# TFR-based Feature Extraction using PCA Approaches for Discrimination of Heart Murmurs

D. Avendaño-Valencia[1], F. Martinez-Tabares[1], D. Acosta-Medina[1],
I. Godino-Llorente[2], G. Castellanos-Dominguez[1]

*Abstract*— Discrimination of murmurs in heart sounds is accomplished by means of time–frequency representations (TFR) which help to deal with non–stationarity. Nevertheless, classification with TFR is not straightforward given their large dimension and redundancy. In this paper we compare several methodologies to apply Principal Component Analysis (PCA) to TFR as a dimensional reduction scheme, which differ in the form that features are represented. Besides, we propose a method which maximizes information among TFR preserving information within TFRs. Results show that the methodologies that represent TFRs as matrices improve discrimination of heart murmurs, and that the proposed methodology shrinks variability of the results.

## I. INTRODUCTION

Cardiac murmurs are non–stationary signals that exhibit sudden frequency changes and transients, therefore, the time–frequency representation (TFR) has been proposed before to investigate the correlation between the time–frequency ($t$–$f$) characteristics of murmurs and the subjacent cardiac pathologies [1]. For that matter, parametric estimations of TFR that are based upon parameterized expressions of the time–dependent autoregressive modeling are generally employed [2]. Due to its intrinsic generality, parametric time varying autoregressive (TVAR) models had provided useful empirical representations of non–stationary time series in biomedical signal analysis [3].

Despite of the appealing features of TFR to deal with non–stationary signals, their major drawback to use them in classification is the large quantity of redundant data which they contain. Thus, there is a growing need for new data reduction methods that can accurately parameterize the activity in TFR of biosignals [4], being PCA a widely used technique which performs a singular

value decomposition $t$–$f$ domains. PCA transformation produces an uncorrelated feature set by projecting the data onto the eigenvectors of maximum variability, providing a mean of dimensionality reduction. The question arises on how PCA should be accomplished in TFR in order to reduce dimension of $t$–$f$ planes keeping information to maximize accuracy rate of classification.

A major motivation in this work is to generate a set of parametric TFR–based features extracted from PCG recordings, capable of detecting murmurs with higher accuracy than using static features. So, the aim of the present work is to evaluate the best set of dynamic features, estimated from parametric–based TFR and extracted with different forms of linear decomposition methods, suitable for the classification of heart murmurs. As criteria of comparison classifier accuracy is suggested, namely, by using the well-known $k$-nearest neighbors ($k$–nn) approach, that is assumed to be adequate, since it directly measures the distance from a test set item to each of the training set items immersed in Euclidean $t$–$f$ planes.

## II. BACKGROUND

### A. TFR–based Feature Extraction using PCA

*Conventional PCA*. Let $\Theta = \{\theta_j : j = 1, \ldots, n\}$ be a set of objects described for $p$ random variables $\{\xi_i, i = 1, \ldots, p, \}$. That is, for each object we have the data set $(\xi_1(\theta_j), \xi_2(\theta_j), \cdots, \xi_n(\theta_j))^T \in \mathbb{R}^p$, and can build the centralized data matrix:

$$\boldsymbol{X} = [\theta_1 - \overline{\theta}|\theta_2 - \overline{\theta}| \cdots |\theta_n - \overline{\theta}]^\mathsf{T}, \quad \overline{\theta} = \frac{1}{n}\sum_{j=1}^{n} \theta_j \quad (1)$$

The conventional PCA looks for an orthogonal transformation ($\boldsymbol{W}^\mathsf{T}\boldsymbol{W} = \boldsymbol{I}_q$), $\boldsymbol{W}_{p \times q}$, projecting the data onto a new set of variables with maximum variance. For that, we set $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{W}$ when

$$\boldsymbol{W} = \arg\max_{\boldsymbol{W}} \operatorname{tr}(\boldsymbol{W}^\mathsf{T}\boldsymbol{X}^\mathsf{T}\boldsymbol{X}\boldsymbol{W}).$$

[1]G. Control y Procesamiento Digital de Señales, Universidad Nacional de Colombia `ldavendanov@unal.edu.co` [2]Universidad Politécnica de Madrid

In practice, the solution is found by setting the columns of $\boldsymbol{W}$ to the $q$ leading eigenvectors of the covariance matrix $\boldsymbol{X}^\mathsf{T}\boldsymbol{X}$.

