College of Engineering

DETECTING THE STAGES OF HYPERPLASIA FORMATION IN THE BREAST DUCTS USING ULTRASOUND B-SCAN IMAGES

Ezgi Taslidere[1], Fernand S. Cohen[1], and Georgia Georgiou[2]

[1]Department of Electrical and Computer Engineering
Drexel University, 3141 Chestnut Street, Philadelphia, PA 19104, USA
[2]European Patent Office, The Hague, The Netherlands

ABSTRACT

A stochastic decomposition algorithm of the RF Echo into its coherent and diffuse components is used towards estimating the structural parameters of the hyperplastic stages of the breast tissue leading to early breast cancer detection. The discrimination power of the various parameters is studied under a host of conditions such as varying resolution and SNR values using a point scatterer model simulator that mimics epithelium hyperplastic growth in the breast ducts. It is shown that three parameters, in particular, the number of coherent scatterers, the Rayleigh scattering degree and the energy of the diffuse scatterers, prove to show very high ability to discriminate between various stages of hyperplasia even in cases of low resolution and SNR values. Values of Az>0.942 were obtained for resolution less than or equal to 0.4mm even in low SNR values, then it drops below the 0.9 range as the resolution exceeds the 0.4mm range.

1. INTRODUCTION

Breast cancer is the second cause of death in women behind lung cancer. One out of eight women will be diagnosed with this deadly disease. Recent studies show that only certain very specific types of microscopic changes put a woman at higher risk according to the National Cancer Institute (NCI). These changes feature excessive cell growth, or hyperplasia. The detection of this growth at its early stages is crucial in terms of early breast cancer detection. *Early-stage* invasive cancer is considered very treatable because the tumor is relatively small and the cancer cells have not spread to the lymph nodes. However, when a tumor has become very large or has spread to other organs (such as the liver, lungs, or bones), it is considered *advanced-stage* invasive cancer and is far less treatable. 90% of cell growths causing breast cancer originate in the ductal epithelium, where they produce ductal carcinoma in situ (DCIS) or, when invasive, infiltrating ductal carcinoma [1].

Researchers have been investigating procedures for sampling breast duct tissue to search for cancer. Procedures used include either Breast Duct Lavage, which is a method used to collect cells from milk ducts in the breast to check for cancer, or Needle Biopsy, which is the removal of tissue or fluid with a needle for examination under a microscope. In order to prevent this kind of painful procedures performed on patients, ultrasonic imaging can be used to differentiate between several stages of hyperplastic evolution in the breast tissue. However, the diagnostic value of B-scan images is limited because of the low resolution of the imaging system with respect to the tissue structures. To improve the usefulness of ultrasound in early diagnosis of cancer, significant attention needs to be focused on developing quantitative methods for extracting additional information from the returned RF echoes for tissue characterization.

In this work, we aim to detect different stages of cell growth in the ducts using features extracted from ultrasound RF Echo. Since there is no database of ultrasound images available where the cases of hyperplasia of different stages are present, we simulate the RF echo reflected from the breast ducts. We extract tissue related parameters using a tissue characterization method, namely the Continuous Wavelet Transform Decomposition (CWTD) algorithm [2] and then we classify these estimated parameters to differentiate between various stages of hyperplastic evolution in the breast tissue. To our knowledge, our technique is the first attempt to detect the cell growth in breast ducts using ultrasound. The ultimate goal of this work is to use reflected RF echo to get additional information for classification of various stages of hyperplastic evolution, when it is not visually obvious in a conventional B-mode ultrasound imaging. The work in this paper uses our previous work [10], by studying which of these extracted parameters are specifically geared towards capturing hyperplastic growth in ductal epithelium and analyzing why they do.

