SONOHISTOLOGY FOR THE COMPUTERIZED DIFFERENTIATION OF PAROTID GLAND TUMORS

Ulrich Scheipers,* Stefan Siebers,* Frank Gottwald,† Mohammad Ashfaq,* Alessandro Bozzato,† Johannes Zenk,† Heinrich Iro,† and Helmut Ermert*
*Institute of High-Frequency Engineering, Ruhr-University Bochum, Bochum, Germany; and †University Hospital of Otorhinolaryngology, University Erlangen, Erlangen, Germany

INTRODUCTION

The incidence of parotid gland tumors is one of the highest of all incidences of tumors in otorhinolaryngology. In some cases, differentiation between benign and malignant lesions is difficult. At later stages of the disease, some types of benign parotid gland tumors, such as the pleomorphic adenoma, may even become malignant. Therefore, the possibility of early detection, differentiation and excision is important. The initial characterization of the tumor will define the extent of surgical intervention and treatment. The excision of parotid gland tumors is extensive, time-consuming and, therefore, costly because several facial nerves are situated within the extent of the parotid gland; thus, requiring precise surgery to prevent damage to the nerves.

Usually, B-mode medical ultrasound (US) is applied as the main diagnostic modality to differentiate among the various types of parotid gland tumors. However, the results of medical US imaging of the parotid gland are dependent on the skills of the examining physician. Two typical examples of tumors of the parotid gland are shown in Fig. 1. Doppler-based US modes, such as color Doppler and power Doppler, can help differentiate between different types of parotid gland tumors, if used in combination with conventional B-mode US (Izzo et al. 2004; Martinoli et al. 1994; Schade et al. 1998; Schick et al. 1998). However, the lack of distinctive vessel patterns prevents the exclusive application of Doppler modes for the differentiation of tumors of the parotid gland (Schade et al. 1998). The application of US contrast agents can only increase the ability to differentiate between certain types of parotid gland tumors (Steinhart et al. 2003).

Dr. Scheipers' present address is Resonant Medical Inc., Montreal, QUE, Canada.
Address corrrespondence to: Dr.-Ing. Ulrich Scheipers, Resonant Medical Inc., 2050 Bleury, Suite 200, Montreal, QUE H3A 2J5 Canada. E-mail: Ulrich@Scheipers.org

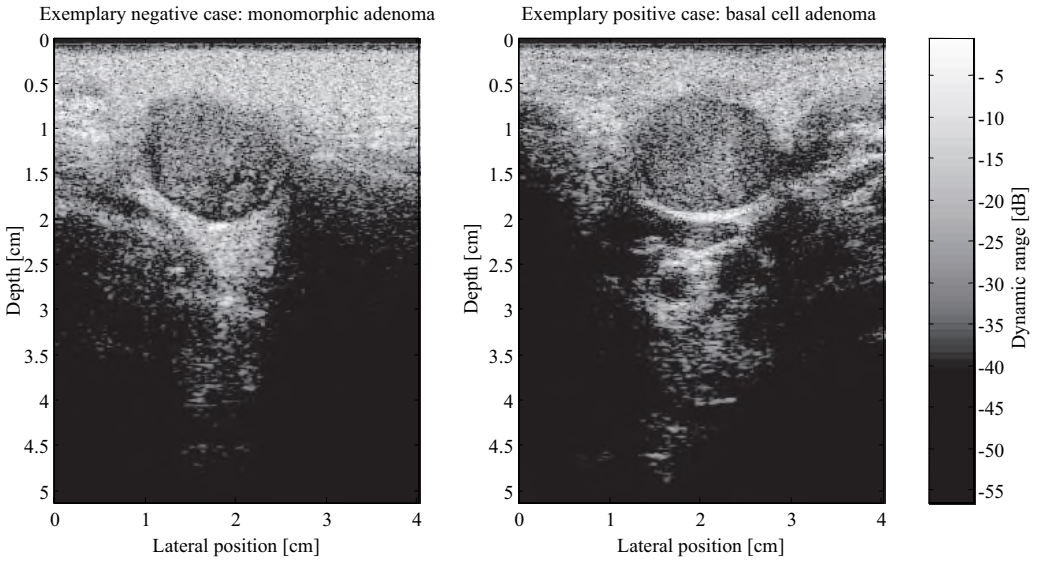

Fig. 1. Exemplary B-mode images of (left) a negative case (monomorphic adenoma) and (right) a positive case (basal cell adenoma). Differentiation between different types of parotid gland tumors and parotid gland alterations requires experienced physicians.

Overall, the current methods of diagnostics of parotid-gland tumors, including medical US and magnetic resonance imaging (Izzo et al. 2004), lack accuracy, especially when it comes to differentiate between benign and malignant forms of alterations. Currently, only the histopathologic examination of tissue specimen can provide the information needed for a definite diagnosis. The application of ultrasonic tissue characterization can add additional information to the currently available methods of diagnosis. In the future, ultrasonic tissue characterization might replace histopathologic examinations.

Methods of ultrasonic tissue characterization have been evaluated on various organs. Representative overview articles on ultrasonic tissue characterization in general have been published by Thijssen (2000) and Delorme and Zuna (2000). Shimizu et al. (1999) compared sonographic and histopathologic findings of parotid gland tumors, providing a baseline for ultrasonic tissue characterization of parotid glands. Texture analysis of parotid glands was formerly performed by Ariji et al. (1996) and Yonetsu et al. (2004).

The ultrasonic tissue characterization system used for the approach described in this work was formerly developed at Ruhr-University Bochum for the early detection of prostate carcinoma (Scheipers et al. 2001, 2002, 2003a, 2003b, 2004). The whole system was adapted for application in diagnostics of parotid glands.

