Technische Universität Dresden
Institut für Technische Akustik
Lehrstuhl für Sprachkommunikation
Prof. Dr. habil. R. Hoffmann

Diplomarbeit

A Learnable Signal Transformation as Nonlinear Generalization of Linear Discriminant Analysis and its Application in Speech Recognition

Eine Lernbare Signaltransformation als Nichtlineare Verallgemeinerung der Linearen Discriminanz Analyse und ihre Anwendung in der Spracherkennung

Carnegie Mellon University
School of Computer Science
Center for Machine Translation
Pittsburgh, PA 15213, USA

under supervision of
Prof. Dr. Alexander Waibel

Author: Thorsten Schüler
Rudolf Leonhard Strasse 22, 01097 Dresden, Matrikelnummer 89090950
Supervisor: Prof. Dr. habil. Rüdiger Hoffmann (TU Dresden)
Issue Date: 04/01/94
Deadline: 09/31/94

Abstract

An important aspect in speech recognition and other classification tasks is the feature extraction. In this paper it is shown how to derive a nonlinear generalization of the LDA concept. Therefore, a seperation quality measure based on within-class scatter matrix and between-class scatter matrix is introduced. Looking for a linear operator wich minimizes this measure leads to the well known Linear Discriminant Analysis (LDA). It is shown how to find a nonlinear operator which is able to minimize this quality measure better than the LDA operator. In order to find this nonlinear operator, connectionist methods are applied. The so found transformation is tested with a continuous speech recognizer and a large vocabulary database. In this tests monophone and triphone acoustical models were applied as classes in the classification task. With this experiments it could be shown that the new nonlinear transformation indeed increases class seperability. However, this improvement did not result in an improvement in word recognition performance in every experiment. In the experiments where the word recognition performance could be improved, the improvement was rather small. For the best experiment with monophone acousitcal models the realtive improvement in error rate was 11% compared with LDA. In the case of triphone acoustical models the realtive improvement in error rate was 4.4% for the best experiment.

Acknowledgements

In the first place I would like to thank Ursula for putting up with me while I wrote this thesis. I would like to thank Professor Alex Waibel for the advise, the host, and for many intresting conversations, Professor Ruediger Hoffmann for the supervision at the University in Dresden, Ivica Rogina for the help with the recognizer and an always open ear and all the folks of the nnspeech groups in Pittsburgh and Karlsruhe. Thanks to the German National Scholarship Foundation for founding me while I stayed at the CMU.

$\mathbf{X}$	random vector
$\mathbf{x_i}$	i-th coefficient of a random vector, one dimensional random variable
X	sample of a random vector
x_i	i-th coefficient of a sample of a random vector
$E\{\mathbf{X}\}$	expected vector of a random vector
$p(X)$	density function of a random vector
$p(X\|\omega_i)$	conditional density function
M	expected vector
ω_i	class i
L	number of classes
$E\{\mathbf{X}\|\omega_\mathbf{i}\}$	conditional expected vector for class i
P_i	a priori probability of class i
q_i	a posteriori probability of class i
N	totatl number of samples
N_i	number of samples assigned to class i
Σ	covariance matrix
Σ_i	covaraince matrix for class i
c_{ij}	i-th row and j-th column element of Σ
S_w	within-class scatter matrix
S_b	between-class scatter matrix
S_m	mixture or total scatter matrix
$\|S_b\|$	determinant of S_b (or magnitude if argument is a vector)
$(s_{ij})_b$	ij-th element of S_b
$\hat{\Sigma}$	an estimate of Σ
m, n	dimensionalities
K	LDA kernel
A	a matrix
A^T	the transposed of the matrix
A^{-1}	the inverse of the matrix
lr	backpropagation learning rate
m	backpropagation momentum
Q_μ	seperation quality measure considering μ classes
α, β, γ	moving target parameter

Contents

Chapter 1

Introduction

Speech recognition becomes increasingly important in today's technological society. With this, there are higher demands on the quality of speech recognition systems. Nobody in the real world wants to adapt his way of speaking in order to make a machine understand. (That kind of sacrifice is for researchers only.) Therefore the domains of continuous speech and spontaneous speech recognition become more and more important in the vast area of speech recognition. Furthermore, modern speech recognition systems are required to understand large vocabularies.

With increasing the vocabulary and the use of continuous or even spontaneous speech, the complexity of the speech recognition task increases drastically. This results not only in the demand to have better classifiers but also demands new ways for feature extraction. On the one hand this is because in those tasks the computational time becomes important. Due to large vocabulary and unknown word boundaries, the set of possibilities to be searched through increases dramatically. In order to save time at another stage, one looks for features with as small as possible dimensionality. On the other hand the features used for the recognition can influence the performance of the whole recognition system drastically.

Speech recognition can be seen as a multi stage classification problem. Depending on the structure of the recognizer, the classes may be more or less related to each other. As every classification problem, speech recognition can be splitted in a feature extraction step and a discrimination step. This split is very coarse but may serve to guide an approach. Naturally, feature extraction and discrimination are not orthogonal to each other but highly dependend upon each other. A feature representation which performs well with one type of discriminator can give terrible results with a different type of discriminator. Unfortunately only for simple discriminators (for instance linear ones) is there a straight forward procedure to derive features. In speech recognition, due to the very complex structure of the discriminators, one is usually forced to handle the feature extraction as an independent step. However, one should at least keep in mind that these two processes are not independent.

This thesis deals with the first step: the feature extraction. There are two major requests on the features:

- The features should work well with the discriminator in order to achieve the best recognition performance on the final stage, usually the word level. The performance measure is then the *word accuracy*.
- The dimensionality of the feature vectors should be as small as possible in order to save computational time.

One way to find signal transformation in order to achieve good features in terms of recognition performance is to look at the way the human auditory systems does its feature extraction. That leads to spectral representations of the speech signal as features. This spectral representation can be achieved by applying a Fourier Transform on the speech signal. This kind of transformation should be called a primary transformation.

Unfortunately, this leads to rather high dimensional feature vectors. In order to reduce the dimensionlity, usually a secondary transformation is applied. Thus, from this point of view, the feature extraction itself can be seen as a two stage process. The focus of this thesis is on the second stage of the feature extraction, the dimension reducing secondary transformation.

There are a number of ways known to reduce the dimensionality of the feature vectors. Some of them are derived by heuristic approaches inspired by the analogy to the human auditory system such as the projection onto melscale coefficients (chapter 5). Sometimes there is no strong borderline between primary and secondary transformation as in the case of filterbanks. There, the dimensionality of the feature vectors can be chosen small by increasing the channel bandwidth with the frequency. However, underlying there is still the theoretical split into primary and secondary transformation.

Another way to achive dimension reducing secondary transformations is the use of statistical methods. This leads first to the well known *Linear Discriminant Analysis* (LDA) which gives a linear projection from a high dimensional space onto a low dimensional space with minimized loss of classification information. This projection is derived by linear optimization of a norm based on measures for the within-class and between-class distances.

The question we ask here is whether it is possible to derive, by a nonlinear optimization of this norm, a transformation which gives not only an optimized projection but also improves class seperability in the feature space.

Since class seperability does not necesarilly improve word recognition performance, it is then still to be shown that this transformation, once found, improves also the *word accuracy*, the important measure of the quality of a speech recognition system.

Chapter 2 gives a brief introduction in the view of a spectral speech representation as random vectors. There, the terms covariance matrix and scatter matrix are briefly reviewed. It is shown, how to derive estimates for the parameters of random vectors.

Chapter 3 gives a brief derivation of the concept of Linear Discriminant Analysis (LDA). Therefore, a measure for class seperability is derived by using within-class and between-class scatter matrices. Some properties of the measure are discussed. Based on this measure, LDA is derived by searching for a linear operator which minimizes the measure.

In chapter 4, the aforementioned measure is minimized nonlinearly, thus leading to NonLinear Discriminant Analysis (NLDA). The nonlinear operator which performs the minimization is found by applying connectionist methods. Three different neural network architectures able to provide the nonlinear operator are introduced and their properties are discussed. All three networks are trained with an error backpropagation algorithm. A special method for calculating the targets for the training (the concept of moving targets) is applied.

In order to see whether NLDA brings any advantages over LDA in terms of word recognition performance, experiments with a hidden markov model based continuous speech recognizer have been carried out. As database, the well known Resource Management Database was used. Due to the big number of parameters, which had to be varied, experiments have been carried out on a restricted cross validation set in the testing first. Chapter 5 describes the experiments with monophones as acoustical modeling, in Chapter 6 the experiments with triphones as acoustical modeling are evaluated.

Then, in Chapter 7, experiments on the official test set with the best parameter combinations, investigated on the cross validation set, are described for both, monophones and triphones as acoustical modeling.

Chapter 8 gives a summary and a conclusive discussion. Some ideas for further work are given there, too.

All experiments are evaluated in detail in the appendices D and E. In appendix E, the experiments with monophone acoustical modeling are evaluated. In appendix D, the experiments with triphone acoustical modeling are reviewed.

Chapter 2

Spectral Representation of Speech as Random Vectors

Since the vectors of a spectral representation of speech data are somehow clustered in the pattern space, it seems useful to interpret them as *conditional random vectors*. This chapter gives an explanation of the random vector concept and describes some of the random vector properties.

2.1 Random Vectors

Let the n-dimensional vector

$$\mathbf{X} = [\mathbf{x}_1 \mathbf{x}_2 \ldots \mathbf{x}_n]^T \tag{2.1}$$

be a *random vector* with n variables $\mathbf{x}_1, \ldots, \mathbf{x}_n$. Then this random vector may be characterized by a *probability distribution function* (pdf)

$$P(x_1, \ldots, x_n) = prob\,\{\mathbf{x}_1 \leq x_1, \ldots, \mathbf{x}_n \leq x_n\} \tag{2.2}$$

or in shorthand form,

$$P(X) = prob\,\{\mathbf{X} \leq X\}\,, \tag{2.3}$$

where $prob\,\{\xi\}$ stands for the probability that event ξ occurs. Another characteristic function of a random vector is the *density function* defined by :

$$p(X) = \partial^n P(X)/\partial x_1 \ldots \partial x_n. \tag{2.4}$$

It should be mentioned that the density function itself is not a probability but would need to be integrated over a certain region ΔX to obtain a probability.