We show the results on simulations that address the issue of early detection of hyperplastic cell formation as a direct support to our research that clearly clarifies the link and correlation that exist between the evolution of hyperplastic growth in ductal epithelium and the signal decomposition backscatter echo parameters. As indicated by the simulation results, some of these parameters turn out to be very strong parameters to track the evolution of hyperplasia and to differentiate between its various stages. Our results verify the potential of our system to discriminate between various stages of hyperplastic evolution in the breast tissue. The results are presented in the form of Receiver Operating Characteristics (ROC) curves.

The paper is organized as follows. Structure of breast duct, and hyperplasia formation in the breast ducts are presented in Section 2. Section 3 presents the details of the simulation of the RF echo reflected from the breast ducts. The CWTD algorithm used for tissue related parameter extraction and the features related to the anatomy of the breast ducts are described in Section 4. The results of the classification performance of these features for the detection of various stages of hyperplastic evolution are given in Section 5. We conclude the paper in Section 6.

2. STRUCTURE OF BREAST DUCT AND HYPERPLASIA FORMATION

The breast ducts are tubes which, following pregnancy carry breast milk to the nipple. Hyperplasia is the term describing the excessive accumulation (or proliferation) of cells. It is usually found on the inside of the lobules or the ducts in the breast tissue. We concentrate on the hyperplasia formation inside of the ducts, because over 90% of breast cancers start at the duct before infiltrating the surrounding tissue [1]. As cancer grows, the ductal epithelium undergoes hyperplastic growth (as seen in Figure 1). Hyperplasia is associated with approximately a double risk for breast cancer. Atypical hyperplasia is associated with an approximately 4-5 times increased risk of developing breast cancer, and ductal carcinoma in situ is associated with approximately 8-10 times increased risk of developing breast cancer [3].

Although there is some disagreement as to when histological changes can be defined as cancer, it is generally accepted that atypical hyperplasia is a precancerous entity [4]. Ductal carcinoma in situ and lobular carcinoma in situ are referred to as noninvasive because the cells do not infiltrate the surrounding tissues. The four stages of transformation from a noncancerous condition, which we called as stage 1 to a cancerous condition, stage 5, are depicted in Figure 1. We numbered the stages of hyperplasia formation inside the breast ducts starting from the normal duct. Stage 1 represents the *normal duct*. Stage 2 represents *hyperplasia* in which there is an increase in the production of cells inside the duct. Stage 3 represents *atypical hyperplasia* which is a condition of the cells in the lining of the milk ducts in the breast. Cells have abnormal features and are increased in number. Stage 4 represents *ductal carcinoma in situ* for which there are abnormal cells in the breast ducts, which over time could develop into breast cancer. Stage 5 represents *infiltrating ductal carcinoma* which means that a cancer that has started to invade the tissues surrounding it. The first three stages of hyperplasia formation are reversible [1] which makes the detection of early stages extremely important.

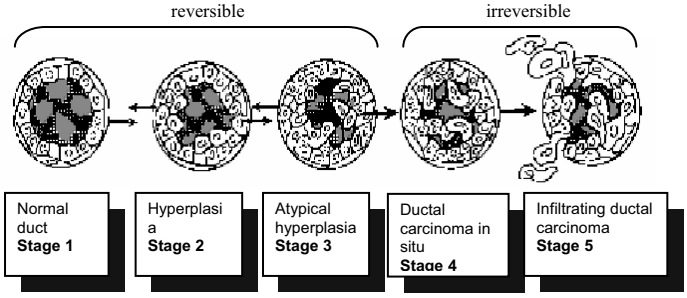


Fig. 1. The stages of hyperplastic evolution [5].

