Ultrasound Thyroid Elastography as a Pre-FNA Screening Tool: Algorithms and Clinical Evaluation

Si Luo

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Yongmin Kim, Chair

Linda Shapiro

Satoshi Minoshima

Program Authorized to Offer Degree:

Electrical Engineering

University of Washington

Abstract

Ultrasound thyroid elastography as a pre-FNA screening tool:
algorithms and clinical evaluation

Si Luo

Chair of the Supervisory Committee:
Professor Yongmin Kim
Departments of Electrical Engineering and Bioengineering

Thyroid nodules are a common medical problem, with studies reporting as high as 50% of the population having a thyroid nodule at autopsy. Although the majority of thyroid nodules are benign, it is clinically important to diagnose the small malignant population from the rest of the asymptomatic benign nodules. A fine needle aspiration (FNA) biopsy is used to evaluate a thyroid nodule's malignancy and determine whether a surgical removal is warranted. It is estimated that somewhere between 250,000 and 300,000 thyroid FNA biopsies are performed annually in the United States. However, a large percentage (approximately 70%) of these biopsies turn out to be benign. Thus, considering the increasing number of thyroid nodules being detected and the vast number of benign nodules undergoing FNA biopsies, the challenge lies in judiciously deciding which nodules should be aspirated.

Ultrasound (US) elastography measures the deformation of tissue in response to stress and derives and displays its stiffness. Recent studies demonstrated the potential of applying elastography to the thyroid gland in noninvasively differentiating benign and malignant thyroid nodules. However, lack of reproducibility in data acquisition and scoring has been a major limitation for US elastography using external

compression. In this dissertation, we have utilized the carotid artery pulsation as an intrinsic compression source and developed quantitative scoring methods to reduce the variability due to the externally-applied compression and subjective assessment of elastography images. We have performed studies with human subjects to evaluate the clinical usability (i.e., interobserver agreement and intraobserver reproducibility, and influence of a nodule's parameters on the diagnostic performance) of the developed elastography technique. We have also demonstrated the good diagnostic performance of our elastography method in differentiating benign and malignant thyroid nodules in clinical settings. The elastography technique described in this dissertation has been implemented and available in commercial ultrasound machines. We believe that our work would be of use to clinicians in deciding whether a thyroid nodule should be referred for an FNA biopsy or follow-up with a higher level of confidence, and providing better care at a lower cost for patients with thyroid nodules.

Table of Contents

List of Figures

Figure Number Page

List of Tables

Acknowledgements

First, I would like to thank my advisor, Professor Yongmin Kim, for providing me with this research opportunity and guiding me throughout the entire Ph.D. study. Without his help and support, this dissertation would not have been possible. His passion for excellence in research and his commitment to translating our research into clinical benefits for patients will inspire me for a lifetime.

I am also thankful to my supervisory committee, Dr. Linda Shapiro, Dr. Satoshi Minoshima, Dr. Manjiri Dighe, Dr. Ravi Managuli and Dr. Christopher Li for their interest and time.

I would like to specially thank Dr. Dong-Jun Lim, Dr. Min-Hee Kim and Dr. Sun-Hee Ko in the Seoul St. Mary's Hospital, Korea. Without their commitment and countless hours spent on patient enrollment, data acquisition and analysis, completing this research would not have been possible.

I would also like to thank all the radiologists and sonographers in the Department of Radiology at the University of Washington Medical Center for their help and support.

I would like to acknowledge the work and support from Mr. Dong-Kuk Shin, Mr. Chulan Kim, Mr. Joon-Sik Kim and Mr. Bong-Koo Seo from Samsung Medison Co. Ltd., Korea. Without it, our intrinsic compression elastography method would not have been commercialized.

I thank all the patients who voluntarily participated in the clinical studies performed at the University of Washington Medical Center and Seoul St. Mary's Hospital over several years.

I am grateful to the students at the Image Computing Systems Laboratory for their support and fellowship.

DEDICATION

To my parents, my wife and son.

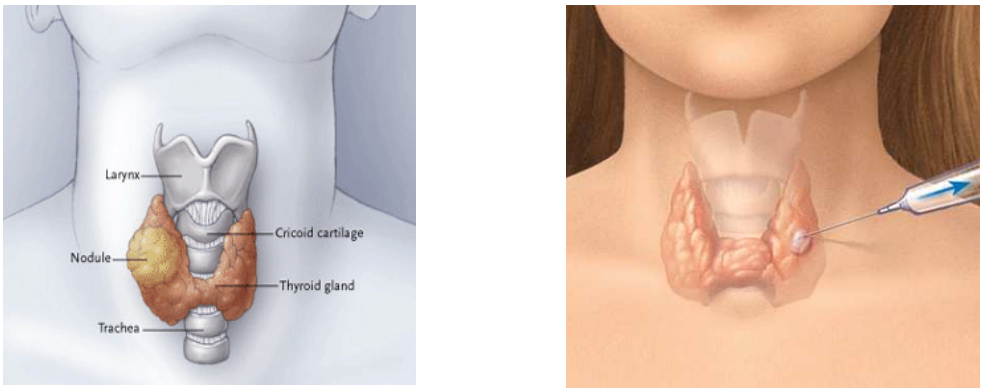
Chapter 1 - Introduction

1.1 Thyroid nodules and fine needle aspiration

A thyroid nodule (Fig. 1.1(a)) is an abnormal growth of cells within the thyroid gland and can be non-cancerous (benign) or cancerous (malignant). Thyroid nodules are a common medical problem, with approximately 8% of adults in the United States having palpable thyroid nodules (Wong and Wheeler 2000; Yeung and Serpell 2008). With the more frequent use of various imaging modalities, the prevalence of thyroid nodules has increased to a range of 20-67% in randomly-selected populations (Mazzaferri 1993; Tan and Gharib 1997; Wang and Crapo 1997; Youserm et al. 1997). On the other hand, most of these asymptomatic thyroid nodules (~90%) are benign (Mazzaferri 1993; Hegedus et al. 2003). The clinical importance of thyroid nodule management rests with the need to detect malignant nodules that occurs in 5%-10% among asymptomatic nodules depending on age, gender, radiation exposure history, family history and other factors (Mazzaferri 1993; Frates et al. 2005; Yeung and Serpell 2008).

Computed tomography (CT), magnetic resonance (MR) imaging, and ultrasound (US) are the most commonly used imaging techniques on thyroid nodules (Weber et al. 2000; Hoang et al. 2007). CT and MR imaging give structural information about the thyroid, identify the location, size and number of nodules and establish the relationship to adjacent structures, such as the carotid artery, jugular vein and trachea. However, the role of CT and MR imaging for thyroid nodule diagnosis is limited because of their inability to distinguish benign and malignant thyroid lesions and the relatively high costs (Youserm et al. 1997). Currently, US is widely used to identify the composition of thyroid nodules, differentiate malignant from benign nodules and guide the fine needle aspiration (FNA) biopsies (Mazzaferri 1993; Frates et al. 2005; Hoang et al. 2007; Fish et al. 2008; Moon et al. 2008). US features that are predictive of malignant nodules include the presence of microcalcifications, hypoechogenicity, irregular margins and intranodular vascularity (Hoang et al. 2007;

Moon et al. 2008). However, no single feature is highly predictive for malignancy (Hoang et al. 2007).



(a) (b)

Figure 1.1. (a) A thyroid nodule and (b) a fine needle aspiration (FNA) biopsy.

Due to indefinite diagnosis with current imaging methods, an FNA biopsy (Fig. 1.1(b)) has been established as a safe and practical method that can differentiate benign and malignant nodules (Gharib and Goellner 1993; Rausch et al. 2001; Cappelli and Rosei 2006; Mahar et al. 2006). To perform an FNA biopsy, typically a 25-gauge needle is placed into the nodule 5~6 times to aspire adequate cell samples, which are then placed on a microscope slide, stained and examined by a cytologist. FNA is currently considered as the most accurate and cost-effective procedure for the preoperative diagnosis of thyroid nodules (Gharib and Goellner 1993; Rausch et al. 2001). The sensitivity of detecting malignancy with FNA ranges from 65% to 98%, while the specificity ranges from 72% to 100% according to various studies (Gharib and Goellner 1993).

Cytodiagnosis from FNA could be one of the following four categories: benign (negative), suspicious (indeterminate), malignant (positive), and nondiagnostic (unsatisfactory) (Lowhagen et al. 1979; Altavilla et al. 1990; Castro and Gharib 2003;

Cappelli et al. 2007). The rate of benign cytological results ranges from 53% to 90% with the average of 69%. Patients with suspicious (indeterminate) cytodiagnosis have specimens showing hypercellularity and patterns suggestive of follicular or Hurthle-cell neoplasms or atypical features suggestive of, but not diagnostic for, malignancy. The rate in this category ranges from 5% to 23% with the average of 10%. Patients with malignant cytological findings (the presence of malignant cells consistent with primary or metastatic thyroid carcinoma) account for 1% to 10% with the average of 4%. In patients with nondiagnostic (unsatisfactory) cytological results, specimens are found to be inadequate for proper cytopathological interpretation, usually due to the presence of cystic fluid or hemorrhagic materials. The rate of nondiagnostic cytological results varies from 2% to 21% with the average of 17%.

Wang and Crapo (1997) estimated that ~300,000 thyroid FNA biopsies are performed annually in the United States. However, a large percentage (approximately 70%) of these biopsies turn out to be benign (Gharib and Goellner 1993; Wang and Crapo 1997). Thus, considering the increasing number of thyroid nodules being detected and the vast number of benign nodules undergoing FNA biopsies, the challenge lies in judiciously deciding which of these nodules should be aspirated.

1.2 Fine needle aspiration biopsy guidelines

Currently, multiple guidelines exist for managing incidentally-discovered and clinically-indicated thyroid nodules for an FNA biopsy (Frates et al. 2005; Gharib et al. 2006; Cooper et al. 2009). According to the American Thyroid Association (ATA) guidelines (Cooper et al. 2009), any thyroid nodule that is 10 mm or greater in diameter should be evaluated via an FNA biopsy, while the Society of Radiologists in Ultrasound (SRU) guidelines strongly recommend an FNA biopsy on solid nodules greater than 10 mm when microcalcifications are present, greater than 15 mm if a nodule is solid or if coarse calcifications are present, and 20 mm or greater if mixed solid and cystic components exist within a nodule (Frates et al. 2005). In addition,

the American Association of Clinical Endocrinologists (AACE) guidelines state an FNA biopsy to be performed on all hypoechoic nodules 10 mm or greater with irregular margins, chaotic intranodular vascular spots, taller-than-wide shape or microcalcifications, all of which have been associated with increased risk, but are not diagnostic of a nodule's malignancy (Gharib et al. 2006).

While these guidelines remain discordant on whether patients with certain nodules (between 10 and 14 mm in diameter) should receive an FNA biopsy or not, all of them rely on a nodule's size and conventional US characteristics to determine the need for an FNA biopsy. However, the predictive value of US features for detecting malignant nodules is typically low, and features characteristic of malignant nodules (e.g., microcalcifications and hypoechogenicity) are also present in benign nodules (Hoang et al. 2007; Moon et al. 2008). This results in a large number of benign nodules undergoing an FNA biopsy. Thus, there exists a clinical need to noninvasively detect benign nodules and reduce the number of FNA biopsies performed on patients with benign nodule(s), which would lead to improving the utilization of FNA in diagnosing malignant nodules while reducing the healthcare cost associated with managing thyroid nodules.

1.3 US elastography

Clinicians often use palpation to detect pathological tissue that is stiffer than its surroundings. Over the last two decades, various methods have been proposed to estimate tissue stiffness (i.e., tissue elasticity) with diagnostic ultrasound (Ophir et al. 1991; O'Donnell et al. 1994; Cespedes et al. 1995; Ophir et al. 1999; Sandrin et al. 2003; Bercoff et al. 2004; Chen et al. 2007). In general, these methods can be categorized into two groups based on how the tissue is stressed. The first category, called elastography or the static method, uses quasi-static compression and estimates resulting tissue strain (Ophir et al. 1991; O'Donnell et al. 1994; Cespedes et al. 1995; Ophir et al. 1999; Bae and Kim 2007). Under the applied stress, stiffer tissues that have a higher elasticity modulus show less strain than softer tissues with a lower

elasticity modulus. The estimated strain is often displayed directly or can be used for the reconstruction of tissue elasticity modulus. The other category is shear wave elasticity imaging, which generates low-frequency shear waves in the target tissue by employing focused acoustic beams (Sandrin et al. 2003; Bercoff et al. 2004). As shear waves travel through the tissue, they are altered by changes in tissue stiffness. When shear waves pass through stiffer tissue, the propagation speed increases. The propagation speed of shear waves is kept track by a high frame rate ultrasound system and then converted to the tissue elasticity (Young's Modulus).

In creating an ultrasound strain image using the quasi-static compression method, ultrasound data are acquired before and after the application of stress. The stress may be applied externally by compressing the tissue with an ultrasound transducer or intrinsically by internal deformation, e.g., the carotid artery pulsation or myocardium movement. Regardless of external or internal compression, one-dimensional strain between two points a and b can be computed by:

$$Strain = \frac{\Delta L_{ab}}{L_{ab}} = \frac{d_a - d_b}{L_{ab}} \tag{1.1}$$

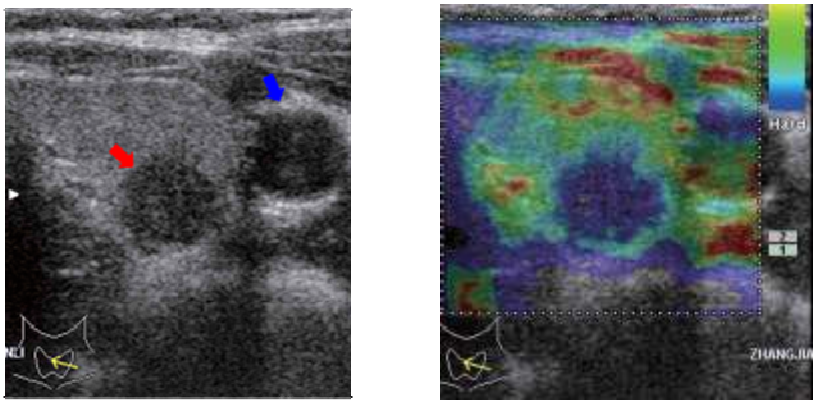
where L_{ab} is the original distance between points a and b and the application of stress changes the distance between them by ΔL_{ab} through their different displacements by d_a and d_b (Ophir et al. 1991). Eq. (1.1) shows that the derivative of displacement can be used to estimate the strain. The crosscorrelation method and the autocorrelation method are widely used to estimate displacement from ultrasound data. The crosscorrelation method is a time-domain technique that estimates displacement by finding the maximum of the crosscorrelation function between the precompression and postcompression echoes (Ophir et al. 1991; Cespedes et al. 1995; Ophir et al. 1999). The crosscorrelation method is generally compute-intensive as it requires an interpolation step for sub-sample accuracy. The autocorrelation method estimates displacement from the phase of the autocorrelation function between the precompression and postcompression ultrasound echoes (Shiina et al. 1996; Yamakawa and Shiina 2001; Bae and Kim 2007). Compared to the

crosscorrelation method, the autocorrelation method has a lower computational requirement while limiting the maximum displacement estimation to one fourth of an ultrasound wavelength to avoid phase aliasing.

To overcome the main disadvantages of the crosscorrelation- and autocorrelation-based methods, Bae and Kim (2007) proposed the angular strain estimation method (ASM), which does not require an interpolation step and can estimate the displacement without being restricted to one fourth of the ultrasound wavelength. This is made possible by using a combination of gross motion estimation and two-stage complex correlation. By computing the complex correlation both temporally and spatially, ASM has been shown to produce more accurate strain (three to five times in terms of signal-to-noise ratio in elastograms) than both the crosscorrelation- and autocorrelation-based methods.

1.4 US elastography in noninvasive diagnosis of thyroid nodules

Figure 1.2(a) shows an ultrasound B-mode image of an invasive papillary thyroid carcinoma, which is indicated by a red arrow. The common carotid artery is pointed by a blue arrow. Figure 1.2(b) shows the corresponding elastography image, which is based on strain computed between a pair of consecutive US frames acquired (one is a precompression frame while the other is a postcompression frame). A rectangular region-of-interest (ROI) in Fig. 1.2(b) indicates the area where the strain is calculated. Within the ROI, each pixel is typically assigned to one of 256 colors, depending on the strain magnitude at that location and an RGB color map used (Lyshchik et al. 2005; Rago et al. 2007). One common color map uses red for high strain (i.e., softer tissues) and blue for low strain (i.e., harder tissues), while green indicates average strain. The whole nodule area in Fig. 1.2(b) is mainly rendered as blue in the elastography image due to the increased stiffness of this papillary carcinoma. To provide both stiffness and anatomical information, the final elastography image is formed by blending the strain ROI with the B-mode image.



(a) (b)

Figure 1.2. (a) Ultrasound B-mode image and (b) the corresponding elastography image of a papillary thyroid carcinoma. An invasive papillary thyroid carcinoma and the carotid artery are pointed by red and blue arrows, respectively.

Various studies demonstrated the potential of US elastography in noninvasively differentiating between benign and malignant thyroid nodules (Lyshchik et al. 2005; Lyshchik et al. 2005; Bae et al. 2007; Rago et al. 2007; Asteria et al. 2008; Dighe et al. 2008; Hong et al. 2009). In an *ex vivo* study, Lyshchik et al. (2005) measured stiffness (elastic modulus) of excised thyroid tissues from 36 patients. They found that papillary carcinomas, the most common thyroid cancer, are 5 times stiffer than normal thyroid tissues and benign thyroid lesions are 1.7 times stiffer than normal thyroid tissues at various compression levels. The results from this study indicate that the stiffness information can be utilized for differential diagnosis of thyroid nodules. As shown in Fig. 1.2(b), the stiffness difference between the normal thyroid tissue and the invasive papillary carcinoma exists and is clearly visualized.

In a follow-up *in vivo* study, Lyshchik et al. (2005) recruited patients who were scheduled for surgical treatment and evaluated the following criteria for elastography: visualization score, relative brightness score, margin regularity score, margin definition score, and tumor area ratio between B-mode and elastography. Their results showed that only two elastographic criteria are significantly associated with

malignancy, i.e., margin regularity and tumor area ratio. These criteria have a high specificity (86 and 92%, respectively), but a low sensitivity (36 and 46%, respectively). On off-line (data processed later on a separate computer) analysis using more sophisticated elastography algorithms, they calculated the strain ratio between a thyroid tumor and its surrounding normal thyroid tissues. Among the various criteria evaluated off-line, only the tumor-to-normal tissue strain ratio was significantly predictive of malignant thyroid nodules ($p < 0.001$) with a sensitivity of 82% and specificity of 96%. On the other hand, they also observed that the quality of strain image and the diagnostic performance were significantly affected by 1) out-of-plane motion caused by the use of external compression (i.e., compressing the thyroid gland using hand-induced transducer motion) and 2) the pulsation of the nearby carotid artery in conjunction with external compression being applied out-of-sync and uncoordinatedly to the thyroid.

In a study by Tranquart et al. (2008), 96 patients with 108 nodules were enrolled for elastography prior to their FNA. They used a scoring scheme similar to that shown in Fig. 1.3 to classify nodules into different grades as follows: 1) grade I: uniformly soft (homogeneously green, Pattern 1); 2) grade II: soft with presence of very soft peripheral zones (green with peripheral red color, Pattern 2); 3) grade III: heterogeneous with soft and hard zones (areas of green and blue color, Pattern 3); 4) grade IV: uniformly hard (mainly blue color, Pattern 4). 95 nodules were soft (grades I and II), and 13 nodules were hard (grades III and IV). No malignant nodule was detected in grade I and II lesions, while 6 malignant nodules were detected in grade III and IV lesions, resulting in a sensitivity of 100% in detecting malignant thyroid nodules.

Rago et al. (2007) evaluated 92 preoperative patients with a single nodule using elastography by assessing an elasticity color scale together with the elasticity score proposed by Itoh et al. (2006). Using these two criteria, a 1-5 scoring scheme was used for grading elastography images. A score of 1 or 2 was assigned in 49 cases, all benign lesions, a score of 3 in 13 cases, one carcinoma and 12 benign lesions,

and a score of 4 or 5 in 30 cases, all carcinomas. By classifying a nodule with a score 4 or 5 as malignant, their method resulted in a sensitivity of 97%, a specificity of 100%, and a positive predictive value of 100%.

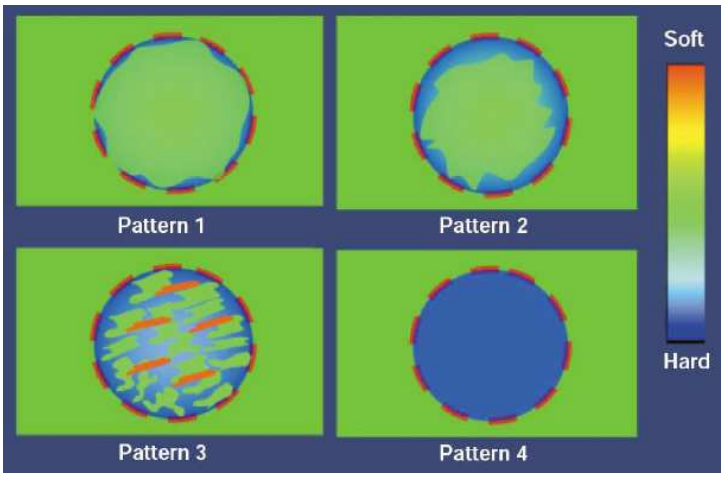


Figure 1.3. Pseudo-color patterns for thyroid nodule classification.

1.5 Variability sources of US elastography

In spite of very good sensitivity and specificity values reported in the literature on thyroid US elastography with external compression, its clinical acceptance has been limited. One major reason is the variability in data acquisition and scoring. The amount of compression used in generating elasticity images is highly operator-dependent. The application of strong compression may cause wrong diagnosis, while soft compression could lead to the decreased contrast in a resulting elasticity image. In detecting malignant nodules by thyroid elastography, the stiffness of a nodule is mostly inferred by visually inspecting the pseudo-color pattern in an elasticity image relative to the surrounding tissues. To make a diagnosis, a clinician needs to categorize the pseudo-color pattern into one of 4~6 different scores

subjectively. Due to the variability in data acquisition and scoring, it requires a high level of experience and training for clinicians to obtain reliable elastography results.

Figure 1.4 shows the detailed steps involved in a thyroid elastography examination using external compression. An imaging plane through a nodule is first selected from US B-mode, after which periodic compression is applied with the transducer for several seconds. Both elastography and B-mode images of the nodule are displayed on the screen in real time while compression is applied. The acquired elastography images are subsequently stored, from which a single elastography image is selected for scoring. Finally, an observer after examining the selected elastography image assigns a score.

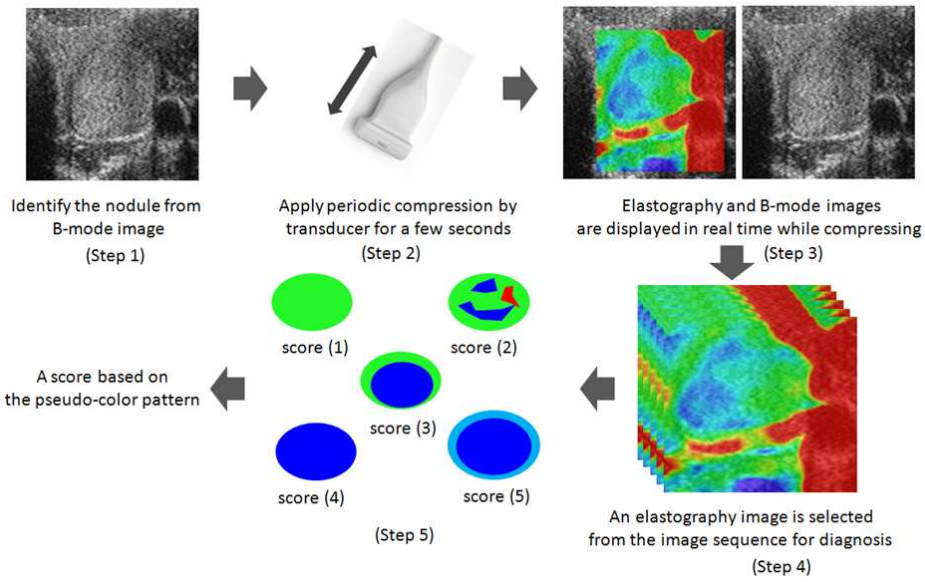


Figure 1.4. Steps in an external compression elastography examination.

There are several places in Fig. 1.4 where variability can be introduced. The first source is in selecting an imaging plane by an operator (step 1). Second, varying compression levels applied by the operator(s) could generate different pseudo-color patterns (step 2). Third, the manual selection of an elastography image from a

dynamic sequence (step 4) adds variability. Another source of variability is due to the subjective nature of assigning a score based on the pseudo-color pattern (step 5). As a result, the reliability with external compression elastography suffers since the pseudo-color pattern could vary depending on an imaging plane selected, the amount of externally-applied compression and manual selection of an elastography image in addition to the variability in score assignment.

Park et al. (2009) evaluated the interobserver agreement of US elastography using external compression. Compared to previous studies (Rago et al. 2007; Hong et al. 2009), where the interobserver agreement was evaluated by using the acquired US elastography images or sequences, they performed the interobserver agreement study where three radiologists independently acquired elastography data and assigned scores. So, their study included all the variability sources shown in Fig. 1.4 in contrast to only one source (step 5 in Fig. 1.4) considered previously (Rago et al. 2007; Hong et al. 2009). They found no statistically significant interobserver agreement for thyroid elastography.

