Differentiation of the breast lesions using statistics of backscattered echoes

Hanna PIOTRKOWSKA-WROBLEWSKA¹, Katarzyna DOBRUCH-SOBczAK¹,², Jerzy LITNIEWSKI¹, Eryk CHRAPowICKI², Katarzyna ROSZKOWSKA-PURSKA², Andrzej NOWICKI¹

¹Institute of Fundamental Technological Research, Department of Ultrasound
Pawinskiego 5b, 02-106 Warsaw, Poland
hpiorz@ippt.pan.pl
²Center of Oncology Memorial Institute
Wawelska 15B, 02-034 Warsaw, Poland

The purpose of this study was to evaluate the accuracy of statistical properties of the backscattered ultrasound in differential diagnosis of the breast lesions. The B-mode images, together with the appropriate RF echoes from the breast lesions and surrounding tissues were collected. The RF data was processed for the statistics of the backscattered echo signals, using K and Nakagami distributions characterized by the \( M \) and \( m \) parameters, respectively. Based on both, \( M \) and \( m \) parameters, a set of 18 parameters was derived.

From the point of view of the sensitivity of detection of cancer, the best score was obtained using maximum value of \( M \) parameter, the best specificity was received using the differential Nakagami parameter (the differential values between lesions and surrounding tissues). In conclusion, quantitative sonography is a method which has potential to be a complementary tool for classification of the breast lesions.

**Keywords:** quantitative ultrasound, breast cancer, Nakagami distribution, K distribution

1. Introduction

Breast cancer is the most frequently diagnosed cancer among women between 40 and 60 years old. According to the report of the World Health Organization WHO, 0.5 million women die each year because of it. In Poland, data from 2010 indicates that there were 15,784 new cases of breast cancer, and 5226 deaths of women [1]. In the world, breast cancer is also considered as the most widespread neoplasm – based on the reports, as much as 34% of women with neoplasm are suffering because of breast cancer. This number puts the breast cancer in first place among the causes of mortality in middle-aged women.

Lowering the scale of the disease incidence can be achieved by prevention, early detection of lesions, as well as proper treatment [2]. According to the American Cancer
Society, the mammographic examination is now one of the main screening imaging methods used for diagnosis of breast cancer. Its use has reduced the mortality of women with breast cancer. However, the sensitivity of the mammography in the breast glandular structure is 30-48 % and 80-90 % in women with fatty breasts [3, 4]. About 50 % of women under 50 years old and 30 % of women over 50 years old have glandular breasts. In addition, this type of breast tissue quadruples the risk of breast cancer compared to women with fatty breast tissue. Especially in that case, ultrasonic examination seems to be recommended for the diagnosis of the breast lesion. To standardize the US investigation, The American College of Radiology has proposed the classification BI-RADS [5]. BI-RADS take into account features of the lesion (among others: echogenicity, orientation, shape, margin, elasticity assessment or vascularity) and based on them, allows prediction of the probability of malignancy of a nodule. When a lesion is classified as BI-RADS 4a, b, c or 5, the biopsy and histological examination is required. However, the proper classification of the breast nodule in many cases is very hard, because it strongly depends on the quality of the applied US scanning machine and the radiologist’s experience.

Recently, there is an increased interest in applying the quantitative ultrasound techniques, allowing us to make the ultrasonic examination less dependent on the operator’s experience, and in some cases (especially in the group of lesions category 4a), reducing the number of unnecessary biopsies.

The main trend of the ultrasound research refers to classify the breast tumors using the features determined from the texture of the B-mode images. Chen et al. used the fractal features with accuracy of 88.8%, Wu and Moon who used morphologic features, reported accuracy, specificity and sensitivity of 92.8%, 91.67% and 94.44%, respectively [6, 7]. Zhang et al. indicated that it is possible to distinguish malignant and benign lesions with accuracy of 91.0% (±3.8%) using texture features extracted by a shearlet transform [8]. However, methods based only on the features determined from the ultrasonographical images have some limitation because part of the information, encoded in the backscattered echo signals, is lost in the pre/post-processing of B mode images. In may be stated, that the classification based on the B-mode images only does not always provide an adequate and unambiguous diagnosis.