By using US baseband data instead of video data, various tissue-describing parameters that contain sufficient information to classify the underlying tissue with satisfactory accuracy, can be calculated. Often, only the use of US baseband or radiofrequency data can provide the information needed to calculate parameters that characterize biologic tissue in an adequate way.

Because the different parameters used in this approach have a highly nonlinear interdependence, only a nonlinear model is able to combine the parameters and, thus, lead to reliable classification results (Scheipers et al. 2001, 2002, 2003a, 2003b, 2004). Both radial basis neural networks, which are based on symmetrical model functions of typically Gaussian shape, and network-based fuzzy inference systems perform reliably when applied in pattern recognition for ultrasonic tissue characterization (Bishop 1995). In this approach, a network-based fuzzy inference system was used (Zadeh 1973, 1989; Jang 1993; Mendel 1995; Furuhashi 2001). Most classification approaches only work satisfactorily if the underlying parameters are normally distributed. Some parameters used in this approach cannot be assumed to be distributed normally (Lizzi et al. 1997a, 1997b). Under these circumstances, a network-based fuzzy inference system that is quite robust to the distribution of input vectors, still performs sufficiently well (Blackmore 1994).

For successful application of spectral parameters, the compensation of depth-dependent diffraction and attenuation effects was found essential (Oosterveld et al. 1991; Huisman and Thijssen 1996; Lizzi et al. 1997a, 1997b; Schmitz et al. 1994, 1999). For estimation of some texture parameters, the compensation of these effects was found to be less important. However, these

effects should also not be ignored when using texture parameters.

Because the information of tissue parameters of different parameter groups (*e.g.*, spectral parameters, first-order texture parameters or second-order texture parameters) is highly uncorrelated, the combination of tissue parameters originating from different parameter groups can lead to better classification results (Hartman et al. 1993; Thijssen et al. 1993; Schmitz et al. 1994, 1999; Siebers et al. 2002). From our point of view, only a combination of different groups of parameters (*e.g.*, spectral and textural parameters) can provide the classification system with enough information to support an accurate and reliable decision.

METHODS

Data acquisition

Baseband US echo data of the parotid gland were captured during routine examination of the patient using standard US imaging equipment. Patient compliance to the procedure was high because the new method does not extend the normal examination time when applying US imaging of the neck and facial region. The proposed system is operator-independent, which means that no special knowledge or training is necessary for a successful application of the system.

For the clinical study, a Siemens Elegra (Issaquah, WA, USA) digital US scanner equipped with a research interface was used. The conventional linear model 7.5L40 probe was set to a center frequency of 7.2 MHz. For every frame, 2400 samples were recorded for 360 lines. The approximate size of the images was 5.1 cm in the axial direction and 4 cm in the lateral direction, respectively. Only one transmit focus was applied using a depth of 2 cm. Time-gain compensation was set to a neutral position. Internally, the dynamic low-pass filter and the frequency of the dynamic local oscillator were set to a constant value. Radiofrequency echo data were acquired at 36 MHz and 12 bits. Baseband data were provided by the research interface for downloading to a personal computer using the local area network.

Two orthogonal frames per lesion were recorded successively and every data frame was subdivided into numerous regions of interest (ROIs) using the sliding window technique to attain spatially distributed tissue characterization maps. Each ROI comprises an area of approximately 2.7 mm in the axial direction and 3.5 mm in the lateral direction. Thus, the smallest region that can be analyzed by the system covers approximately 9.5 mm^2. The ROIs consist of 128 sample points in the axial direction and of 16 scan lines in the lateral direction. The axial and lateral overlaps of the ROIs were 75% and 50%, respectively. For attenuation measurements, adjacent ROIs were combined, because the sizes of the ROIs mentioned above were too small to support reliable attenuation estimation.

In the following, all ROIs were transformed into frequency domain using Fourier transform on every scan line of the ROI. Before applying the Fourier transform, all ROIs were windowed by a Hamming window of the ROI length to avoid spectral leakage. Spectral leakage can affect the estimation of spectral parameters, especially when analyzing short time signals (Kroschel 1996; Proakis and Manolakis 1996).

The baseband data were compensated for system- and depth-dependent effects using the system transfer function over depth as an inverse filter within the effective band width of the system. The effective band width was determined as the frequency range of the spectrum that could clearly be distinguished from the noise floor. Using this approach, the systems effects caused by focusing and the electromechanical characteristics of the transducer can be partly compensated (Thijssen 2000; Huisman and Thijssen 1996). An error was induced because the system transfer function was constructed by interpolating echo sequences recorded at different distances of the transducer from a wire phantom in degassed water (Angelsen 2000). A detailed discussion of the data-correction method implemented here can be found in Schmitz et al. (1999).

Parameter extraction

Several tissue-describing parameters were calculated for each ROI. The authors do not claim that the extracted parameters are completely independent of the US imaging equipment. The parameters used for classification were calculated from the frequency spectrum and from the time domain, partly before and partly after compression and envelope detection of the baseband data.

