DIMENSIONALITY REDUCTION FOR FEAR EMOTION DETECTION FROM SPEECH

*Safa Chebbi,**

University of Carthage
Higher School of Communication of Tunis
Research Lab. COSIM
safa.chebbi@supcom.tn

Sofia Ben Jebara

University of Carthage
Higher School of Communication of Tunis
Research Lab. COSIM
sofia.benjebara@supcom.tn

ABSTRACT

In this paper, we propose to reduce the relatively high-dimension of pitch-based features for fear emotion recognition from speech. To do so, the K-nearest neighbors algorithm has been used to classify three emotion classes: fear, neutral and 'other emotions'. Many techniques of dimensionality reduction are explored. First of all, optimal features ensuring better emotion classification are determined. Next, several families of dimensionality reduction, namely PCA, LDA and LPP, are tested in order to reveal the suitable dimension range guaranteeing the highest overall and fear recognition rates. Results show that the optimal features group permits 93.34% and 78.7% as overall and fear accuracy rates respectively. Using dimensionality reduction, Principal Component Analysis (PCA) has given the best results: 92% as overall accuracy rate and 93.3% as fear recognition percentage.

1. INTRODUCTION

Emotion is one of the main drivers of human thoughts and actions. It manifests itself through several modalities: speech, body gesture, facial expression, eyes contact,... As speech is a simple and natural way of communication, emotion recognition from speech is widely used (see for example [1][2]). In this paper, we deal with fear emotion recognition through the classification of speech into neutral, fear and other emotions. We are mainly interested in fear emotion because it has many applications. In our considered research, we aim to detect suspicious behavior which risks to be a terrorism attack, as part of civil safety. Therefore, we are particularly interested in detecting fear state which may characterize such person (before the action) in order to protect victims and limit damage [3].

In order to design a reliable emotion recognition system, the following questions should be answered: *i)* How to select appropriate features to extract from speech? *ii)*Which classification techniques to use? *iii)*How to select the most relevant and discriminatory features? and *iv)* How to reduce a high dimension feature set into a meaningful representation of reduced dimensionality? With regards to the first point, the speech production system consists of two principal organs: vocal folds, which are responsible for the production of sounds used for speech, and vocal tract related to the movement of the tongue tip, the jaw and the lip during the voice production. In our study, we are interested in studying vocal-folds related features and more precisely the pitch. Indeed, pitch expresses the vibration frequency of vocal folds during the production of voiced sounds. This choice is justified by the fact that, on the one hand, vocal folds vibrate, in a similar way, for all the phonemes unlike vocal tract, whose behaviour varies from one phoneme to another. On the other hand, the voice presents many modifications during the fear state such as oscillation, tremor, irregularity and stammering [4]. These changes are due to the vibration of vocal folds.

For the second point in the context of classification techniques, many classifiers are developped in the litterature based on machine learning approach. We quote for example Neural Network, K-nearest Neighbors, Random Forest, Decision Tree, Gaussian Mixture Model, among others [5]. In a previous work dealing with fear emotion detection [6], we performed the classification using four classifiers : Support Vector Machine (SVM) [7], Decision Tree (DT) [8], Subspace Discriminant [9] and K-nearest Neighbors (KNN) [10]. The highest fear emotion detection has been obtained using KNN. Therefore, KNN has been the classification tool of our study in this paper.

According to the third point related to discriminatory features selection, a large pool of techniques has been proposed for such purpose. We relate for example Fisher discriminant ratio, scatter matrices, statistical tests, the Receiver Operating Characteristic (ROC) curve, Bhattacharyya distance, RELIEF-F algorithm (see for example [11][12][13]). This has been the interest of our previous work for fear emotion detection [14]. In this work, four different relevance indexes have been used to select most relevant ones from a list of 27 features: Fisher Discriminant Ratio, probability divergence, scatter measure and ANOVA statistical test. Features with highest classification accuracy appearing in all relevance indexes are retained. Thanks to this approach, the fear emotion recognition results reached 86.7%.

Finally, the feature dimensionality reduction would be the objective of this paper. Indeed, when dealing with a high dimension data, classification problems become significantly harder and may lead to lower classification accuracy and poor quality of clusters. In the literature, this phenomenon is referred to as the curse of dimensionality [15]. This aspect has been a fertile field of research and development for over a century. In this context, many techniques have been proposed for this task. They are organized into two groups: linear methods such as principal component analysis [16], linear discriminant analysis [17], locality preserving projection [18], factor analysis [19], classical scaling [20] and non-linear ones including Kernel PCA [21], kernel discriminant analysis [22], Isomap [23] and multilayer autoencoders [24] among others.

The aim of this paper is the investigation of the effect of feature dimensionality reduction on classification performance. To this end, two approches have been adopted. The first one consists on performing the classification for all possible combinations of

* This work has been carried out as part of a federated research project entitled: Sensitive Supervision of Sensitive Multi-sensor Sites, funded by the Ministry of Higher Education and Scientific Research, Tunisia.

the whole pitch-based set of features using K-nearest neighbors algorithm. The approach is called 'N to N'. The goal here is to obtain the best combination of features and the suitable dimension range giving the best accuracy rate. The second one is to test many techniques for feature space reduction. They are *i)* Principal Component Analysis (PCA) and its variants (Kernel PCA denoted KPCA and probabilistic PCA denoted PPCA); *ii)* Linear Discriminant Analysis (LDA) and its kernelized version (Kernel DA denoted KDA); *iii)* Locality Preserving Projection (LPP) and its kernelized version (Projection denoted KLPP) and *iv)* many others which will be listed latter. The classification is carried out separately in the reduced space for each technique and the effect of dimension variation is analyzed. Finally, the best tradeoff between dimension reduction and classification performance is revealed.