From now, we consider the variables $\{\xi_i\}$ are also time–dependent and have been mesaured upon a set of $m$ time instants. So, for each object we have the data set $\left\{\xi_i^k(\theta_j) : i = 1, \ldots, p, k = 1, \ldots, m\right\}$, where notation $\xi_i^k(\theta_j)$ stands for $i$−th variable, measured for $j$−th object, at $k$ instant of time.

*Eigenplane–approach.* This approach deals with the stochastic nature of variables by assuming that each instant of time $\xi_i^k(\theta_j), \forall j$, constitutes a new random variable. Therefore, each object is described by:

$$\boldsymbol{\theta}_j = \left[\xi_1^1(\theta_j), \ldots, \xi_1^m(\theta_j), \cdots, \xi_p^1(\theta_j), \ldots, \xi_p^m(\theta_j)\right] \quad (2)$$

and conventional PCA is carried out over the rewritten from (1) centralized data matrix.

*2DPCA–enhancement.* Further refinement of object description (2) can be achieved if the above vector object representation, $\boldsymbol{\theta}_j \in \Theta$, is enhanced by the next matrix, taking into account variability of the whole variable set, as suggested in [5]:

$$\boldsymbol{\theta}_j = \begin{bmatrix} \xi_1^1(\theta_j) & \xi_2^1(\theta_j) & \cdots & \xi_p^1(\theta_j) \\ \xi_1^2(\theta_j) & \xi_2^2(\theta_j) & \cdots & \xi_p^2(\theta_j) \\ \vdots & \vdots & \ddots & \vdots \\ \xi_1^m(\theta_j) & \xi_2^m(\theta_j) & \cdots & \xi_p^m(\theta_j) \end{bmatrix} \quad (3)$$

In this case, matrix of projected data $\boldsymbol{Y} = [\boldsymbol{\vartheta}_1^\mathsf{T}, \cdots, \boldsymbol{\vartheta}_n^\mathsf{T}]^\mathsf{T}$ is described by elemental matrixes $\boldsymbol{\vartheta}_j = \boldsymbol{\theta}_j\boldsymbol{W} \in \mathbb{R}^{m \times q}$. Reduction of model (3) is carried out over the column of the objects, which implies that projected variables are capturing variability of each object in time.

Nonetheless, object description (2) can be transposed, as considered in [6], to compute a transformation matrix $\boldsymbol{Z}_{m \times r}$ for reduction dimension over the rows of $\boldsymbol{\theta}_j$, and hence $\boldsymbol{\vartheta}_j = \boldsymbol{Z}^\mathsf{T}\boldsymbol{\theta}_j$, where the matrix $\boldsymbol{Z}$ is calculated over the matrix set $\left\{\boldsymbol{\phi}_j = \boldsymbol{\theta}_j^\mathsf{T} : j = 1, \ldots, n\right\}$. After calculation of arrangements $\boldsymbol{W}_{p \times q}, \boldsymbol{Z}_{m \times r}$, column-row based reduction of dimension is carried out for each $\boldsymbol{\theta}_j$, that is, $\boldsymbol{\vartheta}_{j_{r \times q}} = \boldsymbol{Z}^\mathsf{T}\boldsymbol{\theta}_j\boldsymbol{W}$. As a result, dimension reduction takes into account not only instant–by–instant variability of each random variable, given by model (2), but also check for information variability through the frequency spectra.

In 2D-PCA the projected data only reflect row or column variations on each $\boldsymbol{\theta}_j$, which implies that not all the contained information in $\boldsymbol{\theta}_j$ is covered. To solve this, another variant known as Diag 2D-PCA is to take information from diagonals on TFR matrices.