Spectrum parameters were calculated after converting the power spectrum to dB. Spectral results of each scan line were averaged to form an estimate of the average power spectrum (Lizzi et al. 1983, 1997a; Lang et al. 1994; Thijssen 1989). The primary set of spectrum parameters consisted of five measures of backscatter calculated for the signal band width. The parameters used in this approach were: axis intercept, slope, midband value, deviation and normalized square deviation of the linear regression spectrum fit (Scheipers et al. 2001, 2002, 2003a, 2003b, 2004; Feleppa et al. 2001, 2004). The estimations of backscatter were compensated for attenuation effects using an attenuation model that is based on the multi-narrow-band method (Cloostermans and Thijssen 1983; Oosterveld et al. 1991; Thijssen et al. 1993). Three measures of this attenuation model were also evaluated by the system. The three attenuation pa-

rameters used in this approach were: axis intercept, slope and midband value. These three parameters were derived from the frequency-dependent attenuation coefficient over depth. As was shown by Oosterveld et al. (1991) and Thijssen et al. (1993), the exclusion of all ROIs with overflows, underflows or severe inhomogeneities (*e.g.*, specular reflections or strong shadows) is important for the calculation of attenuation parameters. ROIs containing these properties were discarded before calculation. A statistical framework for ultrasonic spectral parameter imaging was proposed by Huisman and Thijssen (1996) and Lizzi et al. (1983, 1997a, 1997b) and was taken into consideration in this approach.

Texture parameters evaluated in this work consist of first and second-order (*i.e.*, co-occurrence) parameters. Texture parameters were calculated after envelope detection of the complex baseband data using rectification.

First-order texture parameters consist of eight different estimates of echo amplitude, maximum, minimum, mean, variance, kurtosis, signal-to-noise ratio (SNR), ratio of squares and the full-width at half-maximum of the grey-level histogram.

Common co-occurrence parameters as proposed by Haralick et al. (1973); Valckx and Thijssen (1997) and Valckx et al. (2000) were calculated in the spatial domain of demodulated data for different distances (*i.e.*, step sizes between samples) up to 0.5 of the full-width at half-maximum measure (Scheipers et al. 2002; Basset et al. 1993). Sizes of co-occurrence matrices (*i.e.*, the number of grey levels incorporated) were varied in the range of 16 to 64. In contrast to first-order texture parameters, second-order texture parameters are based on spatial relationships between grey levels of the samples and can, therefore, describe spatial distributions of information in data. The nine second-order texture parameters evaluated in this approach were: angular second moment, which is a measure of the local homogeneity of the data; contrast, which is a measure of the amount of local variations present in the data; correlation, which is a measure of local linear dependencies; dimension, inverse difference moment, kappa, peak density, variance and the SNR of co-occurrence matrices. The lateral resolution of US data changes over depth, thus leading to an increase in lateral speckle size for deeper imaging positions. Grey-tone spatial dependency matrices were calculated only in the axial direction to keep effects of focusing and diffraction as low as possible. Another problem that occurs when texture parameters are evaluated is the dependence of some co-occurrence parameters on the linear attenuation of the system. Depth-variant parameters of this kind used here are: correlation, inverse difference moment and variance. It is possible to compensate partly for this type of depth-dependence by normalizing the data as proposed in Haralick et al. (1973) or by normalizing each ROI locally. All other co-occurrence parameters evaluated in this approach are independent of local grey levels.

Selection of parameters

A preselection of parameters was made by covariance matrix analysis. Parameter vectors that were highly linearly dependent on others were found and discarded. In addition, parameter vectors that have a very small influence on the classification results also were found and discarded. For the detection of these parameters, each single parameter was evaluated on its own, using separate classification experiments. During the preselection procedure, the number of tissue parameters was reduced to six, using a stepwise selection algorithm based on hypothesis testing. The parameter selection procedure starts by calculating the classification performance of each single tissue parameter using fourfold cross-validation over cases. Fourfold cross-validation was performed by dividing the whole number of cases into four data sets. Three data sets were combined and used as the training data and the remaining fourth data set was used as the test data. Every iteration of the classification system was performed 4 times in a row, with all different combinations of data sets involved. While dividing the whole data into four separate data sets, strict separation between patients was reserved to keep the results unbiased.

To use a decision-based criterion (Kung and Taur 1995) instead of an approximation-based criterion (*e.g.*, mean squared error) as used in other approaches, the classification power is expressed as the area A_{ROC} under the ROC curve (Swets 1982; Kroschel 1996).

The overall performance of a classification system that is not dependent on a separation threshold or a cut-off point, can easily be provided using the area A_{ROC} under the ROC curve as a measure (Metz 1986; Obuchowski 1997). The ROC curve sometimes is constructed by plotting the sensitivity against the specificity for all possible separation thresholds. The ROC curve generally is constructed by plotting the sensitivity against one minus specificity, which is the probability of a false alarm. However, both methods yield the same value for the area under the ROC curve.

After the area under the ROC curve was calculated for every tissue parameter, the tissue parameter with the largest area A_{ROC} was chosen as the first feature of choice. During the next step, this parameter was combined with all other remaining tissue parameters and the parameters of the pair with the largest area A_{ROC} were then chosen as the features of choice. This procedure was repeated until the area A_{ROC} decreases or stalls as the total number of parameters increases. Because this procedure is quite time-consuming, a reduction procedure

was proposed by Sugeno and Yasukawa (1993) that, in successive steps, only combines the current tissue parameters of choice with additional tissue parameters that, during the last step, performed as the best combination of tissue parameters two steps ago. In other words, if any combination of N parameters performed worse than the best combination of $N - 1$ parameters, the last parameter of the combination of N parameters is no longer used during the iterative search for the optimal parameter combination. In most cases, this reduction was proven useful because the optimum combination of parameters could be found with a decreased number of evaluations. However, this procedure of selecting the optimum set of tissue parameters is not the best method from a statistical point of view. The best set of tissue parameters can only be estimated when every possible combination of tissue parameters is evaluated and compared with all other possible combinations. At any rate, the statistically best method is too time-consuming to be practical. The optimum method that was used in this approach is considered the best alternative.