Recently, a diagnostic performance comparison study between elastography and US B-mode features was reported by Moon et al. (2012). A total of 703 solid thyroid nodules from 676 patients were enrolled. Real-time US B-mode and elastography were performed by eight radiologists. It was found that the sensitivity and specificity values of US elastography were 15.7% and 95.3% (using Rago's scoring criterion (Rago et al. 2007)) or 65.4% and 58.2% (using Asteria's scoring criterion (Asteria et al. 2008)), which are substantially lower than those reported in previous studies (Lyshchik et al. 2005; Rago et al. 2007; Asteria et al. 2008; Tranquart et al. 2008; Hong et al. 2009). They found that elastography with and without US B-mode features has inferior performance in differentiating benign and malignant thyroid nodules compared to US B-mode features only. Therefore, they concluded that elastography is not useful as an adjunctive diagnostic tool to traditional US nor as a separate clinical tool. They attributed the inferior performance of US elastography

mainly to the interobserver variability as the elastography examinations were performed by multiple radiologists.

While various studies have shown US elastography has the potential of becoming a clinical tool in noninvasive differentiation between benign and malignant thyroid nodules, substantial variability could be introduced in elastography data acquisition and scoring, which would impair the diagnostic performance of elastography. Thus, to be able to apply US elastography to the management of thyroid nodules in clinical practice, it is critical to minimize the variability in data acquisition and scoring.

1.6 US elastography using intrinsic compression

Bae et al. (2007) came up with an innovative approach in minimizing the variability in data acquisition where the carotid artery is used as an intrinsic compression source, taking advantage of its inherent periodic pulsation (e.g., expansion of carotid artery lumen diameter during systole) and location (adjacent to the thyroid). Using this approach, furthermore, they were able to eliminate the interference caused by two independent compression sources on the thyroid (i.e., external compression and carotid artery pulsation) by not requiring any external compression.

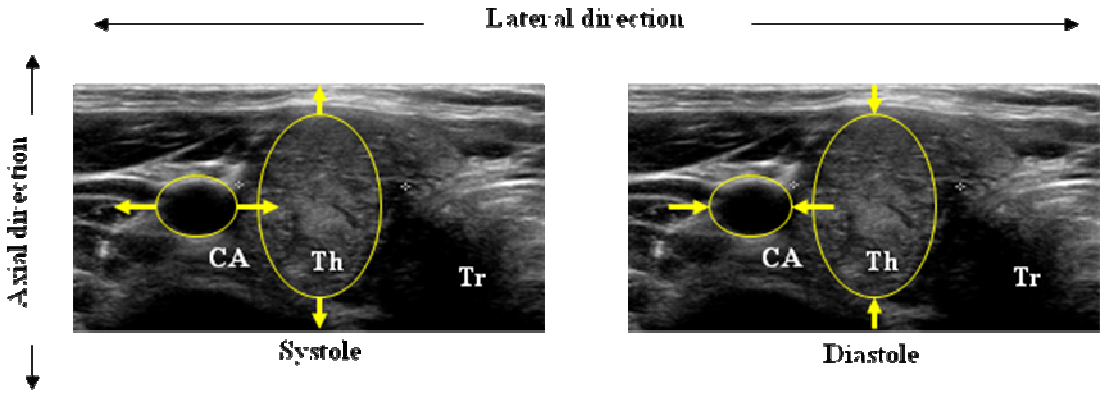


Figure 1.5. Transverse view of the thyroid during systole and diastole when the pulsation of

the carotid artery induces a thyroid's expansion and compression in the axial direction, respectively. The thyroid (Th) is located between the common carotid artery (CCA) and the trachea (Tr).

During systole, the high systolic blood pressure in the carotid artery lumen causes an expansion in the lumen diameter. As the lumen diameter increases, the outer wall of the carotid artery presses against the thyroid, compressing it against the trachea in the lateral direction. Since soft tissues are nearly incompressible, compression in the lateral direction (between the trachea and carotid artery) during systole causes expansion in the axial direction of the thyroid. During diastole, the thyroid expands in the lateral direction when the carotid artery lumen diameter decreases, causing compression of the thyroid in the axial direction. Figure 1.5 illustrates how the thyroid expands and compresses in the axial direction during systole and diastole. By measuring the strain induced by the pulsating carotid artery, Dighe et al. (2008) were able to differentiate between benign and malignant thyroid nodules with a sensitivity of 87.8% and a specificity of 77.5% on 53 nodules, suggesting that differentiation of thyroid nodules is feasible with only the carotid artery as the compression source. Even though we achieved high sensitivity and specificity values in the study, there were several limitations. First, the study population (i.e., 53 nodules) was relatively small. Second, the elastography data were processed off-line, where it took 1.5 hours to process one data set. For routine clinical use, the processing time needs to be reduced significantly. Ideally, the real-time analysis of patients' data is warranted while the patient is still on an examination table. Third, the elasticity scores were manually assigned by a researcher after off-line processing, and no interobserver agreement was evaluated.

In this dissertation, we have continued to use the carotid artery pulsation as an intrinsic compression source for thyroid elastography, because it provides several unique advantages over the traditional external compression method:

I. The external compression method is highly operator-dependent, where the applied force varies from one compression to the next, from one US exam to the next, and among different sonographers. Utilizing the pulsation from the carotid artery can reduce this variability.

II. When external compression is applied, two compression sources (i.e., external freehand compression and carotid artery pulsation) work independently on the thyroid, thus causing error in strain estimation due to both sources compressing the thyroid in an out-of-sync and uncoordinated manner. Using only the carotid artery compression avoids this interference and eliminates resultant noise.

III. Excessive external compression can cause out-of-plane motion and large strain (> 2%) within the thyroid, which can cause signal decorrelation, leading to errors in strain estimation. In the study by Bae et al. (2007), the average correlation coefficient between consecutive frames using intrinsic compression is higher than that of external compression (0.98 vs. 0.92), which indicates that the intrinsic compression would reduce signal decorrelation, thus leading to more accurate estimation of strain.

1.7 Study overview and thesis organization

The goal of our research is to develop a pre-FNA screening tool using intrinsic compression US elastography, which could provide diagnostically useful information (i.e., nodule stiffness) to clinicians with good interobserver and intraobserver agreement in deciding whether a thyroid nodule should go through an FNA biopsy or not.

To achieve this goal, we first addressed the variability in scoring for intrinsic compression elastography. We have developed quantitative scoring methods to reduce the variability and improve the reproducibility in interpreting elastography

images. Due to the difference in stiffness, different frequency characteristics of tissue deformation under the carotid artery compression can be observed in benign and malignant thyroid nodules. The first algorithm extracts the strain rate waveform of the stiffest area within a nodule and converts it to the frequency domain, where linear discriminant analysis is used for nodule classification. In addition to utilizing the temporal strain information (i.e., strain rate waveform), the second quantitative scoring algorithm analyzes the spatial strain contrast within a nodule. We have evaluated the diagnostic performance of these algorithms based on FNA-referred patients' data. Even though very good diagnostic performance was obtained with these algorithms, the results were not quite generalizable in the endocrinology clinics. First, the elastography data were acquired by radiologists and/or sonographers who were very experienced (over 2 years) in performing US elastography examinations. Additionally, the elasticity scores of thyroid nodules were calculated by a researcher off-line in the laboratory (rather than by clinicians in real time), where it took 1.5 hours to process a data set and assign a score. Thus, the results did not reflect much variability (e.g., interobserver variability), typically seen in a clinical setting.

After assessing the diagnostic performance of intrinsic compression elastography in a controlled laboratory setting, the second stage of our study was to fully evaluate the developed elastography method in real clinical practice. To perform clinical evaluation and make the intrinsic compression elastography available for clinicians, we implemented and integrated our algorithms in commercial US machines (Accuvix V20, XG and A30, Samsung Medison Co., Ltd., Seoul, Korea).

To evaluate the variability in data acquisition and scoring with intrinsic compression elastography, we conducted a clinical study to evaluate the interobserver agreement and intraobserver reproducibility of the developed quantitative elastography method, where multiple operators independently performed the US elastography examinations on the same thyroid nodule including both data acquisition and scoring with a commercial US machine. Significant interobserver and intraobserver agreement was observed with intrinsic compression elastography. We

also evaluated the influence of a thyroid nodule's parameters (i.e., size, depth, and distance to the carotid artery) on the diagnostic performance of elastography using intrinsic compression.

Different from the first stage of our study, where radiologists and/or sonographers who were experienced in elastography performed the data acquisition, three endocrinologists who had little experience in US imaging performed the elastography exams. Also, the elasticity scores were interactively calculated on-line by each operator while the patient is still on an examination table rather than by a researcher off-line in the laboratory. Thus, we believe that the results obtained in the second stage would more typically represent the diagnostic performance of intrinsic compression elastography in a clinical setting. Another difference is the size of nodules was smaller than that in the first stage. As these clinical studies were performed in South Korea, where a more aggressive scheme is used in managing thyroid nodules compared to the United States, many small thyroid nodules with the maximum diameter less than 1 cm were enrolled. Since the operators were endocrinologists instead of well-trained radiologists or sonographers, it was difficult for them to hold the transducer still during data acquisition from those small thyroid nodules, which would lead to inconsistencies and variability in the diagnostic performance of elastography.

With good interobserver and intraobserver agreement for intrinsic compression elastography, we have conducted another clinical study with a large patient population to evaluate the diagnostic performance of intrinsic compression elastography in differentiating benign and malignant thyroid nodules in routine clinical practice. While Moon et al. (2012) found external compression elastography has a low sensitivity and specificity in detecting malignant nodules, we have found that the elastography using intrinsic compression could achieve a high sensitivity and good specificity in differentiating benign and malignant thyroid nodules in an endocrinology clinic.

This dissertation is organized into seven chapters. Chapters 2 and 3 present

the quantitative elastography scoring methods using intrinsic compression. The interobserver and intraobserver study is covered in Chapters 4. The influence of a nodule's parameters on the diagnostic performance of intrinsic compression elastography is discussed in Chapter 5. The evaluation of the diagnostic performance of intrinsic compression elastography in real clinical practice is presented in Chapter 6, followed by conclusions in Chapter 7.

Chapter 2 – Thyroid nodules classification using temporal strain information

2.1 Introduction

Previously, the quantitative scoring for thyroid elastography using intrinsic compression was performed by using thyroid stiffness index (TSI) (Bae et al. 2007; Dighe et al. 2008). TSI was computed as the ratio of the highest strain near the carotid artery to the lowest strain within a nodule as indicated in Eq. (2.1).

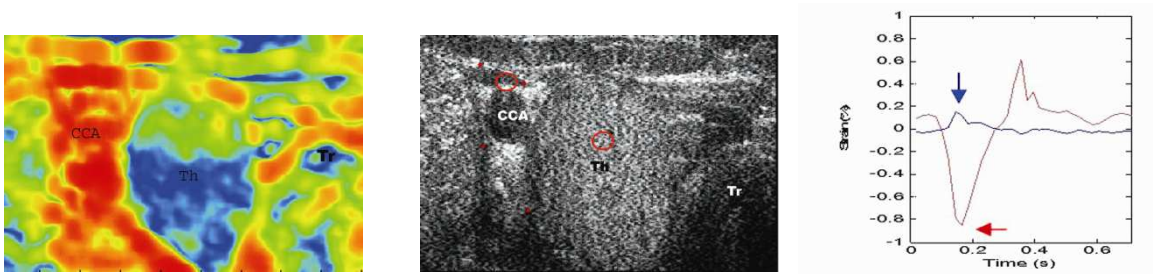
$$TSI = \frac{\text{strain near carotid artery (high strain area)}}{\text{thyroid strain}} \tag{2.1}$$

Due to the increased stiffness (low strain) of a malignant nodule, a higher TSI value would indicate a higher probability of a nodule being malignant. For TSI calculation, all of the strain frames generated (~200) were averaged. From this single averaged frame, the TSI was calculated as the ratio of the strain near the carotid artery to the strain in a nodule. Elastography measures the amount of tissue deformation under applied stress, and in the thyroid, the maximum deformation (strain) occurs during systole when the carotid artery lumen diameter increases maximally because of the high systolic pressure. A better signal-to-noise ratio (SNR) could be achieved at systole due to the increased compression level applied by the lumen during systole (Dighe et al. 2010). By only utilizing strain values at systole, the systolic thyroid stiffness index (STSI) can be defined as:

$$STSI = \frac{\text{strain near carotid artery at systole (highest strain)}}{\text{strain within thyroid nodule at systole (lowest strain)}} \tag{2.2}$$

Figure 2.1 illustrates several steps in computing the STSI value of a thyroid nodule. Figure 2.1(a) shows a typical strain frame during systole. Two ROIs, one near the carotid artery (for the highest strain) and the other within the thyroid nodule

(for the lowest strain), are selected utilizing both the strain (Fig. 2.1(a)) and B-mode (Fig. 2.1(b)) images during systole. The B-mode image is used by an operator to identify the boundary of a nodule and the area near the carotid artery, while the strain image provides the strain distribution information, e.g., within the nodule. The strain value for each ROI is determined by averaging all the strain values in the neighborhood of 2 mm × 2 mm. By repeating this averaging at the same location over multiple strain images, a strain vs. time plot similar to Fig. 2.1(c) is generated. The STSI value is derived by dividing the strain value near the carotid artery at systole (indicated by the red arrow in Fig. 2.1(c)) by the corresponding strain value in the nodule (indicated by the blue arrow in Fig. 2.1(c)).



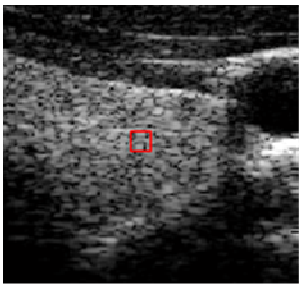
(a) (b) (c)

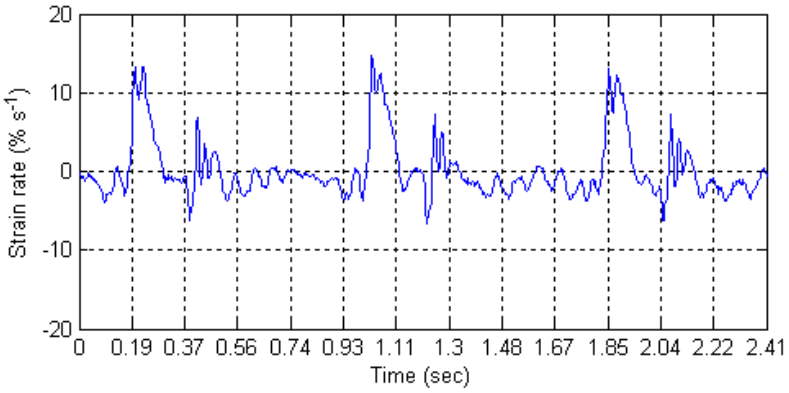
Figure 2.1. The procedure of computing the STSI value of a thyroid nodule: (a) Estimate strain frames from quadrature-demodulated IQ data using angular strain method, (b) place ROIs near the common carotid artery (CCA) wall, where strain is high (area between the red dot), and within a thyroid nodule, where strain is low (the thyroid (Th) is located between the CCA and the trachea (Tr)), and (c) generate the axial strain vs. time plot for each ROI (red corresponds to strain near CCA and blue corresponds to strain within the nodule) and select the peak strain during systole from both strain plots for STSI calculation.

In using the STSI, one limitation is the manual placement of 2 ROIs to estimate the strain near the carotid artery wall and in a thyroid nodule, which may introduce some variability. Because the STSI is calculated as the division of 2 strain values, which are sensitive to noise (e.g., respiratory motion or transducer motion), a small variation in strain value could be amplified, leading to a larger STSI variation. To decrease the variability due to the manual placement of 2 ROIs, we developed a semi-automatic classification method by utilizing the temporal strain information extracted within a thyroid nodule (Luo et al. 2011).

2.2 Strain rate waveform within thyroid

Figure 2.2(b) shows an example thyroid strain rate waveform induced by pulsation of the carotid artery. The data were acquired from a healthy volunteer at 270 frames per second (fps). An ROI (a red rectangle in Fig. 2.2(a)) was first placed within the thyroid gland, and its strain was computed by averaging all the strain values in the neighborhood of 2 mm × 2 mm. By repeating this over multiple strain images, a strain vs. time waveform was produced. By dividing the time interval between two consecutive frames (e.g., 3.7 ms in case of 270 fps), the strain rate waveform was generated as shown in Fig. 2.2(b), which reveals the dynamic deformation of thyroid tissue in response to the compression from the carotid artery pulsation. For example, the thyroid expansion in the axial direction during systole results in positive peaks in Fig. 2.2(b).





(a) (b)

Figure 2.2. (a) Example B-mode image of a thyroid gland and (b) strain rate waveform for an ROI within the thyroid. Positive strain indicates expansion in the axial direction and negative strain indicates axial compression.

2.3 Distinct frequency characteristics of benign and malignant nodules

Figure 2.3 shows the strain rate waveform and its power spectrum for a benign nodule and another set for a malignant nodule. For the strain rate waveforms in Figs. 2.3(a) and 2.3(b), the peaks caused by the systolic blood pressure in the carotid artery

lumen at the heart beat frequency (~1 Hz) are clearly visible. Because of the increased stiffness in a malignant nodule, its peak strain magnitude is smaller than that of a benign nodule. Another difference between benign and malignant nodules is the oscillation in the strain rate waveform. The strain rate waveform of a benign nodule (Fig. 2.3(a)) shows the noticeable oscillation, especially during the diastolic period, while the variation in the strain rate of a malignant nodule is minimum as shown in Fig. 2.3(b). The oscillation frequencies are higher than the heart beat frequency as can be seen in the corresponding power spectrum (Figs. 2.3(c) and 2.3(d)).

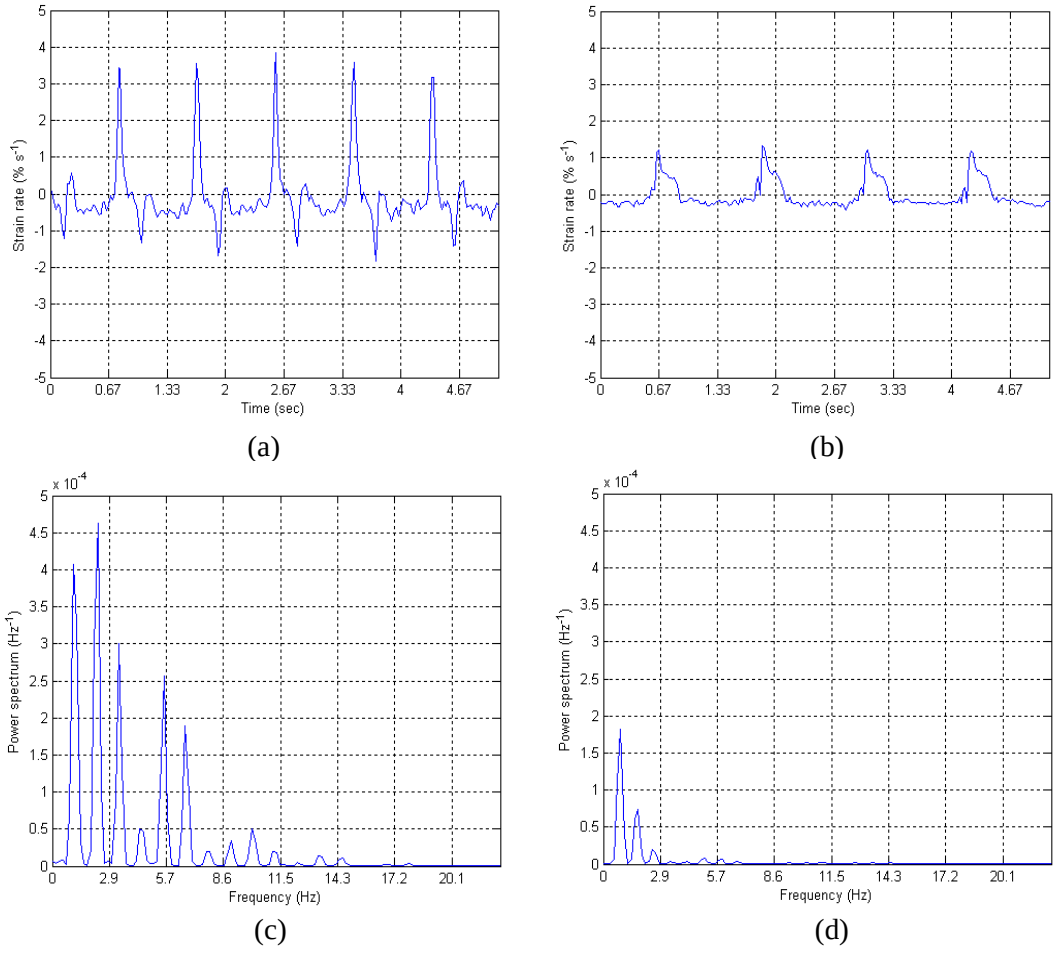


Figure 2.3. Strain rate waveform for (a) a benign nodule and (b) a malignant nodule, (c) the corresponding power spectrum to (a), and (d) the corresponding power spectrum to (b).

Previous studies (Lerner et al. 1990; Parker et al. 1990; Yamakoshi et al. 1990)

found that differences in the power spectrum as shown in Figs. 2.3(c) and 2.3(d) are mainly caused by different stiffness values between benign and malignant nodules. The pulsation from the carotid artery results in the periodic displacement $s(t)$ observed in a thyroid nodule, which can be written as

$$s(t) = s_M \sin(\omega_b t + \phi_b) \tag{2.3}$$

where s_M is the maximum displacement of tissue, ϕ_b the phase and ω_b the angular frequency of the carotid artery pulsation. From the differentiation of Eq. (2.3), the instantaneous velocity of thyroid tissue can be derived as

$$v(t) = s_M \omega_b \cos(\omega_b t + \phi_b) \tag{2.4}$$

Because of the motion of thyroid tissue relative to the ultrasound transducer, the backscattered ultrasound signal is subject to the frequency modulation due to the Doppler effect, which results in an angular frequency shift that can be given by

$$\Delta\omega(t) = \frac{2\omega_0 v(t)\cos\theta}{c} \tag{2.5}$$

where c is the speed of sound, ω_0 the ultrasound carrier angular frequency, θ the angle between the direction of ultrasound wave and thyroid tissue movement. Assuming θ is 0 degree and substituting $v(t)$ using Eq. (2.4), Eq. (2.5) can be written as

$$\Delta\omega(t) = \Delta\omega_M \cos(\omega_b t + \phi_b) \tag{2.6}$$

where

$$\Delta\omega_M = \frac{2\omega_0 s_M \omega_b}{c} \tag{2.7}$$

Based on Eq. (2.6), the phase of the backscattered ultrasound signal can be derived as

$$\theta = \int (\omega_0 + \Delta\omega(t))dt = \omega_0 t + m_f \sin(\omega_0 t + \phi_b) + \phi \tag{2.8}$$

where ϕ is the phase shift due to the wave propagation and m_f is the modulation index that is given by

$$m_f = \frac{2\omega_0 s_M}{c} \tag{2.9}$$

Then, the received backscattered US signal can be written as

$$b(t) = b_0 \sin(\omega_0 t + m_f \sin(\omega_b t + \phi_b) + \phi) \tag{2.10}$$

where b_0 is the maximum amplitude of the backscattered signal. Eq. (2.10) suggests that the phase of the received signal (i.e., $m_f \sin(\omega_b t + \phi_b) + \phi$) is modulated due to the periodic movement of thyroid tissue.