### B. Discriminant PCA–approach

It must be quoted, that PCA transformation is intended to produce an uncorrelated feature set by projecting the data onto the eigenvectors capturing the highest variability. So far, all above considered PCA-based approaches take into consideration variability along the time and among the variables themselves, as well. But, one might consider the information among objects, and therefore projected features can also reflect that variance. Pursuing of such an end, we propose to perform a PCA transformation in a discriminant way. The idea is to project the data by holding the maximum information among objects but preserving constant the variance within each object. For reaching that, we use a Regularized Discriminant Analysis (RDA)-based procedure [7].

Given the couple of sets: the object's one $\Theta = \{\boldsymbol{\theta}_j\}_{j=1}^n$, and time–dependent random variables $\xi_1, \ldots, \xi_p$, where object representation, $\boldsymbol{\theta}_j \in \Theta$, can be assumed the same as (3). Then, ortogonal transformation $\boldsymbol{W}_{p \times q}$ is accomplished when maximizing the next relation:

$$J = \frac{\mathrm{tr}(\boldsymbol{W}^\mathsf{T}\boldsymbol{G}_e\boldsymbol{W})}{\mathrm{tr}(\boldsymbol{W}^\mathsf{T}\boldsymbol{G}_i\boldsymbol{W})} \quad (4)$$

where

$$\boldsymbol{G}_i = \frac{1}{nm}\sum_{j=1}^n\sum_{k=1}^m \left(\theta_j^k - \overline{\theta_j}\right)^\mathsf{T}\left(\theta_j^k - \overline{\theta_j}\right); \quad \overline{\theta_j} = \sum_{k=1}^m \theta_j^k$$

$$\boldsymbol{G}_e = \sum_{j=1}^n \left(\overline{\theta_j} - \overline{\theta}\right)^\mathsf{T}\left(\overline{\theta_j} - \overline{\theta}\right); \quad \overline{\theta} = \sum_{j=1}^n \overline{\theta_j}$$

Notation $\theta_j^k$ stands for $k$–th row of matrix $\boldsymbol{\theta}_j$. Columns of $\boldsymbol{W}$ are obtained by $q$ eigenvectors regarding to highest $q$ eigenvalues of $(\boldsymbol{G}_i)^{-1}\boldsymbol{G}_e$. Likewise this decomposition can be carried out in rows and columns as in 2DPCA.

### III. EXPERIMENTAL SETUP

#### A. Database

The database is made up of 45 de–identified adult subjects who gave their informed consent and underwent a medical examination. A diagnosis is carried out for each patient and the severity of the valve lesion is evaluated by cardiologists according to clinical routine. A set of 26 patients is labeled as normal, while another 19 are tagged as pathological ones with evidence of systolic or diastolic murmur, caused by valve disorders

(see details in [8]). 8 recordings of 12 s corresponding to the four focuses of auscultation (mitral, tricuspid, aortic and pulmonary areas) are taken from each patient being either in post–expiratory or post–inspiratory apnea stage. As a whole, database holds 548 heart beats in total: 274 with murmurs (73 of diastolic class and 201 systolic) and 274 that are labeled as normal class.

### B. TVAR model structure

A specific TVAR model is defined by the model order $p$, the parameter vector $\boldsymbol{\alpha}[t]$, and innovations variance $\sigma_e^2[t]$; related with spectral content in $x[t]$ by [2]:

$$\boldsymbol{S}_x(t,f) = \frac{\sigma_e^2[t]}{\left|1 + \sum\limits_{k=1}^{p} \alpha_i[t]e^{-j\omega kt/f_s}\right|^2}, \quad \boldsymbol{S}_x(t,f) \subset \mathbb{R} \quad (5)$$

that can be assumed as the time–varying power spectral density of the response signal if the system were made stationary at the time instant $t$. The model order is taken $p = 7$ using Bayesian Information Criteria. Representative illustrations of TFR, estimated by Kalman filter approach [9], are shown for typical normal (Figure 1(a)) and murmur (Figure 1(b)) recordings, as well.
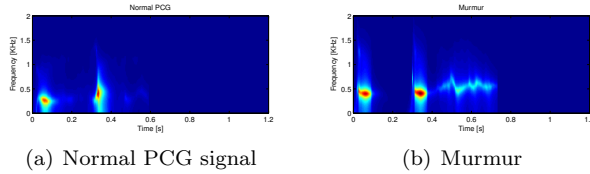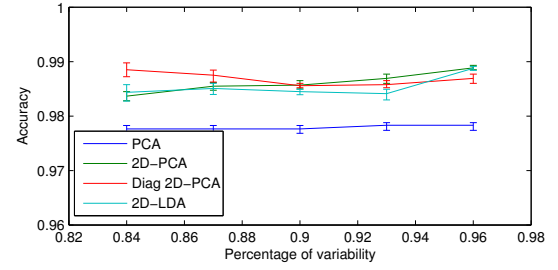