3. SIMULATION OF REFLECTION OF RF ECHO FROM BREAST DUCTS

Based on the information on the structure of breast ducts, the ultrasound RF echo simulation is performed. The simulator is based on a point scatterer model driven by the anatomy of the hyperplastic stages in the breast, and as such it mimics the epithelium hyperplastic growth in the breast ducts. The point scatterer model is described in detail in [2]. Under the assumptions of weak scattering, narrow ultrasound beam, and linear propagation, the point scatterer model is an adequate model and has been widely used in the literature [2], [5], [6-8]. The RF Echo *y(t)* is modeled as the interaction of the pulse with the scattering structure,

$$y(t) = h(t) * s(t) \tag{1}$$

where *s(t)* is the scatterer distribution, and *h(t)* is the ultrasound pulse shape. Scatterer distribution can be represented as the sum of diffuse and coherent components:

$$s(t) = \sum_{n=1}^{N_s} d_n(t-\tau_n) + \sum_{n=1}^{N_c} c_n(t-\theta_n) \tag{2}$$

where N_c is the number of coherent scatterers; N_s is the number of unresolvable scattering centers; τ_n and θ_n are time delays from point scatterers to the receiver; and $c_n(t-\theta_n)$ and $d_n(t-\tau_n)$ are the relative strengths of the point scatterers. The RF echo *y(t)* is sampled, resulting in an A-line *y(n)*. The resolvable scattering structure can be viewed as a summation of delta functions, located at the resolvable scatterers' locations, and of random strength [4] [9]. The transducer pulse is well approximated by a Gaussian modulated sinusoid [4]. The mathematical description of the pulse and tissue interaction is the convolution.

In the simulation presenting the reflection of RF echo from breast ducts, the following parameters are considered: the scatterer densities; the scatterer spatial/temporal distributions; the central frequency of the pulse; the SNR (ratio of average coherent to diffuse energy); sampling frequency; propagation velocity of sound in tissue; and resolution. The parameters of the simulator are within the possible range of ultrasound. The breast duct is assumed to be 2mm in diameter which is the average size of a duct.

Here, we consider the diffuse and coherent components of the ultrasonic RF echo reflected from the breast ducts. At this point, the physical structure of breast ducts is crucial for the simulation of the reflected echo. The echoes of the ducts boundaries and the cells inside the ducts are resolvable. The coherent scatterers are known to be the resolvable part. As a result, coherent scatterers are generated at the duct boundary and the cells inside the ducts. As the duct goes under hyperplastic growth, the anatomy of the duct changes, the number of cells inside the duct increases affecting the number of coherent scatterers. The remaining scatterers produce the non-resolvable part which is known as diffuse scatterers. In our simulation setting, the diffuse scatterers are uniformly distributed throughout the beam of the A-line. In order to obtain the RF echo reflected from the breast ducts, we first created the diffuse and coherent scatterers based on statistical distributions then we add them as in Equation 2, then we convolve the result with the transducer pulse *h(t)*. The statistics of the number of diffuse scatterers per resolution cell is presented using a Poisson distribution. The mean value of the distribution is taken to be 32 scatterers per resolution cell for outside the hyperplastic cells, and 48 for inside the hyperplastic cells. The distribution of the strength of scatterers is approximated by a Γ-probability density function (pdf) with a different set of parameters. The ratio σ/μ where μ is the mean and σ is the standard deviation of the Γ-pdf, controls the regularity of the samples. Large σ/μ means that the strength of the coherent or the diffused scatterers is very irregular whereas small σ/μ means that it is regular. Since coherent scatterers show more regularity than the diffuse scatterers, the ratio σ/μ is taken as 0.1 for coherent scatterers whereas the ratio is taken as 0.2, 0.35, and 0.4 for diffuse scatterers inside the cells, outside the cells, and outside the duct respectively. The pulse center frequency is taken as 7.5MHz, the sampling rate is 20MHz, and the propagation velocity in the tissue is 1540m/s. The simulation is performed for

various values of SNR which are 0, -1.25, -3, -4, -6, -8, and -10 dB. Also, the resolution is taken to be 0.25, 0.4, 0.6, and 0.8mm. For each stage 60 different realizations of the reflected RF echo are generated based on the randomness of the coherent and diffuse scatterer parameters.

