Comparison of Pattern Recognition Methods for Computer-Assisted Classification of Spectra of Heart Sounds in Patients with a Porcine Bioprosthetic Valve Implanted in the Mitral Position

LOUIS-GILLES DURAND, SENIOR MEMBER, IEEE, MICHEL BLANCHARD, MEMBER, IEEE, GUY CLOUTIER, STUDENT MEMBER, IEEE, HANI N. SABBAH, AND PAUL D. STEIN

Abstract—**The diagnostic performance of two pattern recognition methods (or classifiers) to detect valvular degeneration was evaluated in 48 patients with a porcine bioprosthetic heart valve inserted in the mitral position. Twenty patients had a normal porcine bioprosthetic valve and 28 patients had a degenerated bioprosthetic valve. One method was based on the Gaussian-Bayes model and the second on the "nearest neighbor" algorithm using three distance measurements. Eighteen diagnostic features were extracted from the sound spectrum of each patient and, for each method, a two-class supervised learning approach was used to determine the most discriminant diagnostic patterns composed of 6 features or less. The probability of error of the classifiers was estimated with the leave-one-out approach. The performance of each method to discriminate between normal and degenerated bioprosthetic valves was verified by clinical evaluation of the valves. The best performance in evaluation of the sound spectrum (98% correct classifications) was obtained with the Bayes classifier and two patterns of six features each. The percentage of false positive classifications of valve degeneration was 0% and the percentage of false negative classifications was 4%. Sensitivity for the detection of valve degeneration was 96%, specificity was 100%, positive predictive value was 100%, and negative predictive value was 95%. The best performance of the nearest neighbor method (94% correct classifications) was obtained by using the Mahalanobis distance and five patterns composed of three, four, five, or six diagnostic features. Using a pattern composed of only three features, the percentage of false positive classifications for degeneration was 10% and the percentage of false negative classifications was 4%. Sensitivity was 96%, specificity was 90%, positive predictive value was 93%, and the negative predictive value was 95%.**

I. Introduction

The principal advantage of porcine bioprosthetic heart valves is a low incidence of thromboembolism and, therefore, the absence of a need for prophylactic therapy with anticoagulants [1]. However, the main disadvantage of bioprosthetic valves is their low durability (7–12 years). Morphological studies by Thiene *et al.* [2], Levy *et al.* [3], and Stein *et al.* [4] have shown that the deposition of calcium, which leads to stiffening and stenosis, is an important contributor to bioprosthetic valve failure. Because of the inevitable degeneration of bioprosthetic valves, periodic evaluation of their structural and functional integrity is essential to detect early malfunction. Such a followup is ideally based upon a method of evaluation which is both noninvasive and sensitive.

Previous studies by Stein *et al.* [5]–[7], Foale *et al.* [8], and Joo *et al.* [9] emphasized the diagnostic potential of spectral analysis of bioprosthetic valve closure sounds for the detection of valvular degeneration. These studies showed that the dominant frequency (or the two dominant frequencies) of porcine valve closure sounds shifts toward higher frequencies following valve leaflet calcification and stiffening.

More recently, Durand *et al.* [10] developed a pattern recognition algorithm using a Gaussian-Bayes model for automatic detection of valvular degeneration. This investigation was performed by use of nine diagnostic features (features 1–9, Table II) extracted from the sound spectrum of 57 patients with normal porcine bioprosthetic valves and 49 patients with degenerated porcine bioprosthetic valves. Results showed that the mean correct classification rate of the classifier, evaluated with the "holdout" approach, was 74% for the automatic classification of normal and degenerated bioprosthetic valves. The percentage of false positives was 13% while that of false negatives was 40%. Separating mitral and aortic bioprosthetic valves in two subgroups allowed reduction of the percentage of false negatives to approximately 30%. The performance of the classifier was improved to 77% correct classifications for mitral bioprosthetic valves, and reduced to 71% correct classifications for aortic bioprosthetic valves.

Manuscript received March 6, 1989; revised March 7, 1990. This work was supported by the National Sciences and Engineering Research Council of Canada, Ottawa, Ont., and the U.S. Public Health Service, National Institutes of Health, National Heart, Lung, and Blood Institute, Grant R01-Hl23669-08, Bethesda, MD.

L.-G. Durand and M. Blanchard are with the Laboratory of Biomedical Engineering, Clinical Research Institute of Montreal, Montreal, P.Q., Canada, H2W 1R7.

G. Cloutier was with the Laboratory of Biomedical Engineering, Clinical Research Institute of Montreal, Montreal, P.Q., Canada H2W 1R7. He is now with the Laboratory of Medical Ultrasonic, Bioengineering Program, The Pennsylvania State University, University Park, PA 16802.

H. N. Sabbah and P. D. Stein are with the Henry Ford Heart and Vascular Institute, Detroit, MI 48202.

0018-9294/90/1200-1121$01.00 © 1990 IEEE

Stein *et al.* [11] evaluated in combination with the dominant frequency, nine different features of the sound spectrum (features 10–18 in Table II) and showed that the use of these characteristics of the spectrum of the closure sound of bioprosthetic valves in the mitral position markedly improved the ability to identify degeneration.

In the present study, the nine features described by Durand *et al.* [10] as well as those identified by Stein *et al.* [11] were evaluated by two pattern recognition methods, the optimal Bayes' rule and the nearest neighbor method. This data base of 18 diagnostic features of the spectrum of the first heart sound was tested to determine if the performance of the classifier could be improved for the detection of degenerated porcine valves implanted in the mitral position.

II. Classifier Design

The most popular parametric technique of supervised pattern recognition, the Bayes classifier, is the optimal technique when the probability density functions of the patterns in the feature space are known (a pattern is defined as a D-dimensional vector composed of D features). Even though this classifier is robust, it is generally worthwhile to evaluate a nonparametric approach like the nearest neighbor classifier, especially when the probability density functions are difficult to estimate or cannot be estimated. The nearest neighbor method is an intuitive approach based upon distance measurements and motivated by the fact that patterns belonging to the same class should be near each other in the feature space. In this method, the nearest neighbors of a pattern of unknown status are found by computing distance measurements (Euclidian, Mahalanobis, etc.) between the unkown pattern and a given number of patterns from the training set. The status of the unknown pattern is then associated with that of the K nearest neighbors. It has been shown by Bailey and Jain [12] that increasing the K value does not monotonically increase the performance of the classifier. A useful guideline in practice is to select a value of K proportional to the square root of the number of patients in the training set [13]. The choice of the distance measurement is not easy since various distance measurements have been developed and tested. The Euclidian distance is the most commonly used but the Mahalanobis distance provides better performance when the statistical properties of the data are explicitly considered [14].