(a) Normal PCG signal      (b) Murmur

Fig. 1. Examples of TFR for the considered methods of estimation
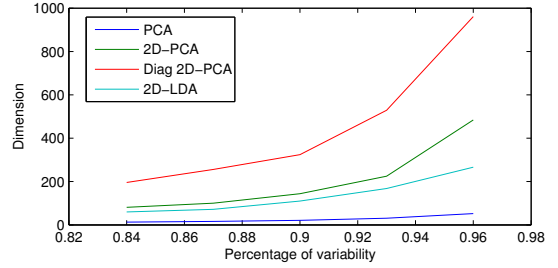
### C. Dimension reduction

Prior to accomplish any of suggested transformations, a straightforward observation of TFR in Figure 1 makes clear a big amount of nil content areas (with no informativity). Therefore, it is strongly convenient to crop nil content areas that lie adjacent to the border of the TFR. In this work, the working relevant rectangle is allocated within framework described by the time interval $0 \leq t \leq 0.6$ s and frequency band $0 \leq f \leq 2000$ Hz, leading to a selected area of $300 \times 128$ $[t \times f]$.

Effectiveness of the dimension reduction approaches (PCA, 2D-PCA, diagonal 2D-PCA and 2D-LDA) is tested by comparing the percentage of variability explained with the accuracy of the classifier. The percentage of variability is also linked with the dimension of the feature space, and the effectiveness of dimension reduction can be related to the number of components required to explain a fraction of the total variance on the dataset.



(a) Classification accuracy vs. percentage of variability explained by base vectors



(b) Dimension of feature space vs. percentage of variability explained by base vectors

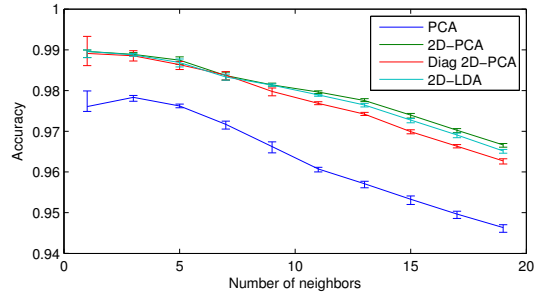Fig. 2. Tuning of number of components accomplished by the methodologies of dimension reduction



Fig. 3. Tuning of $k$–nn discriminator for the studied methodologies of dimension reduction

In Fig. 2 results of performing such test are shown. Fig. 2(a) shows that as the percentage of variance increases, the performance improves as well, except in the case of diagonal 2D-PCA, whose performance decreases. This can be explained by the excessive number of components needed by this methodology to explain the same quantity of variance. In this case, diagonal 2D-PCA might not be so recommendable to dimensionality reduction in TFR of PCG signals. PCA accomplishes the worst performance, but uses the least dimension of all the methodologies. 2D-PCA and 2D-LDA have best performance, but the dimension of 2D-LDA is slightly lower than that of 2D-PCA.
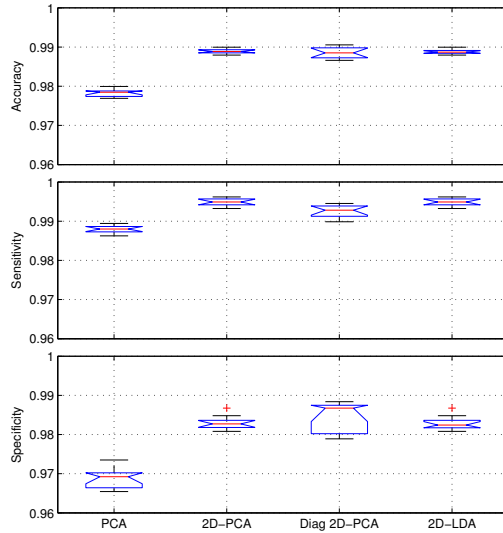
After setting up of the number of components in