4. DECOMPOSITION METHOD AND THE FEATURES RELATED TO THE ANATOMY OF BREAST DUCT

WOLD-decomposition theorem [6] is used to decompose the RF signal *y(n)* into its two components where diffused component *d(n)* is modeled as a zero-mean autoregressive (AR) process of order 5 driven by a zero-mean white noise sequence *w(n)* (not necessarily Gaussian) with variance σ^2 [6] [8], and the coherent component *c(n)* is modeled by a superposition of Gaussian modulated sinusoids [2]. The decomposition of RF echo is achieved by the continuous wavelet transform (CWT) and was thoroughly described and tested on simulated RF data in [2] [10]. The CWTD algorithm first checks the existence of coherent scatterers in the window of tissue under examination by testing the hypothesis of Rayleigh scattering. The rejection of the Rayleigh scattering hypothesis indicates the existence of a coherent component. If a coherent component exists, RF echo *y(n)* is decomposed into its diffuse *d(n)* and coherent *c(n).* After the decomposition, the CWTD algorithm estimates concurrently parameters related to the number and energy of the resolvable scatterers, the energy of the unresolvable scatterers, and the correlations between neighboring diffuse scatterers. If the test of presence of a coherent component fails, the diffuse component *d(n)* is the RF echo *y(n)* itself and modeled as an AR process. Next, the two components (or the one in the case of no coherent component) are used to estimate the features used in the tissue characterization.

When hyperplasia formation begins, the duct structure changes, the features, estimated from the RF signal components, should reflect those changes so that the discrimination between various stages of hyperplastic growth in ductal epithelium in breast tissue can be possible. In the following, we explain a number of features estimated from the coherent and diffuse components separately [10]. The feature related to the coherent component is explained as follows:

Number of detected coherent scatterers, N_c: The number of coherent scatterers is related to the number of resolvable scatterers in the tissue (boundaries of ducts and cells inside the duct). This parameter is directly related to the formation of hyperplastic cell and as the resolution decreases in size so would the ability of that parameter to differentiate between the various stages of hyperplasia formation. This parameter was shown to possess weak discrimination ability between benign, normal and malignant tissues, however, in our case; since the resolution is higher, it turned out to be a very strong parameter for discrimination of hyperplastic stages. We observe that number of detected coherent scatterers goes down as the resolution decreases.

Features related to the diffuse component are explained as follows:

Residual error variance of the diffuse component (energy of the diffuse component), σ^2: The unconditional variance is the equivalent of the energy of the coherent scatterers for the diffuse component of the RF echo. It is related to the residual error variance (conditional variance of the diffuse component) through the AR parameters, the values of the unconditional variance are reflected mainly on the residual error variance. This parameter was shown to possess excellent discrimination ability between benign, normal and malignant tissues. It remains a strong parameter to differentiate between early and late stages of hyperplasia.

Rayleigh scattering degree of the diffuse component, *D*: This parameter describes the discrepancy between the empirical distribution of the innovation process of the diffuse component and the Gaussian distribution. The lower the value of *D* the closer to Rayleigh scattering the diffuse component is. We anticipate this parameter to play a major role in differentiating between the various stages of hyperplasia formation as it directly relates to formation of new hyperplastic cells and the increase in the hyperplastic density that directly reflects on that Kolmogorov Smirnov (KS) distance *D*.

5. CLASSIFICATION OF HYPERPLASTIC EVOLUTION STAGES

We simulate 60 different realizations of the reflected RF echo for each hyperplastic evolution stage and obtain 300 realizations for five stages. We calculated three features, N_c, D, σ^2 for each realization. The classification performance of each feature is evaluated separately using a quadratic classifier. N_c, D, σ^2 performed well for discrimination between the hyperplastic stages of the breast tissue.