The received signal $b(t)$ is then demodulated by multiplying the following reference signals

$$r_I(t) = r_0 \sin \omega_0 t \tag{2.11}$$

$$r_Q(t) = r_0 \cos \omega_0 t \tag{2.12}$$

where r_0 is the maximum amplitude of reference signals, after which a lowpass filter is applied to eliminate the high frequency (i.e., $2\omega_0$) components from $b(t)r_I(t)$ and $b(t)r_Q(t)$. The demodulated baseband signal can be written as

$$\begin{aligned} I(t) = b_0 r_0 \{ & \cos\phi [J_0(m_f) + 2\sum_{n=1}^{\infty} J_{2n}(m_f) \cos 2n(\omega_b t + \phi_b)] \\ & - \sin\phi [2\sum_{n=0}^{\infty} J_{2n+1}(m_f) \sin(2n+1)(\omega_b t + \phi_b)] \} \end{aligned} \tag{2.13}$$

$$\begin{aligned} Q(t) = b_0 r_0 \{ & \sin\phi [J_0(m_f) + 2\sum_{n=1}^{\infty} J_{2n}(m_f) \cos 2n(\omega_b t + \phi_b)] \\ & - \cos\phi [2\sum_{n=0}^{\infty} J_{2n+1}(m_f) \sin(2n+1)(\omega_b t + \phi_b)] \} \end{aligned} \tag{2.14}$$

where $J_i(x)$ is the *i*th order Bessel function of the first kind. In the derivation of Eqs. (2.13) and (2.14), the following relations are used

$$\cos(z\sin x) = J_0(z) + 2\sum_{n=1}^{\infty} J_{2n}(z)\cos 2nx \tag{2.15}$$

$$\sin(z\sin x) = 2\sum_{n=0}^{\infty} J_{2n+1}(z)\sin(2n+1)x \tag{2.16}$$

Eqs. (2.13) and (2.14) show that the baseband signal has a D.C. component and an infinite number of harmonics that are integral multiples of the carotid artery pulsation frequency ω_b. The magnitude of these frequency components is the function of the modulation index m_f, which can be calculated as

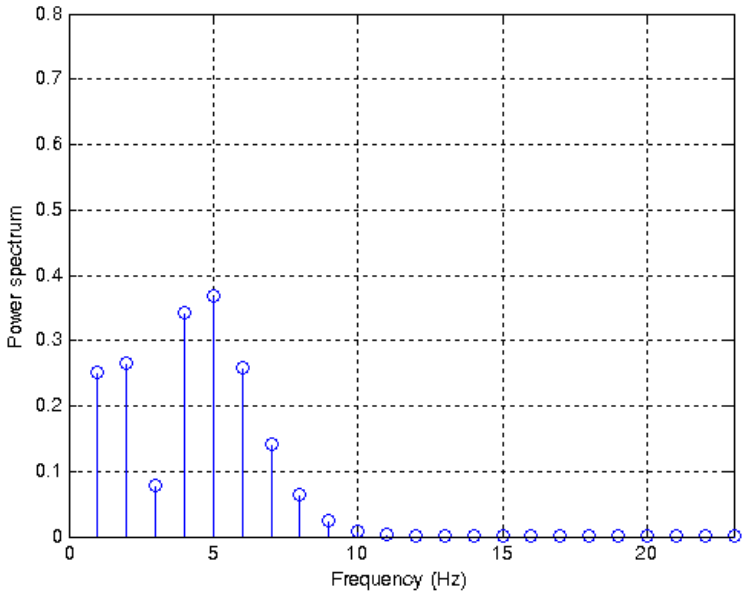
$$A_0 = b_0 r_0 \left| J_0(m_f) \right| \tag{2.17}$$

$$A_i = 2b_0 r_0 \left| J_i(m_f) \right| \quad i = 1, 2, \cdots \tag{2.18}$$

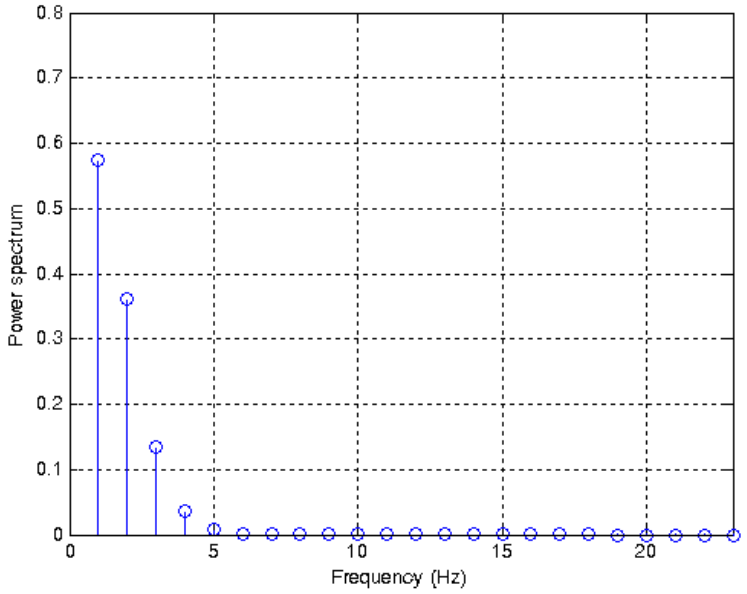
where A_0 is the D.C. component amplitude and A_i is the amplitude of harmonics. The modulation index m_f that is given by Eq. (2.9) is proportional to the maximum displacement (i.e., s_M) of thyroid tissue. A larger displacement of thyroid tissue leads to a larger modulation index m_f. Since malignant thyroid nodules are ~3 times stiffer than benign thyroid nodules, they have a smaller displacement and consequently a smaller modulation index m_f compared to benign nodules. Due to the difference in modulation indices, the power spectrum of benign and malignant thyroid nodules would be different as shown in Figs. 2.3(c) and 2.3(d).

Based on Eqs. (2.17) and (2.18), we can theoretically calculate the power spectrum of benign and malignant thyroid nodules. Since we are mostly interested in the proportional relationship of the magnitudes of harmonics, we assume that for both benign and malignant nodules, $b_0 r_0$ is equal to 1 even though it would be different due to the different maximum amplitudes of backscattered signals. The ultrasound

carrier angular frequency ω_0 was 4.71×10^7 Hz ($2\pi\times7.5$ MHz) for the ultrasound machine used in this study. Assuming the speed of ultrasound c is 1540 m/s, the carotid artery pulsation frequency is 1 Hz and the maximum displacement s_M for a benign thyroid nodule is 0.1 mm, the modulation index m_f can be calculated as 6.12 using Eq. (2.9). For a malignant thyroid nodule that is ~3 times stiffer than a benign nodule, the modulation index m_f is 2.04 (s_M = 0.033 mm). Figures 2.4(a) and 2.4(b) show the power spectrum when m_f = 6.12 and 2.04, respectively.



(a)



(b)

Figure 2.4. Theoretically-calculated power spectrum of (a) benign and (b) malignant thyroid nodules.

As can be seen in Fig. 2.4(a), when the tissue is soft and its displacement is large, a large magnitude can be observed in high frequency harmonics (e.g., 5 to 9 Hz). On the other hand, the magnitude of high frequency harmonics is mostly zero when the tissue is stiff and displacement is small as shown in Fig. 2.4(b). The results are in good agreement with the actual power spectrum observed in benign and malignant nodules as given in Figs. 2.3(c) and 2.3(d), where significantly larger magnitudes are found in high frequency components for a benign thyroid nodule while they are more or less absent in a malignant thyroid nodule.

Because of the increased stiffness in malignant nodules, we can observe the frequency-domain characteristics in tissue deformation that are different from those of benign nodules. We can utilize this difference in performing the thyroid nodule classification. In the following sections, we present how to utilize these frequency characteristics in differentiating benign and malignant nodules.

2.4 Thyroid nodule classification using temporal strain information

Thyroid nodule classification starts with deriving the strain rate waveform and evaluating its frequency characteristics. The flowchart of our classification algorithm is shown in Fig. 2.5. Since a strain rate waveform represents the tissue deformation at a specific location over multiple frames, there are thousands of waveforms to be analyzed depending on the nodule size and parameters used in US scanning. The first step in the classification algorithm is to cluster together the waveforms with similar response to the carotid artery pulsation. We used k-means clustering (MacQueen 1967) to perform this preprocessing with the following steps: 1) the boundary of a thyroid nodule is first delineated in the US B-mode image, and the segmented nodule is considered as a ROI; 2) the strain rate waveforms from the ROI are organized into an n (rows) by p (columns) matrix, where n is the number of pixels or measurements within the ROI and p is the number of elastography frames; 3) k rows are randomly selected from the matrix as the mean of each cluster, and k clusters are created by associating each measurement with the nearest mean based on the squared Euclidean distance; 4) the new mean of each of the k clusters is calculated; and 5) steps 3 and 4 are repeated until convergence is reached.

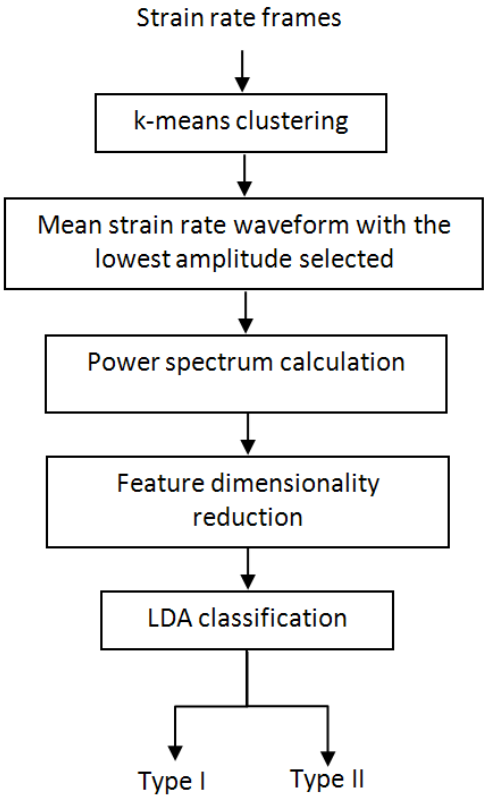


Figure 2.5. Flowchart of our classification algorithm.

Figures 2.6(a) and 2.6(b) show the B-mode image of a thyroid nodule and the corresponding clustering results, where the pixels coded in the same color have the similar response to the carotid artery pulsation. The dashed area in Fig. 2.6(a) indicates a nodule's ROI. In Fig. 2.6(c), the blue line represents the mean strain rate waveform for the region coded in blue in Fig. 2.6(b) while the red line corresponds to the red region in Fig 2.6(b). As can be seen, the mean strain for the blue region is much lower than that for the red region, which indicates that the blue region is stiffer and more suspicious for malignancy. Figure 2.6(d) shows the power spectrum of the mean strain rate waveform corresponding to the blue color region. It represents the thyroid tissue deformation at a specific frequency ranging from 0 Hz to a half of the frame rate and is used as features in classification. If we denote a feature set as Γ with N elements, then a thyroid nodule can be considered as a point Γ in the

N-dimensional feature space. For example, the power spectrum of a thyroid nodule with 128 bins can be considered as a vector of 128 dimensions or equivalently a point in a 128-dimensional space.

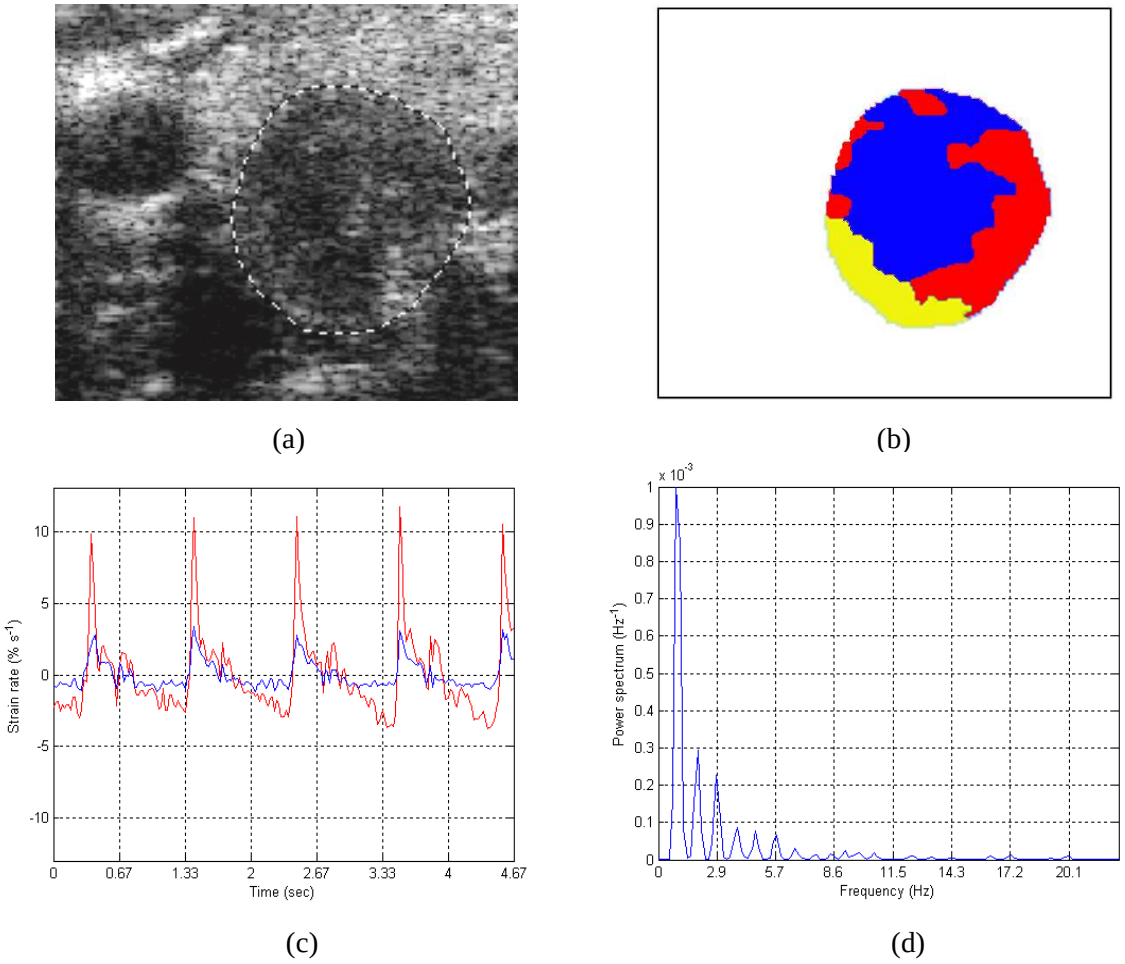


Figure 2.6. (a) B-mode image of a thyroid nodule and (b) the corresponding k-means clustering results, (c) the mean strain rate waveform of the blue and red regions in (b), and (d) the power spectrum of the mean strain rate waveform of the blue region.

To perform classification directly in a high-dimensional space is difficult with the limited number of training cases (e.g., 98 nodules in our study) because many parameters need to be determined in a high-dimensional space compared to the number of parameters needed in a low-dimensional space. Typically, the

high-dimensional feature space is converted into a space with fewer dimensions, where the classification can be performed more efficiently. Principal component analysis (PCA) is a technique used to approximate the original data with low-dimensional feature vectors (Jolliffe 2002). The goal of PCA is to identify a set of orthogonal basis vectors for a new coordinate system to represent the original data set while preserving those characteristics that contribute the most to the data variance. By performing feature dimensionality reduction using PCA, the original N-dimensional feature vector Γ was transformed to a new feature vector Ω with M ($M < N$) dimensions. This feature vector Ω with M dimensions was used for nodule classification via linear discriminant analysis. In this study, we used M = 15 while N = 128.

Linear discriminant analysis (LDA) is a well-established pattern classification method. For a two-class problem, it determines a projection vector W to maximize the between-class scatter matrix while minimizing the within-class scatter matrix in the feature space (Duda et al. 2000). The projected value of a nodule's feature vector Ω along W leads to a discriminant score for that nodule (Chan et al. 1995). Since the feature vector Ω is calculated from the power spectrum of the mean strain rate waveform of the most suspicious region, it represents the stiffness information of the stiffest part of a nodule. Smaller spectral power in high frequency components of the feature vector leads to a larger discriminant score, which suggests the increased likelihood of malignancy.

The k-fold cross-validation method was used to verify the performance of our classification algorithm. Typically, k was set to 10 (Duda et al. 2000). All data sets were partitioned into k subsets, and the cross-validation process was repeated k times. Each time, one of the k subsets was retained as the validation set for testing the classifier, and the remaining k-1 subsets were used for training. The classification performance in term of sensitivity and specificity in all k trials was averaged.

2.5 Patients

92 patients (98 nodules) who were referred for an FNA biopsy at the University of Washington Medical Center following the Society of Radiologists in Ultrasound (SRU) guidelines were recruited for the study. The mean age of patients was 52±13 years (range from 20 to 84, 72 females). The mean nodule size was 2.3 × 1.7 × 1.9 cm, ranging from 0.9 × 0.7 × 0.5 cm to 6.2 × 4.6 × 5.3 cm. US elastography was performed prior to the FNA procedure with a clinical ultrasound machine (Hi Vision 5500, Hitachi Medical Systems America, Twinsburg, OH) with a 7.5-MHz linear array transducer. The data set included 82 benign and 16 malignant nodules. Unless a patient subsequently underwent surgery, FNA results were used as the diagnosis for thyroid nodules. For 19 patients who underwent surgery, the final diagnosis was based on the histopathological examination of the excised thyroid nodule (n=22). One surgery patient had 3 papillary carcinomas, while another surgery patient had 2 benign nodules. Of the 22 excised nodules, 16 were diagnosed as papillary carcinoma, while 6 were diagnosed as benign. The study was approved by the Institutional Review Board of the University of Washington. Before enrollment, informed consent was obtained from each participant.

2.6 Results

Table 2.1 summarizes the mean and standard deviation of discriminant scores given by LDA. The mean score of malignant nodules (1.05±0.09, n=16) is significantly higher than that of benign nodules (0±1.02, n=82) (p = 0.0004). The boxplot distribution of discriminant scores for benign and malignant nodules is shown in Fig. 2.7.

Table 2.1. Discriminant scores for malignant and benign nodules

Parameter	**Malignant** (n=16)	**Benign** (n=82)	***p***
Discriminant score	1.05 ± 0.09	0 ± 1.02	0.0004

Figure 2.8 shows an ROC curve to differentiate between benign and malignant nodules. The area under the ROC curve is 0.88. If we use a discriminant score of 0.86 as a threshold, we can obtain a sensitivity of 100% and specificity of 75.6% in detecting malignant nodules. Any nodule with a discriminant score less than 0.86 is classified as Type I (no FNA, observation-only), while any nodule with the score equal to or greater than 0.86 is classified as Type II (FNA). All of the malignant thyroid nodules were correctly classified, giving the sensitivity of 100% in detecting the malignant nodules. On the other hand, 20 benign thyroid nodules (5 nodular goiters, 11 thyroid adenomas and 4 follicular lesions) were misclassified as Type-II.

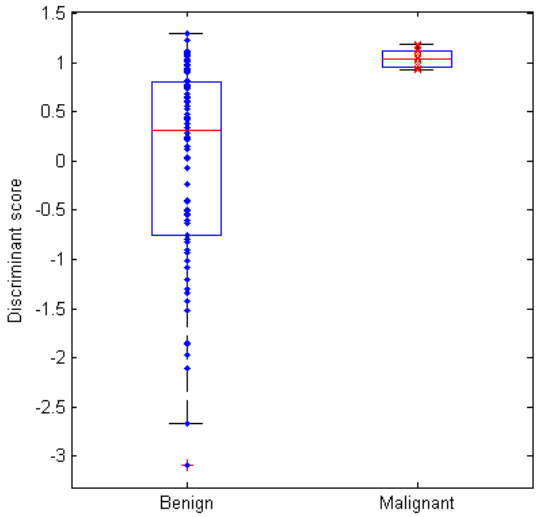


Figure 2.7. Boxplot distribution of discriminant scores for benign nodules (blue dots) and malignant nodules (red x).

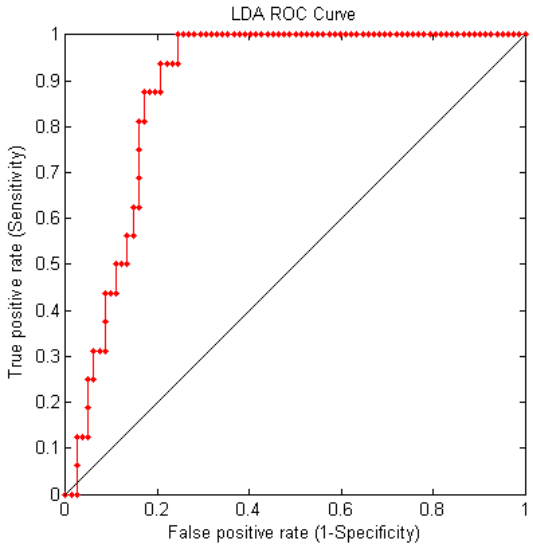


Figure 2.8. ROC curve in differentiating between benign (n=82) and malignant nodules (n=16).

2.7 Discussion and conclusions

In our previous studies (Bae et al. 2007; Dighe et al. 2008), thyroid stiffness index (TSI) was utilized to quantify the stiffness of a nodule. All the strain frames generated (~200) were averaged. By averaging, TSI calculation was simplified, but the temporal information of a nodule's response to the carotid artery pulsation was lost. In this study, we have found that the strain rate waveform of a thyroid nodule is strongly correlated to its stiffness and this temporal information could be used for nodule classification. Instead of directly using the strain rate waveform of a nodule as a feature, the power spectrum of the waveform was calculated and used in classification. By transforming the time-domain signal to the frequency domain, the temporal deformation of a nodule caused by the carotid artery pulsation was represented as a uniform-length vector with the frequency range from 0 Hz to one half of the frame rate. Compared with averaging to get a single stiffness index, the frequency-domain feature can provide richer information as it retains more information on a nodule's response to the carotid artery pulsation, which leads to improved sensitivity and specificity in differentiating benign and malignant nodules

compared to the TSI-based approach.

Compared to other thyroid elastography studies (Rago et al. 2007; Tranquart et al. 2008; Hong et al. 2009), our method can produce more quantitative and repeatable results because the stiffness of a nodule is scored by a classification algorithm rather than by an observer. For quite a few ultrasound elastography studies, the stiffness of a nodule was typically scored using 5 or 6 different grades based on the pseudo-color pattern in an elastography image. Due to the subjective nature of this scoring method, the intra- and interobserver reliabilities need to be evaluated to assess the diagnostic consistency in the same observer and between observers. The kappa coefficient is a commonly-used statistic for this purpose (Viera and Garrett 2005). A kappa coefficient of 1 indicates perfect agreement, whereas a kappa coefficient of 0 indicates chance agreement. In a recent prospective study with 193 breast lesions, Schaefer et al. (2011) reported that the intra- and interobserver reliabilities for ultrasound elastography scoring (excluding data acquisition) using the pseudo-color pattern are 0.720/0.561, where the kappa coefficient of 0.561 corresponds to moderate agreement between observers (Viera and Garrett 2005). In this study, the stiffness of a nodule was calculated and used by a classification algorithm, where there was no need for further interpretation by an observer, leading to more consistent diagnosis.

To diagnose with external compression elastography as shown in Fig. 1.4, an elastography image is first manually selected from an image sequence, where there are dozens of images. Then, a clinician assigns a score based on the pseudo-color pattern in the selected elastography image. Sizable variability would be introduced in selecting the image and assigning the score. In this study, we found that different characteristics in tissue deformation exist between benign and malignant thyroid nodules due to their different stiffness values, which can be utilized for nodule differentiation. As strain rate waveforms exacted from the entire image sequence are utilized, no manual selection of an elastography image is needed with our method. Also, the elasticity score is semi-automatically assigned by a classifier instead of by a clinician. Compared to other thyroid elastography methods, our approach would give

more reproducible results as it substantially reduces the variability in scoring as shown in steps 4 and 5 of Fig. 1.4.

There are some limitations for this study. First, the selection of parameters (e.g., the number of clusters in k-means clustering and points of Fourier transform in calculating the power spectrum) could influence the resultant discriminant scores and introduce the variability to the classification results. The sensitivity of classification results to these parameters needs to be evaluated to establish the robustness of the algorithm. Second, as discussed in Chapter 1, the elasticity scores were calculated off-line by a researcher in the laboratory, where it took 1.5 hours to process a data set and assign a score. For the clinical deployment, the real-time analysis of patients' data is needed. Third, even though we believe that this algorithm can substantially reduce the interobserver and intraobserver variability in performing elastography examinations, no inter- and intraobserver agreement study was performed.

One issue of using LDA or other classifiers for thyroid nodule classification is that the classifier trained using the data sets acquired from one ultrasound machine may not be directly used in other machines. Also, parameter changes during scanning, such as transducer's central frequency and frame rate, could also influence the classification results. In Chapter 3, an improved quantitative scoring method is presented, which utilizes the co-occurrence matrix to analyze the local strain contrast within a nodule and a classifier is not required.

Chapter 3 - Objective elastography scoring using spatiotemporal strain information

3.1 Introduction

In the previous chapter, we presented a method for classifying thyroid nodules based on US elastography using intrinsic compression. By utilizing temporal information from all the acquired elastography frames, the algorithm extracted the strain rate waveform of the stiffest area within a nodule and converted it to the frequency domain, where a uniform-length feature vector with the frequency range from 0 Hz to one half of the frame rate was obtained. Linear discriminant analysis was then used for classification.

In this chapter, we present a new quantitative scoring method for thyroid US elastography (Luo et al. 2012). In addition to utilizing the temporal strain information, the new algorithm analyzes the spatial distribution of strain, from which the local stiffness contrast is computed.

3.2 Materials and methods

3.2.1 Patients

Between April 2009 and December 2009, 113 patients, who were referred for an FNA biopsy at the University of Washington Medical Center following the Society of Radiologists in Ultrasound (SRU) guidelines, were consecutively recruited for the study. A total of 132 nodules were initially included (17 patients had 2 nodules each and one patient had 3 nodules). 9 nodules were later excluded due to inadequate FNA samples (n=3), incomplete elastography data acquisition (n=4), or cystic composition of a nodule (n=2). Thus, our final study population consisted of 106 patients with 123 nodules. The study was approved by the Institutional Review Board of the University of Washington. Before enrollment, informed consent was

obtained from each participant. With the frame rate of 45 frames/s, transverse thyroid US data for ~5 seconds were acquired with intrinsic compression for off-line analysis prior to the FNA procedure with a clinical ultrasound machine (Hi Vision 5500, Hitachi Medical Systems America, Twinsburg, OH) with a 7.5-MHz linear array transducer in a transverse orientation (Dighe et al. 2008). There were 103 benign and 20 malignant nodules. The mean nodule size was 2.3 × 1.8 × 2.1 cm (ranging from 0.5 × 0.6 × 0.6 cm to 4.1 × 2.3 × 3.5 cm). We used the FNA result as the diagnosis for a thyroid nodule unless a patient subsequently underwent surgery. For patients who underwent surgery, the final diagnosis was based on the histopathological examination of the excised thyroid nodule.

