VISUALIZATION OF NORMAL AND PATHOLOGICAL SPEECH DATA

J. Goddard[1], F. Martínez[1], G. Schlotthauer[2], M.E. Torres[2], H.L. Rufiner[2,3]
[1]Departamento de Ingeniería Eléctrica, Universidad Autonoma Metropolitana, Iztapalapa, Mexico
[2]Facultad de Ingeniería, Universidad Nacional de Entre Ríos, Paraná, Argentina
[3]Facultad de Ingeniería y Ciencias Hídricas, Universidad Nacional de Litoral, Santa Fe, Argentina

Abstract: **Techniques for the visualization of high-dimensional data are common in exploratory data analysis and can be very useful for gaining an intuition into the structure of a data set. The classical method of principal component analysis is the one most often employed, however in recent years a number of other nonlinear techniques have been introduced. In the present paper, principal component analysis, and two newer methods, are applied to a set of speech data and their results are compared.**
Keywords : **PCA, LLE, Kernel PCA**

I. INTRODUCTION

Techniques which transform a high-dimensional space into a space of fewer dimensions, often with one, two or three-dimensions, are collectively known as dimensionality reduction techniques. They can be very useful in helping us visualize data sets which we are trying to analyze, often providing clues about properties of the data, such as possible clusters within the data.

The most commonly used classical method for dimensionality reduction is perhaps principal component analysis (PCA), also known as the Karhunen-Loève transform, or singular value decomposition [1]. PCA performs a linear mapping of the data to a lower dimensional space in such a way, that the variance of the data in the low-dimensional representation is maximized. A disadvantage of PCA is that the embedded subspace has to be linear. For example, if the data are located on a circle in a 3-dimensional Euclidean space, R^3, PCA will not be able to identify this structure. Another disadvantage is that PCA depends critically on the units in which the features are measured.

In recent years, a number of other visualization techniques have become available, and their application to data sets, such as those involving speech, is just being conducted [2,3,4]. Among these methods, those of kernel PCA (KPCA) [5] and local linear embedding (LLE) [6,7] are particularly relevant for our purposes in the present paper.

KPCA is a (usually) nonlinear extension of PCA using kernel methods. Kernel methods have been successfully applied in the fields of pattern analysis and pattern recognition [8], often providing better classification performance than other methods, and frequently playing a vital part in the nonlinear extension of classical algorithms.. LLE, on the other hand, provides low-dimensional, neighborhood-preserving embeddings. This means that points which are 'close' to one another in a data space will also be close when projected onto the low-dimensional space.

In the paper, our aim is to briefly describe these methods, and then apply and compare them on a set of normal and pathological speech data.

II. METHODOS

PCA is an unsupervised learning algorithm that attempts to efficiently represent the data by finding orthonormal axes which maximally decorrelate the data. The data is then projected onto these orthogonal axes. The principal components are precisely this set of q orthonormal vectors, where q is often 2 or 3.

There are several equivalent ways to find the principal components, one being that of finding the first q eigenvectors w of the covariance matrix C of the data set, corresponding to the q largest eigenvalues. Mathematically, if $\{x_1,\ldots,x_N\}$ is a zero mean data set from the Euclidean space R^n, then the covariance matrix is given by:

$$C = \frac{1}{N}\sum_{j=1}^{N} x_j x_j^T \qquad (1)$$

and the corresponding eigenvalue equation is

$$Cw = \lambda w \qquad (2)$$

PCA provides a linear mapping of the data onto the lower q-dimensional space, and suffers from several problems, some of which have been mentioned in the introduction. In order to define a nonlinear extension of PCA, KPCA has been introduced. KPCA uses the notion of a kernel to modify the corresponding algorithm. Generally, if X is a data set, then a (positive-definite) kernel k on XxX is defined as a real-valued function:

$$k: \mathrm{XxX} \rightarrow \mathrm{R} \qquad (3)$$

such that:
(i) k is symmetric: $k(x,y) = k(y,x) \quad \forall\, x,y \in \mathrm{X}$, and
(ii) k is positive definite: $\forall\ n \geq 1$

$$\sum_{i,j=1}^{N} a_i a_j k(x_i, x_j) \geq 0 \quad (4)$$

$\forall\ a_1,\ldots,a_N \in \mathrm{R}$ and $x_1,\ldots,x_N \in \mathrm{X}$

It can be shown that given a kernel k, there exists a (Reproducing Kernel) Hilbert space H and a transformation ϕ: $\mathrm{X} \rightarrow \mathrm{H}$ such that

$$k(x,y) = <\phi(x), \phi(y)> \quad (5)$$

holds. H is often referred to as feature space and is often infinite-dimensional.