In speech recognition, one distinguishes between different classes ω_i and tries to map an instantaneous event X to one of these classes. Therefore it is advantageous to introduce the concept of a *conditional random vector*. A conditional random vector is characterized not by only one density function but by L *conditional density functions* (cdf)

$$p(X|\omega_i) \quad \text{or} \quad p_i(X) \qquad (i = 1, \ldots, L). \tag{2.5}$$

Each cdf describes the density function over one class, where L is the total number of classes. The sum over these conditional density functions, weighted with the *a priori probability* P_i for each class, leads then to the unconditional or *mixture* density function

$$p(X) = \sum_{i=1}^{L} P_i p_i(X). \tag{2.6}$$

In view of the recognition (or classification) task, it is useful to mention here the notion of the *a posteriori probability* q_i of a class ω_i for a given X, which is determined by the *Bayes Theorem*:

$$q_i = P(\omega_i|X) = \frac{P_i p_i(X)}{p(X)}. \tag{2.7}$$

A more detailed introduction to the concept of random vectors can be found in [Fuk90], [Wil63].

2.2 Parameter of Random Vector Distributions

Although a random vector is fully determined by its distribution or its density function, one often has problems to obtain these functions or may not be able to use them due to their mathematical complexity. Therefore it is preferable to use a less complete but easier to obtain characterisation, namely the *expected vector* and the *covariance matrix*.

2.2.1 Expected Vector of a Distribution

The *expected vector* or *mean* of a random vector $\mathbf{X}$ is defined by

$$M = E\{\mathbf{X}\} = \int_X X p(X) dX. \tag{2.8}$$

One can show that each component of M can be calculated independently from the other components as the expected value of an individual variable with one-dimensional density

$$m_i = \int_{x_i} x_i p(x_i) dx_i, \tag{2.9}$$

where $p(x_i)$ is the *marginal density* of the i-th component of X given by

$$p(x_i) = \int_{x_1} \cdots \int_{x_n} p(X) dx_1 \cdots dx_{i-1} dx_{i+1} \cdots dx_n. \tag{2.10}$$

Given a conditional random vector $\mathbf{X}$, one can define the *conditional expected vector* for class ω_i by

$$M_i = E\{\mathbf{X}|\omega_i\} = \int_X X p_i(X) dX. \tag{2.11}$$

The unconditional expected vector is then the weighted sum over all conditional expected vectors

$$M = \sum_{i=1}^{L} P_i M_i, \tag{2.12}$$

where the P_i's again are the *a priori* probabilities of different classes.

2.2.2 Covariance Matrix of a Distribution

Another inportant set of paramters, the *covariance matrix*, measures the dispersion of the distribution and is defined by

$$\begin{aligned}
\Sigma &= E\{(\mathbf{X} - M)(\mathbf{X} - M)^T\} \\
&= \begin{bmatrix} E\{(\mathbf{x}_1 - m_1)(\mathbf{x}_1 - m_1\} & \cdots & E\{(\mathbf{x}_1 - m_1)(\mathbf{x}_n - m_n\} \\ E\{(\mathbf{x}_2 - m_2)(\mathbf{x}_1 - m_1\} & \cdots & E\{(\mathbf{x}_2 - m_2)(\mathbf{x}_n - m_n\} \\ \vdots & & \vdots \\ E\{(\mathbf{x}_n - m_n)(\mathbf{x}_1 - m_1\} & \cdots & E\{(\mathbf{x}_n - m_n)(\mathbf{x}_n - m_n\} \end{bmatrix} \\
&= \begin{bmatrix} c_{11} & \cdots & c_{1n} \\ \vdots & & \vdots \\ c_{n1} & \cdots & c_{nn} \end{bmatrix}.
\end{aligned} \tag{2.13}$$

The components c_{ij} of the matrix Σ are

$$c_{ij} = E\{(\mathbf{x}_i - m_i)(\mathbf{x}_j - m_j)\} \tag{2.14}$$

and they obey the equation

$$c_{ij} = c_{ji}. \tag{2.15}$$

The diagonal elements of Σ are the *varainces* of the individual random varaibles; the off diagonal elements measure the *covariance* between two random variables $\mathbf{x}_i$ and $\mathbf{x}_j$. There is a strong connection between the covaraince of two random variables and the *correlation coefficient* ρ_{ij} of these variables

$$\rho_{ij} = \frac{c_{ij}}{\sigma_i \sigma_j}, \tag{2.16}$$

where σ_i is the *standard derivation* of the varaible $\mathbf{x}_i$ and $\sigma_i^2 = c_{ii}$.

It is important to distinguish carefully between the notion of covariance and correlation: while the covariances c_{ij} depend on the scale of the coordinate system, the correlation coefficients ρ_{ij} are invariant under scaling. Subsequently, the correlation coefficients ρ_{ij} retain the essential information of the relation between random variables.

Again, for a conditional random vector $\mathbf{X}$ we can define *conditional covariance matrices* or *class covaraince matrices* Σ_i for class ω_i by

$$\Sigma_i = E\left\{(\mathbf{X} - M_i)(\mathbf{X} - M_i)^T|\omega_i\right\}. \tag{2.17}$$

In view of the classification task, it is useful to measure somehow the dispersion between the classes and and within the classes. Therefore we introduce the notion of *scatter matrices* here.

2.2.3 Scatter Matrices

A *within-class scatter matrix* S_w shows the dispersion of samples around their respective class expected vector, given by (2.11):

$$S_w = \sum_{i=1}^{L} P_i E\left\{(\mathbf{X} - M_i)(\mathbf{X} - M_i)^T|\omega_i\right\} = \sum_{i=1}^{L} P_i \Sigma_i. \tag{2.18}$$

Therefore, the within-class scatter matrix is an integrated representation of the different class covariance matrices. This integration gives the advantage of dealing with one single matrix instead of L covariance matrices.

By defining the *between-class scatter matrix* S_b

$$S_b = \sum_{i=1}^{L} P_i (M_i - M_0)(M_i - M_0)^T, \tag{2.19}$$

we can measure the dispersion of the class expected vectors around the expected vector of the mixture distribution M_0, given by (2.12).

The *mixture scatter matrix* S_m, which is identical with the covariance matrix of all samples regardless of their class assignments, is the sum of the within-class and between-class scatter matrices

$$S_m = E\left\{(\mathbf{X} - M_0)(\mathbf{X} - M_0^T\right\} = S_w + S_b. \tag{2.20}$$

All scatter matrices are invariant with respect to coordinate shifts, but variant with respect to scaling of the coordinate system.

2.3 Estimation of Parameters

The mean and the covariance matrix give a different characterization of a distribution, but both are still unknown at this point and need to be obtained by some estimation technique. Fortunately, it is usually easier to obtain an estimate of the mean and the covariance matrix rather than an estimate of the density function.

A normaly used estimation technique is the *sample estimation technique* described in [KNF75], [TSM85]. This technique estimates the expected vector of a distribution as the average vector of all N samples drawn from this distribution

$$\hat{\mathbf{M}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{X}_i. \tag{2.21}$$

In the same way, we can obtain an estimate for the covariance matrix by summing over the vector product of the difference vector between the estimated mean $\hat{\mathbf{M}}$ and the sample $\mathbf{X}_i$ with its transpose for all drawn samples

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1}\sum_{i=1}^{N}(\mathbf{X}_i - \hat{\mathbf{M}})(\mathbf{X}_i - \hat{\mathbf{M}})^T. \tag{2.22}$$

The use of $N-1$ rather than N in the denominator of (2.22) is in order to obtain an unbiased estimate for Σ rather than a biased one[Fuk90]; that means, the expected matrix of $\hat{\boldsymbol{\Sigma}}$ is Σ.

Note that both, $\hat{\mathbf{M}}$ and $\hat{\boldsymbol{\Sigma}}$, are printed bold, since both are themselves random vectors and therefore each have a distribution and a density function. The distributions of the estimates of the mean and the covariance matrix and their properties are discussed in detail in [Fuk90]. There it is also shown that the expected values of $\hat{\mathbf{M}}$ and $\hat{\boldsymbol{\Sigma}}$ are indeed M and Σ.

Equation (2.22) may be rewritten as

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{N-1}\sum_{i=1}^{N}\mathbf{X}_i\mathbf{X}_i^T - \frac{N}{N-1}\hat{\mathbf{M}}\hat{\mathbf{M}}^T. \tag{2.23}$$

The use of (2.21) and (2.23) gives a computational advantage over using (2.22), since the estimates for M and Σ can be obtained in one pass over the data instead of two passes.

For a mixture distribution, we can obtain estimates for the scatter matrices by using (2.21) and (2.23) with respect to the class assignments of the samples:

$$\begin{aligned}
\hat{S}_w &= \sum_{i=1}^{L}\frac{N_i}{N}\left(\frac{1}{N_i-1}\sum_{j=1}^{N_i}\mathbf{X}_j^i\mathbf{X}_j^{iT} - \frac{N_i}{N_i-1}\hat{\mathbf{M}}_i\hat{\mathbf{M}}_i^T\right) \\
&\approx \sum_{i=1}^{L}\frac{1}{N}\left(\sum_{j=1}^{N_i}\mathbf{X}_j^i\mathbf{X}_j^{iT} - N_i\hat{\mathbf{M}}_i\hat{\mathbf{M}}_i^T\right), && (2.24) \\
\hat{S}_b &= \sum_{i=1}^{L}\frac{N_i}{N}(\hat{\mathbf{M}}_i - \hat{\mathbf{M}}_0)(\hat{\mathbf{M}}_i - \hat{\mathbf{M}}_0)^T, && (2.25) \\
\hat{S}_m &= \frac{1}{N-1}\sum_{i=1}^{L}\sum_{j=1}^{N_i}\mathbf{X}_j^i\mathbf{X}_j^{iT} - \frac{N}{N-1}\hat{\mathbf{M}}_0\hat{\mathbf{M}}_0^T. && (2.26)
\end{aligned}$$

In (2.24)-(2.26), the superscript i in $\mathbf{X}_j^i$ determines the assignment of this sample to class ω_i, the N_i stands for the number of samples belonging to this class.