The quantitative techniques, which are based on the raw (RF) US signals, seem to be a first step on the way to develop the methods which are able to deliver more information on the microstructural properties of the soft tissues, and to improve the accuracy of identification of the tissue type.

The analysis of the envelope of the backscattered signal is an important approach to the modeling of the tissue as a medium with a stochastically variable density/speckle pattern related to the specific tissue. The statistics of the echo depend on both, medium structure and spatial resolution (resolution cell) of the scanning. Resolution cell size varies with probing pulse length, and acoustic beam cross section. The soft tissue is often modeled as a collection of spatially distributed small scatterers. It can be shown that for the large number of scatterers, uniformly distributed within the resolution cell, the signal amplitude statistics follow the Rayleigh distribution. However, in biological tissue Rayleigh’s conditions are rather not fully-fulfilled.

In the case of morphological analysis of the benign and malignant breast lesions, many scattering variants should be considered. Malignant lesions are characterized by very rich cellularity, which means that the cancerous cells may have different shapes, and they may create bigger scattering centers (clusters). They may also contain stromal components, calcifications, or irregular blood vessels, which may provide the source of a coherent signal.
These features result in more complex scatterers’ composition in malignant tumours, exhibiting specific signatures (statistical distributions) in backscattered echoes. In the case of benign lesions, the homogeneity is not always observed. Two of the most common benign lesions are fibrocystic changes, and fibroadenoma consisting of the epithelial and stromal components. The epithelial component of fibroadenoma presented varying degrees of proliferation. The stromal component can show sometimes focal or diffuse hypercellularity, bizarre giant cells, hyalinization or calcification.

Some of them are fibroadenoma complex containing cysts, adenosis sclerosans, epithelial calcifications and fibrosis, but generally benign lesions have more regular arrangement of the cells [9, 10]. Therefore, there is no established standard of the scatterer-arrangement in benign and malignant lesions.

Also, the scatterer density and size vary throughout the tissue. Thus, the effective number of scatterers in the resolution cell may not fulfill Rayleigh statistics, particularly when the resolution cell is small. In this case different non-Rayleigh statistics are considered. Nakagami distribution has been the most frequently adopted model, probably due to its simplicity. The Nakagami model was then systematically used in various medical ultrasound imaging fields; ophthalmology, vascular and breast cancer [10-13].

Molthen et al. showed that, in the case of the breast tissues, the Rayleigh distribution commonly used in statistical analysis of ultrasound signals, is not appropriate [14]. They decided to use the K distribution as a function which is much more applicable in the case of modeling statistics of backscattered RF signals from the breast. A similar approach was presented by Shankar et al [15-17]. They applied the K and Nakagami distributions in order to distinguish benign changes from the malignant lesions occurring in breasts.

The Nakagami distribution, besides providing purely quantitative information, was applied also as a tool for the development of the so-called Nakagami images. Procedures for developing the images of this type were developed, and next applied, by Tsui et al [18-19]. It also provides more precise determination of the contours of lesions [20].

In recent paper Shankar has defined a new model for tissue with microcalcifications. He modified the Nakagami model to the McKay type-I density function for speckle factor exceeding 2, so for the situation approaching breast calcification [21].

The goal of this study was to find the quantitative measure of the breast tissue backscattering properties, for differentiating the changes of tissue structure induced by benign and malignant breast lesions. The approach presented here is based on the shape parameters of the K and Nakagami distributions.

2. Materials and methods

2.1 Patients

The project was approved by the Ethical Review Board of the Medical University of Warsaw; and informed consent from all patients participating in the study, for the scientific analysis of their imaging data, was obtained before the examination.

All patients were sent to us from the Cancer Center, and from the Brodnowski Hospital in Warsaw, after the conventional US examinations, which were performed by experienced radiologist. Ultrasound images of 107 focal breast lesions in 78 women aged between 24 years and 75 (average age 49.2) were assessed. 32 lesions were verified as malignant lesions, and 75 were verified as benign. The most common neoplasm was the invasive ductal carcinoma (n=12), followed by mixed (n=6) and other carcinomas (n=14). Four patients had multifocal malignant lesions. In the group of the benign lesions the most common lesion was
fibroadenoma (n=53) followed by fibrocystic dysplasia (n=9) and others (n=13). Each lesion was histologically assessed by the core needle biopsy or fine-needle aspiration biopsy.