The paper is organized as follows. Section 2 will give a brief description of the extracted pitch-based feature set, the considered emotional corpus and the emotion grouping adopted in this study. Section 3 will present a description of the 'N to N' approach as well as the classification results obtained using this process. Section 4 and 5 will provide an investigation about the use of correlation-based (resp. non-correlation based) techniques for dimensionality reduction and will display their classification results.

2. PRELIMINARIES

2.1. Features Set

The pitch is related to the vocal folds vibration, determining the periodicity of voiced sounds. More precisely, it translates the opening-closing frequency of vocal folds during the production of voiced sounds. Note that pitch is calculated only for voiced frames as vocal folds do not vibrate during the production of unvoiced sounds. In order to extract the set of pitch-based features, speech utterances are first decomposed into frames whose duration is 10ms. Next, voiced and unvoiced frames are identified and pitch values are calculated using the rapt algorithm [25]. Based on these pitch values, a whole set of global features are calculated. They are classified into four groups: usual measures, features related to pitch's derivative and second derivative as they are linked to the vibration speed and acceleration of vocal folds, features related to speech voicing since voicing rate varies from one emotion to another. The whole set of features has a 27 dimensionality. It is summarized in Table 1:

Table 1: *Feature set.*

FAMILY	DESCRIPTION	ABREVIATION
Usual measures	Mean, Maximum, Minimum,	-
	Variance, Median,	-
	Normalised standard deviation	Norm_STD
Speech voicing	Ratio of voiced frames on the total frames	Rat_Voic_tot
	Ratio of unvoiced frames on the total frames	Rat_UnVoic_tot
	Ratio of voiced frames on unvoiced frames	Rat_Voic_UnVoic
	First voiced frame	1st frm
	Second voiced frame	2nd frm
	Voiced frame in the middle frame	Middle frm
	Before last voiced frame	Bef_lst_frm
	Last voiced frame	Lst_frm
Pitch contour derivative	Mean of pitch's derivative	Mean_DRV
	Mean of the absolute value of pitch's derivative	Mean_ABS_DRV
	Variance of pitch's derivative	Var_DRV
	Variance of the absolute value of pitch's derivative	Var_ABS_DRV
	Maximum of pitch's derivative	Max_DRV
	Maximum of the absolute value of pitch's derivative	Max_ABS_DRV
	Mean of pitch's second derivative	Mean_Sec_DRV
	Maximum of pitch's second derivative	Max_Sec_DRV
Others	Ratio of pitch's mean on its maximum	flatness
	Ratio of pitch's mean on its minimum	Vehemence
	Ratio of peaks's number on total frames	Num_Peaks
	Minimum position	Min_Pos
	Maximum position	Max_Pos

2.2. Emotional Database and Selected Emotion Classes

EMO database, which is a German emotional database publicly accessible, has been used during this study [26]. It includes 800 utterances simulated by 10 professional actors (5 males and females). It consists of seven emotion states namely: neutral, fear, anger, joy, sadness, disgust and boredom. Recordings were taken in an anechoic chamber, under supervised conditions with a sampling frequency of 48 kHz and later downsampled to 16 kHz. A human perception test to recognize various emotions with 20 participants resulted in a mean accuracy of 84.3%.
The adopted emotion grouping considers 3 groups: fear, neutral and other emotions. The 'Other emotions' class includes the five remaining states (joy, anger, disgust, sadness and boredom). The classes repartition through the corpus is the following: 14% for fear, 14% for the neutral class and 72% for other emotions.

2.3. Adopted Criteria for Evaluating the classification quality

In this study, we performed the classification using K-nearest neighbors (KNN) algorithm. KNN has been chosen according to our previous study dealing with a comparison between many classifiers [6]. This study has revealed that KNN is the best tradeoff between classification performance and computational cost. The database was trained and tested using the holdout validation method where 70% of the data were used for training while 30% were used for testing. The classification was judged by two criteria in order to have a clear idea of classification model performance:

✓**The overall accuracy rate** : it translates the percentage of well predicted emotion sequences among the total number of emotion speech sequences. It is calculated by dividing the number of well predicted samples on the total number of samples.

✓**The fear accuracy rate** : This rate indicates the proportion of fear recognition among others. It is calculated by dividing the number of well predicted fear samples on the total number of samples in fear class.

3. DIMENSIONALITY REDUCTION BASED ON 'N TO N' COMBINATION

3.1. Approach

The aim of this section is to extract the optimal feature list ensuring maximal overall emotion detection rate from the whole set of selected features. To this end, the adopted approach was to test all the possible combinations of the 27 features already extracted and to identify, as a result, the group giving the best classification accuracy. In the first iteration, we looked for the best accuracy reached by one feature. Then, we looked for the best combination of two features (2 by 2 among the 27 possible ones) giving the best accuracy. The process is re-iterated for all possible values of N (N = 1,..,27) until reaching the whole set of 27 features. This process for each value of N is called 'N to N' combination of features.

