normal echogenic hilum but these features are not sufficiently accurate for predicting the presence of metastatic disease to use in making decisions on clinical management. Both FNAC and ultrasound core biopsy are currently used for sampling axillary lymph nodes\textsuperscript{21} and recent published studies show that in approximately 50\% of patients with node positive invasive cancer, a malignant aspirate or core sample can be obtained from the axilla. If axillary metastatic disease can be confirmed preoperatively, axillary node surgery can be planned for such patients and a sentinel node biopsy procedure avoided. There are currently insufficient data available to show whether ultrasound FNAC or ultrasound core biopsy achieves the higher sensitivity or specificity. The techniques use the same principles as described for sampling breast lesions. Particular care must be taken to ensure that the needle tip is clearly visualised in order to prevent damage to axillary vascular structures.

ULTRASOUND-GUIDED PREOPERATIVE LOCALISATION\textsuperscript{22}

Preoperative marking is required for non-palpable lesions undergoing surgical excision. Ultrasound is the imaging method of choice if the lesion is ultrasound visible and is used for most soft tissue lesions and some microcalcification clusters. The needle containing the localising wire/hook device is introduced into the breast using the same technique and principles as for ultrasound needle biopsy. When a satisfactory position has been achieved with the needle adjacent to the lesion, the needle sheath is withdrawn leaving the wire in position. For mammographically detected lesions, it should be ensured that craniocaudal and lateral mammograms are obtained and are sent with the patient to the operating theatre. A specimen radiograph of the surgically excised tissue should be obtained to confirm that the target lesion has been successfully removed and to radiologically assess the margins around the tumour.

REFERENCES


8

Techniques of ultrasound-guided vacuum-assisted biopsy

Mike Shere

INTRODUCTION

The commonest type of vacuum-assisted large-bore biopsy system used with ultrasound guidance is the Mammotome™ (Ethicon, Endosurgery, Cincinnati, OH). There are other systems such as the SenoCore 360™ (SenoRx Inc, Aliso Viejo, CA) which uses a radiofrequency tip to penetrate the tissue and the Vacora™ (Bard Biopsy Systems, Tempe, AZ) which is self-contained. All three systems use a rotating circular blade to cut the core of tissue. The Mammotome has the advantage of being able to take multiple cores without having to reinsert the needle. There is also the ABBI (Advanced Breast Biopsy Instrumentation System; United States Surgical Corporation, Norwalk, CT), which does not employ vacuum, but uses much bigger cannulae (up to 20 mm) to take specimens with the rotating blade.

Techniques for using the three vacuum systems are very similar, other than the initial use of a cannula to guide the biopsy probe with the SenoCore 360 and the Vacora, and the ease of penetration afforded by the radiofrequency tip of the SenoCore 360. The SenoCore 360 comes in three sizes: 6, 8 and 9 G, the larger taking up to a 260 mg sample. It may not be suitable for some of the indications such as duct excision or cystadenomas because of the need to insert the probe into the lesion before taking the sample. The Vacora only comes in a 10 G size and therefore may not be suitable for large lesions and does not have such a powerful vacuum, being battery powered.

The Mammotome was developed in 1995.¹ There are two versions using different probes: one for stereotactic biopsies of impalpable lesions, and the other for hand-held use under ultrasound control. The original system used an articulated arm to hold the probe, but the next generation was literally hand-held. The third generation (EX), for hand-held use only, is faster and has done away with the mechanical drive cables to the probe. The probes come in three sizes, 14, 11 and 8 G, capable of removing 30, 80 and 240 mg of tissue respectively, per core (Figure 8.1).

The Mammotome was developed as an alternative biopsy system to the manual or semi-automated core biopsy systems which are usually 14 G. Particularly with small impalpable lesions, the idea was that the larger the biopsy, the more likely that it would contain the target lesion. It was soon realised, however, that with some conditions,
Interventional Ultrasound of the Breast

the process could be therapeutic as well, with for example nipple discharge and fibroadenomata. This chapter discusses the techniques not only of ultrasound-guided biopsy, but also of the therapeutic use of vacuum-biopsy systems, mainly the Mammotome.

**PATIENT SELECTION AND INDICATIONS**

**General**

There are many indications for the use of the Mammotome. Indeed, in our unit, most of what is usually regarded as day-case surgery is done in the clinic with the Mammotome, thus freeing valuable theatre time for cancer surgery.

Common to all conditions is the ability of the patient to tolerate the procedure under local anaesthetic. If the patient wants to have a general anaesthetic, then whilst the Mammotome biopsy could be done under general anaesthetic, this adds considerably to the cost of the procedure. Most of the cost of general anaesthetic day-case surgery comes from the staff and facilities, not the disposables, and therefore one is just adding £250 to the cost. The only advantage would be the much smaller incision, but the use of peri-areolar incisions in surgery rarely leaves an appreciable scar. However, using the Mammotome whilst the patient is under general anaesthesia could be useful when starting to learn the procedures, without the added difficulty of giving local anaesthesia.

A general contraindication to Mammotome use is an anticoagulated patient. It has been reported that there were few significant problems with the use of an 11 G probe for stereotactic biopsies in anticoagulated patients. However, with the use of 8 G probes and more extensive biopsies or excisions, it would seem sensible to treat the procedure in the same way as minor surgery and stop anticoagulation beforehand.
Nipple discharge

The indications for the use of the Mammotome in nipple discharge are the same as those for doing an operative microdochectomy for single duct discharge or major duct excision for multiduct discharge. These are single duct discharge, bloody discharge and troublesome multiduct discharge which is for example staining the clothes in the presence of a normal mammogram. Indications also include, with the increasing use of high frequency ultrasound, when there is a visible intraduct papilloma on ultrasound. The aim is also the same: to biopsy the area adequately and to stop the discharge.

Fibroadenoma

The indications for Mammotome biopsy/excision for these lesions are again the same as for operative excision or excision biopsy. The policies for these lesions vary from unit to unit, but the aim is to exclude carcinoma, and most people would agree that in the older woman, it would be unwise to make a diagnosis of fibroadenoma on imaging alone and that a core biopsy should be done. Some policies are made on size as well as age. There may be a diagnostic role for the Mammotome when there is a B3 result from the core biopsy, with its ability to remove most or all of the lesion when the likelihood is that the lesion is benign. However, the main role of the Mammotome is to excise fibroadenomata in women who have a histologically proven fibroadenoma and have opted not to have conservative management, and who want to have the lump excised. Very large fibroadenomata are probably not suitable for excision with the Mammotome as it can take a very long time even with an 8 G probe, and an appreciable amount of normal surrounding breast tissue is excised with the lesion. The limit is probably 25–30 mm in largest diameter as measured by ultrasound. The authors of the paper first discussing this indication suggest that with fibroadenomata over 15–20 mm there is quite a high rate of incomplete excision. However, they only used the 11 G rather than the 8 G probe which makes excision of larger lesions much easier. Another study using both 8 G and 11 G probes showed an incomplete removal rate of 1%. Until the results of randomised trials comparing Mammotome excision with surgical excision, it is wise when counselling these women, to stress that choosing the minimally invasive Mammotome method may have a greater risk of leaving some of the fibroadenoma behind.

Subareolar biopsy before nipple sparing subcutaneous mastectomy

This is a more recent indication designed for situations where a subcutaneous mastectomy is contemplated in, for example extensive ductal carcinoma in situ (DCIS), and it is not known whether it is safe to preserve the nipple. The historical
method is to request frozen section histology of a subareolar operative biopsy whilst the patient is under anaesthetic. This is less than perfect, as frozen section histology is more difficult to interpret and the woman does not know what operation she is having until she wakes up.

**Cystadenoma**

These lesions are usually diagnosed by the ultrasound appearance of a filling defect in a cyst or from the operative excision of a cyst which has refilled or contained blood-stained fluid. The Mammotome is a good way of excising the adenoma under direct ultrasound vision both for the purpose of diagnosis and stopping the cyst refilling.

**B3 core biopsy results**

As mentioned under fibroadenoma, when a core biopsy has been done of an ultrasound visible lesion, with a B3 result, the Mammotome can be used for diagnosis. The proviso is that there should be high probability that the lesion is benign, otherwise an excision biopsy might be preferred which has a greater certainty of complete excision. Included here would be those situations where suspicious mammographic microcalcification can also be seen on ultrasound and the Mammotome can be used for a diagnostic biopsy.

**Abscess**

The majority of lactating breast abscesses can be resolved with repeated aspirations. However, there are some occasions when the pus is too thick to aspirate through a 19 G (cream coloured) needle or when the abscess has become loculated, usually as the result of prolonged antibiotic therapy. We have shown that the Mammotome with its powerful vacuum can be useful to evacuate these abscesses. The abscess usually resolves without needing further aspirations or surgical drainage, at an important time for mother and baby. It is also useful when one needs to take a biopsy of the cavity wall as well. The indication is also being extended to non-lactational abscesses, particularly when there is a chronic subareolar abscess or mammary fistula, possibly following previous surgery for the condition. It has the advantage over surgery through a periareolar incision, in that the Mammotome probe can be inserted well away from the infected tissue.

**Gynaecomastia**

The Mammotome is a useful tool for the minimally invasive treatment of gynaecomastia. Suitable patients for this would be those with mostly glandular tissue, rather than fat which could be dealt with by liposuction. They should also be screened for pathological causes of gynaecomastia.
Fat necrosis

Areas of hard fat necrosis, histologically or cytologically proven, whether caused by trauma or by surgery, can be excised with the Mammatome. These areas are often too hard for the traditional tool of the plastic surgeon, liposuction, and can be excised in the same way as gynaecomastia.

Diagnosis of inflammatory breast cancer

Fine needle aspiration cytology and, occasionally, core biopsy can fail to diagnose inflammatory breast cancer because of the diffuse nature of the disease. The Mammatome, by virtue of taking larger samples of tissue, can be used for diagnosis and for analysis of accurate receptor status in suspicious cases.11

TECHNIQUES

Discussion and informed consent

The procedure can be quite frightening, as there is a large probe with a big needle being thrust into the breast in front of patient’s eyes and there is a roomful of equipment. It is, therefore, worthwhile spending some time explaining it all to the patient in the clinic before arranging the procedure and supplying them with a leaflet explaining why it is being done, how it is done, the equipment being used, what to expect afterwards and how the results of the biopsy will be communicated.

Most people take only verbal consent for an ordinary core biopsy of the breast. However, with this rather more invasive procedure it is probably worth having written consent as well, although there is no legal basis for this.

Preparation and equipment

The preparation and equipment for vacuum-assisted biopsy are shown in Table 8.1.

Positioning

Positioning is probably more important with Mammatome biopsy, than for doing an ordinary core biopsy. Procedures take longer, and a poor position can affect the patient’s comfort and the operator’s back and neck! As well as the relative positions of the probe and patient, there are two screens to look at: the Mammatome machine screen and the ultrasound screen (Figure 8.2). Full use needs to be made of wedges under the patient’s shoulders and back particularly for lateral lesions, and she needs to be comfortable. The height of the table needs to be right for the operator’s comfort. Some lesions may be easier to biopsy with the table down low and the operator seated.
Interventional Ultrasound of the Breast

The assistant also needs to be able to access the Mammutome probe to extract the samples from it as they appear.

Thought needs to be given to the positioning of the screens for each individual procedure, but the author has found that it is generally easier to have the Mammutome and ultrasound machines on the opposite side of the patient to the operator so that he can glance up at them without having to turn his head. Another method is to use a binocular headset so that the ultrasound image is right in front of the operator’s eyes.

### Table 8.1 Preparation and equipment for vacuum-assisted biopsy

<table>
<thead>
<tr>
<th>Assistant (only one person is required in addition to the operator)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vacuum-assisted biopsy equipment (set up and tested before patient arrives)</td>
</tr>
<tr>
<td>Ultrasound machine (plus sterile probe covers if probe cannot be sterilised)</td>
</tr>
<tr>
<td>Gloves</td>
</tr>
<tr>
<td>Skin cleaning solution (this is a clean procedure rather than a sterile one)</td>
</tr>
<tr>
<td>Local anaesthetic (as below)</td>
</tr>
<tr>
<td>20 ml syringe with short 25 G and long 21 G needles</td>
</tr>
<tr>
<td>Size 11 blade</td>
</tr>
<tr>
<td>Gauze swabs</td>
</tr>
<tr>
<td>Steri-Strips</td>
</tr>
<tr>
<td>Elasticated crepe bandages</td>
</tr>
<tr>
<td>Saline solution for cleaning skin after procedure</td>
</tr>
<tr>
<td>Forceps (for removing samples from probe)</td>
</tr>
<tr>
<td>Specimen container containing formalin</td>
</tr>
</tbody>
</table>

![Clinic room set up for Mammutomes. Operator stands on patient’s right opposite machines.](image)

**Figure 8.2** Clinic room set up for Mammutomes. Operator stands on patient’s right opposite machines.

The assistant also needs to be able to access the Mammutome probe to extract the samples from it as they appear.

Thought needs to be given to the positioning of the screens for each individual procedure, but the author has found that it is generally easier to have the Mammutome and ultrasound machines on the opposite side of the patient to the operator so that he can glance up at them without having to turn his head. Another method is to use a binocular headset so that the ultrasound image is right in front of the operator’s eyes.

### Local anaesthesia

Vacuum-biopsy systems have large needles and are often inserted into the breast tissue at some distance from the lesion for both cosmetic reasons and to avoid a steep
angle towards the chest wall. Local anaesthesia therefore needs to be extensive and adequate.

The usual local anaesthetics used are lidocaine or bupivacaine, the latter having an advantage with longer duration of action but slower onset of effect. Because of the use of a large-bore needle in vascular tissue and the strong likelihood of haematoma and bruising, it is advisable to use adrenaline in solution with the local anaesthetic. The author uses 1% lidocaine with adrenaline 1 in 200 000.

The volume required for adequate anaesthesia depends to some extent on the nature of the lesion being biopsied or excised, but 20 ml is usually sufficient. One should be careful not to exceed the maximum recommended dose which is 20 ml of 1% plain lidocaine or 50 ml if given in solutions containing adrenaline. A contraindication to adrenaline-containing solutions is their use in appendages because of the danger of necrosis from reduced blood supply and some have suggested that this applies to the nipple. However, the authors have performed many hundreds of procedures using lidocaine and adrenaline to anaesthetise the nipple area without any problems whatsoever.

One of the unpleasant effects of infiltration with local anaesthetic is the initial stinging. There are two ways in which this can be alleviated. One is to warm the solution to body temperature immediately before use, a convenient way of doing this is to use an ultrasound gel warmer if gel is being used. The other method is to buffer the solution immediately before use with 8.4% sodium bicarbonate solution, as the stinging pain is mostly due to the acidic solutions used to prolong shelf life.

The methods used are similar to any form of local anaesthetic infiltration. The skin is cleaned and a subdermal bleb is first raised using a 25 G (orange) needle, bearing in mind the need for a 5–8 mm stab incision to insert the Mammotome probe. The needle is then swapped to a longer larger-diameter needle. A spinal needle can be used; however, the author uses an ordinary, but longer, 21 G (green) needle 50 mm long which has the advantage of being long enough and robust enough not to bend.

The injection can be used to form a track for the Mammotome probe to follow by the hydraulic pressure of the anaesthetic and the pressure of the anaesthetic can also be used to develop a plane underneath a lesion to lift it away from the chest wall to allow safe insertion of the probe.

**Nipple discharge**

Particularly if the aim is to perform a major duct excision to stop troublesome multiduct discharge then the entry point for the Mammotome will need to be further away from the nipple than the areola–normal skin junction, to allow excision of ducts from the subareolar region right around the clock face. If one specific duct is being targeted, for example if it has an ultrasound visible papilloma in it or is dilated and pressure over it reproduces the single duct discharge (Figure 8.3), or the patient has a
Interventional Ultrasound of the Breast

particularly large areola, then the entry point can be periareolar but from the opposite side of the areola to the abnormal duct.

After the local anaesthetic bleb is raised (Figure 8.4), the longer needle is used to infiltrate the local anaesthetic using the whole 20 ml. The subareolar area is anaesthetised by angling the needle at approximately 30º to the skin, infiltrating all the way as it is pushed down to the far side of the areola (Figure 8.5). It is important to infiltrate a little bit further as there is about 4 mm of probe beyond the cutting groove on the MammoTome probe. The needle is almost withdrawn and the process repeated to anaesthetise under a different part of the areola. This carries on in a fan-shaped way until the whole area is covered (Figure 8.6). Another technique is to try to block the lateral cutaneous branch of the fourth intercostal nerve; however, there is evidence to
show that this is far from being the only nerve supply to the nipple area.\textsuperscript{14} Infiltration should take place under direct ultrasound vision as it is very easy to inject anaesthetic into one of the ducts where it not only does not anaesthetise the area, but also may obscure the papilloma and occasionally comes out of the nipple under pressure into the operator’s eye! Before starting, it is also worth having a look for any large vessels which could be avoided, using colour Doppler.

A stab incision is then made for the Mammotome probe to be inserted. The author uses a straight edged size 11 blade, hand-held, rather than in a handle, because it is rather less frightening than being approached with a whole scalpel (Figure 8.7). The incision length needs to about 5 mm long for an 11 G probe and about 8 mm for the 8 G probe. It is not worth trying to make the incision smaller as the insertion

Figure 8.5  Infiltrating local anaesthetic under the areola.

Figure 8.6  Infiltrating local anaesthetic under subaerolar ducts.
and movement will macerate the skin edges resulting in poor wound healing. For the same reason, the incision needs to be along the line of the probe insertion as in Figure 8.6, rather than antiradial.

An 11 G probe is usually adequate for duct excisions unless there are widespread dilated ducts on ultrasound or it is a repeat procedure, after either a previous Mammotome or operative duct excision, when an 8 G probe might be considered.

The probe is inserted into the stab incision. It might be worth suggesting to the patient that she close her eyes or turn her head at this stage if she is anxious about the size of the needle. The needle is pushed down at an angle of about 30° to the skin, right under the nipple, under ultrasound vision. The aim is to position it under and parallel to the ducts on the opposite side of the nipple (Figures 8.8 and 8.9). It is also worth placing a finger on the skin of this part to gauge where the tip of the probe is located.
If a specific duct is being targeted, for example one with a papilloma, then the probe will have been inserted directly under this duct from the opposite side of the nipple so that it is parallel to the duct and the cutting groove is directly under the lesion. Samples are then taken, cutting upwards towards the duct, until the lesion and/or the duct have been completely removed ultrasonically. A few more samples are then taken to ensure that the duct is obliterated. If a papilloma is being targeted, it is important to aim at this first because if the probe cuts into the duct first, then the fluid in the duct will be sucked out by the vacuum and the papilloma may no longer be visible on ultrasound.

If the aim is a major duct excision, then one starts as above with the ducts directly opposite the insertion point, say at 3 o’clock in Figure 8.8. The probe is then turned 90° anti-clockwise and more excision takes place of ducts in the 1 to 2 o’clock region. The probe is then turned back the other way to excise ducts in the 4 to 5 o’clock region. The probe is then pulled back to excise the ducts directly under the nipple and those at 6 and 12 o’clock, and it is then pulled back further to complete the ducts on the side nearest the insertion point in the same way. It is not usually necessary to reposition the probe, as by just turning it to the side, different areas of tissue are sucked in by the vacuum and excised.

The major danger in this procedure is, when cutting upwards towards the nipple or areolar skin, that the vacuum sucks skin into the cutting groove and cuts a hole in the skin. When one knows that the skin is close, it is important to watch the skin over the probe and if it is sucked down, the cut can be aborted instantly by pressing the back button before the rotary cutting blade reaches the area. At worst the defect would need a suture to repair it.

**Subareolar biopsy before nipple sparing subcutaneous mastectomy**

This procedure is very similar to duct excision. The difference is that the aim is to adequately biopsy the subareolar area, without biopsying too deep into the area.
Interventional Ultrasound of the Breast

which is going to be excised by the subcutaneous mastectomy anyway. As a result, one needs to take the biopsies very close to the skin of the areola and there is a greater danger of sucking in and excising some of the skin. It is therefore important to watch the skin closely, but another technique is to insert the probe immediately under the skin of the nipple–areolar complex and then turn it through 180°, so that it cuts the biopsies downwards, rather than up towards the skin. Another important factor here, from the cosmesis point of view, is to make the entry point of the probe in a position where it will be excised by the incision for the subsequent mastectomy. An 11 G probe is adequate for this procedure, and after a look with ultrasound for any obvious ducts at the beginning, it can be mostly performed free hand. Indeed it is probably more useful to be able to feel where the probe is and to be able to watch the skin.

**Fibroadenoma**

The stab incision made to allow insertion of the Mammotome probe usually heals very well, but there will inevitably be some visible mark. It is therefore important to carefully plan where the insertion point is going to be made. It will depend on where the fibroadenoma is located, but the best two sites are either in the inframammary fold or, for lesions in the upper breast, the normal skin–areolar skin junction. Other considerations include being at least 3 or 4 cm away from the lesion so that the probe can be inserted reasonably parallel to the chest wall, and operator comfort.

Local anaesthesia is important with fibroadenomata. After the bleb is raised under the skin the larger anaesthetic needle can be used under ultrasound vision to develop a track from the incision down to a plane deep to the fibroadenoma. The importance of this is that these lesions are usually in young women with dense breasts in which it is hard to get a large needle into position, even with the cutting blade on the tip of the Mammotome. The hydraulic pressure of injecting the anaesthetic is surprisingly good at developing a path of least resistance for the probe. Likewise, the anaesthetic is useful in creating a space under the lesion for the probe, particularly if the lesion is on the chest wall, when it can be ‘lifted off’. Two further points worth making are that the anaesthetic needs to be infiltrated at least 5 mm beyond the fibroadenoma as the sharp tip of the probe will be in this part when the cutting part is under the fibroadenoma, and the anaesthetic needs to be infiltrated under the lateral parts of the lesion as well as in a fan-like pattern.

The probe, usually an 8 G, unless in the case of a very small fibroadenoma, is now placed in position. It is worth examining the area with colour Doppler ultrasound to see whether there are any large vessels around the fibroadenoma, but it is difficult to avoid these when one is excising the whole lesion. The Mammotome probe needs to be directly under the lesion (Figure 8.10). Despite the preparation with local anaesthetic, it can sometimes need quite a lot of force to get it into position and one may sometimes need to put the ultrasound probe down and hold the fibroadenoma with that hand to steady it whilst pushing the Mammotome probe underneath it with
the other hand. It can sometimes be difficult to see where the cutting groove is on ultrasound imaging, but if the cutting blade is withdrawn with the back button, an area of accentuation is seen just behind the groove (Figure 8.10).

The excision can now begin. One has to remember that the view with ultrasound is a two-dimensional slice of a three-dimensional object, i.e. all the work is being undertaken viewing the screen which shows only a slice of the lesion and there is more fibroadenoma on either side of this slice. For this reason one starts off by cutting large cores up through the centre of the fibroadenoma to the superficial surface and the lesion seems to have disappeared on ultrasound. However, moving the ultrasound probe to either side reveals the rest of the fibroadenoma. One can either reposition the probe under each of the two side pieces and excise them or more easily, rotate the probe to face the side piece and move the ultrasound probe laterally over it to view and excise it (Figure 8.11). The exercise is then repeated with the other lateral fragment. In practice this is a bit easier than it sounds, as the vacuum system sucks the fibroadenoma down into the groove before the blade cuts it off. After this process, there is usually a small fragment of fibroadenoma left behind somewhere and it may be necessary to reposition the probe. Again care needs to be taken, if the fibroadenoma is superficial, not to excise any skin.

**Cystadenoma**

The principle is the same as for excising a duct papilloma, with the same difficulties, in that neither the local anaesthetic needle or the Mammotome needle should enter the cyst before cutting commences, otherwise the cyst might deflate and the cystadenoma may no longer be visible on ultrasound. The Mammotome is inserted carefully
Interventional Ultrasound of the Breast

under the cyst directly under the adenoma (Figure 8.12). When in position, samples are cut upwards towards the adenoma through its base in the cyst wall. At some stage, the Mammotome will cut through the cyst wall and the vacuum will suck the fluid from the cyst, dramatically deflating it. Hopefully, by this stage the adenoma will have been excised, but it is probably worth cutting a few more samples to excise some more of the cyst wall.

Biopsy of B3 lesions and inflammatory breast cancer

The principle of biopsying these lesions with the Mammotome is the same as for ordinary core biopsy. The only real differences are that more local anaesthetic is required and the Mammotome needle is inserted under the lesion, rather than into it, to take samples cutting up into it.
Gynaecomastia

The principle with gynaecomastia is really a combination of the techniques used for excising fibroadenomata, with the safeguards needed when using the Mammatome directly behind the nipple. Local anaesthesia needs to be infiltrated quite widely as the aim is to contour the whole breast area, not to leave a scooped out dent by just removing the breast disk. The best entry point is probably in the submammary area laterally, using an 8 G probe. The breast tissue and fat are excised methodically in a fan-shaped pattern using a combination of ultrasound control and direct feel. The cutting side of the Mammatome should probably be facing towards the skin or side-ways at all times, to avoid excising pieces of the pectoral muscle. The Mammatome is probably less efficient at removing soft fat, and it may be necessary to complete the smooth contouring process with liposuction.

Fat necrosis

Fat necrosis can leave a persistent hard area in the breast and is common after major breast surgery such as reconstruction with autologous material or reduction mammoplasty as well as bruising following trauma. Techniques for excision of these areas with the Mammatome are similar to those of gynaecomastia, the aim being to remove the hard area and recontour it. One major caveat is not to excise the vascular bundle supplying an autologous reconstruction!

Abscess

Large abscesses in lactating women (Figure 8.13) are difficult to anaesthetise adequately for a Mammatome procedure. The full 20 ml should be used and care should

Figure 8.13  Lactating abscess which was successfully treated by Mammatome.
be taken not to get too much actually into the abscess cavity as it does not have much
effect there. The entry point needs to be carefully sited, not just for cosmesis, but well
away from the abscess and not through inflamed skin. Going through several cen-
timetres of normal breast tissue, before entering infected tissue, stops the formation
of a track draining pus afterwards. The Mammotome is pushed into the abscess
cavity and vacuum can then be applied (Figure 8.14). The Mammotome system has
several modes, but the author has found that the best mode for maximum suction is
in ‘position’ mode with the rotary blade positioned halfway down the probe. Vacuum
is then applied by pressing the ‘vac’ button. The Mammotome tip often needs to be
moved around in the cavity to evacuate all the pus and then biopsies can be taken of
the cavity wall using ‘sample’ mode, as a very small number of even lactating
abscesses can be malignant.

The Mammotome can also be used for chronic non-lactating abscesses around the
nipple, including those with chronic fistulae and those which have had operative
drainage procedures done before. The technique is again to site the entry point well
away from the infected area, and to excise as much of it as possible, going up as close
to the skin as possible.

**Postprocedure**

Care needs to be taken withdrawing the Mammotome from the entry point as it has
a very sharp bladed tip and it would be easy to damage the edges of the skin stab inci-
sion. Pressure needs to be applied quite strongly with several gauze swabs over the

![Mammotome needle in abscess evacuating pus.](image)

**Figure 8.14** Mammotome needle in abscess evacuating pus.
area where the biopsy/excision has been performed and the incision and needle track, for a timed 5 minutes (Figure 8.15). This is very important in terms of minimising the amount of haematoma and bruising. Closure of the incision is made with a Steri-Strip; one is usually sufficient (Figure 8.16). The patient is then sat up and the pressure bandage applied. One needs to be careful at this stage as patients often feel faint in the upright position. A folded gauze swab pad is placed over the biopsy/excision area and held in place by the bandages. The bandages should be as tight as is comfortable for the patient and she is instructed to leave them on overnight (Figure 8.17).

The patient is kept in the unit for 30 minutes after the procedure to check for any immediate problems and then allowed home with an appointment for the following week to receive the results. We recommend that someone else drives them home.

Figure 8.15  Applying pressure after Mammotome has been withdrawn.

Figure 8.16  Closure of incision with a Steri-Strip.
Interventional Ultrasound of the Breast

Complications are few and minor. Most patients will get some degree of bruising and some will have a palpable haematoma, but in the approximately 1000 procedures performed in our unit, we have never had to perform a surgical evacuation of a haematoma. We have also never had a problem with nipple necrosis or any significant infection.

REFERENCES


Figure 8.17  Applying pressure bandage after procedure.
9

Future developments

Obi Iwuchukwu and Philip Drew

THREE-DIMENSIONAL ULTRASOUND

Three-dimensional (3D) ultrasonography is rapidly gaining popularity as it moves out of the research environment and into the clinical setting. This modality offers several distinct advantages over conventional ultrasonography. Conventional radiography merges 3D data into two-dimensional (2D) summation images. Radiologists mentally reverse this process by forming 3D impressions of the underlying anatomy and disease. Tomographic data from ultrasonography, computed tomography (CT) and magnetic resonance (MR) imaging have made this mental reformatting process easier and more accurate.

One of the major disadvantages of conventional ultrasound is operator dependency. The operator sweeps the ultrasound beam back and forth across an organ many times while mentally integrating multiple 2D images into a 3D impression of the underlying anatomy and disease. This process is universally acknowledged as time-consuming and inefficient, and there is considerable interobserver variability. In contrast, 3D images can be reconstructed from data obtained with a single sweep of the ultrasound beam across the involved organ. Both the ultrasonography information and the relative position of each tomographic section are accurately recorded. As a result, the exact relationship between anatomic structures is accurately recorded in the 3D image.

Conventional ultrasonography is very useful for distinguishing breast cysts, but 3D ultrasonography has the capacity to demonstrate lesion margins and topography. This information can be invaluable in differentiating benign from malignant masses. The breast tissue must be immobilised during imaging to ensure accurate registration of the imaging sections.

Cho et al. compared prospectively obtained static 2D and 3D ultrasonographic images for the diagnostic performance of radiologists with respect to the differentiation of benign from malignant solid breast masses with histopathological examination as the reference standard. Conventional 2D and 3D ultrasonography images were obtained from 141 patients with 150 solid breast masses (60 cancers and 90 benign lesions) before excision or needle biopsy. Four radiologists who had not performed the examinations independently reviewed 2D ultrasonography images and stored 3D ultrasonography data, and provided a level of suspicion concerning probability
of malignancy. The sensitivity, specificity and negative predictive values of 2D images were then compared with those of 3D ultrasonography images. For all readers, 3D ultrasonography images were the same as or better than 2D ultrasonography images in terms of sensitivity, although the differences were not statistically significant ($p > 0.05$).

The development of an innovative 3D visualisation and navigation tool for use before, during and after the procedure provides surgeons with a tool to give additional insight into the physical configuration of the surgical site and the requirements of the procedure, in turn enabling more effective surgery.

**Diagnostic applications**

The use of 3D ultrasound may help to determine the exact position of the needle during breast biopsy, thereby reducing the number of core samples that are needed to achieve a reliable histological diagnosis. 3D ultrasonography has already been used by several groups in a diagnostic capacity and found to be superior to 2D ultrasonography.

Lell and colleagues evaluated 3D ultrasound in core biopsies of suspicious breast lesions. Biopsies were taken from 59 probably benign and 48 malignant lesions. 3D ultrasonography revealed 61 central, 44 eccentric and two marginal needle positions after the initial 2D guidance. Central repositioning of the needle was achievable under 3D guidance in all patients. In one lesion, the carcinoma was missed in the core biopsy, despite central hits of the biopsy needle. This led to a sensitivity of 98%, specificity of 100%, positive predictive value of 100%, negative predictive value of 98.5% and accuracy of 99% for the diagnosis of a malignant lesion. Thus, 3D ultrasonography improves needle positioning as well as the depiction of correct needle placement in free-hand core biopsy.

Weismann et al evaluated the role of 3D ultrasound following needle breast biopsy under 2D needle guidance. A total of 188 core needle biopsies and 24 fine needle aspiration biopsies were 3D ultrasonography correlated after typical ‘free-hand’ ultrasonography needle guidance. After core needle stroke or localisation of fine needle, a 3D ultrasonography data volume set was acquired and a multiplanar analysis performed. In 117 core needle strokes of benign ($n = 21$) and malignant ($n = 12$) lesions 3D targeting prospectively revealed 95 lesion hits, 12 marginal lesion hits and nine out-of-lesion hits. In one case after the initial large core needle path a 5-mm lesion was disguised by air bubbles, therefore 3D targeting failed during the second biopsy procedure. Thus, 3D ultrasonography combined with 3D targeting technique is a more reliable and objective tool, demonstrating exact spatial positioning of core and fine needle during biopsy procedure when compared with standard 2D ultrasound.

Furthermore, Delle Chiaie and Terinde in a separate study demonstrated the efficacy of 3D ultrasound-validated large-core needle biopsy (LCNB) of the breast. 3D ultrasound visualisation of the needle in the postfiring position was used to classify the biopsy as central, marginal or outside the lesion. Based on this classification it was decided whether another sample had to be obtained. A total of 360 core needle
biopsies were obtained from 169 breast lesions in 146 patients. A median of two core samples per lesion provided for all the lesions a sensitivity for malignancy of 96.9%, specificity of 100%, false positive rate of 0% and false negative rate of 3.1%, and for the excised lesions a sensitivity of 96.5%, specificity of 100%, false positive rate of 0%, false negative rate of 3.5% and an underestimation rate of 3.4%. Delle Chiaie and Terinde concluded that 3D ultrasound validation of the postfiring needle position is an efficient adjunct to ultrasound-guided LCNB. The advantages of 3D ultrasound validation are likely to include a reduction in the number of core samples needed to achieve a reliable histological diagnosis (and a possible reduction in the risk of tumour cell displacement), reduced procedure time and lower costs.

The 3D ultrasound navigation system is useful in visualizing breast tumour extension and is more accurate than palpation. The system is expected to be helpful in deciding on the appropriate surgical margin in breast cancer surgery, resulting in a better cosmetic outcome without increasing the risk of surgical margin positivity.

To perform optimal tumour resection of breast cancer, preoperative information concerning intraductal spread of cancer (ISC) is very important. Tamaki et al. evaluated the sensitivity and specificity of three different 3D modes including helical CT, MRI and ultrasound in patients with primary breast cancer by comparison with multisliced pathological specimens. The sensitivity of each modality for detecting ISC was 64.7%, 90.2% and 78.6%, and the specificity was 97.1%, 62.9% and 100%, respectively. Thus, 3D ultrasonography was the most specific modality for diagnosing ISC.

Meyberg-Solomayer and colleagues evaluated whether 3D ultrasound brings any advantage to breast diagnostics. A total of 65 women with breast lesions (42 malignant, 23 benign) were examined preoperatively with 2D and 3D ultrasonography. It was found that 3D ultrasonography of breast lesions as an adjunct to 2D sonography offers a better assessment of the infiltrative zone. In addition, the coronal plane was of benefit when the infiltrative zone was unclear. Furthermore, the surface mode was advantageous for imaging complex structures such as a multinodular fibroadenoma.

**Real-time three-dimensional ultrasound image guidance for breast surgery**

During breast cancer surgery the primary goal of the surgeon is to remove the entire cancerous tumour as well as a small margin of surrounding healthy tissue to ensure that no cancerous cells are left behind. In general, the surgeon operates without any image feedback during breast cancer surgery. Recently, research has provided surgeons with an ultrasound-based 3D visualisation and navigation tool that can be used prior to, during and after the procedure to ensure successful removal of the cancer. Such an advanced imaging system will increase the likelihood of success by assisting with the surgeons’ primary tasks: locating the target tumour prior to the operation, planning and executing an intervention along a minimally invasive path, removal of
the cancerous tissue, and evaluation of the region after surgery to ensure that no can-
cerous tissue remains.

A prototype free-hand 3D ultrasound system that uses a readily available commercial 2D ultrasound system for image acquisition, an optical tracking system for probe and surgical instrument tracking, and a personal computer for 3D image reconstruction and visualisation in real time has been developed by workers at Stamford University, USA. It is prepared for use in the operating room.