The number of ROIs used in this work was relatively high (Table 1). Thus, using up to six tissue parameters for the classification procedure was still a safe approach (Foley 1972; Chan et al. 1997). The distributions of the tissue parameters used in the classification procedure could be estimated with a sufficiently low variance, although the number of ROIs that were completely independent of each other may be smaller than the total number of ROIs of this approach, because the ROIs overlap and because several ROIs usually originate from the same case.

Fuzzy inference systems

Network-based fuzzy inference systems (FIS) were used to classify and separate the ROIs into two classes (negative = benign, positive = malignant). Fundamentals of fuzzy logic and the idea behind this approach were published by Zadeh (1973, 1989) and Mendel (1995). An overview of network-based fuzzy inference systems is given by Jang and Sun (1995). The mathematical background, especially the learning process of the FIS, is also described by Jang (1993). The fuzzy inference systems used in this work are based on first-order Sugeno-type systems (Sugeno and Yasukawa 1993) with Gaussian membership functions and up to eight rules to model the feature space. The number of rules was adaptively chosen by the system (Yager and Filev 1994). Subtractive clustering, which is an extension of the mountain clustering method proposed by Yager and Filev (1994), was used as the initial step in the supervised learning procedure to find natural clusters in the data space (Chiu 1994). During this step, the center position of the Gaussians in the feature space was determined. Subtractive clustering is a realization of the "scatter partitioning" method described by Jang and Sun (1995). A Sugeno-type system was chosen because this type of fuzzy inference system is computationally efficient and easily allows the use of propagation algorithms.

Table 1. Occurrence of different types of parotid gland tumors and parotid gland alterations during the clinical study

Type of tumor	Cases in study	Amount of ROIs
Basal cell adenoma	2	1320
Monomorphic adenoma	14	8444
Pleomorphic adenoma	3	1843
Adenoid cyst	1	48
Cyst	1	783
Canaliculous adenoma	1	776
Lymph nodes	1	1357
Sum	23	14571

A hybrid, adaptive-training algorithm based on backpropagation and least-square error estimation with adaptive step sizes was used to train the system (Jang 1991, 1993). Batch learning was applied to set the width of the Gaussians and to refine their center position. Network-based fuzzy inference systems were chosen for the classification of US data because fuzzy inference systems can model nonlinear functions of arbitrary complexity and tend to provide satisfying results when considering generalization performance. During the training procedure, the best combination of tissue parameters and the appropriate membership functions were stored in a rule base, which was used during the evaluation procedure and during further application of the system.

Postprocessing

The fuzzy output maps of the network-based fuzzy inference systems were transformed into binary maps applying a separation threshold. The separation threshold is used to separate the quasicontinuous output vectors into two classes or target groups. The separation threshold can be chosen freely by the operator because the implemented system is a quantitative system. The binary output maps were averaged to provide a mean decision criterion for the complete parotid gland. The decision criterion can be scaled according to probabilities for certain target groups. The classification results of the system before and after the postprocessing procedure will be presented and discussed later.

Clinical study

During 3.5 months, between March and July 2004, US baseband data of 23 parotid glands originating from 18 patients were recorded. The parotid glands were contoured in the US B-mode images using custom software

Table 2. Overview of tissue parameters with initial mean values (mean) and standard deviations (SD)

Parameter	Negative target group mean ± SD	Positive target group mean ± SD
Contrast	−0.1963 ± 1.0049	0.2706 ± 0.9273
Correlation	−0.0527 ± 0.6518	0.0726 ± 1.3355
Dimension	−0.0398 ± 0.7528	0.0549 ± 1.2617
Variance	−0.0509 ± 0.7213	0.0702 ± 1.2855
Midband value	−0.0232 ± 1.0532	0.0320 ± 0.9207
Slope	−0.1710 ± 1.0439	0.1241 ± 0.9478

to make sure that only signals originating from the glands and not from surrounding tissue were analyzed. All patients were scheduled to have parotid surgery during the next day or during the next week after acquisition of the US data.

At the day of the examination, the youngest patient was 25 years old, the eldest 88 years old. The mean age of the patients was 65 years, and the median age was 70 years. Over all, 11 of the patients were women and 7 were men. Histopathological examinations after parotidectomy were used as the "gold standard." (Histologic analysis is usually accepted as the "gold standard.") The results of the histologic examinations are presented in Table 1.

In this first approach, the system was trained to differentiate between a first target group containing all cases of monomorphic adenoma and a second target group containing all cases of other types of parotid gland alterations. The second target group contained all non-monomorphic adenoma types of parotid gland alterations that occurred during the clinical study, not making any difference between benign and malignant parotid gland alterations. The first target group was called "negative" because the incident of monomorphic adenoma is considered benign. The second target group was called "positive" because all other types of tumors and alterations of the parotid gland would obviously be considered to be malignant. Even those diseases that are actually of benign nature were counted in the second group, because they occur too seldom to achieve a high probability for being considered safe if left untreated.

The baseband data sets were divided into numerous ROIs, as described above, resulting in a sum of 6127 negative and 8444 positive ROIs. An overview of all different types of tumors and alterations of the parotid glands analyzed during this study together with their number of ROIs is shown in Table 1.

RESULTS

Six tissue-describing parameters were automatically chosen by the classification system. Three parameters originate from the second-order texture parameter group: contrast, correlation and dimension. Two spectral parameters were chosen: midband value and slope. In addition, one parameter of the first-order texture parameter group was included: variance. An overview of the tissue parameters used in the final classification system, together with their mean values and standard deviations (SD) estimated over all data sets, is given in Table 2. The optimum number of rules of the network-based fuzzy inference system was iteratively found to be two.

The fuzzy inference system was trained using the histologic findings as the "gold standard" or "teacher data." The fuzzy inference system provides a fuzzy value for each ROI of the US data set. The fuzzy value is a measure of the probability of the ROI belonging to the positive or negative target group.