The 'N to N' dimensionality reduction technique requires laborious and complex calculation that has lasted many weeks. Indeed, for each iteration of 'N to N' combinations among the 27 features, the classification algorithm is applied C_{27}^{N} times (where C is the combinatory operator). In order to make aware of the heaviness of computational cost, the number of combinations varies between 27 and 20 millions. The second line in Table 2 displays the total number of combinations for each 'N to N' combination of features. The run time of each 'N to N' combination is given in line 3 of Table 2 using a machine with an intel core CPU i3, 64 bits and having 1.70 GHz as a clock speed and 4 Go of RAM.

3.2. Classification Results

Figure 1 represents the evolution of the classification results for each feature vector size in terms of overall and fear accuracy rates. The solid line indicates the variation of overall accuracy rate while dashed line is reserved for fear accuracy. The first value indicates the best overall accuracy rate obtained using only one feature. The second provides the best accuracy rate obtained for the combination of 2 features among the 27 ones, and so on. Note that the best feature group has been extracted according to the overall accuracy rate optimization and not fear accuracy rate.

One can notice that the range of overall accuracy varies between

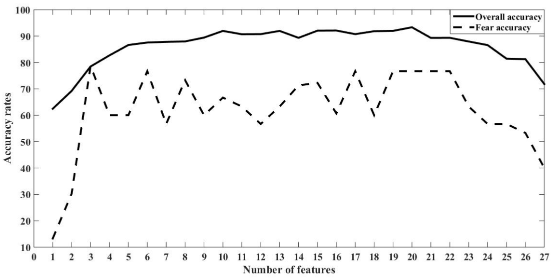

Figure 1: *Classification results according to 'N to N' approach.*

62% and 93,34%. The best value is obtained using 20 features for which the accuracy rate is equal to 93.34%. Also, we can note a stabilization at classification quality for a dimensionality between 10 and 22. On the other hand, fear recognition rate varies between 13.3% and 78.7%. The best one is obtained with a 3-features combination. Note that the quality varies enormously in the ascending and descending order for a dimension range between 3 and 19 features. Moreover, choosing a dimensionality range between 19 and 22 would be the best tradeoff between overall accuracy and fear accuracy rates. Indeed, the classification quality is among the best ones in that interval according to the two criteria.

3.3. Optimal features with reduced dimensionality

Table 3 indicates the list of relevant features giving the best accuracy rate obtained for each 'N to N' combination. The 20 features giving the best overall performance (93.34%) are: mean, median ,variance, normalised standard deviation, flatness, number of peaks, minimum, maximum, the ratio of voiced on unvoiced frames, the ratio of unvoiced frames on the total frames, mean of the absolute value of pitch's derivative, maximum position, variance of derivative, variance of the absolute value of derivative, maximum of the absolute value of derivative, mean of the second derivative, first, second, before last and last voiced frames.

However, one can notice that median, mean of second derivative, mean, last voiced frame and number of peaks are classed on the top-5 according to their presence as optimal features for the other combinations (ticked in bold in Table 3). This fact confirms their usefulness and relevance in discriminating between fear, neutral and other emotion states. If we deal with dimensionality reduction, the reduced vector size of dimensionalities varying from 19 to 22 is considered. Thus, eleven features are revealed as relevant common ones between these ranges. They are the mean, maximum, variance, Rat_Voic_UnVoic, Lst_frm, Mean_ABS_DRV, Var_ABS_DRV, Max_ABS_DRV, Mean_Sec_DRV, flatness and Num_Peaks. Thus, these features seem to be the most relevant ones.

4. CORRELATION-BASED DIMENSIONALITY REDUCTION

Whereas 'N to N' combination approach leads to very significant classification results reaching 93.3%, it remains difficult to apply them in practice because of their complexity and computational cost. Hence, the use of automatic dimension reduction techniques guaranteeing speed and performance are preferred. This section is devoted to investigate dimension reduction methods considering the correlation between features.

4.1. Correlation between features

Referring to the curse of dimensionality, dealing with a redundant and correlated features may lead to poor classification performance. In order to take an idea about the linear dependency between features, the correlation between them has been calculated pairwise and the results are displayed in Table 4. The features' names have been replaced by their corresponding number (1,2, ... 27) due to lack of space. The retained order is the same as the one adopted in Table 3. It means that 1 indicates mean, 2 indicates median and so on.

From Table 4, one can deduce that some pairs of features present strong correlation ($|\rho| > 0.7$). We relate for example the correlation between variance of derivative and mean of absolute value of derivative (0.92). Others are moderately correlated ($0.3<|\rho| <0.7$) such as variance and mean of absolute value of derivative. A good part of the features are weakly correlated. It means that they are quasi independent or totally independent ($|\rho|<0.2$). Thus, we decided to use a dimension reduction technique garanteeing features decorrelation and eliminating dependencies between them in order to obtain better classification results. The most used technique in the literature is Principal Component Analysis (PCA).