The most commonly used kernels are the polynomial and radial base function kernels defined on $\mathrm{R}^m \mathrm{x}\ \mathrm{R}^m$ by:

$$k(x,y) = (<x,y> + 1)^d \quad (5)$$

$$k(x,y) = \exp(-||x-y||^2/2\sigma^2) \quad (6)$$

respectively, where $d = 1,2,\ldots$ and $\sigma \in \mathrm{R}$. For these kernels the transformation ϕ is not defined explicitly, and the kernels are applied directly in the original data space. This is known as the 'kernel trick'.

For the kernels of Eq (5) and (6), it can be shown that KPCA is conceptually the same as performing standard PCA with the data set $\{\phi(x_1),\ldots,\ \phi(x_N)\}$ in the feature space H (with the above notation). Fortunately, the kernel trick, referred to above, can also be applied in this case and the explicit use of ϕ avoided. Instead, the NxN kernel matrix K, is defined through $K_{ij} = k(x_i,x_j)$, and the equation:

$$Ka = N\lambda a \quad (7)$$

is solved for $\lambda \in \mathrm{R}$ and $a = (a_1,\ldots,a_N)^{\mathrm{T}} \in \mathrm{R}^{\mathrm{N}}$.

A projection p of a pattern y in data space onto a principal component in feature space can be found using:

$$p = \sum_{i=1}^{N} a_i k(y, x_i) \quad (8)$$

In order to use KPCA, we have to decide on a kernel function and, as for PCA, the number of dimensions on which to project.

LLE is an unsupervised learning algorithm that computes low-dimensional, neighborhood-preserving embeddings of high-dimensional inputs. LLE does this by applying three steps. First, for each point in the data, it's k nearest neighbors to the other points in the data are found (usually using Euclidean distance, although in the present paper other distance metrics are also tried). Then, each point is approximated by convex combinations of it's k nearest neighbors, to obtain a matrix of reconstruction weights W. Finally, low-dimensional embeddings Y_i (usually in a space of one or two-dimensions) are found such that the local convex representations are preserved. Mathematically, this process can be expressed by: If $\{x_1,\ldots,x_N\}$ is the dataset, and for each vector x_i we let N_i denote the indices of it's k nearest neighbors, then the second step, of finding the reconstruction weights W, corresponds to minimizing the objective function:

$$E(W) = \sum_i \left| x_i - \sum_{j \in N_i} W_{ij} x_j \right|^2 \quad (9)$$

subject to $\sum_j W_{ij} = 1$.

The embeddings $\{y_1,\ldots,y_N\}$ of the original data, corresponding to the third step, are obtained by minimizing the following objective function:

$$O(Y) = \sum_i \left| y_i - \sum_{j \in N_i} W_{ij} y_j \right|^2 \quad (10)$$

An advantage of LLE is that it has few free parameters to set and a non-iterative solution thus avoiding convergence to a local minimum.

Interesting relationships have recently been found between KPCA and LLE, as well as other well-known dimensionality reduction techniques c.f. [10].

III. DATA

In the present paper, the data used consisted of real voice samples of the sustained vowel 'ah' for both normal patients and those with dysphonic speech disorders. The voice samples were taken from the "Disordered Voice Database" [11], acquired at the Massachusetts Eye and Ear Infirmary Voice and Speech Laboratory and distributed by Kay Elemetrics. The clinical information includes diagnostic information along with patient identification, age, sex, smoking status, and more. The files on normal subjects were collected at Kay.

The eight variables used in the paper are the same as those chosen in [12], namely: degree of voice breaks, three variables related to jitter (local, relative average perturbation, five-point period perturbation quotient), three related to shimmer (local, three-point amplitude perturbation, eleven-point amplitude perturbation), and harmonics-to-noise ratio.

For completeness, we include their definitions (c.f. [12] for more details):

1) Degree of voice breaks is the total duration of the breaks between the voiced parts of the signal, divided by

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, G. Schlotthauer, H. M. Torres & H. L. Rufiner; "Visualization of Normal and Pathological Speech Data"
Proc. of the 5th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2007), dec, 2007.

the total duration of the analyzed part of signal. Silences at the beginning and at the end of the signal are not considered breaks.