Baez et al evaluated the use of 3D ultrasonography in the complete excision of benign breast tumours using ultrasound-guided vacuum-assisted core needle biopsy (Mammotome®). Twenty consecutive patients with sonographically benign breast lesions underwent 3D ultrasound-guided Mammotome biopsy under local anaesthesia. Ultrasonographic follow-up examinations were performed on the following day and 3–6 months later to assess residual tissue and scarring. Follow-up examinations revealed complete excision of all lesions of <1.5 ml in volume as assessed by 3D volumetry. 3D ultrasonographic volume assessment was more accurate than 2D. No bleeding or infections occurred postoperatively and no significant scarring was seen ultrasonographically on follow-up examinations. Thus, 3D ultrasound offers the advantage of better preoperative demonstration of the lesions’ margins, resulting in better assessment of volume, improved intraoperative needle location and perioperative identification of residual tumour tissue. Furthermore, it was recommend that 3D sonographically guided biopsy should be integrated into breast cancer screening programmes as a safe therapeutic option for breast lesions presumed to be benign.

**Limitations of three-dimensional ultrasonography**

All of the current 3D ultrasonography data acquisition techniques are more cumbersome than conventional ultrasonography techniques. Most involve adding a localising technology or a mechanical motor to the transducer. The resulting assemblies are typically bigger and therefore more difficult for the user to manipulate. In addition, the data acquisition software usually requires considerably more user input than is required with conventional ultrasonography. The larger data sets produced with 3D ultrasonography also make data archiving and communication more challenging.

Different computer algorithms reconstruct data into 3D volumes at different speeds. Some produce images almost instantaneously, whereas others require several seconds to produce an on-screen image. Waiting for the 3D image to appear can be frustrating for users who are accustomed to having ultrasonography images appear immediately.

The ability to view data with a variety of algorithms and from different perspectives may slow the image interpretation process. Inexperienced users may have to spend extra time finding the best algorithm and perspective for viewing the data. Many of the viewing programs require a considerable amount of image manipulation to obtain high quality results.
Two separate groups have evaluated the diagnostic accuracy of 3D ultrasound in comparison with conventional 2D ultrasound in the characterisation of breast lesions. Ultrasound features on 3D ultrasound differ significantly from those on 2D ultrasound. However, the diagnostic accuracy of both methods is almost identical. 3D ultrasound as an adjunct to conventional 2D ultrasound should be evaluated in larger trials to determine its clinical value in breast imaging.10

Cho et al11 found 3D imaging superior to 2D imaging in terms of lesion contrast and characterisation of masses. This superiority was more apparent in masses which were 10 mm or larger. However, diagnostic accuracy, including sensitivity, specificity, positive predictive value, negative predictive value and false negative rate, for diagnosis of breast cancer of 3D imaging was not different from that of 2D imaging. Cho and colleagues concluded that in spite of superior image quality on 3D ultrasonography, it does not provide additional benefits to diagnostic accuracy for diagnosis of breast cancer.

REFERENCES

8. Image 2006; 39.
10

Future developments: focused ultrasound ablation

Obi Iwuchukwu and Philip Drew

The ideal treatment of localised cancer should cause the complete death of tumour cells without damage to surrounding normal tissue. Focused ultrasound ablation (FUS) is such a potential treatment which can induce complete coagulation necrosis of a targeted tumour, at depth, through the intact skin.

FUS relies on the same principles as conventional ultrasound. It can propagate harmlessly through living tissue, but if the ultrasound beam carries sufficient energy and is brought into a tight focus, the energy within the focal volume can cause a local rise in temperature of sufficient magnitude to cause tissue necrosis. This occurs without damage to surrounding or overlying tissues. High energy FUS destroys cells by denaturation of cell proteins. The absorbed energy results in extremely high tissue gradients between target cells and surrounding tissue, so the effect of the focused energy is concentrated only at the target, leaving the healthy tissue unscathed.

Although breast conservation treatment (BCT) is not a major procedure, it is invasive and cosmetically undesirable in some patients. Furthermore, some patients with significant comorbidities, or elderly patients, are precluded from surgery primarily due to unsuitability for general anaesthesia. A completely non-invasive local therapy would require no anaesthesia, would reduce recovery time, could avoid infections and scar formation, and possibly also reduce cost.

This ideal of non-invasive therapy can be realised, if two prerequisites are fulfilled: imaging technology to provide accurate information on the exact anatomy of the tumour and the surrounding healthy tissue, and the precise delivery of energy to the target. Recently, a variety of minimally invasive therapies have been applied to breast tumours including interstitial laser coagulation, radiofrequency, cryotherapy and interstitial radiotherapy. These techniques, however, still require a probe to be inserted, but percutaneous focused ultrasonic waves have the potential to very precisely deliver energy to a given point in soft tissue within an accuracy of 1 mm through the intact skin. The ultrasound energy can induce temperature elevations of 55–90°C at the focal spot in less than 10 s and instantaneously induce cellular death and vascular obliteration in normal and tumour tissue.¹

The first work to consider potential applications of FUS was published in 1942.² Further, more detailed properties of focused ultrasound conduction and modes of
Interventional Ultrasound of the Breast

destruction in normal tissues were investigated the 1970s and 1980s,\textsuperscript{3,4} and studies using FUS to treat experimental tumours followed.\textsuperscript{5,6}

The FUS system provides a safe, repeatable treatment approach for benign tumours (e.g. uterine fibroid and prostatic lesions)\textsuperscript{7–10} that do not require an aggressive approach. Magnetic resonance (MR)-guided FUS can also be used for debulking cancerous tissue.\textsuperscript{11,12}

MR-guided FUS offers an attractive alternative to conventional surgery because it incorporates intraoperative MR imaging, which provides far more precise target definition than is possible with the surgeon’s direct visualisation of the lesion. MR-guided FUS is undeniably the most promising interventional MR imaging method in the field of image-guided therapy today.\textsuperscript{11} It is applicable not only in the thermal coagulative treatment of tumours, but also in several other medical situations for which invasive surgery or radiation may not be treatment options.

Although use of FUS has been appraised over the years its widespread use was hampered by difficulty with controlling focal spot position, precise target definition and beam dosimetry. However, the technical feasibility of performing MR imaging-guided FUS by using MR imaging to guide and monitor therapy has been established\textsuperscript{14,15} and demonstrated by several research groups.\textsuperscript{16–18}

In MRI-FUS the patient lies on a specialised table with an MRI scanner. Anxiolytics help reduce patient induced movement. No incision is necessary, but the transducer is positioned such that the ultrasound beam is focused directly on specific positions within the tumour, thus outlining its margins accurately. FUS treatment comprises a series of sonications inside the focal point area, thus encompassing the tumour.\textsuperscript{18}

Hynynen and co-workers\textsuperscript{14} first showed that MR imaging-guided FUS could be performed to non-invasively coagulate benign breast fibroadenomas. They treated 12 benign tumours, five of which required further treatment due to initial use of a low power. They also encountered problems when local anaesthetic was liberally injected anterior to the tumour requiring treatment because the resultant microscopic bubbles in the local anaesthetic injected in front of the fibroadenoma probably caused scattering of the ultrasound beam thus limiting the power delivered to the tumour. Two further treatment failures were due to patient motion during treatment with consequent delivery of power to the pectoralis major.

Human studies

Furusawa et al\textsuperscript{19} recently published their experience with FUS. Thirty women with biopsy-proved breast cancer underwent MR-FUS treatment. Gadolinium-enhanced MR images were used for treatment planning and post-treatment radiological assessment of treated tissue. After MR-FUS, all 30 women underwent wide excision or mastectomy. The extent of thermal ablation was assessed with tumour histology. In general, treatment was well tolerated, with a minimum of adverse effects, especially when performed under local anaesthesia. On pathological examination, mean (\pm SD) necrosis of the targeted breast tumours was 96.9 \pm 4\% (median 100\%, range 78–100\%).
of tumour volume. Fifteen (53.5%) of 28 patients had 100% necrosis of the ablated
tumour; only three patients (10.7%) had less than 95% necrosis. In 28 (93.3%) patients,
100% of the malignancy was within the treatment field, and 98% and 95% of tumour
lay within the treatment field in two remaining patients. Retrospective analysis in
two patients with residual tumour showed treatment was not delivered to the full
recommended area, reaffirming the need for precise localisation and the value of
contrast-enhanced images for treatment planning. MR-FUS has great potential to
become a viable non-invasive replacement for lumpectomy. The unit is currently
addressing studies focusing on post-treatment image-based evaluation.

Wu and colleagues20 in a non-randomised prospective trial investigated the safety,
efficacy and feasibility of using FUS as a non-invasive treatment for patients with
breast cancer. Twenty-two patients with breast cancer were enrolled into this study.
Disease tumour, node, metastases (TNM) stage was classified as stage I in four patients,
stage II(A) in nine patients, stage II(B) in eight patients and stage IV in one patient.
Tumour size ranged from 2 to 4.8 cm in diameter (mean 3.4 cm). All patients received
chemotherapy, radiation and tamoxifen, following FUS for the primary lesions.
Outcome measures included radiological and pathological assessment of the treated
tumour, cosmesis and local recurrence. No severe complications were encountered
after FUS. Postoperative imaging demonstrated positive response and regression of
all treated lesions. Follow-up biopsy revealed coagulation necrosis of target tumour
and subsequent replacement by fibroblastic tissue. After a median follow-up of 54.8
months, one patient had died, one was lost to follow-up, and 20 were still alive.
Two of 22 patients developed local recurrence. Five-year disease-free survival and
recurrence-free survival were 95% and 89%, respectively. Cosmetic result was judged as
good to excellent in 94% of patients.

The largest series of FUS comes from China, Wu et al21 reviewed a total of 1038
patients with solid tumours who underwent FUS ablation in China. Pathological
examination showed that the target region presented clear evidence of cellular
destruction. Small blood vessels less than 2 mm in diameter were severely damaged.
Follow-up diagnostic imaging revealed that there was no, or reduced, blood supply,
and no uptake of radioisotope in the treated tumour after FUS, indicating both a pos-
tive therapeutic response and an absence of viable tumour. Imaging at 6–12 months
showed obvious regression of the lesion. Following 4 years of follow-up they noted an
extremely low major complication rate.21

Zippel and Papa22 in a phase one trial, examined the possibility of ablating breast
carcinoma using MRI-FUS. Ten women underwent the procedure; 7–10 days after
the procedure, all patients underwent standard lumpectomy and axillary sampling to
complete standard treatment and to allow pathological evaluation of the procedure.
Two patients had a complete pathological response. The remaining eight patients had
varying amounts of residual tumour: two had microscopic foci of residual carcinoma,
three had 10% residual tumour and three had 10–30% of residual tumour. Further work
is currently ongoing to standardise the optimum energy levels required.

Wu et al23 in a separate study explored the possibility of using FUS to treat patients
with localised breast cancer in a controlled clinical trial. A total of 48 women with
biopsy-proven breast cancer (T(1–2), N(0–2), M0) were randomised to the control group in which modified radical mastectomy was performed, and the FUS group in which an extracorporeal FUS ablation of breast cancer was followed by modified radical mastectomy. Short-term follow-up, pathological and immunohistochemical stains were performed to assess the therapeutic effects on tumour and complications of FUS. The results showed that no severe side-effects were observed in the FUS-treated patients. Pathological findings revealed that FUS-treated tumour cells underwent complete coagulative necrosis, and tumour vascular vessels were severely damaged. Immunohistochemical staining showed that no expression of proliferating cell nuclear antigen (PCNA), matrix metalloproteinase (MMP)-9 and CD44v6 was detected within the treated tumour cells in the FUS group, indicating that the treated tumour cells lost the abilities of proliferation, invasion and metastasis. Furthermore, it was demonstrated that there were significant alterations in expression of PCNA, CD44v6, MMP-9 and erbB2 mRNA, and a dramatic decrease in telomerase activity in the FUS group. Thus malignant tumour characteristics are demonstrably arrested by FUS and biological factors are potential markers for assessing FUS efficacy.

In support of above biological effects of FUS, Huber and colleagues demonstrated that a consequence of the FUS therapy is homogeneous lethal and sublethal tumour damage with subsequent up-regulation of p53 and loss of proliferative activity. Gianfellice et al. evaluated the feasibility of treating breast neoplasms with use of MR-FUS. Twenty-four female patients, each with a single biopsy-proven breast carcinoma, who were considered to be at increased surgical risk or who had refused surgery underwent MR-FUS surgery as an adjunct to tamoxifen therapy. Follow-up included routine studies to rule out metastatic disease and MR studies with and without contrast material infusion in the treated breast, 10 days and 1, 3 and 6 months after MR-FUS. Percutaneous biopsy was performed after 6 months of follow-up, and if residual tumour was present, a second MR-FUS surgery treatment session was performed, followed by repeat biopsy 1 month later. Twenty-three of 24 patients completed the protocol, with only one minor complication associated with the treatment sessions (second-degree skin burn which resolved with local treatment). Follow-up MR studies demonstrated a varying hypointense treatment margin (range 1–11 mm), which represents destruction of tissue beyond the visible tumour. Absence of enhancement indicated destruction of tumour (18 of 19 patients with negative biopsy results), whereas persistent enhancement was due to residual tumour (three of five patients with residual tumour). Overall, 19 of 24 patients (79%) had negative biopsy results after one or two treatment sessions. They demonstrated satisfactory correlation of contrast MRI and tissue necrosis in a substudy.

CONCLUSION

MRI-guided FUS therapy has the potential to replace open surgery in breast cancer patients in carefully selected cases, such as patients with high operative risks, elderly
patients, or patients not willing to undergo surgery. Several groups have independently assessed the safety and efficacy of FUS, and found it safe, effective and feasible for patients with breast cancer. But, large-scale multicentre clinical trials will be needed to determine the future role of this novel modality.

REFERENCES

Index

Page numbers in *italics* refer to tables and figures.

3D ultrasound see three-dimensional (3D) ultrasound
ABBI (Advanced Breast Biopsy Instrumentation System) 89
abscesses see breast abscesses
accessory nipples 11, 12, 13
amastia 11
amazia 11
amplitude (A) scan 24
anatomy of the breast 11–22
blood supply and venous drainage 14, 16
ducts 15–16, 16, 17, 18
lobes and lobules 14–15, 17
lymphatic drainage 16–21, 20
muscles 12–14, 15
nerve supply 14
suspensory ligaments 11, 14
angiosarcomas 55
arteries supplying the breast 14, 16
atypical ductal hyperplasia (ADH) 66
atypical lobular hyperplasia (ALH) 66–7
axilla 17–18, 19
axillary lymph nodes
anatomy 18, 20, 20
classification of appearances 57
sentinal node biopsy 78
ultrasound appearance of malignancy 56, 57, 58
ultrasound-guided sampling 78, 87
B3 lesions, Mammotome biopsy 92, 102
beam size
and contrast resolution 27, 28
effect of focusing on 28–9, 29
benign breast lesions, ultrasound appearance
  cysts 47–8, 47, 48
  diabetic fibrous mastopathy 50
  fat necrosis 49, 54
  fibroadenomas 45–6, 46, 54
  general 45
  lesions mimicking cancers 54
  lipomas 49, 49
  phyllodes tumours 48
  radial scars/complex sclerosing lesions 50
  sebaceous cysts 49–50
  biopsy guns 69, 70, 84
  blood supply to the breast 14, 16
  breast abscesses
differentiation from cancer 52
lactating breast 56, 103
Mammotome excision 92, 103–6, 104, 105, 106
ultrasound-guided drainage 82
breast cancer
core biopsy results 68–9, 68, 69
metastatic spread 21
MRI-guided FUS therapy 116–19
screening for 4–5
breast cancer, ultrasound appearance
  axillary lymph nodes 56, 57, 58
  ductal carcinoma in situ (DCIS) 50, 55, 55
  less common appearances
    apparently well-defined mass 51–2, 52
    diffuse oedema 53
    enlarged axillary nodes 54
    lesion resembling an abscess 52
    papillary lesions 53
    predominantly bright mass 53, 53
    round/finely irregular mass 52
  lymphomas 56
  male breast 59
  metastases 56
  recurrent disease 59, 60
  sarcomas 55
  typical appearances 50, 51
breast clinics 5
breast lesions, benign see benign breast lesions,
  ultrasound appearance
breast surgery
  3D ultrasound guidance 111–12
  diagnostic 5–6
chemotherapy, microbubble delivery of 60
clinical assessment 64
complex sclerosing lesions 50, 68
computed tomography (CT) 3, 111
congenital abnormalities associated with the breast 11, 13
continuing medical education 38
contrast enhanced mammography 3
core biopsy 66–71
accuracy 70
axillary lymph node 87
biopsy guns 69, 70, 84
complications 72–3
histological classification 66–8, 67, 68, 69
information provided by 69
intracystic papilloma investigation 47
procedure 70–1
stereotactic 7, 74–6, 75
ultrasound-guided 7, 73, 74, 82–5, 84
depth compensation 29–30, 30, 31
depth control 30, 31
development of the breast 11, 12
diabetic fibrous mastopathy 50
diffuse oedema of the breast 53
digital mammography 3
diagnostic accuracy 111
fibroadenomas
histological image 67
fat content of the breast 45
fat necrosis
differentiation from cancer 54
Mammotome excision 93, 103
ultrasound appearance 49
fibroadenomas
histological image 67
Mammotome biopsy/excision
indications 91
procedure 100–1, 101
ultrasound appearance 45–6, 46, 54
fibrous tissue 45
fine needle aspiration cytology (FNAC)
axillary lymph nodes 87
factors affecting accuracy 65
procedure 65–6
result classification 64, 65
stereotactic 6–7, 73–4, 75
ultrasound-guided 73, 74, 82, 87
see also three-dimensional (3D) ultrasound
focus control 30–1, 31
focused ultrasound (FUS) ablation 115–20
human studies 116–19
magnetic resonance-guidance 116
principle 115
freeze control 30, 31
frequency 23, 44
galactoceles 47
Gershon-Cohen, J. 1
glandular tissue 45
gynaecomastia
Mammotome treatment 92, 103
ultrasound appearance 58–9
haematoma, postoperative 59
harmonic imaging 60
helical CT 111
histology of the breast 14–16, 16, 17
history of breast diagnosis 1–9
inflammatory breast cancer, Mammotome biopsy 93, 102
Ingleby, Helen 1
internal thoracic nodes 20–1, 20
International Breast Ultrasound School (IBUS)
training recommendations 38
intracystic papillomas 47, 48
intraductal spread of cancer (ISC) detection 111
lactating breast
abscesses 56, 103
Mammotome excision 92, 103–6, 104, 105, 106
fibroadenomas 46, 46
galactoceles 47
lactiferous ducts 11, 12, 15, 16, 17, 18
large-core needle biopsy (LCNB), 3D ultrasound-validated 110–11
linear array scanning 25–6, 26
lipomas 49, 49
liposarcomas 55
lobes 14–15
lobular carcinoma in situ (LCIS) 66–7
lobules 15, 17
lymphatic drainage of the breast 16–21, 20
lymphomas 56
magnetic resonance imaging (MRI)
biopsy guidance 6
breast cancer detection 4
focused ultrasound (FUS) ablation guidance
   116–19
future roles 4
intraductal spread of cancer 111
male breast 58–9
mammary buds 11, 12
mammary lines 11, 12, 13
mammary pits 11, 12
mammography see X-ray mammography
Mammotome 71, 89, 90
   see also ultrasound-guided vacuum-assisted biopsy
mastectomy 63
metastases in the breast 56
metastatic spread of breast cancer 21
microbubble contrast agents 60
microcalcification, core biopsy 69, 75, 75, 85
minimally invasive therapies 115
mucinous carcinoma 52
muscles associated with the breast 12–14, 15
National Health Service Breast Screening Programme (NHS BSP) 5
needle guidance principles
   needle visualisation 32, 33
   scan volume 32–3, 33
nerve supply to the breast 14
neurofibromas 49
nipple discharge, Mammotome biopsy/excision indications 91
   procedure 95–9, 96, 97, 98, 99
nipples
   accessory 11, 12, 13
   anatomy of 15, 17, 18
   nipple sparing subcutaneous mastectomy, subareolar biopsy 91–2, 99–100
   normal breast, age and tissue composition 45
Objective Structured Clinical Examination (OSCE) 38
open biopsies 76–7
overall gain control 30, 31
papillomas
   intracystic 47, 48, 53
   intraductal 53, 68
   peri-ductal mastitis 58, 59
   phyllodes tumours 48, 67–8
pneumothorax during core biopsy 73
polymastia 11, 13
polythelia 11, 12, 13
postmenopausal breast
   fibroadenomas 46, 54
   tissue composition 45
pseudoangiomatous stromal hyperplasia 54
pulse echo principle 24, 25
quadrants of the breast 16–17, 18, 19
radial scars/complex sclerosing lesions 50, 68
reflections from breast tissues 26–7, 28
Royal College of Radiologists training requirements 37
safety of diagnostic ultrasound 40
Salomon, A. 1
sarcoidosis 54
sarcomas 55
sclerosing lymphoepithelial lobulitis 50
screening for breast cancer 4–5
sebaceous cysts 49–50
SenoCore 360 89
Senograph 2
sentinal lymph nodes 21
sentinal node biopsy, axillary nodes 78
seromas, postoperative 59
skin recurrence of breast cancer 59, 60
staging codes for the breast 19
stereotactic breast biopsy 6–7, 73–6, 75
subareolar biopsy 91–2, 99–100
superficial fascia 11
surgery
   3D ultrasound guidance 111–12
   diagnostic 5–6, 76–7
   reconstructive 76
suspensory ligaments 11, 14
teaching aids 39
three-dimensional (3D) ultrasound 109–13
   advantages over conventional techniques 109–10
   diagnostic applications 110–11
   limitations 112–13
   real-time guidance for breast surgery 111–12
   tissue-mimicking phantoms 39, 40
tomosynthesis 3
training in breast ultrasound
   assessment and accreditation 37–8, 41
   continuing medical education/professional development 38
   knowledge of quality assurance 40
   patient safety management 40
   teacher training 38–9
Index

training in breast ultrasound (Continued)
  teaching aids 39
  training provider accreditation 41
  training requirements 35–7, 37, 39, 44
  triple assessment clinics 63–4
  triple diagnosis method 5
  tuberculosis 58

ultrasound
  3D see three-dimensional (3D) ultrasound
    beam size
      and contrast resolution 27, 28
      effect of focusing on 28–9, 29
    breast image optimization 43–4
    depth compensation 29–30, 30
    development of breast imaging 3–4
    echo size and tissue type 26, 27
    elastography 60
    equipment requirement guidelines 39
    frequencies 23, 44
    lateral resolution 29
    needle guidance principles
      needle visualisation 32, 33
      scan volume 32–3, 33
    operator dependency of technique 109
    properties useful in breast imaging 24
    pulse characteristics 24
    pulse echo principle 24, 25
    reflections from breast tissues 26–7, 28
    sensitivity and specificity 43
    see also benign breast lesions, ultrasound appearance; breast cancer, ultrasound appearance
  ultrasound-guided biopsy 7, 79–88
    equipment 79
    explanation and consent 80
    local anaesthesia 80
    multidisciplinary team approach 80
    procedures 73, 79
      abscess aspiration 82
      axillary lymph node sampling 87
      core biopsy 82–6, 84, 86
      cyst aspiration 81, 81
      fine needle aspiration cytology 82, 82
      preoperative localisation 87
      room layout 79, 80
      technique 81
    ultrasound-guided vacuum-assisted biopsy 89–107
      3D ultrasound guidance 112
      anaesthesia 90, 94–5
      anticoagulated patients 90
      biopsy systems 71, 89, 90
      discussion and informed consent 93
      indications and procedures
        abscesses 92, 103–6, 104, 105, 106
        B3 lesions 92, 102
        cystadenomas 92, 101–2, 102
        fat necrosis 93, 103
        fibroadenomas 91, 100–1, 101
        gynaecomastia 92, 103
        inflammatory breast cancer diagnosis 93, 102
        nipple discharge 91, 95–9, 96, 97, 98, 99
        subareolar biopsy 91, 99–100
        positioning and room set up 93–4, 94
        preparation and equipment 94
    ultrasound scanners
      amplitude (A) scans 24
      basic principles 24
      calibration 40
      display screen 25
      operator controls 30–2, 30
      real-time linear array scanning 25–6, 26
      recommended features 39
      safe use 40
    Vacora system 89
    vacuum-assisted core biopsy 7, 71–2, 71, 72
      see also ultrasound-guided vacuum-assisted biopsy
    venous drainage of the breast 14, 16
  WHO curriculum for theoretical training in ultrasound 36
  wire-guided excision biopsy 77, 77

X-ray mammography
  in assessment and diagnosis 5
  biopsy guidance 6–7, 73–6, 75
  contrast enhanced 3
  digital 3
  history
    breast cancer screening 4–5
    early work 1–2
    modern era 2–3
  young breast
    fibroadenomas 45, 46
    tissue composition 45
  zoom control 32
INTERVENTIONAL ULTRASOUND OF THE BREAST

Edited by Phillip Drew • Simon Cawthorn • Michael Michell
With a Foreword by Erika Denton

Over the past decade interventional radiology techniques have replaced surgical and clinically guided diagnostic procedures such that the vast majority of patients have a non surgical diagnosis of a breast problem. Interventional Ultrasound of the Breast, written by a collection of well-established and respected authors from the UK’s leading cancer centres, is an invaluable reference for surgeons, radiologists and all other members of the team involved in the diagnosis and management of breast disease.

Contents
A history of breast diagnosis • Anatomy of the breast • Technology for surgeons • Training and accreditation • Ultrasound appearance of benign and malignant breast lesions • Preoperative diagnosis and biopsy guns • Techniques of ultrasound biopsy • Techniques of ultrasound-guided vacuum-assisted biopsy • Future developments • Future developments: focused ultrasound ablation

About the Editors
PHILIP DREW MD FRCS is Professor of Tissue Engineering and Repair at The Hull York Medical School, University of Hull, Hull, UK

SIMON CAWTHORN MS FRCS is Consultant Breast Surgeon at the Frenchay Hospital, Bristol, UK

MICHAEL MICHELL MA MB BCir FRCR is Clinical Director of the Breast Screening Training Centre, King’s College Hospital, London, UK

informa healthcare
www.informahcare.com
Review

What is ultrasound?

Timothy G. Leighton

Institute of Sound and Vibration Research, Southampton University, Highfield, Southampton, SO17 1BJ, UK

Available online 15 August 2006

Abstract

This paper is based on material presented at the start of a Health Protection Agency meeting on ultrasound and infrasound. In answering the question ‘what is ultrasound?’, it shows that the simple description of a wave which transports mechanical energy through the local vibration of particles at frequencies of 20 kHz or more, with no net transport of the particles themselves, can in every respect be misleading or even incorrect. To explain the complexities responsible for this, the description of ultrasound is first built up from the fundamental properties of these local particle vibrations. This progresses through an exposition of the characteristics of linear waves, in order to explain the propensity for, and properties of, the nonlinear propagation which occurs in many practical ultrasonic fields. Given the Health Protection environment which framed the original presentation, explanation and examples are given of how these complexities affect issues of practical importance. These issues include the measurement and description of fields and exposures, and the ability of ultrasound to affect tissue (through microstreaming, streaming, cavitation, heating, etc.). It is noted that there are two very distinct regimes, in terms of wave characteristics and potential for bioeffect. The first concerns the use of ultrasound in liquids/solids, for measurement or material processing. For biomedical applications (where these two processes are termed diagnosis and therapy, respectively), the issue of hazard has been studied in depth, although this has not been done to such a degree for industrial uses of ultrasound in liquids/solids (sonar, non-destructive testing, ultrasonic processing etc.). However, in the second regime, that of the use of ultrasound in air, although the waves in question tend to be of much lower intensities than those used in liquids/solids, there is a greater mismatch between the extent to which hazard has been studied, and the growth in commercial applications for airborne ultrasound.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Ultrasound; Infrasound; Effects

Contents

1. Introduction .................................................................................................................. 4
2. Simple relationships .................................................................................................... 7
   2.1. Description of a compressional acoustic wave .................................................. 7
   2.2. Simple relationships, impedance, and intensity in the linear limit. .................... 10
   2.3. Reporting field amplitudes .................................................................................. 14
3. Nonlinear propagation ................................................................................................. 20
   3.1. The loss of linearity ............................................................................................ 20
1. Introduction

The question ‘what is ultrasound?’ illustrates perhaps the major difficulty with the subject: the answers seem simple and obvious. This is borne out by the fact that acoustics is a discipline with such an established history that many imprudent physicists consider it to be ‘solved’: the important equations were written, they believe, a century ago, leaving today’s acousticians simply with the task of translating that established physics into useful technology. Such an approach can lead scientists to underestimate complexity, engineers to rely on received wisdom, and industry to complacency when manufacturing or using ultrasonic systems. It is therefore not surprising that ultrasound as a technology for material processing has a reputation for unpredictability, difficulty in scale-up, and a reputation as a ‘black art’ (Mason et al., 1992). Since the assessment of ultrasound safety (a specific topic of this volume) must quite rightly include the involvement of specialists in fields other than ultrasound physics, it is important to raise awareness of the complexities peculiar to ultrasound, in order that the field can be dealt with as a predictable and understandable science, and not an unpredictable ‘black art’.

Another reason for this complacency with respect to the basic science is the ubiquitous and historic nature of our interaction with acoustic waves. With no other radiation do we interact to such a degree, both as a source and a receiver, and this can lead to an unwarranted complacency in the ‘physical feel’ we have for the radiation. We feel familiar with both positive and negative aspects of acoustics. Its ability to annoy might make the news through noise problems associated with roads, aircraft and neighbours, and yet, through speech, acoustics has dominated our communications for millennia. Acoustical engineering underpins not only recorded music, but also effective ‘live’ transmissions, from entertainment in theatres and concert venues, to well-designed public address systems. Many other of our acoustical interactions go unrecognised as such by the public. Often this is because although our experience for millennia has been dominated by audiofrequency sound in air, today we use ultrasound for both detection and material processing (the equivalent terms ‘diagnosis’ and ‘therapy’ being used if the application is biomedical). In terms of diagnosis, in the developed world, ultrasonic scanning of the foetus or other organs is commonplace (Duck et al., 1998; Szabo, 2004). The most familiar ultrasonic diagnostic application is external foetal scanning using 3–10 MHz ultrasound.

The question of the safety of diagnostic ultrasound is regularly and professionally reviewed (Barnett et al., 2000), although every few years or so the issue appears in the popular press. A notable exchange occurred in 1984, by which time it was estimated that probably well over a quarter of a million foetuses were being exposed to ultrasound each year in Britain alone (Chalmers, 1984). Public concern prompted Mr John Patten,
Junior Health Minister in the United Kingdom, to warn against the routine use of ultrasound in pregnancy in a letter to the Association for Improvements in Maternity Services. Quoted in the Daily Mail (22 October 1984) and 5 days later in Lancet, he said, “Given the publicity there has recently been about the possible risk of ultrasound scanning we would not expect any health authority to be advocating screening for all mothers as a routine procedure”. However in the same year Davies (1984) stated in the British Medical Journal that “the accumulated clinical experience of the past quarter century should be reassuring enough”. In response to Davies (1984), Chalmers (1984) however warned against complacency in the light of a lack of evidence to the contrary. He cited the two interesting historical cases. X-rays were first used in obstetrics in 1899, and by 1935 clinicians were recommending their routine use. Then in 1956 it was suggested that such a practice might predispose to the development of leukaemia in children (Stewart et al., 1956). Subsequent research supported this hypothesis (Bithell and Stewart, 1975). It may be noted that both X-rays and ultrasound have at times in the past been perceived to be non-invasive. Citing in addition the history of the use of the drug diethylstilboestrol, which was first used in obstetrics in the early 1940s (Smith et al., 1946), and in 1969 was proposed to predispose young women to vaginal adenocarcinoma (Herbst and Scully, 1970), Chalmers (1984) noted that none of the adverse effects in either case were “clinically obvious” to radiologists and obstetricians, and in fact were identified through the research of non-clinicians. Ziskin (1987) commented that “There is nothing that I’m aware of that has a safer record than that of diagnostic ultrasound”. Foetal ultrasonic scanning is now so established in industrialised nations that it would now be difficult to find a control group for epidemiological studies.

Nowadays many more anatomical sites are benefiting from ultrasonic scanning. Innovations include the development of probes for use whilst inserted into body cavities, and the exploitation of frequencies in excess of 30 MHz for use on shallow sites in dermatological and ophthalmic work (where enhanced spatial resolution is required, but the increased absorption which occurs at these frequencies is not debilitating). Non-imaging diagnostic methods have also developed, for example in the use of ultrasound to investigate bone health and osteoporosis through measurement of sound speed and attenuation (Langton et al., 1984; Hosokawa and Otani, 1997; Strelitzki et al., 1998; Hughes et al., 1999, 2001, 2003; Njeh et al., 1999; Wear and Armstrong, 2001; Lin et al., 2001; Lee et al., 2003; Wear, 2005). The role of ultrasound in non-destructive testing has a long history e.g. for crack detection, but in addition industry has found applications ranging from nuclear power (Watkins et al., 1988) to potteries (Leighton, 2004). The diagnostic requirements of military sonar have driven many of the oceanic developments in acoustic monitoring and measurement. From the Second World War to the present conflicts, sonar has had an unrivalled role in the underwater battlespace (Urick, 1983). The end of the Cold War prompted a move away from the study of low frequency acoustics (a few kHz and below) in deep waters that had been used, for example, to detect nuclear submarines beneath the polar icecaps (Leighton and Heal, 2005). In the last decade military engagements have tended to occur in shallower coastal waters, and research interests have reflected this, leading to the development of the new discipline of acoustical oceanography (Leighton, 1997; Medwin and Clay, 1998; Leighton et al., 2001; Medwin, 2005). As an oceanic sensor, acoustic systems map petrochemical reserves and archaeological sites, and monitor a huge variety of parameters of commercial and environmental importance, from fish stocks to the effect of global warming on the oceans (Medwin and Clay, 1998; Leighton et al., 2001, Leighton, 2004; Medwin, 2005).

Exploration of higher acoustic frequencies in these shallow waters promises many oceanographic spin-offs (Leighton and Heal, 2005). These include monitoring of zooplankton (Holliday, 2001), the seabed and suspended sediment (Thorne and Haynes, 2002; Richards et al., 2003), and archaeological investigations (Dix et al., 2001).

In terms of processing and therapy, the use of ultrasonic cleaning (for jewellery, computer chips, tool sterilization, etc.) is commonplace (Zeqiri et al., 1997; Leighton et al., 2005a). Industry has for example used ultrasound in the preparation of foodstuffs, pharmaceuticals and other domestic products (Leighton, 2004). Biomedical applications range from lithotripsy (the ultrasonic destruction of kidney stones) (Sass et al. 1991; Chaussy et al., 2002; Fedele et al., 2004) to surgery (Bailey et al., 2003), to physiotherapy. Ter Haar et al. (1987) published the results of a survey on the practices of physiotherapists in England and Wales in 1985. They concluded that the number of treatments was very considerable (about a million per year in the English and Welsh National Health Service Departments, and 150,000 in the private practices, that replied to the questionnaire).
Sonochemistry (the enhancement of chemical reactions using ultrasound, almost always in liquids) typifies many of the issues facing the use of ultrasound. It is by necessity a multidisciplinary activity, since effective exploitation of the technology requires knowledge of the chemistry, fluid dynamics, acoustics and transducer technology. The field is therefore a fruitful one for discoveries, but in turn a difficult one for understanding, scale-up and exploitation (see Section 5.2). To the author’s knowledge, there has not been a period within the last forty years when the field was not said by many to be on the verge of realising its commercial potential, but this expectation has, for the most part, not yet materialised. This is probably because of the requirement fully to solve, in a given project, the issues raised by the chemistry, engineering, physics and acoustics of the problem in equal measure.