4.2. Traditional PCA and variants

PCA stills the most used technique for dimensionality reduction. It consists on using an orthogonal transformation to convert a set of possibly correlated features into uncorrelated ones called principal components. The new components of the embedded basis meet the following criteria: *(i)* they are linear combinations of the original features, *(ii)* they form an orthogonal basis that can be viewed as a rotation of the original one, and *(iii)* components are uncorrelated but preserve the maximum amount of variation in the data. In addition to traditional PCA [16], probabilistic PCA (PPCA) [27] and

Table 2: *Calculation complexity of 'N to N' combination approach.*

Features number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
Combinations number	27	351	2925	17550	80730	296010	888030	2220075	4686825	8436285	13037895	17383860	20058300	20058300
Overage calculation time	5.3s	223s	17.4 min	47min	23hours	18hours	3days	7days	15days	28 days	43days	43days	57days	66days
Features number	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	
Combinations number	17383860	13037895	8436285	4686825	2220075	888030	296010	80730	17550	2925	351	27	1	
Overage calculation time	43days	43days	28 days	15days	7days	3days	19hours	1day	56min	18min	250s	7s	12s	

Table 3: *Best feature combinations.*

Feature number / Feature name	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**
Mean				✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓		✓
Median			✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓			✓	✓	✓	✓		✓	✓	✓	✓	✓
Maximum		✓														✓			✓	✓	✓	✓	✓		✓	✓	✓
Minimum												✓					✓			✓		✓	✓	✓	✓	✓	✓
Variance											✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓	✓
Norm_STD																✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Rat_Voic_tot					✓		✓		✓			✓	✓	✓	✓		✓	✓	✓		✓	✓	✓	✓	✓	✓	✓
Rat_UnVoic_tot								✓				✓				✓		✓	✓	✓		✓	✓	✓	✓	✓	✓
Rat_Voic_UnVoic										✓	✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
1st frm						✓	✓							✓			✓	✓	✓	✓	✓			✓	✓	✓	✓
2nd frm													✓			✓		✓	✓	✓	✓		✓	✓		✓	✓
middle frm			✓			✓							✓	✓		✓	✓	✓			✓	✓	✓	✓	✓	✓	✓
Bef_lst_frm									✓		✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓		✓	✓	✓
Lst_frm					✓	✓		✓		✓		✓	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean_DRV										✓					✓			✓			✓	✓			✓	✓	✓
Mean_ABS_DRV								✓			✓			✓	✓				✓	✓	✓	✓	✓	✓	✓	✓	✓
Var_DRV									✓	✓	✓		✓		✓					✓	✓	✓	✓	✓	✓	✓	✓
Var_ABS_DRV									✓	✓			✓				✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Max_DRV				✓													✓	✓				✓	✓	✓	✓	✓	✓
Max_ABS_DRV								✓	✓			✓			✓	✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓
Mean_Sec_DRV			✓		✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓		✓	✓	✓	✓
Max_Sec_DRV																✓		✓			✓		✓	✓	✓	✓	✓
flatness	✓									✓				✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
vehemence		✓				✓	✓			✓	✓			✓	✓		✓	✓	✓		✓	✓		✓	✓	✓	✓
Num_Peaks							✓			✓	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Min_Pos				✓				✓	✓			✓			✓	✓	✓	✓				✓	✓	✓	✓	✓	✓
Max_Pos											✓					✓	✓			✓		✓	✓	✓	✓	✓	✓

kernel PCA (KPCA) [21] have been used for dimension reduction. KPCA is a non-linear reformulation of standard PCA. Indeed it uses a kernel trick to find principal components in a different space. In other words, it performs standard PCA in a new non-linear space. It is applicable for features presenting non-linear correlation between each other [21].

The PPCA is another formulation of standard PCA based upon a probability model [27]. The principal components are determined through maximum-likelihood estimation of parameters from the data principal components.

4.3. Classification Results

First, the embedded subspace is extracted for each technique. Then, the classification is performed with a different number of components each time. That is to say that first, the classification is performed using only the first component. Then the 2 first components are used and so on until using the whole set of componants. Hence, the suitable dimension range is the one giving the best classification performance. As for 'N to N' approach, it is judged using the overall and fear accuracy rates.

Classification results are provided in Figure 2 (resp. Figure 3) for each used technique from the PCA family in terms of overall accuracy rate (resp. fear accuracy rate). The two figures lead to the following interpretations:

✓Using traditional PCA, the best overall accuracy and fear accuracy rates reach 92% and 93.3% respectively with 19 components.

✓Using KPCA, the best overall accuracy reaches 82.7% with 6 components and the best fear accuracy reaches 86.7% with 6 components.

✓Using PPCA, the best overall accuracy and fear accuracy rates are worst. They are equal to 65.3% with 4 components and 33.3% with only one component. This approach should be discarded.