Since most of the experiments in this work are carried out on a large amount of data, the identity between estimate and true value is assumed. This means, there will no longer be a distinction made between, e.g. $\hat{\boldsymbol{\Sigma}}$ and Σ. However, the approximative nature of the parameters should be always kept in mind.

2.4 Multivariate Normal Distribution

An important distribution, although in practice rather too simple to be applied, is the *multrivariate normal distribution* with cdf

$$p(X) = N_x(M, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} exp\left(-\frac{1}{2}d^2(X)\right), \tag{2.27}$$

where d^2 stands for the *squared generalized distance* or *Mahalanobis distance*

$$d^2(X) = (X - M)^T \Sigma^{-1}(X - M), \tag{2.28}$$

with M beeing the expected vector of the distribution, Σ its covaraince matrix, Σ^{-1} the inverse of the covariance matrix and n the dimensionality of the vector space.

There are several reasons why the normal distribution is important [Fuk90], especially the invariance of its quadratic form and its positive definiteness under a nonsingular linear transformation and the posibility to express every other function in terms of itself.

Unfortunately, the conditional distributions of the classes used in Speech Recognition (mostly phonems) are usually not unimodal mulitvariate normal but rather multimodal (i.e., a mixture distribution of unimodal multivariate normal distributions).

Nevertheless it is sometimes useful to assume they are unimodal normal distributed to simplify the mathematical treatment of the problem. However, one should at least be conscious about the error one makes.

Therefore, it is necessary to find a quality measure for the goodness of the unimodal normality assumption. [Fuk90] proposes a quality measure by observing the variance of (2.28). It is shown there, that for unimodal normal $\mathbf{X}$, the varaible $\xi = d^2/(N-1)$ has a beta distribution and

$$Var\{\xi\} = \frac{2n}{(N-1)^2} \frac{1-(n+1)/N}{1+1/N} \tag{2.29}$$

holds for its variance. Therefore the goodness of the unimodal normal assumption can be measured for each class by calculating the variance of (2.28) for all samples of this class, dividing it by $(N-1)$ and comparing the result with the right hand side of (2.29).

Chapter 3

Linear Discriminant Analysis (LDA)

As shown in [Fuk90], the lowest (from a statistical viewpoint) possible classification error for a L class problem is the *Bayes error* ϵ

$$\epsilon = E\left\{1 - \max_i q_i(X)\right\} \qquad (i = 1, \ldots, L) \tag{3.1}$$

where the X are the observations of the random vector $\mathbf{X}$ and q_i is the *a posteriori* probability of class ω_i. Decision is then based on these probabilities where the decision rule is

$$q_k = \max_i q_i(X) \Rightarrow X \in \omega_k. \tag{3.2}$$

Therefore the set $\{q_1(X), \ldots, q_L(X)\}$ carries all necessary information for a classifier based on (3.2). Since

$$\sum_{i=1}^{L} q_i(X) = 1, \tag{3.3}$$

only $(L-1)$ of the *a posteriori* probability functions are linear independent. Subsequently, any $(L-1)$ functions of the set form a basis for a $(L-1)$-dimensional space. The Bayes error in this $(L-1)$-dimensional space is identical to the Bayes error in the original n-dimensional X-Space. It follows that the transformation

$$y_i = q_i(X) \qquad (i = 1, \ldots, L-1) \tag{3.4}$$

from a n-dimensional space onto a $(L-1)$-dimensional space does not result in a loss of classification information. One calls the set $\{q_1(X), \ldots, q_{L-1}(X)\}$ the *ideal feature set for classification*.

3.1 A Quality Measure for Class Separability

Although the *a posteriori* probabilities are ideal features, it is hard to obtain them. Furthermore, known estimation techniques [Fuk90] deliver estimates with severe biases and variances. This same problem exists for the Bayes error, which is the best criterion to evaluate feature sets. Due to its complex form, it is in most cases useless as analytic tool for feature extraction.

In order to derive analytic tools for feature extraction, one needs simpler criteria. One attempt to achieve this is to find upper bounds for the Bayes error, such as the *Bhattacharyya distance*.

Another way to find a simple criterion for feature extraction is to measure the *class separability*, i.e. to find a number which expresses the ability a feature set gives to discriminate between different classes.

Since the one measure the dispersion within and the other between the classes, it seems straight forward to use the scatter matrices (2.24) and (2.25) to obtain such a number, since they measure the dispersion within and between the classes respectivly. Futhermore, both are easy to calculate and are not bothered by biases.

Of course, for good class separability one wants the distance (measured for instance by (2.28)) between the means of the classes to be as large as possible and the classes themselves rather concentrated around their mean vectors.

If one assumes, with probability 1, the nonsingularity of the inner-class scatter matrix [1], one can calculate its inverse and premultiply it with the between-class scatter matrix. If one further assumes the nonsingularity of the between-class scatter matrix [2], one can establish a *separability criterion* by calculating

$$Q = -\log\left|S_w^{-1}S_b\right| = log\,|S_w| - log\,|S_b|\,. \tag{3.5}$$

Sometimes one finds in the literature this criterion without the logarithm. This is in principle a similar criterion. Yet, since for higher dimensionality the determinants of the scatter matrices are small numbers, due to numerical reasons it is advantageous to apply the logarithm.

To show that (3.5) really measures the class seperability, we put forward the following theorem:

Theorem 3.1 *Let* $\mathbf{X}$ *be a conditional random vector. Let* S_w *be the within-class scatter matrix of* $\mathbf{X}$ *and* S_b *the between-class scatter matrix, both nonsingular with probability 1.*

1. *Let* $T : \mathbf{R}^n \mapsto \mathbf{R}^n$ *be a linear nonsingular operator. Then, the quadratic forms* $\tilde{S}_b = T^T S_b T$ *and* $\tilde{S}_w = T^T S_w T$ *fulfill*
$$|\tilde{S}_w^{-1}\tilde{S}_b| = |S_w^{-1}S_b|. \tag{3.6}$$
Thus, the criterion (3.5) is invariant against the transformation T.

2. *Let* $f : \mathbf{R}^n \mapsto \mathbf{R}^n$ *be a mapping, with*
$$Y = f(X) = [c_1x_1, \ldots, c_nx_n]^T \qquad c_i \in \mathbf{R}, c_i \neq 0.$$
Then, for the within-class and the between-class scatter matrices $\tilde{S}_w$ *and* $\tilde{S}_b$ *in the* Y*-space (3.6) holds. Thus, (3.5) is also invariant against transformations of type* f*, i.e., invariant against separate scaling of coordinate axis.*

3. *The criterion (3.5) is invariant against coordinate shifts.*

Proof: *Appendix A*

Suppose now, we have found a linear operator $T : \mathbf{R}^n \mapsto \mathbf{R}^n$ which transforms $(S_w^{-1}S_b)$ into a diagonal matrix. Such an operator will exist due to the nonsingularity of S_w and S_b and is the *eigenvector* matrix of $(S_w^{-1}S_b)$. Moreover, since both matrices are symmetric, it follows that they are positive definite. From the theory of such matrices [BM53] one knows, that the operator T not only transforms $(S_w^{-1}S_b)$ into a diagonal matrix but also diagonalizes the matrices S_w and S_b. In other words, $(S_w^{-1}S_b)$, S_w and S_b share the same *eigenvectors* (but may have different *eigenvalues*).

Denote now the transformed scatter matrices $\tilde{S}_w$ and $\tilde{S}_b$. The criterion (3.5) then simplifies to

$$Q = \sum_{i=1}^{n} \log(\tilde{s}_{ii})_w - \sum_{i=1}^{n} \log(\tilde{s}_{ii})_b, \tag{3.7}$$

where the $(\tilde{s}_{ii})_b$ are the diagonal elements of $\tilde{S}_b$ and the $(\tilde{s}_{ii})_w$ the diagonal elements of $\tilde{S}_w$.

[1] This is a small restriction of generality, but is usually not hard to guarantee

[2] This is a much larger restriction of generality, since from (2.12) and (2.19) it follows, that only $(L-1)$ columns or rows of S_b are linear independent. This in turn restricts the applicability of the derived criterion to original spaces with dimension $n \leq L-1$. Since the dimensions of the problems we treat here are always lower than the number of classes decreased by one, this restriction is always fulfilled here. By taking the total scatter matrix S_m instead of S_b one can annul this restriction if desired.

On the other hand, due to (2.19), for the $(\tilde{s}_{ii})_b$ it holds that

$$(\tilde{s}_{ii})_b = \sum_{k=1}^{L} P_i \left((\widetilde{m}_i)_k - (\widetilde{m}_0)_k \right)^2 . \tag{3.8}$$

Thus, a large between-class distance means large diagonal elements of the between-class scatter matrix $\tilde{S}_b$.

An analog treatment of the within-class scatter matrix leads to the conclusion, that a small within-class distance [1] is equivalent to small diagonal elements of the within-class scatter matrix $\tilde{S}_w$.

Obviously, (3.7) decreases with increasing $(\tilde{s}_{ii})_b$ and decreasing $(\tilde{s}_{ii})_w$. Thus, a smaller value of (3.7) is equivalent with a larger distance between the classes and/or a smaller distance within the classes. Consequently, (3.7) and therefore (3.5) are valid separability criteria, although to this point only for diagonal scatter matrices.

However, if one remembers that the diagonal forms $\tilde{S}_w$ and $\tilde{S}_b$ were results of a linear transformation T, carried out on a space with general scatter matrices S_w and S_b, and if one further applies part 1 of theorem 3.1, it follows immediatly, that (3.5) equivalently measures the separability in the original space with the general scatter matrices S_w and S_b. Therefore, (3.5) is a valid separability criterion for problems with general scatter matrices S_w and S_b.

It should be mentioned here, that the above argument implies that the class distributions are seperated by their means and, that for multimodal distribution, the expected vectors of the partial distributions are somehow close to each other. For speech data, this two implications seem fairly fulfilled.