Table 1. A_z values for pairs of stages (Resolution=0.25mm, SNR=0dB)

	Stage1	Stage2	Stage3	Stage4	Stage5
Stage1	-	0.948	1.000	1.000	1.000
Stage2	0.948	-	0.999	1.000	1.000
Stage3	1.000	0.999	-	1.000	1.000
Stage4	1.000	1.000	1.000	-	1.000
Stage5	1.000	1.000	1.000	1.000	-

The features N_c and D showed outstanding discrimination ability, while σ^2 showed good performance for the discrimination between the early and late stages of hyperplasia (i.e., stage1 vs. stage5). The two features, N_c and D are combined in order to increase the discrimination ability. Table 1 shows the results in terms of the area under the ROC curves for the combined parameters. Each value in the table presents the area under the ROC curve for classification between pairs of stages. In this work, we define SNR as the ratio of the average energies for diffuse and coherent scatterers. In order to evaluate the overall performance of the system, the ROC curves for each pairs of stages are averaged to obtain the overall ROC curve. The overall performance for various SNRs is shown in Table 2. As SNR decreases the performance decreases as well. Even in the case of low SNR, the performance is reliable. Table 3 shows the results in terms of the area under the ROC curves for various resolutions for the combination of parameters N_c and D and for only the parameter N_c. We observe that the A_z values are very high even for low SNR for resolutions up to 0.4mm, then it drops below the 0.9 range for resolution below that. We observe as expected (Section 4) the performance of the N_c goes down from 0.921 to 0.740 as the resolution decreases.

Table 2. A_z values (Resolution=0.25mm)

SNR(dB)	A_z
0.00	0.997
-1.25	0.995
-3.00	0.991
-4.00	0.988
-6.00	0.984
-8.00	0.983

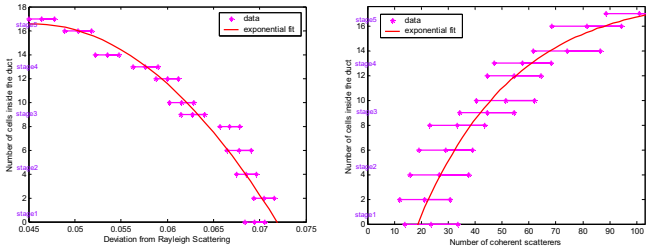
Table 3. A_z values (SNR=0dB)

Resolution(mm)	A_z (N_c&D)	A_z (N_c)
0.25	0.997	0.921
0.40	0.942	0.892
0.60	0.863	0.848
0.80	0.842	0.740

Table 4 shows the performance of σ^2 for the discrimination between the early and late stages of hyperplasia (i.e., stage1 vs stage5). The feature σ^2 did not perform extremely well for discrimination between all the pairs of stages, however it showed good performance for discrimination of the early and late stages of hyperplasia. The features N_c and D are studied in depth in order to see the direct effect of the hyperplasia evolution on the features. These features are directly related to density of cells forming as a result of hyperplasia. All the steps of hyperplastic evolution representing the increase in the density of cells are simulated. For each step, 60 realizations are obtained. The mean value of the features, N_c and D, are plotted as a function of the density of the hyperplastic cells. An exponential function is fitted on the values. The plots are shown in Figure 2. We observe that the number of coherent scatterers increases as the number of cells increases, whereas the *KS* distance D decreases as the number of cells increases since we approach the Rayleigh scattering. The bars in the fitted exponential are one standard deviation away from the mean of the N_c and D as a function of the hyperplastic cells.

Table 4. A_z values for discrimination between early and late stages (SNR=0dB, resolution=0.25mm)

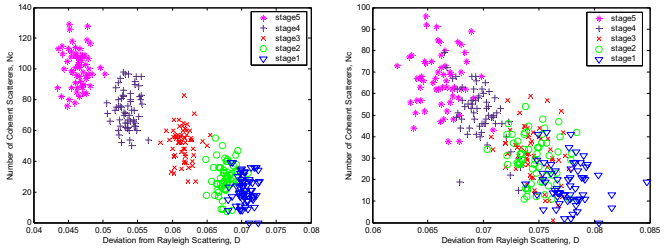
Pairs of stages	A_z (σ^2)
Stage1 vs Stage2	0.837
Stage1 vs Stage4	0.878
Stage1 vs Stage5	0.903



(a) (b)

Fig. 2. Exponential fit for (a) D and (b) N_c versus the number of hyperplastic cells inside the breast duct.