The design and evaluation of a classifier is straightforward when, in addition to the basic training set, an independent test set is available to evaluate its performance. In some instances, and particularly for detection of bioprosthetic valve degeneration, it may be difficult to obtain a large number of samples in each class. In such cases, splitting the sample population into two mutuallyexclusive sets as done by Joo *et al.* [9] (which is known as the holdout method) may appear to be an inefficient use of the data [15]. Other approaches such as the resubstitution method and the leave-one-out method minimize this problem [15], [16].

The resubstitution method uses the entire sample population as both training and test sets. Consequently, the probability of misclassification obtained with this technique is underestimated. The leave-one-out method provides a better estimate of the probability of error because it uses the complete sample population minus one ($N - 1$) as the training set. Once the classifier is trained, the isolated sample is classified. The procedure is repeated N times until all samples have been classified individually. Since this method uses almost all of the data to classify one sample at a time, its bias is small. However, its main disadvantage is that it requires extensive computing time (N classifiers must be designed).

In practice, the performance of the classifier increases up to a given point as the number of features is increased and then begins to deteriorate as further features are added. When the ratio of the sample dimension to the number of features per class is increased above three, the error rate of the classifier approaches the true error rate attained by the minimum probability of error classifier [17]. A number of features (m) of five or more times lower than the number of training samples per class (N_i) ensures that the real performance of the classifier is not overestimated [18], [19].

III. Method

A. Patient Population and Valve Status

Heart sounds were recorded in 48 patients with porcine bioprosthetic valves implanted in the mitral position. Among these patients, 28 patients had clinically apparent bioprosthetic valve degeneration that we detected on the basis of auscultation and echocardiography [20], and confirmed by cardiac catheterization. These patients underwent valve replacement. Their heart sounds were recorded a few days before surgery. The valves in these patients had been inserted for 92 $\pm$ 33 months (mean $\pm$ standard deviation; range 11–135 months). Twenty patients had a normally functioning bioprosthetic valve, as determined by clinical assessment. In these patients, the valve had been inserted for $\leq$6 months (2 $\pm$ 2 months).

The gross appearance of all of the degenerated bioprosthetic valves was examined, and the presence or absence of calcification was evaluated by palpation and inspection. Color photographic slides of the inflow and outflow surfaces, obtained soon after the valves were removed, were also examined to determine the extent of gross calcification. Only bioprostheses that underwent spontaneous degeneration were studied. If the patients had perivalvular regurgitation or if they, at any time, had prosthetic valve endocarditis, the heart sounds were not examined. A bioprosthetic valve was considered to have undergone spontaneous degeneration if there was valvular incompetence or stenosis in the absence of an identifiable cause. Table I describes the distribution of degenerated valves as a function of the degree of calcification. The patients whose valves were inserted $\leq$6 months served as controls. None had regurgitant murmurs. All aspects of

TABLE I
DISTRIBUTION OF THE DEGREE OF CALCIFICATION OF THE DEGENERATED VALVES

Number of valves	Degree of Calcification	Description
6	0	No calcification
4	1	One circumscribed nodule
3	2	Two circumscribed nodule
6	3	Three circumscribed nodule or single confluent equivalent
3	4	One third area of one leaflet
4	5	One third area + one or two nodules
0	6	One third area of two leaflets
2	7	One third area of two leaflets + nodules
0	8	One third area of three leaflets

clinical examination showed normal prosthetic valve function.

B. Recording and Analysis of Phonocardiograms

Although there has been disagreement in the literature over the contribution of the tricuspid valve to the first heart sound, it seems that vibratory motion of both the mitral and tricuspid valves initiate vibrations recognized as the first heart sound [21]. Intracardiac recordings of heart sounds in normal men showed that the mitral component of the first sound is of greater amplitude than the tricuspid component (136 ± 9.5 Pa versus 40 ± 4.6 Pa; 1 Pa = 10 dyn/cm^2) [21]. The power spectrum of the first sound recorded at the point of maximal impulse, therefore, would be dominated by vibrations induced by the mitral valve. For this reason, it seemed reasonable to assess the spectrum of the entire first sound, although we cannot exclude some contribution of the tricuspid valve.

In all patients, heart sounds were recorded with an Irex Medical Systems model 120-131 aircoupled microphone. The microphone was held in position on the chest wall by a suction cup. This permitted a uniform and consistent technique for the application of the microphone. The diameter of the air cavity of the microphone was 24 mm, the height of the air column was 6 mm, the thickness of the housing was 25 mm, and its total mass was 75 g. The sensitivity of the microphone was 12 mV/Pa, and the noise level was 42 dB above the audibility threshold of 20×10^{-6} Pa at 1000 Hz.

In each patient, the phonocardiogram was recorded on an 8-track magnetic tape recorder (Hewlett-Packard model 88688-A). The sound signal was filtered below 50 Hz and above 500 Hz. The frequency response of the sound amplifier and microphone combination was flat (within 1 dB) between 80 and 300 Hz. In the lower frequency range, at 40 Hz there was a 6-dB attenuation and at 20 Hz there was a 14-dB attenuation. In the higher frequency range, at 600 Hz there was a 6-dB attenuation and at 1200 Hz a 20 dB attenuation.

The tape recorder was operated at a speed of 95.25 mm/s and had a passband frequency of 0 to 1250 Hz. The frequency response of the passband was $\pm$ 1.0 dB referenced to 10% of the upper band edge frequency. The signal-to-noise ratio, measured with carrier deviation of $\pm$40% at 10% of the upper passband and without flutter compensation, was 46 dB. The level of flutter at a speed of 95.25 mm/s and a passband of 0.2 to 626 Hz was 0.40% peak-to-peak.