3.2 Derivation of the LDA Kernel

The original idea of LDA was to find *linear discriminate functions* to discriminate between L classes [Fis26]. It was assumed there, that all classes have a unimodal normal distribution with different means but equivalent covariance matrix. The optimal Bayes classifier for this problem is a linear classifier which measures the Mahalanobis distance between a sample and all class means. The sample is then assigned to the class with the smallest Mahalanobis distance (under the assumption that all *a priori* probabilities are equal). Since all class covariance matrices are equal, it is possible to find a linear operator $K : \mathbf{R}^n \mapsto \mathbf{R}^n$ which turns the covaraince matrix into a unity matrix. Then the Mahalanobis distance turns into the euclidean distance and the classifier becomes a distance classifier. The operator K is called *LDA kernel*. The $(L-1)$ necessary linear discriminate functions then have the directions of the systems basis vectors. They are easy to calculate by finding points of equal distance from different means for each direction.

A reduced form of LDA adds a graphical component: One finds the subspace $\mathbf{R}^m$, $m < L-1$, $m < n$ of $\mathbf{R}^n$ in which the class means are most seperated and assignes a sample to the closest mean in this subspace. This results in a dimension reduction of the data with a linear optimized loss of information. The LDA kernel is in this case $K : \mathbf{R}^n \mapsto \mathbf{R}^m$

The derivation of the LDA in the original proposal was, naturally, based on an optimization of the $(L-1)$ linear discriminant functions. We will go here a different way, which does not preassume unimodal normal distributions with equal covariance matrices. Therefore, the optimal Bayes classifier is probably no longer a linear classifier. Since the optimal Bayes classifier for this general problem is unknown, we will apply the separability criterion (3.5) and optimize a linear operator with respect to it.

Therefore, we have to obey the restrictions mentioned above, which again are:

1. Both the between-class scatter S_b and the within-class scatter S_w are nonsingular with propability 1. (It follows from this, that the dimension of the problem is smaller than the number

[1] Since the within-class scatter matrix is in principle a superposition of the class covariance matrices, this distance is an average value, i.e. a small within-class distance does not necessary mean that the distance within *every* class is small

of classes; $n \leq (L-1)$ [1])

2. The classes are seperated by their means.

3. If the class distributions are multimodal, the means of their partial distributions are close.

It should be mentioned here, that this approach is not without dangers, since it does not pay any regard to the classifier used on the feature set. Feature extraction and classification process are not orthogonal to each other. Therefore, one later has to decide carefully, whether a certain classifier is to be employed on top of a feature set obtained by this approach or not. However, this approach leads to the same result as the original way, but seems in our case more straight forward, since one certainly cannot assume equal class covariance matrices and unimodal distributions for the different classes in a speech recognition system.

Consider now, that we want to find a nonsingular linear operator $K : \mathbf{R}^n \mapsto \mathbf{R}^m$ $m < n$, which minimizes (3.5) with respect to the quadratic forms $\tilde{S}_w = K^T S_w K$ and $\tilde{S}_b = K^T S_b K$. Then, the criterion value becomes

$$Q = \log|\tilde{S}_w| - \log|\tilde{S}_b| = \log|K^T S_w K| - \log|K^T S_b K|. \tag{3.9}$$

Taking the derivate of (3.9) with respect to K, one finds, using $(\partial/\partial A)\ln|A^T B A| = 2BA(A^T BA)^{-1}$,

$$\frac{\partial}{\partial K} Q = \frac{2}{\ln 10}\left(S_w K \left(K^T S_w K\right)^{-1} - S_b K \left(K^T S_b K\right)^{-1}\right). \tag{3.10}$$

For the operator K, which minimizes (3.9), it must hold that $(\partial/\partial K)Q = 0$ and therefore

$$S_b K \left(K^T S_b K\right)^{-1} = S_w K \left(K^T S_w K\right)^{-1}. \tag{3.11}$$

Elementary matrix operations transforms this into

$$S_w^{-1} S_b K = K \left(K^T S_w K\right)^{-1} \left(K^T S_b K\right). \tag{3.12}$$

With the abbreviation $\Lambda = \left(K^T S_w K\right)^{-1} \left(K^T S_b K\right)$ this becomes

$$S_w^{-1} S_b K = K\Lambda. \tag{3.13}$$

If one assumes now Λ to be a diagonal matrix, then (3.13) would be an eigenvalue problem and the columns of K would be the eigenvectors of $S_w^{-1} S_b$. Λ will surely be a diagonal matrix, if $K^T S_w K$ and $K^T S_b K$ are diagonal matrices. As mentioned earlier in this chapter, due to the nonsingularity and symmetry of S_w and S_b, the eigenvector matrix K of $S_w^{-1} S_b$ will not only turn $K^T S_w^{-1} S_b K$ into a diagonal matrix, but also $K^T S_w K$ and $K^T S_b K$. This again verifies the diagonality assumption of Λ.

Suppose now, that $S_w^{-1} S_b$ has n (not necessarily different) eigenvalues λ_i. Since K is $\mathbf{R}^n \mapsto \mathbf{R}^m$ with $m < n$, one still has to decide which eigenvalues and corresponding eigenvectors to choose for K. Since $\tilde{S}_w^{-1} \tilde{S}_b = \Lambda$ and the λ_i, $i = 1, \ldots, m$ are the diagonal elements of Λ, (3.5) is clearly minimized by selecting the m largest eigenvalues of $S_w^{-1} S_b$ (since $|K^T S_w^{-1} S_b K|$ is the product of the eigenvalues). The columns of K are then the eigenvectors corresponding to these m eigenvalues.

If $S_w^{-1} S_b$ has less than n, say ξ, eigenvalues, then clearly the dimension m of the transformed space has to be reduced to be equal or smaller than ξ.

Consider now, we have a problem, where the dimension of the original space n exceeds the number of classes, more accurate $n > (L-1)$. Then, by the reasons we gave above, the between-class scatter matrix is singular and one can no longer use (3.5) for the derivation of the LDA kernel. But if we assume the nonsingularity of the total scatter matrix $S_m = S_b + S_w$, which will be true in most

[1] We will remove this restriction later

cases, we can use $S_w^{-1}S_m$ instead of $S_w^{-1}S_b$ in (3.5). By observing $S_w^{-1}S_m = I + S_w^{-1}S_b$ and applying the same strategie as above, it is not hard to show, that

$$\frac{\partial}{\partial K} \log \left| K^T S_w^{-1} S_m K \right| = 0 \tag{3.14}$$

results in (3.13). Thus, we can also apply (3.13) for problems with $n > (L-1)$. Clearly, since the number of nonzero eigenvalues of $S_w^{-1}S_m$ and $S_w^{-1}S_b$ are equal (they have the same eigenvectors!), the dimension m of the transformed space has to be chosen $m \leq (L-1)$.

However, the choice $m = (L-1)$, although optimal in the sense of dimension reduction with preserving relevant information, is only subotpimal in the baysian sense. For $m < (L-1)$ we again have a transformation with linear optimized loss of information.

While the eigenvalues of a matrix are unique, the eigenvectors are not. Therefore we need a boundary condition to obtain a unique solution of (3.13). According to the original approach (where one chose the transformed class covaraince matrices to be unity matrices) one chooses for this purpose the condition

$$K^T S_w K = I. \tag{3.15}$$

So, the LDA transforms the inner-class scatter matrix into a unity matrix. However, it should be mentioned, that, since we dropped the request for equal class covariances matrices, the transformed class covariance matrices will *not* have unity, they will not even be diagonal. Therefore, the on top classifier cannot be a simple distance classifier like in the original approach, but rather a classifier, which takes somehow care of the covariance information. Thus, the LDA on distributions with unequal class covariance matrices looses the advantage of a trivial on top classifier. Still, it gives us the opportunity of a dimension reducing transformation with optimized loss of classification information.

Some further general discussion of LDA can be found in [HTB93]. [HL89] and [HRBA91] discuss the application of LDA in Speech Recognition and give some experimental results in comparison to other feature extraction algorithms. An approach based on different separation criteria (but still using scatter matrices) is given in [Fuk90]. There it is also given the derivation for the case $n > (L-1)$ in a more straight forward manner than it was done here. [Mai94] discusses the results of dimension reduction for the case $m < n < (L-1)$ on the recognition of speech. There it is shown, that a reduction of dimensionality, up to a certain grade but far under $(L-1)$, does not have to result in a loss of recognition perfomance.

Let us summarize the above in an algorithm for computing a LDA feature representation for a L class problem with dimension reduction from n to m:

1. Find the within-class scatter matrix S_w and the between-class scatter matrix S_b by applying (2.24) and (2.25). Both matrices are $n \times n$.
2. Calculate $S_w^{-1}S_b$ and the eigenvalues of this matrix [TW69], [Piz62] and [Ste73].
3. Choose the m largest eigenvalues ($m < n$, $m \leq (L-1)$) and calculate the corresponding eigenvectors under the boundry condition $K^T S_w K = I$.
4. Form the LDA kernel K by taking the eigenvectors as columns of K. K is therefore $m \times n$.
5. Transform an incomming sample X by applying $Y = K^T X$.

Y is then the transformed random vector with dimension m, unity within-class scatter $K^T S_w K$ and diagonal between-class scatter $K^T S_b K$ (which is equal to the eigenvalue matrix Λ).