2.2 Data Acquisition

The ultrasound radio-frequency signals were acquired using Ultrasonix SonixTouch-Research ultrasound scanner (Analogic Corporation Peabody, MA, USA), with the raw RF data digitalized at a 40 MHz sampling rate. Breast tissues were examined in vivo using a 7.5 MHz linear array transducer. 78 women with diagnosed benign and malignant breast lesions participated in this study (70 and 37 cases, respectively). Two types of data: a traditional B-mode image, and a set of RF echo-lines, were acquired for the breast regions where the tumors were localized. The breast US examinations were performed according to the American College of Radiology BIRADS guidelines, and Polish Ultrasound Society guidelines, using radial scanning around the nipple and perpendicular direction [22]. For all the patients, the reference data from the healthy regions of the breast tissues were also recorded. The examples of B-mode images of breast lesions in two sections: horizontal and vertical, are shown in Fig. 1. The received sequences were first logcompressed, and next envelope detected, and displayed. One image consists of up to 500 RF echo lines. The number of samples in every RF signal was depending on the chosen depth of examination. Once the RF data was collected, the further processing and analysis was done off-line, using Mathcad (Mathsoft PTC, USA) software.

![Fig. 1. B-mode images in radial plane (a) and antiradial plane (b) images of the carcinoma ductale.](image)

2.3 Data preprocessing

Attenuation in the examined tissue results in the decay of the wave amplitude with depth. Therefore, the influence of the attenuation on RF echoes was compensated, prior to the statistical properties of the breast backscatter. Two different mean attenuation slopes $\alpha_o$ values, equal 0.7 dB/(cm·MHz), and 0.5dB/(cm·MHz), were used for attenuation compensation, in lesion and healthy tissues, respectively. To compensate amplitude decay, the following algorithm was used: first, the spectrum (FA) of the attenuated signal was calculated. Next, the synthesis of a resulting signal ($F(.)$) on the basis of the spectral components of the backscattered signal was performed. During the synthesis, the amplitudes of spectral components are increasing with the depth $d=vt_i$. This process is described according to the formula:

$$F(t_i) = \sum_k FA_k \exp(\alpha_0 f_k v t_i) \exp -(j2\pi f_k t_i)$$

(1)

where $k$ stands for the index of the spectral component, $f_k$ denotes frequency, $FA_k$ is a complex spectrum of backscattered signal, $v$ denotes phase velocity of the longitudinal acoustic wave in the skin and $\alpha_0$ is the attenuation coefficient, $t_i = i \delta t$ stands for time, where
\( \delta t \) is a time step given by the signal sampling rate. The real part of \( F(.) \) is the desired backscattered signal compensated for attenuation.

For tumors with calcifications, clear echogenic echoes (bright spots in the B-mode image) occur. These types of high amplitude echoes may considerably modify the statistics of the envelope. Therefore, in the next step, the samples of the signal exceeding 95% of the maximum signal amplitude (reflected from microcalcifications), were removed from the analysis.

After attenuation compensation, and high echoes filtration, the envelope of the signal was computed using the Hilbert transform method.

### 2.4 Statistics of echo envelope

Statistics-based parameters were estimated by applying K and Nakagami distributions to the RF data. K distribution is characterized by the \( M \) parameter related to the effective scatterers’ density, and Nakagami distribution is characterized by the shape parameter \( m \).

\( M \) parameter was calculated from the second and fourth order moments of RF echoes’ amplitude values:

\[
M = \frac{2}{r_4 - 2}
\]  

where \( r_4 = \frac{\mathbb{E}[A^4]}{\mathbb{E}[A^2]^2} \), \( \mathbb{E}[A^4] \) and \( \mathbb{E}[A^2] \) are the fourth and second order moments of RF echoes’ amplitude values, respectively.