When dealing with tradeoff between accuracy and dimensionality reduction, KPCA seems to be better than PCA. In fact, the dimensionality is reduced to 6 (versus 19) with a loss of 10% for overall accuracy and 7% for fear rate. Moreover, KPCA has the advantage of presenting a stable variation of quality when dimen-

Table 4: *Correlation between features.*

Correlation	**Feature pairs**
$0.9<\|\rho\|<1$	(7,8);(7,9);(16,18);(17,18);(19,20)
$0.8<\|\rho\|<=0.9$	(8,9);(16,17)
$0.7<\|\rho\|<=0.8$	(1,12);(2,24);(3,22)
$0.6<\|\rho\|<=0.7$	(2,23);(3,17);(3,18);(5,6);(5,12);(5,16);(10,11);(10,15);(17,22);(18,22)
$0.5<\|\rho\|<=0.6$	(3,6);(3,16);(5,17);(5,18);(5,21);(6,16);(6,17); (6,18);(6,22);(14,15);(18,22);(23,24)
$0.4<\|\rho\|<=0.5$	(1,2);(1,5);(1,16);(1,23);(2,25);(3,5);(4,6);(16,22);(22,23);(23,25)
$0.3<\|\rho\|<=0.4$	(1,4);(1,11);(1,21);(2,5);(2,12);(2,13);(2,21);(3,23);(4,12);(5,24);(11,12);(11,16);(12,16);(12,21);(12,23); (12,24);(13,14);(13,17);(13,18);(13,24);(24,25)
$0.2<\|\rho\|<=0.3$	(1,3);(1,6);(1,7);(1,8);(1,9);(1,10);(1,13);(1,14);(1,17);(1,18);(2,4);(2,7);(2,8);(2,9);(2,10); (2,11);(2,14);(3,12);(3,21);(4,9);(4,22);(4,23);(4,25);(5,7);(5,8);(5,9);(5,11);(5,22);(6,12);(6,21); (6,23);(6,24);(8,12);(8,23);(8,24);(8,25);(9,12);(9,23);(10,12);(11,15);(11,23);(12,14);(12,21);(13,16);(13,23); (14,24);(14,27);(16,21);(16,24); (17,21);(17,23);(17,24);(17,27); (18,23);(18,27);(21,24);(22,27)
$0.1<\|\rho\|<=0.2$	(1,25);(2,16);(3,10);(3,11);(3,13);(3,14);(3,24);(3,25);(3,27);(4,5);(4,7);(4,8);(4,10);(4,11);(4,17); (4,18);(4,24);(5,10);(5,13);(5,14);(5,20);(6,7);(6,8);(6,13);(6,20);(6,25);(7,12);(7,16);(7,17);(7,18); (7,21);(7,23);(7,24);(7,25);(8,16);(8,17); (8,18);(8,21);(9,16);(9,17);(9,21);(9,24);(9,25);(10,13);(10,16); (10,17);(10,23);(10,24);(10,25);(10,26);(11,13);(11,14);(11,21);(11,24);(11,25);(11,26);(11,27);(12,13);(12,17);(12,18); (12,22);(12,25);(13,15);(13,22);(13,27);(14,16);(14,17);(14,18);(14,23);(14,25);(15,27); (16,19);(16,20);(16,23);(16,25); (16,27);(17,19);(17,20);(17,25);(18,20);(18,21);(18,24);(18,25);(20,25);(21,22);(21,23);(21,25);(22,25);(24,27)
$\|\rho\|<=0.1$	(1,15);(1,19);(1,20);(1,22);(1,24);(1,26);(1,27); (2,3);(2,6);(2,15);(2,17);(2,18);(2,19);(2,20);(2,22); (2,26);(2,27); (3,4);(3,7);(3,8);(3,9);(3,15);(3,19);(3,20);(3,26);(4,13);(4,14);(4,15);(4,16);(4,19); (4,20);(4,21);(4,26);(4,27);(5,15);(5,19);(5,23);(5,25);(5,26);(5,27);(6,9);(6,10);(6,11);(6,14);(6,15); (6,19);(6,26);(6,27);(7,10);(7,11);(7,13);(7,14);(7,15);(7,16);(7,17);(7,18);(7,19);(7,20);(7,22);(7,26); (7,27); (8,10);(8,11);(8,13);(8,14);(8,15);(8,19);(8,20);(8,22);(8,26);(8,27); (9,10);(9,11);(9,13);(9,14);(9,15); (9,18);(9,19);(9,20);(9,22);(9,26);(9,27); (10,14);(10,18);(10,19);(10,20);(10,21);(10,22);(10,27); (11,17);(11,18); (11,20);(11,19);(11,22);(11,25);(11,26);(11,27); (12,15);(12,19);(12,20);(12,26);(12,27);(13,19);(13,20);(13,21);(13,25); (13,26);(14,19);(14,20);(14,21);(14,22);(14,23);(14,25);(14,26);(15,16);(15,17);(15,18);(15,19);(15,20);(15,21);(15,22); (15,23);(15,24);(15,25);(15,26);(16,26);(17,26);(18,19);(18,26); (19,21);(19,22);(19,23);(19,24);(19,25);(19,26);(19,27); (20,21);(20,22);(20,23);(20,24);(20,26);(20,27);(22,24);(22,26);(23,26);(23,27);(24,26);(25,26);(25,27);(26,27)

sionality changes. Indeed, it appears as a horizontal line of accuracy rate for a feature number varying between 7 and 19.