Sounds were processed through a 12-b data translation analog-to-digital converter board (DT2801-A) contained in an IBM-PC/XT personal computer. The board had a calibrated bandwidth of 10 kHz $\pm$ 0.1 dB and was linear within this bandwidth. All data files were recorded at a sampling rate of 10 kHz. Fast Fourier transform (FFT) analysis of the first sound was performed on the microcomputer using Signal Technology Software (ILS-PC). The entire first sound was analysed. If a systolic murmur began close to the first sound, it was difficult to determine where the first heart sound ended and the murmur began. Occasional beats in patients with a pansystolic murmur showed a space between S1 and the murmur. Patients in whom the first heart sound was not identifiable were not included in the present study. In all 48 patients investigated, the first heart sounds were selected with a rectangular window having a duration varying between 70 and 100 ms. The selected sounds were padded with zeros and a 2048-point FFT computed. The frequency spectrum was obtained in polar form and the magnitude in decibels was transformed into linear coordinates in terms of a power ratio. The amplitude of the power spectrum was then normalized to the amplitude of the dominant frequency on a relative scale varying from 0 to 100%. Examples of power spectra of the first heart sound of a patient with a normal porcine bioprosthetic valve and of a patient with a degenerated porcine bioprosthetic valve in the mitral position are shown in Figs. 1 and 2, respectively.

C. Diagnostic Feature Extraction and Selection

A description of the 18 features extracted from the phonocardiogram of each patient is given in Table II. The first nine features are those of our first study [10]. The dominant frequency F1 is defined as the frequency with the highest magnitude (0 dB) in the sound spectrum. DF1 and Q1 are features related to the damping factor of the heart-valve system. F2, the second-most dominant frequency, is defined as the frequency of the peak of the spectrum having the second highest magnitude. No restriction was imposed on the minimum value separating the two lobes of F1 and F2. F-3, F-10, and F-20 are features associated with the frequency bandwidth of the spectra (as defined in Table II) while IM20 is an area measurement used to take into account the morphology, or profile, of the spectra. IR12 is used to extract information from the relative peak-to-peak intensity of S1 and S2 (the second heart sound).

The selection of these nine diagnostic features was based upon analyses of valve vibration which produce heart sounds [22]–[25] and it was based upon studies by Stein *et al.* [5]–[7], Foale *et al.* [8], and Joo *et al.* [9].

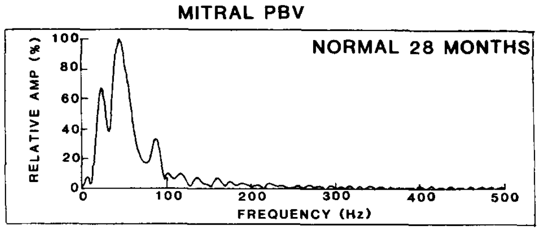


Fig. 1. Relative power spectrum of the first heart sound in a patient with a normally functioning porcine bioprosthetic valve (PBV) in the mitral position. The valve had been inserted for 2 years, 4 months. The dominant frequency was 45 Hz.

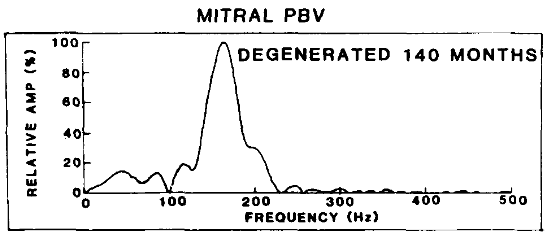


Fig. 2. Relative power spectrum of the first heart sound in a patient with a degenerated porcine bioprosthetic valve (PBV) in the mitral position. The valve was removed a few days after the sounds were recorded; it had been inserted for 11 years 8 months. The dominant frequency was 164 Hz.

TABLE II
DESCRIPTION OF VALVE SOUND DIAGNOSTIC FEATURES

No.	Feature	Description
1	F1	: The dominant frequency peak (0 dB)
2	DF1	: Bandwidth of F1 at −3 dB
3	Q1	: Quality factor of F1 (Q1 = F1/DF1)
4	F2	: The second dominant frequency peak
5	F-3	: Highest frequency found at −3 dB
6	F-10	: Highest frequency found at −10 dB
7	F-20	: Highest frequency found at −20 dB
8	IM20	: Integrated mean area above −20 dB
9	IR12	: Intensity ratio of the first heart sound to second heart sound ($S1/S2$)
10	AREA	: Total area of the spectrum
11	A201	: Area in the 20–100 Hz bandwidth
12	A1002	: Area in the 100–200 Hz bandwidth
13	A2003	: Area in the 200–300 Hz bandwidth
14	rms	: rms value of the spectrum
15	R201	: rms value in the 20–100 Hz bandwidth
16	R1002	: rms value in the 100–200 Hz bandwidth
17	R2003	: rms value in the 200–300 Hz bandwidth
18	FM	: The median frequency

The 0 dB intensity reference corresponds to the maximal value of spectrum (100% on the linear scale; see Figs. 1 and 2).

The analysis of valve vibration and factors that effect the frequency of heart sounds showed that the natural modes of resonance of biological valves increase with stiffening of the valve leaflets [22]–[25]. Clinical studies confirmed this and also showed that the second dominant frequency shifts toward higher frequencies [8], [9].

The second series of features was derived on the basis of visual inspection of several sound spectra from normal and degenerated valves [11]. It appeared that area measurements of the sound spectra in three frequency bands could provide additional discriminant features. The area features were obtained by summing the values of the spectral coefficients in the various frequency ranges of interest while the rms values were obtained by summing the squared value of the spectral coefficients in the same frequency ranges, dividing by the number of frequency coefficients included in the computation, and taking the square root of this mean value. The median frequency was also investigated. It was computed as the frequency that corresponds to the median of the spectral coefficients. For each patient, the 18 diagnostic features were extracted from the spectrum of the first heart sound. Because of the large amount of data involved in the analysis of these spectra, we visually selected three typical appearing beats in each patient. Each spectrum was then analyzed separately and the values of the diagnostic features averaged.

Feature selection is a process used to choose the best patterns from the ensemble of features considered to have a discriminant power among the different classes. The approach used in the present study was to test the discriminant value of all diagnostic patterns composed of a number of features varying between two and the optimal ratio $m = N_i/5$. Since the population of the degenerated valves was 28, the optimal ratio was chosen as $m = 6$. This represents a huge amount of computation since each classifier had to be designed and tested 31 161 times (all combinations of 2, 3, 4, 5, and 6 features chosen among 18). The algorithms were written in the C language and run on an Intel 301 computer system (a 16 MHz 386 IBM-PC/AT compatible personal computer).