3.3 An Example

In this place, we want to introduce a small example, to which we will also refer to later when introducing the nonlinear generalization of LDA. The example is derived from the *resource management*

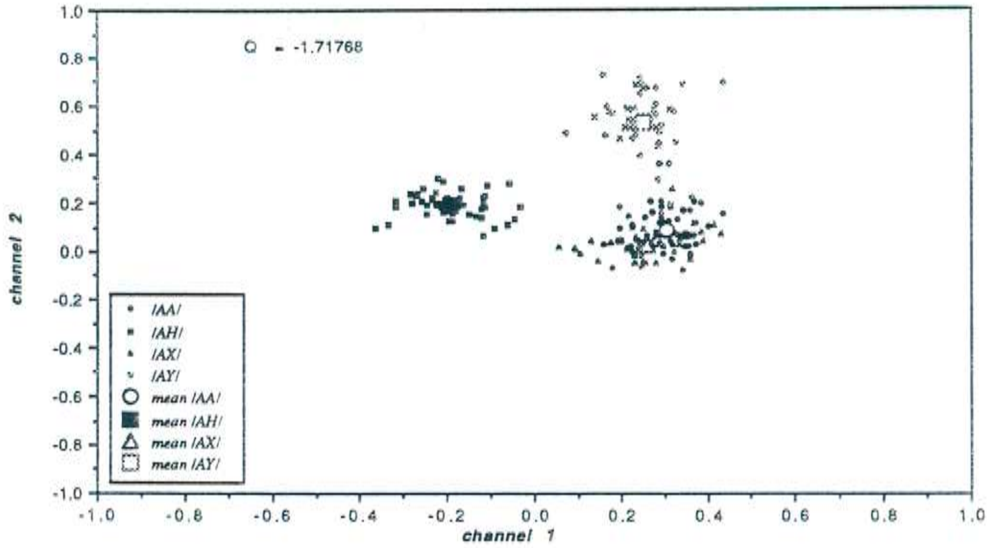


Figure 3.1: Plot of the LDA transformed data.

database [PWFP88], on which we will later carry out the 'real' experiments. The database is fourier transformed; the FFT-vectors have each 16 melscale coefficients. The database is labeled, so the assignment of each sample to one class is known. [1]

The example consists of four classes, each represented by 50 samples, drawn randomly from all samples belonging to the actual class. In order to avoid triviality, the classes are selected to be acousticly somehow similar; we choose four 'a'-sounds, namely /AA/, /AH/, /AX/ and /AY/. Since we are going to use this data futher through this work, it is printed in Appendix B together with its within-class and between-class scatter matrices.

On this data, we carried out a LDA. As expected, the number of nonvanishing eigenvalues of $S_w^{-1}S_b$ is three (since $L = 4$), namely

$$\lambda_1 = 8.701, \quad \lambda_2 = 6.002 \quad \text{and} \quad \lambda_3 = 2.018.$$

The matrix $S_w^{-1}S_b$ and the eigenvectors belonging to the eigenvalues $\lambda_{1\ldots3}$ can be found in Appendix C.

This small dataset was derived to demonstrate how LDA, and later its generalization, works. Therefore, we choose the dimension of the image space $m = 2$. This enables us to make the results graphically visible. The LDA kernel K consists therefore of the two eigenvectors corresponding to λ_1 and λ_2. The seperation measure (3.5) for the transformed data is

$$Q = -1.717.$$

This value may serve for later comparisons (we cannot give Q for the original space since $n > (L-1)$). Figure 2.1 shows a plot of the transformed data. The samples of the classes /AH/ and /AY/ are already very good seperated, while the the samples of classes /AA/ and /AX/ are highly overlapped.

[1] We will talk in detail about the database and the primary transformation further down.

Chapter 4

Nonlinear Discrimiant Analysis (NLDA)

In the previous chapter it was shown how to derive a linear transformation in order to extract dimension reduced features with a minimalized loss of classification information. Now one could ask, whether it is possible to develope a transformation which further minimalizes this loss of classification information.

Of course, since LDA was derived by linear optimization, there is no other linear operator, which minimizes (3.5) better than the LDA kernel K. Subsequently, one has to look for a nonlinear transformation. Unfortunately, since there is no closed nonlinear theory, it is not possible to derive such a transformation in a general way.

4.1 Motivation for Nonlinear Generlization of LDA

Clearly, observing (3.5), it holds that

$$\frac{\partial Q}{\partial |S_w|} = \frac{1}{|S_w|} \qquad \text{and} \qquad \frac{\partial Q}{\partial |S_b|} = -\frac{1}{|S_b|}. \tag{4.1}$$

That means Q increases with $|S_w|$ and decreases with $|S_b|$. Thus, one way to to minimize Q is the simultaneous minimization of $|S_w|$ and maximization of $|S_b|$.

One can show (the way is rather long and avoided here) that for any matrix $A = [a_{ij}]$ with

$$a_{ij} = \sum_{\nu} h_i^{\nu} h_j^{\nu} \tag{4.2}$$

it holds that

$$|A| \sim (h_{\mu}^{\xi})^2. \tag{4.3}$$

The within and between class scatter matrices $|S_w|$ and $|S_b|$ are matrices of the type A. For $|S_w|$, $h_i^{\nu} = x_i^{\nu} - m_i^l$ where x_i^{ν} is the i-th component of the ν-th sample vector and m_i^l is the i-th component of the mean of the class l the sample is assigned to. In the case of $|S_b|$, $h_i^{\nu} = m_i^{\nu} - m_i$, now with ν the class label and m_i the i-th component of the overall mean M_0.

Thus, minimization of Q can be achieved by simultaneous minimization of

$$(h_i^{\nu})^2 = (x_i^{\nu} - m_i^l)^2 \qquad \nu = 1, \ldots, N; \quad l = 1, \ldots, L \tag{4.4}$$

and maximization of

$$(v_i^{\nu})^2 = (m_i^l - m_i)^2 \qquad l = 1, \ldots, L. \tag{4.5}$$

4.2 Minimization of the Within-Class Scatter

The geometric interpretation of minimizing (4.4) can be seen as moving each sample to the corresponding class mean. That is, one tries to concentrate the clouds around the class means.

Mathematically, one is looking for a nonlinear operator $O : \mathbf{R}^n \mapsto \mathbf{R}^m$ with $m \leq n$ which minimizes (4.4) in $\mathbf{R}^m$. Here, $\mathbf{R}^n$ is the original vector space of dimension n and $\mathbf{R}^m$ the space of the transformed vectors with dimension m. The case $m < n$ allows similar to LDA a dimension reduced set of feature vectors.

We need to find the nonlinear operator O. An often used method for finding such an operator under the boundary condition of minimizing a difference like (4.4) is applying a neural network trained by error backpropagation using the mean squared error. The target vectors for the error backpropagation are then the means of the different classes. An introduction into the theory of neural networks and the description of the backpropagation algorithm can be found in [HKP91].

In order to accelerate the convergence and to ensure that the result of the transformation is indeed better than a LDA in the sense of class seperability we demand that the initial state of the network already performs a LDA (we will discuss later how this can be achieved).

In this thesis, we will to introduce three different network types which seem able to perform the desired kind of nonlinear transformation. Two of these will be investigated more in detail. All three approaches work with feed forward networks and with the sigmoid function

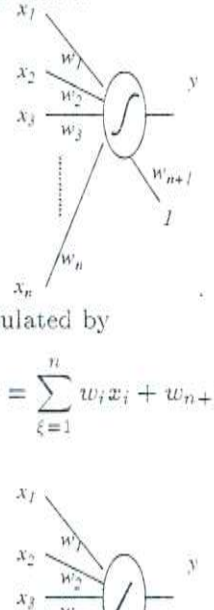
$$s(x) = \frac{1}{1+e^{-x}}$$

as transfer function for nonlinear units.

A nonlinear unit performs the function

$$y = s\left(\sum_{\xi=1}^{n} w_i x_i + w_{n+1}\right) \qquad (4.6)$$

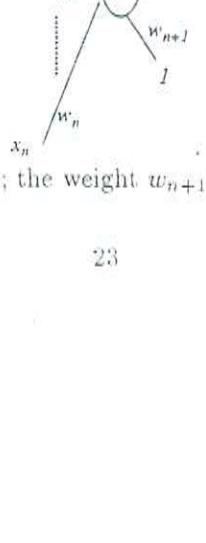
and its graphical symbol throughout this text is



The output y for a linear unit is calculated by

$$y = \sum_{\xi=1}^{n} w_i x_i + w_{n+1}, \qquad (4.7)$$

its graphical symbol is



The coefficients w_i are called weights; the weight w_{n+1} is the bias weight.

4.2.1 Approach 1

Figure 4.1 shows the network to perform the first NLDA approach. The network has two layers. The hidden layer consists of n nonlinear units (again, n is the dimension of the original vector space). Its output layer has m linear units. Therefore its function is

$$Y = A^T \acute{X}, \tag{4.8}$$

where $A : \mathbf{R}^n \mapsto \mathbf{R}^m$ is a linear operator and $\acute{X}$ the output vector of the hidden units with

$$\acute{X} = f(X) \tag{4.9}$$

and $f : \mathbf{R}^n \mapsto \mathbf{R}^n$ is the nonlinear function performed by the hidden units. Thus, the network function can also be written as

$$Y = A^T f(X). \tag{4.10}$$

To satisfy the boundary condition of the initial state LDA, the weights of the hidden units are initially chosen to be

$$w^1_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

and the bias weights are initially chosen to be

$$w^1_{in+1} = -0.5.$$

The reason for this choice is that the data coefficients here range in the interval [0..1]. Therefore this choice of initial weights degenerates the nonlinear function f of the hidden units to the linear unity operator. In other words, for the initial state of the network it holds that $\acute{X} \approx X$.

It is easy to adapt the initialization to data that ranges in different intervals simply by changing the value of the bias weights and introducing a gain into the transfer function.

It is clear that we can achieve the initial LDA state now by chosing the linear operator A to be the LDA kernel $K = [k_{ij}]$. Thus, the initial weights of the output units are

$$w^2_{ij} = k_{ij}.$$

The bias weights of the output layer units are initial set to zero.

Thus, equation (4.10) can be interpreted in the following way: we are looking for a nonlinear transformation which deforms the space in a way that allows a following LDA to perform better than it would be possible in the original space. This nonlinear transformation is performed by the hidden units of the network the linear output units perform the LDA.

We should mention here, that the linear operator A does not really performs a LDA but works in a similar way. The within and between class scatter matrices of the transformed data are not expected to be diagonal [1]. It should be also kept in mind, that the operators A and f are changing continuously while the network is trained.

4.2.2 Approach 2

The second approach to the nonlinear generalization of LDA is very similar to the first one. Its mathematical expression is

$$Y = f(A^T X). \tag{4.11}$$

Figure 4.3 shows the network able to perform this kind of transformation. Again $A : \mathbf{R}^n \mapsto \mathbf{R}^m$ is a linear operator and $f : \mathbf{R}^m \mapsto \mathbf{R}^m$ is a nonlinear operator. However, the order of application has changed. Here, first the linear operator is applied to perform a LDA like transformation and then the space of vectors $\acute{X} = A^T X$ is deformed nonlinearly creating a space of feature vectors Y.