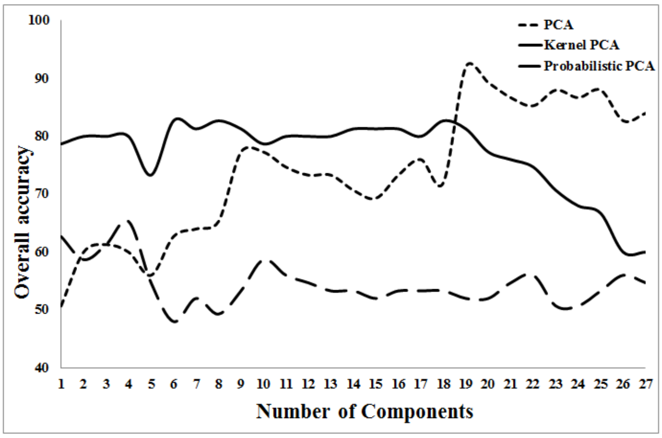

Figure 2: *Overall classification results according to PCA family techniques.*

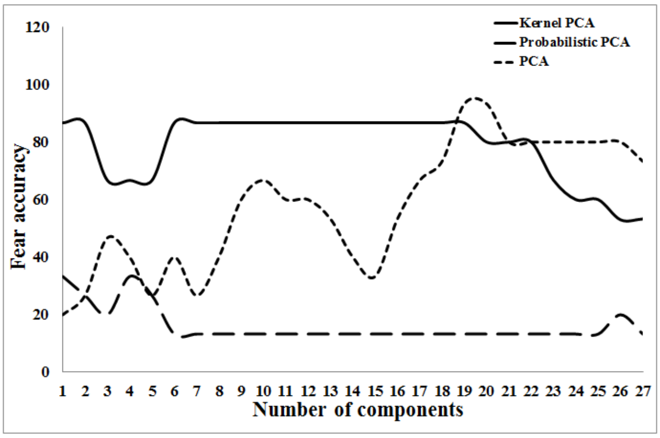

Figure 3: *Fear accuracy rates according to PCA family techniques.*

5. NON-CORRELATION BASED DIMENSIONALITY REDUCTION

5.1. Linear Discriminant Analysis Family

In contrast to most other dimensionality reduction methods, LDA is a supervised technique as it takes into consideration the class labels when constructing the embedded feature space [17]. It attempts to find a new feature space to project the data in order to maximize classes separability. It is based on the concept of maximizing the Fisher ratio. This latter is calculated by dividing the between-class variability on the within-class variability.

Standard LDA and kernel LDA have been tested in this study in order to reduce feature space. Standard LDA attempts to maximize the linear separability between classes. It reduces dimensionality from original number of feature to C-1 features, where C is the number of classes. In our study, the new feature space will be only a 2-dimensional space as we have 3 emotion classes.

KDA is a kernelized version of LDA using the kernel trick [22]. Standard LDA is performed in a new feature space which allows non-linear mapping. Contrary to LDA, it has the advantage of allowing the variation of dimensionality from 1 to the total number of features (27 here).

Classification results are provided in Figures 4 and 5 for each used technique from the LDA family in terms of overall accuracy rate and fear accuracy rate respectively. It leads to the following results:

✓Using standard LDA, the best overall accuracy and fear accuracy rates reach 77.3% and 60% with 2 components.

✓Using KDA, the best overall accuracy reaches 80% with 9 components and the best fear accuracy rates reaches and 60% with 5 components.

Moreover, one can conclude that the LDA family seems to be not stable as the accuracy rate presents important variations when increasing the dimensionality. When dealing with trade-off between accuracy and dimensionality reduction, standard LDA seems to be better than KDA. In fact, the dimensionality is reduced to 2 (versus 9) with a loss of 3% for overall accuracy. For fear accuracy rate, they present the same accuracy rate with different dimensionality (2 for LDA versus 5 for KDA).

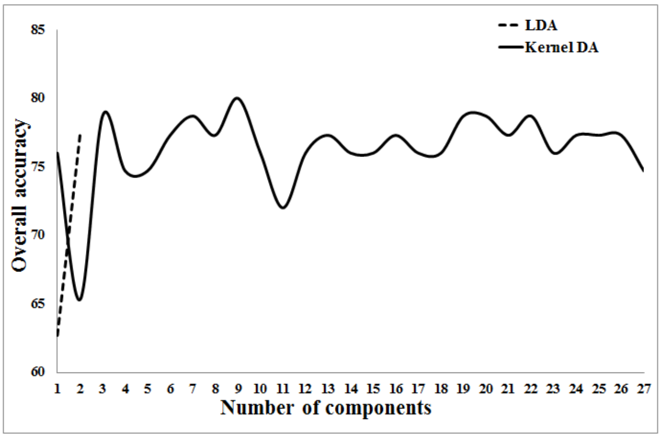

Figure 4: *Overall classification results according to LDA family.*

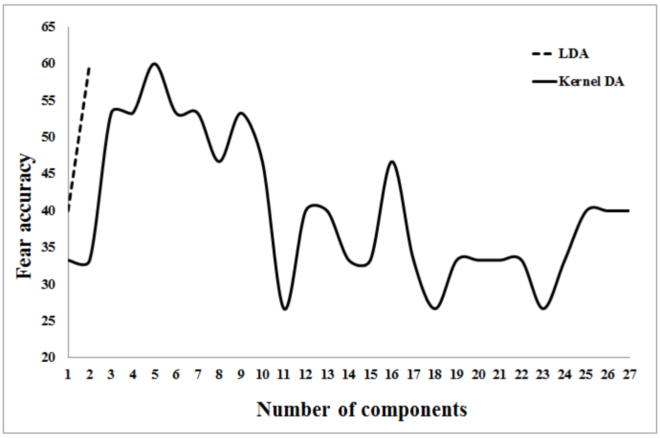

Figure 5: *Fear accuracy rates according to LDA family.*

5.2. Locality Preserving Projection Family

LPP is an unsupervised family based on mapping the data in a low dimensional space preserving the neighborhood structure of the dataset [18]. This mapping is obtained by constructing first the adjacency graph, then attempting to minimize an objective function. This latter ensures that if two data points are close in the original space, then their transformation in the embedded space are also close.