3.2.2 Algorithm overview

Figure 3.1 shows the flowchart of our algorithm. The acquired quadrature-demodulated IQ (in-phase and quadrature) data were used for generating strain frames using the angular strain estimation method (Bae and Kim 2007). The strain oscillation index (SOI) at every pixel location within the delineated ROI was calculated from the corresponding strain rate waveform. Placing the SOI values in their proper locations formed a strain oscillation map (SOM), from which a co-occurrence matrix was derived. Finally, the elasticity contrast index (ECI) was calculated from the co-occurrence matrix. A larger ECI value suggests a high probability of a nodule being malignant.

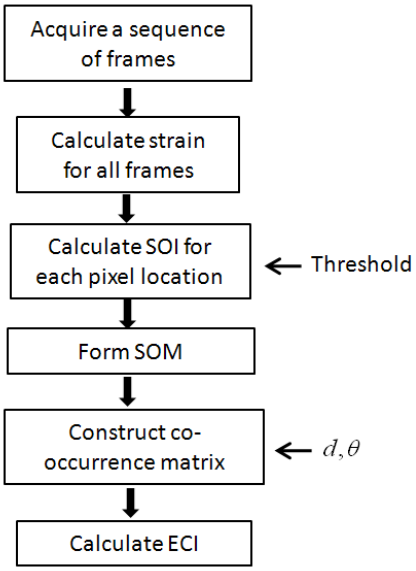


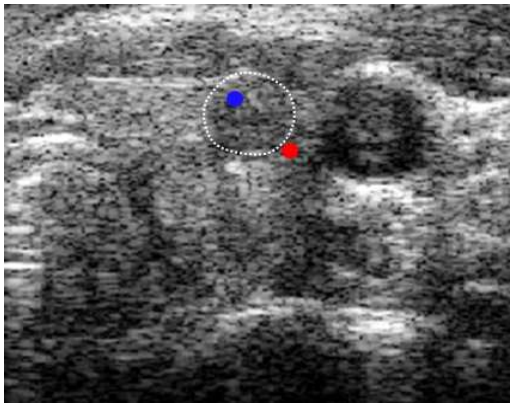
Figure 3.1. Flowchart of our algorithm.

3.2.3 Strain oscillation map

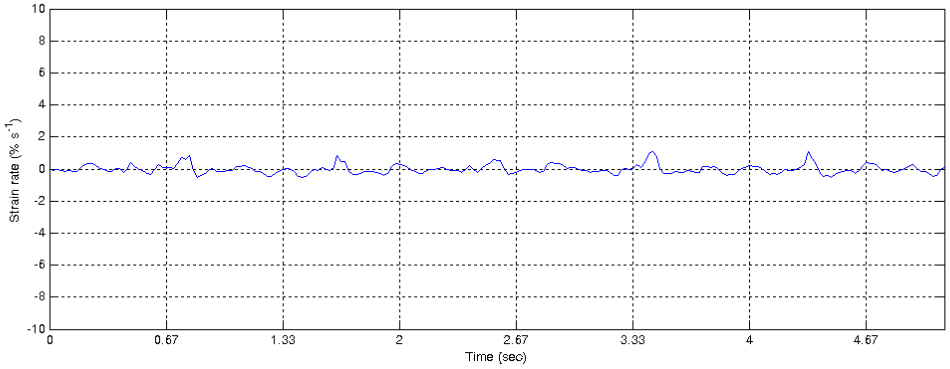
Figure 3.2(a) shows a B-mode image from a small malignant thyroid nodule close to the carotid artery. The nodule's boundary is indicated by white dashed lines. Figures 3.2(b) and 3.2(c) are the strain rate waveform corresponding to the blue and red color regions identified in Fig. 3.2(a), respectively. The disparate strain rate waveforms in Figs. 3.2(b) and 3.2(c) are mostly due to different stiffness values in these blue and red color regions as discussed in the previous chapter. Because of the increased tissue stiffness in the blue region of this malignant nodule, the peak strain magnitude is much smaller than that from the surrounding thyroid tissue (red region). Also, the strain rate waveform in the red color region shows noticeable oscillations, especially during diastole (black ellipse in Fig. 3.2(c)), while that in the blue color region shows minimum oscillations. As a measure of oscillation, we define SOI as follows:

$$SOI(x,y) = \frac{\left\| abs(sr(x,y,n)) \geq threshold \right\|}{N} \times 100\%, \quad n = 0,1,\ldots,N-1 \qquad (3.1)$$

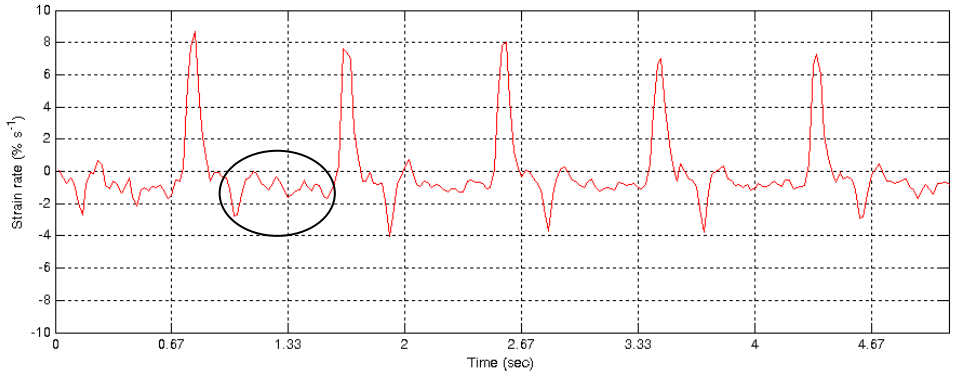
where $abs(z)$ is the absolute value of z, N is the total number of frames in a sequence, $sr(x,y,n)$ represents the strain rate value at a pixel location (x,y) in the nth frame, and $\|\cdot\|$ is the number of frames where the condition inside is satisfied. For each location (x,y), SOI is the ratio of the number of frames where the strain rate value exceeds a given threshold to the total number of frames. The range of SOI is from 0 to 100%. A low SOI value means that a low level of oscillations is present in the strain rate waveform, suggesting stiff tissue. A high SOI value indicates increased strain oscillations, suggesting soft tissue. For the strain rate waveform in Figs. 3.2(b) and 3.2(c), SOI is 12.4% and 84%, respectively, when 1%/s is used as a strain rate threshold in Eq. (3.1).



(a)



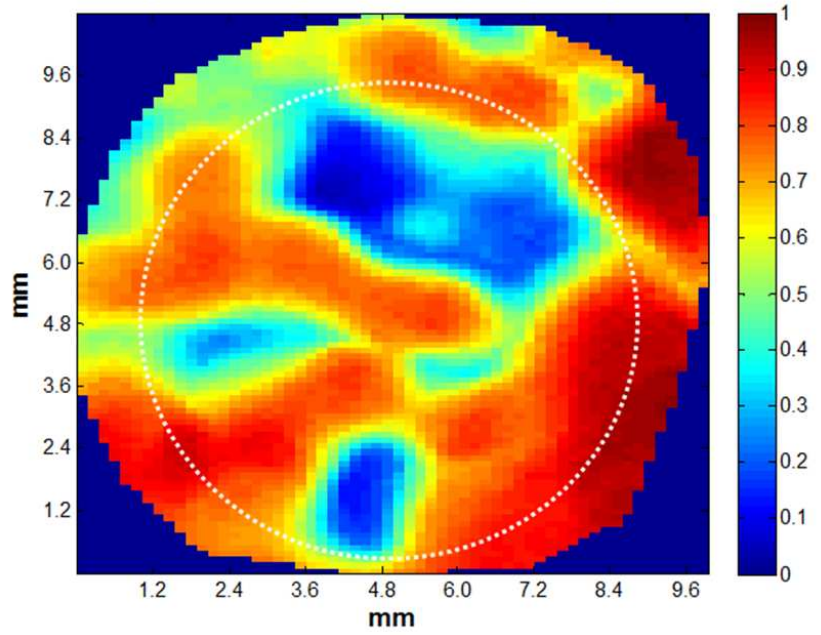
(b)



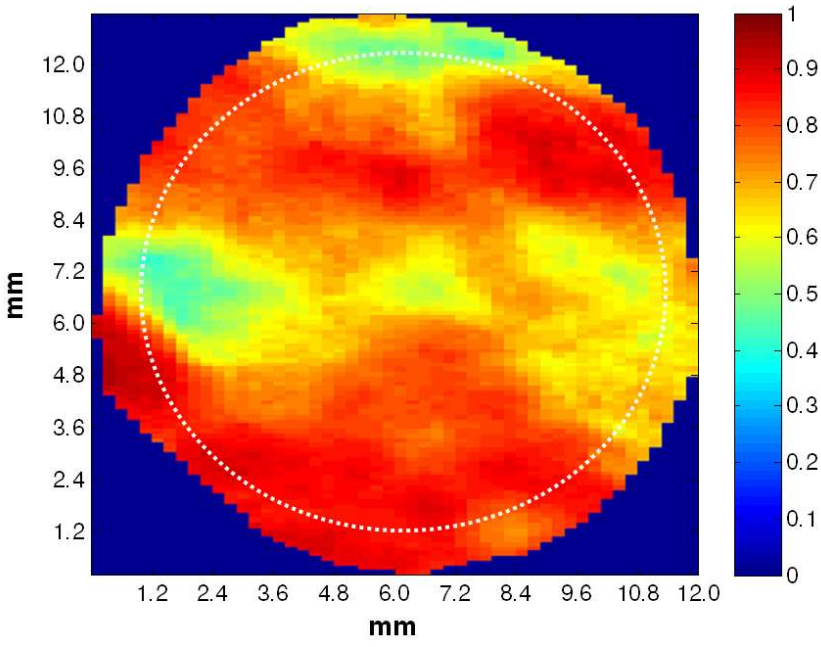
(c)

Figure 3.2. (a) B-mode image of a malignant thyroid nodule, whose boundary is delineated by white dashed line, (b) the strain rate waveform of the blue region and (c) the strain rate waveform of the red region, where the black ellipse in Fig. 3.2(c) highlights strain oscillations during diastole.

By calculating SOI at every pixel location within the white dashed line in Fig. 3.2(a), we were able to condense the entire strain sequence's spatiotemporal information to a single image, which is called a strain oscillation map (SOM). Figure 3.3(a) is the SOM of the small malignant nodule shown in Fig. 3.2(a). As can be seen in Fig. 3.3(a), in the lower-right side of SOM, SOI is as high as 80%, indicating that this portion is soft. At the same time, SOI in the upper part of the map is as low as 10%, suggesting that this area is stiff. On the other hand, Fig. 3.3(b) shows the SOM of a small benign thyroid nodule. A distinct difference between SOMs of the malignant (Fig. 3.3(a)) and benign (Fig. 3.3(b)) nodules is the local contrast. In a malignant nodule, the differences between low and high SOI areas are large, and the transitions from high to low (or low to high) SOIs happen over a short distance, leading to a high local contrast. On the other hand, as the stiffness difference between a benign nodule and normal thyroid tissue is relatively small, a high local contrast is much less likely to occur in a benign nodule as shown in Fig. 3.3(b).



(a)



(b)

Figure 3.3. (a) Strain oscillation map (SOM) corresponding to Fig. 3.2(a), where white dashed lines indicate the nodule boundary, and (b) SOM for a benign thyroid nodule.

3.2.4 Elasticity contrast index

The local contrast is one of the second-order texture features of an image. A co-occurrence matrix has been used in medical image analysis for quantitative measurements of image texture (Haralick et al. 1973; Chan et al. 1998; Mudigonda et al. 2000). Each element in the co-occurrence matrix $C_{d,\theta}(i,j)$ is defined as the number of times (or the probability) where a specific pixel value pair combination (i,j) occurs when two pixels are separated by a Euclidean distance d and along the direction of θ:

$$C_{d,\theta}(i,j) = \left\| x_2 - x_1 = d\cos\theta, y_2 - y_1 = d\sin\theta, I(x_1,y_1) = i, I(x_2,y_2) = j \right\| \quad (3.2)$$

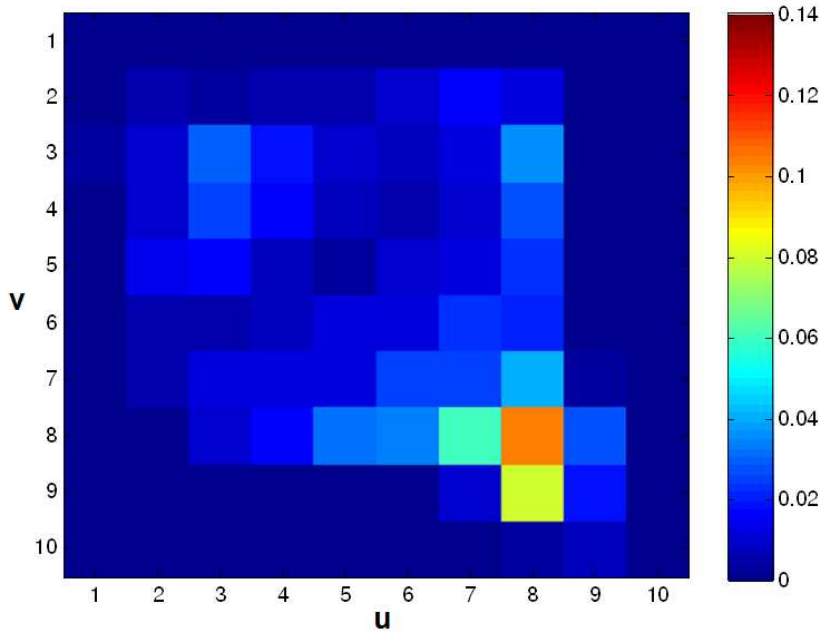
where (x_1,y_1) and (x_2,y_2) are the locations in SOM of the first and second pixels, $I(.)$ is a pixel's SOI value and $\|.\|$ is the number of pixel pairs in SOM satisfying the conditions inside. Note that the dimension of the co-occurrence matrix is $N_g \times N_g$, where N_g is determined by the number of discrete levels used in quantizing SOI values. We used 10 for N_g.

To quantify an SOM's local contrast for a given d and θ combination, we define an elasticity contrast index (ECI) as follows:

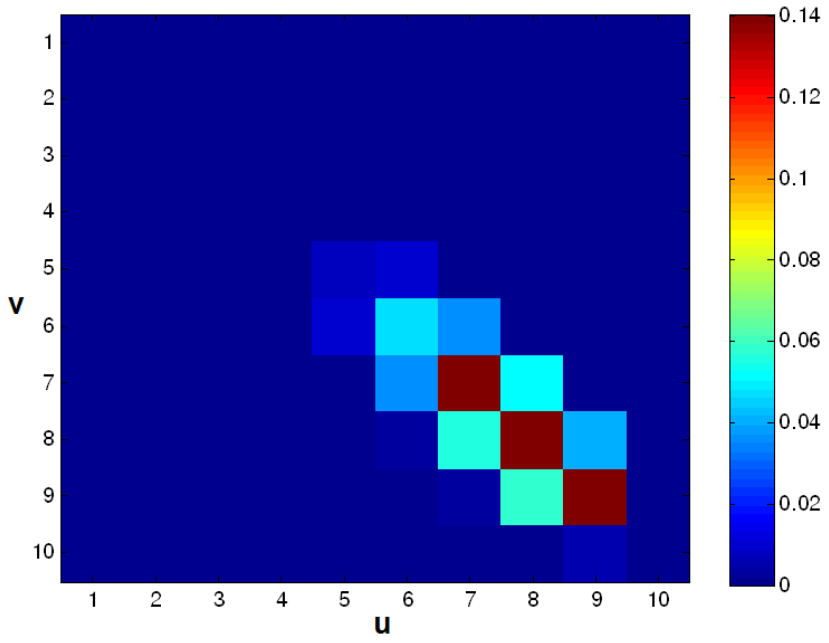
$$ECI = \sum_{u=1}^{10}\sum_{v=1}^{10} |u-v|^2 \, p(u,v), \;\; where \sum_{u=1}^{10}\sum_{v=1}^{10} p(u,v) = 1 \quad (3.3)$$

where $p(u,v)$ is a value at (u,v) in the normalized co-occurrence matrix of 10 × 10, which is the probability derived from $C_{d,\theta}(i,j)$ in Eq. (3.2). Figures 3.4(a) and 3.4(b) show the normalized co-occurrence matrix for the nodules shown in Figs. 3.3(a) and 3.3(b), respectively, when d = 2 pixels and θ = 45 degrees were used. The co-occurrence matrix for the benign nodule (Fig. 3.4(b)) has most of its nonzero elements in the lower-right side, suggesting that the pixel pairs separated by 2 pixels with the angle of 45 degrees have similar SOI values. The elements in the upper-right and lower-left parts of the co-occurrence matrix are all zero because of the

slow SOI transitions in Fig. 3.3(b). As a result, the ECI value calculated with Eq. (3.3) is 0.32. On the other hand, quite a few nonzero elements in the co-occurrence matrix are seen in the upper-right and lower-left parts of Fig. 3.4(a) due to the rapid SOI transitions in Fig. 3.3(a). The ECI value for this nodule is 1.29.



(a)



(b)

Figure 3.4. (a) The co-occurrence matrix for Fig. 3.3(a) where the ECI value is 1.29 and (b) the co-occurrence matrix for Fig. 3.3(b) where the ECI value is 0.32.

3.2.5 Parameter selection

In Fig. 3.1, there are three parameters that could influence the ECI value. The first parameter is the threshold value used in Eq. (3.1) for calculating SOI. Different threshold values would lead to different SOMs. The other two parameters are d and θ used in constructing a co-occurrence matrix from SOM. In this study, we used 4 directions (0, 45, 90 and 135 degrees) and 30 distances (1 to 30 pixels), thus a total of 120 (4 × 30) different co-occurrence matrices were generated for each SOM, leading to 120 ECI values for each threshold value.

To evaluate which parameter combinations (threshold, d and θ) achieve better detection performance, the area under the ROC curve (AUC) in detecting malignant thyroid nodules was used. Figure 3.5 schematically shows the parameter evaluation process. For each threshold value, a total of 120 ECI values corresponding to all the possible d and θ combinations were derived for each

nodule. This was repeated for all 123 thyroid nodules. To evaluate the malignant nodule detection performance with the first ECI value (corresponding to d = 1 pixel and θ = 0 degree), we extracted $ECI_{1,1}$ ~ $ECI_{1,123}$ from all 123 thyroid nodules. Using these 123 ECI values and the results from cytopathology/histopathology exams, the true positive and false positive rates in detecting malignant nodules were determined by varying the ECI cut-off value, from which the ROC curve for d = 1 and θ = 0 was generated and the area under the curve (AUC) was then calculated (van Erkel and Pattynama 1998). This process was repeated 120 times in calculating the AUCs for all the possible d and θ combinations with a given threshold value. A parameter set producing a larger AUC has a better discriminant power in differentiating benign and malignant nodules (Metz 1978). Thus, we selected the threshold, d and θ values that produced the maximum AUC.

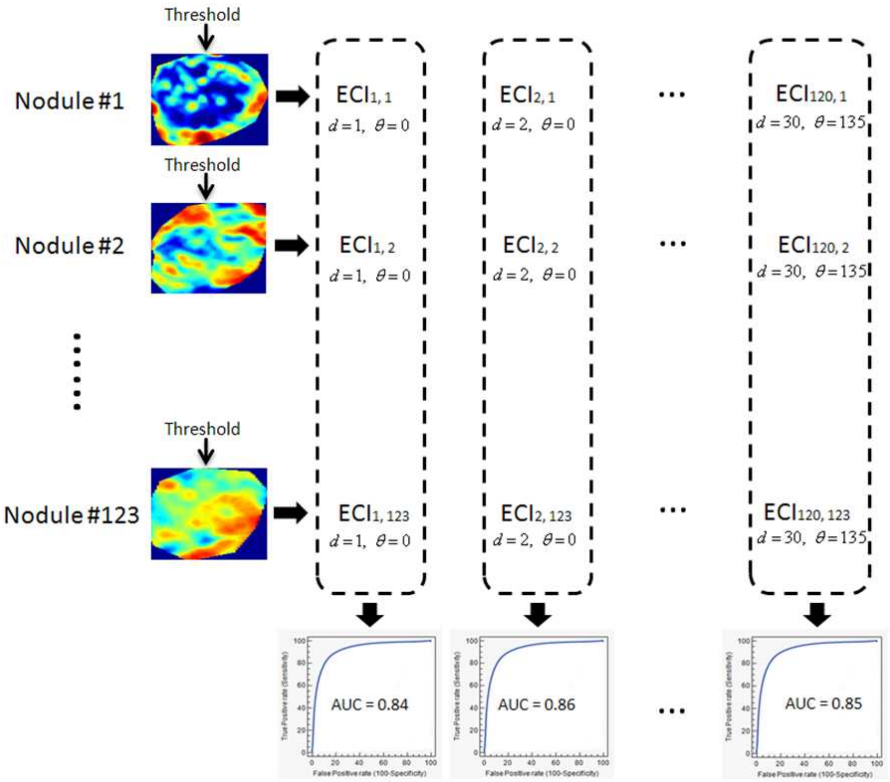


Figure 3.5. Evaluating different parameter sets (threshold, d and θ) based on the area under the curve (AUC) criterion.

3.3 Results

Figure 3.6 shows the AUC for 120 d and θ combinations with different threshold values. The horizontal axis of Fig. 3.6 corresponds to 120 different d and θ combinations, which are arranged by first setting θ = 0 degree and varying d from 1 to 30 then setting θ = 45 degrees and varying d from 1 to 30, and repeating this for θ = 90 and 135 degrees. Four plots in Fig. 3.6 correspond to different threshold values, i.e., 0.7%/s, 1.0%/s, 1.3%/s and 1.6%/s. As shown, the AUC is large when θ = 45 degrees, d = 1~16 and the threshold value is 0.7%/s or 1.0%/s. The average AUC with each threshold value is listed in Table 3.1, and it does not vary much as the threshold value is changed from 0.7%/s to 1.3%/s, which indicates that the performance in detecting malignant nodules is reasonably insensitive to a particular threshold value used in this range.

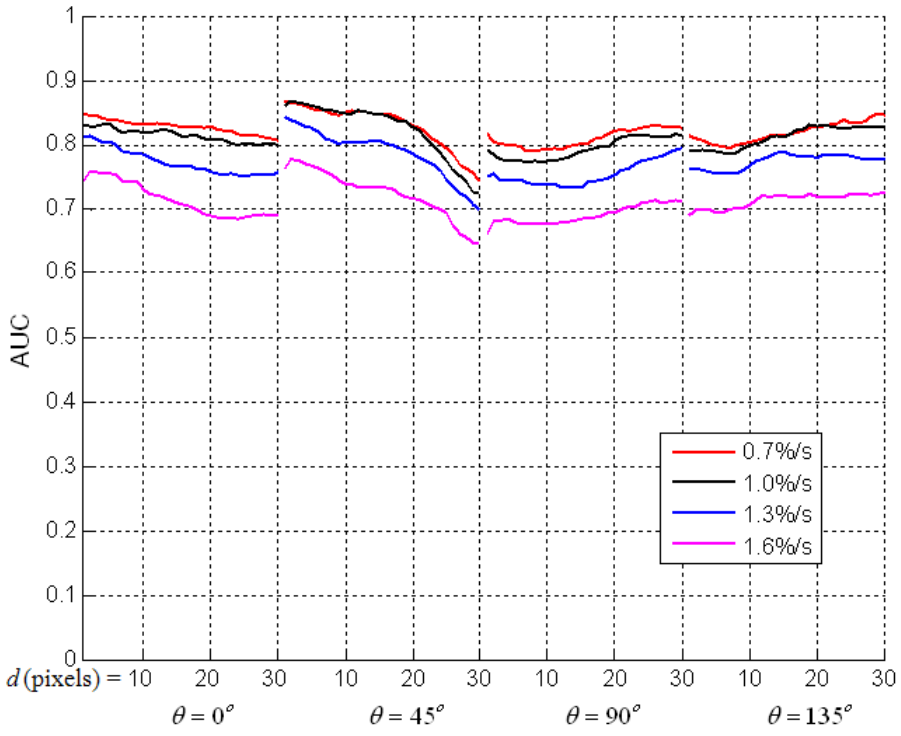


Figure 3.6. The area under the curve (AUC) for different threshold, d and θ combinations.

Table 3.1. Average AUC with different threshold values

Threshold	0.7%/ s	1% /s	1.3% /s	1.6% /s
AUC	0.82	0.81	0.77	0.71

The maximum AUC of 0.87 (95% confidence interval, 0.76~0.97) was obtained with the threshold value of 1.0%/s, d = 2 and θ = 45 degrees. This parameter combination was used in all subsequent analysis. Table 3.2 lists the mean and standard deviation of ECI values with this selected parameter set. The mean ECI value of malignant nodules (0.83±0.23, n=20) is significantly higher than that of benign nodules (0.53±0.18, n=103) (p = 0.00002). The boxplot distribution of ECI values for benign and malignant nodules is shown in Fig. 3.7. Table 3.3 lists the sensitivity, specificity, accuracy and geometric mean of sensitivity and specificity (i.e., $\sqrt{sensitivity \times specificity}$) with different ECI cut-off values. Increasing the cut-off value increases the specificity while the sensitivity decreases. We selected the cut-off value that maximized the geometric mean. Thus, we used an ECI value of 0.60 (corresponding to the geometric mean of 0.837) as a cut-off in malignant/benign decision, which led to a sensitivity of 95%, specificity of 73.8% and accuracy of 77.2%.