D. Classifier Evaluation

The performance of the Bayes method (assuming that the probability density functions were Gaussian) and that of the nearest neighbor method to identify the status of the bioprostheses were compared with the clinical classification of the bioprostheses. The leave-one-out algorithm was utilized to estimate the probability of error of the classifiers. The Euclidean distance (ED), the normalized Euclidean distance (NED), and the Mahalanobis distance (MD) were investigated to find the best measurement of pattern similarity for the nearest neighbor method. These distance measurements, computed between two patterns X_1 and X_2 were evaluated by using the following formulae:

$$\mathrm{ED} = \sum_{i=1}^{D} (x_{1i} - x_{2i})^2 \tag{1}$$

$$\mathrm{NED} = \sum_{i=1}^{D} (x_{1i} - x_{2i})^2 / \boldsymbol{x}_i^2 \tag{2}$$

$$\mathrm{MD} = (X_1 - X_2)^{t} C^{-1} (X_1 - X_2) \tag{3}$$

where D is the dimension of the patterns, $\boldsymbol{x}_i$ is defined by

$$\boldsymbol{x}_i = \frac{1}{N} \sum_{j=1}^{N} (x_{ji}), \tag{4}$$

C^{-1} is the inverse of the covariance matrix estimated by

$$C = \frac{\sum_{j=1}^{N} X_j X_j^t - \sum_{i=1}^{N} MM^t}{N - 1} \tag{5}$$

and

$$M = \frac{1}{N} \sum_{j=1}^{N} X_j. \tag{6}$$

Parameter N is the total number of vectors (bioprostheses) included in the study while the vectors used in these equations are defined by

$$X_i^t = (x_{i1}, x_{i2}, \cdots, x_{iD}).$$

The number of neighbors included in the decision rule of the nearest neighbor algorithm was varied from 1 to 13 (range of $\sqrt{N} \pm 6$) to find the best value of K.

For both methods, the performance of the classifiers was evaluated by computing the percentage of correct classifications (%CC's), false positives (%FP's), and false negatives (%FN's) by using

$$\%\text{CC's} = 100 * (\text{TP} + \text{TN})/N \tag{7}$$

$$\%\text{FP's} = 100 * \text{FP}/(\text{TN} + \text{FP}) \tag{8}$$

$$\%\text{FN's} = 100 * \text{FN}/(\text{TP} + \text{FN}) \tag{9}$$

where TP was the number of true positives, FP the number of false positives, TN the number of true negatives, and FN the number of false negatives. A true positive was a degenerated valve classified as degenerated while a false positive was a normal valve classified as degenerated. Similarly, a true negative was a normal valve classified as normal while a false negative was a degenerated valve classified as normal. Sensitivity (SE), specificity (SP), positive predictive value (PPV), and negative predictive value (NPV) were also estimated by using the following equations [26]:

$$\text{SE} = 100 * \text{TP}/(\text{TP} + \text{FN}) \tag{10}$$

$$\text{SP} = 100 * \text{TN}/(\text{TN} + \text{FP}) \tag{11}$$

$$\text{PPV} = 100 * \text{TP}/(\text{TP} + \text{FP}) \tag{12}$$

$$\text{NPV} = 100 * \text{TN}/(\text{TN} + \text{FN}). \tag{13}$$

IV. Results

As a first step, we tested, using the holdout method described in our previous study [10], the performance of patterns composed of various subset combinations of the nine features (features 10 to 18 in Table II) evaluated by Stein *et al.* [11]. The results (80% of CC's, 14% of FP's, and 25.5% of FN's) were only slightly better than those obtained during the first study (77% of CC's, 15% of FP's, and 30% of FN's). In a second step, we decided to combine all features thus creating the data base used in the present study.

The numbers of patterns providing a percentage of correct classifications of 90% or higher for the automatic classification of normal and degenerated mitral bioprosthetic valves for the different methods evaluated are shown in Table III. Results are presented for four different classifiers. Each classifier had two to six features of the sound spectra selected from Table II. Table III also indicates how many patterns provide the highest performance of each classifier. As shown by this table, the performance of all classifiers improved as the number of features was increased. The best performance of the classifiers varied between 85 and 98% correct classifications. With the Bayes classifier, a total of 511 different patterns can generate a performance equal or higher than 90% correct classifications. For the nearest neighbor method, a percentage of correct classifications higher than or equal to 90% can be reached by using seven different patterns with the Euclidean distance, 169 patterns with the normalized Euclidean distance, and 111 patterns with the Mahalanobis distance. The Mahalanobis and the normalized Euclidean distances were the best measurements of similarity between the patterns of each class. They provided the best results (94% correct classifications) for the nearest neighbor method. However, the Mahalanobis distance was slightly superior to the normalized Euclidean distance because five patterns, each containing between three and six features, provided 94% correct classifications compared to only one of dimension 6 with the normalized Euclidean distance. In most cases, the best results of the nearest neighbor method were obtained with a K value of 1, 3, or 5, irrespective of the type of distance measurement used.

The Bayes method provided the best performance (98% correct classification) but required patterns composed of six features to reach this level. However, with this number of features in the patterns, there were 500 combinations of features that provided 90% or more correct classifications. On the other hand, the nearest neighbor algorithm based on the Mahalanobis distance showed a good performance (94% correct classifications) with a pattern composed of only three diagnostic features. In addition, the performance of this algorithm was unchanged irrespective of whether 3, 4, 5, or 6 features were used. The detailed descriptions of the combinations of features providing the best performances for the nearest neighbor algorithm using the Mahalanobis distance and the Bayes algorithm are given in Tables IV and V, respectively.

Table IV shows that the most discriminant features to use with the Mahalanobis distance were: the highest frequency at −10 dB (F−10), the total area of the spectrum (AREA) and the rms value in the 20–100 Hz band (R201). These three features appeared in four of the five most discriminant features (94% correct classifications). Among the other discriminant features, we found the integrated mean area above −20 dB (IM20), the area in the 20–100 Hz band (A201), that in the 100–200 Hz band (A1002) and that in the 200–300 Hz band (A2003). Sensitivity, specificity, positive predictive value and negative predictive value were always higher than 85%.