The linear property of classical LPP may lead to modeling failure when the data structure is non-linear. The basic idea of kernel LPP is to non-linearly map the data into a reduced feature space by using the non-linear structure of the features. To this end, the kernel trick is applied to extract nonlinear kernel model.

Classification results for LPP and KLPP are provided in Figures 6 and 7 and lead to the following results:

✓Using standard LPP, the best overall accuracy rate reaches 90.7% using 20 components. As for the fear accuracy, the best rate is obtained using 14 components reaching 86.7%.

✓Using KLPP, the best overall accuracy reaches 73.3% with 20 components and the best fear accuracy rate reaches 53.3% with 19 components.

One can deduce that the classification quality presents an increasing variation according to LPP as well as KLPP. Also, they both stabilize in the high dimensionality for which they reach their highest quality accuracies. Moreover, LPP seems to be better than KLPP in terms of classification performance for a fixed value of dimensionality greater than 9.

5.3. Other Techniques for dimension Reduction

In addition to the mentioned families, many other different techniques have been tested in this study namely Isomap, Landmark Isomap, Factor Analysis, Sammon Mapping, Locally Linear Embedding, Laplacian Eigenmaps, Local Tangent Space Alignment, Diffusion Maps, Stochastic Neighbor Embedding, Manifold Charting, Gaussian Process Latent Variable Model, Deep Autoencoders and Neighborhood Components Analysis. Their best classification results in terms of the overall accuracy rate and fear accuracy rate and their corresponding dimensions are summarized in Table 5. One can notice that they lead to worst results compared to previous ones.

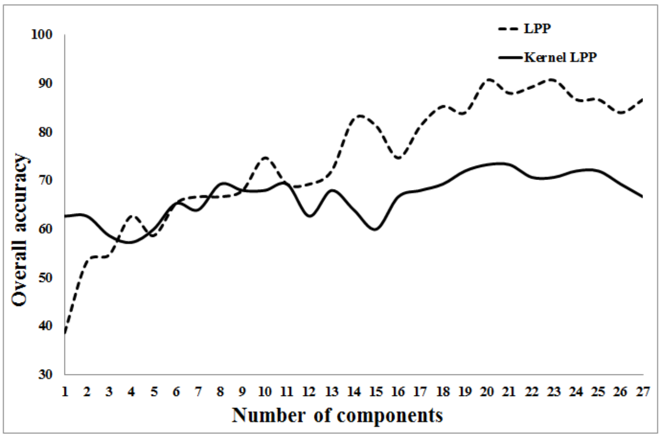

Figure 6: *Overall classification results according to LPP family techniques.*

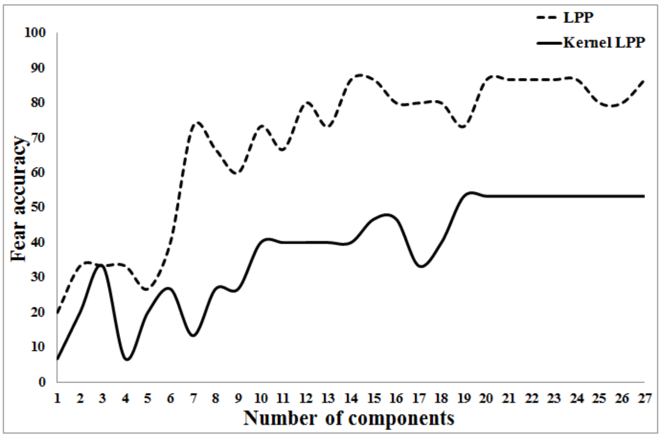

Figure 7: *Fear accuracy rates according to LPP family techniques.*

6. CONCLUSION

In this paper, we tested different techniques to reduce the relatively high dimensional feature set in order to guarantee a high overall classification rate and a high fear recognition rate. The first tested approach is manual and based on 'N to N' combinations. It leads to good results reaching 93.34% as an overall accuracy rate and 78.7% as a fear recognition rate. The other approaches are automatic. A comparative study between them was presented. It shows that the best fear recognition rate is obtained using principal components analysis reaching 93,3% using 19 components, which is practically the same result obtained for 'N to N' approach. If we aim to reduce more the dimensionality, we can use KPCA but we loose in terms of classification performance.