Table 3.2. ECI values (d = 2, θ = 45 degrees and threshold = 1.0%/s)

Malignant (n=20)	**Benign** (n=103)	***p***
0.83 ± 0.23	0.53 ± 0.18	0.00002

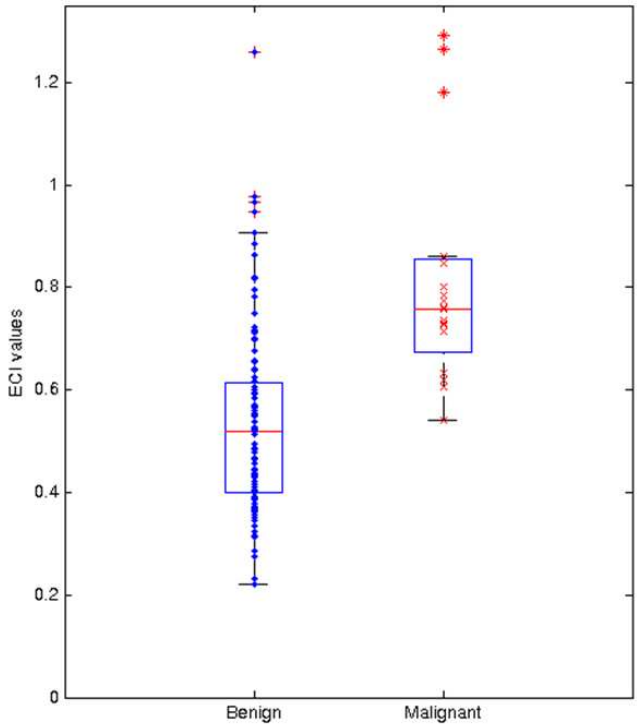


Figure 3.7. Boxplot distribution of ECI values for benign (blue dots) and malignant nodules (red x).

Table 3.3. Detection performance with different ECI cut-off values

Cut-off value	Sensitivity	Specificity	Accuracy	Geometric mean of sensitivity and specificity
0.54	100%	54.4%	61.8%	73.8%
0.55	95%	54.4%	60.9%	71.9%
0.56	95%	61.2%	66.7%	76.2%
0.57	95%	66.8%	70.7%	79.7%
0.58	95%	66.0%	70.7%	79.2%
0.59	95%	68.0%	72.4%	80.4%
0.60	95%	73.8%	77.2%	83.7%
0.61	90%	74.8%	77.2%	82.0%
0.62	85%	75.7%	77.2%	80.2%
0.63	85%	76.7%	78.1%	80.7%
0.64	75%	77.7%	77.2%	76.3%

3.4 Discussion

In Chapter 2, we developed a thyroid nodule classifier utilizing the temporal strain information, which demonstrated the feasibility that the temporal strain information of a nodule induced by the carotid artery pulsation could be used in differentiating benign and malignant thyroid nodules. Compared with other thyroid elastography methods, our classifier-based approach would render more reproducible results. However, one issue in utilizing a classifier for thyroid nodule differentiation is that the classifier trained using the data sets acquired from one ultrasound machine may not be directly used in other machines. For example, the classifier presented in Chapter 2 was trained using 98 data sets acquired from a Hitachi Hi Vision 5500 US machine. If we implement and integrate our algorithm in a Philips US machine, then it is most likely that the classifier needs to be re-trained by acquiring another 98 cases or so with the Philips US machine, which is impractical if not impossible. To improve the generalizability and clinical applicability, the current algorithm utilizes the co-occurrence matrix to analyze the local strain contrast within a nodule, where a classifier is not required. Compared to the previous classifier-based method, this new algorithm can be more readily integrated in commercial US machines for clinical deployment. Another issue with the previous study is that the sensitivity of classification performance to the selection of parameters (e.g., the number of clusters in k-means clustering) was not evaluated. In this study, the area under the ROC curve in detecting malignant nodules was evaluated with different d and θ combinations and threshold values, where the diagnostic performance of the algorithm was found to be reasonably insensitive to a particular d and θ combination and threshold value.

Still, one limitation with the current method is off-line processing of data. For each nodule in this study, it took 1.5 hours to process the elastography data and calculate the ECI score using MATLAB (The MathWorks, Inc., Natick MA) running on a PC with an Intel Core 2 Dual 2.4-GHz processor and 3 Gbytes of memory. As the ECI values of thyroid nodules were calculated off-line by a researcher in the

laboratory rather than by clinicians in real time, the results do not include much variability (e.g., interobserver variability) typically seen in a clinical setting. By collaborating with Samsung Medison, the elastography method presented here was implemented and integrated in commercial US machines (Accuvix V20, XG and A30, Samsung Medison Co., Ltd., Seoul Korea). It now takes ~4 seconds to process one data set, which enables the thyroid elastography to be interactive. In the following chapters, the developed elastography method is evaluated with commercial US machines in real clinical practice.

Utilizing the carotid artery pulsation as an intrinsic compression source and the quantitative scoring method, our intrinsic compression elastography reduces the variability in both data acquisition (step 2 in Fig. 1.4) and scoring (steps 4 and 5 in Fig. 1.4). We believe that our intrinsic compression elastography could render more reproducible results than external compression elastography.

3.5 Conclusion

In this chapter, we presented an improved quantitative scoring method for thyroid US elastography. Compared with other thyroid elastography methods, our method could give more objective and less observer-dependent results, which would be demonstrated in the next chapter. By using our elastography method, we have achieved a sensitivity of 95%, a specificity of 73.8% and an accuracy of 77.2% for detecting malignant nodules in a retrospective study with 123 FNA-bound nodules.

Chapter 4 – Interobserver agreement and intraobserver reproducibility of thyroid elastography

4.1 Introduction

In spite of very good performance that many previous studies reported (Lyshchik et al. 2005; Rago et al. 2007; Asteria et al. 2008; Tranquart et al. 2008; Hong et al. 2009), thyroid US elastography using external compression is not in wide clinical use. One major reason for this is the low interobserver and intraobserver agreement. There are two main sources of variability in US elastography, one in data acquisition and the other in scoring (Yoon et al. 2011). For most previous studies (Rago et al. 2007; Hong et al. 2009), only the variability in scoring was evaluated. The US elastography data sets were acquired by a single operator, and scores were also assigned by the same operator. The second operator only reviewed the already-captured elastography images/sequences and assigned scores.

A study evaluating the interobserver agreement both in data acquisition and scoring for thyroid US elastography was reported by Park et al. (2009). Three radiologists independently performed elastography data acquisition and scoring on 52 surgery-bound thyroid nodules. With external compression elastography, they found no interobserver agreement among three radiologists. They attributed this lack of interobserver agreement in elastography mainly to the fact that the extent of compression influences both the elastography image and, consequently, elasticity score. Their results showed that the variability in acquiring elastography data is larger than that in scoring.

All the thyroid elastography studies mentioned above employed freehand external compression, where periodic force is applied to the thyroid gland by using an US transducer. Utilizing the carotid artery as an intrinsic compression source could

eliminate the variability due to different compression levels applied by the operator(s) (Bae et al. 2007; Dighe et al. 2008). Thus, the objective of this study is to evaluate the interobserver agreement and intraobserver reproducibility in thyroid US elastography using intrinsic compression.

4.2 Materials and methods

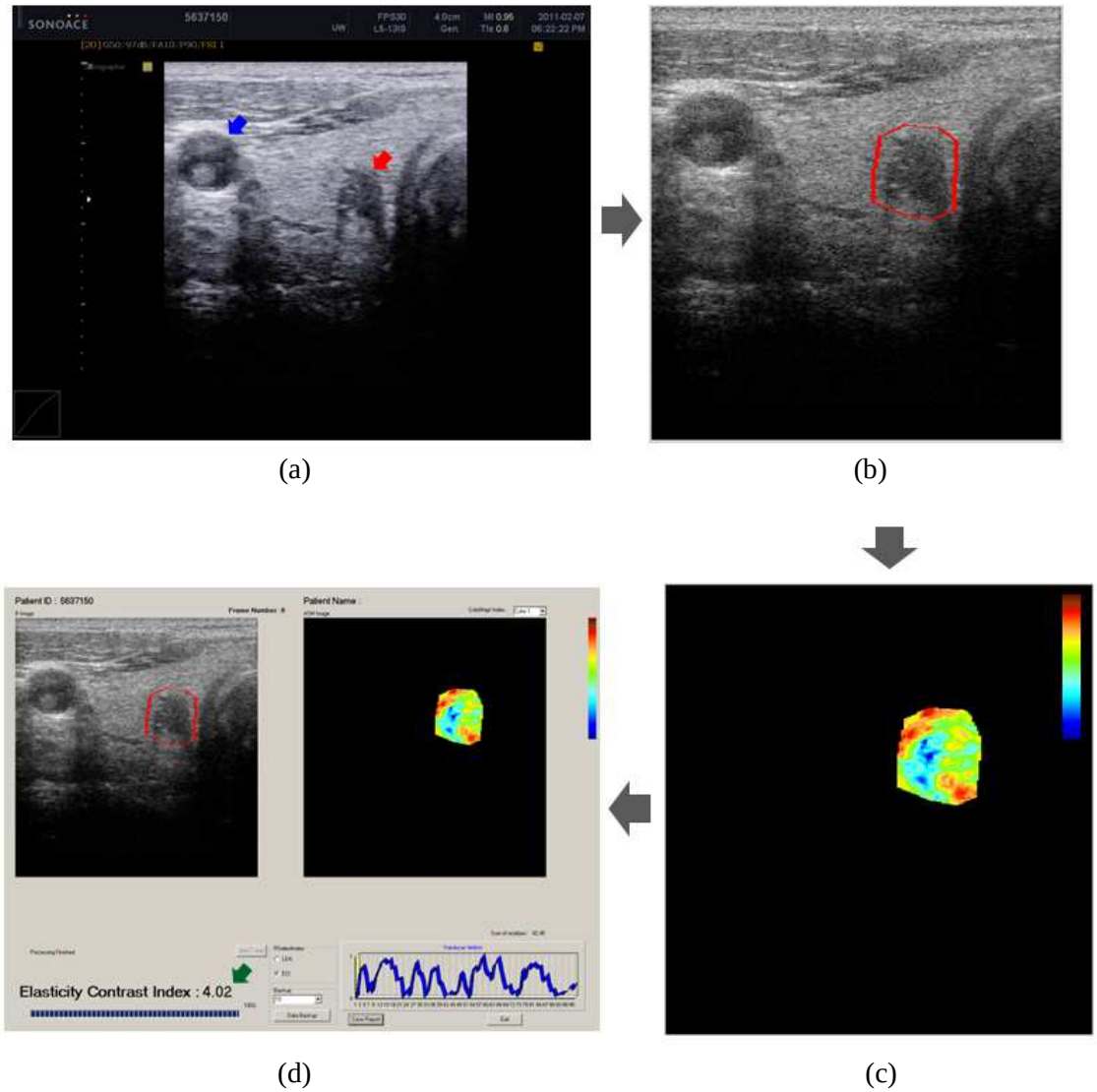
4.2.1 Patients

Between December 2010 and March 2011, 56 surgery-bound patients (14 male and 42 female), whose thyroid nodules had been found via FNA to be malignant, suspicious for malignancy or follicular/Hurthle cell neoplasm, were recruited in the Seoul St. Mary's Hospital for this study. The mean age of patients was 44.1±11.6 years (ranging from 18 to 68 years). US elastography examinations were performed in the evening before the day of thyroidectomy. The final diagnosis was based on the histopathological examination of an excised thyroid nodule. This study was approved by the Institutional Review Board of the Seoul St. Mary's Hospital, and written informed consent was obtained from all participants.

4.2.2 Real-time US elastography examination using intrinsic compression

Ultrasound elastography with intrinsic compression described earlier including ECI calculation was implemented on a commercial US machine (Accuvix V20, Samsung Medison Co. Ltd., Seoul, Korea). Three endocrinologists performed elastography examinations using the US machine with a L5-13 linear transducer. By December 2010, all endocrinologists had less than three months of experience in using US. During a two-month period prior to this study, each endocrinologist practiced thyroid elastography on ~50 nodules to familiarize themselves with data acquisition and ECI scoring. During the study, each endocrinologist performed two separate elastography examinations on each nodule. A total of 6 elastography exams (2 exams per operator × 3 operators) were performed on each nodule within a time span of 30 minutes. The steps used in an elastography examination are illustrated in Fig. 4.1. An operator searched for a transverse plane showing both the common carotid

artery and the largest diameter of a thyroid nodule using B-mode as shown in Fig. 4.1(a). Once the imaging plane was identified, the operator asked the patient to hold his/her breath and acquired the US data for ~4 seconds by placing the transducer over the patient's anterior neck without applying any external compression. The strain frames were generated using the acquired data. After the nodule boundary was delineated by the operator (Fig. 4.1(b)), the SOM (Fig. 4.1(c)) of a nodule and its ECI value were computed and displayed on the screen (Fig. 4.1(d)). A larger ECI value suggests a stiffer nodule, thus indicating an increased likelihood of being malignant.



(a) (b)

(d) (c)

Figure 4.1. Steps used in an elastography examination using intrinsic compression. (a) A

transverse plane showing both the common carotid artery (CCA, pointed by a blue arrow) and the largest diameter of a nodule (pointed by a red arrow) is identified using B-mode and the US data are acquired, (b) an operator delineates the boundary of a nodule using a track ball, (c) the computed SOM of a nodule is displayed, and (d) an ECI value is calculated and displayed on the screen (pointed by a green arrow). The graph on the lower-right corner displays the transducer motion during elastography data acquisition to give immediate feedback to the operator so that the transducer motion can be monitored and/or the operator can repeat the data acquisition if needed.

4.2.3 Statistical analysis

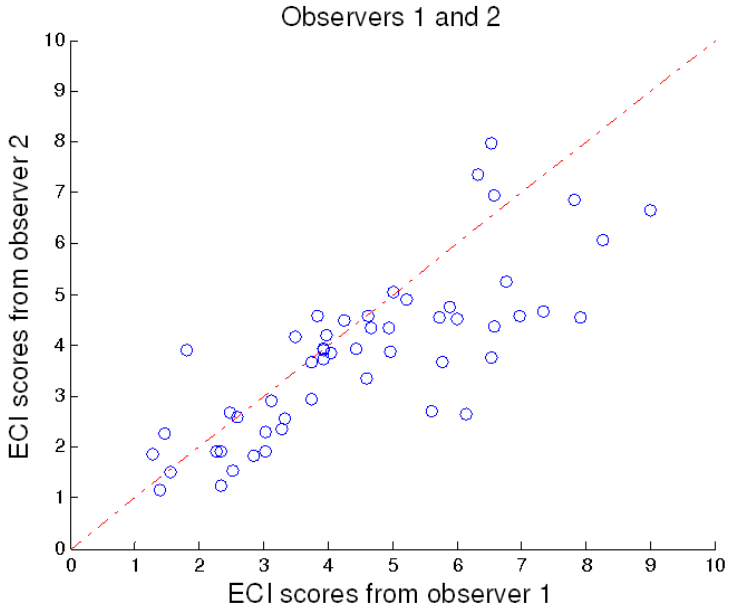
Pearson's correlation coefficient was used to evaluate the interobserver and intraobserver agreement. The collected data were processed with Matlab Statistics Toolbox™ (The MathWorks, Inc., Natick MA). To evaluate the intraobserver reproducibility, Pearson's correlation coefficient was calculated based on two ECI values obtained in two separate elastography examinations by each observer. To evaluate the interobserver agreement, the mean value of two ECI values by each observer was first derived, after which Pearson's correlation coefficients between two observers were calculated. Based on the resultant correlation coefficient value, the interobserver/intraobserver agreement is considered to be poor (0.00-0.20), weak (0.21-0.39), fair to good (0.40-0.75), or excellent (> 0.75) (Fleiss 1986). $p < 0.05$ was considered statistically significant.

4.3 Results

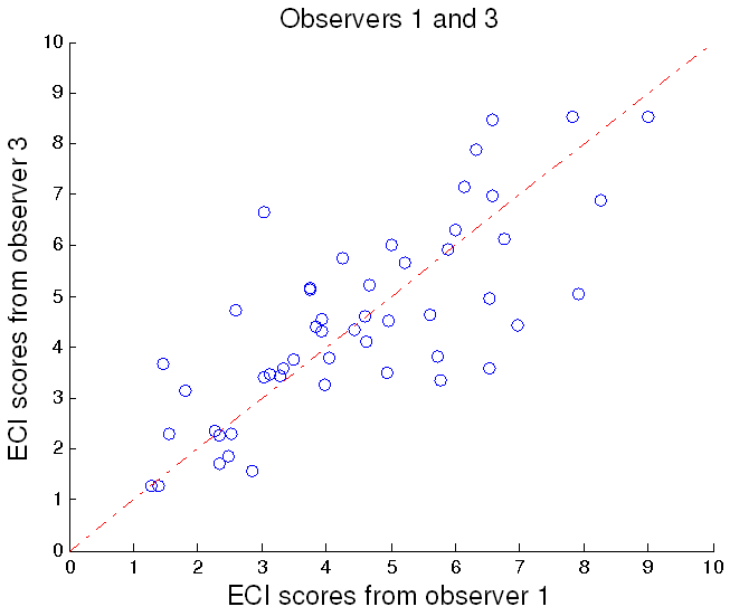
Of the 56 thyroid nodules, 53 were confirmed as malignant by the histopathological exams, where 52 were the classic type of papillary carcinoma and one was a follicular variant. Three nodules were determined to be benign (two cases of nodular hyperplasia and one Hurthle cell adenoma), even though a preoperative FNA biopsy was suspicious for malignancy. The mean nodule size was 9.5±4.4 mm (ranging from 4 mm to 24 mm) in the transverse view. Out of 56 nodules, 33

(58.9%) had the diameter less than 10 mm. All nodules were solid in US and had no cystic component inside.

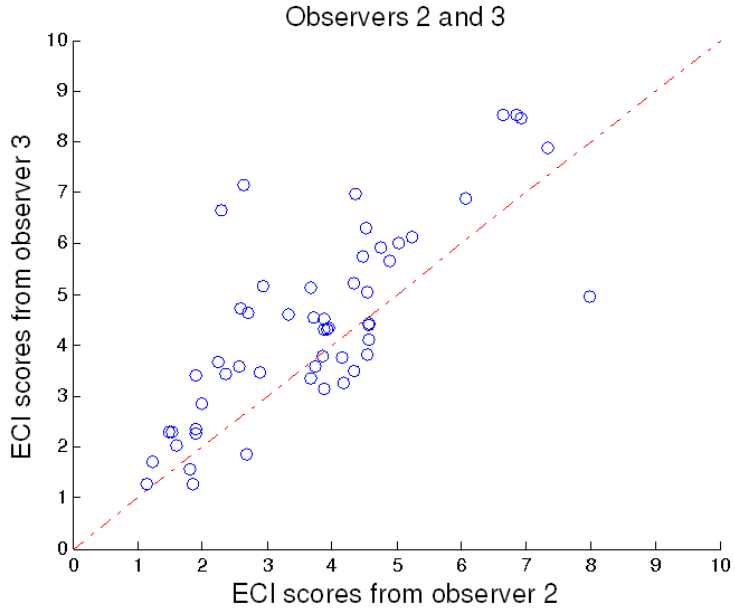
The ECI values of thyroid nodules obtained by observers 1, 2 and 3 were 4.40±1.98 (1.01-9.02), 3.72±1.72 (1.11-9.10), and 4.51±2.06 (1.11-10.66), respectively. Pearson's correlation coefficients for intraobserver agreement were 0.87 ($p < 0.01$), 0.73 ($p < 0.01$) and 0.84 ($p < 0.01$) for observer 1, 2, and 3, respectively. Figure 4.2 shows the scatter plots of ECI values between two observers. Pearson's correlation coefficients between two observers were 0.79 (observers 1 and 2, $p < 0.01$) in Fig. 4.2(a), 0.77 (observers 1 and 3, $p < 0.01$) in Fig. 4.2(b) and 0.73 (observers 2 and 3, $p < 0.01$) in Fig. 4.2(c). It indicates that good to excellent intraobserver and interobserver agreement exists in thyroid elastography with intrinsic compression.



(a)



(b)



(c)

Figure 4.2. Scatter plots of ECI scores from (a) observers 1 and 2, (b) observers 1 and 3, and (c) observers 2 and 3.

4.4 Discussion

With external compression elastography, the variability can be introduced at several places in Fig. 1.4. The first source is in selecting an imaging plane by an operator (step 1). Second, different pseudo-color patterns could be generated by different compression levels applied by the operator (step 2). Other sources of variability include the manual selection of an elastography image from a dynamic sequence (step 4) and the subjective scoring after visually inspecting the pseudo-color pattern (step 5). Due to the variability in both data acquisition and scoring, the reliability with external compression elastography suffers.

Park et al. (2009) evaluated the interobserver agreement of US thyroid elastography using external compression, where three radiologists independently acquired elastography data and assigned scores. They found no statistically significant interobserver agreement for thyroid elastography using external compression. In this study, three endocrinologists performed both elastography data acquisition and nodule delineation for scoring. We found that significant interobserver and intraobserver agreement exists with thyroid US elastography. One major difference in our study was the use of intrinsic compression.

There are several differences between intrinsic (Fig. 4.1) and external (Fig. 1.4) compression elastography methods. The first difference is that during data acquisition, no external compression is applied with intrinsic compression elastography. Because the carotid artery pulsation is used, the variability due to different compression levels used by the operator(s) is reduced. In addition, the out-of-plane motion, which could degrade the quality of elastography images, would be small compared to external compression elastography as the transducer is fixed during data acquisition and the magnitude of intrinsic compression is small. The second difference is in scoring. For external compression elastography, a clinician needs to manually select a single elastography image from an elastography sequence and assign a score. In contrast, we use a quantitative scoring method, where no subjective score assignment is needed except for delineating a nodule's ROI with 7~8

points. These differences in data acquisition and scoring are mainly responsible for a very different (from Park et al. (2009)) finding of the existence of statistically significant interobserver and intraobserver agreement in thyroid US elastography.

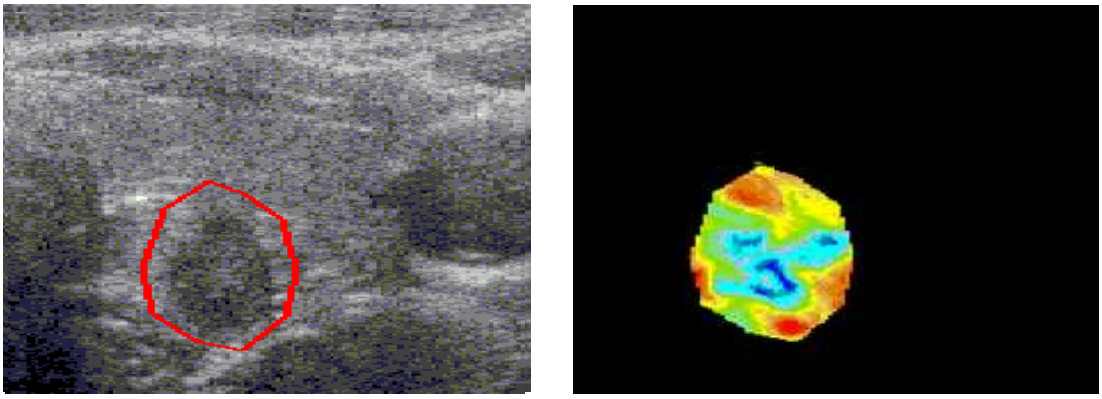
In addition to the interobserver agreement, we also evaluated the intraobserver reproducibility. For a typical intraobserver agreement study (Choi et al. 2010), the same observer retrospectively reviews the acquired images several times separated by a certain time period (e.g., 1 month). Rather than using the acquired images multiple times, we evaluated the intraobserver reproducibility based on two separate elastography examinations performed by the same observer including data acquisition. So, the intraobserver reproducibility in this study encompasses the variability involved in the entire elastography examination process using intrinsic compression, including data acquisition and scoring. In spite of this, we found very good intraobserver reproducibility.

Since each operator produced two ECI values for each nodule, there can be a total of four Pearson's correlation coefficients between two observers instead of one correlation coefficient when the mean of two ECI values was used (which is presented in Results). The mean and standard deviation of these four Pearson's correlation coefficients are 0.71±0.07 (observers 1 and 2), 0.72±0.04 (observers 1 and 3) and 0.65±0.06 (observers 2 and 3), which are about 10% lower than those (0.79 (observers 1 and 2), 0.77 (observers 1 and 3) and 0.73 (observers 2 and 3)) when the mean values were used. When an operator makes multiple ECI measurements on one nodule, it would be clinically prudent to use the maximum ECI value to increase the sensitivity (i.e., malignant nodule detection rate). In that case, the interobserver Pearson's correlation coefficients would be 0.79 (observers 1 and 2), 0.80 (observers 1 and 3) and 0.71 (observers 2 and 3), which are very similar to those when the mean ECI values were used.