TABLE III
NUMBER OF PATTERNS PROVIDING ≥90% CORRECT CLASSIFICATIONS FOR NORMAL AND DEGENERATED MITRAL BIOPROSTHETIC VALVES

Type of Classifier	# of Features per Pattern	# of Patterns Providing ≥ 90% of Correct Classifications	Highest Performance	# of Patterns Providing the Highest Performance
Nearest	2	0	85%	1
Neighbor	3	1	90%	1
	4	2	90%	2
(Euclidean)	5	2	92%	1
(Distance)	6	2	92%	1
Nearest	2	0	88%	1
Neighbor	3	6	90%	6
(Normalized)	4	14	92%	3
(Euclidean)	5	52	92%	12
(Distance)	6	97	94%	1
Nearest	2	1	90%	1
Neighbor	3	6	94%	1
	4	25	94%	2
(Mahalanobis)	5	31	94%	1
(Distance)	6	48	94%	1
	2	1	90%	1
	3	0	88%	9
Bayes	4	3	90%	3
	5	7	90%	7
	6	500	98%	2

TABLE IV
BEST PERFORMANCES OF THE NEAREST NEIGHBOR ALGORITHM WITH MAHALANOBIS DISTANCE FOR AUTOMATIC CLASSIFICATION OF NORMAL AND DEGENERATED MITRAL BIOPROSTHETIC VALVES

κ	Most Discriminant Patterns						Correct Classifications	False Positives	False Negatives	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
1	F-10	IM20	AREA	A201	R201	R2003	94%	10%	4%	96.4%	90.0%	93.1%	94.7%
1	F-10	IM20	AREA	A201	A2003	R2003	92%	10%	7%	92.9%	90.0%	92.9%	90.0%
1	F-10	IM20	AREA	A201	A2003	R201	92%	10%	7%	92.9%	90.0%	92.9%	90.0%
5	IM20	DF1	AREA	A201	A2003	R1002	92%	10%	7%	92.9%	90.0%	92.9%	90.0%
1	F-3	IM20	MF	A201	R201	R1002	92%	10%	7%	92.9%	90.0%	92.9%	90.0%
5	IM20	DF1	AREA	A201	A1002	A2003	92%	10%	7%	92.9%	90.0%	92.9%	90.0%
1	F-10	DF1	rms	A201	A2003	R1002	92%	10%	7%	92.9%	90.0%	92.9%	90.0%
1	F-10	IM20	AREA	A201	R2003		94%	5%	7%	92.9%	95.0%	96.3%	90.0%
1	IM20	Q1	AREA	A1002	R201		92%	10%	7%	92.9%	90.0%	92.9%	90.0%
1	F-3	IM20	MF	A201	R1002		92%	10%	7%	92.9%	90.0%	92.9%	90.0%
3	F-10	MF	AREA	rms	A1002		92%	10%	7%	92.9%	90.0%	92.9%	90.0%
3	F-10	MF	AREA	rms	R201		92%	10%	7%	92.9%	90.0%	92.9%	90.0%
1	F-10	IM20	AREA	A201	A2003		92%	5%	11%	89.3%	95.0%	96.2%	86.4%
1	F-10	IM20	AREA	A2003	R201		92%	5%	11%	89.3%	95.0%	96.2%	86.4%
1	F-10	IM20	AREA	R201	R2003		92%	5%	11%	89.3%	95.0%	96.2%	86.4%
3	F-10	AREA	rms	R201			94%	10%	4%	96.4%	90.0%	93.1%	94.7%
3	F-10	AREA	A1002	R201			94%	10%	4%	96.4%	90.0%	93.1%	94.7%
13	IM20	DF1	AREA	rms			92%	10%	7%	92.9%	90.0%	92.9%	90.0%
3	F-10	AREA	rms	A1002			92%	10%	7%	92.9%	90.0%	92.9%	90.0%
3	F-10	AREA	R201	R1002			92%	10%	7%	92.9%	90.0%	92.9%	90.0%
3	F-10	AREA	R201				94%	10%	4%	96.4%	90.0%	93.1%	94.7%
5	F-10	AREA	rms				92%	15%	4%	96.4%	85.0%	90.0%	94.4%
3	F-10	rms	A1002				92%	10%	7%	92.9%	90.0%	92.9%	90.0%

Table V shows that the best performance of the Bayes classifier (98% correct classifications) was obtained by using patterns composed of the following two sets of diagnostic features: The dominant frequency (F1), the integrated mean area above − 20 dB (IM20) or the total area of the spectrum (AREA), the rms value of the spectrum (rms), the area of the spectrum in the 100–200 Hz frequency band (A1002), the rms value of the spectrum

TABLE V
BEST PERFORMANCES OF THE BAYES ALGORITHM FOR AUTOMATIC CLASSIFICATION OF NORMAL AND DEGENERATED MITRAL BIOPROSTHETIC VALVES

Most Discriminant Patterns						Correct Classifications	False Positives	False Negatives	Sensitivity	Specificity	Positive Predictive Value	Negative Predictive Value
F1	IM20	rms	A1002	R1002	R2003	98%	0%	4%	96.4%	100.0%	100.0%	95.2%
F1	AREA	rms	A1002	R1002	R2003	98%	0%	4%	96.4%	100.0%	100.0%	95.2%
F-3	F-10	DF1	Q1	AREA	R2003	96%	10%	0%	100.0%	90.0%	93.3%	100.0%
F1	F2	rms	A1002	R1002	R2003	96%	5%	4%	96.4%	95.0%	96.4%	95.0%
F1	Q1	rms	A1002	R1002	R2003	96%	5%	4%	96.4%	95.0%	96.4%	95.0%
F1	F-3	AREA	A1002	R1002	R2003	96%	5%	4%	96.4%	95.0%	96.4%	95.0%
F1	F2	F-20	AREA	R1002	R2003	96%	5%	4%	96.4%	95.0%	96.4%	95.0%
F1	MF	rms	A1002	R1002	R2003	96%	5%	4%	96.4%	95.0%	96.4%	95.0%
Q1	MF	A1002	R201	R1002	R2003	96%	0%	7%	92.9%	100.0%	100.0%	90.9%
F2	Q1	rms	A1002	R1002	R2003	96%	0%	7%	92.9%	100.0%	100.0%	90.9%
F2	Q1	A1002	R201	R1002	R2003	96%	0%	7%	92.9%	100.0%	100.0%	90.9%
DF1	Q1	A1002	R201	R1002	R2003	96%	0%	7%	92.9%	100.0%	100.0%	90.9%
F-3	Q1	A1002	R201	R1002	R2003	96%	0%	7%	92.9%	100.0%	100.0%	90.9%
IM20	Q1	A1002	R201	R1002	R2003	96%	0%	7%	92.9%	100.0%	100.0%	90.9%
F1	DF1	AREA	A1002	R1002	R2003	96%	0%	7%	92.9%	100.0%	100.0%	90.9%