The area under the ROC curve given as the cross-validation mean and the cross-validation SD is $A_{ROC} = 0.95 \pm 0.07$ when using fourfold cross-validation over cases and differentiating between monomorphic adenoma as the first target group and all other types of parotid gland alterations as the second target group. In two cross-validation cases, exceptional ideal classifications of A_{ROC} were performed.

Intermediate results were evaluated for comparison with the final classification results. Without the postprocessing step, which provides the final score for each case, the area under the ROC curve given as the cross-validation mean and the cross-validation SD is A^*_{ROC}, where the asterisk simply denotes the "intermediate" status. The ROC curve area was calculated by continuously varying the separation threshold (Obuchowski 1997).

Table 3. Overview of intermediate and final classification results

	Intermediate results estimated over ROIs			Final results estimated over cases		
Cross-validation set	A^*_{ROC}	SE^*_{ROC}	E^*_{EER}	A_{ROC}	SE_{ROC}	E_{EER}
1/4	0.80	0.01	0.73	0.96	0.00	0.85
2/4	0.64	0.02	0.60	0.83	0.01	0.83
3/4	0.65	0.01	0.62	1.00	0.00	1.00
4/4	0.75	0.01	0.68	1.00	0.00	1.00
Mean ± SD	0.71 ± 0.07	–	0.66 ± 0.05	0.95 ± 0.07	–	0.92 ± 0.08

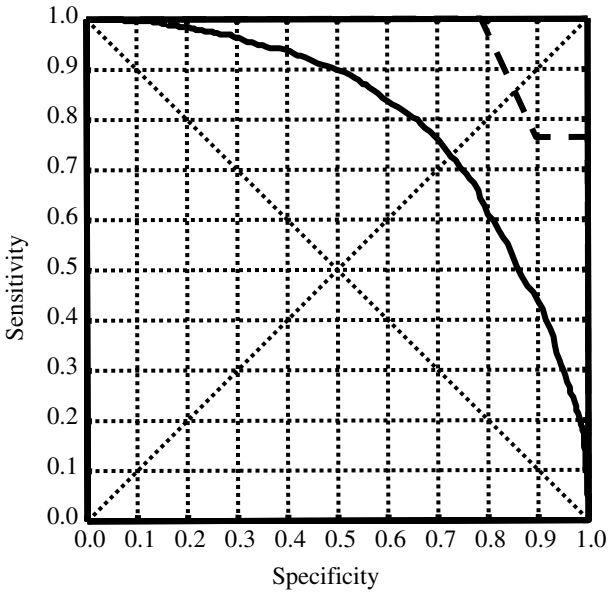

Fig. 2. ROC curves for cross-validation subset 1 of 4. (—) Intermediate classification results estimated over ROIs are plotted as a solid line: $A^*_{ROC} = 0.80, E^*_{EER} = 0.73$. (----) Final classification results estimated over cases are plotted as a dashed line: $A_{ROC} = 0.96, E_{EER} = 0.85$.

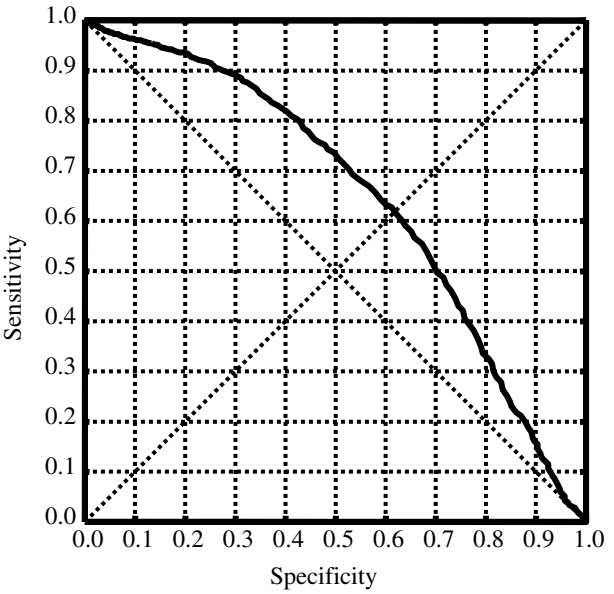

Fig. 4. ROC curves for cross-validation subset 3 of 4. (—) Intermediate classification results estimated over ROIs are plotted as a solid line: $A^*_{ROC} = 0.65, E^*_{ERR} = 0.62$. Final classification results estimated over cases: $A_{ROC} = 1.00, E_{EER} = 1.00$.

All classification results are presented in Table 3. In addition to the area under the ROC curve A_{ROC}, the standard error of the ROC estimate SE_{ROC} and the equal error rate E_{EER} are shown in the table. The standard error was estimated over the number of ROIs involved in the study (Hanley and McNeil 1982). Because the ROIs are corre-

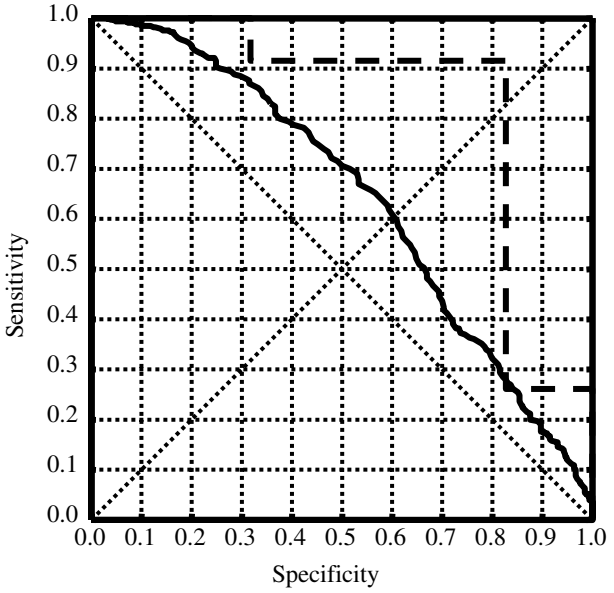

Fig. 3. ROC curves for cross-validation subset 2 of 4. (—) Intermediate classification results estimated over ROIs are plotted as a solid line: $A^*_{ROC} = 0.64, E^*_{EER} = 0.60$. (----) Final classification results estimated over cases are plotted as a dashed line: $A_{ROC} = 0.83, E^*_{EER} = 0.83$.