Almost all of the above applications refer to the use of ultrasound in liquids or tissue, with a few (such as ultrasonic non-destructive testing, NDT) relating to the use of ultrasound in solids (Krautkramer and Krautkramer, 1977). Historically there have been fewer applications of ultrasound in air. Whilst some diagnostic applications have appeared over the decades (such as in range finding or intruder detection), there have often been competing technologies which have not suffered the high signal attenuation experienced by ultrasound in air (see Section 2.2 and Table 2). Whilst diagnostic applications in air are rare, those relating to processing are more scarce yet. Some, whilst they may appear at first sight to be processing applications of ultrasound in air (such as the ultrasonic knife, or the dental drills, filing apparatus and scalers—also known as scalers), they in fact rely on the vibration of the working solid tool, and are neither acoustic nor ultrasonic (in that their successful operation does not require the propagation of ultrasound in air, although this may be unwanted side-effect). Some proposed applications would use ultrasonic radiation forces (see Section 3.5) to agglomerate, filter or fractionate objects, such as particles suspended in air (de Sarabia et al., 2000; Kogan et al., 2004; Goddard and Kaduchak, 2005; Riera et al., 2006). These are techniques which, in liquids at least, might become more common as engineers attempt to fabricate microscopic scale fluid handling and analysis systems (Hawkes and Coakley, 2001; Harris et al., 2003; Hill, 2003; Townsend et al., 2004; Kuznetsova and Coakley, 2004; Wiklund et al., 2004; Pangu and Feke, 2004; Martin et al., 2005; Peterson et al., 2005a, b; Lilliehorn et al., 2005; Haake et al., 2005; Kumar et al., 2005).

The limiting feature for the use of ultrasound in air is the severe absorption which rapidly reduces the amplitude of the field, as it propagates away from the source, to levels which are too low for most processing activities, or even to provide sufficient signal-to-noise ratios (SNRs) for many diagnostic applications. There is however one exception, the manifestation of which illustrates a key point which must be appreciated in the assessment of the safety of ultrasound in air.

The human ear is an extremely sensitive sensor for acoustic waves. Intensities which are low by the standards used for ultrasonic diagnostic technology, and certainly for ultrasonic processing, are generally very much higher than the maximum intensities which the human ear can sustain at audio frequencies without damage. Therefore when ultrasound is used to generate signals to which the ear can respond (which may not necessarily be restricted to audiofrequencies—see Section 6), whilst the resulting intensities may be thought of as ‘low’ from the perspective of many ultrasonic technologies, they may be ‘high’ from the perspective of the ear. This point is discussed further in Section 7.

As a final word on applications of ultrasound, those new to the area should be aware that, in addition to the established applications, there have always been candidate uses for ultrasound advocated with enthusiasm, but which have yet to stand the test of time.

This paper proceeds through a series of sections to provide an answer to the question ‘what is ultrasound?’. Simple linear representations of acoustic waves, which are adequate for the vast majority of audio-frequency (20–20,000 Hz) acoustics, are often inaccurate for analyses in the ultrasonic frequency regime. This is because the ear is such a sensitive detector that the intensities at the ear\(^1\) at which audio-frequency waves cause pain or even hearing damage, are in physical terms very low (see Section 7), and usually place those waves in the linear regimes: any nonlinearities in their propagation are usually too small to be easily detectable.

\(^{1}\)Of course, amplitudes high enough to generate nonlinearities during propagation are often generated at audio-frequencies, by for example construction equipment and jet engines, but if human ears are close enough to the source to be in a region where amplitudes are this high, hearing protection should be worn to attenuate the wave before it reaches the ear.
However at ultrasonic frequencies the ear is not such a sensitive sensor. As a result, the higher amplitudes (needed for example, for food processing, or to provide strong SNRs during non-destructive testing) can be generated without, in the majority of cases, causing pain to the ear. Indeed, the need to generate such high amplitudes at the source becomes increasingly essential as the frequency increases. This is because of the need to maintain a satisfactory SNR, since the absorption of ultrasound by most media usually increases with frequency. However whilst they are strictly inaccurate in the nonlinear regime that such high amplitudes induce, if used cautiously the linear relationships can provide useful rough guides, and so will be presented in Section 2 to demonstrate some of the basic features of acoustic waves.

The formulations of Section 2 assume that the wave propagates linearly. However in many applications, the amplitude of the ultrasound is sufficiently great, and the dissipation sufficiently low, for the inherent nonlinearity of the propagation to reveal itself. Such high amplitudes are very characteristic of ultrasonic applications: where MHz frequencies are used (for example to achieve tight spatial or range resolution, or focusing), the attenuation is so great that high source amplitudes are often used. At the lower ultrasonic frequencies, instrumentation often exploits the assumed insensitivity of the human ear to frequencies above 20 kHz to generate ultrasound (either directly or as a by-product of the vibration of some instrument, such as a dental scaler) which has an amplitude that is sufficiently high to cause acoustic processing of a material, in a way which would not be possible at audio frequencies without hearing hazard. Section 3 describes such nonlinear propagation and its effects.

Such nonlinearities need to be appreciated since their effects tend to run contrary to the ‘common sense’ expectations of those who have experience only with linear regimes (Section 5.2 illustrates examples where the author has been asked to solve problems with ultrasonic instrumentation which is not behaving as expected). Because of this, Section 4 describes perhaps the most potent source of nonlinearity, cavitation. This is not to say that other ultrasonic effects upon tissue (streaming and microstreaming, radiation forces, etc.), which have less than an entire section devoted to them, are less important: indeed, for foetal scanning, hyperthermia is very likely a more important issue than cavitation. However cavitation is also the most difficult to understand, perhaps because of its inherent propensity for nonlinearity, and hence is the topic of Section 4. After a discussion in Section 5 of the scales, in terms of space and time, encountered in typical ultrasonic applications, Section 6 describes the use of ultrasound in air, before Section 7 provides the conclusions of the paper.

2. Simple relationships

2.1. Description of a compressional acoustic wave

Acoustic waves can come in a variety of forms (Fig. 1). The energy contained in one form of wave can be converted to another, for example at interfaces between two media. These waves have different propagation characteristics (phase and group velocity, dispersion, attenuation, etc.). Indeed there are some forms of wave which do not propagate in the strict sense, such as evanescent waves and hydrodynamic pressure signals (the strength of which can readily be seen by moving one’s head underwater when submerged in the domestic bath or swimming pool—the pressure signatures cause by flow at the pinna give rise to an apparently loud ‘sound’, but these do not propagate to distance and cannot be detected by a nearby observer).

However probably the most common and familiar forms of acoustic wave are the longitudinal compressional waves, in which the particles are displaced parallel to the direction of motion of the wave (Fig. 1(a)). It is important to note that in both cases, the particles themselves are merely displaced locally, or oscillate: it is the wave that travels from source to detector, not the particles. Therefore if one sings a loud, steady note at a lighted candle from a distance of a few centimeters, the flame barely flickers, since it is local
vibrations which are transmitted: there is no net flow of air, which would correspond to an extinguishing ‘blow’.

Fig. 1 shows a small subset of the types of acoustic waves that exist, and because there are so many types, it is useful first to consider a very simple analogue of a compressional longitudinal wave, shown in Fig. 2. This analogue consists of a series of bobs of equal mass, connected in a line by massless, lossless springs. The model
therefore comprises the two necessary elements of any medium through which a sound wave will pass: inertia (invested in the bobs) and elasticity (invested in the springs). Only a section of the infinite line of bobs is shown.

Fig. 2(a) shows the bobs equally spaced in the equilibrium position. In Fig. 2(b) a one-dimensional single-frequency longitudinal wave is passing through the medium, and the bobs are shown frozen at an instant in time (as if photographed). Almost all of the bobs are displaced either to the left or right, the exceptions being those bobs at the centres of the rarefactions and compressions so created. The wavelength is much greater than the bob spacing and, in order to plot the particle velocity from this, we will need to know that the wave is travelling from left to right with a single phase speed \( c \).

Arrows between parts (a) and (b) shows how each bob has been displaced. The bobs and springs in (a) and (b) represent the inertia and stiffness of a continuum, and interpolation between these bobs allows the characteristics of that continuum (through which the wave is passing) to be identified. Therefore in Fig. 2(c) we can represent the concentration of particles as a continuous change in density, the darker the regions the greater the density. This continuous change in density can be related to one in pressure, through the equation of state (Section 3.2), and this is plotted in Fig. 2(e). Regions of high pressure (compressions) in Fig. 2(e) correspond to points of high population density in Fig. 2(c). Similarly, low-pressure regions (rarefactions) occur at points with low concentrations of particles.

The solid line in Fig. 2(d) shows the displacement as a function of the equilibrium position for the continuum. This can be found by interpolating the discrete displacements required to go from Fig. 2(a) to Fig. 2(b), and plotting them (with right as positive, and left as negative, on the mantissa) as a function (on the abscissa) of the equilibrium position of the bob shown in Fig. 2(a). The displacement and pressure plots are in quadrature.\(^4\) This schematic demonstrates an important point in acoustics, that one must take care to specify whether one is referring to pressure or displacement: in the figure, positions of zero displacement correspond to maximum or minimum pressure. If unqualified, common terms such as ‘amplitude’, ‘node’, or ‘antinode’ could apply to either displacement or pressure (Walton and Reynolds, 1984). An early example of one such ambiguity can be found by Paounoff (1939): “la lumière est plus intense aux plans nodaux des ondes stationnaires”: one cannot tell from this phraseology whether the luminescence in the standing-wave field described by Paounoff occurred at the pressure nodes or the displacement nodes. For more on this topic, see Fig. 13(a).

Fig. 2 illustrates the concept of wavelength, that is the distance between two points on a wave (here, a sinusoidal wave) showing the same disturbance and doing the same thing (i.e. the disturbance is increasing in both, or decreasing, or stationary). The wavelength \( \lambda \) is shown on the figure, and is related to the phase speed

\(^4\)That is, if one is a sine wave, the other is a cosine wave, such that they plot with a phase difference of \( \pi/2 \).
\( c_\phi \) by \( c_\phi = \gamma \lambda \) where \( \gamma \) is the linear frequency (i.e., in Hertz). This is in turn related to the circular frequency \( \omega \) (measured in radians per second) by \( 2\pi \gamma = \omega = c_\phi k \), where \( k = 2\pi / \lambda \) is the wavenumber. Similarly, the group velocity \( c_g \) is defined as \( c_g = \partial \omega / \partial k \) in the usual manner. For linear waves, a complex wavenumber can be used to describe the absorption of the wave by the medium for processes which convert the acoustic energy ultimately into heat. Such processes include viscous, ionic, etc. mechanisms. In contrast geometrical factors are incorporated by a range-dependence in the amplitude term. Other losses, such as diffraction losses and scatter, require application of some form of propagation model. These factors are discussed further in Section 2.2.

As stated above, this wave is travelling to the right, and a small time \( \Delta t \) later the displacement curve has moved to the dashed curve in Fig. 2(d). For each position along the wave \( x \), the difference in the mantissa between the solid and dashed curves in Fig. 2(d) is proportional to the particle velocity \( v \) (it equals \( v\Delta t \)). The particles in the compressed regions are moving forwards, and those in rarefaction moving backwards. Particle speed is greatest at the regions of zero displacement, and reduces to zero at the points of maximum and minimum displacement. Comparison of the two curves of Fig. 2(d) in this way allows the particle velocity to be plotted, and this is shown in Fig. 2(f). Note that, the particle velocity and acoustic pressure are in phase, but they are in quadrature with the particle displacement. This finding for loseless linear waves will now be derived analytically as part of a discussion of the characteristics of waves which propagate linearly.

### 2.2. Simple relationships, impedance, and intensity in the linear limit

For the simple example discussed in the preceding section, the acoustic pressure \( P \) (Fig. 2(e)) and particle velocity \( v \) (Fig. 2(f)) are in phase. This would imply that the specific acoustic impedance of the medium \( Z \) is real, and for the moment this is the only type of wave which will be considered.3 This particular impedance is defined through the ratio \( Z = P/v \), and for such waves is numerically equal to the product of the mass density of a medium at equilibrium \((\rho_0)\) and the phase speed of the compressional wave:

\[
Z = P/v = \rho_0 c_\phi. \tag{1}
\]

If the wave is linear, Eq. (1) allows unsophisticated estimations of the phase and magnitude relationships between displacement \( \varepsilon \), acoustic pressure \( P \) and particle velocity \( v \). If for example the displacement were a simple harmonic wave, of zero-to-peak displacement amplitude \( \varepsilon_0 \), then

\[
\varepsilon = \varepsilon_0 e^{i(\omega t-kx)},
\]

\[
v = \dot{\varepsilon} = j\omega \varepsilon_0 e^{i(\omega t-kx)} = \omega e^{i\pi/2},
\]

\[
P = Zv = j\rho_0 c_\phi \varepsilon_0 e^{i(\omega t-kx)} = P_\Lambda e^{i(\omega t-kx)},
\]

\[
\ddot{\varepsilon} = -\omega^2 \varepsilon,
\]

where \( P_\Lambda \) is the acoustic pressure amplitude of the wave, recalling that the assumption is still maintained that \( Z \) is real. Eq. (2) confirms the deductions made in Section 2.1 using Fig. 2. That is to say, if the propagation is lossless and \( Z \) is real, then whilst acoustic pressure \( P \) and particle velocity \( v \) are in phase, they are in quadrature with the displacement (since \( j = e^{i\pi/2} \)).

Eq. (2) also allows the relative magnitudes of the various wave properties to be estimated in common materials. The basic properties of air, water and aluminium are shown in Table 1.

Table 1 also compares the properties of a sinusoidal ultrasonic wave of rms acoustic pressure amplitude 100 Pa in these three materials. From Eq. (2) it is simple to calculate the magnitudes of the particle velocity \(|v| = |P|/|Z|\), the displacement \(|\varepsilon| = |P|/(\omega |Z|)\) and the particle acceleration \(|\ddot{\varepsilon}| = \omega |P|/|Z|\), and to see the roles that the different sound speeds and densities play in these calculations. Within a given material, the frequency dependence is important, which is illustrated by performing the calculations for Table 1 at 20 kHz and 1 MHz. As introduced qualitatively in Section 2.1, the reason why one can sing a vowel loudly at a candle flame and not perceive it to flicker is because the displacements are tiny: from Eq. (2), a sung note ‘A’ at

---

3In practice, there is a finite phase difference between \( P \) and \( v \) in lossy media, such as that for simple linear waves, \( Z \) is complex (see Eq. (8)). Such absorption will be introduced later in this section.
444 Hz, having an acoustic pressure amplitude of 0.02 Pa (0-peak, i.e. half the peak-to-peak amplitude in a sinusoidal wave) would generate displacement amplitudes in room air of only about 15 nm.

To return to the discussion of acoustic impedance, note that the specific acoustic impedance is just one of a range of impedances that can be defined in acoustics (depending, for example, on whether one uses the particle velocity or the volume velocity). These impedances take the usual role in physics in describing how simple waves are transmitted from one medium (where $Z = Z_1$) to another (where $Z = Z_2$), since they can be used to describe the constraining boundary conditions in a very simple form. For example if a plane compressional linear wave is normally incident on a plane boundary between two media, then continuity of the pressure gives us the first boundary condition, and continuity of the particle velocity gives us the second (Leighton, 1994, Section 1.1.5).

From these boundary conditions the pressure amplitude reflection coefficient for normally incidence plane compressional wave (note that it is important to have all these qualifications) is defined as the ratio $R = P_r/P_i$, where $P_i$ is the acoustic pressure amplitude of the wave that is incident on the boundary, and $P_r$ is the acoustic pressure amplitude of the wave that is reflected off it. It equals:

$$R = \frac{Z_2 - Z_1}{Z_2 + Z_1}. \quad (3)$$

Formulations for arbitrary angles of incidence exist (Kinsler et al., 1982; Leighton 1994, Section 1.1.5), which reduce to (3) at normal incidence. It should be remembered that the underlying model for (3) is based on a planar interface between two fluids, and that a fluid/solid or solid/solid boundary can give more complicated effects (such as the generation of shear waves) when a longitudinal wave is incident upon it at an arbitrary angle of incidence (Kinsler et al., 1982). It is however sufficient to illustrate to what degree differences in acoustic impedance between two media can lead to strong reflection of the incident wave. This process underlies a range of common practices in using ultrasonics, from the operation of range-finders in air, to the production of sonar and biomedical images (using the time-of-flight of the echo for range, and the intensity to indicate acoustic impedance mismatch at the reflecting boundary). Because the acoustic signal one expects to detect from reflection often resembles the electrical signal used to drive the transmitter, to ensure that the

The first three rows of numbers show material properties. The next six rows of numbers (italicized) show the parameter values in each category for an acoustic wave having an rms acoustic pressure amplitude of 100 Pa. The final three rows form a matrix expressing the percentage of energy transmitted when a plane wave is normally incident at a plane interface between the medium shown in the row and the medium shown in the corresponding column (see text).

The degree of resemblance depends on the transfer functions and ring-up/ring-down characteristics, but is often close enough to fool the unwary observer. There are techniques to increase the resemblance if ultrasonic high-fidelity systems are required (Doust and Dix, 2001).
detected signal is the result of acoustic propagation and not direct electrical pick-up of the driving signal, it is good practice in underwater ultrasonics to ensure that the detected signal disappears when the transducers are taken out of the water or, if that is not possible, when a sheet of expanded polystyrene is placed in the propagation path. This is because the impedances of air and expanded polystyrene are about 0.03% of the impedance of water, such that signals in water which are normally incident upon large sheets of expanded polystyrene exhibit values of $R$ close to $-1$, indicating that the wave is almost entirely reflected with a $\pi$ phase change. This type of interface is termed ‘pressure release’. Waves in air which encounter walls of aluminium exhibit values of $R$ close to $+1$ (the waves are almost entirely reflected, without a phase change), and the boundary is termed ‘rigid’ (Kinsler et al., 1982; Leighton, 1994, Section 1.1.5).

Whilst they do not hold for waves more complicated than the simple linear waves discussed in this section, relationships such as Eq. (2) are nevertheless frequently used to estimate the magnitude of various acoustic parameters in the manner used for Table 1. In similar vein, estimates are often made based on assumptions that the acoustic intensity of a wave is proportional to the square of its acoustic pressure, based on the correspondence for simple plane and spherical waves:

$$I = \frac{P_A^2}{2Z} = \frac{P_{\text{rms}}^2}{Z},$$

(4)

where $P_A$ and $P_{\text{rms}}$ are, respectively, the zero-to-peak and the rms acoustic pressure amplitudes of the wave (Leighton, 1994, Section 1.1.3, 3.2.1(c)(iii)).

Given that the intensities of the incident and reflected waves of the simple plane waves of Eq. (3) are, from (4), proportional to the square of their respective acoustic pressure amplitudes, then the intensity reflection coefficient for normally incidence plane compressional wave is $R^2$. The percentage of incident energy which is reflected by the boundary is $|R|^2 \times 100\%$, and the percentage of energy which is transmitted is $(1-|R|^2) \times 100\%$, such that $1-|R|^2$ is known as the normal incidence intensity transmission coefficient. Insertion into Eq. (3) of the appropriate values from Table 1 provides the normal incidence values of the normal incidence intensity transmission coefficient in percentage terms (i.e. $(1-|R|^2) \times 100\%$) for interfaces involving boundaries between air, water and aluminium. Whilst 100% of the energy is transmitted, of course, for the trivial cases of perfect interfaces between a given material and itself (where the materials are said to be ‘impedance matched’), the other cases calculated show considerable depreciation in the transmitted energy. It is particularly low for interfaces involving air, because of the low density of gas compared to fluid and solid. This is the reason for numerous phenomena in acoustics, including the use of coupling gel to eliminate air gaps when biomedical ultrasonic transducers are placed on skins for foetal scanning, physiotherapy, etc. The low transmission between liquid/soft tissue and gas also accounts for the restrictions on the use of ultrasound when large bodies of gas (in for example the lungs or gut) are present, which in part accounts for the advent of intracavity transducers. It accounts for the almost casual treatment of safety for many common aspects of industrial power ultrasonics: there can be little transmission of the ultrasound from an ultrasonic cleaning bath to the human body if an air gap is present between them. It also accounts for the way the assessment of hazard must change when contact is made, be it through hand contact with a transducer, or when whole body immersion occurs (in which case the close impedance matching between water and the human body can account for numerous transmission paths for acoustic energy to anatomical structure, e.g. from knees to ear). Therefore any assessments of the effect of ultrasound in a given circumstance must take account of (i) the various transmission paths to the organ in question, and (ii) the possibility that more than one organ should be considered. The discussion of dental ultrasonics in Section 6.2 illustrates these points.

The transmission losses calculated in Table 1 refer to that component of attenuation which is caused by the reflection of acoustic energy back from the interface. Again, this accounts for the efficacy with which diagnostic ultrasound detects structures with which the host medium forms a strong impedance mismatch (examples range from the detection of air-filled cracks in solids during ultrasonic non-destructive testing (NDT), to the clarity of bone structure in foetal scanning).

In general terms, the loss of energy from the propagating wave through reflection at interfaces enters the formulation through use of the reflection coefficients (themselves derived from the boundary conditions) to calculate the amplitudes of the reflected and transmitted waves from the amplitude of the incident wave. It is important to recall that there are other sources of loss, mentioned in Section 2.1. These enter the above
formulations through other routes. Other sources of loss would enhance the attenuation of the wave. Absorption (the conversion of acoustic energy ultimately into heat) can enter the formulation through use of a complex wavenumber. If now $k = q - jb$, then the above displacement becomes $\varepsilon = (i_0/r) a^{(at - qr + jb)} = (i_0/r) a^{(at - qr)} e^{-br}$, where in addition to the absorption loss $e^{-br}$ the amplitude has been allowed to fall off as $1/r$ to incorporate inverse-square geometrical spreading losses in intensity with increasing range $r$ (see later).

The parameter $b$ characterises the absorption of the medium, and in general it increases strongly with frequency. For plane waves, absorption will contribute to attenuation an amount such that the amplitudes (pressure, displacement or particle velocity) decay with propagation distance $x$ as $e^{-bx}$ whilst, for propagation obeying (4), the wave intensity will fall off as $e^{-2bx}$. For example, the pressure and intensity for a plane wave could follow:

$$P = P_\lambda e^{i(at-qx)} e^{-bx} \Rightarrow \frac{dI}{T} = -2bx \Rightarrow I_2 = I_1 e^{-2b(x_2-x_1)},$$

where $x_1$ and $x_2$ are the locations where the intensity takes values of $I_1$ and $I_2$ respectively. Two intensities can be compared via the decibel scale (Section 2.3), such that the logarithmic absorption coefficient $\alpha_{ab}$ is:

$$\alpha_{ab} = \frac{10 \log_{10}(I_1/I_2)}{x_2 - x_1} = 20 \log_{10} e$$

such that the choice of $x_2 - x_1$ determines the units of $\alpha_{ab}$ (e.g. dB/cm, dB/km). In turn the e-folding distance (the distance a wave must travel before its intensity decays to $e^{-1}$ of its original value) is found by setting $I_2/I_1$ in Eq. (5) equal to $e^{-1}$, giving an e-folding distance for energy ($L_e$) of:

$$L_e = 1/(2b).$$

Values of the amplitude attenuation coefficient (in $b$ and dB/m), along with the e-folding distance, are tabulated for fresh water, seawater and air at 10 kHz and 1 MHz for three materials in Table 2. The absorption increases with frequency, and in air it is very much greater than that in water (seawater being more attenuating than freshwater). Note that these are plane wave calculations, so that no amplitude changes associated with geometrical spreading or converging have been included (see below). The high absorption seen in air accounts for the fact that, whilst there are numerous devices which exploit the small wavelengths afforded by high frequency ultrasound to obtain good spatial resolution in water, tissue, and many homogeneous solids, applications of ultrasound in air are limited to the lower ultrasonic frequencies. For comparison, at 1 MHz the amplitude attenuation coefficient $b$ for aluminium is 0.0207 neper m$^{-1}$, giving it an e-folding depth of 24 m.

<table>
<thead>
<tr>
<th></th>
<th>Water</th>
<th>Seawater</th>
<th>Air</th>
</tr>
</thead>
<tbody>
<tr>
<td>$10,\text{kHz}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{ab}$ (dB/m$^{-1}$)</td>
<td>$2 \times 10^{-5}$</td>
<td>$4 \times 10^{-4}$</td>
<td>$1 \times 10^{-1}$</td>
</tr>
<tr>
<td>$b$ (Np m$^{-1}$)</td>
<td>$2 \times 10^{-6}$</td>
<td>$5 \times 10^{-5}$</td>
<td>$1 \times 10^{-2}$</td>
</tr>
<tr>
<td>e-folding distance for energy</td>
<td>250 km</td>
<td>10 km</td>
<td>50 m</td>
</tr>
<tr>
<td>$1,\text{MHz}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha_{ab}$ (dB/m$^{-1}$)</td>
<td>$2 \times 10^{-1}$</td>
<td>$3 \times 10^{-1}$</td>
<td>$1 \times 10^{3}$</td>
</tr>
<tr>
<td>$b$ (Np m$^{-1}$)</td>
<td>$2 \times 10^{-2}$</td>
<td>$3 \times 10^{-2}$</td>
<td>$1 \times 10^{2}$</td>
</tr>
<tr>
<td>e-folding distance for energy</td>
<td>25 m</td>
<td>17 m</td>
<td>5 mm</td>
</tr>
</tbody>
</table>

Symbols are defined in the text. Sources: Piercy et al. (1977), Leighton (1994), Medwin (2005). Whilst air has greater absorption than either type of water, the differences between fresh and seawater are revealing. These are not primarily due to the presence of sodium chloride, but rather to magnesium sulphate and boric acid. This illustrates two things. First, it shows the importance of properly characterising all of the sources of loss, even from structures or chemicals whose presence may at first sight be minor. Second, it illustrates the potential problems of using water as an in vitro medium for measurement. Whilst this is recognised for many traditional applications of ultrasound, and hence has led for example to derating procedures, we should remain aware of the potential problems as new methodologies and measurements are introduced (the interdependency of nonlinear propagation and absorption being one example).
In contrast, castor oil is often used as an absorber for MHz ultrasound, since at 1 MHz it takes a value of $b = 10.9$ giving it an e-folding depth of 46 mm.

Of the many sources of loss in an ultrasonic wave, one deserved special discussion. It is often (e.g. at low ultrasonic frequencies) called ‘geometrical spreading loss’ (although this is in some ways a misnomer—see below). It describes the loss in the locally measured wave amplitude as its energy is spread over a wider area, as the wave propagates away from the source. However unlike absorption, geometric spreading does not change the total energy invested in the wave: it simply changes the intensity or pressure measured at a single point depending on the location of that point with respect to the source.

However because of this, it is often considered to be a source of ‘loss’, particularly at low ultrasonic frequencies. This is because in many cases the wavelength is much greater than the size of the source (see Section 5.2). In such circumstances, in the free field it would be impossible to generate a plane wave (which has no geometrical spreading losses). Most sound fields would disperse away from the source, the wavefront expanding as it propagates and spreading the energy of the wave over an area which increases with distance from the source. Consider for example how such an effect is usually incorporated into the above formulation by use of a range-dependent amplitude. Spherical spreading from a monopole source would contribute an inverse square law to the loss in intensity with range $r$. As a result, the displacement in Eq. (2) would be characterised by an amplitude term which falls off as $r^{-1}$ i.e. $e = (\xi_0/r) e^{(-\delta r)}$. (Note that if intensity falls off as $r^{-2}$, then from (4) the pressure amplitude would fall off as $r^{-1}$ and hence, from (2), so would the displacement amplitude).

Therefore at low ultrasonic frequencies (such as are used in much of sonar), users tend to discuss ‘geometrical spreading losses’. However if the source size is large compared to the wavelength (as with MHz biomedical ultrasonics), geometrical factors are viewed very differently. With large values of $ka$ (see Section 5.2), sources can produce focused fields. This is vital to the operation of many biomedical ultrasonics devices, both diagnostic (e.g. imaging) and therapeutic (e.g. HIFU, lithotripsy etc.). Under such circumstances the field tends first to undertake geometrical convergence (focusing), and only after the focus does geometrical spreading occur.

### 2.3. Reporting field amplitudes

The preceding discussion outlines the ways in which the amplitude of a wave can change, and because the magnitude of these changes can be so great, the decibel (dB) scale is commonly used. Misuse of this scale is responsible for errors, ambiguities and misreporting in the literature.

The amplitude of a wave will be attenuated through the absorption illustrated in Table 2, and also by scattering and diffraction (see Section 5.2 and Fig. 27). Loss at interfaces also occurs: only around 0.1% of the wave intensity would be transmitted across a plane air–water interface by a normally incident wave (Table 1). However we know from our experience with audio acoustics that we cannot set this value equal to zero, since a submerged person can hear sounds which originated from the air. Hence to cope with these large dynamic ranges it is usual to use a dB scale to express differences in amplitude or intensity. For two simple waves of the type described above, of intensities $I_1$ and $I_2$, the dB difference between them is $10 \log_{10}(I_1/I_2)$. Footnote 2 used this when referring to the intensities at the limits of a ± 3 dB uncertainty range, and Eq. (6) used it to compare the intensities measured at two locations as the amplitude of a given propagating wave attenuated with distance. However more generally the dB scale can be used to compare any two intensities (one of which may, or may not, be a reference pressure, an important case which will be discussed below). The dB difference can also be used to compare two acoustic pressures (both being measured the same way, e.g. 0-peak amplitude, rms, etc.). Consider again the two simple waves discussed above, of intensities $I_1$ and $I_2$ and a dB difference of $10 \log_{10}(I_1/I_2)$. Since for such waves the intensity is proportional to the square of the acoustic pressure amplitude (Eq. (4)), then the equivalent dB difference between them, in terms of their respective acoustic pressure amplitudes $P_1$ and $P_2$, is $20 \log_{10}(P_1/P_2)$.

In acoustics it is not uncommon (although it is very bad practice) to cite the dB as if it were an absolute unit: in fact this is not the case, and the dB level cited is actually relative to some reference pressure. That is to say, in the formulation above, the dB level of the signal $P_1$ would be expressed by using for $P_2$ the appropriate reference level. Further confusion can be caused to the unwary by the fact that in common practice,
the reference level for waves in air\textsuperscript{7} is different to that used for waves in liquids or liquid-like tissue (1 \(\mu\)Pa rms). Therefore in this regrettably colloquial usage, the example acoustic wave in air used to discuss the implications of Eq. (2), having an rms acoustic pressure amplitude of 100 Pa, would by some be said to have an ‘amplitude’ of about 134 dB (i.e. \(20 \log_{10}(100/20 \times 10^{-6})\)). However were a wave of exactly the same acoustic pressure amplitude to be measured in water, colloquial usage would attribute to it an ‘amplitude’ of 160 dB (i.e. \(20 \log_{10}(100/10^{-6})\)). Proper usage would give these amplitudes, respectively, as 134 dB re 20 \(\mu\)Pa, and 160 dB re 1 \(\mu\)Pa. Failure to adhere to such a protocol, or indeed to appreciate that the dB is not an absolute unit and indeed has different reference pressures in different media, has led to some notable cases of inappropriate assessment of hazard. Chapman and Ellis (1998) discuss a current example, specifically the concern over the effect of sonar on marine mammals. They analyse a quote from The Economist (1998), which arose following scientific correspondence in Nature (Frantzis, 1998). Referring to a sonar source designed to produce low-frequency sound, The Economist stated that “It has a maximum output of 230 dB, compared with 100 dB for a jumbo jet.” Chapman and Ellis (1998) criticise this phrase in the following way: “Regardless of the author’s intention, the implication is that the whale would experience an auditory effect from the sonar that would be substantially greater than that of a person exposed to the jet aircraft.” There are several reasons why this type of comparison is misleading.

First, the reference sound pressure for in-air acoustics (20 \(\mu\)Pa rms) is not the same as that used in underwater acoustics and biomedical ultrasonics (1 \(\mu\)Pa rms). This automatically means that a given rms acoustic pressure measured in water will have a level (in dB re 20 \(\mu\)Pa rms) that is \(20 \log_{10}(20/1) \approx 26\) dB greater than for the same rms acoustic pressure measured in air.

Indeed, many practitioners actually use a rule-of-thumb of subtracting 62 dB from intensity levels in water to estimate the intensities in air for the same acoustic pressure amplitude. This comes from the 26 dB to account for the different reference pressures, plus 36 dB to account for the differences in specific acoustic impedance required to compare intensities (see Eq. (4)):

\[
10 \log_{10}(\rho_w c_w / \rho_g c_g) \approx 10 \log_{10}(1.5 \times 10^6 / 1.23 \times 343)
\]
\[
\approx 10 \log_{10} 3600 = 36\text{ dB}.
\]

To investigate the validity of this factor of 62 dB (or 61.5 dB, as some use), let us translate the underwater noise on a coral reef into an in-air equivalent as rated by the noise rating (NR) curves (a simplistic but widely used system for expressing the magnitude of ambient noise signals as a single number). The noise of snapping shrimp is the dominant source of underwater-sound in many tropical bays away from surf and man-made noise. It sounds like ‘sizzling sausages’ to snorklers swimming over tropical reefs (Ferguson and Cleary, 2001). Let us consider what these snorklers, with their human 20–20,000 Hz range, are hearing. Although the signal from the snapping shrimp contains energy at frequencies in excess of 100 kHz, the following analysis will consider the problem from the perspective of a standard NR (Noise Rating) calculation, which only considers energy up to the octave band centred on 8 kHz.

The white bars in Fig. 3 show the spectrum in this frequency range for the sound from the shrimp, as recorded in Kaneohe Bay, Hawaii (Everest et al., 1948). Similar levels have been confirmed internationally by Readhead (1997) and Au and Banks (1998). The spectrum over the frequency range of concern for calculating NR is shown in Fig. 3, both before (white bars) and after (black bars) the 62 dB air-sea “correction” recommended by many\textsuperscript{8}. Fig. 3 also includes a spectrum of grey bars, indicating the intermediate step whereby the first 26 dB are subtracted to account for the difference in reference pressures.

\textsuperscript{7}It is common to use a reference level in air of 20 \(\mu\)Pa rms. This is based on \(I_{\text{min}}\), the \(10^{-12}\) W m\(^{-2}\) consensus minimum audible intensity at 1 kHz, which corresponds to an acoustic pressure amplitude of \(\sqrt{2} \rho g c \rho_{\text{in}} = 28.9 \mu\)Pa in terms of 0-peak acoustic pressure amplitude, or 20.4 \(\mu\)Pa rms, such that 20 \(\mu\)Pa rms is usually used. Here \(\rho_g\) and \(c_g\) refer respectively to the equilibrium density and sound speed in the gas (here, air); a similar notation will be used for other media, such that \(\rho_w\) and \(c_w\) will refer respectively to the equilibrium density and sound speed in water.

\textsuperscript{8}http://www.pmel.noaa.gov/vents/acoustics/tutorial/8-conversion.html, the conversion of dB between air to water, in Underwater Acoustics Tutorial, US National Oceanic and Atmospheric Administration (NOAA) Vents Program. See also: Taking and importing marine mammals; Taking marine mammals incidental to navy operations of surveillance towed array sensor system low frequency active sonar; final rule, Federal register 2002 (67; 46712). See also: Final overseas environmental impact statement and environmental impact statement for surveillance towed array sensor system; Low frequency active (SURTASS LFA) sonar (Volume 1 of 2), US Department of the Navy, 2001.
The NR rating is determined by the NR curve which just envelopes the measured noise spectrum. Therefore this noise spectrum (the black histogram of Fig. 3) would be interpreted as a NR value of NR 70 (Bies and Hansen, 1996). For a human living space, NR 45 corresponds to the expected noise levels expected inside a living room in a domestic dwelling which is situated in an area of heavy industry. In exceeding this by 25, a rating of NR 70 would suggest that the acoustic environment near a coral reef is comparable to the maximum allowable levels for the machine control room of a ship, as set forth by the UK Maritime and Coastguard Agency for that environment. Snapping shrimp on a coral reef are certainly significant sources of sound. However it is difficult to accept the result of an NR calculation which subjectively likens the perceived acceptability of the noise to that found in a loud industrial setting. We know that, for the data in Fig. 3, in fact the snorkler in question would be hearing the natural acoustic environment that is commonly found away from sources of man-made noise during a shallow dive in say, Hawaii or the Caribbean. This suggests that, as regards the acceptability of noise to humans, subtraction of more than 62 dB may be required in translating from water to air. This may be attributable in part to the differences in the way the ear operates in air and water (including questions of whether the ear canal is filled with water or air).