[1] It is not hard to retransfer them again into diagonal matrices by applying a LDA without dimension reduction on top of the network (as we will do in later experiments).

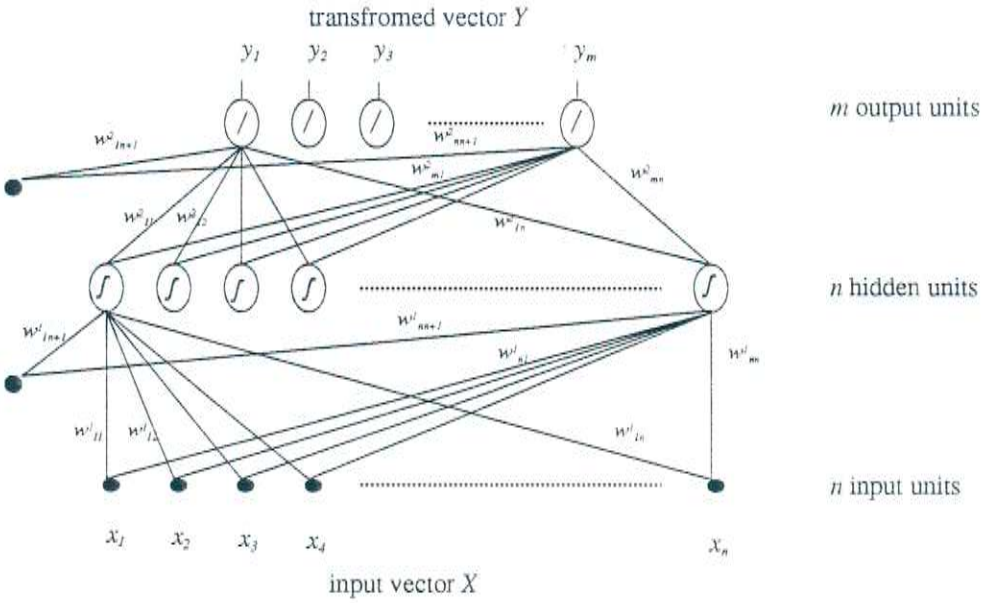


Figure 4.1: Network to perform NLDA $Y = A^T f(X)$.

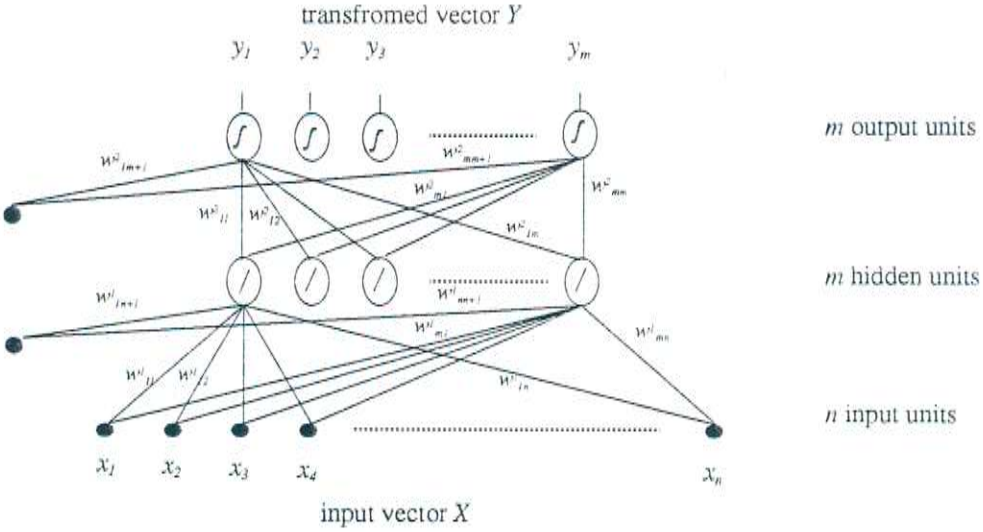


Figure 4.2: Network to perform NLDA $Y = f(A^T X)$.

To satisfy the request of an initial LDA state, suitable initial weights of all units have to be chosen. Again, the weights of the linear units become the coefficients of the LDA kernel $K = [k_{ij}]$ such that

$$w^1_{ij} = k_{ij}.$$

The bias weights of the linear units are set to be the offsets ξ_i so that every coefficient of X is again in [0..1] [1]. The weights of the output units then are again chosen to approximate the unity operator. That is

$$w^2_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

where its biased weights are initialized to be

$$w^2_{in+1} = -0.5.$$

Due to the use of the sigmiod function in the output units, this transformation always creates feature vectors with coefficients in the interval [0..1]. This is an important difference to the former method and (as we shall see later) is also the reason that we do not employ this approach. It is only mentioned here due to completeness.

4.2.3 Approach 3

The third approach to NLDA is essentially different from the former ones. The network to perform it is shown in figure 4.3. The hidden units in this approach 'look' at the input units and feed their responses into the linear units which, in turn, also 'look' at the input units.

The mathematical expression for this approach can be written as

$$Y = A^T X + f(X). \tag{4.12}$$

The main difference from the other two approaches is the superposition of two independent processes as opposed to a two stage process. The linear operator $A : \mathbf{R}^n \mapsto \mathbf{R}^m$ again carries out a LDA-like transformation. The nonlinear operator $f : \mathbf{R}^n \mapsto \mathbf{R}^m$ can be seen as a kind of nonlinear error function. However, the internal representation of the transformation is slightly different: first the original vector space is extended by k dimension, the output from the k hidden units. Then this space is projected on a space with m dimensions by the linear output units.

Suitable choice of the initial weights ensures again initial state LDA. The weights of the hidden units are chosen randomly. The weights of the linear units which are connected to the input units are again set to be the coefficients of the LDA kernel $K = [k_{ij}]$. The weights of the connections from hidden unit to linear unit are set to be zero. Thus,

$$w^2_{ij} = \begin{cases} k_{ij} & \text{if } j \leq m \\ 0 & \text{if } m < j \leq m + k \end{cases}$$

Due to this, for the nonlinear operator f, it holds initially that $f \equiv 0$, and for the linear operator A, it holds that $A = K$. Thus, the initial state is really a LDA. However, in the training process it is useful to accelerate the convergence of training by bootstrapping the weights w^2_{ij} with $m < j \leq m+k$.

In contrast with the other two approaches, here the number k of hidden units can be freely chosen. In the other approaches, the number of hidden units was determined by the demand of the initial LDA state. This is an important advantage of the third approach.

Here it is mentioned again that the scatter matrices of the resulting feature vectors Y are no longer diagonal and that the operators A and f change in the training process.

[1] The dynamic range of the coefficients in each channel $[\xi_i..\xi_i + 1.0]$ can be assured by linear scaling of the LDA kernel with respect to a value m. This changes the within-class scatter matrix from I to mI but does not change any of the LDA's properties. The exact procedure is described later when discussing the experiments with LDA and the resource management database.

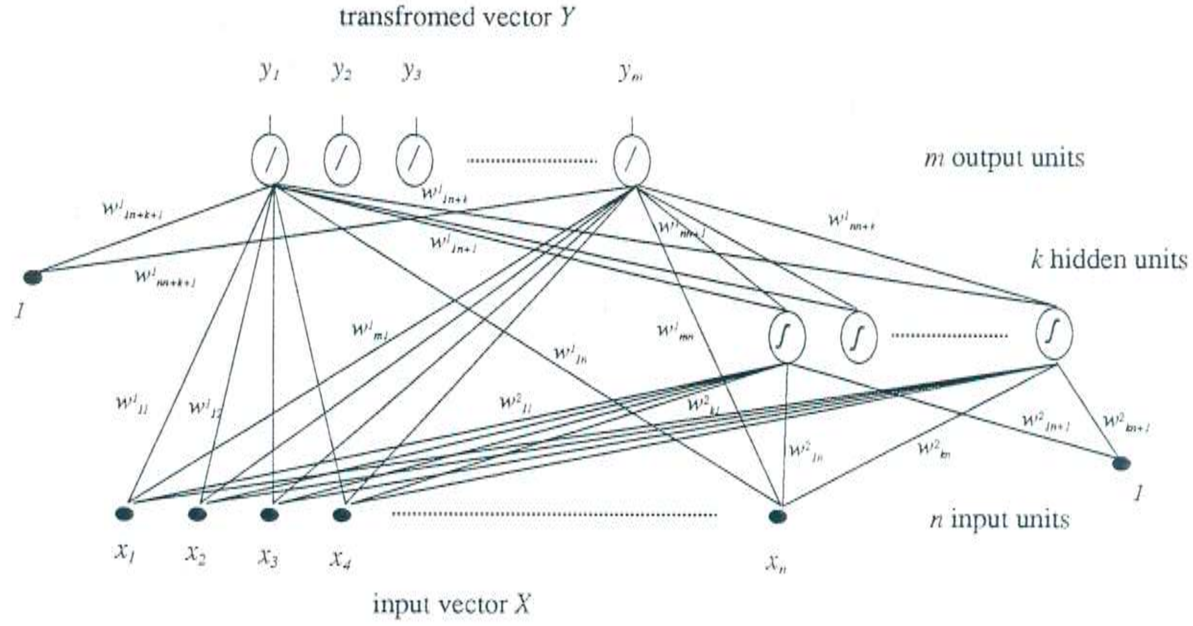


Figure 4.3: Network to perform NLDA $Y = A^T X + f(X)$.

4.3 Maximization of the Between-Class Scatter

For the maximization of (4.5) and therefore of the between-class scatter we use the concept of *moving targets* as intorduced in [Mai94]. The geometrical interpretation is as follows: the target vectors for the backpropagation algorithm are moved away from each other by adding drift vectors to the original targets. The drift vector for one target is calculated by finding directions of freedom to move away from the other targets.