in the 100–200 Hz frequency band (R1002), and the rms value of the spectrum in the 200–300 Hz frequency band (R2003). For both patterns, the percentage of false positives was 0% and that of false negatives was 4%. Sensitivity was 96%, specificity was 100%, positive predictive value was 100%, and negative predictive value was 95%. Various combinations of these and other features provided 96% correct classifications. The rms value in the 200–300 Hz band (R2003) appeared in every discriminant pattern while the rms value in the 100–200 Hz band (R1002) appeared in almost every discriminant pattern (Table V).

Finally, the first-to-second sound intensity ratio (IR12) was the only feature that did not appear in any of the most-discriminant patterns of the four algorithms tested.

V. DISCUSSION

Joo *et al.* [9] demonstrated the diagnostic potential of a Gaussian–Bayesian classifier for detecting degenerated bioprostheses implanted in the aortic position. They estimated the performance of their classifier using the holdout method and the two most-dominant frequencies (F1 and F2) of the bioprosthetic closure sound spectra as diagnostic features. Spectral analysis was done with pole-zero modeling using the Steiglitz–McBride method. Their training set was composed of 14 normal and 6 degenerated bioprostheses while their test set was composed of 13 normal and 7 degenerated bioprostheses. Their results showed that their classifier correctly classified 17 of the 20 patients. The three misclassifications were false positives. Their results were subsequently reevaluated by BeMent and Veeneman [27] and some computational errors corrected. The performance of the corrected classifier was found to be higher: the three false positive classifications were reduced to only one.

As a first step, we attempted to reproduce the results of Joo *et al.* [9] by using only F1 and F2 as diagnostic features and the FFT to estimate the sound spectra. The performance of the Bayesian classifier was disappointing with 45% correct classifications, 0% false positives, and 61% false negatives. Better results (65% correct classifications) were obtained with the nearest neighbor method based on the Mahalanobis distance. By excluding bioprostheses having no calcium deposits from the degenerated class, the performance of the Bayesian classifier was increased to 63% correct classifications, 0% false positives, and 54% false negatives. The performance of the nearest neighbor classifier was decreased to 62% correct classifications, 40% false positives, and 36% false negatives.

The difference between the results of Joo *et al.* [9] and ours, when using the two dominant frequencies of the valve closure sound spectra, can be ascribed to: 1) the better spectral resolution of the Steiglitz–McBride pole-zero method compared to that of the FFT method, and 2) the small number of degenerated valves included in the study of Joo *et al.* [9]. The performance of a classifier can be overestimated when the number of samples in a given class is very small. This is because the true statistical distribution of the patterns can be overestimated or biased at the advantage of the samples available.

Studies by Durand *et al.* [28] and Cloutier *et al.* [29], [30] showed that the FFT is the best method for the estimation of the most dominant frequency (F1) of the prosthetic valve sounds while the Steiglitz–McBride pole-zero method is the best for estimating its second most-dominant frequency (F2). The results of these studies [28]–[30] suggest that the performance of the classifier tested by Joo *et al.* [9] was overestimated by the small number of patients with degenerated bioprostheses (six in the training set and seven in the test set) since the holdout method is known to generate an unreliable estimate of the real performance of the classifier unless the training and test set populations are very large [15].

A secondary conclusion of the study of Joo *et al.* was that the FFT algorithm does not have sufficient frequency resolution to extract the diagnostic information associated with the two dominant frequencies, F1 and F2, of the sound spectrum. The present study shows on the contrary that a simple FFT algorithm can provide significant diagnostic information. However, the diagnostic information is not extracted in the same way. With the FFT algorithm, the diagnostic information is associated with the rms value or the area of the spectrum in various frequency bands combined with point measurements such as F1 or F-10. The advantages of rms or area values, compared to secondary peaks like F2, is that they are less sensitive to the spectral resolution of the power spectrum estimation technique used because they are obtained by computing some averaging functions over the frequency bands of interest.

According to the rule of thumb suggested by Trunk *et al.* [18] and Kalayeh and Langrebe [19], the optimal ratio of sample dimension to the number of features used per class should be five or higher. In the present study, this ratio is respected by the nearest neighbor algorithm (94% correct classification with three features) but not by the Bayes method for the normal population. For example, there were 20 patients in the normal valve population, 28 in the degenerated population, and the best results were obtained with patterns composed of six diagnostic features. Thus, the ratio of sample dimension to the number of features is 3.3 for the normal bioprosthetic valves and 4.7 for the degenerated bioprosthetic valves. This suggests that the best performance of the Bayes classifier obtained in the present study (98% correct classification) may be slightly overestimated. However, these ratios (3.3 and 4.7) are in agreement with the criterion of Foley [17] which states that a ratio of sample dimension to the number of features greater than or equal to three is sufficient to provide a good estimate of the performance of the classifier.

Our study also shows that the number and the type of discriminant diagnostic features that must be extracted from the valve closure sound spectra depend upon the type of pattern recognition algorithm used. Identification of the best discriminant patterns for each algorithm tested requires a careful evaluation of all possible combinations of these diagnostic features. Since this process is computationaly intensive, it has become acceptable only with the advent of more powerful computers. This conclusion has also been reported by Cover and Van Campenhout [31] who showed that all possible combinations of features must be evaluated to determine the best patterns. To avoid this exhaustive search, the heuristics approaches (sequential forward selection, sequential backward selection, and plus 1-take away *r*) described by Devijver and Kittler [13] could be useful. However, these techniques do not guarantee that the best patterns will be found.

The Gaussian-Bayesian classifier and the nearest neighbor classifier using the Mahalanobis distance are the most discriminant classifiers because they both take into consideration the statistical properties of the diagnostic patterns. The estimated performance of the classifiers described in the present study when the number of diagnostic features is four or less, is probably highly conservative because of the following:

1) Care has been taken to respect recognized criteria determining the ratio of the sample dimension to the number of features that should be used in the design of the pattern recognition systems.