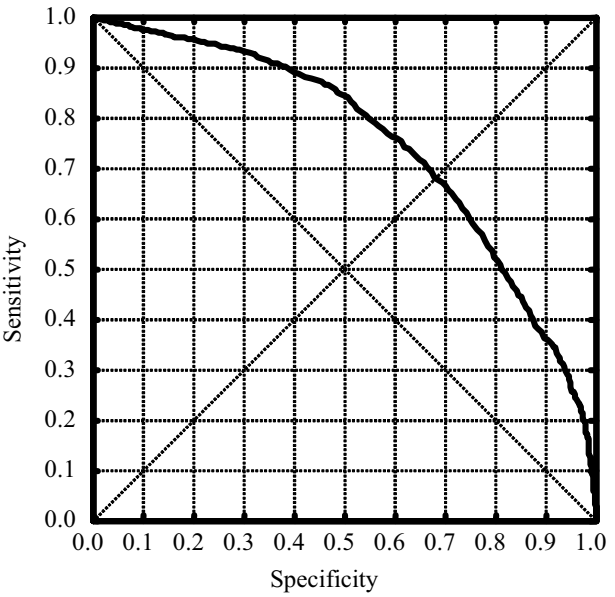

Fig. 5. ROC curves for cross-validation subset 4 of 4. (—) Intermediate classification results estimated over ROIs are plotted as a solid line: $A^*_{ROC} = 0.75, E^*_{ERR} = 0.68$. Final classification results estimated over cases: $A_{ROC} = 1.00, E_{EER} = 1.00$.

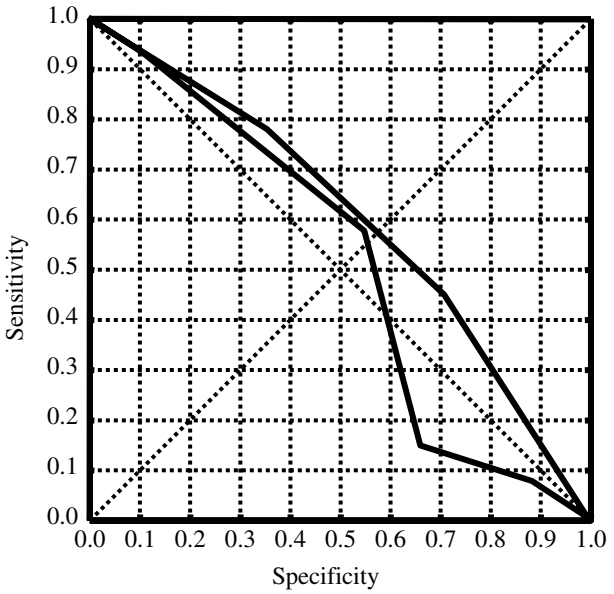

Fig. 6. ROC curves of "blinded" experiment of two experienced physicians. They reviewed the B-mode images generated from recorded US data sets without knowledge of the age of the patients or any other information of the patients' records. The probability of monomorphic adenoma had to be judged with a grade between 1 and 5. The areas under the ROC curves are $A_{ROC,P1} = 0.51$ for the first physician and $A_{ROC,P2} = 0.60$ for the second physician.

lated with each other to an undefined degree, the estimations of standard error tend to underestimate the true error. The equal error rate is the sensitivity or specificity of the system at the operating point where sensitivity equals specificity. For the final system, an equal error rate of $E_{EER} = 0.92 \pm 0.08$ was achieved. Intermediate results yielded an equal error rate of E^*_{EER} 0.66 ± 0.05. Because the classification procedure applied here represents a continuous system, sensitivities and specificities can be chosen freely under dependence of each other, which means that the operating point of the system can be set to any separation threshold requested by the operator or physician.

The ROC curves for all four subsets of the cross-validation experiment are shown in Figs. 2 to 5. In the figures, the intermediate results estimated over ROIs are plotted as a solid line and the final classification results estimated over cases are plotted as a dashed line.

During the clinical study, the diagnoses of an experienced physician were recorded for each case before the US data of the cases was processed by the classification system. The experienced physician achieved a sensitivity of $SE_{P1} = 0.67$ and a specificity of $SP_{P1} = 0.79$. However, because the database is comparably small, the performance of the classification system comparison can only be rudimentary compared with the results of the experienced physician. At a sensitivity of $SE = 0.67$, the final classification system achieves a specificity of $SP = 0.96 \pm 0.08$. At a specificity of $SP = 0.79$, the final classification system achieves a sensitivity of $SE = 0.98 \pm 0.04$, respectively.

As an additional experiment, two experienced physicians were asked to review the B-mode images of the data underlying the study. Both physicians were "blinded" to all additional information, such as the ages and the medical records of the patients. The physicians had to choose one of five possible diagnoses for each patient:

1. Definitely no monomorphic adenoma
2. Probably no monomorphic adenoma
3. Uncertain
4. Probably monomorphic adenoma
5. Definitely monomorphic adenoma.