Although such subjective comparisons are not rigorous, they indicate that it is no simple matter to transfer ‘annoyance’ levels of sound from one medium to another, even when we restrict it to one species: to make such comparisons with an interspecies transfer included (as is frequently done between humans and cetaceans) is unwise.

Fig. 3. This figure uses the noise rating curves to show why it might be inappropriate to suggest that subtracting 62 dB from an underwater sound pressure level “converts” that level to its aeroacoustic equivalent. To make this comparison as transparent as possible, the levels in each octave band are shown at three separate points during conversion. In white bars are shown unadjusted levels recorded in Kaneohe Bay, where the ambient acoustic spectrum is dominated by snapping shrimp. To account for the fact that most acoustic measurements performed in water are referenced to 1 μPa rms, while those performed in air are referenced to the nominal human threshold of hearing of 20μPa rms, a sum of 26 dB is subtracted from the original levels to give the octave band levels illustrated by the grey bars (see text). To account further for the difference in the specific acoustic impedance from air to water, the octave band levels are reduced by an additional 36 dB, giving a net reduction of 62 dB per octave. These results are shown as black bars. The calculation indicates that according to the conversion method indicated, the audible crackle of a coral reef might be rated as NR70; a level prescribed as being the maximum allowable in the control space for a ship’s machine room (figure by DC Finfer and TG Leighton).

The NR rating is determined by the NR curve which just envelopes the measured noise spectrum. Therefore this noise spectrum (the black histogram of Fig. 3) would be interpreted as a NR value of NR 70 (Bies and Hansen, 1996). For a human living space, NR 45 corresponds to the expected noise levels expected inside a living room in a domestic dwelling which is situated in an area of heavy industry. In exceeding this by 25, a rating of NR 70 would suggest that the acoustic environment near a coral reef is comparable to the maximum acceptable levels for the machine control room of a ship, as set forth by the UK Maritime and Coastguard Agency for that environment. Snapping shrimp on a coral reef are certainly significant sources of sound. However it is difficult to accept the result of an NR calculation which subjectively likens the perceived acceptability of the noise to that found in a loud industrial setting. We know that, for the data in Fig. 3, in fact the snorkler in question would be hearing the natural acoustic environment that is commonly found away from sources of man-made noise during a shallow dive in say, Hawaii or the Caribbean. This suggests that, as regards the acceptability of noise to humans, subtraction of more than 62 dB may be required in translating from water to air. This may be attributable in part to the differences in the way the ear operates in air and water (including questions of whether the ear canal is filled with water or air).

Although such subjective comparisons are not rigorous, they indicate that it is no simple matter to transfer ‘annoyance’ levels of sound from one medium to another, even when we restrict it to one species: to make such comparisons with an interspecies transfer included (as is frequently done between humans and cetaceans) is unwise.

10Indeed, since the snorkler would also hear the contributions made by the shrimp at >8 kHz, the reef environment would actually be perceived as ‘louder’. However the trends in Fig. 3 above the 8 kHz band have not been included in the NR calculation.
Second, the values of many of the physical parameters (e.g. particle velocity, displacement, acceleration) associated with two acoustic waves which have the same rms acoustic pressure, one in air and one in water, are not the same. This can be seen from Eq. (2), and from rows 5–10 of Table 1, which are all calculated for a sinusoidal acoustic wave of 100 Pa rms acoustic pressure. It is by no means conclusive that the intensity alone is a sufficient measure of annoyance or hazard when transferring between physically different media.

Third, the quote in *The Economist* unwittingly compares the sonar sound level measured ‘at the source’ (which typically means at a distance of 1 m from an underwater sound source, where no listener, human or cetacean, is likely to be) with a level measured ‘at the receiver’ (i.e. at the position of the listener or microphone/hydrophone, which, for the aircraft case mentioned by *The Economist*, refers to the level averaged over several microphones ranged 100s to 1000s of metres away from the aircraft during take-off, where the intensity of the field has fallen off considerably as a result of geometrical spreading, absorption, scatter, etc. Chapman and Ellis, 1998). For this reason, just as no dB level should be used as an ‘absolute’ measure without quoting the reference level and the nature of the medium in which the measurement was taken, so too should it not be quoted without reference to the measurement position.

Fourth, as Chapman and Ellis (1998) point out, “there is no obvious connection between an annoying or harmful sound level for a human in air and an annoying or harmful sound level for a marine animal in water.”

Finally, the frequency range to which the measurement refers, and the bandwidth of the signal generated by the source, is important. As discussed in footnote 11, the bandwidth of the signal is important when considering how its ‘level’ is interpreted. Furthermore, if the signal has a small bandwidth, as some sonar sources do, then concentrating the acoustic output over a narrow frequency range may hit, or miss, a particularly sensitive part of the spectrum for the subject. Furthermore, to state that a given signal is ‘broadband’ does not impart enough information to make a judgement as to the associated hazards. A pseudorandom sequence (or ‘random’ noise), a chirp, and an impulsive sound can all have the same bandwidth, and could all be designed to produce the same ‘level’ with respect to a given reference intensity or pressure, at the same range from a source. However the subjective effect on the listener of each could be very different.

This last point, regarding the spectrum of the emission, is particularly relevant when discussing infrasound and ultrasound. To be specific, there are particular problems associated with the assessment of potential hazard for humans when exposed to ultrasound and infrasound, because of reliance on the vast body of experience at audio frequencies. There is some overlap between the problems experienced in working with ultrasound and infrasound, but also each has specific problems based on the fact that, whilst the infrasound of interest tends to be treated as ‘environmental noise’, the ultrasound in question tends to be treated as ‘signal’. This point will be discussed further below, using three example problems.

The first example issue relates to the A-weighted sound level, “the simplest and probably most widely used measure of environmental noise” (Kinsler et al., 1982). The measured sound spectrum is divided into frequency bands, and the energy in each band is weighted by a factor that reflects the sensitivity of the human ear to that frequency band (Bies and Hansen, 1996). These weighted energies are then combined to produce a single figure, given as dB(A), in an attempt to characterise the perceived loudness of a given environment. This procedure is so familiar that the steps behind producing a list comparing the ‘noise levels’ in various environments (ranking, for example, lawnmowers, rock concerts and aircraft on a single scale of A-weighted sound levels) are often ignored. As this weighting network was developed to protect against damage in the air due to sound within the audible frequency spectrum, A-weighting ignores energy which is transmitted within those third octave bands centred below 10 Hz or above 20 kHz. If therefore there is a ‘significant’ source of ultrasound (Grigor’eva, 1966; Acton, 1968, 1974, 1975, 1983; International Labour Office, 1977; International Non-ionizing Radiation Committee, 1984; auf der Maur, 1985; Damongeot and André, 1985, 1988; Lawton, 2001) or infrasound (von Gierke and Nixon, 1976) in the environment, then characterizing that environment by quoting the A-weighted decibel level will be inappropriate and misleading (see Section 6.2). The problems are further compounded when one considers what is meant by the word ‘significant’ in this context. For the purposes of the workshop with which this paper is associated, it should mean whether the ‘measured field’ can generate a potential hazard in humans.

This brings us onto our second problem: Both ultrasound and infrasound present difficulties in determining those ‘measured fields’ (Section 6). Furthermore, any past or current measurements of in-air ultrasound must
be critically questioned. This is because there is a lack of traceability for measurements in air at frequencies greater than 20 kHz. Although there is currently interest in increasing the upper frequency limit, primary standards for microphone calibration for sound in air have only been the subject of comparison between national measurement laboratories up to a frequency of 20 kHz. Measurements above this frequency cannot therefore be carried out in a traceable way using methods which have been the subject of international scrutiny and validation through the completion of formal comparisons, termed key comparisons (Zeqiri, 2005). Measurements in water can be traced to internationally validated standards in the frequency range 1–15 MHz, such that it is possible to measure the acoustic pressure amplitude at a given location and relate that back to a primary national standard. However, it is important to note that this does not tell us about the ultrasonic ‘dose’ given to humans in vivo, for example during foetal scanning (Duck, 1987; Duck and Martin, 1991; O’Brien, 1992; Siddiqi et al., 1995; Harris, 1999). Measurements of ultrasonic ‘exposure’ (which for example in the case of biomedical ultrasound refers to the measurements recorded in water) need to be undergo derating to allow estimation of the field levels which would occur in tissue (where acoustic pressure measurements are rarely made because of the invasiveness of hydrophones; Siddiqi et al., 1995). However the ultrasonic ‘dosage’ would refer to the quantitative determination of the interaction of ultrasonic energy with tissue, and this is currently not available. This point will be discussed further in Section 7.

Third, whilst it is not universally true, ultrasonic acoustic emissions tend to be dominated by man-made ‘signals’ (such as short pulses for diagnostic purposes, or the longer tone-burst or even continuous wave signals used for material processing). That is to say, their time-histories tend to be predictable. This is in contrast to emissions at audio or infrasonic frequencies, which tend to be viewed as environmental noise (even if they are impulsive in nature, such as can be generated by ordnance). As such it is very important to consider how the field is measured (and also, how it is then represented). Consider for example a diagnostic ultrasound field used for foetal imaging. Here the sound field consists of a series of pulses, each with a centre oscillatory frequency of a few MHz, and duration of a microsecond or so (the ‘on-time’). The interval between consecutive pulses consists of an ‘off-time’ of around 1 ms, so that there is a pulse repetition frequency (PRF) of about 1 kHz. Even at one specific location, this field has no single ‘intensity’, because (despite the rather loose use of the term ‘instantaneous intensity’) the measurement of intensity requires some time window for the measurement, and the result is an average over the duration of that time window. For example, a plane wave with time-dependent pressure field \( P \) over a given surface, would have an associated power per unit area based on the time-average (say from \( t_1 \) to \( t_2 \) of the product of the acoustic pressure and the particle velocity (Leighton, 1994, Section 1.1.7). Therefore using Eq. (1) to substitute for the particle velocity, the intensity at the given surface can be based on the product of the Real parts of the complex acoustic pressure and complex particle velocity:

\[
I = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \text{Re}(P) \text{Re}(v) \, dt
\]

\[
= \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \left( \frac{P + P^*}{2} \right) \left( \frac{v + v^*}{2} \right) \, dt
\]

\[
= \pm \frac{1}{4(t_2 - t_1)} \int_{t_1}^{t_2} \left( Zv + (Zv)^* \right) \left( v + v^* \right) \, dt,
\]

\[
(8)
\]

\[1\]When plotting a spectrum from the voltage \( V \) time history output of a sensor, there are a number of conventions. With the frequency usually plotted on the abscissa, the four most common options for the mantissa are: \( V \) \( Hz^{-1} \), \( V^2 Hz^{-1} \), \( V \), \( V^2 \). Clearly the representations that use \( V^2 \) in preference to \( V \) are plotting a parameter which reflects the energy of the signal, as opposed to the amplitude. The advantage of using \( Hz^{-1} \) comes from the common interpretation of a spectrum as a histogram, since the frequency bins will be finite. If this is the interpretation, then changing the width of these bins should affect the amplitude of the spectral level plotted. This is certainly appropriate for broadband signals. Environmental noise can fall into this category. However this methodology is problematic if the signal is a sine wave. Whilst not purely sinusoidal, some common ultrasonic fields can be sufficiently close for this to be an issue, for example the field used for ultrasonic physiotherapy or material processing. For a sine wave, the energy in the bin is independent of the bin width, since all the energy is at a single frequency, i.e. it has zero bandwith. Only one of the bins will contain non-zero energy. Division of the energy of that bin by the bandwidth of the bin simply causes the spectral peak corresponding to the sine wave to reduce in amplitude as the bin width increases. As a result, the parameters \( V \) and \( V^2 \) are sometimes used in preference to \( V Hz^{-1} \) and \( V^2 Hz^{-1} \) (a quantity resembling power), if the signal is perceived to more closely resemble a sinewave than a broadband signal (Leighton et al., 2005b).
where the * symbol indicates the complex conjugate. If the wave were sinusoidal\(^{12}\) this would quickly reduce to:

\[
I = \pm \frac{Z + Z^*}{4} v v^* \\
= \pm \frac{v v^*}{2} \text{Re}(Z).
\]

(9)

The intensity of a plane wave can also be interpreted in terms of the mean energy density of the wave, \(\phi_V = \rho_0 v v^* / 2\) (Leighton, 1994, Section 1.1.3). In time \(\Delta t\), a plane wave travelling at speed \(c_0\) will carry energy \(\phi_V \Xi c_0 \Delta t\) across a plane of area \(\Xi\) whose normal is aligned with the direction of propagation. Since the intensity equals the rate at which energy crosses a unit area of such a plane, then

\[
I = \frac{\phi_V \Xi c_0 \Delta t}{\Xi \Delta t} = \phi_V c_0 \\
= \pm \frac{\rho_0 c_0}{2} v v^*,
\]

(10)

which, when compared with (9), indicates that:

\[
\text{Re}(Z) = \rho_0 c_0.
\]

(11)

Therefore, when \(Z\) is real, (9)–(11) reduces to (1) and (4) for the simple harmonic wave discussed there. Fig. 4 shows a schematic of the pressure time history of some ultrasonic signal. Three possible intensities can readily be defined from it, with horizontal arrows to show their corresponding time windows (such that the left extremity of each arrow corresponds to a time which would be used as \(t_1\) in Eq. (8), whilst the right extremity occurs at the time which would be used as \(t_2\)). To calculate the pulse average intensity \((I_{PA})\), the time window (and subsequent average) is taken over the duration of the pulse. There is the temporal average intensity \((I_{TA})\), where the window is taken over many pulses. Hence \(I_{PA} > I_{TA}\) for the foetal imaging field under discussion, since the average which makes up \(I_{TA}\) will include measurement during the off-time. In turn, because the pulse takes time to ring up and ring down, the temporal peak intensity \(I_{TP}\) will be greater than the \(I_{PA}\). It is therefore important to quote which intensity is being used, not only to avoid ambiguity, but because one time window might be far more relevant than another to the process under discussion. For example, probably the two main physical processes of interest relating to hazard during ultrasonic foetal scanning are hyperthermia and the bubble-based activity known as inertial cavitation (see Section 4.3) (Miller and Ziskin, 1989; AIUM/NEMA 1992, 1998; AIUM 1993; Barnett et al., 1998). Hyperthermia is a relatively slow process (compared to the period of a single cycle of the insonifying field), with the potential for significant tissue heating occurring only over many pulses of the foetal imaging field discussed above. Therefore \(I_{TA}\) would be the more relevant intensity to use in discussion of the hyperthermia hazard. In contrast, whilst there are issues (such as the

---

\(^{12}\)Given that \(\cos^2 \omega t = (\cos 2\omega t + 1) / 2\), then if the interval \((t_2 - t_1)\) covers an whole number of oscillatory cycles, the integral of \(\cos 2\omega t\) over this window is zero, giving \(\int_{t_1}^{t_2} \left( \cos^2 \omega t \right) dt = \frac{1}{2} \int_{t_1}^{t_2} \cos 2\omega t dt + \frac{1}{2} (t_2 - t_1) = \frac{1}{2}\).
persistence of cavitation nuclei) which carry over from one pulse to the next (Leighton, 1994, Section 5.3), on the whole the cavitation effect is a far more rapid response than hyperthermia and, what is more, is a threshold phenomenon. Therefore, if only these three intensities were available, $I_{TP}$ would probably be the more useful than either $I_{PA}$ or $I_{TA}$ in discussion of the cavitation hazard. It is therefore no coincidence that the peak rarefaction pressure, the most widely accepted amplitude measurement for the discussion of cavitation hazard at a known frequency, contains much more information in common with the $I_{TP}$ than it does with either $I_{PA}$ or $I_{TA}$.

Of course there are numerous other ways of measuring intensities, and indeed it is important to include spatial as well as temporal information into measurement of intensity. This is not only because ultrasonic fields are often focused (leading for example to the use of the subscripts SA or SP to indicate spatial average or spatial peak measurements). It is also because, at all frequencies, acoustic waves have the propensity to be changed by the environment (Table 1 illustrates the propensity for reflection through the presence of small values for $(1-|R|^2)\times100\%$), which can lead to highly anisotropic fields, a feature which will be discussed further in Section 3.5.

3. Nonlinear propagation

3.1. The loss of linearity

The preceding section illustrated how complex notation can readily be used to describe several important features of linear acoustic wave propagation (as it can for waves of other types). This formulation is very widespread in acoustics because, such is the sensitivity of the human ear at audiofrequencies (Section 7), a great deal of the acoustic propagation that has been modelled is of low enough amplitude to allow for the use of linear theory. However at ultrasonic frequencies, the amplitudes used can be sufficiently high that nonlinear propagation occurs. One reason for this is because high source levels are often used to ameliorate signal-to-noise problems when ultrasonic absorption is high, and it does tend to increase with increasing frequency (see Section 2.2 and Table 2). In addition, without the constraint to avoid pain and hearing damage, ultrasound can be exploited to process the medium, which almost always requires higher acoustic pressure amplitudes than encountered at audio frequencies, where the response of the ear is the primary consideration.

Of course when such high source levels generate fields which propagate nonlinearly, the use of complex representation for a wave (which is ubiquitous throughout linear acoustics\textsuperscript{13}) is invalid. This can be simply illustrated through the quadratic process, which is probably the simplest and most common order of nonlinearity that is considered mathematically (consider the early terms of a Taylor Series). Whilst the square of the complex representation gives signal only at twice the original frequency (i.e. $(e^{i\omega t})^2 = e^{2i\omega t}$), trigonometric representation shows there are additional components in the signal, as can be seen from footnote 12. Indeed comparison with that footnote explicitly shows that the convention of using the real component of the complex entity to represent the measurable quantity is flawed if the process is nonlinear, since it reveals the errors in the above complex example (i.e. $(\text{Re}(e^{i\omega t}))^2 = \cos^2 \omega t \neq \text{Re}(e^{2i\omega t}) = \cos 2\omega t$). Higher powers act in a similar manner, e.g.

\begin{equation}
(\text{Re}(e^{i\omega t}))^3 = \cos^3 \omega t = (\cos 3\omega t + 3 \cos \omega t)/4 \neq \text{Re}(e^{3i\omega t}) = \cos 3\omega t.
\end{equation}

If the regime is indeed nonlinear, then one loses not only the ability to use complex representation of waves, but also a plethora of the most useful mathematical techniques in wave physics, including small amplitude expansions, Green’s function, Fourier transforms, superposition and addition of solutions.

\textsuperscript{13}Consider a parameter $u_1$ that is a function of the time $t$ (the argument applies equally to the spatial coordinates). Say that an operation $G$ transforms $u_1$ into $w_1$, so that $G(u_1) = w_1$. Similarly the operator might act on a second parameter, $u_2$, such that $G(u_2) = w_2$. If $G$ is a linear operator, then $G(u_1 + u_2) = w_1 + w_2$. Thus examples of linear operations on $e^{i\omega t}$ are: multiplication by a real constant, (i.e. $G(u_1) = \gamma u_1$); differentiation by time (i.e. $G(u_1) = \frac{\partial u_1}{\partial \tau}$); double-differentiation by time (i.e. $G(u_1) = \chi^2 \frac{u_1}{\partial \tau^2}$); and displacement in time (i.e. $G(u_1(t)) = u_1(t + \tau)$) where $\chi$ and $\tau$ are real constants. However important operations which are not linear include taking the square of a parameter (i.e. $G(u_1) = u_1^2$) and related functions.
3.2. The source of the nonlinearity

Perhaps the most important equation in acoustics is the one which defines the sound speed in terms of the equation of state for the medium $p = p(\rho, S)$ where $S$ is the entropy and $ho$ is the sum of all steady and unsteady pressures in the medium. In an infinite body of material that contains no dissipation, the sound speed $c$ may be defined through

$$c^2 = \frac{\partial p(\rho, S)}{\partial \rho},$$

(13)

which in turn describes the adiabatic or isentropic bulk modulus $B_S$, and the isothermal bulk modulus $B_T$, of the material$^{14}$:

$$B_S = \rho \left( \frac{\partial p}{\partial \rho} \right)_S, \quad B_T = \rho \left( \frac{\partial p}{\partial \rho} \right)_T.$$ (14)

Nonlinearity in the relationship between reversible, adiabatic variations in pressure $p$ and density $\rho$ in a medium of constant composition, can be characterised at small amplitudes through a Taylor series expansion of the equation of state along the isentrope for which the specific entropy $S$ takes the value $S_0$ (Beyer, 1998):

$$p - p_0 = \left. \frac{\partial p}{\partial \rho} \right|_{S_0} (\rho - \rho_0) + \frac{1}{2!} \left. \frac{\partial^2 p}{\partial \rho^2} \right|_{S_0} (\rho - \rho_0)^2 + \frac{1}{3!} \left. \frac{\partial^3 p}{\partial \rho^3} \right|_{S_0} (\rho - \rho_0)^3 + \cdots,$$ (15)

where $p_0$ and $\rho_0$ are the unperturbed values of pressure and density, respectively. The subscript 0 on the partial derivatives indicates that they are evaluated at the unperturbed state $(\rho_0, S_0)$ (Hamilton and Blackstock, 1998). Equating the perturbation $(p - p_0)$ to the acoustic pressure $P$, and writing the density variation as $\Delta \rho = \rho - \rho_0$ Eq. (15) can be rewritten as:

$$p - p_0 = P = A \left( \frac{\Delta \rho}{\rho_0} \right) + B \left( \frac{\Delta \rho}{\rho_0} \right)^2 + C \left( \frac{\Delta \rho}{\rho_0} \right)^3 + \cdots,$$ (16)

where

$$A = \left. \frac{\partial p}{\partial \rho} \right|_{S_0}, \quad B = \left. \frac{1}{2!} \frac{\partial^2 p}{\partial \rho^2} \right|_{S_0}, \quad C = \left. \frac{1}{3!} \frac{\partial^3 p}{\partial \rho^3} \right|_{S_0}.$$ (17)

Through (13) we can interpret the description of $A$ in (17) as describing the isentropic speed of sound for small amplitude signals, $c_0$:

$$c_0 = \sqrt{\frac{\partial p}{\partial \rho}} \bigg|_{S_0} = \sqrt{\frac{A}{\rho_0}}.$$ (18)

This is the sound speed at which the linear waves inherent in the descriptions of Section 2 propagate. Those wave descriptions stem from the fact that solutions to the plane wave equation, used for most of audio acoustics, are of the form $P = g(t \pm x/c_0)$ for the acoustic pressure and $v = f(t \pm x/c_0)/(\rho_0 c_0)$ for the particle velocity. The particular forms $e^{i(\omega t - kx)}$, used for the example of Eq. (2), form a subset of these general solutions. The $(t - x/c_0)$ argument represents waves travelling towards $x \to \infty$ whilst the $(t + x/c_0)$ argument represents waves travelling in the opposite direction, towards $x \to -\infty$.

Substitution shows that $P = g(t \pm x/c_0)$ and $v = f(t \pm x/c_0)/(\rho_0 c_0)$ are satisfactory solutions of the one dimensional linearised plane wave equation (often called the ‘plane wave equation’ by acousticians because of

$^{14}$The bulk modulus in turn is simply a three-dimensional form of the stiffness familiar from simple spring-bob oscillations, where the negative sign makes the bulk modulus positive for most (but not all) materials, specifically those which are compressed by an increase in pressure. The bulk modulus serves a stiffness-like role, in that the square root of the ratio of it to an inertial term, provides a characteristic time$^{-1}$ parameter for the material (that parameter being the sound speed for an acoustic wave (m s$^{-1}$); in the spring-bob, the root of the ratio of the stiffness to the bob mass provides the natural frequency (rad s$^{-1}$)).
the dominance of the assumption in acoustics that the propagation is linear), which is given below for pressure

\[ \frac{\partial^2 P}{\partial t^2} = c_0^2 \frac{\partial^2 P}{\partial x^2} \]  

(19)

and again for particle velocity:

\[ \frac{\partial^2 v}{\partial t^2} = c_0^2 \frac{\partial^2 v}{\partial x^2}. \]  

(20)

Such substitutions are readily undertaken, given that

\[ \frac{\partial P}{\partial t} = \frac{\partial g}{\partial (t \pm x/c_0)} \frac{\partial (t \pm x/c_0)}{\partial t} = g', \]

\[ \frac{\partial^2 P}{\partial t^2} = \frac{\partial g'}{\partial (t \pm x/c_0)} \frac{\partial (t \pm x/c_0)}{\partial t} = g'', \]

\[ \frac{\partial P}{\partial x} = \frac{\partial g}{\partial (t \pm x/c_0)} \frac{\partial (t \pm x/c_0)}{\partial x} = \frac{\partial g}{c_0} \frac{\partial (t \pm x/c_0)}{\partial x} = g', \]

\[ \frac{\partial^2 P}{\partial x^2} = \pm \frac{1}{c_0} \frac{\partial g'}{\partial (t \pm x/c_0)} \frac{\partial (t \pm x/c_0)}{\partial x} = \pm \frac{g'}{c_0} \frac{\partial (t \pm x/c_0)}{\partial x} = \frac{g''}{c_0}, \]

(21)

where \( g' = \frac{\partial g}{\partial (t \pm x/c_0)} \). Similarly, given \( f' = \frac{\partial f}{\partial (t \pm x/c_0)} \) then:

\[ \frac{\partial v}{\partial t} = \frac{1}{\rho_0 c_0} \frac{\partial f}{\partial (t \pm x/c_0)} \frac{\partial (t \pm x/c_0)}{\partial t} = \frac{f'}{\rho_0 c_0} \frac{\partial (t \pm x/c_0)}{\partial t} = \frac{f'}{\rho_0 c_0}, \]

\[ \frac{\partial^2 v}{\partial t^2} = \frac{1}{\rho_0 c_0} \frac{\partial f'}{\partial (t \pm x/c_0)} \frac{\partial (t \pm x/c_0)}{\partial t} = \frac{f''}{\rho_0 c_0} \frac{\partial (t \pm x/c_0)}{\partial t} = \frac{f''}{\rho_0 c_0}, \]

\[ \frac{\partial v}{\partial x} = \frac{1}{\rho_0 c_0} \frac{\partial f}{\partial (t \pm x/c_0)} \frac{\partial (t \pm x/c_0)}{\partial x} = \frac{f'}{\rho_0 c_0} \frac{\partial (t \pm x/c_0)}{\partial x} = \frac{f'}{\rho_0 c_0^2}, \]

\[ \frac{\partial^2 v}{\partial x^2} = \pm \frac{1}{\rho_0 c_0^2} \frac{\partial f'}{\partial (t \pm x/c_0)} \frac{\partial (t \pm x/c_0)}{\partial x} = \pm \frac{f''}{\rho_0 c_0} \frac{\partial (t \pm x/c_0)}{\partial x} = \frac{f''}{\rho_0 c_0}. \]

(22)

The following analysis will demonstrate two features. First, that the linearized plane wave equation is not valid for waves of finite amplitude, and hence, if we are strictly accurate, linear acoustic waves of the form of (19) and (20) never propagate in fluids (gases and liquids). Most of audiofrequency acoustics, of course, operates on the assumption that the wave amplitude is so small that the linear representation is accurate to within our ability to measure it. As discussed in Section 7, this is often not the case in ultrasonic work. Hence the second purpose of the following analysis is to illustrate the important terms which must be small in order for linear propagation models to be adequate.

The characterization of the propagation of an acoustic wave in a fluid requires three fundamental inputs enshrined in equations reflecting: first, the conservation of mass in a fluid; second, the fluid dynamic properties relating such motions to the pressure gradient which causes them; and third, an equation which shows that pressure gradient to be part of a longitudinal wave. Stated in one-dimensional form, these three equations are, respectively: the equation of continuity:

\[ \frac{\partial \rho}{\partial t} + \frac{\partial (\rho v)}{\partial x} = 0, \]

(23)

where \( v \) is the particle velocity, and \( \rho \) the fluid density; Euler’s equation for an inviscid fluid:

\[ v \frac{\partial v}{\partial x} + \frac{\partial v}{\partial t} = - \frac{1}{\rho} \frac{\partial p}{\partial x}, \]

(24)
where \( p \) is the sum of all steady and unsteady (assumed here to be purely acoustic) pressures; and an equation which relates the wavespeed to the equation of state (Eq. (13)).

Combination of these three equations allows formulation of the propagation of acoustic longitudinal waves in a fluid, linearisation allowing the generation of what is generally termed the (linearised) plane wave equation. In the small amplitude regime where this is applicable, the sound speed is represented by the phase speed in the linear limit \( c_0 \), and the approximation\(^{15}\) made that \( \rho^{-1} \approx \rho_0^{-1} \) where \( \rho_0 \) is the equilibrium density:

\[
\rho = \rho_0(1 + \Delta \rho/\rho_0) \Rightarrow \rho^{-1} = \rho_0^{-1}(1 + \Delta \rho/\rho_0)^{-1} \approx \rho_0^{-1}(1 - \Delta \rho/\rho_0 + \cdots) \\
\Rightarrow \rho^{-1} \approx \rho_0^{-1} - \Delta \rho/\rho_0^2 + \cdots \\
\Rightarrow \Delta \rho/\rho_0 \approx \Delta \rho/\rho_0 - (\Delta \rho/\rho_0)^2 + \cdots \\
\Rightarrow \Delta \rho/\rho_0 \approx \Delta \rho/\rho_0 \Rightarrow \rho^{-1} \approx \rho_0^{-1}, \quad \left(\frac{\Delta \rho}{\rho_0} \ll 1\right). \tag{25}
\]

In this linear limit Eq. (13) gives

\[
\frac{\partial^2 \rho}{\partial t^2} - \frac{\rho_0}{c_0^2} \frac{\partial \rho}{\partial x} = \frac{\partial^2 \rho}{\partial x^2} = \frac{\partial^2 \rho}{\partial x^2} 
\]

and substitution of the first equation of (26) into (23), with (25) (i.e. \( \rho^{-1} \approx \rho_0^{-1} \)), gives:

\[
\frac{1}{c_0^2} \frac{\partial \rho}{\partial t} + \frac{\rho \rho / \partial x}{c_0^2} = 0 \Rightarrow \frac{\partial \rho}{\partial x} + \frac{v}{\rho_0} \frac{\partial \rho}{\partial x} = - \frac{1}{\rho_0 c_0^2} \frac{\partial \rho}{\partial t}. \tag{27}
\]

Similarly, if from (25) we have \( \rho^{-1} \approx \rho_0^{-1} \), then Euler’s Eq. (24) becomes:

\[
\frac{\partial v}{\partial x} + \frac{v}{\rho_0} \frac{\partial v}{\partial x} = - \frac{1}{\rho_0} \frac{\partial \rho}{\partial x}. \tag{28}
\]

Linear acoustics are based on the assumption that two of the terms in the above series of equations are negligible, specifically that \( |(v/\rho_0)(\partial \rho/\partial x)| \ll |\partial v/\partial x| \) when calculating \( (\partial \rho/\partial t)/(\rho_0 c_0^2) \) using (27); and that \( |(v/\rho_0)(\partial \rho/\partial x)| \ll |\partial v/\partial t| \) when calculating \( (\partial \rho/\partial x)/\rho_0 \) using (28). Differentiation of (27) with respect to \( t \), and of (28) with respect to \( x \), gives the one dimensional linearised plane wave equation for pressure (19) if the equivalence \( \left(\frac{\partial^2 v}{\partial t \partial x}\right) = \left(\frac{\partial^2 v}{\partial t \partial x}\right) \) is made. As discussed above, Eq. (19) describes a linear pressure wave propagating at speed \( c_0 \), and has solutions \( P = g(t \pm x/c_0) \), proven using the substitutions of (21) into (19).

Similarly differentiation of (27) with respect to \( x \), and of (28) with respect to \( t \), with the same two terms deemed to be negligible, gives the one dimensional linearised plane wave equation for particle velocity (20), with solutions \( v = f(t \pm x/c_0)/(\rho_0 c_0) \), proven using the substitutions of (22) into (20).

If the linear solutions \( P = g(t \pm x/c_0) \) and \( v = f(t \pm x/c_0)/(\rho_0 c_0) \) are valid, then the conditions which make them valid (i.e., which ensure \( |(v/\rho_0)(\partial \rho/\partial x)| \ll |\partial v/\partial x| \) and \( |(v/\rho_0)(\partial \rho/\partial x)| \ll |\partial v/\partial t| \) are readily found. Taking the first inequality, \( |(v/\rho_0)(\partial \rho/\partial x)| \ll |\partial v/\partial x| \), substitution from (22) gives

\[
\frac{\partial v}{\partial t} = f' \quad \text{and} \quad \frac{\partial v}{\partial x} = \frac{v}{c_0} f' \Rightarrow \frac{\partial v}{\partial x}/\frac{\partial v}{\partial t} = \pm \frac{v}{c_0}. \tag{29}
\]

Consider the second inequality \( |(v/\rho_0)(\partial \rho/\partial x)| \ll |\partial v/\partial x| \) in the same linear limit, where (26) is valid, and recall that all perturbations in pressure are assumed to be acoustic. Since neglect of the nonlinear term in (28) implies

\(^{15}\text{Linear acoustics requires the approximation } \rho^{-1} \approx \rho_0^{-1} \text{ which, as Eq. (25) shows, is obtained by ignoring the quadratic (and higher) terms (i.e. } (\Delta \rho/\rho_0)^2 + \cdots \text{). It is not sufficient to say that, given } \rho = \rho_0 + \Delta \rho \text{ and } (\Delta \rho/\rho_0) \ll 1, \text{ we can in turn make the approximation that } \rho \approx \rho_0, \text{ since that would be to ignore the first order perturbation, an incompressible assumption which is at odds with the concepts behind linear acoustics (Fig. 2) and would, for example, imply infinite sound speeds.}\)
that \( \frac{\partial v}{\partial t} = \frac{\partial p}{\partial x} / \rho_0 \), then substitution from (26) and (18) reduces the second ratio to

\[
\frac{v}{\rho_0 c_0} \frac{\partial p}{\partial x} \frac{\partial v}{\partial x} = \left( \frac{\partial p}{\partial x} / \rho_0 c_0^2 \right) \left( \pm f'/\rho_0 c_0^2 \right)
\]

\[
= -v \rho_0 \frac{\partial v}{\partial t} \left( \pm f' \right)
\]

\[
= -v \rho_0 \left( f' / \rho_0 c_0 \right) \left( \pm f' \right)
\]

\[
= \mp \frac{v}{c_0}.
\]  

Hence the linear limit is approached as the acoustic Mach number \((v/c_0)f\) becomes small and the two nonlinear terms become negligible. The consequence of finite acoustic Mach numbers is that Eqs. (19) and (20) no longer hold, and a waveform which is initially sinusoidal will not remain so because of two nonlinear effects which act co-operatively. Both can be readily understood through the realization that, if dissipation is small, then \( P \) and \( v \) would be in phase (Leighton, 1994, Section 1.2.3(a); Fig. 2). First, there is a convection effect. In simple terms, if \( v/c_0 \) is not negligible, then parts of the wave tend to propagate as \( c + v \). The particle velocity varies throughout the wave, and so the greater the local acoustic pressure, the greater the velocity of migration of that section of the wave. Consider the case when \( P \) and \( v \) are in phase (as occurs if the impedance is real—see Section 2.1). This was shown in Fig. 2, and the particle velocity and pressure plots from there are repeated in Figs. 5(a) and (b). However the particle velocities are superimposed as vector arrows on the pressure plot of Fig. 5(b). Unlike the linear propagation of Fig. 2, during the nonlinear propagation of Fig. 5 the regions of compression (where the \( v \) and \( c \) are in the same direction) would tend to migrate faster than the regions of

Fig. 5. Schematic showing the development of an infinite, sinusoidal plane wave during nonlinear propagation. The first two graphs (repeated from Fig. 2) show the initial sinusoidal waveform in terms of (a) particle velocity and (b) pressure. They are both plotted as a function of position \( x \), with the particle velocities are superimposed as vector arrows on the pressure plot of (b). Unlike during the linear propagation of Fig. 2, these waveforms propagate non-linearly, such that as it propagates away from the source, the pressure waveform distorts (c), and if dissipation is not too great, can eventually form a shock (d). Enhanced absorption of the higher frequencies then causes the amplitude to decrease (e), until eventually the ‘old age’ waveform is sinusoidal once more (f). Parts (a) to (f) of the figure plot the pressure waveforms as a function of \( x \); Part (g) plots the pressure waveform of (d) as a function of time \( t \).
rarefaction (where \( v \) is opposite to \( c \)). The pressure peak propagates with the greatest speed, the trough with the least (Fig. 5(b)).