2) The probability of error of the classifiers was evaluated with the conservative leave-one-out method.

3) All four classifier algorithms tested showed that many patterns can provide a level of 90% correct classifications or more.

4) The bioprosthetic valve status was known more precisely than in the study of Joo *et al.* [9]. This is especially true for the degenerated valves, which were examined and their degree of degeneration quantified after surgical removal.

One interesting aspect of our study is the fact that a simple FFT algorithm can provide discriminant diagnostic features. As emphasized in section II of the paper, once the discriminant function of the Bayes classifier or the distance metric and the K parameter of the nearest neighbor classifier are known, the classification of an unknown bioprosthesis status can be fast (within a few seconds) if the spectral estimate of the valve closure sound does not require a long computational time. This could be an important advantage of the FFT technique over the Steiglitz–McBride pole-zero method which may require between 2 and 5 min of computing time on an IBM-PC/AT personal computer before the diagnostic patterns can be extracted from the valve closure sound spectrum.

The present study shows that reliable diagnostic information about the status of bioprosthetic valves inserted in the mitral position is contained in the spectrum of the first heart sound. The results obtained are encouraging and indicate that it is now possible to design a low-cost noninvasive system for periodic evaluation and detection of degenerated bioprosthetic heart valves.

References

[1] P. D. Stein, J. J. Collins, Jr., and A. Kantrowitz, "Antithrombotic therapy in mechanical and biological prosthetic heart valves and saphenous vein bypass grafts," *Chest*, vol. 89, (supp.), pp. 46S–53S, Feb. 1986.

[2] G. Thiene, E. Arbustini, U. Bortolotti, E. Talenti, A. Milano, M. Valente, G. Molin, and V. Gallucci, "Pathologic substrates of porcine valve dysfunction," in *Cardiac Bioprostheses*. New York: Yorke Med. Books, 1982, ch. 34, pp. 378–400.

[3] R. J. Levy, F. J. Shoen, and G. Golomb, "Bioprosthetic heart valve calcification: Clinical features, pathobiology, and prospect for prevention," *CRC Crit. Rev. Biocom.*, vol. 2, pp. 147–187, 1986.

[4] P. D. Stein, S. R. Kemp, J. M. Riddle, M. W. Lee, J. W. Lewis, Jr., and D. J. Magilligan, Jr., "Relation of calcification to torn leaflets of spontaneously degenerated porcine bioprosthetic valves," *Ann. Thorac. Surg.*, vol. 40, pp. 175–180, July, 1985.

[5] P. D. Stein, H. N. Sabbah, J. B. Lakier, and S. Goldstein, "Frequency spectrum of the aortic component of the second heart sound in patients with normal valves aortic stenosis and aortic porcine xenograph," *Amer. J. Cardiol.*, vol. 46, pp. 48–52, July, 1980.

[6] P. D. Stein, H. N. Sabbah, J. B. Lakier, D. J. Magilligan, Jr., and S. Goldstein, "Frequency of the first heart sound in the assessment of stiffening of mitral bioprosthetic valves," *Circulation*, vol. 63, pp. 200–204, 1981.

[7] P. D. Stein, H. N. Sabbah, J. B. Lakier, S. R. Kemp, and D. J. Magilligan, Jr., "Frequency spectra of the first heart sound and of the aortic component of the second heart sound in patients with degenerated porcine bioprosthetic valves," *Amer. J. Cardiol.*, vol. 53, pp. 557–561, Feb. 1984.

[8] R. A. Foale, T. H. Joo, J. H. McClellan, R. W. Metzinger, G. L. Grant, G. S. Meyers, and R. S. Lees, "Detection of aortic porcine

valve dysfunction by maximum entropy spectral analysis," *Circulation*, vol. 68, pp. 42-49, July 1983.

[9] T. H. Joo, J. H. McClellan, R. A. Foale, G. S. Meyers and R. S. Lees, "Pole-zero modeling and classification of phonocardiograms," *IEEE Trans. Biomed. Eng.*, vol. BME-30, pp. 110-118, Feb. 1983.

[10] L. G. Durand, M. Blanchard, H. N. Sabbath, M. S. Hamid, S. R. Kemp, and P. D. Stein, "A Bayes model for automatic detection and quantification of bioprosthetic valve degeneration," *Math. Comput. Model.*, vol. 11, pp. 158-163, 1988.

[11] P. D. Stein, H. N. Sabbah, M. S. Hamid, and D. J. Magilligan, Jr., "Valve assessment by sound," in *Proc. 11th Annu. Contractors' Grantees Meet. Dev., Technol. Branch*, Nat. Heart, Lung, Blood Inst., Bethesda, MD, 1987, p. 173.

[12] T. Bailey, and A. K. Jain, "A note on distance-weighted K-nearest neighbor rules," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-8, pp. 311-313, Apr. 1978.

[13] P. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach.* Englewood Cliffs, NJ: Prentice Hall, 1982.

[14] J. T. Tou, and R. C. Gonzalez, *Pattern Recognition Principles.* Reading, MA: Addison-Wesley, 1974.

[15] G. T. Toussaint, "Bibliography on estimation of misclassification," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 472-479, July, 1974.

[16] A. K. Jain, "Advances in statistical pattern recognition;" in *Pattern Recognition Theory and Applications.* Berlin, Heidelberg; Springer-Verlag, 1987, vol. F30, pp. 1-19.

[17] D. H. Foley, "Considerations of sample and feature size," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 618-626, Sept. 1972.

[18] G. V. Trunk, "A problem of dimensionality: A simple example," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 1, pp. 306-307, July, 1979.

[19] M. M. Kalayeh and D. A. Landgrebe, "Predicting the required number of training samples," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 5, pp. 644-667, Nov. 1983.

[20] H. S. Rosman, M. Alam, J. B. Lakier, S. R. Kemp, H. N. Sabbah, D. J. Magilligan, Jr., and P. D. Stein, "Utility of physical examination and noninvasive tests in the diagnosis of degeneration of porcine bioprosthetic valves in the mitral position," *Amer. J. Noninvas. Cardiol.*, vol. 2,pp. 48-51, 1988.

[21] P. D. Stein, *A Physical and Physiological Basis for the Interpretation of Cardiac Auscultation: Evaluations Based Primarily on Second Sound and Ejection Murmurs.* Mount Kisco, NY: Futura, 1981, pp. 113-124.