Certainly, the setup of the experiment is far from clinical practice; however, the outcome is interesting because it is directly comparable with the results of the system for the computerized differentiation of parotid gland tumors. From the results of the physicians, two ROC curves were estimated. Both curves are shown in Fig. 6. The classification results of the first physician yield an area under the ROC curve of $A_{ROC,\ P1} = 0.51$, and the classification results of the second physician yield an area under the ROC curve of $A_{ROC,P2} = 0.60$, respectively. Both results are far from the classification results achieved by the computerized differentiation.

DISCUSSION

Although only a relatively small database consisting of $n = 23$ cases was evaluated in this work, the system for ultrasonic multifeature tissue characterization differentiates between monomorphic adenomas and all other types of alterations of parotid glands with a satisfying grade of accuracy. The area under the ROC curve is $A_{ROC} = 0.95 \pm 0.07$ when using fourfold cross-validation methods to evaluate the underlying data. In two of the four cross-validation cases, exceptional ideal classifications were performed. The classification rates will possibly increase if a larger database is available. Several alterations of parotid glands only occurred once during the clinical study, but a total of 14 cases of monomorphic adenoma could be counted. The classification system seems to have learned the typical characteristics of the monomorphic adenoma sufficiently well. However, a differentiation between the other types of tumors would be interesting. This will be evaluated in the future, together with leave-one-out cross-validation tests, when a larger database will be available.

The interobserver variability is a problem of conventional B-mode US for the differentiation of diverse

parotid gland alterations. The interobserver variability is the dependence of the diagnostic results of the experience and expert knowledge of the conducting physician. A highly experienced physician may be able to detect the features of malignant tumors, but a novice physician may not be able to evaluate the echographic properties sufficiently. An ultrasonic tissue characterization system can automate the process of differentiating between various types of parotid gland tumors and, hence, may assist to reduce the gap in diagnostic results between expert and novice physicians. Because the system evaluates characteristics of the ultrasonic echo signal that are not shown in conventional B-mode images or that cannot be visually interpreted by the person operating the US system, the system may also be of great help to the expert, giving additional data in small and polymorphic cases.

REFERENCES

Angelsen BAJ. Ultrasound imaging. Trondheim, Norway: Emantec AS, 2000.

Ariji Y, Ohki M, Eguchi K, et al. Texture analysis of sonographic features of the parotid gland in Sjoegren's syndrome. Am J Roentgen 1996;166:935–941.

Basset O, Sun Z, Mestas JL, Gimenez G. Texture analysis of ultrasonic images of the prostate by means of co-occurrence matrices. Ultrason Imaging 1993;15:218–237.

Bishop CM. Neural networks for pattern recognition. Oxford, UK: Oxford University Press, 1995.

Blackmore S. Intelligent sensing and self-organizing fuzzy-logic techniques used in agricultural automation. Asae/Csae meeting paper no. 931048. Cranfield, UK: Silsoe College, Cranfield University, 1994.

Chan HP, Sahiner B, Wagner RF, Petrick N, Mossoba J. Effects of sample size on classifier design: Quadratic and neural network classifiers. Proc SPIE Med Imaging 1997;3034:1102–1113.

Chiu S. Fuzzy model identification based on cluster estimation. J Intell Fuzzy Syst 1994;2:267–278.

Cloostermans MJTM, Thijssen JM. A beam corrected estimation of the frequency dependent attenuation of biological tissues from backscattered ultrasound. Ultrason Imaging 1983;5:136–147.

Delorme S, Zuna I. Ad multos annos. Ultraschall Med 2000;21:230–232.

Feleppa EJ, Ennis D, Schiff Pb, et al. Spectrum-analysis and neural-networks for imaging to detect and treat prostate cancer. Ultrason Imaging 2001;23:135–146.

Feleppa EJ, Porter CR, Ketterling J, et al. Recent developments in tissue-type imaging (TTI) for planning and monitoring treatment of prostate cancer. Ultrason Imaging 2004;26:163–172.

Foley DH. Considerations of sample and feature size. IEEE Trans Inf Theory 1972;18:618–626.

Furuhashi T. Fusion of fuzzy/neuro/evolutionary computing for knowledge acquisition. Proc IEEE 2001;89:1266–1274.

Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 1982; 143:29–36.

Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;3:768–780.

Hartman PC, Oosterveld BJ, Thijssen JM, Rosenbusch GJ, van den Berg J. Detection and differentiation of diffuse liver disease by quantitative echography. A retrospective assessment. Invest Radiol 1993;28:1–6.

Huisman HJ, Thijssen JM. Precision and accuracy of acoustospectrographic parameters. Ultrasound Med Biol 1996;22:855–871.

Izzo L, Sassayannis PG, Frati R, et al. The role of echo colour/power Doppler and magnetic resonance in expansive parotid lesions. J Exp Clin Cancer Res 2004;23(4):585–592.

Jang JSR. Fuzzy modeling using generalized neural networks and Kalman filter algorithm. In: Dean TL, McKeown K eds. Proceedings of the Ninth National Conference on Artificial Intelligence. Menlo Park, CA: AAAI Press. 1991:762–767.

Jang JSR. ANFIS. Adaptive-network-based fuzzy inference systems. IEEE Trans Syst Man Cyber 1993;23:665–685.

Jang JSR, Sun CT. Neuro fuzzy modeling and control. Proc IEEE 1995;83:378–406.

Kroschel K. Statistische Nachrichtentechnik. Berlin: Springer-Verlag, 1996.

Kung SY, Taur JS. Decision-based neural networks with signal/image classification Applications. IEEE Trans Neural Netw 1995;6:170–181.