Second, there is an effect which arises because, when a fluid is compressed, its bulk modulus and stiffness increase. This results in an increase in sound speed, and this effect too will cause the pressure peaks to travel at greater speed than the troughs, and tend to try to catch up and encroach upon them (Fig. 5(c)).

A continuous wave that is initially sinusoidal (Fig. 5(a) and (b)) will therefore distort. The compressional parts of the wave ‘catch up’ on the rarefaction components as the propagation progresses, such that it becomes distorted\(^{16} \) (Fig. 5(c)). Such distortion of course is accompanied by the appearance in the spectrum of energy at harmonics of the driving frequency. Balancing this tendency is, of course, the absorption of the medium: if it is very great, the wave amplitude will decrease so rapidly that, even at high source levels, high Mach numbers will not be maintained for much of the propagation path away from the source, and the nonlinear distortion is negligible. However where absorption is not so great, such distortion can proceed until a shock wave develops (Fig. 5(d)). The discontinuity length (or shock-formation distance, \( L_{\text{dis}} \) see Eq. (42)) is taken to be the distance propagated by acoustic plane waves in a lossless medium when an infinite slope first appears in the waveform. For a shock to develop in this way, the absorption must not have been so great as to dissipate the energy in the wave before the shock can develop. Just as the discontinuity length is characteristic of the distance over which the wave would propagate in the absence of dissipation in order to exhibit this strong nonlinear feature, so the e-folding length of Eq. (7) typifies the distance over which the wave will propagate in the absence of nonlinearity in order to show the effects of absorption strongly. The competition between the influences of nonlinearity and viscothermal absorption is characterised by the dimensionless Gol'dberg number \( \Gamma_G \) (Gol'dberg, 1956), equal to:

\[
\Gamma_G = \frac{1}{bL_{\text{dis}}} = \frac{2L_e}{L_{\text{dis}}},
\]

where \( b \) is the attenuation coefficient for small amplitude waves (Eq. (5)), and \( L_e \) is the e-folding distance for energy (Eq. (7)). Since for the waves of Section 2.2, the length \( 2L_e \) equals the e-folding distance for the amplitudes of acoustic pressure, displacement and particle velocity, then the Gol’dberg number can be thought of as the ratio of two lengths, one characterising the dissipation and the other characterising the nonlinearity. Both lengths pertain to the distance through which the wave would have to propagate, in the absence of the competing effect, for their characteristic effect to manifest itself strongly. A large value of the Gol’dberg number indicates that the nonlinearity has generated a shock long before the wave has been strongly absorbed.

In Fig. 5(d), the Gol’dberg number has been sufficiently great for a shock to develop. However, as discussed above, the process of nonlinear distortion is, in the frequency domain, associated with the ‘pumping’ of energy from the fundamental frequency up to higher frequencies where the absorption is greater (for most but not all materials—see Fig. 15). There is therefore enhanced absorption of this energy, such that further propagation is accompanied by a decrease in wave amplitude. The waveform amplitude decreases (Fig. 5(e)). Eventually this hypothetical wave is sinusoidal once more, as the only unabsorbed energy is at the fundamental (Fig. 5(f)). If the absorption is greater at higher frequencies, then whilst this ‘old age’ waveform might resemble the sinusoid that would have been detected had only linear propagation occurred, its amplitude is less than would have been found at equivalent distances from the source had the same initial wave propagated only linearly. Similarly the higher absorption would mean more acoustic energy has been converted to heat during this propagation path, than would have occurred had only linear propagation of the same initial waveform taken place.

Fig. 5(g) plots the shocked waveform as a function of time \( t \), instead of position \( x \) (as was used for the rest of Fig. 5 and for Fig. 2). This is because the explanations in both Fig. 2 and Fig. 5 are easier with respect to the spatial coordinate. However most measurements are made with respect to the temporal coordinate. In comparing Figs. 5(d) and (g), note that the appearance of the waveform is reversed: this is because that

\(^{16}\text{An additional feature (not apparent in Fig. 5 because it illustrates the distortion in an infinite plane wave) arises from the action of dissipation and diffraction, which cause phase shifts in the various frequency components of the wave. As a result, the distorted waveform is not symmetrical about the zero-line: in general the trough, which corresponds to negative values of the acoustic pressure, becomes rounded whilst the positive peak is augmented.}\)
portion of a waveform which arrives first at a fixed measuring point in the medium (such as a fixed microphone or hydrophone) is plotted at the earliest arrival times in Fig. 5(g). This has important uses: for example, whilst explanation of nonlinear propagation is most easily done by referring to waveforms plotted as a function of $x$, the waveforms appear to be reversed when monitored using a fixed hydrophone and
the fact that acoustic absorption by a bubble population tends to peak, rather than simply increase with frequency (MHz). Measurements were made with a 2 × 9 μm² bilaminar membrane hydrophone (Marconi, with a high gain, 100 MHz bandwidth preamplifier). This had an active element of diameter 0.5 mm, separated from the transducer face by (a) 2 mm, (b) 24 mm, (c) 68 mm (the position of the last axial maximum, i.e., the ratio of the square of the faceplate radius to the acoustic wavelength; see Eq. (69)), (d) 136 mm, (e) 300 mm, (f) 570 mm, determined from the time relative to the transducer firing. Data were recorded by a Tektronix TDS 784D DPO oscilloscope (50 ns/div. 5000 point waveforms). The experimental system was aligned according to IEC 61102 prior to measurement (International Electrotechnical Commission, 1991): the water temperature was 18.7 °C. (Measurements taken at the request of the author by M. Hodnett and B. Zeqiri, National Physical Laboratory, UK).

Fig. 5 therefore explains the data in Fig. 6, where the waveform is detected by a fixed hydrophone and displayed on an oscilloscope. The real data in (i) illustrate the time history of the wave, measured at increasing distances ((a)–(f)) from these sources. The corresponding spectra are shown in (ii). Close to source, the waveform is initially nearly sinusoidal (a(i)) and single frequency (a(ii)). As it propagates through the medium, each compressive region gains upon the preceding rarefactive half-cycle, the peak positive acoustic pressure appearing earlier and earlier in the time history compared to the peak rarefaction. An accumulated steepness of the waveform between the two develops (b(i)) and harmonics appear in the spectrum (b(ii)). After propagating a distance corresponding to the discontinuity length, the waveform includes a discontinuity, in that a shock wave develops (c(i)). A continuum component may increase in the spectrum (c(ii)). The contribution of higher harmonics to the waveform is clearly visible (c(i)). Note that the amplitude of the time history decreases, because any further compressional advance leads to dissipation and results in a reduction in the amplitude of the shock. This is because the waveform distortion has been equivalent to transferring some of the energy of the initial wave to higher frequencies, which are more strongly absorbed (d,e) The energy transfer is not sufficient to maintain the shock, and the wave approaches a low-amplitude sinusoidal form (termed ‘old age’) (f).

The vastness of the frequency range covered in underwater ultrasonics has several implications (see Section 5.1). No one hydrophone can span it (Fig. 7), and the calibration of hydrophones (which may themselves be invasive, directional etc.) requires very great skill. The nonlinear propagation described in the preceding paragraph is one tool which is exploited to decrease the time taken for a hydrophone calibration. A standard calibrated hydrophone is placed sufficiently far from the source to detect energy at many harmonics (as in Fig. 6(d)). If the sound field remains constant when the calibrated sensor is replaced by the unknown sensor and monitors the same positioning the sound field, comparison of the two measured signals allows many frequencies to be calibrated simultaneously.

To summarise the effect on signal-to-noise ratios (SNRs) during nonlinear propagation, energy is pumped from lower to higher frequencies, where it is preferentially absorbed. This means that the net attenuation over distance will be greater if nonlinear propagation occurs, than if conditions were linear. Furthermore, a narrow band detector tuned to the frequency of the emitted pulse might fail to detect energy in the returned signal which is outside of its bandwidth (and hence ‘invisible’ to it). Both of these will act to reduce the signal-to-noise of the received signal, a problem which may not be alleviated by simply increasing the amplitude of the emitted pulse (since this enhances the nonlinear effects, a phenomenon known as “acoustic saturation”—see Section 3.4).

Propagation such as that described above is just one of the possible sources of nonlinearity: others include the transducer itself, and entities within the water column (see Section 4). As the earlier discussion of the Gol’dberg number showed, the degree to which the effects occur for a given acoustic signal in a given environment of course depends not only on the nonlinearity (and hence the rate at which energy is transferred from lower to higher frequencies) but also on the absorption. Some features, such as bubbles, increase both, and it may be that the absorption is so great that nonlinearity is not present in the received signal (although the fact that acoustic absorption by a bubble population tends to peak, rather than simply increase with
frequency, may hold opportunities for exploitation of nonlinear effects; see Fig. 15(c)). However its possibility during propagation must be appreciated. This is not only because nonlinear propagation might cause enhanced absorption, or may make some energy in the received signal ‘invisible’ if it is outside the bandwidth of the detector, but also because it might be exploited. The ability of the propagating medium to generate multiple frequencies, as illustrated in Fig. 6, could not only be used to diagnose the properties of that medium (e.g., bubbly water). It could also be used to generate a signal containing harmonics across a wide frequency
range for simultaneous testing of the scatter at many frequencies from the seabed or a target (the propagation after scattering, being of lower amplitude, would tend to be linear and so preserve the frequency characteristics of the target within the usual linear constraints of absorption, etc.).

A simple model for the transference of some energy from lower to higher frequencies (which does not include the critical absorption component) can be found in the Taylor series expansion (16). This is adequate to demonstrate the generation of harmonics through a nonlinear process (including propagation). If the nonlinear fluid element (in this case the liquid) is subjected to single-frequency insonification \( P(t) \propto \cos \omega t \), then the second and third harmonics are generated by the quadratic \( B \) and cubic \( C \) terms of Eq. (16), with of course additional frequencies as indicated by the trigonometric expansions in footnote 12 and Eq. (12).

There are many ways in which the nonlinearity has been exploited in biomedical ultrasonics (Duck, 2002). The next section however will concentrate on only one application, that of parametric sonar, since it directly shows use of the quadratic nonlinearity, and also provides the background for one of the in-air applications of ultrasound. This is important because, as will be shown in Section 6, most of the in-air applications of ultrasound tend to be based rather on the generation of high signal amplitudes rather than cogent understanding and exploitation of the nonlinearities so generated.

### 3.3. Parametric sonar

The Taylor series description of the nonlinearity (16) can also be used to illustrate what happens if the medium responds nonlinearly to insonification by an acoustic field consisting of two coherent frequencies. For example, the initial waveform \( P(t) \propto \cos \omega_1 t + \cos \omega_2 t \) can generate combination-frequency signals at \( \omega_1 \pm \omega_2 \), as well as at \( 2\omega_1, 2\omega_2, \omega_1 \pm 2\omega_2, 2\omega_1 \pm \omega_2, \) etc. since the quadratic terms alone gives \( 2\cos \omega_1 t \cos \omega_2 t = \cos(\omega_1 t + \omega_2 t) + \cos(\omega_1 t - \omega_2 t) \). Because of this, nonlinear propagation can be exploited to generate a beam of sound far narrower than would be possible for a given source size, were the propagation to behave linearly. In general, to obtain a highly directional acoustic beam, the size of the source must be very much greater than a wavelength (see section 5). Conversely, if the wavelength is much larger than the acoustic source, the sound field tends to be emitted in an omnidirectional manner. This leads to a number of problems and solutions. For example, if one wishes to produce a narrow beam of ‘low’ frequency sound (i.e., sound for which the wavelength is not significantly less than the source size), linear techniques are precluded by the source size required. For example, a 500 Hz sonar beam would have wavelengths in bubble-free water of \(~3\,\text{m},\) and clearly mounting a source much larger than this on, say, the front of a ship is not feasible. Similarly if one has museum exhibits (e.g., paintings) spaced every 2 m and one wishes to direct a beam of sound from the ceiling to a 1 m\(^2\) region in front of each exhibit in order to carry recorded spoken information about that exhibit, the source size would have to be much greater than the wavelength of, say, 500 Hz in air, which is \(~0.7\,\text{m}.\)

Nonlinear acoustics offers a possible solution to these problems (Bennett and Blackstock, 1975; Yomeyama et al., 1983). Consider two sources, one driven at \( \omega_1 \) and the other at \( \omega_2 \), which are physically big enough to generate narrow ultrasonic beams. These two beams (called the ‘primaries’) are directed such that they overlap in the medium. In the region where they overlap, they can generate a highly directional propagating wave which contains the various frequencies listed above, and more. The signal at the ‘difference frequency’ \( \omega_1 - \omega_2 \) is of particular interest.\(^{17}\) for several reasons. First, the directionality of the emission is far greater than could be obtained with a source of the same size operating at this frequency directly. Second, depending on the medium, the higher frequencies can be more strongly absorbed, so that only the difference frequency propagates to distance (the use of parametric sonar in the oceans in a good example of this). Third, in many practical situations, only the difference frequency is within the bandwidth of the detector (the museum scenario cited above would provide an apt example, the detector here being the human ear). Probably the

\(^{17}\)Note that this generation of a propagating wave at the difference frequency is distinct from the purely linear process of generating beats, which occurs when waveforms are superimposed at the detector. Another distinct process by which difference frequencies (and other signals) can be generated is through nonlinearities in the receiver (e.g. the ear, preamplifier, or microphone; see Section 6.4). These can act without necessarily there being nonlinearity present in the propagation, as can nonlinearities in the transmitter or the data acquisition process.
major drawback of this application, which has implications for the safe use of ultrasound, is that it is in general very difficult to generate high intensities at the difference frequency using this system (because it is a second-order process, based for example on the quadratic term in Eq. (16)). This could lead to the use of very high levels for the primary signals in such applications. Given the paucity of information on the safe levels for human exposure to ultrasound in air (Section 6), and the lack of traceability for the measurement of such fields (see Section 2.3), this could be a safety issue.

There are in fact many commercial systems available which probably exploit nonlinear ultrasonic interactions to generate localised audiofrequency sound. HyperSonic SoundTM advertise: “This ability to direct or focus sound into a tight beam has a wealth of applications. Imagine: directing narration in a museum only to the people standing in front of a specific display; capturing and holding customers’ attention to advertise a product or promote a brand at the point-of-purchase, without disturbing employees or causing unwanted noise pollution; providing information or messages that can offer direction to shorten wait times to people standing in line; directing an alarm or alert only to the intended operator in a control room environment to assist them in making timely, critical decisions.” They cite the following applications: “Digital Signage/NarrowCasting; In-Store Advertising; Museums; Trade Shows; Kiosks; Corporate Lobbies; Command & Control Room; Automotive Dealerships”. As the use of other acoustic technologies (such as active noise control and virtual acoustic systems) increases, there may be a proliferation in the application of such parametric technology (to provide, for example focused control sources) to implement them more effectively. This topic of human exposure to high levels of ultrasound is discussed further in Section 6.

3.4. The material and convective nonlinearities

As outlined in Section 3.2, there are two contributions to the change in phase speed of the wave, the change in stiffness (or bulk modulus) and the convection. These are called the material nonlinearity and the convective nonlinearity, respectively, and are formulated as follows (Leighton, 1994, Section 1.2.3). The propagation velocity of a point in the wave having particle velocity \( v \) is

\[
\text{Propagation velocity} \bigg|_{v=\text{constant}} = \frac{dx}{dt} \bigg|_{v=\text{constant}} = v + c,
\]

where a convection component \( v \) is added to \( c \), the local speed of sound. This local speed has in turn been affected by the change in bulk modulus, and is related to \( c_0 \), the sound speed for waves of infinitesimal amplitude. This effect can be determined by substituting Eq. (16) into (13) to give:

\[
\frac{c}{c_0} = \sqrt{1 + \frac{B}{A} \left( \frac{\Delta \rho}{\rho_0} \right) + \frac{C}{2A} \left( \frac{\Delta \rho}{\rho_0} \right)^2 + \ldots} \\
\approx 1 + \frac{B}{2A} \left( \frac{\Delta \rho}{\rho_0} \right) + \frac{1}{4} \left( \frac{C}{A} - \frac{1}{2} \left( \frac{B}{A} \right)^2 \right) \left( \frac{\Delta \rho}{\rho_0} \right)^2 + \ldots
\]

assuming isentropic conditions and that it is valid to perform this binomial expansion. The ratio \( B/A \) is called the second-order nonlinearity ratio of the liquid and characterises the potential size of the nonlinearity. Following Beyer (1998), we make use of the simple relationship for progressive plane linear waves, in isentropic conditions, that follows from Eq. (1) and (18) by, for example, retaining only the first term of the expansion (16)

\[
\frac{\Delta \rho}{\rho_0} \approx \frac{P}{A} \approx \frac{p_0 c_0 v}{\rho_0 c_0^2} = \frac{v}{c_0}
\]

\[18\]See HyperSonic Sound (http://www.atecsd.com/hss.html); holosonics (http://www.holosonics.com). Note that the web-based material currently available on such products does not allow the author fully to assess the mechanisms or compare the levels with guidelines (Section 6.4), and until this is done any assessment of the potential or otherwise for hazard cannot be made.
using (18) to substitute for \(A\). Retaining only the first two terms in the expansion (33) and substituting for \(\Delta \rho/\rho_0\) from (34) gives:

\[
c \approx c_0 + \frac{B}{2A} v
\]

The equivalent expression for an isentropic gas can readily be obtained by noting that the expansion of the equation state along an isentrope for a perfect gas gives:

\[
\frac{p}{p_0} = \left(\frac{\rho}{\rho_0}\right)^\gamma = \left(1 + \frac{\Delta \rho}{\rho_0}\right)^\gamma = 1 + \gamma\left(\frac{\Delta \rho}{\rho_0}\right) + \frac{\gamma}{2!}(\gamma - 1)\left(\frac{\Delta \rho}{\rho_0}\right)^2 + \ldots
\]

where \(\gamma\) is the ratio of the specific heat at constant pressure to that at constant volume. Term-by-term comparison with (16) gives (Beyer, 1998):

\[
A = \gamma, \quad B_A = \gamma - 1, \quad C_A = (\gamma - 1)(\gamma - 2)
\]

Substitution of (37) into (35) indicates that:

\[
c = c_0 + \frac{\gamma - 1}{2} v
\]

for a perfectly isentropic gas. Therefore, using (35) and (38), Eq. (32) can be rewritten to include both the material and convective nonlinearities explicitly:

\[
\text{Propagation velocity} \bigg|_{\text{constant}} = \frac{dx}{dt} \bigg|_{\text{constant}} = v + c
\]

\[
= c_0 + \left(1 + \frac{\gamma - 1}{2}\right) v \quad \text{for a perfectly isentropic gas, and}
\]

\[
= c_0 + \left(1 + \frac{B}{2A}\right) v \quad \text{for a liquid.}
\]

In the brackets in Eq. (39), the unity term corresponds to the contribution from the convective nonlinearity, and the other term to that from the material nonlinearity. For water the ratio \(B/2A\) equals 2.5, whereas for air the isentropic gas equivalent is \((\gamma-1)/2 = 0.2\). Therefore in water the dominant cause of the waveform distortion is nonlinearity of the equation of state (the material nonlinearity), whereas in air the distortion arises mainly through convection.

The sum of the material and convective nonlinearities is given by the ‘coefficient of nonlinearity’ \(\zeta\), where from Eq. (39):

\[
\text{Propagation velocity} \bigg|_{\text{constant}} = \frac{dx}{dt} \bigg|_{\text{constant}} = c_0 + \zeta v,
\]

where

\[
\zeta = \frac{\gamma + 1}{2} \quad \text{for a perfectly isentropic gas, and}
\]

\[
\zeta = 1 + \frac{B}{2A} \quad \text{for a liquid.}
\]

For air \(\zeta = 1.2\), and for water \(\zeta = 3.5\). From this one might expect the distortion to be more obvious in water than in air; however one must take into account attenuation of high frequencies. The discontinuity length is given by

\[
L_{\text{dis}} = \frac{1}{\zeta Mk},
\]
where $k$ is the wavenumber and where $M$ is the peak acoustic Mach number of the source, the ratio of the amplitude of the particle velocity at the source to $c_0$. Thus (assuming large Gol'dberg numbers; Eq. (31)) the shock forms closer to the source as the coefficient $\zeta$, which indicates the degree of nonlinearity, increases. It also forms sooner as the amplitude increases, because the speed differential between the peaks and the troughs increases. Similarly $L_{\text{dis}}$ decreases with increasing frequency, since this signifies decreasing wavelength, and so the peaks have to travel a shorter distance to catch up with the troughs. If the point of observation is at a distance greater than $L_{\text{dis}}$ then as either amplitude or frequency are increased the discontinuity length decreases, the point where the shock forms moving closer to the source. This means that increasing amounts of energy can be dissipated at the shocks beyond $L_{\text{dis}}$ before the wave arrives at the point of observation. If the energy supplied to the wave at the source is continually increased, there comes a point at which the increased dissipation between the discontinuity point and the point of observation outweighs the increase in energy supplied to the wave at the source. Beyond this critical source power, the wave ceases to be dependent upon the source amplitude: any additional energy supplied to the wave by the source is lost at the shock front, and \textit{acoustic saturation} is said to have occurred (Leighton, 1994, Section 1.2.3; Leighton, 1998; Duck, 1999; Duck, 2002).

3.5. Other phenomena

Consider a plane wave, travelling in the $+x$ direction at speed $c$, approaching a wall in the $yz$ plane of area $\Xi$ (Fig. 8). The wave energy is completely absorbed by the wall. If the wave has intensity $I$, then the energy absorbed by the wall in a time $\Delta t$ is $I\Xi\Delta t$. The wall must have applied a force $F_r$ in the $-x$ direction to stop the wave motion, which in time $\Delta t$ acted over a distance $c\Delta t$. Therefore the work done by the wall on the wave is $F_r c\Delta t$. Equating this to the energy absorbed, we obtain $F_r = I\Xi/c$. From Newton’s third law of motion, this must be equal and opposite to the force exerted by the wave on the wall. Therefore upon absorption the wave exerts a \textit{radiation pressure} ($p_{\text{rad,abs}}$) in the direction of its motion of magnitude

$$p_{\text{rad,abs}} = I/c$$

for normal incidence of plane waves.\textsuperscript{19} The force $F_r$ exerted by the wall can also be thought of as acting upon the wave to absorb its momentum. In time $\Delta t$ the wall absorbs a length $\Delta L = c\Delta t$ of the wave, exerting an

\textsuperscript{19}Demonstrations of radiation pressure are not confined to acoustics laboratories, but can frequently be seen in science fiction films, where ray guns appear to have the ability to impart a force capable of making the victim (usually a Star Wars Imperial stormtrooper) stagger backwards. Likening this motion to that observed when catching a 5 kg bag of potatoes, suggests the radiation force is 50 N. This implies that, during firing, the ray gun would need to project power $W$ equal, from Eq. (43), to a force of (50c) N, where $c$ is the wavespeed. Were the ray-gun to project electromagnetic radiation, such as a laser beam, the power of the gun would have to be $50 \times 3 \times 10^6 = 1.5 \times 10^{10}$ W. This is more than ten times the combined power output of the two 600 MW turbines at the Torness
impulse $F \Delta t = \Xi \Delta L / c^2$ upon the wave, causing a change in momentum of $\Delta \Gamma$. Since after absorption the momentum of wave is zero, then the momentum associated with one wavelength of the wave (setting $\Delta L$ equal to $\lambda$ in the above) is:

$$\Delta \Gamma_{\lambda} = I \Xi \lambda / c^2.$$  \hspace{1cm} (44)

If this normally incident wave is reflected, instead of being absorbed, this momentum must be not simply absorbed but reversed. The wall must exert twice as much force upon the wave, and so the radiation pressure felt by the reflector is:

$$p_{\text{rad, refl}} = 2I / c$$ \hspace{1cm} (45)

for total reflection of normally incident waves back along the line of incidence (Leighton, 1994, Section 1.1.4).

As the acoustic wave travels through the medium, it will be absorbed (Section 2.2). However the momentum absorbed from the acoustic field manifests as a flow of the liquid in the direction of the sound field, termed acoustic streaming (Lighthill, 1978, Leighton, 1994, Section 1.2.3; Trinh and Robey, 1994). Several potential beneficial uses of this phenomenon have been investigated (Nightingale et al., 1995; Rife et al. 2000; Shi et al., 2002), and it has also been assessed for its possible adverse effects (Starrritt et al., 1991; Barnett et al., 1998). Since it is more usual to consider energy than momentum in the context of acoustic waves, the process is conventionally thought of as the setting up of an energy gradient in the direction of propagation when energy is absorbed from the beam during its passage through an attenuating liquid. A gradient in energy corresponds to a force, and when this acts upon the liquid a streaming flow is generated. The force per unit volume ($F / V$) equals the gradient in pressure ($\Delta p_{\text{stream}}$) which causes the liquid to accelerate in the direction of propagation:

$$\Delta p_{\text{stream}} = F / V = 2I / c.$$ \hspace{1cm} (46)

From this equation it is clear that if both intensity and attenuation can vary spatially throughout a sound beam in a uniform medium, then so will the streaming forces and flows. An increase in either parameter will increase the streaming. As will be evident from the preceding section, the pumping of energy into higher frequencies, which are more strongly absorbed than the fundamental, means that attenuation may vary spatially in an acoustic field of finite amplitude. Finite amplitude effects will also affect streaming through the formation of shocks. Starritt et al. (1989) observed the enhancement of streaming in high amplitude diagnostic pulsed ultrasonic fields which have formed shocks in water.

Streaming speeds of up to around 10 cm/s can be demonstrated from clinical ultrasonic equipment. Fig. 9 shows a plan view of dye, carried along by the streaming flow, in three clinical underwater ultrasound beams. The visualisation technique is described by Merzkirch (1987), utilising the electrolysis of water containing dissolved thymol blue indicator. Through the insertion of acoustically-transparent ‘clingfilm’ windows in the diagnostic B-scan field, Starritt et al. (1991) demonstrated that, in addition to local energy absorption, there is a more significant contribution to unimpeded flow in the far-field region beyond the focus. To be specific, this comes from a narrow jet of flowing liquid which is generated near the focus and flows onwards from there with much the same speed. They therefore concluded that the acoustic stream is powered significantly by a near-focus ‘source pump’, which results from the enhanced absorption of the high-frequency components of the distorted finite amplitude pulses there. In materials, such as tissue, which are not free to flow, stresses may still be set up by these processes, and consideration must be given to the response of the medium (Starritt et al., 1991).

There is a second type of streaming associated not with the spatial attenuation of a wave in free space, but which instead occurs near small obstacles placed within a sound field, or near small sound sources, or vibrating membranes or wires (Leighton, 1994). It arises instead from the frictional forces between a boundary and a medium carrying vibrations of circular frequency $\omega$. Unlike the streaming described earlier, this time-independent circulation occurs only in a small region of the fluid, being generally confined to an acoustic boundary layer of thickness:

$$L_{\text{ms}} = \sqrt{2 \eta / \rho \omega},$$ \hspace{1cm} (47)

(footnote continued)

nuclear power station in Scotland. However since the speed of sound is lower than that of light, the equivalent power requirements would be only 17 kW in air and 75 kW in water.
where \( \eta \) and \( \rho \) are the shear viscosity and density respectively of the liquid (Nyborg, 1958). Because of the restricted scale of the circulation, it is commonly termed microstreaming (Fig. 10). Microstreaming can bring about a number of important effects. The shear forces set up within the liquid may disrupt DNA (Williams, 1974), disaggregate bacteria (Williams and Slade, 1971), disrupt human erythrocytes and platelets in vitro and in vivo (Williams et al., 1974; Williams, 1977), and other bioeffects (Rooney, 1972). Microstreaming can specifically occur as a result of the oscillations of an acoustically driven bubble in a sound field, which can also lead to bioeffects (Leighton, 1994, Section 4.4.3c(iii), Section 4.4.4, Section 5.4.2; Clarke and Hill, 1970).

The nonlinear propagation discussed earlier may be indirectly responsible for the relative scarcity of nonlinear acoustic phenomena, compared to optical ones. This is because, in many pure media, the sound
speed is non-dispersive at frequencies for which absorption is small over the distance of a wavelength. As a result, high-pressure fields are converted into shock waves as energy is pumped into the higher harmonics, as outlined earlier. Since absorption tends to be greater at higher frequencies, the high frequency oscillations are

Fig. 10. (a) A microtube (labelled 2) is inserted into a 10 kHz sound field of approximate 0-peak acoustic pressure amplitude 0.2 MPa in tap water, and air is injected into the tube, generating bubbles at the tip of the nozzle (labelled 3). One, labelled 1, displays the characteristic rippled surface shimmer of surface waves (including the Faraday wave), which can pinch off microbubbles from the main bubble (Leighton, 1994, Section 4.4.1(b)). The result of such 'pinching off' can be seen in (b). The arrows in (a) follow the flow of such small bubbles as they are transported within the microstreaming circulation; and their motion is evident through the streaks they produce in the picture under the 1 ms flash exposure. Arrows are aligned with the local streaks to show more clearly the local direction of the microstreaming flow. Given that streaks of up to 0.6 mm are made during the 1 ms flash, flow speeds of up to about 0.6 m s\(^{-1}\) are evident, and can reverse direction over a span of 4 mm (see upper right corner), giving rise to shear. (Picture: Leighton TG). (b) Picture showing the fragmentation of a gas bubble (of radius \(\approx 2.1\) mm) driven at 1.5006 kHz (139 Pa). The bubble is held against buoyant rise against the base or a vertical glass rod (labelled 1). Microbubbles (labelled 2 and 3, the latter being much smaller) are arrowed. They have been caused by fragmentation at the tips of the peaks of the surface waves (Leighton, 2004). (Picture: Birkin PR, Watson YE, Leighton TG).
strongly absorbed, so that strong sound waves are rapidly dissipated in the liquid. This effect can be reduced by engineering dispersion into the medium, for example by propagating the sound through a waveguide, or by introducing gas bubbles. Nevertheless there are several other nonlinear effects associated with the propagation of ultrasound, such as self-interaction and parametric phenomena, stimulated scattering and phase conjugation (Leighton, 1994, Section 1.2.3). Examples of some of these are given below.

A variety of self-interaction effects are shown in (Fig. 11). One is the self-focusing of acoustic beams. Thermal self-focusing is a consequence of the heating which occurs in a medium as a result of acoustic absorption. In most liquids such heating causes the sound-speed to fall (or equivalently causes the acoustic refractive index to increase) so that the beam is focused in towards the axis as a result of total internal reflection at its perimeter. In water, however, the sound speed passes through a maximum at 74°C (Fig. 12), so that thermal self-focusing only occurs at temperatures in excess of this. At lower temperatures, the inverse phenomenon of self-defocusing occurs. Acoustic streaming imparts a defocusing effect by increasing the acoustic propagation speed near the beam axis.

There are mechanisms other than thermal which can bring about self-focusing and self-defocusing. Because the presence of bubbles can strongly affect sound speed (Section 4.2), an inhomogeneity in the distribution of
bubbles can cause self-focusing and self-defocusing, and other spectacular refractive effects (as exploited by humpback whales—Leighton, 2004). In addition, since the acoustic impedance of bubbly water can be very different from that of bubble-free water, reflection can occur at the boundary between bubbly and bubble-free water. These phenomena become self-interaction effects when the sound field itself shapes the distribution of bubbles, which then modify the sound field (also shown in Fig. 11). Self-interaction effects involving bubbles are very common in liquids, because of the potency of the interaction between bubbles and sound (see the discussion of Fig. 19 in Section 4.2). For example, Bjerknes forces (Stephens and Bate, 1966; Leighton, 1994, Section 4.4.1) cause bubbles to accumulate at regions where the time-varying pressure oscillations are greatest or least, depending on bubble size. In Fig. 13, the clustering of bubbles at the acoustic pressure amplitudes of a standing wave field is indicated by the bands of sonoluminescence measured there. These bands can in turn modify the sound field through scattering. A similar effect is illustrated in Fig. 11, where the bubbles cluster at the focus, scattering and refracting the sound. The presence of bubbles can also bring about self-transparency, where the absorption decreases with increasing intensity, though this can arise in bubble-free media, such as glycerin, owing to the temperature-dependence of the absorption coefficient.

Self-focusing can occur not just in the bulk of a liquid, but at an interface between two media where, for example, distortion of the surface of a water sample by the beam can lead to a focusing effect simply as a result of local angling of the reflecting surface (Leighton, 1994, Section 1.2.3). In the extreme case, when the reflected beams are directed upwards, it can lead to *acoustic self-concentration* and the formation of a fountain.

Therefore both heating and bubbles can affect beam focusing, and of course can affect each other (Section 4). It would be important to consider these effects in circumstances where a tight beam focus is required, but where conditions promote both heating and bubble activity, and nonlinear propagation (Frizzell et al., 1983; Hynynen, 1987, 1991; Lee and Frizzell, 1988; Billard et al., 1990). One example of where such conditions occur, for example, is HIFU (High Intensity Focused Ultrasound, or Focused Ultrasound Surgery, FUS—Yang et al., 1993; Vaezy et al., 1998; Wu et al., 2002). Having grown from a long history of experiment and

---

**Fig. 12.** A plot of the sound speed as a function of temperature in pure water under 1 atmosphere of static pressure.
research (ter Haar, 1986; ter Haar, 2001; Leighton 1994, Section 4.4.2 (b)(i)), this technology (as other papers in this issue will describe) has grown to a therapy where the sound field of high $I_{TA}$ (e.g. continuous-wave) is focused into a region which is scanned across a tumour (Kennedy, 2005). The typical size of this focus is a few millimetres in each dimension. Tissue destruction is the result of heating (Barnett 1980a, b), and the generation of bubble activity can occur and can enhance heating (Holt and Roy 2001; Bailey et al., 2001; Thomas et al., 2005; Rabkin et al., 2005). Treatment times might be reduced if this effect could be controlled in the clinical environment.