[22] E. Blick, H. N. Sabbah, and P. D. Stein, "One-dimensional model of diastolic semilunar valve vibrations productive of heart sounds," *J. Biomechan.*, vol. 12, pp. 223-227, 1979.

[23] M. S. Hamid, H. N. Sabbah, and P. D. Stein, "Vibrational analysis of bioprosthetic heart valve leaflets using numerical models: Effects of leaflets stiffening, calcification and perforation," *Cir. Res.* 61:687-694, 1987.

[24] T. C. Hearn, D. N. Gopal, D. N. Ghista, J. Robinson, H. Tihal, J. Mazumdar, and R. Bogner, "Towards quantitative determination of valvular calcification and pressure drops across malfunctioning heart valves," in *Proc. IEEE Frontiers Eng. Comput. Health Care*, 1983, pp. 127-131.

[25] D. L. Sikarskie, P. D. Stein, and M. Vable, "A mathematical model of aortic valve vibration," *J. Biomechan.*, vol. 17, pp. 831-837, 1984.

[26] A. F. Feinstein, "On the sensitivity, specificity, and discrimination of diagnostic tests," in *Clinical Biostatistics.* St. Louis: C. V. Mosby 1977, ch. 15, pp. 214-226.

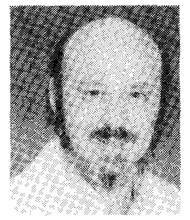
[27] S. L. BeMent and D. E. Veeneman, "Comments on Pole-zero modeling and classification of phonocardiograms," *IEEE Trans. Biomed Eng.*, vol. 31, pp. 473-474, June, 1984.

[28] L. G. Durand, J. De Guise, G. Cloutier, R. Guardo, and M. Brais, "Evaluation of FFT-based and modern parametric methods for the spectral analysis of bioprosthetic valve sounds," *IEEE Trans. Biomed. Eng.*, vol. 33, pp. 572-578, June, 1986.

[29] G. Cloutier, R. Guardo, and L. G. Durand, "Spectral analysis of closing sounds produced by Ionescu-Shiley bioprosthetic heart valves. Part 3: Performance of FFT-based and parametric methods for extracting diagnostic spectral parameters," *Med. Biol. Eng. Comput.*, vol. 25, pp. 497-503, Sept. 1987.

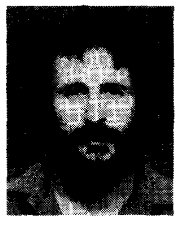
[30] G. Cloutier, L. G. Durand, R. Guardo, H. N. Sabbah, and P. D.Stein, "Bias and variability of diagnostic spectral parameters extracted from closing sounds produced by bioprosthetic valves implanted in the mitral position," *IEEE Trans. Biomed. Eng.*, vol. 36, Aug. 1989.

[31] T. M. Cover and J. M. Campenhout, "On the possible ordering in the measurement selection problem," *IEEE Trans. Syst. Man Cybern.*, vol. SMC-7, pp. 657-661, Sept. 1977.



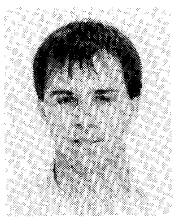
Louis-Gilles Durand (M'78-SM'85) was born in St-Jean de Matha, P. Q., Canada, in 1949. He received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Ecole Polytechnique, University of Montreal in 1975, 1979, and 1983, respectively.

Currently, he is Director of the Biomedical Engineering Laboratory at the Clinical Research Institute of Montreal, Research Associate Professor in the Department of Medicine at the University of Montreal and Adjunct Professor at the Institute of Biomedical Engineering of the Ecole Polytechnique, Montreal, Canada. His current interests are digital signal processing of the phonocardiogram and echo-Doppler signal, modeling of the heart-thorax acoustic system, and medical instrumentation.



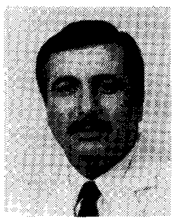
Michel Blanchard (M'85) was born in Pont-Viau, P. Q., Canada, in 1957. He received the B.Eng., and M.Sc. degrees in electrical engineering from the Ecole Polytechnique, University of Montreal in 1981 and 1983, respectively.

From 1983 to 1989 he was a research engineer at the Laboratory of Biomedical Engineering of the Clinical Research Institute of Montreal. He is currently a development engineer at Modular Vision System Inc., Montreal, Canada. His major interests are digital signal processing, computer sciences and medical instrumentation.



Guy Cloutier (S'88) was born in Trois-Rivières, Québec, Canada, in 1961. He received the B.Eng. degree in electrical engineering from the Universitè du Québec à Trois-Rivières in 1984, and the M.Sc. and Ph.D. degrees in biomedical engineering from the Ecole Polytechnique, University of Montreal in 1986 and 1990, respectively.

Currently, he is a research fellow of the Heart and Stroke Foundation of Canada and pursues postdoctoral researches at the Bioengineering Program, Pennsylvania State University, University Park. His principal research interests are ultrasonic tissue characterization, digital signal processing and pattern recognition applied to Doppler echocardiographic and phonocardiographic signals.



Hani N. Sabbah received the Bachelor of Science degree in aerospace engineering from the University of Oklahoma, Norman, in 1972, and the doctorate in biomedical sciences from Oakland University, Rochester, MI, in 1988.

He is Associate Director of Cardiovascular Research at the Henry Ford Heart and Vascular Institute. He has worked in Cardiovascular Research since 1972, both at the University of Oklahoma, and subsequently, at Henry Ford Hospital. He is the author of over 180 scientific publications.

Paul D. Stein received the Bachelor of Science degree in physics in 1955 and the M.D. degree in 1959, both from the University of Cincinnati, Cincinnati, OH.

Following internship and residency in internal medicine, he took post-graduate fellowships Cardiology at the University of Cincinnati Medical Center, Mount Sinai Medical Center, and Harvard Medical School. He was at the Univeristy of Oklahoma from 1969 to 1976, where he was Professor of Research Medicine and Director of the Cardiac Catheterization Laboratory. Since 1976, he has been at Henry Ford Hospital in Detroit, where he is Director of Cardiovascular Research. He is also Adjunct Professor of Physics at Oakland University in Rochester, Michigan.