2) Jitter or period perturbation quotient

a) Jitter ratio (local) or jitt is defined as:

$$jitt = 1000 \frac{\frac{1}{n-1}\sum_{i=1}^{n-1} P_i - P_{i+1}}{\frac{1}{n}\sum_{i=1}^{n} P_i} \quad (11)$$

where P_i is the period of the i[th] cycle, in ms, and n is the number of periods in the sample.

b) Relative average perturbation (RAP):

$$RAP = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\left|\frac{P_{i-1}+P_i+P_{i+1}}{3} - P_i\right|}{\frac{1}{n}\sum_{i=1}^{n} P_i} \quad (12)$$

c) Five-point period perturbation quotient (ppq5):

$$ppq5 = \frac{\frac{1}{n-4}\sum_{i=3}^{n-2}\left|\frac{\sum_{j=-2}^{2} P_{i+j}}{3} - P_i\right|}{\frac{1}{n}\sum_{i=1}^{n} P_i} \quad (13)$$

3) Shimmer or amplitude perturbation quotient

a) Shimmer (shimm):

$$shimm = \frac{\frac{1}{n-1}\sum_{i=1}^{n-1}|A_i - A_{i+1}|}{\frac{1}{n}\sum_{i=1}^{n} A_i} \quad (14)$$

where A_i is the amplitude of the i[th] cycle, and n is the number of periods in the sample.

b) Three-point amplitude perturbation quotient (apq3):

$$apq3 = \frac{\frac{1}{n-2}\sum_{i=2}^{n-1}\left|\frac{A_{i-1}+A_i+A_{i+1}}{3} - A_i\right|}{\frac{1}{n}\sum_{i=1}^{n} A_i} \quad (15)$$

c) Eleven-point ampitude perturbation quotient (apq11).

$$apq11 = \frac{\frac{1}{n-10}\sum_{i=6}^{n-5}\left|\frac{\sum_{j=-5}^{5} A_{i+j}}{11} - A_i\right|}{\frac{1}{n}\sum_{i=1}^{n} A_i} \quad (16)$$

4) Harmonics-to-noise ratio: This parameter quantifies the amount of glottal noise in the vowel waveform. In contrast to perturbation measures, it attempts to resolve the vowel waveform into signal and noise components, computing their energies ratio.

In total there were 34 subjects with dysphonic speech disorders, and a further 53 normal subjects. For each subject, an 8-variable vector was associated. The minimum, maximum and standard deviation for each of the eight variables is given in Table 1.

Table 1. minimum, maximum and standard deviation for each of the 8 variables for the normal and pathological data

Normal							
0.105	0.048	0.070	0.064	0.375	0.567	0	17.52
0.682	0.368	0.447	0.463	3.011	3.770	0	30.37
0.11	0.069	0.067	0.088	0.589	0.744	0	2.941
Pathological							
0.131	0.064	0.074	0.119	0.654	0.937	0	2.515
6.061	3.701	4.783	1.756	10.80	16.63	0.164	28.04
1.4233	0.8221	1.0783	0.431	2.524	3.304	0.035	6.83

IV. RESULTS

PCA, KPCA, and LLE were applied to the real voice samples described in the previous section. Software for

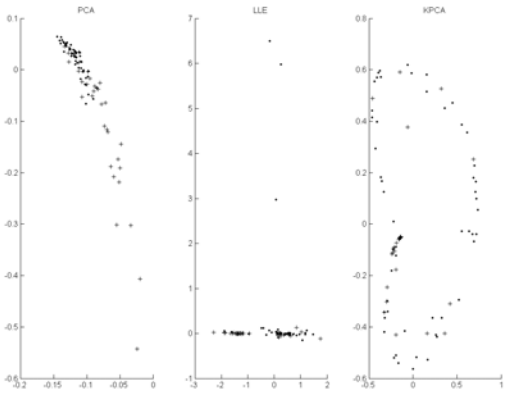

Fig. 1 PCA, LLE, and KPCA applied to the data with 2 dimensions.

these techniques has been developed by [13,14]. Fig.1 shows the three techniques applied to the data and projected onto two-dimensions. In this case, k=8 was chosen for LLE, and a radial base function kernels with σ=1 for KPCA.

In Fig.2, the same parameters are used but with the data projected onto three-dimensions.

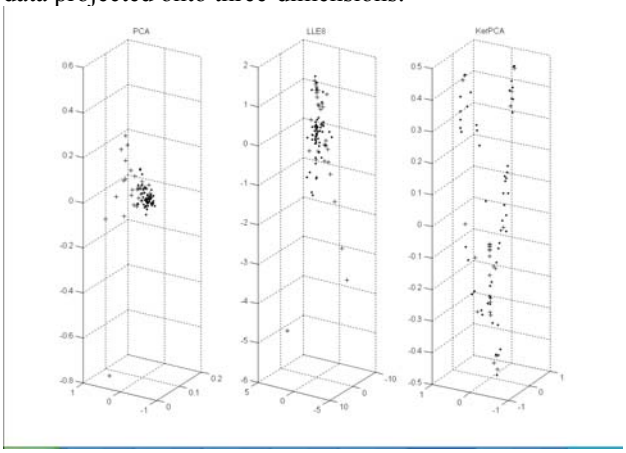

Fig 2. PCA, LLE, and KPCA applied to the data with 3 dimensions.