20Experiments of the effects of high $I_{TA}$ focused ultrasound on tumours have been ongoing from the 1950s (Herrick 1953, Nightingale 1959, Fry and Johnson 1978, ter Haar et al. 1991; Hand et al., 1992). Research into the possible use in other clinical procedures has a similar long history. Early experiments (dating from the 1950s) were on neurosurgical research (Fry et al., 1954; Fry and Dunn 1956; Bausauri and Lele 1962; Warwick and Pond, 1968), then in ophthalmology work (Lizzi et al., 1978, 1984, 1992; Coleman et al., 1986; Polack et al., 1991) and studies of the nervous system (Fry et al., 1970; Lizzi and Ostromogilsky, 1987), Meniere's disease (Kossof et al., 1967; Barnett, 1980a, b) and the reduction of bleeding through ultrasonic haemostasis (Hwang et al., 1998; Vaery et al., 1999, 2001).
4. Acoustic cavitation

The previous section demonstrated the presence of nonlinearity during acoustic propagation, and ended with discussion of examples of how it can affect focusing. By far the most effective way of generating a nonlinearity, and of focusing acoustic energy both in terms of time and space, is through the interaction of acoustic waves with bubbles, a phenomenon known as acoustic cavitation.

Gas bubbles are the most potent naturally occurring entities that influence the acoustic environment in liquids. Upon entrainment under breaking waves, waterfalls, or rainfall over water, each bubble undergoes small amplitude decaying pulsations with a natural frequency that varies approximately inversely with the bubble radius, giving rise to the ‘plink’ of a dripping tap or the roar of a cataract. When they occur in their millions per cubic metre in the top few metres of the ocean, bubbles can dominate the underwater sound field. Similarly, when driven by an incident sound field, bubbles exhibit a strong pulsation resonance. Acoustic scatter by bubbles can confound sonar in the shallow waters which typify many modern maritime military operations. If they are driven by sound fields of sufficient amplitude, the bubble pulsations can become highly nonlinear. These nonlinearities might be exploited to enhance sonar, or to monitor the bubble population. Such oceanic monitoring is important, for example, because of the significant contribution made by bubbles to the greenhouse gas budget (Haugan and Drange, 1992; Stewart, 1992; Broecker and Siems, 1984; Farmer et al., 1993; Deane and Stokes, 1999; Leighton et al., 2004). In industry, bubble monitoring is required for sparging, electrochemical processes, the production of paints, pharmaceuticals and foodstuffs. At yet higher amplitudes of pulsation, gas compression within the collapsing bubble can generate temperatures of several thousand Kelvin whilst, in the liquid, shock waves and shear can produce erosion and bioeffects. Not only can these effects be exploited in industrial cleaning and manufacturing, and research into novel chemical processes, but moreover we need to understand (and if possible control) their occurrence when biomedical ultrasound is passed through the body. This is because the potential of such bubble-related physical and chemical processes to damage tissue will be desirable in some circumstances (e.g., ultrasonic kidney stone therapy), and undesirable in others (e.g., foetal scanning) (Leighton, 2004).

This list of example applications illustrates the distinctions which are made between the different forms of cavitation. The sound of a breaking wave, mentioned below, is generated through the pulsations of bubbles, decaying in amplitude following their entrainment. However when driven by a sound field, the pulsations, rather than decaying, can persist, the bubble pulsating in what is termed ‘stable cavitation’. Associated with such behaviour are the effects of bubbles on the sound speed, and the scattering of the sound field (Section 4.2). A range of other behaviours can affect the medium. These include Bjerknes forces (Fig. 13), which are radiation forces associated with bubble oscillation, and cause translational movements of bodies suspended in the liquid (Leighton, 1994, Section 4.4.1). For example, platelets might be caused to accumulate...
around a pulsating bubble (Miller et al., 1979); bubbles might group together, forming acoustic ‘shields’ which hinder the penetration of sound further into the liquid (Leighton, 1994, Section 5.3; Leighton, 1995; Fig. 14); and bubbles may translate through the medium generating hydrodynamic shear which can cause cell lysis (Miller and Williams, 1989). Microstreaming currents generated in the liquid close to the oscillating bubble wall can cause cell lysis (Vivino et al., 1985). These examples illustrate an important point: such is the potency of the interaction of bubbles with acoustic waves that the presence of bubbles can drastically affect the lengthscales, timescales, and magnitudes of the effects we expect from that wave. In the free field, diffraction limits the lengthscales over which we might expect significant stresses to be exerted on tissue, because it determines the lengthscales one must travel in order to see significant differences in particle velocity or displacement. However when bubbles are present, such changes can occur over much smaller distances (Fig. 10).

During rectified diffusion (Leighton, 1994, Section 4.4.3) the equilibrium radius of the bubble grows as the pulsations draw previously-dissolved gas out from the surrounding liquid into the bubble interior. This could conceivably cause a bioeffect through mechanical action or depletion of the gas reservoir, but probably is most important in that it affects the size distribution of bubbles in the population, which can affect the type of cavitation behaviour they undertake.

For many decades the term ‘stable cavitation’ was taken to imply, not only the fact that the bubbles pulsate over many cycles and remain physically intact, but also to imply that the physical effects generated by these pulsations are characteristic of ‘low energy’ sources. Therefore whilst nonlinear acoustic emissions from stable cavitation could be detected, shocks and luminescence could not. Similarly, whilst vigorous stable cavitation might cause rupture of delicate membranes (through microstreaming or radiation forces), erosion of steel does not occur. Whilst delicate long-chain polymers might be fragmented by microstreaming associated with stable cavitation, free radicals were not produced. Such ‘high energy’ effects were rather associated with a phenomenon known as ‘transient’ cavitation, so-named because the bubble pulsation which generated these effects was thought always to result in break-up of the bubble. Clearly it is unsatisfactory to have a definition which depends on two factors (stability of bubble and whether the effects caused by the cavitation were ‘high energy’ or not). This is because the scheme would break down when ‘high energy’ phenomena were observed from bubbles which pulsed stably over many cycles. Indications (specifically the observation of sonoluminescence from a stable bubble) that this might occur were first reported by Saksena and Nyborg (1970), but definitive proof was not forthcoming until later (Gaitan and Crum, 1990; Gaitan et al., 1992). Following this, a new scheme was adopted, whereby most of ‘stable’ cavitation (i.e., excluding the type which could produce sonoluminescence) was renamed ‘non-inertial cavitation’; all cavitation which generated the afore-mentioned ‘high-energy’ effects was termed ‘inertial cavitation’. Whilst this provided a useful working definition (primarily because researchers rarely observed the ‘cavitation’ itself, or could even agree on a definition, but rather worked with the observed effects of cavitation), a rigorous definition was found by re-examining the work of Flynn (1975a, b). The acceleration term in the equation of motion of the bubble is divided into two parts, a pressure function (PF) and an inertial function (IF), such that:

$$\dot{R} = IF + PF.$$

The inertial function is so-called since it is reminiscent of a kinetic energy gradient, in that it represents the effect of inertial forces in the liquid. For a free collapse, for example, it would be negative, representing the spherical convergence of the liquid. The PF represents the acceleration which is due to the summed pressure forces, as will now be shown.

Non-inertial cavitation is said to occur when PF dominates the dynamics of the collapse: inertial cavitation is said to occur when IF dominates. Approximate forms (e.g., ignoring radiation an thermal losses—see below) of the PF and IF are:

$$IF = - \frac{3R^2}{2R^2}$$

$$PF = \frac{1}{\rho_0 R} \left\{ \left( p_0 + \frac{2\sigma}{R_0} - p_v \right) \left( \frac{R_0}{R} \right)^{3k} + p_v - \frac{2\sigma}{R} - \frac{4\eta \dot{R}}{R} - p_0 - P(t) \right\}.$$  

(49)
which are found by comparing (48) with the well-known Rayleigh–Plesset equation for bubble dynamics:

\[
R \ddot{R} + \frac{3}{2} \dot{R}^2 = \frac{1}{\rho_0} \left( p_0 + \frac{2\sigma}{R_0} - p_v \right) \left( \frac{R_0}{R} \right)^{3\kappa} + p_v - \frac{2\sigma}{R} - \frac{4\eta \dot{R}}{R} - p_0 - P(t) \right) + O(\dot{R}/c),
\]

(50)

where \( p_0 \) is the static pressure in the liquid outside the bubble, \( \sigma \) is the surface tension (causing the \( 2\sigma/R \) Laplace pressure term—Leighton 1994, Section 2.1.1), and \( p_v \) is the vapour pressure. The mass density of the liquid (\( \rho_0 \)) is assumed to be incompressible. This is indicated by the neglected terms \( O(\dot{R}/c) \), which indicates that acoustic radiation losses are not included in this equation. Given that thermal losses are also not included (see the comments on \( \kappa \), below), the only dissipation mechanism in this equation is through the shear viscosity \( \eta \).

The term \( \kappa \) is the so-called polytropic index. This engineering term is not a fundamental quantity, but takes an intermediate value between \( \gamma \) (the ratio of the specific heat of the gas at constant pressure to that at constant volume) and unity, depending on whether the gas is behaving adiabatically, isothermally, or in some intermediate manner (such that the ideal gas relationship between the bubble volume (\( V \)) and its gas pressure (\( p_g \)) can vary between \( p_g V^\gamma = \) constant and \( p_g V^\gamma = \) constant). Note that the use of a polytropic law only adjusts the way gas pressure changes in response to volume changes to account for heat flow between the gas and its surroundings. In most bubble acoustics models where it is used, \( \kappa \) takes a constant value over the oscillatory cycle and, used in this way, can never describe net thermal damping during the oscillatory cycle of a bubble (Leighton et al., 2004). However the polytropic index does adjust the bubble stiffness for this heat flow.

The following sections will explain typical bubble dynamical behaviour, progressing from its oscillator characteristics (Section 4.1) through non-inertial cavitation (Section 4.2) to inertial cavitation (Section 4.3).

4.1. The bubble as an oscillator

The ideal spherical pulsating bubble acts as a damped oscillator: the stiffness comes from the bubble gas; and the inertia is invested primarily in the surrounding liquid, which is set into motion when the bubble wall moves. Viscous, thermal and acoustic radiation losses contribute to the damping.

For spherical bubbles, at least four possible classes of equation of motion can be constructed, depending on the ‘frame’ (Leighton, 1994, Section 3.2.1(a)) chosen in which to represent the system (i.e. whether the position of the bubble wall is expressed in terms of bubble volume \( V(t) \) or radius \( R(t) \), and whether the external driving field is expressed in terms of the acoustic pressure \( P(t) \) or force \( F(t) \)). In the radius/force frame, if the displacement of the bubble wall is \( R_\alpha \), such that

\[
R(t) = R_0 + R_\alpha(t)
\]

(51)

then the equation of motion is:

\[
m_{RF}^\text{rad} \ddot{R}_\alpha + b_{RF}^\text{tot} \dot{R}_\alpha + k_{RF} R_\alpha = -F(t) = -4\pi R_0^2 P(t),
\]

(52)

where the force \( F(t) \) is assumed to come from an applied pressure field \( P(t) \) which gives uniform pressure over the bubble wall at any instant (e.g. as occurs if all the wavenumbers \( k \) associated with \( P(t) \) are such that \( k R_0 \ll 1 \)). The negative sign ensures that a quasi-static reduction in pressure produces an increase in the radial coordinate \( r \) of the wall. The stiffness in this frame is \( k_{RF} \approx 12\pi \kappa p_0 R_0 \) (Leighton, 1994, Section 3.2.1(b)) and the inertia (the so-called “radiation mass”) is \( m_{RF} = 4\pi R_0^3 \rho_0 \) (Leighton, 1994, Section 3.2.1(c)(iii)) The term \( b_{RF}^\text{tot} \) summarises all the dissipation mechanisms (Leighton, 1994, Section 3.4).

As linear oscillators at low amplitudes of pulsation, gas bubbles in liquids are abundant, and responsible for many of the sounds we associate with liquids in the natural world. When driven by external sound fields, a bubble exhibits a powerful pulsation resonance, plus numerous resonances associated with higher order spherical harmonic shape perturbations. At finite amplitudes, bubbles will not only ‘process’ the driving sound field by generating harmonics, subharmonics and combination frequencies; they will also process the surrounding medium, producing physical, chemical, and biological changes. Furthermore, all these phenomena are not simply interesting in their own right: they can be exploited as tools. As a result,
problems in bubble acoustics can sometimes lead to complete solutions, starting with the fundamental physics and ending with a product or device in the clinic, laboratory, or market.

To examine the stiffness element in detail, note that the gas, if slowly compressed, exerts a force which resists that compression, and would tend to make the bubble expand (and vice versa). Potential energy is stored in the gas as the bubble volume changes. When the bubble wall moves, the surrounding liquid must also move: If the system has spherical symmetry, then the velocity of an incompressible liquid falls off as an inverse square law away from the bubble wall (Minnaert 1933). Therefore there is a kinetic energy associated with bubble pulsations which is characterised by the moving matter. Since the liquid is so much denser than the gas, this is primarily invested in the liquid (though motion of the gas contributes to a much smaller extent Leighton et al., 1995). Comparison of the potential and kinetic energies (which is in effect a consideration of the relative effects of the gas stiffness and the liquid inertia) allows the formulation of the natural frequency \( \gamma_0 \) of the oscillator. A simple calculation, based on the linear pulsations of a spherical air bubble of equilibrium radius \( R_0 \) in water, which is assumed to be incompressible and inviscid, gives

\[
\gamma_0 R_0 \approx 3.3 \text{ Hz m}, \quad R_0 > \sim 10 \mu\text{m}
\]

assuming one atmosphere of static pressure. Eq. (53) neglects surface tension, making this formulation less valid for smaller bubbles. Throughout this paper, the long wavelength limit \( kR_0 \ll 1 \) is assumed.

Eq. (53) is found by substituting sea surface parameters (for water and air) into the resonance frequency, found by a small-amplitude expansion of Eq. (50):

\[
\gamma_0 = \frac{\omega_0}{2\pi} = \frac{1}{2\pi R_0 \sqrt{\rho_0}} \sqrt{3\kappa \left( p_0 - p_v + \frac{2\sigma}{R_0} \right) - \frac{2\sigma}{R_0} + p_v - \frac{4\eta^2}{\rho_0 R_0^2} \approx \frac{1}{2\pi R_0} \sqrt{\frac{3k p_0}{\rho_0}},}
\]

where \( \omega_0 \) is the natural circular frequency of the bubble, and where the final approximation neglects the effects of vapour pressure \( p_v \), surface tension \( \sigma \) and shear viscosity \( \eta \).

The bubble is of course displaying the features expected of a single-degree-of-freedom linear resonance (Leighton, 1994, Section 4.1). For example, in steady-state, the amplitude of pulsation tends to be largest close to the resonance condition. Bubbles just larger than the size which is resonant with the sound field will pulsate in antiphase to those just smaller than resonance size. This leads to somewhat unusual features, such that for most separation distances, bubbles which are either both larger than, or both smaller than, resonance size, tend to attract; whilst repulsion occurs if one bubble is greater than resonance size and the other is smaller (Leighton, 1994, Section 4.4.1). However in order to calculate the frequency-dependence of any of these resonant effects, the bubble damping must be known.

The characteristic values in Eq. (53) are responsible for the ubiquitous nature of the bubble acoustics when sound is passed through liquids. Take for example a breaking ocean wave: this can generate bubbles having radii typically ranging from millimetres to microns (Leighton et al., 2004). From Eq. (53), these provide pulsation natural frequencies in the frequency range of at least \( \sim 1–500 \text{ kHz} \), respectively (with commensurate quality factors of roughly 30–5). Similarly, a biomedical sound field of a few MHz might resonate with a micron-sized bubble.

Eq. (54) does not incorporate thermal or acoustic radiation losses, which require more sophisticated models (Leighton, 1994, Section 4.2), bubble–bubble interactions (Foldy, 1945; Kargl, 2002) reverberation (Leighton et al., 2002), or the fact that the bubble may be constrained by surrounding structures, such as pipes, tissue, etc. (Miller, 1985; Quain et al., 1991; Leighton et al. 1995; Oguz and Prosperetti, 1998; Geng et al., 1999; Miller and Quddus, 2000; Leighton et al., 2000; Symons, 2004; Yang and Church, 2005.). Indeed, whilst the integral which expresses the inertia associated with bubble expansion converges in three-dimensions (e.g. for the pulsation of a bubble in free space), if the bubble is constrained to expand along the length of a uniform pipe without fracturing it, the ‘radiation mass’ is proportional to the length of the pipe and so can become very great indeed (Leighton et al., 1995). Therefore, for example, conceptual models of the ultrasonically-induced expansion of bubbles contained within blood vessels, to sizes sufficiently large that the liquid flow becomes more 1D than 3D, could well be precluded by the associated liquid inertia, causing wall rupture (although of 21Outside of this quasi-static limit, the phase relation between the driving force and the bubble volume of course changes in the expected manner in the pseudolinear limit (Leighton, 1994, Section 4.1).
course cavitation need not always be the main cause when such effects are observed (O’Brien and Frizzell, 2000).

Section 4.2 examines the effect which bubbles have on acoustic propagation, and Section 4.3 discusses inertial cavitation. Of all the bubble-based phenomenon, this is the one which has historically attracted most interest with respect to ultrasonic bioeffect (Carstensen 1987). This fact, and its complexity, warrants a devoted section. However in the last decade interest has increased in other bubble-based phenomena, such as the use of ultrasonic echocontrast agents (Cosgrove, 1996, 1997). Many of these are engineered forms of microbubbles. From initially being an agent designed to enhance the echogenicity of blood during ultrasonic imaging (Yeh et al., 2004), the technology is now being investigated for use in ultrasonically medicated targeted drug delivery (Shortencarier et al. 2004; Pitt et al., 2004; Zhao et al., 2004; Postema et al., 2004; Dayton et al., 2004).

4.2. The effect of bubbles on acoustic propagation

Fig. 15(a) shows a bubble size distribution, measured in the ocean, along with the associated sound speed and the component of attenuation for which bubbles are responsible (Fig. 15(c)). Although a wide range of bubble sizes are present (from at least microns to millimetres) in the ocean, the population as a whole tends to impart to the ocean characteristics such that, for frequencies below about 20 kHz, the bubbles reduce the sound speed to less than that of bubble-free water (~1480 m s\(^{-1}\)), whilst for frequencies above about 40 kHz, the bubbles may tend to increase the sound speed (Fig. 15(b)). The magnitude of the change to sound speed increases the closer the insonifying frequency is to the critical 30–50 kHz range. The additional attenuation caused by bubbles (over and above that which occurs in bubble-free water) also peaks in this range (Fig. 15(c)).

These features are caused by the oscillatory behaviour described in Section 4.1. Bubbles may either increase or decrease the sound speed compared to that of bubble-free liquid, because of the characteristics imparted by the distinction between a stiffness-controlled and an inertia-controlled regimes. In short, the bubble pulsation...
undergoes a $\pi$ phase change (compared to the phase of the driving pressure field) in passing from one regime to the other, and this imparts a sign change to the perturbation of the sound speed.

Consider a volume $V_c$ of bubbly water, which is made up of a volume $V_w$ of bubble-free water and a volume $V_g$ of gas (distributed in an unspecified way between an unspecified number of bubbles of arbitrary size). Conservation of volume gives:

$$V_c = V_w + V_g,$$

(55)

where the subscripts will be taken to refer to the gas (g), bubble-free water (w) and the cloud (c). Mass conservation is simply expressed by multiplication of the volumes with the respective densities (of cloud, $\rho_c$; bubble free water, $\rho_w$ and gas, $\rho_g$), i.e.

$$\rho_c V_c = \rho_w V_w + \rho_g V_g.$$  

Assume that the bulk moduli and sound speeds of the components can be defined through Eqs. (13) and (14), such that for example

$$c_c^2 = \frac{B_c}{\rho_c} = \left(\frac{\partial p(\rho, S)}{\partial \rho}\right)_c, \quad \epsilon = w, g,$$

(57)

where the subscript $\epsilon$ can refer to application to gas (g) or bubble-free water (w). Differentiation of Eq. (55) with respect to the applied pressure gives, with (57), the relationship between the bulk moduli

$$\frac{1}{B_c} = V_w \frac{1}{V_c B_w} + V_g \frac{1}{V_c B_g},$$

(58)

and hence, noting Eq. (57), we define a function (which is not an inherent property of the bubble cloud in the thermodynamic sense), equal to the root of the ratio of the bulk modulus of the cloud to its density (Leighton et al., 2004):

$$\xi_c = \sqrt{\frac{B_c}{\rho_c}} = \sqrt{\left(\frac{\rho_w V_w + \rho_g V_g}{V_c B_w + \frac{V_g}{V_c B_g}}\right)} \approx c_w \left(1 + \frac{B_w V_g}{V_c B_g}\right)^{-\frac{1}{2}} \approx c_w \left(1 - \frac{B_w V_g}{2 V_c B_g}\right),$$

(59)

where use is made of the small-perturbation approximations: $1/\rho_c \approx 1/\rho_w$ (Eq. (25)) and $\rho_c V_c = \rho_w V_w + \rho_g V_g \approx \rho_w V_w$ and $V_c \approx V_w$. Finally, the substitution $c_w = \sqrt{B_w/\rho_w}$ is made from Eq. (57) and a binomial expansion performed under the assumption that $B_w V_q(t) \ll B_c V_q(t)$. This is valid if the void fraction $V_g/V_c$ is very low If the cloud were not dissipative, this function would equal the sound speed in the cloud. However since dissipation does occur there, such an identity would not be rigorous. How therefore sound speeds can be estimated from this function is discussed below.

To simplify the expression further, assume that all the bubbles are of the same size (i.e. the population is monodisperse). Note that it is not a difficult extension to include a polydispers population (Leighton et al., 2004), although that is not necessary for the illustrative purposes of this section. Assume that the bubble population is monodisperse, containing $N_b$ bubbles each of volume $V_b$ and equilibrium radius $R_0$. The instantaneous bubble radius is $R$, and the bubble number density is $n_b = N_b/V_c$. Applying Eq. (57) to the gas and water, and noting that $V_b \propto R^3$, implies that $B_w = \rho_w c_w^2$, that $V_g = n_b V_c V_b$, and that under these conditions $B_g = \partial P/\partial (\mu_\rho_g/\rho_g) = -\partial P/(3 \partial R/R)$. Substitution of these into Eq. (59) gives:

$$\xi_c \approx c_w \left(1 + \frac{3 \rho_w c_w^2 N_b V_b}{2 R V_c} \frac{dR}{dP}\right) \approx c_w \left(1 + \frac{3 \rho_w c_w^2 n_b V_b}{2 R} \frac{dR}{dP}\right),$$

(60)

where the use of the differential symbol $d$ here implies an intention to calculate the result numerically (Leighton et al., 2004). The importance of the $dR/dP$ (or, equivalently, the $dV/dP$) term of Eq. (60) in determining the sound speed can be explained using plots of the applied pressure $P$ against bubble volume $V$ (Leighton et al., 2004).
First consider a monodisperse bubble population (i.e. all bubbles have the same equilibrium radius) pulsating in the linear steady state when driven by an acoustic field of circular frequency $\omega$. If the propagation were linear and lossless, the graphs of applied pressure ($P$) against bubble volume ($V$) would take the form of straight lines, the location of the bubble wall being plotted by the translation of the point of interest up and down these lines at the driving frequency (Fig. 16, top row). Since a positive applied pressure compresses a bubble in the stiffness-controlled regime, here $dP/dV > 0$ (Fig. 16, top row, right). However since a phase change of $\pi$ radians occurs across the resonance, the opposite is true in the inertia-controlled regime ($dP/dV < 0$, Fig. 16, top row, left) (Leighton, 2004; Leighton and Dumbrell, 2004).

The behaviour of the top row of Fig. 16 reflects the trend indicated at the start of Section 4.2, and because of this the sound speed in bubbly water ($c_c$) is increased in the inertia-controlled regime, and decreased under stiffness-control. Obviously, a bubble which is driven in the inertia-controlled regime will be expanding during the compressive half-cycle of the driving pulse, which contributes a component increase in volume to the bubbly water during compression. The sign of $dP/dV$ implies, through (60), that inertia-controlled bubbles will increase the sound speed in bubbly water ($c_c$) above that found in bubble-free water ($c_w$). In the stiffness-controlled mode, the bubble will compress to a greater extent than the volume of water it replaces during this compressive half-cycle. Hence through (60), the usual phase change which occurs across resonance means that, in the stiffness-controlled regime, $dP/dV < 0$. This in turn changes the sign of the contribution made by the bubbles to the sound speed in the mixture, and $c_c < c_w$. If the bubble population contains a range of equilibrium radii, an appropriate summation is required (Leighton et al., 2004).

If conditions are linear and lossy (Fig. 16, second row), each acoustic cycle in the steady-state must map out a finite area which is equal to the energy loss per cycle from the First Law of Thermodynamics. The sound speed can be estimated using the $dP/dV$ gradient of the spine of the loop (shown in Fig. 16 as a dashed line). Assume the gas is perfect. Its internal energy $U$ is a state function, such that whenever an orbit crosses its previous path, at both moments represented by the intersection the value of $U$ is the same. More specifically, consider that:

$$dU = d\tilde{Q} + d\tilde{W} = \tilde{d}Q - PdV,$$

where the notation indicates that both the incremental heat supplied to the bubble ($\tilde{d}Q$) and the work done on the bubble ($d\tilde{W}$) are not exact differentials, while $dU$ is.

Because Fig. 16 (and later, Fig. 17) use the applied acoustic pressure $P(t)$, the area mapped out by any loop represents the energy subtracted from the acoustic wave by the bubble in the time interval corresponding to the perimeter of the loop. This is because the bubble dynamics (such as used here) may be interpreted simply as...
a statement of the equality between that pressure difference \( \Delta p \) which is uniform across the entire bubble wall, and a summation of other. These terms relate to the pressure within the gas/vapour mixture inside the bubble \( p_i \), surface tension pressures \( p_\sigma \), and the dynamic terms resulting from the motion of the liquid required when the bubble wall is displaced, which will here be termed \( p_{\text{dyn}} \). Thus

\[
\Delta p = p_i - p_{\text{dyn}} - p_\sigma. \tag{62}
\]

The energy \( E_{\text{loop}} \) subtracted from the sound field by the pulsating bubble in each circuit of a loop is given by

\[
E_{\text{loop}} = - \oint p_i \, dV + \oint p_{\text{dyn}} \, dV + \oint p_\sigma \, dV, \tag{63}
\]
noting that the details of the chemistry on the bubble wall may make the final integral non-zero. However \( \Delta p \) equals the spatial average over the bubble wall of the blocked pressure \( \langle P_{\text{blocked}} \rangle \), which in the long-wavelength limit equals the applied acoustic pressure \( P(t) \) that would be present at the bubble centre were the bubble not present. Substituting Eq. (62) into (63) therefore shows that the area mapped in a loop in the applied pressure-volume plane is the energy subtracted from the acoustic wave in the time interval corresponding to that loop:

\[
E_{\text{loop}} = - \oint \Delta p \, dV = - \oint \langle P_{\text{blocked}} \rangle \, dV \approx - \oint P \, dV. \tag{64}
\]

Therefore, the rate at which the acoustic field does work on the bubble can be found by integrating the area in the pressure–volume plane enclosed by the loops formed by the intersections described above, and dividing energy so obtained by the time interval taken to map out that loop. In this way the rate at which the bubble subtracts energy from the driving acoustic field can be calculated.

Traditionally in bubble acoustics, researchers have found greatest imprecision and difficulty in defining a sound speed near resonance. The second row of Fig. 17 illustrates how this will coincide with conditions where not only is the area mapped out very large, but the characteristic gradient of \( dP/dV \) is very difficult to identify (in keeping with known through-resonance behaviour of sound speed of the type shown in Fig. 15(b)).

If conditions are nonlinear and lossless (Fig. 16, third row), in steady-state the \( P-V \) plots must encompass zero area, but they will depart from straight-lines (for example because the degree of compression cannot scale indefinitely). The gradient \( dP/dV \) varies throughout the acoustic cycle in a manner familiar from nonlinear acoustic propagation, and this can appropriately describe nonlinear propagation and the associated waveform distortion in the usual manner (Morse and Ingard, 1986). As before, if the bubble population contains a range of equilibrium radii, an appropriate summation is required (Leighton et al., 2004). Furthermore, since such a summation would provide an approximation for the relationship between pressure and density for the sample of bubbly water, treating it as an effective medium would provide the information required, (through Sections 3.2 and 3.3) to predict the effectiveness of bubbles in enhancing parametric sonar, an observed effect which has previously been modelled using analysis which assume finite-amplitude expansions of the bubble pulsation and simplified models for damping (Kozyaev and Naugol'nykh, 1980; Kustov et al., 1982; Lerner and Sutin, 1983; Kotel’nikov and Stupakov, 1983).

If conditions are nonlinear and lossy (Fig. 16, bottom row), finite areas are mapped out, and whilst the characteristic spines may present significant challenges, nonlinear propagation may again be identified (the example of the right of the bottom row in Fig. 16 illustrates a strong third harmonic, where the steady-state volume pulsation undertakes three cycles for each period of the driving field).

This scheme can now be used to interpret Fig. 17, which uses a nonlinear model (see Leighton et al., 2004) to predict the response of a single air bubble (of equilibrium radius 49 \( \mu \)m) in water, subjected to a semi-infinite sinusoidal driving pulse (starting at \( t = 0 \)). Under linear sea surface conditions this bubble has a resonance of 65.7 kHz. The left column, (a), corresponds to insonification at 84.2 kHz, a frequency greater than resonance, i.e. the inertia-controlled regime. The middle column, (b), corresponds roughly to a bubble at resonance (65.7 kHz). The column on the right, (c), shows insonification at 31.5 kHz, a frequency less than resonance (i.e. the stiffness-controlled regime).

The top row shows the volume time history of the bubble, as predicted by a nonlinear model (see Leighton et al., 2004). The middle row plots the same data in the plane of the applied pressure versus bubble volume. The locus of this plot consists of a single point until the onset of insonification. From this moment on, the locus describes orbits until reaching steady-state, after which it repeatedly maps out a given orbit. The time-dependent rate at which each bubble in the population subtracts energy from the driving acoustic field can be calculated in the manner described for Fig. 16, with steady state being achieved as \( t \to \infty \) (Fig. 17, middle row).

Of particular interest is the bottom row of Fig. 17, which superimposes the steady-state nonlinear loops of the middle row (thin line) with the corresponding linear steady-state solution; thick line. At frequencies much greater than or less than resonance (not shown), both models predict loci indistinguishable from straight lines (dissipation and nonlinearities being negligible at such extremes, the area mapped out by each loop is very small). The gradients of these lines have opposite signs, in keeping with the phase change of \( \pi \) radians which takes place between the stiffness- and inertia-controlled regimes.
Closer to resonance, increasing dissipation imparts finite areas to the loops, and the sound speed must be inferred from the spine of the loop. While in some cases the nonlinear model would impart a similar spine to its loop as would that of linear theory (Fig. 17(a), bottom row), closer to resonance identification of the optimum spine becomes more difficult (Fig. 17(b), bottom row; be aware that the conditions for resonance in the nonlinear and linear models are slightly different). The increasing dissipation and indistinct nature of the spine near resonance may lead to inaccuracies, as discussed above. The different losses predicted by linear and nonlinear theories in the steady state are readily determined by comparing their respective loop areas in the bottom row of Fig. 17. Of particular interest is Fig. 17(c), where the nonlinear model displays a second harmonic (which is of course not apparent in the linear result). In calculating the losses, the area of the clockwise loops must be subtracted from the anticlockwise loops (Leighton et al., 2004). Clearly when bubbles undergo nonlinear pulsations, the propagation conditions may be very different from the predictions of linear theory, and indeed this may be exploited by dolphins and porpoises (Leighton, 2004; Leighton et al., 2004; Leighton et al., 2005b).

This method of visualising the losses from the sound field through the areas of the P–V loops provides a method, not only of calculating the losses during steady state, but also during the ring-up period (an alternative method must be used during ring-down - Leighton et al., 2004; Leighton and Dumbrell, 2004). The exact loss mechanisms included, and the accuracy with which they predict losses, depends on the quality of the terms in the equation of motion for the bubble which encapsulate these processes (the plots in Fig. 17 include thermal, viscous and acoustic radiation losses which encapsulate the nonlinearity; the more commonly used Rayleigh–Plesset equation only includes viscous losses, and whilst there have been attempts to add radiation and thermal losses by augmenting the viscous terms, these can only be partially successful). If linear bubble oscillations only are assumed to occur, and only steady-state oscillations are to be considered, then analytical expressions for the power lost to acoustic radiation, viscous and thermal effects can be readily derived (Leighton, 1994, Section 4.1.2(d)).

The ability of bubbles to affect the sound speed is evident in Fig. 19, where the acoustic modes of a vessel are made manifest by the ability of ultrasonically induced chemiluminescence to reveal the position of acoustic pressure antinodes. Since for a vessel of fixed geometry and boundary conditions, the frequency of the modes depends on the sound speed (Birkin et al., 2003a). Therefore the frequency at which the various modes occurs can be used to give the sound speed (e.g. between 868 and 1063 m s$^{-1}$—compared to $\sim$1500 m s$^{-1}$ for bubble-free water-in the study of Birkin et al., 2003a).

However, whilst we might exploit such patterns as Fig. 19 to estimate sound speeds, the existence of such patterns and ‘characteristic sound speeds’ is in some ways very strange. This is because the bubble population in Fig. 19 was generated by a ‘power ultrasound’ transducer. As such, it therefore contained not just non-inertial cavitation, but also inertial cavitation. The number, dynamics, and locations of such bubbles are in large parts governed by the random conditions of the nucleation (Church, 2002), and so the actual bubble population will be rapidly changing on the scale of an acoustic cycle. Furthermore, the distribution of the cavitation is clearly highly inhomogeneous on the scale of a wavelength. How therefore this inhomogeneous, rapidly changing and stochastic population manages to generate a highly stable characteristic sound speed for the medium is not entirely clear. It is surprising that there appears to be present a bubble population, the statistics of which have sufficient longevity to give a sound speed that could lock the insonification into a specific mode of the vessel with a mode. This is another example of a self-interaction effect, as introduced in Section 3.5. The phenomenon of inertial cavitation is the topic of the next section.

4.3. Inertial cavitation

When induced by a changing pressure field, inertial cavitation requires that the bubble undergo significant expansion, prior to a more rapid collapse, and then a rebound (which emits a pressure pulse). This can generate a range of effects, both chemical and biological, but perhaps the earliest observables attributed to inertial cavitation were erosion (see Fig. 18), as noted in the ‘pitting’ in ship propellers, and the associated generation of cavitation noise (primarily the rebound pressure pulses). Here the changing pressure field which causes cavitation is generated hydrodynamically, not acoustically, but the principles are the same. An interesting observation in the behaviour of submariners illustrates well the importance to inertial cavitation of
Fig. 18. A back-lighted sample of aluminised mylar sheet, which has been placed for about one minute in a cavitating field (acoustic pressure amplitude $\sim 0.2$ MPa, 10 kHz). The dark regions show where the sheet is intact. Large light circles indicate where the aluminium has been removed through cavitation erosion (Leighton, 1994).
both the growth and collapse phases. Submariners wish to suppress any cavitation noise generated by their propellers since it can give away their location. They know that submerging their vessel will tend to reduce cavitation noise by suppressing the propeller cavitation. However when the cavitation is strong and the vessel is at high speed, increasing the depth of the vessel will first cause an increase in the cavitation noise, before suppression occurs. This so-called ‘anomalous depth effect’ is due to the fact that, an increase in static pressure increases the violence of each individual bubble collapse, before (at greater depths) it suppresses the growth phase. ‘Pitting’ can also be caused by an aspherical form of collapse which can occur when the bubble is close to an inhomogeneity such as a solid boundary: The bubble can invert on collapse such that a high-speed liquid jet can pass through the bubble and, on impacting a nearby boundary, can create damage (Leighton, 1994, Section 5.4.1; Leighton, 2004). Inertial cavitation produces shear in the liquid, and generates free radicals through the compression of gas during the collapse stage.