The mathematical expression for this behavior is

$$d_i^\nu = \frac{1}{L} \sum_{k=1,k\neq\nu}^{L} \frac{m_i^\nu - m_i^k}{|M^\nu - M^k|} \qquad 0 < \nu \leq L, \quad 0 < i \leq m. \tag{4.13}$$

Here, $D^\nu = [d_1^\nu, \ldots, d_m^\nu]$ is the drift vector belonging the the ν-th target vector. The vectors $M^\nu = [m_1^\nu, \ldots, m_m^\nu]$ are the L class centroids in the transfromed space. The drift vector D^ν can be also written as

$$D^\nu = \frac{1}{L} \sum_{k=1,k\neq\nu}^{L} \frac{M^\nu - M^k}{|M^\nu - M^k|} \qquad 0 < \nu \leq L. \tag{4.14}$$

Therefore, each drift vector D^ν is the sum over the $L - 1$ unity vectors in the directions of the difference vectors between the corresponding centroid M^ν and the other centroids. The unity of the direction vectors is assured by the nominator in the sum in (4.13) and (4.14).

It seems useful to have a larger drift for classes closer to each other. That can be achieved by introducing a weighting function $b(|M^\nu - M^k|)$ into the sum of (4.14). To preserve the metric of the space, this weighting function has to satisify the following conditions:

- $b(|M^\nu - M^k|)$ has to be a monotonously decreasing function.
- From $|M^\nu - M^k| < |M^\nu - M^l|$, it has to follow that

$$\left| M^\nu - M^k + b(|M^\nu - M^k|)\frac{M^k - M^\nu}{|M^\nu - M^k|} \right| < \left| M^\nu - M^l + b(|M^\nu - M^l|)\frac{M^l - M^\nu}{|M^\nu - M^l|} \right|.$$

This prevents a target centroid for a centroid closer to M^ν from becoming more distant than the target centroid for a centroid further apart from M^ν

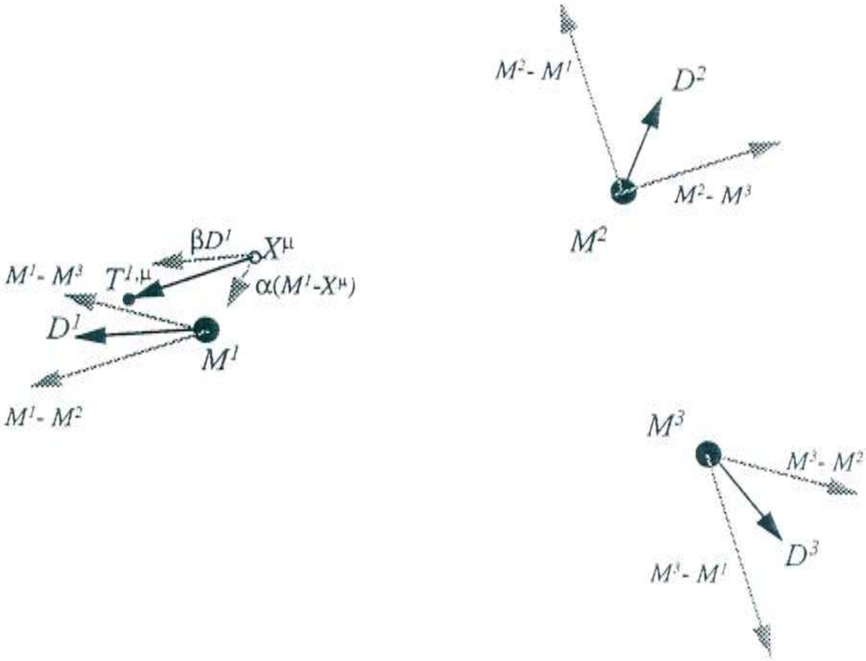


Figure 4.4: Demonstration of the concept of moving targets.

A little geometric and analytic considaration of the problems leads to a family of functions which fulfill these conditions [Mai94], namely

$$b(|M^\nu - M^k|) = \frac{1}{(1 + |M^\nu - M^k|)^\gamma} \qquad \gamma > 0. \tag{4.15}$$

The parameter γ in (4.15) is a gain which controls how strongly smaller distances are prefered in the weighting. Equation (4.14) is therefore updated to

$$D^\nu = \frac{1}{L} \sum_{k=1,k\neq\nu}^{L} \frac{1}{(1 + |M^\nu - M^k|)^\gamma} \frac{M^\nu - M^k}{|M^\nu - M^k|} \qquad 0 < \nu \leq L. \tag{4.16}$$

The actual target for the μ-th sample belonging to class ν, which is given to the backpropagation algorithm, is thus

$$T^{\nu,\mu} = \alpha M^\nu + (1 - \alpha) Y^\mu + \beta D^\nu \qquad 0 \leq \alpha \leq 1, \quad 0 \leq \beta \leq L. \tag{4.17}$$

In (4.17) Y^μ is the net output for the μ-th sample vector X^μ. The parameter β allows an overall weighting of the drift. The restriction of β to the interval $[0..L]$ follows from the conditions and from the special type of the weighting function (4.15) [Mai94]. However, β will usually be much smaller than L. A good way to find intervals for β is to compare the magnitude of the centroids with the magnitude of the drift vectors. The parameter α controls how strong the targets for each sample are scattered around the target centroid $M^\nu + \beta D^\nu$. In the one extreme case, $\alpha = 1$, the target vector for every sample of one class is the same : $M^\nu + \beta D^\nu$. Thus, multimodal distributions are eventually transformed into unimodal distributions. In the other extreme, $\alpha = 0$, the target vector for each sample vector is the sample vector itself shiftet by the weigthed drift vector. This should give the possibility to preserve multimodality in the distributions.

From the above follows that the process of maximization of the between-class scatter is a parametric process controlled by the parameter triple (β, α, γ).

Figure 4.4 makes the whole procedure clear for a simple example.

From (4.13) and (4.17) is easy to see that this procedure maximizes the distance between the class means (4.5). Therefore the between-class scatter is also maximized.

For this procedure of maximizing the between-class scatter, the above mentioned reason for discarding approach 2 (4.11) to NLDA holds: That approach restricts the range of the components of the output vectors. Using the concept of moving targets, it is not desirable to do this since one is interested in allowing the centroids to drift freely. This is achieved by using linear units in the output layer of the network. However, [Mai94] proposes to solve this problem for nonlinear output units by projecting the drift vectors onto the surface of a sphere in the restricted output space. While this approach solves the aforementioned problem, it seems to be dangerous because it can result in a drastic change of the class centroids neighbourhood relations. The metric of the original space could therefore not be preserved in the output space.

4.4 Combination of the Simultaneous Optimization of the Scatters to a NLDA Algorithm

The processses for minimizing the within-class scatter and maximizing the between-class scatter as described in the above sections can be melted together into an algorithm for a Nonlinear Discriminant Analysis with dimension reduction from n to m dimensions performed by a neural network as follows (the algorithm's parameter l and k are explained after the algorithm):

1. Choose a neural network corresponding to the different types of NLDA (4.10) or (4.12).
2. Choose the number of the network input units to be n and the number of output units to be m.
3. If the NLDA approach is (4.12), choose the number of hidden units.
4. Choose a parameter triple (β, α, γ).
5. Initialize the network to achieve an initial LDA state.
6. Pass all training data through the network and calculate the class centroids in the output space, the scatter matrices in the output space, and the criterion (3.5).
7. If there is no significant change in the value of (3.5) compared to the previous iteration or if this is the k-th iteration go to step 11.
8. If the number of iterations is a multiple of l, store the class centroids of the output space and calculate the drift vectors for each class by using (4.16).
9. For each sample:
   (a) Calculate target for error backpropagation using (4.17) and the cached class centroids.
   (b) Perform error backpropagation using mean squared error as error measure.
   (c) Update the network weights.
10. Go to step 6.
11. Calculate a additional LDA without dimension reduction on the output data to return the scatter matrices into diagonal matrices (this step can be skipped if diagonal matrices are not needed).
12. STOP.

This algorithm has two additional parameters: the number of iterations k and the parameter l. l determines after how many iterations the class centroid and drift vectors for the target computation are updated. This gives the possibility to train more than one iteration against the same targets and facilitates convergence (a too frequent update of centroids and drifts can result in runaway of the targets).

4.5 Other Approaches to NLDA

One could think of a number of other approaches to achieve a NLDA that we do not consider here. In [WL90], for example, it is shown that a neural network with a linear output layer but nonlinear hidden layer trained as a one-from-N classifier minimizes a criterion similar to (3.5) in the space spanned by the hidden units. Therefore, the space spanned by the hidden units is a NLDA representation of the original space.

Another possibility is the nonlinear principal component analysis performed by the hidden layer of an auto-assoziative trained network. Again the space spanned by the hidden units can be seen as a NLDA representation of the input space.

A way to achieve a generalization of the LDA idea which takes care of the classifier following the feature extraction is shown in [HTB93]. The name of the resulting method is *Flexible Discriminant Analysis* (FDA). This approach works without neural networks in an analytic way.

4.6 Demonstration of the NLDA Abilities with the Test Database

Again we use the small database introduced in chapter 3 to shown whether and how the NLDA approaches (4.10) and (4.12) work. Since there are only 250 samples in this database many iterations are needed to train the networks. For each experiment, we have made 500 iterations over the whole database. Thus the network has seen 125000 samples in the training process. The samples have been laid on the network input units in a random fashion. The centroids and drift vectors for the target calculation have been updated every 50 iterations for the approach (4.10) and every 25 iterations for the approach (4.12). For each network type, 3 experiments for 3 different values of γ were made. Here γ is 1, 2 or 3. Since these experiments are only for demonstration purposes, the parameters α and β have not been varied but have been chosen. The parameter α in all experiments has been assigned the value 1.0. The paramter β has been chosen to assure that the magnitude of the drift vectors is approximately one third of the magnitude of the centroids in the output space.

In all experiments, the dimension of the original space was 16 (16 input units) and the dimension of the output space was 2 (2 output units). The initial LDA state has been achieved by using the LDA kernel K derived from the data (Appendix C).

The learning rate for the backpropagation was $lr = 0.008$ in the experiments corresponding to (4.10) and $lr = 0.004$ in the experiments corresponding to (4.12). The backpropagation momentum was $m = 0.9$ in all experiments.