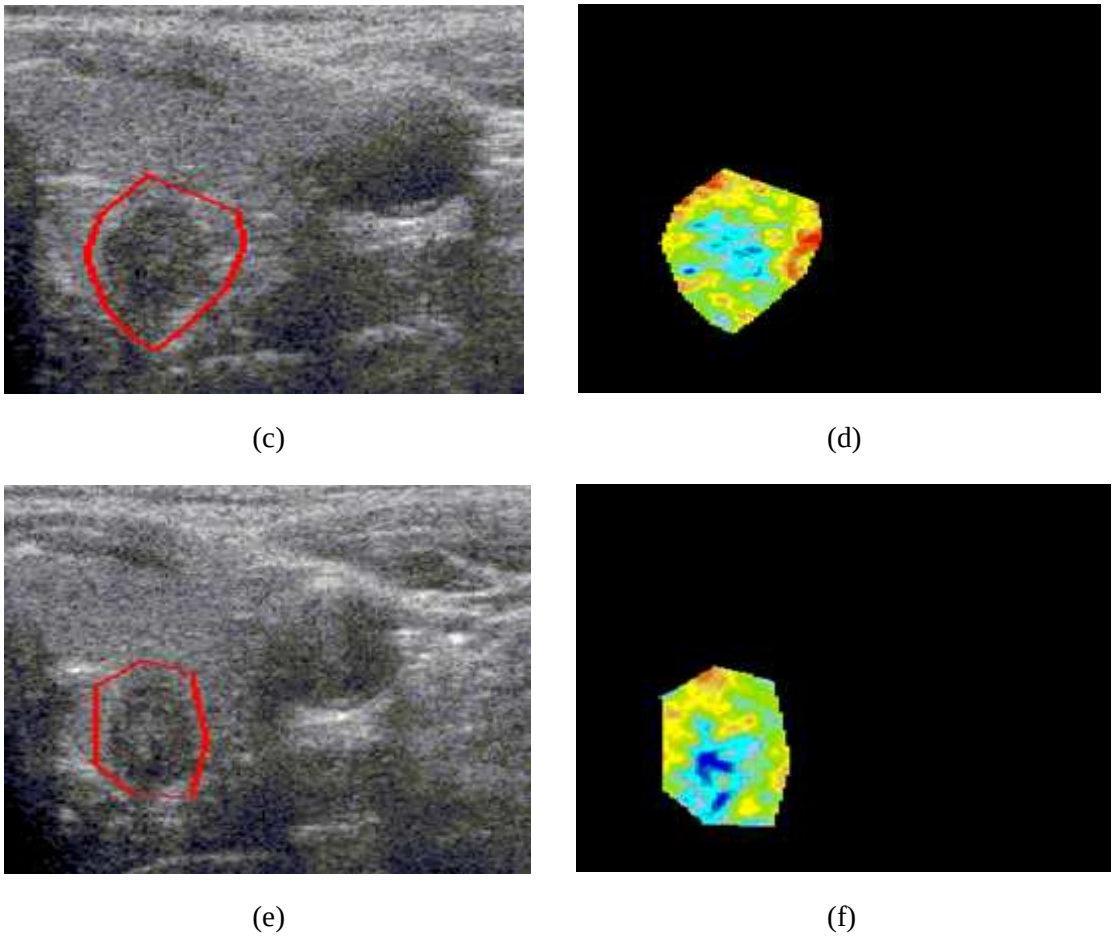
In spite of the fact that all three endocrinologists in this study had limited experience in using US to scan thyroid nodules, good to excellent interobserver and intraobserver agreement was obtained. It suggests that the reproducibility with

intrinsic compression elastography may be less dependent on an operator's prior experience in US. While more training is likely to improve the reproducibility further, it might not require as much training and experience to get reliable elastography results compared to external compression thyroid elastography.

Although the magnitude of both interobserver and intraobserver variability was reduced with intrinsic compression elastography, a certain level of variability still remains. There is always some variability in selecting an imaging plane during an elastography exam as in other US modes, e.g., B-mode and spectral Doppler. A transverse plane showing the maximum diameter of a nodule was selected for US elastography acquisition in our study. Since our operators were not much experienced in examining thyroid nodules with US, they might have used somewhat different imaging planes. Figure 4.3 shows an example where different ECI values were obtained from the same nodule by different observers. SOMs show different patterns, indicating that three operators acquired data from somewhat different portions of the nodule. Furthermore, it was not easy for them to go back to the same imaging plane for the second acquisition. As a result, there would be a difference in the ECI values obtained by the same observer at two separate exams. The inherent cardiac variability could also contribute to interobserver and intraobserver variability. For example, the blood pressure and/or the heart rate of a patient could be somewhat different during multiple elastography examinations, which could lead to a difference in the measured ECI scores.



(a) (b)



(c) (d)

(e) (f)

Figure 4.3. 62-year-old woman with a 6-mm hypoechoic nodule in the left thyroid lobe. Although B-mode images acquired by three observers look similar (Figs. 4.4(a), 4.4(c) and 4.4(e)), the SOMs obtained by observers 1 (Fig. 4.4(b)), 2 (Fig. 4.4(d)) and 3 (Fig. 4.4(f)) look somewhat different. The corresponding ECI value is 5.75, 3.51 and 7.47, respectively. This nodule was confirmed as papillary carcinoma by histopathology.

There are several limitations in this study. First, a relatively small number of nodules were enrolled. Second, two elastography exams performed by the same operator were only separated by less than 10 minutes, so the memory effect could have played a role in the second acquisition. However, since the ECI score was generated only after an operator acquired the data and delineated the nodule boundary, the memory effect would not be large. Third, the reproducibility of ECI values over time, which could be affected by the blood pressure, arterial compliance and cardiac output, was not evaluated in this study. Fourth, the study population was limited to

surgery-bound patients with predominantly malignant nodules, thus the diagnostic performance of intrinsic compression elastography was not evaluated in this study.

4.5 Conclusion

We obtained good interobserver and intraobserver agreement with intrinsic compression elastography. Compared to external compression elastography, which would take a high level of training and experience for sonographers and radiologists to use, the intrinsic compression thyroid elastography method would be able to produce more reliable results with less experience and training.

Chapter 5 – Thyroid nodule parameters influencing performance of ultrasound elastography using intrinsic compression

5.1 Introduction

As discussed in Chapter 4, a major limitation of US elastography with externally-applied compression is low reproducibility. When the intrinsic carotid artery pulsation is used as a compression source, the interobserver and intraobserver variability could be substantially reduced. It suggests that the thyroid elastography with intrinsic compression could render more reproducible results compared to the external compression elastography (Lim et al. 2012).

Another issue with US elastography using external compression is that the elastography image quality and diagnostic accuracy could be influenced by a nodule's parameters, such as size and location. For example, the elastography image quality degrades for deep-seated nodules (from the compression source) due to less compression they receive. Chang et al. (2011) studied the factors influencing the elastography image quality in evaluating suspicious breast masses. They found the lesion size, lesion depth and breast thickness are significantly associated with the elastography image quality. Sensitivity in detecting malignant breast masses increases significantly with higher image quality.

By using the quantitative scoring method in intrinsic compression elastography, the diagnostic accuracy would be less influenced by the image quality and physician's skills in interpreting the elastography image compared to subjective scoring (Luo et al. 2011; Lim et al. 2012; Luo et al. 2012). However, whether the diagnostic accuracy of intrinsic compression elastography would be influenced by nodule parameters has not been studied. Identifying those parameters that influence the predictive value of US elastography would be beneficial for wider use of this technique by clinicians in

managing thyroid nodules. Thus, the aim of this study was to evaluate several nodule parameters that could influence the diagnostic performance of thyroid US elastography using intrinsic compression.

5.2 Materials and methods

5.2.1 Patients

From August 2010 to March 2011, 167 patients were recruited for US B-mode and elastography examinations in the Seoul St. Mary's Hospital. Seven patients with unsatisfactory FNA results and 4 patients with atypical cells of undetermined significance (ACUS) were excluded from the study. Thus, a total of 156 patients with 176 nodules formed the study population. The mean age of patients was 47.3±12.7 years, and 82.1% were female.

Based on the FNA biopsy results, nodules were classified using The Bethesda System for Reporting Thyroid Cytopathology (TBSRTC) as negative for malignancy (79 nodules), follicular neoplasm (5 nodules), Hurthle cell neoplasm (3 nodules), or malignant (including suspicious for malignancy, 89 nodules). Patients with follicular neoplasm or Hurthle cell neoplasm went through surgery, where 3 papillary thyroid carcinomas (PTC), 3 follicular adenomas, 1 Hurthle cell adenoma and 1 follicular thyroid carcinoma were found based on the histopathological results. 86 nodules (81 patients) out of 89 with cytological findings being malignant or suspicious for malignancy went through surgery in the Seoul St. Mary's Hospital. The histopathological findings confirmed that all 86 nodules were PTCs. Since the histopathological findings of 3 nodules diagnosed as PTCs by FNA biopsy were not available (surgery was not performed in the Seoul St. Mary's Hospital), their cytological results were used in the study. Thus, there were a total of 83 benign and 93 malignant nodules (92 papillary thyroid carcinomas and 1 follicular carcinoma) enrolled in the study. There was no significant difference in the mean age (49.7±13.2

vs. 45.5±12.1, p = 0.055) and proportion of women (87% vs. 78%, p = 0.210) between patient groups with benign and malignant nodules. This study was approved by the Institutional Review Board of the Seoul St. Mary's Hospital, and written informed consent was obtained from all participants.

5.2.2 Real-time US elastography examination and nodule parameter measurements

B-mode US and elastography examinations were performed by one of three endocrinologists using a commercial US machine (Accuvix V20, Samsung Medison Co. Ltd., Seoul, Korea) with a L5-13 linear transducer.

ECI was calculated to quantify the local strain contrast within a nodule. As discussed in Chapter 3, for the elastography image of a malignant nodule (Fig. 5.1), the difference between low and high strain areas is large, and the transition from high to low (or low to high) strain occurs over a short distance, leading to a high local strain contrast and a large ECI value. On the other hand, a benign nodule shown in Fig. 5.2, which has a relatively small stiffness difference to normal thyroid tissue, typically has a small ECI value.

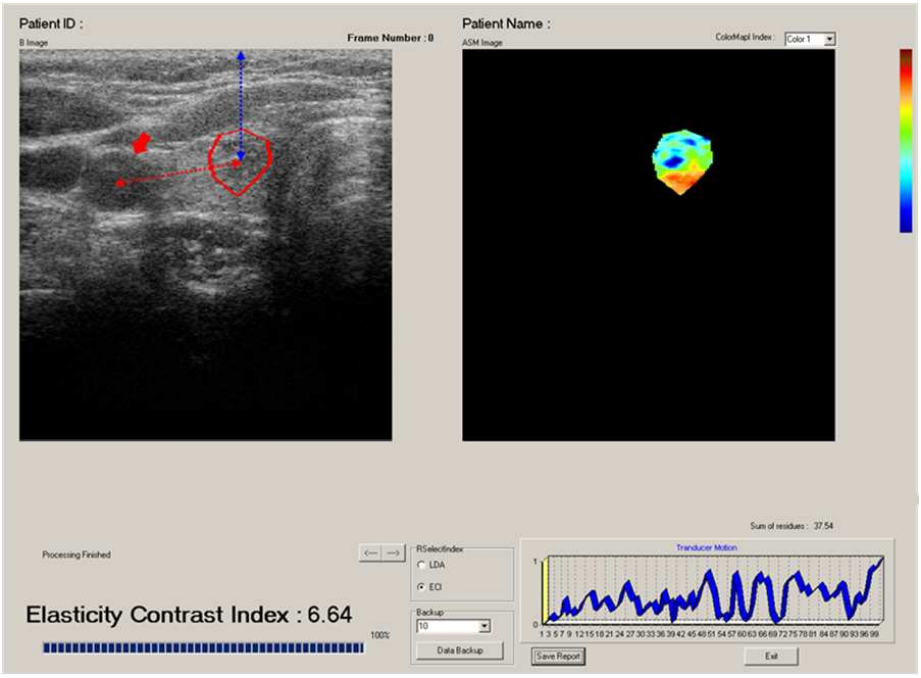


Figure 5.1. US B-mode (left) and elastography (right) images for a 6-mm papillary thyroid carcinoma in the right thyroid lobe. In the B-mode image, the carotid artery is pointed by a thick red arrow. The red line represents the distance between the nodule and the carotid artery, and the blue line represents the nodule depth. For the elastography image, red indicates high strain (i.e., soft tissue) while blue indicates low strain (i.e., stiff tissue). A large ECI value (i.e., 6.64) indicates that this is likely to be a malignant nodule.

To perform an elastography examination, an operator searched for a transverse plane showing both the common carotid artery and the largest diameter of a thyroid nodule using B-mode. The operator asked the patient to hold his/her breath and acquired the US data for ~4 seconds once the imaging plane was identified. No external compression was applied during data acquisition as the intrinsic carotid artery pulsation was used as a compression source. The strain frames were generated using the acquired data. The ECI value was computed and displayed on the monitor of the US machine after a nodule's boundary was delineated by the operator.

Three nodule parameters were measured based on the B-mode image: (1) size, (2) depth and (3) distance to the carotid artery. The size of a nodule was measured as the maximal diameter in the transverse view. The depth of a nodule was calculated as the vertical distance from the skin to the center of the nodule (blue vertical line in

Figs. 5.1 and 5.2). The Euclidean distance between the center of a nodule and the center of the carotid artery was measured as the nodule's distance to the carotid artery (red line in Figs. 5.1 and 5.2).

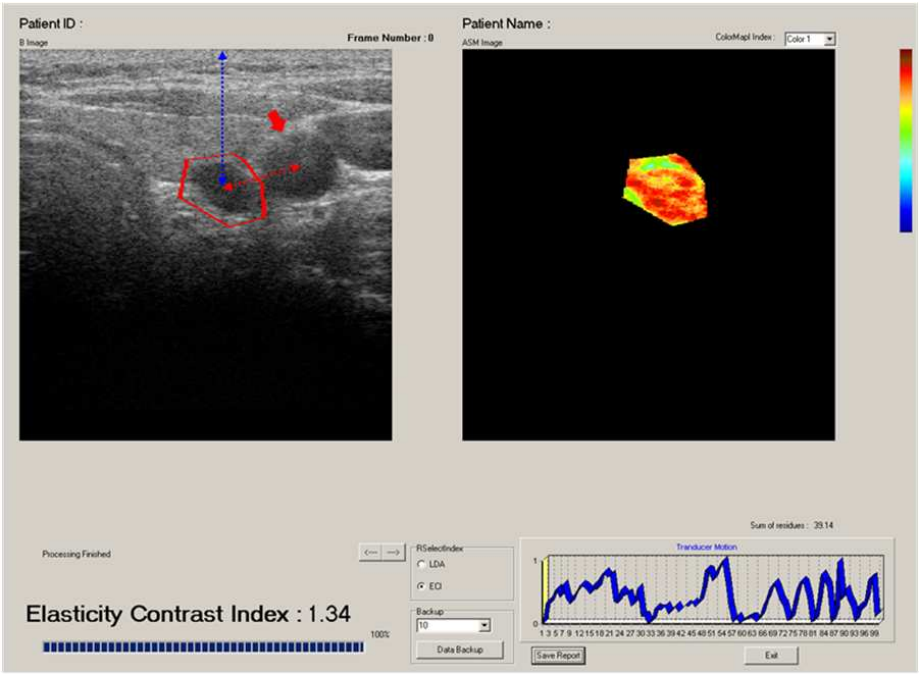


Figure 5.2. US B-mode (left) and elastography (right) images for an 8-mm benign nodule in the left thyroid lobe. In the B-mode image, the carotid artery is pointed by a thick red arrow. The red line represents the distance between the nodule and the carotid artery, and the blue line represents the nodule depth. For the elastography image, red indicates high strain (i.e., soft tissue) while blue indicates low strain (i.e., stiff tissue). Almost the entire nodule is coded as red in the elastography image. A small ECI value (i.e., 1.34) was obtained for this benign nodule.

5.2.3 Statistical analysis

Nodule parameters influencing the ECI value

Parameters influencing the ECI value were examined by univariate regression analysis with the ECI value as the dependent variable. The analysis was performed separately for benign and malignant thyroid nodules. Multivariate regression analysis was used to determine the most important parameter influencing the ECI

value.

Nodule parameters associated with the diagnostic accuracy of elastography

Multivariate logistic regression analysis was used to evaluate several nodule parameters associated with the diagnostic accuracy of US elastography, where the dependent variable was the classification result (correct or incorrect). A thyroid nodule was considered to be correctly classified when the elastography result was concordant with the cytological or histopathological (if available) finding. A larger odds ratio implies a higher probability in correctly differentiating benign and malignant nodules.

Statistical analyses were performed by using Matlab Statistics Toolbox™ (The MathWorks, Inc., Natick MA). $p < 0.05$ was considered as statistically significant.

5.3 Results

5.3.1 Nodule parameters influencing the ECI value

The mean nodule size was 7.8±3.3 mm, the mean distance to the carotid artery was 13.2±4.5 mm, and the mean depth was 12.4±3.5 mm. There was no significant difference in the nodule size, distance to the carotid artery, or nodule depth between benign and malignant nodules. In univariate analysis, the distance to the carotid artery was found to be significantly associated with the ECI value for malignant nodules, but not for benign nodules (Table 5.1). On the other hand, there was no correlation between the ECI value and the size and depth of both benign and malignant nodules. Multivariate regression analysis also showed that the distance to the carotid artery was the only factor influencing the ECI value ($p < 0.0001$) for malignant nodules, which was concordant with the univariate regression analysis result.

Table 5.1. Nodule parameters influencing the ECI value (univariate regression analysis)

	Benign (n=83)		Malignant (n=93)	
	Coefficient	*p* value	Coefficient	*p* value
Size	-0.11	0.16	0.01	0.97
Depth	0.02	0.66	-0.04	0.41
Distance to the carotid artery	0.05	0.32	0.13	< 0.0001

Table 5.2 lists the mean ECI value of two equally-divided subgroups based on the size, depth and distance to the carotid artery for benign and malignant nodules. The mean ECI value (3.87±1.41) for malignant nodules close to the carotid artery is significantly lower than that (4.91±1.76) for malignant nodules distant from the carotid artery (p = 0.0039). On the other hand, no statistically significant difference exists between the mean ECI values of other subgroups.

Since the nodule size is an important factor in deciding on FNA referral in clinical settings, we also separated benign and malignant nodules into two subgroups based on whether the nodule size is equal to or great than 10 mm or not. We found no significant difference in the mean ECI values of two subgroups as well.

Table 5.2. The mean ECI value for nodules grouped based on the nodule size, depth and distance to the carotid artery

	Benign			Malignant	
	Mean ECI	*p* value		Mean ECI	*p* value
Nodule size			Nodule size		
< 7.2 mm	2.69±1.67	> 0.05	< 8 mm	4.44±1.92	> 0.05
≥ 7.2 mm	2.54±1.15		≥ 8 mm	4.35±1.40	
Nodule depth			Nodule depth		
< 12 mm	2.32±1.34	> 0.05	< 11.8 mm	4.45±1.64	> 0.05
≥ 12 mm	2.87±1.50		≥ 11.8 mm	4.30±1.71	
Distance to the carotid artery			Distance to the carotid artery		
< 11.5 mm	2.42±1.29	> 0.05	< 13 mm	3.87±1.41	0.0039
≥ 11.5 mm	2.79±1.56		≥ 13 mm	4.91±1.76	

5.3.2 Nodule parameters associated with the diagnostic accuracy of elastography

To assess the diagnostic accuracy of US elastography, we varied an ECI cut-off value and determined the sensitivity and specificity in detecting malignant nodules. With an ECI cut-off value of 2.9, the US elastography achieved the maximum geometric mean (i.e., $\sqrt{sensitivity \times specificity}$) with a sensitivity of 85.0% and specificity of 67.5%. The mean ECI value of malignant nodules (4.39±1.66) is significantly higher than that of benign nodules (2.59±1.43) ($p < 0.0001$). The area under the receiver operator characteristic (ROC) curve is 0.79 (95% confidence interval: 0.72~0.85).

Table 5.3 shows the association between the nodule parameters and the diagnostic accuracy of US elastography. The diagnostic accuracy of US elastography is significantly associated with the distance to the carotid artery ($p < 0.05$). On the other hand, the nodule size and depth are not significantly associated with the diagnostic accuracy. Table 5.4 lists the odds ratios for three groups of nodules with different distances to the carotid artery. The odds ratio is higher for the nodules with a distance greater than 10 mm from the carotid artery.

Table 5.3. Nodule parameters associated with the diagnostic accuracy of US elastography (multivariate logistic regression analysis)

	Odds ratio	95% CI	*p* value
Nodule size	1.14	0.92 – 1.40	0.23
Nodule depth	0.98	0.91 – 1.06	0.67
Distance to the carotid artery	1.07	1.02 – 1.18	< 0.05

Table 5.4. Odds ratios in correctly differentiating benign and malignant nodules for three groups of nodules with different distances to the carotid artery

	Odds ratio	95% CI	*p* value
Distance to carotid artery			
< 10 mm (n=43)	2.31	1.20 – 4.42	< 0.05
10 - 15 mm (n=90)	4.29	2.53 – 7.38	< 0.05
> 15 mm (n=43)	3.30	1.63 – 6.70	< 0.05

5.4 Discussion

In Chapter 4, significant interobserver and intraobserver agreement with US elastography using the carotid artery pulsation was found. However, since US thyroid elastography requires compression (either external or intrinsic) to a nodule in deriving its stiffness, the performance could be influenced by several nodule parameters as these parameters would affect the amount of compression delivered to the nodule. Therefore, before US elastography can be accepted as a routine clinical tool in managing thyroid nodules, it is important to assess how nodule parameters influence its performance.

In this study, we investigated three nodule parameters (i.e., size, depth and distance to the carotid artery) that could influence the performance of US elastography using intrinsic compression. We found that a nodule's distance to the carotid artery is the only parameter significantly influencing the ECI value and the diagnostic accuracy of US elastography. The mean ECI value of malignant nodules close to the carotid artery (3.87±1.41) is significantly lower than that of malignant nodules distant from the carotid artery (4.91±1.76). It indicates that the amount of force applied to a thyroid nodule is dependent on its distance to the compression source. A nodule close to the compression source would receive a larger amount of compression, leading to a larger strain magnitude and a lower ECI score. By analyzing the local strain contrast instead of absolute strain magnitude, the ECI algorithm reduces the effects of a nodule's location with respect to the carotid artery to some extent (Luo et al. 2012). However, the influence of a nodule's distance to the carotid artery on ECI could not be completely eliminated, especially for those small nodules very close to the carotid artery. Thus, it is recommended that US elastography with intrinsic compression should be used cautiously in evaluating thyroid nodules located very close to the carotid artery (e.g., less than 10 mm).

Our results suggest that the nodule depth has no significant influence on the ECI value and the diagnostic accuracy of thyroid US elastography. Chang et al. (2011) evaluated the factors influencing the performance of breast US elastography

with external compression. They found that shallower lesion depth is significantly associated with higher elastography image quality and diagnostic accuracy as the distance to the compression source (i.e., US transducer) is smaller. Since the force is internally provided by the carotid artery pulsation instead of an US transducer, the elastography performance using intrinsic compression would be less influenced by the thyroid nodule depth compared to that using external compression. However, Chang et al. (2011) and our study results agree that the distance of lesions to the compression source (e.g., transducer and carotid artery) is an important factor influencing the elastography performance regardless of external or intrinsic compression.

We found that the diagnostic accuracy of thyroid US elastography using intrinsic compression was not significantly associated with the nodule size. On 133 nodules (59 benign and 74 malignant) less than 10 mm in size, the area under the ROC curve is 0.79 (95% confidence interval: 0.69~0.88), which is as good as the overall diagnostic performance. On the other hand, as the size of a nodule was measured as the maximal diameter in the transverse view, only the mediolateral and anteroposterior dimension of a nodule was considered in this study. A nodule, which has a diameter equal to or greater than 10 mm in the craniocaudal but less than 10 mm in the mediolateral and anteroposterior dimension, would be still considered as < 10 mm in this study. Thus, the influence of nodule size might not have been fully evaluated in this study. To fully evaluate the influence of nodule size on the diagnostic performance of elastography, the maximum diameter of a nodule among all three directions needs to be considered in future studies.

There are several limitations in the current study. First, the patient factors (e.g., weight, blood pressure and combined diseases, such as diabetes mellitus, hypertension or atherosclerosis) that might influence the diagnostic accuracy of US elastography were not analyzed. Because the carotid artery pulsation was used as the compression source, the factors associated with the arterial stiffness could have impacted elastography results. Second, the presence of autoimmune thyroiditis, which may modify the stiffness of thyroid tissue, was not evaluated in the study.

Third, there are 56 surgery-bound nodules included in the current study, where 53 are malignant nodules. As a result, the malignant rate is 52.8%, which is substantially higher than ~15% typically seen among FNA-bound nodules. Also, for a surgery-bound nodule, an endocrinologist was aware of the FNA findings of the nodule before performing the elastography exams. Therefore, certain bias would be introduced into the elastography results due to the operator's prior knowledge of a nodule's cytodiagnosis. To fully evaluate the diagnostic performance of intrinsic compression elastography on FNA-bound nodules, future studies that only enroll patients referred for an FNA biopsy would be needed.

5.5 Conclusion

In this chapter, we evaluated the influence of a nodule's parameters on the ECI value and the diagnostic performance of elastography using intrinsic compression. We found that a nodule's distance to the carotid artery significantly influences the ECI value and the diagnostic performance of US thyroid elastography using intrinsic compression. No significant correlation was found between the diagnostic accuracy of elastography and a nodule's size and depth. With good interobserver and intraobserver agreement and less influence from a nodule's parameters on the diagnostic performance, the diagnostic performance of intrinsic compression elastography in a real clinical setting is evaluated in Chapter 6.