In order to obtain a simple comparison between the three methods, a k-nearest neighbor classifier was applied to the projected data using k=1,3. For this, the data was split randomly into training and test sets subsets with sizes of 66% and 33%, respectively. The classification results are shown in Table 2, where in the first row, LLEn means that k=n was taken, and Gn means that σ=n was used.

Table 2. Results of applying knn to the projected data

	PCA	LLE3	LLE5	LLE8	G0.5	G1	G5
Two-dimensions							
k=1	68.97	55.17	68.97	**79.31**	68.97	65.52	75.86
k=3	**79.31**	68.97	**79.31**	72.41	65.52	75.86	68.97
Three-dimensions							
k=1	**79.31**	55.17	68.97	72.41	65.52	75.86	75.86
k=3	65.52	65.52	**79.31**	72.41	65.52	75.86	72.41

V. Conclusions

In the present paper, the dimensionality reduction techniques of PCA, KPCA, and LLE were applied to speech data from both normal and pathological subjects. The data has been projected onto both two and three-dimensional Euclidean spaces, and different parameters occurring in KPCA and LLE have been varied. The projected data is shown in Figs.1 and 2.

In order to obtain a simple comparison between the three methods, a k-nearest neighbor classifier was introduced and applied to the projected data. In Table 2 it can be seen that LLE, along with PCA, achieve the best classification performances. Whilst this is obviously not a definitive result, and will depend on the data set and parameters employed, it is encouraging and provides motivation to continue the exploration of alternative methods to PCA in the case of speech data.

References

[1] I.T. Jolliffe, *Principal Component Analysis*, Springer, 1986.

[2] M.A. Carreira-Perpinan, "Continuous latent variable models for dimensionality reduction and sequential data reconstruction," PhD thesis, Dept. of Computer Science, University of Sheffield, UK, 2001.

[3] V. Jain and L. K. Saul, "Exploratory analysis and visualization of speech and music by locally linear embedding," In Proceedings of the International Conference of Speech, Acoustics, and Signal Processing (ICASSP-04), vol.3, pp.984-987, Canada, 2004.

[4] A. Kocsor and L. Tóth, "Kernel-Based Feature Extraction with a Speech Technology Application," IEEE Transaction on Signal Processing, Vol. 52, No. 8, pp.2250-2263.

[5] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel Principal Component Analysis," In B. Schölkopf, C. J. C. Burges, and A.J. Smola, editors, *Advances in Kernel Methods---Support Vector Learning*, pp. 327–352, MIT Press, Cambridge, MA, 1999.

[6] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, v.290, no.5500, pp.2323-2326, 2000.

[7] Locally linear embedding homepage: http://www.cs.toronto.edu/~roweis/lle/.

[8] J. Shawe-Taylor and N. Cristianini, *Kernel Methods for Pattern Analysis*, Cambridge University Press, 2004.

[10] J. Ham, D.D. Lee, S. Mika, and B. Schölkopf, "A kernel view of the dimensionality reduction of manifolds," Proceedings of the Twenty-First International Conference on Machine Learning, pp. 369-376, (Eds.) R. Greiner and D. Schuurmans, 2004.

[11] Kay Elemetrics Corporation. Disordered Voice Database Model 4337. Massachusetts Eye and Ear Infirmary Voice and Speech Lab, Boston, MA., 1994.

[12] G. Schlotthauer, M. E. Torres, y C. Jackson-Menaldi, "Automatic classification of dysphonic voices," WSEAS Transactions on Signal Processing, vol. 2, no. 9, pp. 1260-1267, September 2006.

[13] L.J.P. van der Maaten, "An Introduction to Dimensionality Reduction Using Matlab," Technical Report MICC 07-07. Maastricht University, Maastricht, The Netherlands, 2007.

[14] T. Wittman, "Manifold Learning Matlab Demo," Department of Mathematics, University of Minnesota, 2005.

sinc(*i*) Research Center for Signals, Systems and Computational Intelligence (fich.unl.edu.ar/sinc)
J. Goddard, F. M. Martínez, G. Schlotthauer, H. M. Torres & H. L. Rufiner; "Visualization of Normal and Pathological Speech Data"
Proc. of the 5th International Workshop on Models and Analysis of Vocal Emissions for Biomedical Applications (MAVEBA 2007), dec, 2007.