Table 5: *Best classification results using other dimensionality reduction techniques.*

Reduction Techniques	Overall detection		Fear detection	
	Overall Accuracy	*Dimension*	*Fear Accuracy*	*Dimension*
Isomap	45	2	40	3
Landmark Isomap	50	3	55	10
Factor Analysis	52	10	40	12
Sammon Mapping	54.7	22	53.3	20
Locally Linear Embedding	54.3	25	47.3	24
Laplacian Eigenmaps	49.3	25	46.3	22
Local Tangent Space Alignment	44.3	12	33.3	20
Diffusion Maps	40.3	6	33.3	8
Stochastic Neighbor Embedding	52.4	24	33.3	23
Deep Autoencoders	44.3	12	39.7	16
Neighborhood Components Analysis	53.3	22	55.7	24

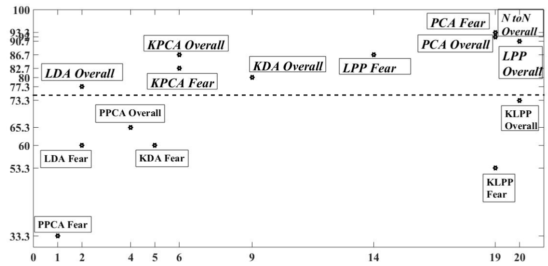

Figure 8: *Classification results according to all dimensionality reduction families and 'N to N' approach.*

7. REFERENCES

[1] Rupayan Chakraborty, Meghna Pandharipande, and Sunil Kumar Kopparapu, *Analyzing Emotion in Spontaneous Speech*, Springer, 2018.

[2] Moataz El Ayadi, Mohamed S Kamel, and Fakhri Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.

[3] Paul K Davis, Walter L Perry, Ryan Andrew Brown, Douglas Yeung, Parisa Roshan, and Phoenix Voorhies, *Using behavioral indicators to help detect potential violent acts*, RAND Corporation, 2013.

[4] Beatrice De Gelder, Jeffrey S Morris, and Raymond J Dolan, "Unconscious fear influences emotional awareness of faces and voices," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 51, pp. 18682–18687, 2005.

[5] Yuichiro Anzai, *Pattern recognition and machine learning*, Elsevier, 2012.

[6] S. Chebbi and S. Ben Jebara, "On the use of pitch-based features for fear emotion detection from speech," in *4th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, 2018.

[7] Johan AK Suykens and Joos Vandewalle, "Least squares sup-

port vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.

[8] S Rasoul Safavian and David Landgrebe, "A survey of decision tree classifier methodology," *IEEE transactions on systems, man, and cybernetics*, vol. 21, no. 3, pp. 660–674, 1991.

[9] Gaston Baudat and Fatiha Anouar, "Generalized discriminant analysis using a kernel approach," *Neural computation*, vol. 12, no. 10, pp. 2385–2404, 2000.

[10] Padraig Cunningham and Sarah Jane Delany, "k-nearest neighbour classifiers," *Multiple Classifier Systems*, vol. 34, pp. 1–17, 2007.

[11] Andrew R Webb, *Statistical pattern recognition*, John Wiley & Sons, 2003.

[12] Sergios Theodoridis, Aggelos Pikrakis, Konstantinos Koutroumbas, and Dionisis Cavouras, *Introduction to pattern recognition: a matlab approach*, Academic Press, 2010.

[13] Igor Kononenko, "Estimating attributes: analysis and extensions of relief," in *European conference on machine learning*. Springer, 1994, pp. 171–182.

[14] S. Chebbi and S. Ben Jebara, "On the selection of relevant features for fear emotion detection from speech," in *submitted in European Signal Processing Conference (EUSIPCO)*, 2018.

[15] Luis O Jimenez and David A Landgrebe, "Supervised classification in high-dimensional space: geometrical, statistical, and asymptotical properties of multivariate data," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 28, no. 1, pp. 39–54, 1998.

[16] Svante Wold, Kim Esbensen, and Paul Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.

[17] Petros Xanthopoulos, Panos M Pardalos, and Theodore B Trafalis, "Linear discriminant analysis," in *Robust data mining*, pp. 27–33. Springer, 2013.

[18] Xiaofei He and Partha Niyogi, "Locality preserving projections," in *Advances in neural information processing systems*, 2004, pp. 153–160.

[19] Bruce Thompson, *Exploratory and confirmatory factor analysis: Understanding concepts and applications.*, American Psychological Association, 2004.

[20] Warren S Torgerson, "Multidimensional scaling: I. theory and method," *Psychometrika*, vol. 17, no. 4, pp. 401–419, 1952.

[21] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural computation*, vol. 10, no. 5, pp. 1299–1319, 1998.

[22] Sebastian Mika, Gunnar Ratsch, Jason Weston, Bernhard Scholkopf, and Klaus-Robert Mullers, "Fisher discriminant analysis with kernels," in *Neural networks for signal processing IX, 1999. Proceedings of the 1999 IEEE signal processing society workshop.* Ieee, 1999, pp. 41–48.

[23] Joshua B Tenenbaum, Vin De Silva, and John C Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[24] David DeMers and Garrison W Cottrell, "Non-linear dimensionality reduction," in *Advances in neural information processing systems*, 1993, pp. 580–587.

[25] David Talkin, "A robust algorithm for pitch tracking (rapt)," *Speech coding and synthesis*, vol. 495, pp. 518, 1995.

[26] Felix Burkhardt, Astrid Paeschke, Miriam Rolfes, Walter F Sendlmeier, and Benjamin Weiss, "A database of german emotional speech," in *9th European Conference on Speech Communication and Technology*, 2005.

[27] Sam T Roweis, "Em algorithms for pca and spca," in *Advances in neural information processing systems*, 1998, pp. 626–632.