Chapter 6 – Diagnostic performance of intrinsic compression elastography in real clinical practice

6.1 Introduction

There have been numerous studies to evaluate the diagnostic performance of US elastography in noninvasive differentiation of benign and malignant thyroid nodules, where many reported very good diagnostic performance (Lyshchik et al. 2005; Rago et al. 2007; Asteria et al. 2008; Tranquart et al. 2008; Hong et al. 2009). In spite of high sensitivity and specificity values reported with US elastography, elastography is still not widely used in clinical practice in managing thyroid nodules.

Recently, it was reported by Moon et al. (2012) that external compression elastography with and without US B-mode features has inferior performance in differentiating benign and malignant thyroid nodules compared to US B-mode features only. The sensitivity and specificity values of US elastography were 15.7% and 95.3% (using Rago's scoring criterion (Rago et al. 2007)) or 65.4% and 58.2% (using Asteria's scoring criterion (Asteria et al. 2008)), which are substantially lower than those (e.g., sensitivity of 97% and specificity of 100% by Rago et al. (2007)) reported in previous studies. With this lowered performance, Moon et al. (2012) concluded that elastography is not useful as an adjunctive diagnostic tool to traditional US nor as a separate clinical tool.

Unluturk et al. (2012) evaluated the diagnostic performance of external compression elastography on 237 FNA-bound thyroid nodules from 194 patients. They reported a sensitivity of 47% and specificity of 80% in detecting malignant nodules. In explaining the difference in the diagnostic performance of elastography compared to previous studies, they attributed to the small number of patients (i.e., less than 100) in many previous thyroid elastography studies in that the diagnostic

accuracy of elastography could have been overestimated due to a small study population.

We evaluated the diagnostic performance of intrinsic compression elastography in Chapters 2 and 3, where very good sensitivity and specificity values were obtained. However, the results were not quite generalizable in endocrinology clinics. First, the elastography data were acquired by radiologists and/or sonographers who were very experienced in performing US elastography examinations. For endocrinologists or primary care doctors who have little experience in using US, the quality of data would degrade due to lower skills, and larger variability would be introduced in performing elastography examinations, resulting in lower diagnostic performance. Additionally, the elasticity scores of thyroid nodules were calculated off-line by the same researcher in the laboratory, taking 1.5 hours to process a data set and assign a score. Thus, the results obtained in Chapters 2 and 3 represent the best-case scenario and cannot be reproduced in clinics since lack of expertise and various variability (e.g., interobserver variability) typically encountered in an endocrinology clinic were not considered.

By collaborating with Samsung Medison, our intrinsic compression elastography was implemented and integrated in commercial US machines. With our elastography method in commercial US machines, we evaluated the interobserver and intraobserver agreement and the influence of a nodule's parameters on the diagnostic performance of elastography in Chapters 4 and 5. In this study, we have evaluated the diagnostic performance of intrinsic compression elastography in clinical practice with a large number of FNA-bound patients, where three endocrinologists performed the elastography exams and obtained the ECI scores in real time. We believe that the results obtained in this study would more typically represent the diagnostic performance of intrinsic compression elastography in realistic clinical settings.

6.2 Materials and methods

6.2.1 Patients

196 patients (237 nodules) who were referred to the Seoul St. Mary's Hospital by primary care physicians for an FNA biopsy from May 2011 to January 2012 were recruited for the study. After elastography examination, all patients went through an FNA biopsy regardless of the results of elastography. 26 nodules from 18 patients with atypical squamous cells of undetermined significance (ASCUS) or unsatisfactory FNA results were excluded from the study. Additionally, 9 cystic (> 90% of content) and 6 dominant cystic (> 50% of content) nodules were also excluded. For patients with multiple nodules, only the nodule(s) that had been biopsied was used in the study. Overall, 196 nodules from 165 patients were included in this study. We used the FNA result as the diagnosis for a thyroid nodule unless a patient subsequently underwent surgery. For patients who underwent surgery, the final diagnosis was based on the histopathological examination of the excised thyroid nodule. This study was approved by the Institutional Review Board of the Seoul St. Mary's Hospital, and written informed consent was obtained from all participants.

6.2.2 Real-time US elastography examination

An elastography examination was performed by one of three endocrinologists using a commercial US machine (Accuvix XG, Samsung Medison Co. Ltd., Seoul, Korea) with a L5-13 linear transducer. To perform an elastography examination, an endocrinologist searched for an imaging plane in the transverse view. Once an appropriate imaging plane was identified, the operator asked the patient to hold his/her breath and acquired the US data for ~4 seconds (Lim et al. 2012). No external compression was applied during data acquisition as the carotid artery pulsation was used as an intrinsic compression source. The ECI value was computed and displayed on the monitor of the US machine after a nodule's boundary was delineated by the operator (Luo et al. 2012).

For each nodule, a minimum of two measurements were taken at the imaging plane showing the largest diameter of a thyroid nodule in the transverse view. If the operator found other planes showing different US characteristics (e.g., echogenicity, presence of calcifications) from the plane with the largest diameter, additional measurements were taken at those image planes. For data analysis, the maximum ECI value was used among ECI values obtained from multiple measurements.

6.2.3 Statistical analysis

To assess the diagnostic performance of US elastography, we varied the ECI cut-off value and determined the sensitivity and specificity combination for the maximum geometric mean (i.e., $\sqrt{sensitivity \times specificity}$) in detecting malignant nodules. Statistical analyses were performed by using Matlab Statistics Toolbox™ (The MathWorks, Inc., Natick MA). $p < 0.05$ was considered as statistically significant.

6.3 Results

6.3.1 Baseline characteristics of patients and nodules

There were 154 benign and 42 malignant nodules. All malignant nodules were diagnosed as papillary thyroid carcinomas. The median age (46, range from 24 to 72) of patients with malignant nodules was not significantly different from that (52, range from 23 to 78) with benign nodules ($p > 0.05$). No difference of gender was found between patients with benign and malignant nodules ($p > 0.05$).

The mean size of nodules was 1.1 × 0.9 × 0.7 cm, ranging from 0.3 × 0.3 × 0.3 cm to 3.3 × 3.4 × 1.8 cm. There were 92 nodules (66 benign and 26 malignant) with the maximal diameter less than 1 cm, where the mean size of these sub-centimeter nodules was 0.6 × 0.6 × 0.5 cm. 104 nodules (88 benign and 16 malignant) had the

maximal diameter equal to or greater than 1 cm with their mean size of 1.5 × 1.1 × 0.9 cm. There were 184 solid nodules and 12 dominant solid nodules (solid in more than 50% of internal content).

6.3.2 Diagnostic performance of elastography

The mean ECI value (4.51±2.22) of malignant nodules was significantly larger than that (2.98±1.47) of benign nodules ($p < 0.001$). The sensitivity and specificity of elastography are 81.0% and 64.3% with an ECI cut-off value of 3.11, respectively. Figure 6.1 shows the ROC curve in detecting malignant thyroid nodules with our elastography. The area under the ROC curve is 0.74 (95% confidence interval: 0.64~0.83).

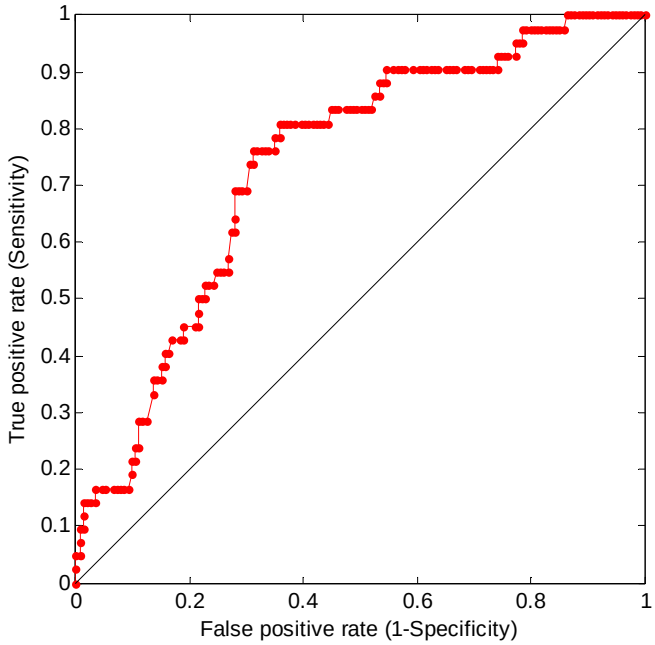


Figure 6.1. ROC curve in detecting malignant nodules with elastography.

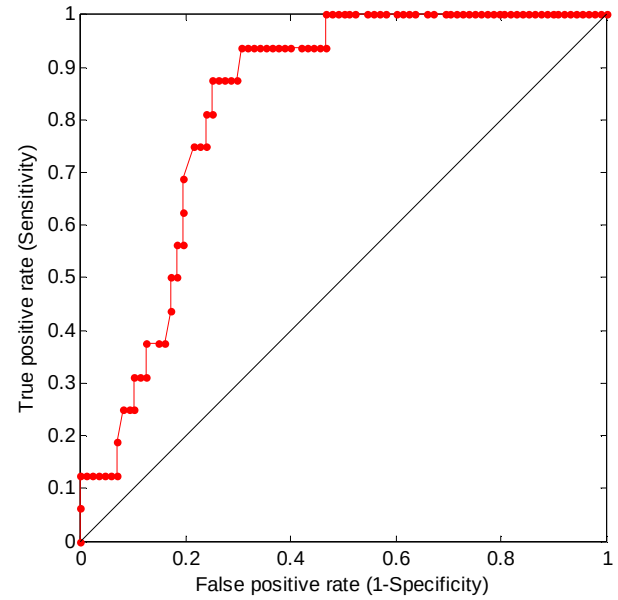
There are a total of 8 false negatives and 55 false positives. As shown in Table 6.1, 7 out of 8 false negative cases occur when the nodule diameter is less than 1

cm. We further evaluated the diagnostic performance of elastography on nodules ≥ 1 cm (n=104) and < 1 cm (n=92).

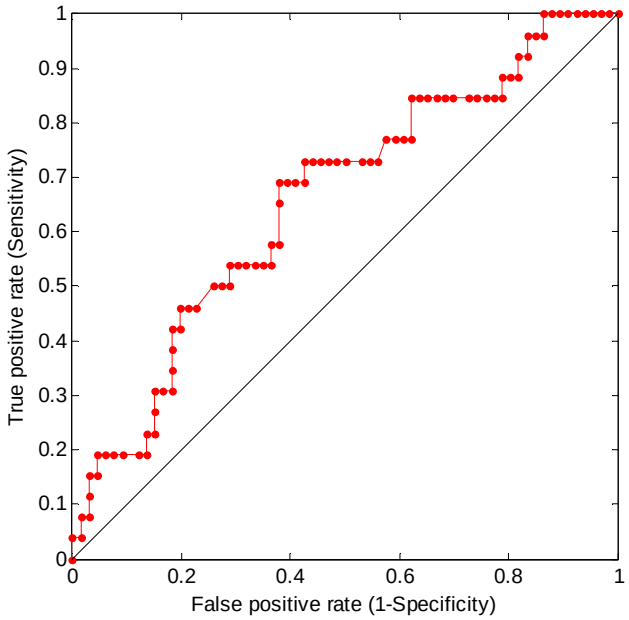
Table 6.1. False negative cases with elastography

	ECI value	Maximal diameter (mm)
Case 1	1.84	9
Case 2	2.57	6
Case 3	2.51	7
Case 4	2.79	5
Case 5	2.29	13
Case 6	1.75	6
Case 7	1.50	4
Case 8	1.93	4

The sensitivity and specificity of elastography on 104 nodules ≥ 1 cm are 93.8% and 69.3%, respectively. Figure 6.2(a) shows the ROC curve for elastography in detecting malignant nodules ≥ 1 cm, where the area under the ROC curve is 0.83 (95% confidence interval: 0.70~0.96). On the other hand, elastography obtained a sensitivity of 73.1% and specificity of 56.1% on 92 nodules < 1 cm. The ROC curve in detecting malignant nodules < 1 cm is shown in Fig. 6.2(b). The area under the ROC curve is 0.66 (95% confidence interval: 0.53~0.79), which is significantly less than that (0.83) on nodules ≥ 1 cm ($p < 0.05$).



(a)



(b)

Figure 6.2. ROC curve for elastography in detecting malignant nodules (a) ≥ 1 cm and (b) < 1 cm.

6.4 Discussion

Many previous studies claimed US elastography achieving a high sensitivity and specificity in differentiating benign and malignant thyroid nodules (Rago et al. 2007; Asteria et al. 2008; Tranquart et al. 2008; Hong et al. 2009). However, several recent studies reported that external compression elastography has a low sensitivity in detecting malignant nodules. Moon el al. (2012) evaluated the diagnostic performance of external compression elastography on 703 FNA-bound nodules, where the elastography examinations were performed by eight radiologists. The sensitivity and specificity values of US elastography were 15.7% and 95.3% (using Rago's scoring criterion (Rago et al. 2007)) or 65.4% and 58.2% (using Asteria's scoring criterion (Asteria et al. 2008)), respectively. Compared to the sensitivity of 97% and specificity of 100% reported by Rago et al. (2007) or the sensitivity of 94.1% and specificity of 81% by Asteria et al. (2008), the diagnostic performance reported by Moon et al. (2012) is substantially lower.

In Chapters 2 and 3, we obtained very good sensitivity and specificity values with intrinsic compression elastography. A sensitivity of 100% and specificity of 75.6% were obtained on 98 nodules (92 patients) in Chapter 2. Utilizing the elasticity contrast index (ECI), a sensitivity of 95% and specificity of 73.8% were achieved on 123 nodules (106 patients) in Chapter 3. In the current study, a sensitivity of 81.0% and specificity of 64.3% are obtained on 196 nodules (165 patients), which are lower than the diagnostic performance obtained in Chapters 2 and 3. There are several differences between the current and previous studies, which would contribute to the difference in the diagnostic performance of intrinsic compression elastography. First, the mean nodule size is 1.1 × 0.9 × 0.7 cm, which is substantially smaller than 2.3 × 1.7 × 1.9 and 2.3 × 1.8 × 2.1 cm in Chapters 2 and 3, respectively. 46.9% of nodules in the current study have the maximal diameter < 1 cm compared to only ~10% in Chapters 2 and 3. On nodules ≥ 1 cm (n=104), a sensitivity of 93.8% and specificity of 69.3% were obtained, which are comparable to the sensitivity of 95% and specificity of 73.8% in Chapter 3. Second, three

endocrinologists with little previous experience in US imaging performed the elastography examinations while radiologists and/or sonographers performed the data acquisition in previous studies. As many smaller thyroid nodules with the diameter < 1 cm were enrolled in the current study, it is difficult for the inexperienced endocrinologists to hold the transducer still for ~4 seconds in acquiring data from those small nodules, which would lead to the lower signal-to-noise ratio in estimating strain, resulting in the inferior diagnostic performance on nodules < 1 cm. Third, in the current study, the ECI values were interactively calculated on-line by each endocrinologist while the patient is still on an examination table rather than by a researcher off-line in the laboratory. Thus, much variability (e.g., interobserver variability) commonly seen in a clinical setting would be introduced compared to our previous studies, leading to the lower diagnostic performance of elastography in the current study.

We can compare the diagnostic performance of external compression elastography reported in Moon et al.' study with that of intrinsic compression elastography even though ideally they should be compared with the same operator(s) performing both external and intrinsic compression elastography on the same study population. As there are ~50% of nodules with the maximal diameter < 1 cm in both Moon et al.'s and our current study, the diagnostic performance would be compared on: 1) all nodules, 2) nodules ≥ 1 cm and 3) nodules < 1 cm. The sensitivity and specificity values of external compression elastography were 65.4% and 58.2% using Asteria's scoring criterion on 703 nodules. The sensitivity and specificity of intrinsic compression elastography were 81.0% and 64.3%, which are higher than those of external compression elastography. On nodules ≥ 1 cm, intrinsic compression elastography achieves a sensitivity of 93.8% and specificity of 69.3% compared to a sensitivity of 76.8% and specificity of 64.9% with external compression elastography. On nodules < 1 cm, the intrinsic compression elastography achieves a higher sensitivity (73.1% vs. 60.1%) and specificity (56.1% vs. 51.8%) than external compression elastography.

In another study performed by Unluturk et al. (2012), where the total number of nodules (237) is similar to this study, a sensitivity of 47% and specificity of 80% was obtained with external compression elastography, where there were 190 nodules ≥ 1 cm (80.2%) and 47 nodules < 1 cm (19.8%). Intrinsic compression elastography has a much higher sensitivity (81.0% vs. 47%) and somewhat lower specificity (64.3% vs. 80%) compared to external compression elastography. On nodules ≥ 1 cm, the sensitivity (93.8%) of intrinsic compression elastography is more than two times higher than that (42.1%) of external compression elastography with somewhat lower specificity (69.3% vs. 80.9%). On nodules < 1 cm, the sensitivity (73.1%) of intrinsic compression elastography is higher than that (55%) of external compression elastography while the specificity is lower (56.1% vs. 77.8%).

In comparing our current study with Moon et al. and Unluturk et al.'s study, where all three studies were performed on a large number of FNA-bound nodules in a real clinical setting, we found that the diagnostic performance, especially the sensitivity of intrinsic compression elastography is superior compared to external compression elastography. We believe that good interobserver and intraobserver agreement of intrinsic compression elastography mainly accounts for its superior diagnostic performance to external compression elastography. To confirm this result, a side-by-side comparison study, where the same operator(s) performs both external and intrinsic compression elastography on the same patient population, needs to be conducted in the future.

In the current study, we found that intrinsic compression elastography obtained a high sensitivity (i.e., 93.8%) and good specificity (i.e., 69.3%) in detecting malignant nodules ≥ 1 cm. Based on the current major guidelines for managing thyroid nodules, a nodule ≥ 1 cm would be strongly recommended to be referred to an FNA biopsy. For example, according to the American Thyroid Association (ATA) guidelines, any thyroid nodules that is 1 cm or greater in diameter should be evaluated using an FNA biopsy (Cooper et al. 2009). The Society of Radiologists in Ultrasound (SRU) guidelines strongly recommend an FNA biopsy on solid nodules

greater than 1 cm when microcalcifications are present and solid nodules greater than 1.5 cm if coarse calcifications are present (Frates et al. 2005). The American Association of Clinical Endocrinologists (AACE) guidelines state an FNA biopsy to be performed on all nodules 1 cm or greater with suspicious US features (Gharib et al. 2006). However, the predictive value of nodule size and US features in detecting malignant nodules is typically low, and features that are characteristical of malignant nodules (e.g., microcalcifications) are also present in benign nodules. This results in a large number of benign nodules undergoing an FNA biopsy. As can be seen in this study, 88 (84.6%) benign out of 104 nodules ≥ 1 cm received an unnecessary FNA. In the new management scheme as shown in Fig. 6.3, a clinician could rely on the elastography findings in judiciously deciding which nodules should be referred for an FNA biopsy. If a nodule ≥ 1 cm has an elastography finding indicative for malignancy, then a clinician should refer the nodule for an FNA. On the other hand, if the elastography finding suggests benignancy, a nodule even greater than 1 cm and with the presence of suspicious US features could be simply followed up in 6, 12 and 24 months, which is a standard practice, instead of being aspirated.

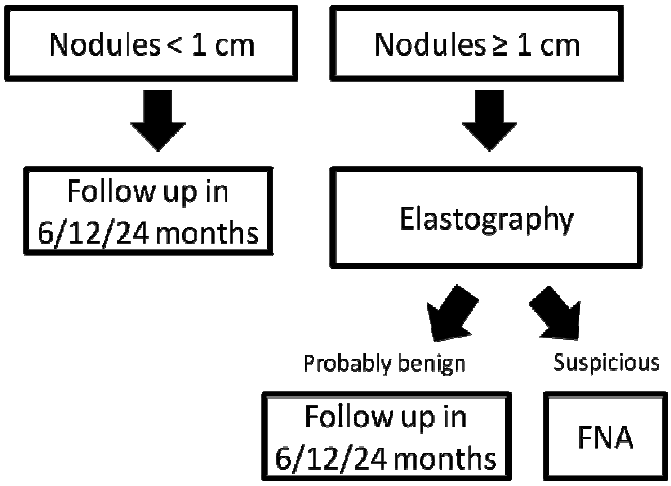


Figure 6.3. Proposed thyroid nodule management scheme with the use of elastography.

It is difficult for clinicians to differentiate benign and malignant nodules solely based on US B-mode features. Meanwhile, clinicians tend to refer nodules ≥ 1 cm to

an FNA biopsy as all major guidelines recommend an FNA on nodules ≥ 1 cm with or without suspicious US features. As a result, many benign thyroid nodules ≥ 1 cm go through FNAs. By using elastography on nodules ≥ 1 cm, a clinician could substantially increase the specificity while still achieving a high sensitivity in detecting malignant thyroid nodules as demonstrated in the current study. Thus, they could have more confidence in recommending nodules ≥ 1 cm for a follow-up instead of an FNA biopsy with the aid of elastography. The number of FNAs being performed on benign thyroid nodules would be substantially decreased (~60%), leading to better use of healthcare expenditures.

In the current study, the diagnostic performance of our intrinsic compression elastography was evaluated in an endocrinology clinic, where three endocrinologists with little experience in US imaging performed the elastography exams. Based on the study results, we suggest an example use of intrinsic compression elastography in managing thyroid nodules ≥ 1 cm. On the other hand, our intrinsic compression elastography is not limited to the use shown in Fig. 6.3. For example, US B-mode can provide diagnostically useful information (e.g., echogenicity, margin, calcification and/or concomitant lymph node enlargement) in differentiating benign and malignant thyroid nodules. Thus, intrinsic compression elastography can be used in conjunction with US B-mode, where better diagnostic performance could be achieved in differentiating thyroid nodules. Elastography could also be used by well-trained radiologists and/or sonographers in detecting malignant thyroid nodules < 1 cm as they have a higher level of skills in acquiring data from those small nodules. The proposed management scheme in Fig. 6.3 can be considered as a starting point in utilizing intrinsic compression elastography in clinical practice. As our elastography technique has been commercialized in ultrasound machines, the utility of intrinsic compression elastography in improving the quality of patient care can be further explored by other clinicians and researchers.

There are still some limitations for the current study. First, even though the ECI values were prospectively obtained by three endocrinologists, the ECI cut-off

value (i.e., 3.11) was determined after all the data were acquired. As shown in Fig. 6.3, to utilize intrinsic compression elastography prospectively on nodules ≥ 1 cm in deciding whether a nodule should be referred to an FNA biopsy or follow-up, the ECI cut-off value needs to be predetermined. In future studies, the diagnostic performance of intrinsic compression elastography needs to be evaluated with a predetermined ECI cut-off value. Additionally, the ECI cut-off value in the current study is somewhat different from that (i.e., 2.9) in Chapter 5. One reason is that different ultrasound machines were used in the current and previous study. Due to the different signal characteristics between two ultrasound machines, the ECI cut-off value would be different. Another reason could be due to different malignant rates between the current and previous study. 56 surgery-bound nodules were included in Chapter 5, resulting in a malignant rate of 52.8%. In contrast, the malignant rate of the current study is 21.4% as all the nodules are bound for FNAs. The ECI cut-off value could be varied due to different percentages of malignant nodules between two studies. The sensitivity of the ECI cut-off value to the malignant rate and/or other factors needs be evaluated in future studies.

Another limitation is that the role of intrinsic compression elastography in managing small sub-centimeter thyroid nodules was not fully evaluated in the current study. First, the small sub-centimeter thyroid nodules enrolled in the current study were pre-selected by primary care physicians, where only those highly suspicious for malignancy were referred for an FNA biopsy. As can be seen in the current study, the malignant rate (28.3%) of sub-centimeter nodules is almost two times higher than that (15.4%) of nodules ≥ 1 cm. All major guidelines recommend an FNA for small sub-centimeter nodules only if they are from patients with high-risk history (family history of cancer or previous exposure to ionizing radiation) and/or where abnormal cervical lymph nodes are found. Thus, a clinician is cautious in aspirating a sub-centimeter nodule and only considers an FNA when it is highly suspicious for malignancy. In contrast, a clinician has a tendency to refer a nodule ≥ 1 cm for an FNA even though it does not have any suspicious history and/or US features as all major guidelines use 1 cm as a major criterion in referring FNA. To fully evaluate

the diagnostic performance of intrinsic compression elastography on nodules < 1 cm, future studies need to be performed on sub-centimeter nodules that are not pre-selected for FNAs. Second, as mentioned previously, three endocrinologists who had little experience in US imaging performed the elastography exams in the current study. It was difficult for them to hold the transducer still during data acquisition from those small thyroid nodules. Thus, the diagnostic performance of elastography on nodules < 1 cm could be underestimated in the current study due to the operators' lack of experience and skills in using US. In future studies, the diagnostic performance of intrinsic compression elastography on sub-centimeter nodules need to be evaluated with well-trained operators.

6.5 Conclusion

In this chapter, we evaluated the diagnostic performance of intrinsic compression elastography in real clinical practice. Three endocrinologists performed the elastography examinations and obtained the ECI scores on 196 FNA-bound thyroid nodules, where a sensitivity of 81.0% and specificity of 64.3% were obtained. On 104 nodules ≥ 1 cm, intrinsic compression elastography obtained a sensitivity of 93.8% and specificity of 69.3% in detecting malignant nodules. With high sensitivity and good specificity on nodules ≥ 1 cm, elastography could be incorporated into the current major guidelines in reducing the number of FNAs performed on benign nodules.