Lang M, Ermert H, Heuser L. In vivo study of on-line liver tissue classification based on envelope power spectrum analysis. Ultrason Imaging 1994;16:77–86.

Lizzi FL, Astor M, Feleppa EJ, Shao M, Kalisz A. Statistical framework for ultrasonic spectral parameter imaging. Ultrasound Med Biol 1997a;23:1371–1382.

Lizzi FL, Feleppa EJ, Astor M, Kalisz A. Statistics of ultrasonic spectral parameters and liver examinations. IEEE Trans Ultrason Ferroelec Freq Control 1997b;44:935–942.

Lizzi FL, Greenebaum M, Feleppa EJ, Elbaum M. Theoretical framework for spectrum analysis in ultrasonic tissue characterization. JASA 1983;73:1366–1373.

Martinoli C, Derchi LE, Solbiati L, et al. Color Doppler sonography of salivary glands. Am J Roentgen 1994;163:933–941.

Mendel JM. Fuzzy logic systems for engineering: A tutorial. Proc IEEE 1995;83:345–377.

Metz CE. ROC methodology in radiologic imaging. Invest Radiol 1986;21:720–733.

Obuchowski NA. Nonparametric analysis of clustered ROC curve data. Biometrics 1997;53:567–578.

Oosterveld BJ, Thijssen JM, Hartman PC, Romijn RL, Rosenbusch GJE. Ultrasound attenuation and texture analysis of diffuse liver disease: Methods and preliminary results. Phys Med Biol 1991;36: 1039–1064.

Proakis GJ, Manolakis DG. Digital signal processing. Upper Saddle River, NJ: Prentice Hall, 1996.

Schade G, Ussmuller J, Leuwer R. Value of duplex ultrasound in diagnosis of parotid-gland tumors. Laryngorhinootologie 1998;77(6):337–341.

Scheipers U, Ermert H. Neuro-fuzzy inference system for ultrasonic multifeature tissue characterization for prostate diagnostics. IEEE Ultrason Sympos 2002:1347–1350.

Scheipers U, Ermert H, et al. Ultrasonic tissue characterization for prostate diagnostics: Spectral parameters vs. texture parameters. Biomed Technik 2003a;48:122–129.

Scheipers U, Ermert H, et al. Ultrasonic multifeature tissue characterization for prostate diagnostics. Ultrasound Med Biol 2003b;29: 1137–1149.

Scheipers U, Koenig K, et al. Diagnostics of prostate cancer based on ultrasonic multifeature tissue characterization. IEEE Ultrason Sympos 2004:2153–2156.

Scheipers U, Lorenz A, Pesavento A, et al. Ultrasonic multifeature tissue characterization for the early detection of prostate cancer. IEEE Ultrason Sympos 2001:1265–1278.

Schick S, Steiner E, Gahleitner A, et al. Differentiation of benign and malignant tumors of the parotid gland: Value of pulsed Doppler and color Doppler sonography. Eur Radiol 1998;8(8): 1462–1467.

Schmitz G, Ermert H, Senge T. Tissue characterization of the prostate using Kohonen-maps. IEEE Ultrason Sympos 1994:1487–1490.

Schmitz G, Ermert H, Senge T. Tissue characterization and imaging of the prostate using radio frequency ultrasonic signals. IEEE Trans Ultrason Ferroelec Freq Control 1999;46:126–138.

Shimizu M, Ussmueller J, Hartwein J, Donath K. A comparative study of sonographic and histopathologic findings of tumorous lesions in the parotid gland. Oral Surg Oral Med Oral Pathol Oral Radiol Endod 1999;88:723–737.

Siebers S, Welp C, Werner J, Ermert H. Ultrasound-based imaging modalities for thermal therapy monitoring. Biomed Tech 2002;47: 438–440.

Steinhart H, Zenk J, Sprang K, Bozzato A, Iro H. Contrast-enhanced color Doppler sonography of parotid gland tumors. Eur Arch Otorhinolaryngol 2003;260(6):344–348.

Sugeno M, Yasukawa T. A fuzzy-logic-based approach to qualitative modeling. IEEE Trans Fuzzy Syst 1993;1:7–31.

Swets J. Evaluation of diagnostic systems: Methods from signal detection theory. New York: Academic Press, 1982.

Thijssen JM. Ultrasonic tissue characterization and echographic imaging. Phys Med Biol 1989;34:1667–1674.

Thijssen JM. Spectroscopy and image texture analysis. Ultrasound Med Biol 2000;26:41–44.

Thijssen JM, Oosterveld BJ, Hartman PC, Rosenbusch GJ. Correlations between acoustic and texture parameters from RF and B-mode liver echograms. Ultrasound Med Biol 1993;19:13–20.

Valckx FMJ, Thijssen JM. Characterization of echographic image texture by cooccurrence matrix parameters. Ultrasound Med Biol 1997;23:559–571.

Valckx FMJ, Thijssen JM, van Geemen AJ, Rotteveel JJ, Mullaart R. Calibrated parametric medical ultrasound imaging. Ultrason Imaging 2000;22:57–75.

Yager R, Filev D. Generation of fuzzy rules by mountain clustering. J Intell Fuzzy Syst 1994;2:209–219.

Yonetsu K, Ohki M, Kumazawa S, et al. Parotid tumors: Differentiation of benign and malignant tumors with quantitative sonographic analyses. Ultrasound Med Biol 2004;30(5):567–574.

Zadeh LA. Outline of a new approach to the analysis of complex systems and decision processes. IEEE Trans Syst Man Cyber 1973; 3:28–44.

Zadeh LA. Knowledge representation in fuzzy logic. IEEE Trans Knowledge Data Eng 1989;1:89–100.