We also provide data scatter information on how these parameters change as hyperplasia is developing through its stages. The data scatterer plot is shown in Figure 3 for the resolutions 0.25mm and 0.40mm. As can be seen from Figure 3, features are well separated for different stages for two resolutions. For high resolution, features are perfectly separated (Figure 3-a).



(a) (b)

Fig. 3. Data Scatter for N_c and D for stages 1 through 5 for resolution (a) 0.25mm. (b) 0.40mm.

6. CONCLUSION

What makes this work unique is its portability and systematic way of extracting image parameters that closely correlate with the various structures of the various stages of hyperplastic growth in ductal epithelium in breast tissue, as well as establishing formal ways of accessing the discrimination capability of these parameters in various stages. Another important aspect is the ability to extract structurally related parameters that go well beyond those that can be visibly extracted from the ultrasonic scan. Therefore, it presents a potential for a complimentary tool to the radiologist's assessment and might prove to even exceed that assessment especially in the early formation of cancer where the ductal epithelium undergoes hyperplastic growth. Moreover, our work shows a potential for reduction of biopsies and detection of precancerous entities at the reversible stages depending on its performance on the simulated data.

As a conclusion, our study indicates that the classification between various stages of hyperplasic evolution in the breast tissue can be done reliably using the number of coherent scatterers N_c and the *KS* distance D based on the simulations. The results demonstrate the effectiveness of the features even in cases of low resolution and SNR values. We report on the ability of the parameters to differentiate between the various hyperplasia stages (A_z>0.942 for resolution less than or equal to 0.4mm).

7. REFERENCES

[1] D. B. Kopans, *Breast Imaging*. J. B. Lippincott Company, 1998.

[2] G. Georgiou and F. S. Cohen, "Tissue characterization using the continuous wavelet transform Part I: Decomposition method," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 48, no. 2. pp. 355-363, Mar. 2001.

[3] Bland KI, and Copeland EM, *The Breast: Comprehensive Management of Benign and Malignant Disorders*. W.B. Saunders Company, 3rd edition, 2003.

[4] Institute of Medicine Staff, *Review of the Department of Defense's Program for Breast Cancer Research.* National Academy Press, 1997

[5] K. D. Donohue. F. Forsberg, C. W. Piccoli, and B. B. Goldberg, "Classification of breast masses with ultrasonic scattering structure templates," *IEEE Trans. Ultrason. Ferroelect., Freq. Contr.*, vol.46, no. 2, pp. 300-310, Mar. 1999.

[6] F. S. Cohen, G. Georgiou, and E. Halpern, "WOLD decomposition of the backscatter echo in ultrasound images of soft tissue organs." *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 44, no. 2, pp. 460-472. Mar. 1997.

[7] R. F. Wagner. M. F. Insana, and D. G. Brown, "Unified approach to the detection and classification of speckle texture in diagnostic ultrasound." *Opt. Eng.*, vol. 25, pp. 738-742, Jun. 1986.

[8] G. Georgiou and F. S. Cohen, "Statistical characterization of diffuse scattering in ultrasound images," *IEEE Trans. Ultrason,, Ferroelect., Freq. Contr.*, vol. 45. pp. 57-64, Jan. 1998.

[9] L. Landini and L. Verrazanni. "Spectral characterization of tissue microstructure by ultrasounds: A stochastic approach," *IEEE Trans. Ultrason., Ferroelect,, Freq. Contr.*, vol. .37, no. 5, pp. 448-456. Sep. 1990.

[10] G. Georgiou. F. S. Cohen, C. W. Piccoli, F. Forsberg, and B. B. Goldberg, "Tissue characterization using the continuous wavelet transform, Part II: Application on breast RF data," *IEEE Trans. Ultrason., Ferroelect., Freq. Contr.*, vol. 48, no. 2, pp.364-373, Mar. 2001.