Inertial cavitation is a threshold phenomenon. The threshold is defined in terms of the amplitude of the driving sound field (usually the peak negative pressure), its frequency, and the size of the bubble which is pre-existing and available to nucleate the inertial cavitation event. The prevalence of such ‘cavitation nuclei’ is evident from the fact that, even with the most extensive measures in place to remove such nuclei from the

Fig. 19. The acoustic pressure antinodes within reverberant water-filled cylinders (with vertical axis of symmetry, and the sound source at the cylinder base) are made visible through the chemiluminescence which occurs there. (a) Plan and (b) side views of luminescence (which occurs at pressure antinodes) in a water-filled cell which had a polymethylmethacrylate wall (9.4 cm internal diameter, 10 cm external diameter; height of aqueous solution = 14 cm) for insonification at 132.44 kHz where the spatial peak temporal fluctuation in pressure in the liquid was 75 kPa (all quoted zero-to-peak). Frames (c)–(f) (to which the scale bar of length 5.8 cm in frame (c) refers) were taken in a double-walled, water-jacketed cell (5.8 cm internal diameter, 8.5 cm external diameter, and liquid height 8 cm) which was maintained at a constant liquid temperature of 25 °C. For a constant applied drive voltage, as the insonifying frequency changed, so too did the spatial peak acoustic pressure, providing the following combinations: (c) 121 kHz; 139 kPa; (d) 122 kHz; 150 kPa; (e) 123 kHz; 180 kPa; (f) 124 kHz; 200 kPa. The effect of tuning into particular acoustic modes is evident: a 1 kHz change in frequency can dramatically alter the amount and distribution of the luminescence. Hence the not uncommon practice of incrementing frequencies by O(100 kHz) when testing for the ‘optimal processing frequency’ in such arrangements is nonsensical. Similarly, if calorimetry were used to estimate the ultrasonic field, the change of sound speed resulting from the rise in liquid temperature could detune the mode. By noting the modal resonance frequencies in these and similar cylinders, the sound speed in this bubbly water was found to be in the range 868–1063 m/s, implying void fractions of 2.9–4.2 × 10⁻³%. Frames selected from several figures in Birkin et al. (2003a). (Figure courtesy P.R. Birkin, J.F. Power, T.G. Leighton and A.M.L. Vincotte).
liquid, tensile strength tests on liquids hardly ever measure the actual tensile strength of the pure liquid. Rather, the liquid ‘fails’ through the growth of these pre-existing microscopic bubbles, which are stabilised against dissolution by hydrophobic contaminants in the liquid, or in cracks in the container walls or suspended solids; or even freshly-created by cosmic rays (Leighton 1994, Section 2.1.2). Because of this dependence of the threshold on the size of the nuclei available (as well, to a lesser extent, on the amplitude of the sound field, which may be inhomogeneous (see Fig. 19)), unless special measures are taken (Gaitan and Crum, 1990), then for most sound fields in which inertial cavitation is occurring, some non-inertial cavitation is also occurring (see Figs. 20 and 21).

When bubbles within a certain size range (that tends to be smaller than resonance) are driven strongly, they undergo a dramatic change in behaviour (Flynn, 1975a, b; Apfel, 1981; Apfel and Holland, 1991; Church, 1993; Leighton, 1994, Sections 4.2 and 4.3; Church, 2002). This transition from non-inertial to inertial cavitation (where the bubble undergoes rapid explosive growth, followed by a violent collapse, generating high gas temperatures, gas shocks, and liquid shocks) is characterised by the onset of effects such as erosion, sonoluminescence, chemiluminescence etc. Both the high gas temperatures and the gas shocks may generate free radicals (e.g. in aqueous solution, H atoms, OH radicals, and products such as H2O2, hydrogen peroxide). These are highly reactive and may represent a hazard (they are sources of sonochemical reaction). As regards the hot-spot itself, although its temperature is high, it does not contain significant energy or last for a long time. The rebound pressure pulse emitted into the liquid as the bubble rebounds from minimum size may cause mechanical damage to structures close to it. However in a cloud of bubbles the shocks may act to greater distances through a co-operation effect between many bubbles (cloud concentration).

The threshold is such that there is a critical size range, which increases with increasing acoustic pressure amplitude and decreasing frequency, over which bubbles will undergo inertial cavitation if insonified by an appropriate sound field. This is illustrated in Fig. 20, which contains three curves. Each curve represents the threshold condition if a bubble of radius given on the abscissa is insonified by a sinusoidal wave of peak negative pressure as shown on the mantissa. It is usually (but not always) the case that the sound frequency and amplitude are known or, if necessary (e.g. in vivo) can be estimated, but the radii of bubbles present is not. This is particularly so in the in vivo scenario. Suppose therefore that a continuum of bubble sizes is present. For a given frequency, the parameter space above the curve indicates that inertial cavitation will occur, within the limits of the model and for the specific definition of inertial cavitation chosen (Neppiras, 1980; Flynn and Church, 1988; Leighton 1994, Section 4.3; Holland and Apfel, 1989; Apfel and Holland, 1991; Allen et al.,

![Fig. 20. Illustration of the threshold for inertial cavitation, after the calculation of Apfel and Holland (1991), in terms of the variation in the peak negative pressure required to generate inertial cavitation from a free-floating spherical gas bubble nucleus, as a function of the initial radius of that bubble. Above the curve, inertial cavitation will occur. As the acoustic frequency increases, the threshold tends to increase for all initial bubble sizes, and the radius range which will nucleate inertial cavitation decreases. Redrawn by permission from Ultrasound in Medicine and Biology, vol., 17, pp. 179–185; Copyright © 1991 Pergamon Press Ltd.](image-url)
Therefore, for a given frequency and sound pressure amplitude, bubbles within a certain radius range will undergo inertial cavitation, and those outside of it will not. This range is illustrated in Fig. 20 for 10 MHz insonification with a peak negative pressure of 1.5 MPa.

Why should bubbles outside this radius range not undergo inertial cavitation? The simple answer is that inertial cavitation comprises both explosive bubble growth, followed by a sufficiently rapid collapse. If the bubble is too small, then surface tension forces prevent the initial sudden growth, and inertial cavitation does not occur (Leighton, 1998). This is because the Laplace pressure (the $2\sigma/R$ term in Eq. (50)) varies inversely with bubble size, and therefore increases rapidly with decreasing radius (Leighton, 1994, Section 2.1.1).

Conversely, if the equilibrium size $R_0$ of the bubble nucleus is initially too large, then it may grow, but insufficiently to then concentrate the energy sufficiently on collapse to generate free radicals etc. There are several ways of understanding this. For example, the timescales on which such large bubbles respond to pressure (i.e. grow during a rarefaction) are relatively slow compared to smaller bubbles (as evidenced by the approximately inverse relationship between bubble radius and natural frequency in Eq. (54). Approximate analytical expressions for this time were given by Holland and Apfel (1989), who considered the delay times in bubble response to be the summation of three components, corresponding to contributions caused by surface tension ($\Delta t_\sigma$), inertia ($\Delta t_I$) and viscosity ($\Delta t_\eta$), their sum being:

$$\Delta t_\sigma + \Delta t_I + \Delta t_\eta \approx \frac{2\sigma}{P_A - P_B} \sqrt{\frac{3\rho_0}{2(P_A - P_B)}} + \frac{2R_0}{3} \sqrt{\frac{\rho_0}{\Delta P_{\text{wall}}}} + \frac{4\eta}{\Delta P_{\text{wall}}}, \quad (65)$$

where $P_A$ is the acoustic pressure amplitude of the insonifying field (assumed to be sinusoidal), and where $P_B$ is the Blake threshold pressure, the difference between the hydrostatic pressure which exists in the liquid at equilibrium, and the critical tension in the liquid which must be generated in order to produce explosive growth in bubbles which are initially very small (Leighton, 1994, Section 2.1.3.(b))

$$P_B \approx p_0 + \frac{8\sigma}{9} \sqrt{\frac{3\sigma}{2R_0(p_0 + 2\sigma/R_0)}} \quad (66)$$

and where $\Delta P_{\text{wall}}$ is the time-averaged pressure difference across the bubble wall:

$$\Delta P_{\text{wall}} \approx (P_A + P_B - 2p_0 + \sqrt{(P_A - p_0)(P_A - P_B)})/3. \quad (67)$$

Since we are considering the large-bubble limit of the range of bubble radii which can nucleate inertial cavitation, the issue is not with $P_B$, and hence the dependence in (65) of the time for growth on initial bubble radius is primarily through the inertial term $\Delta t_I \approx (2R_0/3)\sqrt{\rho/\Delta P_{\text{wall}}}$, which is in this regime approximately proportional to $R_0$. Therefore the larger the bubble, the more slowly it grows, and so during a given rarefaction cycle, the less the degree of growth it achieves. To put this another way, the maximum radius $R_{\text{max}}$ achieved by a bubble during the growth phase of inertial cavitation is:

$$R_{\text{max}} \approx \frac{4}{3\rho_0}(P - p_0) \sqrt{\frac{2}{\rho_0 P_A} \left[1 + \frac{2(P_A - p_0)}{3p_0}\right]^{1/3}} \quad (68)$$

(Apfel, 1981; Leighton, 1994, Section 4.3.1(b)(ii)) which is independent of the initial bubble radius $R_0$ (a point in agreement with high speed photography—Fig. 21). Again, it is because large bubbles do not grow to such a great extent as small bubbles (assuming the small bubbles are sufficiently large to grow at all—Eq. (66)), that bubbles larger than the threshold size in Fig. 20 do not have subsequent collapses which attain conditions sufficiently extreme to be termed ‘inertial cavitation’. This is because, if one considers a bubble which has expanded to a maximum radius $R_{\text{max}}$, then during the subsequent collapse the wall accelerates inwards under the external pressure in the liquid $p_{\infty}$. The kinetic energy acquired by the liquid when the wall speed has reached $\dot{R}$ is

$$\phi_{\text{KE}} = \frac{1}{2}p_0 \int_{R_0}^{\infty} 4\pi r^2 \dot{r}^2 \, dr = 2\pi \rho_0 \dot{R}^2 R. \quad \text{This must equal the work done by the difference in pressure between that found far from the bubble ($p_{\infty}$) and the pressure in the liquid at the bubble wall ($p_L$)},$$
which is \( \int_{R_{\text{max}}}^{R} (p_L - p_\infty) 4\pi R^2 \, dR = 2\pi \rho_0 R^3 \dot{R}^2 \). In the limit that the cavity contains no gas at all, the liquid pressure \( p_L \) just outside the cavity is zero (if surface tension is assumed to be negligible): this is the Rayleigh collapse, and bubbles which were initially very small would tend to this in the early stages, since their gas pressure when \( R = R_{\text{max}} \) is very low. However bubbles which are initially much larger than this will contain a significant gas pressure when \( R = R_{\text{max}} \) since they have not undergone such an extreme expansion. As a result, \( p_L \) will take a finite value in the above energy balance (Leighton, 1994, Section 4.2.1(a)), and the kinetic energy achieved by the liquid on collapse will be less.
There is therefore a critical size range\footnote{An imperfect analogy of this can be found by imagining shooting a stone from a catapult. As discussed in the text, it is a feature of inertial cavitation that those bubbles which undergo it in a given sound field tend to grow to a maximum size (prior to collapse) which is similar for all such bubbles (Leighton, 1994, Section 4.3.1(b)(iii)). In the analogy, this is similar to the stretching of the catapult elastic to a length of approximately 1 m: the Y-shaped stick is held out in front, in one hand, whilst the stone is held in the other hand, which is drawn back behind the shoulder. The tension in the elastic is an imperfect analogue of the gas within the bubble which, because of the expansion, is at low pressure. If one were to be considering a very small bubble, this might be represented by very short elastic: the forces causing expansion are insufficient to produce growth to the degree required (one cannot draw the elastic to 1 m length, just as the surface tension forces in the bubble hinder growth). The draw in the catapult is small, and the stone does not project far. If one were to consider a large bubble, this might be like having elastic which initially (i.e. when relaxed) is too long (say, 80 cm). Though it can readily expand to the full 1 m, when it does so the elastic energy stored is insufficient to project the stone very far. There is a critical range of relaxed lengths for the elastic which will cause the stone to be projected to far distance.} in which, for a given sound field, the initial size of the bubble must fall if it is to nucleate inertial cavitation (Apfel, 1981; Flynn and Church, 1984; Holland and Apfel, 1989; Leighton, 1994, Section 4.3.1). The lower the frequency, the wider this range. This is one reason why ultrasonic cleaning baths exploit relatively low ultrasonic frequencies (20–30 kHz), in order to generate as much cavitation is possible. To take an example from the opposite end of the frequency scale, the original plot on which Fig. 20 is based played a part in defining the mechanical index (MI) (Holland and Apfel, 1989; Apfel and Holland 1991). The MI is defined as the ratio of the peak rarefractional pressure (expressed in MPa) to the square root of the centre frequency of the pulse (in MHz). The MI can be used to provide a rough form of estimation of the likelihood of cavitation during MHz insonification (Barnett et al., 1993; Meltzer, 1996; Carstensen et al., 1999; Duck, 1999; Abbott, 1999). Since at MHz frequencies the U-shaped curves of Fig. 20 are narrow, the MI is defined with the assumption that a bubble nucleus having a radius corresponding to the minimum of the curve will be present. However it is important to appreciate the additional assumptions which have been made now that the MI is now part of the AIUM/NEMA Real Time Output Display Standard for the on-screen labelling of acoustic output on diagnostic ultrasound systems (thermal effects are characterised by the Thermal Index; AIUM/NEMA, 1992, 1998). These assumptions arise primarily because in clinical use there is no direct measurement of the peak rarefractional pressure (and indeed, since this pressure varies throughout the field, the cited MI refers to one point in the field, usually near the transmit focus of the transducer and near the centre of the scanned plane). The value of MI which is displayed on-screen is automatically estimated by the circuitry in the scanner using the output to the transducer, and as such it assumes that the medium is simple and uniform and has an attenuation of 0.3 dB/cm MHz. The circuitry does this, of course, without having any direct knowledge of the medium which is being scanned (i.e. whether it is in vivo or in vitro, or whether it contains contrast agent or not), such that it is possible to set up conditions (e.g. in vitro) where the ability of the on-screen MI to estimate the likelihood of cavitation is compromised.

5. Scales in space and time

5.1. Frequency ranges

Fig. 22 shows several identifiable features on a scale of the frequencies. It is based on a demonstration of such scales, where five markers (indicating, respectively, 20, 40, 60, 80, and 100 kHz) were placed in a lecture hall at 1 m intervals from a datum (representing 0 Hz). The low frequency limit of human hearing, which is taken to be at 20 Hz, would occur 1 mm from the datum on this scale. The upper frequency limit of human hearing, taken to be 20 kHz, occurs 1 m from the datum, and represents the lower frequency limit for the ultrasonic range.

It is in this low kHz regime that some key technologies operate, for physical reasons which are different depending on whether the modality is being used for diagnosis or material processing (termed ‘therapy’ for biomedical applications).

Consider diagnostic modalities. Sub-bottom profiling of the seabed, for example, uses audio and ultrasonic frequencies up to a few tens of kHz for geophysical surveying for industries involved in harbour construction, petrochemical prospecting etc. (Fig. 23). The frequency chosen illustrates a compromise found in much of diagnostic ultrasonics. The first component of this compromise arises because an increase in ultrasonic frequency promises better spatial resolution (for example in imaging). This is based not only on the Rayleigh criterion for...
the separation of objects laterally (which in its most basic terms requires that, for good imaging, the wavelength be significantly smaller than the object to be resolved,) but also for resolution in terms of the range from the sensor. This is determined by the duration of the ultrasonic pulse, which in general can be more brief (increasing range resolution) the greater the ultrasonic frequency. However with increasing frequency comes, in general, an increase in absorption, and hence a decrease in the depth from which echoes can be received with an acceptable SNR. A compromise must be found between this loss of 'penetration depth' and the ability to resolve the size of object of interest. For geophysical work, this compromise is illustrated in Fig. 23 and its caption: the highest resolution profilers currently available, with resolutions of the order of 1 cm (i.e. O(1 cm)) require a frequency of 150 kHz, which gives a penetration depth of about 3 m (Mindell and Bingham, 2001). In contrast standard geophysical surveying (where resolutions of O(10 cm) or less are adequate) uses frequencies of <10 kHz, and commensurately has greater penetration depths (up to ~20 m depending on the seabed type). The wavelengths in pure air ($\lambda_a$) and pure water ($\lambda_w$) under STP at key frequencies are shown in Fig. 22.
For material processing, the compromise is based on other criteria, some of which overlap with the above. Penetration depth is also an issue here, particularly as the material to be processed is often contained in pipework. If this is the case, the insonifying field must propagate with sufficient amplitude to the point of interest, often crossing numerous interfaces. For example, if the transducers are mounted on the outside of a pipe, then the sound needs not only to propagate through the material of the pipe and its contents without severe attenuation: it must in addition not suffer too great attenuation at the material interfaces between transducer faceplate and the outer wall of the pipe, and between the inner wall of the pipe and its contents (note the values of $\left| 1 - R^2 \right|$ shown in Table 1). Whilst to first order these reflection losses are frequency independent, there are numerous complexities, such as the generation of other modes (e.g. shear waves) when a compressional wave is incident on an interface. If the choice of frequency is done carefully, other features might come into account: the pipe example illustrates this well, when the choice of frequency may be made to coincide with, or avoid, modes of the pipe (Leighton et al., 2002; Birkin et al., 2003a, b). Alternatively, if the ultrasound field is to be used for processing the material in the pipe, for example to cause crystallisation, it may be very important to avoid leakage of the sound from the region where processing is intended to occur, to other regions of the pipe (where premature crystallisation might cause blocking): this can be prevented by careful engineering, which includes consideration of the choice of frequencies.

If the objective is to process the material, then the mechanism by which that processing occurs may influence the choice of frequency. If heating (hyperthermia) is required for the processing, a high frequency may be chosen to generate this. Increasing the frequency tends to reduce the likelihood of cavitation (Section 4.3). Therefore for ultrasonic cleaning baths, ultrasonic frequencies of 20–30 kHz are chosen. Alternatively, some processes (e.g. in food production) may require that the ultrasonic processing (e.g. crystallisation) be undertaken while eliminating the risk of cavitation, as it can oxidise the product and generate ‘off-flavours’. For such a field the frequency should be high (Section 4.3), the $I_{TP}$ should be low, but the $I_{TA}$ should be sufficiently high to generate the processing. Conversely lithotripsy requires cavitation without hyperthermia. As a result the $I_{TP}$ for lithotripsy is high, but the $I_{TA}$ is made low by having a duty cycle of 1:10^5 (i.e. 1 s off-time for every 10 μs on-time), with an oscillatory frequency centred on several hundred kHz (Fig. 24). Note that, although cavitation is required, the chosen frequency is rather higher than the 30 kHz used in the cleaning bath example above, where cavitation was also required. The reason why lithotripsy uses higher frequencies is the need to focus the sound field onto the kidney stone in order to fragment it, without damaging the surrounding tissue. Therefore the ability to focus, and to resolve (in both space and time/range), are also factors in the choice of frequency.
In recent years there have been many papers investigating which is the ‘best’ frequency for a given ultrasonic processing application, by ‘spot-checking’ the sonochemical effect at a half-a-dozen frequencies covering the range 20 kHz to over 1 MHz (Mark et al., 1998; Hung and Hoffmann, 1999; Sato et al., 2000; Kojima et al., 2001; Becket and Hua, 2001). Such approaches do not convey the message that, because net frequency response of the system is dependent on the frequency responses of several particular pieces of apparatus, the results could be highly laboratory-specific. As a specific example, the results of Fig. 19 show that a change of 1 kHz (≈1% of the driving frequency) can dramatically change the chemiluminescence produced by cavitation: in such a system, spot-checking the chemical activity at half-a-dozen frequencies spaced 10s or even 100s of kHz apart would provide a experiment-specific result for the ‘best’ frequency. This is because in most applications, the amount of processing generated is strongly dependent on the amplitude of the sound field in the material which is to be processed,\(^{23}\) a point upon which the remainder of this subsection will expand.

Having said that the amplitude is important, how this manifests itself depends on the mechanisms in question (pressure or intensity, spatial and temporal averaging, etc.). In many cases, that acoustic pressure which corresponds both to the spatial peak and to the temporal peak is a useful indicator of the processing conditions. However the value of this depends on many factors. In the words of Apfel (1984): “Know thy liquid; know thy sound field; know when something happens.” As shown in Fig. 25, the chemical, physical and biological effects of cavitation depend on both the type of cavitation (e.g. inertial, non-inertial, jetting, fragmentary) and its location (Leighton, 1995; Leighton et al., 2005a). Both of these factors depend strongly on the local sound field at the bubble and on the sizes of bubble present in the population. These two together, for example, characterise the inertial cavitation threshold, and also the locations to where bubbles might migrate and accumulate under radiation forces (depending on the other relevant forces present, e.g. flow, turbulence, streaming, etc.). Such accumulations will in turn affect the local sound field, through the processes of channelling, dispersion, absorption, diffraction, scattering and shielding (Fig. 13). They will in turn affect the bubble size distribution through their influence on the processes of coalescence, fragmentation, rectified

\(^{23}\)That is not to say that the amplitude is the only factor: Section 5 illustrates in addition how the frequency can contribute to the spatial and temporal characteristics of the ultrasound field. As shown in Section 2, the frequency is important for considering hyperthermia and, with cavitation, the initial distribution of bubble nuclei is also key (Section 4.2). The nuclei distribution is often ignored, partly because it is often (but not always—Shortencarier et al., 2004) outside of the control of the user, and also because if many pulses are to be projected into the liquid, each causing cavitation, the initial distribution might not be as important as the hysteretic effect generated by the nuclei which survive from one pulse to the next—Leighton et al., 1994, Section 5.3).
diffusion and shape stability. As a result, the observed effect depends on the characteristics of the cavitation, which are determined by the local sound field and the bubble size distribution. However there is feedback from the cavitation, which influences these two key parameters.

It should be noted that the words of Apfel, quoted above, specify the minimum criteria that must be met if a reference for cavitation is to be produced. This is because they refer to the threshold for cavitation, and not to the degree of cavitational ‘activity’. The latter is quantified through measurement of some effect produced by cavitation, be it luminescent, chemical, acoustic, erosive, biological, etc. (Leighton et al., 2005a). Hence to measure cavitional ‘activity’, it is necessary to understand what is being measured, and its place amongst the other effects. It is for example no use basing a sensor to monitor the efficacy of cleaning baths on acoustic emission if that emission bears no relation to the amount of cleaning which occurs. Birkin et al. (2003b) attempted to identify the existence, or otherwise, of such correlations. A cylindrical sonochemical reactor was driven at frequencies from 20 to 160 kHz, with frequency increments as small as 1 kHz. Over the same frequency band, a hydrophone was used to monitor the acoustic pressure at three locations within the cell. A number of experimental parameters which reflect cavitation were monitored: multibubble sonoluminescence (MBSL), multibubble sonochemiluminescence, degradation of an organic species (medola blue), the Fricke reaction, the Weissler reaction, hydrogen radical trapping using the formation of CuCl₂⁻ and the emission of in-air broadband acoustic signals across the audio frequency range as the drive frequency varied from 20 kHz to 160 kHz. Strong correlations were observed between both luminescences and the sonochemical reaction rates. The peaks in activity followed the frequencies at which strong modal structures occurred within the reactor, rather than reflecting the resonances of the drive transducer (Birkin et al., 2003b). The in-air acoustic emission did not correlate so well with the luminescences and sonochemical indicators. Measurements of the drive acoustic pressure amplitude, which was measured at only three locations in the cell, could be used to predict at which frequencies cavitation (and hence sonochemical activity) would be initiated: when a peak in one occurred, a peak in the other was seen. However the relative magnitudes of the peaks in drive pressure amplitudes did not reflect the relative magnitudes of the peaks in sonochemical activity. These results were interpreted in terms of the frequency dependencies of the various components of the system (Fig. 26).
This section outlined the range of frequencies used for ultrasonic applications, and the reasons for choosing a particular frequency. Having made this choice because of a given priority (such as generating a resonance, cavitation, hyperthermia, or a particular resolution), there are implications of this choice with respect to how it affects the process in question. Perhaps the most important of these is through the spatial interactions of the sound field with its surroundings, as characterised by the value of $ka$. This is the topic of next section.

5.2. Implications of the choice of $ka$

If $a$ is the characteristic linear dimension of a physical inhomogeneity in a medium, its interaction with the sound field depends, amongst other quantities, on the value of $ka$, where $k$ is the acoustic wavenumber.

The degree to which consideration of this parameter is very common in acoustics may surprise readers who are more familiar with consideration of electromagnetic radiation. This is because the speed of sound is so much less than the speed of light. If $ka = \omega a/c = 2\pi Ya/c$ is much less than unity, the body is considered to be too small to disturb the wave field: if such a body is placed in the sound field, barely any diffraction occurs.\footnote{Although the issue may not be simple. Birkin et al. (2005) examined an ultrasonic horn, operating at 23 kHz and with a transducer faceplate of radius $a = 1.5$ mm. A simple calculation for such a horn would show that $ka = 0.14$ for the horn. In the free field it would be an approximately omnidirectional transmitter, although since it is normally dipped into water from above, its emissions would resemble a dipole depending on how close the tip was to the surface. Into the sound field was placed an erosion sensor for which the characteristic dimension $a$ was $\sim 1$ mm: hence for this sensor $ka \sim 0.1$. Therefore the sound field from the horn should not have been directional (although the presence of the pressure-release air/water interface through which the horn tip was dipped would in practice lend a dipole character to the field), and the sensor should have been too small to perturb the field significantly. However these calculated values of $ka$ are made assuming that the wavenumber of interest is that associated with the 23 kHz sound field from the horn. In fact, the ultrasonic cavitational effect detected by the sensor was strongly affected by the shock waves produced by the cavitation immediately preceding the measurement. For these shock waves, the sensor has sensor $ka > 0.5$. Reflection of these shock waves by the sensor greatly influenced the subsequent cavitation. This demonstrated how a sensor of radius $a$, where $ka < 1$ as regards the wavenumber of the insonifying field from an ultrasonic horn, nevertheless strongly influences the ultrasonic processing of the medium because $ka$ can be much larger for those shock waves generated by the cavitation collapse which is produced by the field from the ultrasonic horn.}

Fig. 26. Diagram showing the components and their various frequency dependencies commonly employed in sonochemical experiments (after Birkin et al., 2003b).
and if such a body is a sound source or receiver, its shape has no effect on the sound field, and it projects and
receives omnidirectionally (Wells, 1977; Kinsler et al., 1982; Leighton, 1994, Section 3.3.2(b)). A 1 MHz radio
wave has a wavelength of 300 m: most bodies smaller than buildings would present this wave with a \( \frac{ka}{5} \ll 1 \), and
hence reflection and diffraction of this wave by them would be small. In contrast, a 5.3 MHz ultrasonic wave
in water has a wavelength of \( \sim 280 \mu \text{m} \). Even a 400 \( \mu \text{m} \) radius needle hydrophone is seen as large by the sound
field (\( ka \sim 9 \)) (Fig. 27). Such sensors can therefore be invasive with respect to perturbing the ultrasonic field
they are in place to measure. Even if they were not invasive, the large \( ka \) values means that if the directionality
of the sensor is not measured (Fig. 28), and its orientation precisely known, it would be impossible to be apply
a calibration to measured data. That some researchers have obtained data using hydrophones in inaccessible
places (such as the vagina) where its orientation is difficult to monitor and control, is a testament to the care
with which the experimentation was undertaken (Daft et al., 1990).

The \( ka \) issue is further complicated for the propagating wave, since not only will it often encounter physical
objects for which \( ka > 1 \), but the acoustic impedance mismatches between the object and the host medium will
often be large (Table 1). Consider a 30 kHz bench-top cleaning bath field in which the wavelength\(^{25} \) is 5 cm,
such that each side of the bath will be approximately several half-wavelengths long. This will make the field
inhomogeneous (Fig. 29), complicating its measurement (even if the users are aware of the inhomogeneity) and
making its performance very dependent on the exact position within it of the item to be cleaned. When the
investigations are extended to the studies which describe the ‘best’ frequency for an ultrasonic process (Section
5.1), it is clear that not only are the transfer functions of Fig. 26 a consideration, but also the frequency
response(s) of the particular vessel(s) used for the test can dominate the measurement.

The final example of the implications of having sound speeds such that \( ka \) will commonly exceed unity, is
found in the transducers used to generate sound fields. Consider first the opposing case, that of a transducer
for which \( ka \ll 1 \) with respect to the driving oscillatory field in the free field. A simple calculation (i.e. one
which disregards the cautionary tale of footnote 24) would suggest that such a transducer would act as a
monopole source, and exhibit no near field. However for transducers with larger values of \( ka \), there will be a
measurable near field (Fig. 30). This is characterised by a regions where, for constant driving conditions, even
in the free field the sound field amplitude can change dramatically for very small changes in the relative
position of the sensor with respect to the transducer. Similarly, any perturbations to the field (such as the
passage through the beam of inhomogeneities in the medium which change the sound speed or scatter
the field), can give rise to very large fluctuations in amplitude in the detected signal. As one moves away from
the transducer faceplate along the axis of the transducer, the near field region is characterised by a series of
closely-spaced maxima and minima. The near-field extends out from the transducer to a distance of around:

\[
L_{\text{near}} \approx \frac{\alpha^2}{\lambda} = \frac{ka^2}{2\pi}.
\]

This corresponds to the distance from the faceplate to the axial acoustic pressure maximum which is furthest
from the faceplate (Wells, 1977; Fig. 30). At further ranges, in the free field the beam amplitude should fall off
steadily with increasing on-axis range from the source.

Because the propagation speed of sound is much less than that of light, these near field complications can be
very much more commonly a problem in ultrasonics than they are in electromagnetic studies. Even in
ostensibly unfocused beams, apparently negative attenuations can readily be detected, since the amplitude
further from the source can be greater than that close to the source. If the attenuation of the beam is being
attributed to absorption by the medium, and then inverted to determine physical properties of that medium
(as is done for bubble detection, osteoporosis monitoring, the quality of pharmaceuticals, foodstuffs,
domestics, paint and pottery, etc. (Leighton, 2004)) then lack of appreciation of near field effects would lead to
nonsensical answers. A typical scenario, unfortunately very common in both experimentation and commercial
ultrasonic technology, outlines several of these problems (all the issues here are not merely fictitious examples,
but rather, have been identified by the author in sensors he has been asked to validate). Consider a thin-walled
pipe of internal diameter 30 mm. A source and a receiver transducer (each of 10 mm faceplate radius) are to be

\(^{25}\)Whilst it is acceptable for the purposes of the thought-experiment described here (such that at 30 kHz the wavelength is 5 cm), it is not
safe to assume that the sound speed in a cavitating field is that of bubbly water (see (Fig. 19)).
mounted opposite each other, ostensibly to interrogate the cross-section of the pipe (Fig. 31). A coupler is added to each faceplate to mould the shape of the flat transducer faceplates to the curved wall of the pipe, and to match their impedances and ensure no air gap between transducer and pipe (which would introduce large impedance mismatch (Table 2)). A practical example of the use of such a coupler is shown in Fig. 32. In the hypothetical situation of Fig. 31, the source emits a signal (continuous wave, tone-burst or chirp) and this waveform is seen in the output from the receiver. The amplitude of this signal is inverted using a model (let us say of acoustic absorption by bubbles, the inversion being used to estimate the bubble population in the pipe, by projecting a series of tones extending from 30 to 300 kHz). All seems to go well, and the experimenters progress to emitting a series of pulses, like the ones shown in Fig. 4, so that the inversion can be based on both attenuation and sound speed. Again, all goes well, until the discovery that the signals are also detected when the receiver is not in contact with the pipe. Being acoustically uncoupled from the pipe, it should detect nothing, but in fact its output is the result of electromagnetic pick-up from the amplifier/transmitter ((i) in Fig. 31). Unless a decoupling test is undertaken, this can be overlooked as the EM pick-up closely resembles the transmitted waveform and so is as the experimenter expects. However the problem should have been spotted from the lack of an appropriate delay introduced by the transit time of the pulses (of roughly 0.03 m/1500 m s\(^{-1}\) ~20 \(\mu\)s). Suitable electromagnetic shielding is added, and the detector now receives far less ‘clean’ waveforms, as expected for such a reverberant environment. Again, no problems are found in either

Fig. 27. Simulation using acoustoelastic finite difference solver (AfiDS—Hurrell, 2002) showing the effect on the acoustic field of a needle hydrophone. The left side of the graph is an axis of rotational symmetry. A plane pulse waveform (having centre frequency 5.3 MHz) has travelled through water, downwards from the top of the figure. It is shown at the moment when it has passed the base of the tapering in the stainless steel tubular tip of the needle hydrophone. Contained within this steel are (working from the outside towards the centre) a thin layer of polymer, a thicker layer of polymer, and finally, in the centre, a rod of copper. At the very tip of the needle is the active sensing element of the hydrophone, a disc of radius 200 \(\mu\)m. The pressure perturbations of the pulse are broader in the copper, and have travelled further downwards, than in the water, because the sound speed there is greater than in the water. Amongst the pressure field in the steel can be seen the antisymmetric Lamb wave, and its radiation into the liquid can be seen ahead of (i.e. below) the original pulse in the water. With a radius of 400 \(\mu\)m, the \(ka\) of the hydrophone to this waveform in water is about 9. Reflected waves, travelling back upwards from the hydrophone tip, are clearly visible and, even with tapering, diffraction from the tip perimeter is evident. The waves is shown at a time 0.825 \(\mu\)s after the incident pulse first appears at the top of the figure (simulation by A Hurrell). A schematic to the left of the graph clarifies the geometry and features.
continuous-wave or pulsed operations, but after a while of apparently satisfactory operation, the pipe is one day emptied and the waveforms are still detected (and the inversion still uses them to invert for a now obviously fictitious bubble population). The problem is identified as being the result of waves (for which there can be several types—see Fig. 1) propagating through the pipe walls between transmitter and receiver, which dominate the signal ((ii) in Fig. 31). As a result the transducers are designed to be embedded in the pipe wall, and insulated from it acoustically. This new set-up appears to work and the detected waveform is inverted to generate a bubble population, but the now cautious users ask for ground truthing, a separate measurement of the bubble population, which disagrees with that of the original device. This exposes a number of problems in the arrangement. There is reverberation in the pipe ((iii) in Fig. 31), so that the measured attenuation contains contributions from much longer path lengths than that provided by the aligned axes of the transducer, making conversion of such an attenuation into a dB/m estimation difficult. Attempts to remove reverberation by time-gating the received signals prove to be unsuccessful: the path difference between the direct path (30 mm) and the first echo from the pipe wall \(2 \times \sqrt{15^2 + 15^2} \approx 42.4\) mm corresponds to an interval in pure water of \(\left(\frac{42.4 - 30}{30}\right) \times \frac{15}{C} \approx 8\) μs, i.e. less than three cycles at 300 kHz, and only one quarter of a cycle at 30 kHz. This difficulty is compounded if the system has a time-response (such as a ring-up in the response of the bubbles, noting that the inversion assumes they are at steady state). Even without reverberation, the beam pattern that would occur in free field would be problematic: As a simple application of equation (Eq. (69)) would have shown, the near-field for this transducer can extend well into the measurement volume (at 300 kHz, \(a^2/\lambda = 20\) mm, Fig. 30). As a result, the introduction of small inhomogeneities like bubbles into the measurement volume can give changes which are difficult to interpret. For example, since the path difference (between the path along the transducer axis and the path from the edge of the transmitter to the centre of the receiver in Fig. 31) is \(\sqrt{30^2 + 10^2} - 30 \approx 1.6\) mm, equivalent to \(> 30\%\) of a wavelength at 300 kHz, then a small change in sound speed along one of those paths could dramatically change the received signal, such that a sound speed change will be interpreted as a change in absorption.\(^{26,27}\) Indeed, those changes can be of either sign: the problem of apparent negative absorption in the near field was introduced earlier, and it is insufficient for the operator to reject such nonsensical results but retain those which, whilst similarly compromised, are

\[^{26}\text{This is because of phase changes induced along some propagation paths, which affect the signal on summation at the receiver. Indeed even in the far field, changes in sound speed can affect beam patterns sufficiently to appear as additional attenuations, although here it is far simpler to correct for the effect (Robb et al., 2006).}\]

\[^{27}\text{The assumption is made here that receivers are ‘phase sensitive’ rather than ‘phase insensitive’ (Busse and Miller, 1981).}\]
not so obviously erroneous. Indeed, it is apparent that measurement of attenuation and pulse propagation speed using a single source-receiver pair is difficult, even in the far field, as both quantities are realised through the propagation of a signal past two measurement points. If only one hydrophone is used, assumptions must be made about the signal amplitude and timing at the source, which can be simple to make but very difficult to verify.