The shape parameter of the Nakagami distribution \( m \) was calculated using the Lorents estimator.

\[
m = \frac{4.4}{\sqrt{u_2 - (u_1)^2}} + \frac{17.4}{(u_2 - (u_1)^2)^{1.29}}
\]  

where

\[
u_k = \frac{1}{N} \sum_{i=1}^{N} (20 \log_{10} A_i)^k
\]  

The correct estimation of the shape parameters of the probability density functions requires a large number of data samples. The minimum size of sampling space resulting in size-independent statistics, was found from the data recorded in a homogeneous tissue-mimicking phantom. A set of concentric ROI-like squares was selected, and each square was analyzed for their respected \( M \) and \( m \) parameters. It was found that a 3 mm × 3mm ROI box, corresponding to a 2.8s long time window with 28 adjacent scan lines, is sufficient to obtain reliable results. The total number of samples in such a data set was not less than 4200 (150 samples along each line).

In the next step, each lesion was divided into several sub-ROIs covering the lesion. The number of sub-ROIs depended on the size of the lesion. The lesions which were smaller than 10mm² were not taken under consideration. The shape parameters were calculated for each sub-ROI separately. These steps were done for each lesion, in two orthogonal (radial and antiradial scans). In the same way, the shape parameters were calculated in the “healthy” region surrounding the lesion. The subsequent calculations steps are shown in Fig. 2.
2.5 Statistical analysis

Statistical evaluation was performed using software Statistica ver.10 (StatSoft Ltd., USA). The reference standards were histopathological or cytopathological findings. The values of 18 parameters were considered, using three statistical tests: t-Student, Cochran-Cox and Mann-Whitney, as markers useful in differentiating the malignant and benign breast lesions.

For the parameters exhibiting the statistical difference ($p<0.05$), between the benign and malignant lesions, the sensitivity, specificity and accuracy were calculated. The specificity and sensitivity are defined as:

$$SENSITIVITY = \frac{TP}{TP + FN}$$

$$SPECIFICITY = \frac{TN}{TN + FP}$$

where, $TN, TP, FP, FN$ mean true negative (correctly rejected), true negative (correctly identified), false positive (incorrectly identified) and false negative (incorrectly rejected) and indicate the percentage of lesions which were correctly indicated as a true positive (in the case of the presented results, the real malignant) and true negative rate (the percentage of benign lesions, which were correctly identified as benign).

The accuracy is a measure of precision (true results) and is defined as:

$$ACCURACY = \frac{TP + TN}{TP + TN + FP + FN}$$

The optimal cut-off value was determined by using the Youden index (maximum sum of sensitivity and specificity -1).
3. Results

The results obtained showed that the differences in the values of the shape parameters determined for the radial and antiradial scans were not statistically significant (p>0.05, t-Student test). In consequence, the values of the shape parameters that were estimated for each lesion in both planes, were considered as one set of data. Similar results were obtained for M and m values for the surrounding tissues. The set of 18 parameters defined in Table 1 was calculated and analyzed for each lesion and the surrounding tissues.

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>maxML</td>
<td>maximum value of the M parameter determined for lesion</td>
</tr>
<tr>
<td>minML</td>
<td>minimum value of the M parameter determined for lesion</td>
</tr>
<tr>
<td>avgML</td>
<td>average value of the M parameter determined for lesion</td>
</tr>
<tr>
<td>maxMS</td>
<td>maximum value of the M parameter determined for surrounding tissues</td>
</tr>
<tr>
<td>minMS</td>
<td>minimum value of the M parameter determined for surrounding tissues</td>
</tr>
<tr>
<td>avgMS</td>
<td>average value of the M parameter determined for surrounding tissues</td>
</tr>
<tr>
<td>maxmL</td>
<td>maximum value of the m parameter determined for lesion</td>
</tr>
<tr>
<td>minmL</td>
<td>minimum value of the m parameter determined for lesion</td>
</tr>
<tr>
<td>avgmL</td>
<td>average value of the m parameter determined for lesion</td>
</tr>
<tr>
<td>maxmS</td>
<td>maximum value of the m parameter determined for surrounding tissues</td>
</tr>
<tr>
<td>minmS</td>
<td>minimum value of the m parameter determined for surrounding tissues</td>
</tr>
<tr>
<td>avgmS</td>
<td>average value of the m parameter determined for surrounding tissues</td>
</tr>
<tr>
<td>dmaxM</td>
<td></td>
</tr>
<tr>
<td>dminM</td>
<td></td>
</tr>
<tr>
<td>davgM</td>
<td></td>
</tr>
<tr>
<td>dmaxm</td>
<td></td>
</tr>
<tr>
<td>dminm</td>
<td></td>
</tr>
<tr>
<td>davgm</td>
<td></td>
</tr>
</tbody>
</table>