Fig. 4. Comparison of accuracy, sensitivity and specificity for the best configurations of the studied methodologies

each methodology, we study the sensibility of the $k$–nn classifier as the number of neighbors increases. We test odd values from 1 to 19 neighbors and measure the accuracy of the classifier. In Fig.3 we show the results of this experiment. The overall performance of classifiers is similar to that described in the previous test. Nonetheless, it can be seen that the performance of all methodologies has a maximum for 3 neighbors, and then performance sharply decreases. This means that the decision boundary on the feature space is very irregular and shows nonlinearity of the features.

Finally, in Fig.4 we show the accuracy, sensitivity and specificity of the methodologies in their best set up. Here we conclude that 2D-PCA and 2D-LDA have the best performance with best accuracy and sensitivity and least variation. Diagonal 2D-PCA has similar median values, nevertheless its dispersion is higher. PCA has the worst performance of the discussed methods. In general, we see that the overall values of sensitivity are higher than the overall values of specificity. This means that in general, the methodologies explain better the features corresponding to murmurs than the normal heart sounds. This can be attributed to the larger variability contained on heart murmur records.

## IV. CONCLUSIONS

In this work several methodologies of dimensionality reduction in TFR were compared. The proposed methods proven be useful in dimensionality reduction,

given that the dimension of the original data is reduced in a factor lower than 2.5%. The difference among methodologies is how data is represented. Classification is improved by taking into account variability in time and frequency axes in the dimensionality reduction schemes. Also, the performance is improved when a discrimination restriction is posed on PCA, in this case, when the variability within TFR is preserved while the variability among TFR is maximized. As a future work, the proposed methodologies of object representation will be applied in more refined algorithms of data projection for dimension reduction so that information of labels can be used to improve accuracy of classification.

## V. ACKNOWLEDGMENTS

## REFERENCES

[1] E. Sejdic, I. Djurovic, and J. Jiang, "Time–frequency feature representation using energy concentration: An overview of recent advances," *Digital Signal Processing*, vol. 19, no. 1, pp. 153–183, 2009.

[2] M. Tarvainen, J. Hiltunen, P. Ranta-aho, and P. Karjalainen, "Estimation of nonstationary eeg with kalman smoother approach: An application to event-related synchronization," *IEEE Transactions On Biomedical Engineering*, vol. 51, no. 3, pp. 516–524, 2004.

[3] M. Cassidy and W. Penny, "Bayesian nonstationary autoregressive models for biomedical signal analysis," *Biomedical Engineering, IEEE Transactions on*, vol. 49, no. 10, pp. 1142 – 1152, Oct. 2002.

[4] E. Bernat, W. Williams, and W. Gehring, "Decomposing erp time–frequency energy using pca," *Clinical Neurophysiology*, vol. 116, p. 1314–1334, 2005.

[5] J. Yang, D. Zhang, A. F. Frangi, and J. yu Yang, "Two-dimensional pca: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[6] D. Zhang and Z.-H. Zhou, "(2d)2pca: Two-directional two-dimensional pca for efficient face representation and recognition," *Neurocomputing*, vol. 69, no. 1-3, pp. 224 – 231, 2005, neural Networks in Signal Processing.

[7] J. Friedman, "Regularized discriminant analysis," *Journal of the American Statistical Association*, no. 84, pp. 165–175, 1989.

[8] E. Delgado-Trejos, A. Quiceno-Manrique, J. G.-L. J. M. B.-V. M., and G. Castellanos-Dominguez, "Digital auscultation analysis for heart murmur detection," *Annals of Biomedical Engineering*, vol. 37, no. 2, pp. 337–53, 2009.

[9] L. Avendano-Valencia, J. Ferrero, and G. Castellanos-Dominguez, "Improved parametric estimation of time frequency representations for cardiac murmur discrimination," in *CINC 08, Computers in Cardiology*, Bolonia, Italy, 2008, pp. 157 – 160.