Fig. 29. Sonochemiluminescence pictures recorded in an ultrasonic bath as a function of the volumetric power density. The volumetric power density increases, respectively, from picture (a) to (h). The power settings were 2.5, 5, 7.5, 10, 20, 32.5, 42.5 and 47.5 \( \text{W dm}^{-3} \), respectively. The temperature of the bath was 21°C. The height of the camera above the surface of the 18 l luminol solution was 45.5 cm ± 1 cm. Regions of high cavitation activity (‘hot’ spot) and low cavitation activity (‘cold’ spot) are labeled. Also labelled is the cross-section of the finger of a latex glove which was used to contain sonochemical reactions (from Leighton et al., 2005b).
As a result of these problems, the device is drastically redesigned to use two hydrophones, and to place them in the far-field zone of the transducer. Even here, there are problems. The measurement is of group velocity, which enters the inversion theory as phase velocity. In this hypothetical scenario, the theory assumes that free-field conditions exist (Commander and Prosperetti, 1989; Commander and McDonald 1991), whereas in the pipe they do not (Leighton et al., 2002). The theory assumes steady state, but there is considerable ring-up and ring-down (Leighton et al., 2004). Experimentally, when the attenuation is high (at the highest frequencies, say) the source amplitude is increased to obtain a good SNR, but no corresponding increase is seen in the detected signals. The problem is that the propagation is becoming nonlinear, such that energy is being pumped out of the fundamental frequencies to higher harmonics (in the manner discussed in Section 3). These are at higher frequencies than the upper limit of the receiver bandwidth (300 kHz), and so this energy is becoming ‘invisible’ to the detector. Increasing the source amplitude increases the nonlinearity and augments the effect.

6. Ultrasound in air

6.1. Categories of exposure of humans to ultrasound in air

The circumstances under which humans and animals can be subjected to ultrasound in air fall into three categories. First there is the unintended exposure because some process (such as an engine, ultrasonic dental tool, or ultrasonic cleaning bath) generates ultrasound as a by-product of its operation. This might be termed ‘ultrasonic noise exposure’. Second, there is the unintended exposure because some process (such as an...
ultrasonic range finder) requires the generation of a specific ultrasonic signal as key to completing its task, but in addition to insonifying its inanimate target, it also exposes a human or animal to ultrasound. This would correspond to the in-air equivalent of the unintentional exposure of cetaceans to sonar discussed in Section 2.3. Third, there are devices which are designed to expose humans and/or animals to ultrasound in air in order to elicit some subjective response.

As regards the first category, whilst numerous examples exist, as well as anecdotal evidence of varying quality, to the author’s knowledge there has been no census of such ‘ultrasonic noise exposure’. Where the generation of ultrasonic noise also generates audiofrequency noise, the level of audible noise might stimulate the use of hearing protection which could act fortuitously to protect against ultrasonic hazard. The third category has shown a rapid increase in products, many with the intention of generating discomfort and with little ability to undertake measurements to comply with what few safety guidelines there are. These will be given special consideration in Section 6.4. With respect to the second category, as has been stated earlier the applications of ultrasound in air are not as numerous as those in liquids and solids. This is in part because of the plethora of alternatives available for use in air, most notably optical and other EM methods. Where they do occur, they are restricted to the low ultrasonic frequency range, because of the high absorption in air and its increase with frequency (Table 2).

In the vast majority of circumstances, the hazard issues associated with ultrasound in air relate to their interaction with the hearing and balance organs, because otherwise the acoustic impedance mismatch between air and tissue hinders penetration (Section 2). The issues associated with ultrasound in liquids and tissue (hyperthermia, cavitation, microstreaming, etc.) would not normally be an issue unless direct contact is made between the user and the source. Whilst this may at first sight appear to be a remote possibility, the presence of ultrasound in air might not just be the result of a ‘category two’ application with exploits the ultrasound as a signal (such as in range finding, intruder detection, or pest deterrence). It might also be the result of a ‘category one’ noise leakage from apparatus designed to generate ultrasound in solids, liquids or even other organs in the body (such ultrasonic apparatus include cleaning baths, welding, drills, NDT equipment and

---

**Fig. 31.** Consider a thin-walled pipe of internal diameter 3 cm. A source and a receiver transducer (each of 1 cm faceplate radius) are to be mounted opposite each other, ostensibly to interrogate the cross-section of the pipe (the beam pattern of one such transducer, in a rigid baffle and radiating into the free field, is shown in Fig. 30). However the received signal can also receive contribution from path other than along the transducer axis, including (i) EM pickup of the driving signal, (ii) waves within the pipe wall, and (iii) reverberation in the pipe. In addition, the beam pattern for one of these transducers, operating at 300 kHz, is modelled in Fig. 30, indicating that much of the interior of the pipe would be in the near field at this frequency.
lithotripters—Dawson et al., 1994). In such circumstances the contact hazard is more obvious, and the user may be unaware of the presence of ultrasound in the air. In other circumstances, whilst ultrasound can be generated, it is accompanied by significantly high levels of audiofrequency noise, which would deter users from hazardous locations (or, as discussed above for ‘category one’ noise, it might force the use of hearing protection). The jet engine, which was (probably erroneously) cited in the popular press as a source of “ultrasonic sickness”, falls into this category (Parrack, 1966; Lawton, 2001). Whilst the earliest studies arose from concerns about occupational exposure to industrial ultrasound, there has been a growth in domestic, and even recreational, devices which could generate ultrasound. These could generate category two or category three exposures. Section 3.3 introduced devices to generate a localised sound field by utilising nonlinearity in the propagation. Other commercial devices include humidifiers, positioning devices, echo rangers, garden pest repellents and ultrasonic weapons to be used by joggers against dogs. “Ultrasonic weapons” are now advertised. Note that with a wavelength of more than 1.5 cm in air, a 22 kHz handheld device would be unlikely to be large enough to generate a tight beam of ultrasound, without leakage into directions other than

Fig. 32. An ultrasonic device for the detection of bubbles in the opaque liquid ‘casting slip’ which is used by the pottery industry. The detector consists of ultrasonic source (S) and receiver (R) transducers mounted across from each other on the pipe (P). The body of each transducer is labelled ‘T’, and the faceplate of each is coupled to the pipe by an insert (M) which can be used to overcome mismatches both in shape and acoustic impedance. The system held in place by a clamp (C) which is easily portable, and can be mounted at any position along the pipe, so allowing the operator to track down the source of the bubbles (Leighton, 2004; Yim and Leighton, 2006).
that in which is it ‘fired’. The author could find no census of devices which quantified their ultrasonic output. In addition, surveys of the effectiveness of ultrasonic devices have found evidence of claims in advertising which are not warranted by the proven effectiveness.28

6.2. Contact and non-contact exposure—the example of dental ultrasonics

For several decades guidelines have been proposed for the maximum recommended exposures for humans to low frequency ultrasound (i.e. having frequencies of between about 20 and 60 kHz, although the upper limit is not fixed). These are discussed in Section 6.3. However these discussions have two major assumptions which are seldom explicitly stated. First, they refer to in-air, non-contact exposure. Second, barring subjective effects for which a mechanisms of generation is unknown, the assumption is made that the hazard is best monitored by the effect of the ultrasound on the ear (usually quantified through the ability of ultrasound to produce a temporary threshold shift for hearing at some low-kHz audiofrequency).

Before the review in Section 6.3, it is enlightening to recall the discussions regarding the differences between contact and non-contact ultrasonic exposure, as introduced in Section 2.2. That section stressed the importance, of including in the assessment of hazard, (i) the various transmission paths to the organ in question and (ii) the possibility that more than one organ should be considered. Examples have been given throughout this paper of where it is obvious that contact ultrasound should be considered (e.g. with MHz biomedical ultrasonics). In many applications, it is clear that only non-contact exposure will occur in normal practice (such as in the ‘museum’ example of Section 3.3), although the possibility of contact exposure should be considered with respect to accidental touching of the transducer. Whilst this might seem remote in normal operation for some devices, for others (such as hand-held devices to deter dogs from joggers) it is not.

This point is well illustrated by discussion of the considerations which should be undertaken to assess the hazard caused by dental ultrasonic apparatus (which includes drills, files, and scalers). There is the obvious question of what acoustic levels occur at the ear as a result of through-air transmission (Wilson et al., 2002). This might be thought of as an issue primarily for the dental practitioners who, unlike the patients, are routinely close to such apparatus and so receive day-to-day exposure. However the patients themselves are in physical contact with the instruments, so that conduction pathways other than airborne should be considered: the acoustic impedance mismatch between the tool and the tooth, the tooth and the jawbone, the jaw and the skull, and the skull to cochlea, are all very likely to be less severe than the tool-air-to-skin pathway, and without detailed consideration it would be difficult to assess the relative importance of the various conduction routes (Walmsley et al., 1987a). Other potential sources of hazard from ultrasonic scalers have been raised, including heating of the tooth during scaling, vibrational hazards causing cell disruption, possible platelet damage by cavitation, associated electromagnetic fields that can interrupt pacemakers, auditory damage to patient and clinician and the release of aerosols containing dangerous bacteria (Trenter and Walmsley, 2003). The possibility of such hazards is not unexpected given that this ultrasonic device is designed to produce beneficial effects through aggressive action at the interface of solid, liquid and gas (Walmsley et al., 1987b Ahmad et al., 1987; Roy et al., 1994; O’Leary et al., 1997).

Given the complexities which should be considered, the simplistic approach taken in the literature is surprising. As discussed in Section 2.3, it is extremely difficult and potentially very misleading to assign a single dB ‘sound level’ to a signal which contains ultrasonic or infrasonic components. It is possible to obtain a false sense of security from such statements as “Kilpatrick (1981) has listed the decibel ratings for various office instruments and equipment, which amount to 70–92 dB for high-speed turbine handpieces, 91 dB for ultrasonic cleaners, 86 dB for ultrasonic scalers, 84 dB for stone mixers and 74 dB for low-speed handpieces” (to quote from Szymańska, 2000). This was written (without reference levels for the dB scale) in an article assessing the hearing hazard associated with dental surgeries, and goes on to state “The dentist should maintain a proper distance from the operating field. Kilpatrick (1981) recommends the distance from the dentist’s eye to the patient’s mouth to be 14 in, i.e. about 35 cm. When the operator is closer, decibel rating increases.” The attention is wholly on the through-air transmission path to the dentist’s pinna. There is no

reference made to the route by which sound may affect the patient’s cochlea through, for example, bone conduction from the jaw, a route for which the transmission losses from the working tip may be much less (Table 1). Furthermore, not only is that transmission path uncharacterised, but the source level is unknown. To explain, an ultrasonic dental drill operating in air would most likely be an inefficient source of radiated acoustic energy. It becomes a more powerful source once the drill is in contact with the tooth, but whilst we may well measure the acoustic energy in the air which results from that contact, it is very difficult to measure the levels in the tooth, jaw, skull and hearing apparatus. The levels at audio-frequencies might be subjectively estimated using loudness matching (although the bandwidth of the signal might make this difficult), by measuring the thresholds by which the noise masks of a series of tones of different frequencies, but these would not be possible for the ultrasonic components. This might require measuring the acoustic signal radiated into the ear canal through the canal walls and drum during drilling, and then placing a bone vibrator on the teeth and measuring the transfer function between the bone vibrator and the ear canal pressure. The analogy of a masonry drill is illuminating: it generates more sound when in contact with a brick to drill a hole than when running in air, but whilst we can measure the noise in air resulting from such drilling, it is no simple matter non-invasively to measure the levels in the brick.

6.3. Guidelines for non-contact in-air exposure to ultrasound

The mechanisms by which ultrasound interacts with the various sensing organs of the ear is not as well understood as the various physical interactions with less receptive matter (involving cavitation, hyperthermia, microstreaming, etc.). Unlike the latter case, in-air ultrasonic hazard has tended to be assessed more by audiologists than biophysicists.

Grigor’eva (1966) found no shift in hearing threshold in the range 250–10,000 Hz, when an unreported number of subjects were exposed to 20 kHz at 110 dB re 20 μPa for an hour (for comparison, a 90 dB insonification for 1 h at 5 kHz did produce a substantial shift). Notable amongst other studies in the 1960s (Smith, 1967; Parrack, 1966) is the work of Acton (1968) who, like Grigor’eva, was also concerned about hearing damage as a result of occupational exposure to industrial ultrasonic equipment. He concluded that 8 h exposures to 110 dB re 20 μPa in the one-third octave bands centred on 20, 25 and 31.5 kHz should not cause hearing loss at audio frequencies. Lawton (2001) points out that, stated in this way, a broadband signal could exhibit levels which satisfy the criteria for each of the one-third-octave bands, but when these levels are added to give the octave level it could exceed the limit. Later, Acton (1975) revised the limit to 75 dB re 20 μPa for the lowest of these frequency bands, as its frequency range extended from 22.5 kHz down to 17.6 kHz, i.e. to within the audio frequency range (and certainly audible to many young females). It was already known that exposure to high frequencies in the audible range could produce subjective effects, including nausea, fatigue, tinnitus, persistent headaches, and “fullness in the ears” (Knight, 1968, Acton, 1973, 1974; International Non-Ionizing Radiation Committee, 1984; Damongeot and André, 1988; von Gierke and Nixon, 1992; Lawton, 2001; National Occupational Health and Safety Commission, 2002). In addition, criteria for narrowband emissions were introduced (Acton, 1975; Acton and Hill, 1977).

From such studies in the 1960s, tentative Damage Risk Criteria, and Maximum Permissible Levels, were therefore first proposed, although without thorough investigations into dose–response relationships (which are still far from complete, given that there is only partial data on occupational—and other—sources of ultrasound). The limits for ultrasound were set to avoid hearing damage at audio frequencies, with the need to avoid any subjective effects also being introduced. This led to further guidelines of exposure to ultrasonic (>20 kHz) and ‘very high frequency’ audible sound (10–20 kHz) (International Labour Office, 1977; International Non-Ionizing Radiation Committee, 1984; Damongeot and André, 1985; Health Canada, 1991). Amongst these works Lawton (2001) also discusses one which stands out as being more lenient in its limits (American Conference of Governmental Industrial Hygienists, 1988), of which he states “This reviewer believes that the ACGIH has pushed its acceptable exposure limits to the very edge of potentially injurious exposure”.

No temporary hearing loss was produced by ultrasound limited to one-third-octave band levels of 105–115 dB re 20 μPa, levels which were then taken to be non-hazardous with respect to generating permanent hearing damage (Lawton, 2001). A temporary threshold shift was found in subjects exposed to 150 dB re
20\,\mu\text{Pa} at 18\,kHz for about 5\,min (Acton and Carson, 1967). According to Lawton (2001) “since the introduction of these recommended limits, there have been no reports showing systematic hearing loss associated with occupational exposure to very high frequency noise”. It is important, of course, to understand fully the implications of treatment of ultrasonic exposure as ‘noise’. This term implies that the emission is a by-product of some other process, and considered to be an ‘occupational exposure’. Such emission might be broadband, or narrow band (Skillern, 1965; Roscin et al., 1967; Dobroserdov, 1967; Acton and Carson, 1967; Herman and Powell, 1981; Holmberg et al., 1995). In the author’s opinion, these historical studies focused on category one and category two exposures, and even here the temporal and frequency characteristic of the emissions may vary greatly, from broadband noise to tonal. The growth in category three exposures in recent years (such as the museum example given in Section 3.3) has probably escaped these historical studies.

As noted in Section 2.3, the dB(A) scale is wholly inadequate for providing guidelines of hazard from in-air ultrasound. Korpert and Vanek (1987) proposed a scheme to introduce a single-number quantification of the levels of airborne ultrasound, both for convenience and to attempt to overcome drawbacks with the Acton guidelines. Their approach has however been criticised by Lawton (2001).

Howard et al. (2005) reviewed the current recommended acceptable exposure limits from standards organisations around the world. They note the general consensus amongst standards bodies on limits for the exposure of ultrasound. However one feature of particular concern to them was the revision by the United States of America’s Occupational Health and Safety Administration to increase the exposure limit by an additional 30\,dB under some conditions (equivalent to a factor of 1000 in intensity).

6.4. Category three exposures: deliberate exposure of humans and animals to ultrasound to elicit some subjective response

Of the three categories of exposure defined in Section 6.1, category three exposures are unique in that the exposure is intentional, and indeed is driven by a commercial imperative. The revision by the United States of America’s Occupational Health and Safety Administration, discussed at the end of the preceding section, is of particular concern given the proliferation of commercial ‘category three’ devices which could expose humans (intentionally or through misuse of, say, a dog deterrent) to high levels of ultrasound. The in-air equivalent of parametric sonar, for which the ‘museum’ example was given in Section 3.3, is one manifestation. Because the generation of the difference frequency is a second-order process (Eq. (16)), the power of the primary frequencies (here, the ultrasonic ones) needs to be very great in order to generate a loud signal at audio frequencies. Whilst the analysis of Section 3.3 describes the role played by propagation in generating nonlinearity, it is not clear to what extent the ear, when driven at high levels, is capable of its own nonlinear response, such that the level of difference frequency generated in the ear may be different from that monitored by a microphone. Whilst power series expansions of the type discussed in of Section 3.3 will never of course describe a subharmonic, a nonlinearity in the ear which is capable of doing so could of course generate an audible response as a result of exposure to a single ultrasonic frequency.

Similarly, there has recently been a proliferation of devices whose mechanism of operation is not clear (even to the extent of whether they generate ultrasound or very high frequency sound). However many are described as ‘ultrasonic’, and advertised for their ability to generate unpleasant subjective effects and engender discomfort, for example for homeland or personal defence. The advertised ability to be able to generate “140 + dB” (for a “Pain Field Generator”, for “potential crowd control applications”) at an unspecified ultrasonic frequency is alarming, even given the comment in Section 2.3 that it would currently be impossible to substantiate the ‘140 + dB’ measurements above 20\,kHz in a traceable way using methods which have been internationally validated. Other devices include a “Blast wave pistol”, advertised as producing “130\,dB” (again, with no reference pressure or indication of the location of the measurement) for “animal control, predators for bird feeders, control of unruly dogs, cats even people”. Whilst it is of course difficult to verify the output of such devices, or even to assess the operating frequency, the intention of using high outputs to generate discomfort is clear.

Indeed, the fact that the sensitivity of the ear to high frequencies is personal, and often age-related, is now being exploited to target specific demographic groups, for example by using high frequencies below 20\,kHz to prevent teenagers loitering around shops: “Police have given their backing to a gadget that sends out an ultra
high-pitched noise that can be heard only by those under 20 and is so distressing it forces them to clutch their ears in discomfort” (Alleyne, 2006). However a mother with babe-in-arms would not be sensitive to the radiation which distresses her child, and therefore would require extra protection to avoid holding it in close proximity to the ‘silent’ (to her) transducer. Whilst the definition of ultrasound cites a 20 kHz boundary, based on the statistics of populations, for the practical purpose of protecting individuals the distinction between ultrasonic and sonic is not so sharp.

7. Conclusions

In answer to the question ‘what is ultrasound?’ there are some simple answers. There is a range of ultrasonic waves which can propagate (Fig. 1). They are commonly described as waves which transport mechanical energy through the local vibration of particles, with no net transport of the particles themselves. The term "ultrasonic" is generally taken to mean that the "frequency" of the wave is greater than the upper limit of human hearing (usually taken to be 20 kHz).

In practice these simple ideas do not translate readily into our experience of ultrasonics. This is in part a result of the extreme sensitivity of the ear. Because of this, audio frequency sounds are restricted to "intensities" (noting from Section 2.3 that it is important to define the intensity with care) which, for most physical interactions of sound with matter other than the ear, would be thought of as being extremely low (Fig. 33). Since the vast majority of our experience with acoustics occurs at such audio frequencies, acousticians have been as culpable as any in propagating, as general truths, those characteristics of acoustics which relate to linear propagation only (Section 2). In practice, the high amplitudes used for much of ultrasonic work means that even the two apparently self-evident statements made in preceding paragraph are questionable: there may indeed be a net transport of particles (see Figs. 9 and 10), and the frequency content of the tonal wave may change in such a way that it becomes misleading to assign to it a single frequency (Section 3) and, further, to compare it to 20 kHz when protecting individuals (see end of Section 6).

This acute sensitivity of the ear leads to the conclusion that there is no single satisfactory way to discuss “exposure”. Consider the range of acoustic pressure amplitudes which are shown in Fig. 33. As for Fig. 22, the schematic is based on a lecture demonstration that was undertaken. It is based on a linear scale, such that the marks corresponding to 20, 40, 60, 80 and 100 kPa occur at respective distances of 1, 2, 3, 4 and 5 metres from the datum.

As discussed earlier (Sections 1 and 6), the particular sensitivity of the ear ensures that the acoustic pressures associated with hearing all occur close to the datum. The commonly accepted threshold for human hearing at 1 kHz, 28.9 μPa (zero-to-peak—see footnote 7) occurs at about 1 nm from the datum on this scale. An amplitude of 20 Pa (which occurs at 1 mm from the datum) would cause pain, and hearing damage would occur at 200 Pa (1 cm from the wall). The range of human hearing is therefore typified by the first cm on this scale.

In contrast, even applications considered by physicists and chemists to be ‘low-amplitude’ (such that, for example, they are used only for material diagnosis and not processing) occur much further from the datum on this scale. Although the signal amplitudes are not often recorded, it is likely than many ultrasonic systems used in ocean monitoring employ amplitudes at the sensor29 of not more than 10 kPa (50 cm from the datum on this scale). As one increases in frequency (Fig. 22) the source amplitudes tend to increase to ensure an adequate SNR at reception.

Under sea surface conditions, an amplitude of just over 100 kPa (5 m from the datum) corresponds to the cavitation threshold for low frequencies (tens of kHz—see Section 4.3). Ultrasonic cleaning baths, sonochemical reactors, dental tools, humidifiers and other devices designed to operate with continuous-wave and pulsed ultrasound at a few tens of kHz, operate in this range of 100–300 kPa (i.e. 5–15 m from the datum on this scale). We will assign to these the inexact label ‘power ultrasound devices’. However the fact that they operate in a frequency range designed to promote cavitation has an interesting consequence. Specifically, the source amplitudes used by such devices does not increase without limit because, whilst cavitation is required for effective operation, increasing the source amplitude for such devices will not deliver

29Of course the amplitudes at the source can be much greater (Leighton et al., 2001; Leighton 2004)—see Sections 2.2 and 2.3.
enhanced cavitation in such devices. This is because the high amplitudes which occur at the faceplate can cause such active cavitation there that the ultrasound is prevented from propagating in to the body of the liquid, and the device ceases to function (Fig. 14).

Most biomedical applications, both diagnosis and therapy, require better localisation/focusing of the ultrasound than is required for the ‘power ultrasound devices’ discussed above. As a result, they tend to use...
higher frequencies (Fig. 22). Because of this, although the acoustic pressure amplitudes they use tend to be greater than those employed by the ‘power ultrasound devices’ (100–300 kPa at 10s of kHz), the risk of cavitation has not increased proportionally\(^{30}\) (Fig. 20). The range 10–100 m from the datum on this scale covers those devices operating with peak acoustic pressure amplitudes of 0.2–2 MPa. These include diagnostic devices, both for imaging (like foetal scanners) and measurement through attenuation and sound speed (as used for bone health monitoring). The only therapeutic devices operating in this range are considered to be ‘low power’, such as physiotherapy devices. Note that the values assigned to acoustic pressures on this scale are tentative (see caption), and that they reflect measurements in water as opposed to tissue (Duck, 1987; Harris, 1999; Duck and Martin, 1991).

A distance of 5 km from the datum corresponds to 100 MPa, typical of the acoustic pressure amplitudes used in High Intensity Focused Ultrasound (HIFU) and lithotripsy. Given the range from 1 nm to 5 km on this scale, it is difficult to find general ideas which span the entire 12 decades in acoustic pressure (or 24 decades in intensity) covered by common interactions of man with acoustic fields. This is particularly the case given that such phenomenon as cavitation can occur, and can generate sound speed fluctuations of a factor of 2 or more on rapid timescales. Furthermore, cavitation can focus the energy both in terms of space and time to create extreme conditions of shear, temperature and pressure. In particular, it would be imprudent to extrapolate from the bulk of our experience with acoustics (i.e. linear acoustics at audiofrequencies) to provide a baseline understanding of what occurs in the ultrasonic regime.

To close the article, some salient points for addressing the question ‘what is ultrasound’ (with respect to the safety topic of this journal issue) are now listed:

(i) Ultrasound affects tissue through a variety of mechanisms. Whilst there are commonly recognised regimes in which certain mechanisms produce noticeable effects (in approximate order of ascending acoustic pressure amplitude: microstreaming, streaming, radiation forces (with and without bubbles and particles), hyperthermia, cavitation), there is a special regime at very low intensities (Fig. 33) for consideration of the effect of ultrasound on the hearing and balance organs of the ear. The degree to which these ear-related effects are understood is generally much less than the degree to which the higher-amplitude effects are understood. The conduction of ultrasound to the ear could be very different if the body is submerged (see (v), below).

(ii) We do not measure ‘ultrasound’ itself, but rather monitor the effects it produces (see (i)). The closest we get to monitoring ‘ultrasound’ is perhaps optically measuring the vibration of a fixed pellicle in a field (Bickley et al., 2004), which is not dissimilar to the processes involved in hearing. Even here, however, what we actually measure is some effect of ultrasound. The nature of this effect often reflects the interests of the user: to give three examples, photographers use Schlieren photography to monitor ultrasonically induced changes in refractive index (Kudo et al., 2004); chemists commonly use calorimetry; electrical engineers estimate the acoustical energy input through knowledge of the electrical measurements and estimations of transfer efficiencies. The latter two illustrate the problems of this approach well, and indeed compared poorly in a controlled comparison of the two when measuring the same field (Leighton et al., 2005a). Both have features in common. First, they use indirect measurements (temperature rise and electrical power, respectively) and make certain assumptions regarding the efficiencies with which that power is converted from/to (respectively) acoustical power. Second, both give only a spatial average (and for calorimetry, a temporal average) measure. Third, both have been accepted as satisfactory, indeed the best, methods of estimating the acoustical power for many years by different groups of users. For any system of ultrasonic measurement to be precise, the mechanism by which the measured effect occurs, and how it relates to the acoustic parameters, must be understood. However a disturbing feature of ultrasonics is that this mechanism is not always even identified. For example, the heating measured in calorimetry could come about through ultrasonic absorption in the liquid, or in the container walls, or through direct conduction of heat from the transducer, whose temperature can rise significantly during operation. Hence the occurrence and use of the term “ultrasonic exposure” should be subjected to

\(^{30}\)Although it is possible to use such devices to generate cavitation in water, the ability of them to generate cavitation in tissue is generally less (Leighton et al., 1990).
rigorous scientific examination of the underlying measurement method, and the use of the term “ultrasonic dose” should be treated with scientific scepticism (see Section 2.3). There is no clear understanding of the dosimetric unit in ultrasound. Quite apart from the difficulties in measuring ultrasonic fields, cited above, and the problems with the in vivo environment, this is in part because there is a wide range of mechanisms by which ultrasound can affect tissue. Because of this, there is no simple measure which can be used, for example in bioeffect studies. In particular this has implications for the ease with which epidemiological studies of foetal scanning can be interpreted (Ziskin and Pittiti, 1988): there is no quantitative ‘dose’ against which effects can be correlated. This difficulty is further compounded if no record is kept of even the duration, or of nominal (manufacturer’s) power setting, for the scan. Currently the duration and ‘output intensity’ for each foetal scan are not recorded.

(iii) Currently the two most useful effects exploited to measure ‘ultrasound’ are the piezoelectric development of charge in a hydrophone, and the measurement of radiation force on a target (ranging from precise local measurements of pressure (Bickley et al., 2004) to spatial average measures for checking the output of physiotherapeutic devices (Zeqiri et al., 2004)). Calibrated systems based on one or other should certainly be used to monitor current sound fields in most cases (the only viable exceptions being where the probes are too invasive). From marine zoologists to cell biologists, from chemists to food scientists, in the vast majority of cases the measurement and understanding of the acoustic field comes second to characterisation of the biological and chemical conditions of the experiment. Hence we find many papers where there are paragraphs devoted to listing the provenance and purity of the purchased chemicals or cells, but where these items are then exposed to a sound field about which nothing is quoted other than the frequency. This makes it impossible to repeat the experiment and very difficult to interpret the meticulously recorded biological or chemical effects. The tendency to quote only the on-screen MI (or a peak negative pressure which is derived solely from the on-screen MI) with in vitro fields is nonsensical at best, and in many cases misleading. The MI is popularly used to characterise fields involving contrast agents, even though their presence reduces the validity of the MI in representing the conditions. However such poor reporting is prevalent because an MI value is available on-screen when clinical instruments are, for example, used as sources in in vitro experiments.

(iv) Even if hydrophones are used to monitor the field, it is important fully to appreciate the ability of the environment to change the sound field through, for example, nonlinear propagation, or through reflection and diffraction. The importance of considering the interaction between the hydrophone, the ultrasonic field, and the environment is in part a reflection of the small wavelengths involved, such that mm-sized hydrophones and cm-sized vessels can represent large scattering or diffracting targets to an ultrasound field, and the measurement can be complicated by directionality, spatial averaging etc. As a result, a single hydrophone measurement in a reverberant environment (where the sound field might be highly inhomogeneous—see Figs. 19 and 29) can be misleading. Similarly, the ability of the sound field to affect the environment can lead to a cycle where the environment and the sound field continually change each other. For example, in Fig. 19 a <1% change in the frequency produces a dramatic change in the sound field and its subsequent luminescent effect, because of the tuning of the modes of the vessel. This same effect could be produced if the vessel were used at a fixed frequency, but the liquid temperature changed by 2.5°C (as might occur during the course of an experiment, or even be used as the source for a calorimetry measurement) (Fig. 12). The generation of a bubble population can change the sound speed by much more (Fig. 19), and that population can in turn be affected by the sound field, which it then changes (e.g. by scattering, introducing impedance mismatches, or refraction—Figs. 11 and 25).

(v) The acoustic reflections which are important to the generation of reverberant fields in (iv) are a specific example of the importance of the acoustic impedance matches and mismatches that can occur (Table 1). The sensitivity of the ear mentioned in (i) is in part an exercise in overcoming the impedance mismatch between the air and the cochlear. As an example, consideration of the submerged body introduced in (i) must determine whether the ear canal is air- or fluid-filled, and whether other conduction routes become important (such as from surrounding liquid to cochlea via skull, jawbone, etc.—see Section 6.2). The degree of protection afforded by the acoustic impedance mismatch introduced by partial covering by a wet-suit or a dry-suit should be considered (Fig. 34). In particular, the level of ultrasound at the tissue in question will dramatically increase as the impedance conditions change, as indeed may the tissue of
interest. Consider the differences that occur when an ultrasonic dental scaler is in contact with the patient’s tooth, compared to the conditions when it operates in air. The issue to consider when a human is close to (but not touching) a 30 kHz ultrasonic cleaning bath will almost wholly be a question of the effect of the in-air conduction path to, and effect on, the ear. However if he/she touches the transducer or liquid, the issue may change to one of cavitation in the tissue of the finger.

(vi) When ultrasound is passed through liquid or liquid-like media, bubbles represent the most potent natural sources for interfering with the sound field: if you use ultrasound in liquids and do not think you...
have a problem with bubbles, you probably do not completely understand the problem. The tensile strength of water has never been experimentally measured, because even with filtering, degassing and shielding in place to remove bubble nucleation by cosmic rays, a small number of microbubbles have been present which fail at lower tensions than would the water–water molecular bond. Even a low amplitude sound field can be refracted, absorbed and scattered by sparse populations of bubbles. High amplitude fields can generate cavitation. Bubbles concentrate the energy of the acoustic wave, and focus the timescales and lengthscales over which effects are observed (see the introduction to Section 4). If acoustic fields are difficult to measure directly (see Section 5), cavitation is very much more difficult. Recall that we do not measure cavitation itself, but the effects of cavitation: Biologists might measure haemolysis; Chemists chose a wide range of reactions (Leighton et al., 2005a); Industry favours erosion tests which are very difficult to standardise; Physicists sometimes prefer luminescence. Whilst each of these measurement technologies has its supporters who claim that one technique or another is ideal, only one cross-comparison has been performed to identify the strengths and weaknesses of each method (Leighton et al., 2005a).

(vii) The perception that acoustics in general, and ultrasound in particular, are uncomplicated disciplines with established technologies, has produced numerous errors both in practice and in the literature. This paper provided example scenarios of how such misconceptions arise. These included casual and misleading use of the dB scale (Sections 2.3 and 6.2); the ambiguities inherent in a simple source-to-receiver transmission sensor (Fig. 31); the way in which the frequency of interest might not be the frequency at which the source is driven (Section 3; footnote 24), such that energy may become ‘invisible’ to the detector and saturation might occur (Section 3.2); the complexities inherent in the frequency transfer function of the ultrasonic system as a whole (caption to Fig. 19 and (iv), above; Figs. 25 and 26); the prevalence of self-interaction effects (Fig. 11); the importance of identifying the transmission paths and all tissues of interest (Section 6; Fig. 34; (v), above); the directionality of sensors and the effect of the environment on the sound field (Section 5.2). As a result, ultrasonic sensors can rarely be used off-the-shelf with the confidence one would apply to sensors of many other radiations; a power ultrasound transducer is not an uncomplicated combination heater-stirrer for a chemical reaction; and the parameters by which we describe the waves (intensity, pressure, frequency and the decibel) are not a simple and foolproof.

(viii) As regards the use of ultrasound in liquid and tissue, there are many established applications, and much work has been done on providing international guidelines for safe use. Whilst the mechanisms by which ultrasound can interact with tissue are very many, there is a considerable body of research on these to aid our understanding of them. In contrast, the use of ultrasound in air has had relatively few applications which do not involve a subjective human or animal response (in part because of the attenuation issues discussed in Section 2). Historically the assessment of safety guidelines for ultrasound in air is a much smaller enterprise than that undertaken for the assessment of the safety of foetal ultrasonic scanning. There has recently been an increase in available products, primarily based upon the evocation of a subjective human or animal response. Those which exploit the nonlinearity to generate audiofrequency signals from an in-air equivalent of parametric sonar should be critically assessed, given that the inefficiency of the conversion requires high signal levels in the primary beams. Another category exploits the discomforting effects of in-air ultrasound (to pests for whom it is within their audible frequency range, or to humans for whom it is not, but who can experience unpleasant subjective effects and, potentially, shifts in the hearing threshold). Commercial products are advertised with cited levels which cannot be critically accepted, given that there is a lack of traceability (Sections 2.3 and 6.4) for measurements of ultrasound in air, and little understanding of the mechanism by which they may represent a hazard.

Acknowledgements

The author is grateful to his student Daniel C. Finfer for reading this manuscript and checking the author’s calculations, and providing helpful comments and suggestions (particular for his assistance with the material associated with Fig. 3); to Andrew Hurrell and Geun Tae Yim for providing data for the figures where their
Contributions are acknowledged, and to Martyn Hill, Bajram Zeqiri, Alan McAlpine and Ben Lineton for useful discussions.

References