Statistical analysis showed that the values of parameters: maxML, minML, avgML, maxmL, minmL, avgmL, dmaxM, dminM, davgM, dmaxm, dminm, davgm, were statistically significant (p<0.05), and were able to distinguish the malignant, and benign, tumors.

For tissues surrounding the lesion, six parameters were considered (maxMS, minMS, avgMS, maxmS, minmS, avgmS). The obtained results showed that the values of the parameters were not statistically significant, and were not able to distinguish malignant from benign (p>0.05).

In the next step, the significance of maxML, minML, avgML, maxmL, minmL, avgmL, dmaxM, dminM, davgM, dmaxm, dminm and davgm parameters (Table 2), in differentiation of the character of the lesion, was investigated. For this purpose the cut-off values for all parameters were calculated, and the levels of sensitivity and specificity were determined.

The results presented in Table 2 indicate that the davgM parameter is the most accurate (83%) in diagnosing the breast lesions. For the davgM parameter, the value of the sum of the sensitivity and specificity was the highest; being equal to 161.67 (sensitivity 75%, specificity 86.7%). For this parameter, the cut-off value was equal to 1.18, and there were 10/75 (13.3%) false positive cases among benign lesions verified as fibroadenoma (n=9) and fat tissue (n=1). Among the malignant lesions, there were 8 out of 32 (25%) lesions diagnosed as false negative cases. There were verified, as mixed carcinoma (n=5), carcinoma ductale in situ (n=1), lobulare carcinoma (n=1), non-specific type (n=1).
For the \( \text{avgm} \) parameter, the value of the sum of the sensitivity and specificity was equal to 154.5% (sensitivity 62.5%, specificity 92%), and for \( \maxML \) 155.9% (sensitivity 81.3% and specificity 74.7%).

Tab. 2. Cut-off, sensitivity and specificity values determined for analyzed parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Cut-off</th>
<th>Sensitivity</th>
<th>Specificity</th>
<th>Sensitivity + Specificity</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \maxML )</td>
<td>2.52</td>
<td>81.25</td>
<td>74.67</td>
<td>155.92</td>
<td>76.64</td>
</tr>
<tr>
<td>( \minML )</td>
<td>0.59</td>
<td>71.88</td>
<td>76.00</td>
<td>147.88</td>
<td>74.77</td>
</tr>
<tr>
<td>( \text{avgML} )</td>
<td>1.89</td>
<td>68.75</td>
<td>84.00</td>
<td>152.75</td>
<td>79.44</td>
</tr>
<tr>
<td>( \maxmL )</td>
<td>0.54</td>
<td>81.25</td>
<td>65.33</td>
<td>146.58</td>
<td>70.09</td>
</tr>
<tr>
<td>( \minmL )</td>
<td>0.22</td>
<td>71.88</td>
<td>73.33</td>
<td>145.21</td>
<td>72.9</td>
</tr>
<tr>
<td>( \text{avgmL} )</td>
<td>0.45</td>
<td>65.63</td>
<td>86.67</td>
<td>152.29</td>
<td>80.37</td>
</tr>
<tr>
<td>( \text{dmaxM} )</td>
<td>1.81</td>
<td>68.75</td>
<td>82.67</td>
<td>151.42</td>
<td>78.5</td>
</tr>
<tr>
<td>( \text{dminM} )</td>
<td>0.51</td>
<td>65.63</td>
<td>84.00</td>
<td>149.63</td>
<td>78.5</td>
</tr>
<tr>
<td>( \text{davgM} )</td>
<td>1.18</td>
<td>75.00</td>
<td>86.67</td>
<td>161.67</td>
<td>83.18</td>
</tr>
<tr>
<td>( \text{dmaxm} )</td>
<td>0.19</td>
<td>71.88</td>
<td>70.67</td>
<td>142.54</td>
<td>71.03</td>
</tr>
<tr>
<td>( \text{dminm} )</td>
<td>0.18</td>
<td>56.25</td>
<td>89.33</td>
<td>145.58</td>
<td>79.44</td>
</tr>
<tr>
<td>( \text{davgm} )</td>
<td>0.21</td>
<td>62.50</td>
<td>92.00</td>
<td>154.50</td>
<td>83.18</td>
</tr>
</tbody>
</table>