Chapter 7 – Conclusion

7. 1 Introduction

Thyroid nodules are considered as an epidemic due to a large number of imaging studies performed and the increasing incidental detection of these nodules (Mazzaferri 1993; Wang and Crapo 1997; Yeung and Serpell 2008). Since the existing imaging modalities (CT, MR imaging and US) cannot accurately differentiate malignant from benign nodules (Youserm et al. 1997; Frates et al. 2005), an FNA biopsy, which costs typically $1,500 in the U.S., is performed on nodules showing suspicious features in US (e.g., shape, microcalcifications and/or irregular margins). However, the majority of FNA procedures (~70%) are performed on benign nodules (Gharib and Goellner 1993). Thus, by detecting many benign nodules and removing them from an FNA biopsy altogether, the quality of patient care would improve. Costs associated with FNA biopsies on patients with benign nodules could be substantially reduced. At the same time, although an FNA biopsy is a minimally-invasive procedure, it can take a toll on the patients emotionally, especially those who are fearful of procedures requiring the use of needles. Thus, the noninvasive screening of benign nodules with a high level of confidence can be of significant benefit to the patients and healthcare expenditures (Dighe et al. 2010).

While various studies have shown US elastography using external compression has the potential being a pre-FNA screening tool (Lyshchik et al. 2005; Rago et al. 2007; Asteria et al. 2008; Tranquart et al. 2008; Hong et al. 2009), there are some limitations of this elastography method. One limitation is that the amount of compression influences the elastography image and consequently its score. The optimum compression that should be used to generate a diagnostic elastography image is highly operator dependent. Another limitation is the subjective scoring method.

In detecting malignant nodules by elastography, the stiffness of a thyroid nodule is mostly inferred by visually inspecting the pseudo-color pattern in a thyroid elastography image relative to the surrounding tissues, after which a score is assigned. The variability in scoring would be introduced due to the subjective nature of this method. Because of the variability involved in the elastography data acquisition and scoring, no interobserver agreement was observed for US elastography using external compression in evaluating malignant thyroid nodules (Park et al. 2009).

The research in this dissertation has focused on the development and clinical evaluation of intrinsic compression elastography technique, which would overcome the limitations of previous methods with superior clinical usability. The scope of this dissertation includes developing new quantitative scoring methods for thyroid elastography using intrinsic compression, evaluating the interobserver and intraobserver agreement of the developed elastography technique, assessing the influence of a nodule's parameters on the diagnostic performance, and evaluating the diagnostic performance of intrinsic compression elastography in differentiating benign and malignant thyroid nodules in real clinical practice.

7. 2 Contributions

7.2.1 Quantitative scoring methods

To reduce the variability in scoring, we developed quantitative scoring algorithms for thyroid elastography using intrinsic compression. We found distinctively different strain rate waveforms for benign and malignant thyroid nodules as a result of the carotid artery pulsation due to their different stiffness values, which was utilized for nodule classification (Luo et al. 2011). Another quantitative scoring method (i.e., elasticity contrast index, ECI) was developed by utilizing the co-occurrence matrix to analyze the local strain contrast within a thyroid nodule (Luo et al. 2012), where we do not have to use a classifier.

7.2.2 Evaluation of the interobserver and intraobserver agreement of the developed elastography technique

To perform clinical studies using the developed elastography technique, we implemented and integrated our algorithms in commercial US machines. We performed a clinical study to evaluate the interobserver and intraobserver agreement of our elastography method including both data acquisition and scoring, where significant interobserver and intraobserver agreement was observed (Lim et al. 2012). We identified the variability sources for elastography using external and intrinsic compression and demonstrated that our intrinsic compression elastography could achieve more reproducible results compared to external compression elastography because it reduces the variability in both data acquisition and scoring. Compared to external compression elastography, which would take a high level of training and experience for clinicians to use, the intrinsic compression thyroid elastography method would be able to produce more reliable results with less experience and training.

7.2.3 Assessment of the influence of a nodule's parameters on the diagnostic performance of elastography

We have evaluated the influence of a nodule's parameters on the diagnostic performance of elastography using intrinsic compression (Kim et al. 2012). It was found that a nodule's distance to the carotid artery significantly influences the ECI value and the diagnostic performance of thyroid elastography using intrinsic compression. No significant correlation was found between the diagnostic accuracy of elastography and a nodule's size and depth. One finding from this study is that US elastography with intrinsic compression should be used cautiously in evaluating thyroid nodules located close to the carotid artery (e.g., less than 10 mm).

7.2.4 Evaluation of diagnostic performance of intrinsic compression elastography in real clinical practice

We performed a clinical study to evaluate the diagnostic performance of intrinsic compression elastography, where three endocrinologists performed the elastography examinations and obtained the ECI scores in a real clinical setting.

Intrinsic compression elastography achieved a sensitivity of 81.0% and specificity of 64.3% on 196 FNA-bound nodules. On 104 nodules ≥ 1 cm, intrinsic compression elastography obtained a sensitivity of 93.8% and specificity of 69.3% in detecting malignant nodules. With high sensitivity and good specificity on nodules ≥ 1 cm, intrinsic compression elastography would increase a clinician's level of confidence in deciding whether a thyroid nodule should go through an FNA or not.

7.3 Conclusions

To improve the clinical applicability of thyroid elastography, we have developed quantitative scoring methods to reduce the variability due to the subjective assessment of elastography images in Chapters 2 and 3. We have evaluated the clinical usability (i.e., interobserver agreement and intraobserver reproducibility, influence of a nodule's parameters on the diagnostic performance) of the developed elastography technique in Chapters 4 and 5. In Chapter 6, we have demonstrated the superior diagnostic performance of intrinsic compression elastography in differentiating benign and malignant thyroid nodules in clinical practice.

Due to the increased use of various imaging modalities, many incidental and subclinical thyroid nodules are being detected. All the current major guidelines for managing thyroid nodules recommend an FNA biopsy on a nodule ≥ 1 cm with or without suspicious US features. Due to the low specificity of nodule size and US features in predicting malignancy, many benign nodules go through an FNA biopsy. Thus, there exists a clinical need to noninvasively detect benign nodules and reduce the number of FNA biopsies performed on patients with benign nodules, which would lead to improving the quality of patient care while reducing the healthcare expenditures associated with managing thyroid nodules. As demonstrated in Chapters 2, 3 and 6, ~60% of FNAs performed on benign thyroid nodules ≥ 1 cm could have been consistently avoided with the use of our elastography technique.

We have demonstrated that intrinsic compression elastography can effectively differentiate benign and malignant thyroid nodules with superior clinical usability in this dissertation. Our research can be considered as an initial step in utilizing this new technique in the management of thyroid nodules. For the wide use of intrinsic compression elastography in routine clinical practice, more clinical evaluation needs to be performed. Currently, the elastography technique described in this dissertation has been commercialized in ultrasound machines and available to clinicians and researchers all over the world. We believe that the utility of intrinsic compression elastography in thyroid nodules management can be further explored by other clinicians and researchers, which would eventually lead to improving the quality of care for patients with thyroid nodules while reducing the healthcare cost associated with managing thyroid nodules.

Bibliography

Altavilla G, Pascale M, Nenci I, Fine needle aspiration cytology of thyroid gland diseases. Acta Cytol 1990;34:251-6.

Asteria C, Giovanardi A, Pizzocaro A, Cozzaglio L, Morabito A, Somalvico F, Zoppo A, US-elastography in the differential diagnosis of benign and malignant thyroid nodules. Thyroid 2008;18:523-31.

Bae U, Dighe M, Dubinsky T, Minoshima S, Shamdasani V, Kim Y, Ultrasound thyroid elastography using carotid artery pulsation: preliminary study. J Ultrasound Med 2007;26:797-805.

Bae U, Kim Y, Angular strain estimation method for elastography. IEEE Trans Ultrason Ferroelectr Freq Control 2007;54:2653-61.

Bercoff J, Tanter M, Fink M, Supersonic shear imaging: a new technique for soft tissue elasticity mapping. Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on 2004;51:396-409.

Cappelli C, Pirola I, Castellano M, Gandossi E, De Martino E, Delbarba A, Agosti B, Tironi A, Rosei EA, Fine needle cytology of complex thyroid nodules. Eur J Endocrinol 2007;157:529-32.

Cappelli C, Rosei EA, Fine needle aspiration cytology of thyroid nodule. Thyroid 2006;16:818.

Castro MR, Gharib H, Thyroid fine-needle aspiration biopsy: progress, practice, and pitfalls. Endocr Pract 2003;9:128-36.

Cespedes I, Huang Y, Ophir J, Spratt S, Methods for estimation of subsample time delays of digitized echo signals. Ultrason Imaging 1995;17:142-71.

Chan HP, Sahiner B, Lam KL, Petrick N, Helvie MA, Goodsitt MM, Adler DD, Computerized analysis of mammographic microcalcifications in morphological and texture feature spaces. Med Phys 1998;25:2007-19.

Chan HP, Wei D, Helvie MA, Sahiner B, Adler DD, Goodsitt MM, Petrick N, Computer-aided classification of mammographic masses and normal tissue: linear discriminant analysis in texture feature space. Phys Med Biol 1995;40:857-76.

Chang JM, Moon WK, Cho N, Kim SJ, Breast mass evaluation: factors influencing the quality of US elastography. Radiology 2011;259:59-64.

Chen H, Shi H, Varghese T, Improvement of elastographic displacement estimation using a two-step cross-correlation method. Ultrasound Med Biol 2007;33:48-56.

Choi SH, Kim EK, Kwak JY, Kim MJ, Son EJ, Interobserver and intraobserver variations in ultrasound assessment of thyroid nodules. Thyroid 2010;20:167-72.

Cooper DS, Doherty GM, Haugen BR, Kloos RT, Lee SL, Mandel SJ, Mazzaferri EL, McIver B, Pacini F, Schlumberger M, Sherman SI, Steward DL, Tuttle RM, Revised American Thyroid Association management guidelines for patients with thyroid nodules and differentiated thyroid cancer. Thyroid 2009;19:1167-214.

Dighe M, Bae U, Richardson ML, Dubinsky TJ, Minoshima S, Kim Y, Differential diagnosis of thyroid nodules with US elastography using carotid artery pulsation. Radiology 2008;248:662-9.

Dighe M, Kim J, Luo S, Kim Y, Utility of the ultrasound elastographic systolic thyroid stiffness index in reducing fine-needle aspirations. J Ultrasound Med 2010;29:565-74.

Duda RO, Hart PE, Stork DG. Pattern Classification. New York: Wiley, 2000.

Fish SA, Langer JE, Mandel SJ, Sonographic imaging of thyroid nodules and cervical lymph nodes. Endocrinol Metab Clin North Am 2008;37:401-17, ix.

Fleiss JL. The design and analysis of clinical experiments. New York: Wiley, 1986.

Frates MC, Benson CB, Charboneau JW, Cibas ES, Clark OH, Coleman BG, Cronan JJ, Doubilet PM, Evans DB, Goellner JR, Hay ID, Hertzberg BS, Intenzo CM, Jeffrey RB, Langer JE, Larsen PR, Mandel SJ, Middleton WD, Reading CC, Sherman SI, Tessler FN, Management of thyroid nodules detected at US: Society of Radiologists in Ultrasound consensus conference statement. Radiology 2005;237:794-800.

Gharib H, Goellner JR, Fine-needle aspiration biopsy of the thyroid: an appraisal. Ann Intern Med 1993;118:282-9.

Gharib H, Papini E, Valcavi R, Baskin HJ, Crescenzi A, Dottorini ME, Duick DS, Guglielmi R, Hamilton CR, Jr., Zeiger MA, Zini M, American Association of Clinical Endocrinologists and Associazione Medici Endocrinologi medical guidelines for

clinical practice for the diagnosis and management of thyroid nodules. Endocr Pract 2006;12:63-102.

Haralick RM, Shanmugam K, Dinstein IH, Textural Features for Image Classification. Systems, Man and Cybernetics, IEEE Transactions on 1973;3:610-21.

Hegedus L, Bonnema SJ, Bennedbaek FN, Management of simple nodular goiter: current status and future perspectives. Endocr Rev 2003;24:102-32.

Hoang JK, Lee WK, Lee M, Johnson D, Farrell S, US Features of thyroid malignancy: pearls and pitfalls. Radiographics 2007;27:847-60; discussion 61-5.

Hong Y, Liu X, Li Z, Zhang X, Chen M, Luo Z, Real-time ultrasound elastography in the differential diagnosis of benign and malignant thyroid nodules. J Ultrasound Med 2009;28:861-7.

Itoh A, Ueno E, Tohno E, Kamma H, Takahashi H, Shiina T, Yamakawa M, Matsumura T, Breast disease: clinical application of US elastography for diagnosis. Radiology 2006;239:341-50.

Jolliffe IT. Principal Component Analysis. New York: Springer-Verlag, 2002.

Kim MH, Luo S, Ko SH, Bae JS, Lim DJ, Kim Y, Thyroid nodule parameters influencing performance of ultrasound elastography using intrinsic compression. AJR Am J Roentgenol 2012. (submitted)

Lerner RM, Huang S, Parker KJ, Sonoelasticity images derived from ultrasound signals in mechanically vibrated tissues. Ultrasound in Medicine & Biology 1990;16:231-39.

Lim DJ, Luo S, Kim MH, Ko SH, Kim Y, Interobserver agreement and intraobserver reproducibility in thyroid ultrasound elastography. AJR Am J Roentgenol 2012;198:896-901.

Lowhagen T, Granberg PO, Lundell G, Skinnari P, Sundblad R, Willems JS, Aspiration biopsy cytology (ABC) in nodules of the thyroid gland suspected to be malignant. Surg Clin North Am 1979;59:3-18.

Luo S, Kim EH, Dighe M, Kim Y, Thyroid nodule classification using ultrasound elastography via linear discriminant analysis. Ultrasonics 2011;51:425-31.

Luo S, Lim DJ, Kim Y, Objective ultrasound elastography scoring of thyroid nodules using

spatiotemporal strain information. Med Phys 2012;39:1182-9.

Lyshchik A, Higashi T, Asato R, Tanaka S, Ito J, Hiraoka M, Brill AB, Saga T, Togashi K, Elastic moduli of thyroid tissues under compression. Ultrason Imaging 2005;27:101-10.

Lyshchik A, Higashi T, Asato R, Tanaka S, Ito J, Mai JJ, Pellot-Barakat C, Insana MF, Brill AB, Saga T, Hiraoka M, Togashi K, Thyroid gland tumor diagnosis at US elastography. Radiology 2005;237:202-11.

MacQueen J. Some methods for classification and analysis of multivariate data. Proc. Fifth Berkeley Symp. on Math. Statist. and Prob., 1967. pp. 281-97.

Mahar SA, Husain A, Islam N, Fine needle aspiration cytology of thyroid nodule: diagnostic accuracy and pitfalls. J Ayub Med Coll Abbottabad 2006;18:26-9.

Mazzaferri EL, Management of a solitary thyroid nodule. N Engl J Med 1993;328:553-9.

Metz CE, Basic principles of ROC analysis. Semin Nucl Med 1978;8:283-98.

Moon HJ, Sung JM, Kim EK, Yoon JH, Youk JH, Kwak JY, Diagnostic performance of gray-scale US and elastography in solid thyroid nodules. Radiology 2012;262:1002-13.

Moon WJ, Jung SL, Lee JH, Na DG, Baek JH, Lee YH, Kim J, Kim HS, Byun JS, Lee DH, Benign and malignant thyroid nodules: US differentiation--multicenter retrospective study. Radiology 2008;247:762-70.

Mudigonda NR, Rangayyan RM, Desautels JE, Gradient and texture analysis for the classification of mammographic masses. IEEE Trans Med Imaging 2000;19:1032-43.

O'Donnell M, Skovoroda AR, Shapo BM, Emelianov SY, Internal displacement and strain imaging using ultrasonic speckle tracking. Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on 1994;41:314-25.

Ophir J, Alam S, Garra B, Kallel F, Konofagou E, Krouskop T, Varghese T, Elastography: ultrasonic estimation and imaging of the elastic properties of tissues. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine 1999;213:203-33.

Ophir J, Cespedes I, Ponnekanti H, Yazdi Y, Li X, Elastography: a quantitative method for

imaging the elasticity of biological tissues. Ultrason Imaging 1991;13:111-34.

Park SH, Kim SJ, Kim EK, Kim MJ, Son EJ, Kwak JY, Interobserver agreement in assessing the sonographic and elastographic features of malignant thyroid nodules. AJR Am J Roentgenol 2009;193:W416-23.

Parker K, Huang S, Musulin R, Lerner R, Tissue response to mechanical vibrations for sonoelasticity imaging. Ultrasound in Medicine & Biology 1990;16:241-46.

Rago T, Santini F, Scutari M, Pinchera A, Vitti P, Elastography: new developments in ultrasound for predicting malignancy in thyroid nodules. J Clin Endocrinol Metab 2007;92:2917-22.

Rausch P, Nowels K, Jeffrey RB, Jr., Ultrasonographically guided thyroid biopsy: a review with emphasis on technique. J Ultrasound Med 2001;20:79-85.

Sandrin L, Fourquet B, Hasquenoph JM, Yon S, Fournier C, Mal F, Christidis C, Ziol M, Poulet B, Kazemi F, Transient elastography: a new noninvasive method for assessment of hepatic fibrosis. Ultrasound in Medicine & Biology 2003;29:1705-13.

Schaefer FK, Heer I, Schaefer PJ, Mundhenke C, Osterholz S, Order BM, Hofheinz N, Hedderich J, Heller M, Jonat W, Schreer I, Breast ultrasound elastography--results of 193 breast lesions in a prospective study with histopathologic correlation. Eur J Radiol 2011;77:450-6.

Shiina T, Doyley M, Bamber J. Strain imaging using combined RF and envelope autocorrelation processing: IEEE, 1996. pp. 1331-36 vol. 2.

Tan GH, Gharib H, Thyroid incidentalomas: management approaches to nonpalpable nodules discovered incidentally on thyroid imaging. Ann Intern Med 1997;126:226-31.

Tranquart F, Bleuzen A, Pierre-Renoult P, Chabrolle C, Sam Giao M, Lecomte P, Elastosonography of thyroid lesions. J Radiol 2008;89:35-9.

Unluturk U, Erdogan MF, Demir O, Gullu S, Baskal N, Ultrasound Elastography Is Not Superior to Grayscale Ultrasound in Predicting Malignancy in Thyroid Nodules. Thyroid 2012 (in press).

van Erkel AR, Pattynama PM, Receiver operating characteristic (ROC) analysis: basic principles and applications in radiology. Eur J Radiol 1998;27:88-94.

Viera AJ, Garrett JM, Understanding interobserver agreement: the kappa statistic. Fam Med 2005;37:360-3.

Wang C, Crapo LM, The epidemiology of thyroid disease and implications for screening. Endocrinol Metab Clin North Am 1997;26:189-218.

Weber AL, Randolph G, Aksoy FG, The thyroid and parathyroid glands. CT and MR imaging and correlation with pathology and clinical findings. Radiol Clin North Am 2000;38:1105-29.

Wong CK, Wheeler MH, Thyroid nodules: rational management. World J Surg 2000;24:934-41.

Yamakawa M, Shiina T, Strain estimation using the extended combined autocorrelation method. Japanese Journal of Applied Physics 2001;40:3872.

Yamakoshi Y, Sato J, Sato T, Ultrasonic imaging of internal vibration of soft tissue under forced vibration. Ultrasonics, Ferroelectrics and Frequency Control, IEEE Transactions on 1990;37:45-53.

Yeung MJ, Serpell JW, Management of the solitary thyroid nodule. Oncologist 2008;13:105-12.

Yoon JH, Kim MH, Kim EK, Moon HJ, Kwak JY, Kim MJ, Interobserver variability of ultrasound elastography: how it affects the diagnosis of breast lesions. AJR Am J Roentgenol 2011;196:730-6.

Youserm DM, Huang T, Loevner LA, Langlotz CP, Clinical and economic impact of incidental thyroid lesions found with CT and MR. AJNR Am J Neuroradiol 1997;18:1423-8.

Curriculum Vitae

Education

09/06 – 12/12 Ph.D., Department of Electrical Engineering, University of Washington at Seattle, WA, U.S.A.

09/03 – 07/06 M.Eng, Department of Electronic Engineering, Tsinghua University, Beijing, P.R. China

09/99 – 07/03 B.Eng, Department of Electronic Engineering, Sun Yat-sen University, Guangzhou, P.R. China

Research/development

09/06 – 12/12 **Research Assistant**, Imaging Computing Systems Lab, University of Washington

Ultrasound elastography in noninvasive differentiation of thyroid nodules using intrinsic compression

- Development of quantitative scoring methods for thyroid ultrasound elastography in improving its clinical usability
- Real-time implementation of developed algorithms on commercial ultrasound machines (licensed to Samsung Medison Co. Ltd.)
- Performing clinical evaluation in University of Washington Medical Center and the Seoul St. Mary's Hospital (Korea)
- Analysis of patients' data in demonstrating the superior performance of the developed elastography method over the traditional approach

Programmable ultrasound system

- Optimal mapping of beamforming, B-mode, spectral Doppler and color flow algorithms on a high performance digital signal processor (Texas Instruments C6455) for a low-cost portable ultrasound system (licensed to Verathon, Inc.)
- Optimal implementation of 3D boxcar filtering and spatial compounding for a commercial ultrasound machine based on IBM CELL processor

09/03 – 07/06 **Research Assistant**, Digital Video and Intelligent Surveillance Lab, Tsinghua University

- Development of background estimation algorithms for automated visual surveillance

Publications

Journal

- **Luo S**, Lim DJ, Kim Y, Objective ultrasound elastography scoring of thyroid nodules using spatiotemporal strain information, Medical Physics 2012; 39(3): 1182-9
- **Luo S**, Kim EH, Dighe M, Kim Y, Thyroid nodule classification using ultrasound

elastography via linear discriminant analysis, Ultrasonics. 2011; 51(4): 425-31

- Lim DJ, **Luo S**, Kim MH, Ko SH, Kim Y, Interobserver agreement and intraobserver reproducibility in thyroid ultrasound elastography, AJR Am J Roentgenol. 2012; 198(4): 896-901
- Dighe M, Kim J, **Luo S**, Kim Y, Utility of the ultrasound elastographic systolic thyroid stiffness index in reducing fine-needle aspirations, J Ultrasound Med. 2010; 29(4):565-74
- Kim MH, **Luo S**, Ko SH, Bae JS, Lim DJ, Kim Y, Thyroid nodule parameters influencing diagnostic accuracy of ultrasound elastography using intrinsic compression, AJR Am J Roentgenol. 2012; submitted.
- Dighe M, **Luo S**, Cuevas C, Kim Y, Efficacy of thyroid ultrasound elastography in differential diagnosis of small sub-cm thyroid nodules, Eur J Radiol. 2012; in revision.
- Kim MH, **Luo S**, Lim DJ, Kim Y, Complementary role of elastography to US B-mode in managing thyroid nodules, in preparation.

Conference

- Lim DJ, **Luo S**, Kim MH, Ko SH, Bae J, Kim Y, Interobserver and intraobserver agreement of thyroid ultrasound elastography using in vivo compression, Annual Meeting of Radiological Society of North America (RSNA), Dec. 2011
- Dighe M, **Luo S**, Kim Y, Inter and intraobserver variability in real-time thyroid elastography using in vivo Compression, Annual Meeting of Radiological Society of North America (RSNA), Dec. 2011
- **Luo S**, Dighe M, Lim DJ, Kim Y, Objective elastography scoring on thyroid nodule, Annual Meeting of Radiological Society of North America (RSNA), Dec. 2010
- Dighe M**, Luo S**, Jordan D, Kim Y, Efficacy of thyroid elastography in differential diagnosis of small < 1 cm thyroid nodule, Annual Meeting of Radiological Society of North America (RSNA), Dec. 2010
- **Luo S**, Kim J, Dighe M, Kim Y, Variability in metric-based ultrasound thyroid elastography, Annual Meeting of Radiological Society of North America (RSNA), Dec. 2009
- Ma J, **Luo S**, Dighe M, Lim DJ, Kim Y, Differential diagnosis of thyroid nodule with ultrasound elastography based on support vector machines, IEEE International Ultrasonics Symposium, pp. 1372-5, Oct. 2010
- **Luo S**, Kim EH, Dighe M, Kim Y, Screening of thyroid nodules by ultrasound elastography using diastolic strain variation. Conf Proc IEEE Eng Med Biol Soc, pp. 4420-3, Oct. 2009
- **Luo S,** Shamdasani V, Xu CX, Managuli R, Kim Y, 3D boxcar filtering on CELL processor, Workshop of solving computational challenges in medical imaging, Seattle, Aug. 2007
- Zhao XL, He W, **Luo S**, Zhang L, MRF-based adaptive approach for foreground segmentation under sudden illumination change, Information, Communications & Signal Processing, 2007 6th International Conference on, pp. 1-4, Dec. 2007
- **Luo S**, Zhang L, A Background model estimation algorithm based on analysis of local motion for video surveillance, Information, Communications and Signal Processing, 2005 5th International Conference on, pp.700-4, Dec. 2005

Patent

- Kim Y, Dighe M, **Luo S**, Kim J, Kim EH, Nodule screening using ultrasound elastography, Appl. No. 12/779,903