4. Discussion

Breast cancer is the most common cancer among women. Imaging techniques, leading to improved effectiveness of breast cancer detection, are continuously sought, especially in the group of young women, having mammographically dense breasts. Mammography is, in Poland, the screening test, which is used in women over 50 years old; but the risk of breast cancer increases in women up to 30 years old. In this group of patients, also in older women with dense breasts, ultrasound examination is an alternative tool: especially with a high resolution linear transducer, which allows description of the characteristics of the breast lesions. This examination is operator-dependent, therefore, we conduct the research which is connected by ultrasound, and is operator-independent. To our knowledge, there have been no published results regarding the statistical parameters, used by us, in the differential diagnosis of breast lesions.

In this paper a method based on statistical processing of the RF echoes, backscattered in the breast tissues, is discussed. Two probability distributions, K and Nakagami were used to answer the question: as to whether, the benign and malignant lesions, can be differentiated using the backscattering properties of breast tissue. Based on the shape parameters of K and Nakagami distributions, 18 parameters have been defined, and their value in distinguishing the type of lesion, was analyzed.

In the first step, it was shown that the values of the parameters determined for tissues surrounding the lesion, are not feasible to differentiate the character of the breast lesions.

On other hand, the parameters calculated for the lesions’ backscattered echoes indicated that all of them might be helpful in improving the diagnosing of breast lesions, using quantitative ultrasound. In this set of parameters, the highest value of the sum of sensitivity and specificity was obtained using the \( \maxML \) for K distribution. The sum of sensitivity and specificity was equal to 155.92 (sensitivity 81.25%, specificity 74.67%).

Breasts may have various structure, and consist of three major tissue types; fat, connective, and epithelial. Furthermore, there are many various variations of normal breast
tissue, that sometimes it is easy to misclassify the anatomic structure as a pathological one. Consequently, the modifications of the tissue microstructure, as a result of disease, may also have a various character. For that reason, six differential parameters (Table 2), from the lesion and surrounding tissues, were also analyzed. Parameter defined as $d_{avgM}$ (the difference between $avgM_L$ for lesion and $avgM_S$ for surrounding tissue), and $d_{avgm}$ (the difference between $avgm_L$ for lesion and $avgm_S$ for surrounding tissue), were able to differentiate malignant and benign lesions with a sensitivity and specificity equal to 75% and 86.67%, and 62.50% and 90%, respectively.

Presented results are similar to that obtained by Tsui et al [18], who have demonstrated the usefulness of the Nakagami parameter in characterizing breast lesions. They have obtained (in a set of 100 lesions (50 malignant and 50 benign)) 82% of the diagnostic accuracy (the sensitivity was equal to 92% (46 out of 50) and specificity was 72% (36 out of 50)). They also stressed the fact that the quantitative ultrasound reduces the operator dependencies of the classification of breast tumors; which is especially important when B-mode images are analyzed by a less experienced physician. Authors have demonstrated that, in that case, the application of the statistical analysis of ultrasonic signals can increase the diagnostic accuracy from 66% to 82%.

In conclusion, the presented results confirm that quantitative ultrasound provides a new source of information: helping to differentiate between the malignant, and benign, breast lesions in vivo. In the future we are going to continue this research by validating the value of other probability distribution functions. (e.g. homodyne K) in the diagnosis of cancer. The combining of quantitative ultrasound, with the methods which use sonographic image features, (e.g. BIRADS classification) will be examined also.

Acknowledgment

This work was partially supported by the National Science Centre Grant Number UMO-2014/13/B/ST7/01271.

References