Before describing the results of the different experiments, one common feature of all experiments should be mentioned. All experiments showed a convergence behaviour as qualitativly shown in figure 4.5. Here the value of N_c is a critical number of iterations that need to be carried out before

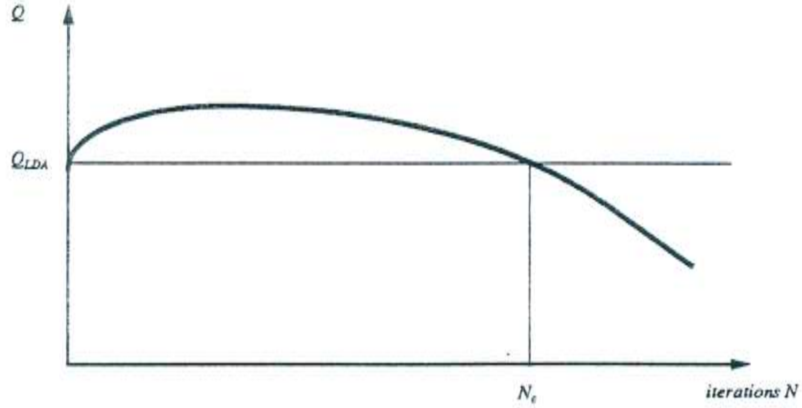


Figure 4.5: Convergence behaviour of a NLDA network.

an improvement in the criterion value for the class seperability is achieved. This behaviour is due to the strong initial deformation of the space before convergence starts. In all experiments, the value of N_c was approximately 100.

However, real databases usually have many more samples than our small example. Thus, this initial convergence behaviour is not crucial in real experiments as will be seen in chapter 5 (N_c is smaller than 1).

4.6.1 NLDA $Y = A^T f(X)$

For the three experiments with this network type, the parameters and results are listed in the following table:

	β	α	γ	iterations	Q	
LDA	-	-	-	-	-1.717	figure 3.1
NLDA	2.0	1.0	1	500	-3.200	figure 4.6
NLDA	3.2	1.0	2	500	-3.789	figure 4.8
NLDA	4.0	1.0	3	500	-3.425	figure 4.10

Obviously all three experiments minimize Q very strongly (since Q is a logarithmic value). The plots of the output data for the experiments can be found in the figures 4.6, 4.8 and 4.10. Indeed, class seperability has improved strongly, especially with respect to the classes /AA/ and /AX/. While in the LDA case (cp. figure 3.1) both these classes were strongly overlapped, they are seperated well in all three NLDA experiments.

Furthermore, with increasing value of γ, the drift for closer classes becomes larger and subsequently their seperability in the output space increases while the classes which were initially well seperated drift less.

One feature of the concept of moving targets can be observed in all three experiments: due to the movement of the targets further and further away there is a tendency in each class to develop a tail. A way to get rid of this tail is to train the network after the 500 iterations by some additional iterations with the centroids of the classes as targets. Thus, the classes are finally concentrated around their means. This procedure corresponds to the parameter set (0.0, 1.0, γ) where the value of γ is unimportant because it influences only the drift vectors and the weighting factor for the drift vectors, β, is zero.

This procedure was applied and the results are plotted in the figures 4.7, 4.9 and 4.11. Since this procedure is equal to a further minimization of the within-class scatter, the value of Q decreases a little more:

moving target parameter (β, α, γ)	Q after 500 iterations	Q after additional 100 iterations with centroids as targets
(2.0, 1.0, 1)	-3.200	-3.990
(3.2, 1.0, 2)	-3.789	-4.153
(4.0, 1.0, 3)	-3.425	-3.736

Altogether this approach to NLDA seems promising.

4.6.2 NLDA $Y = A^T X + f(X)$

The three experiments for this network type were carried out in the same way as the experiments described above. The number of hidden units in these three experiments is 10. That is less than the 16 hidden units in the network for the former experiments. Despite the smaller number of hidden units the results of the experiments are good. The parameters and results are listed below:

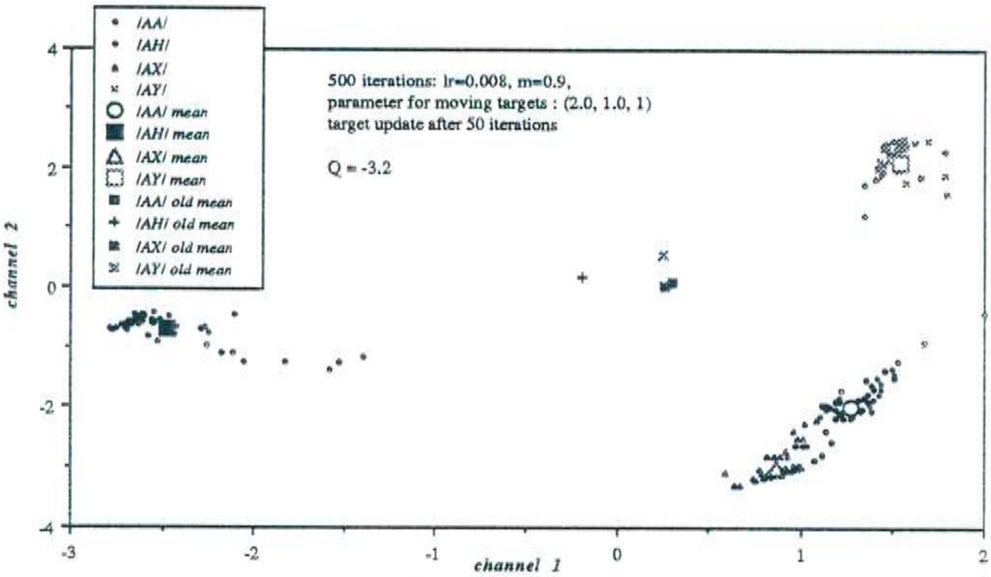


Figure 4.6: Data after NLDA $Y = A^T f(X)$ with 500 iterations, moving target parameter (2.0, 1.0, 1), and centroid and drift vector update every 50 iterations.

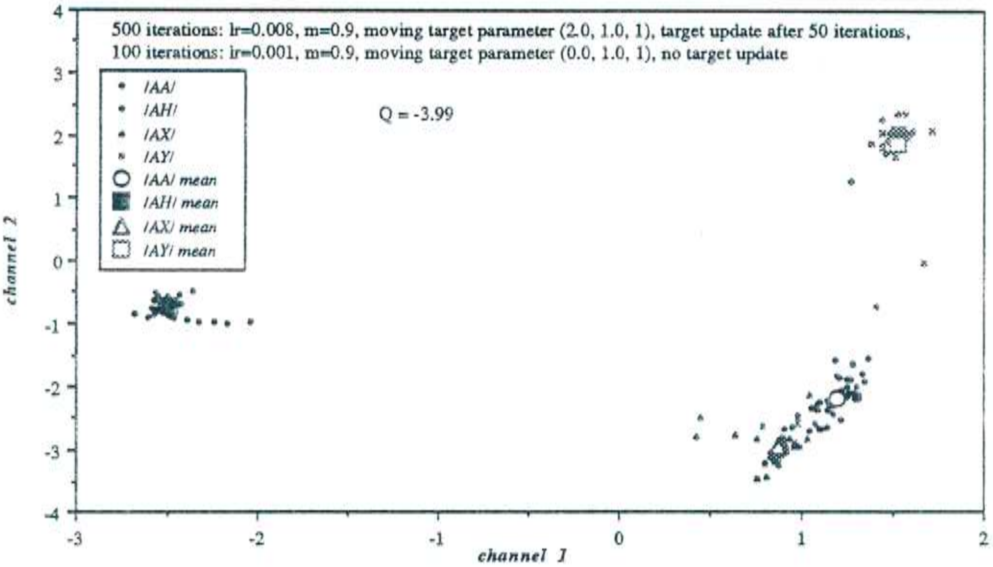


Figure 4.7: Data after NLDA $Y = A^T f(X)$ with 500 iterations, moving target parameter (2.0, 1.0, 1), centroid and drift vector update every 50 iterations, and additional 100 iterations with centroids as targets.

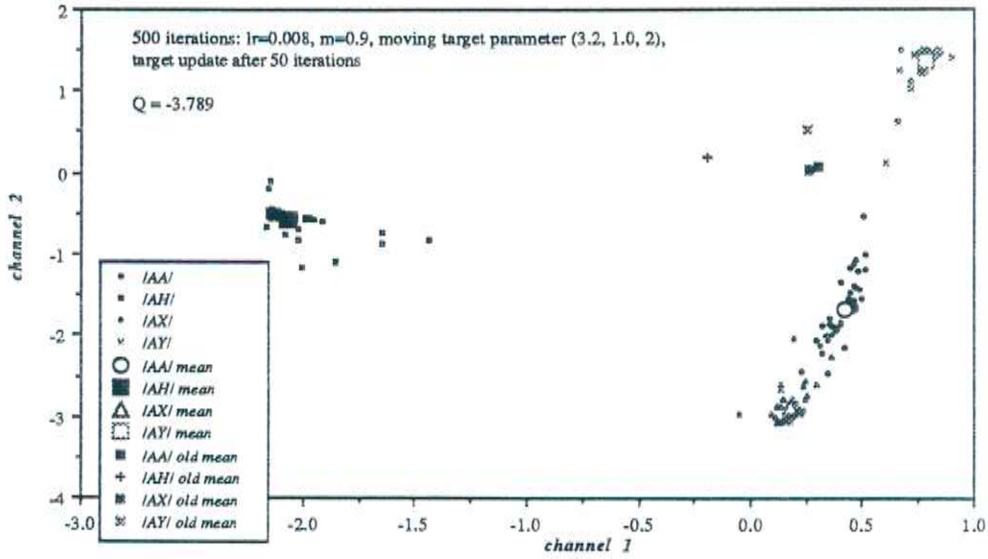


Figure 4.8: Data after NLDA $Y = A^T f(X)$ with 500 iterations, moving target parameter (3.2, 1.0, 2), and centroid and drift vector update every 50 iterations.

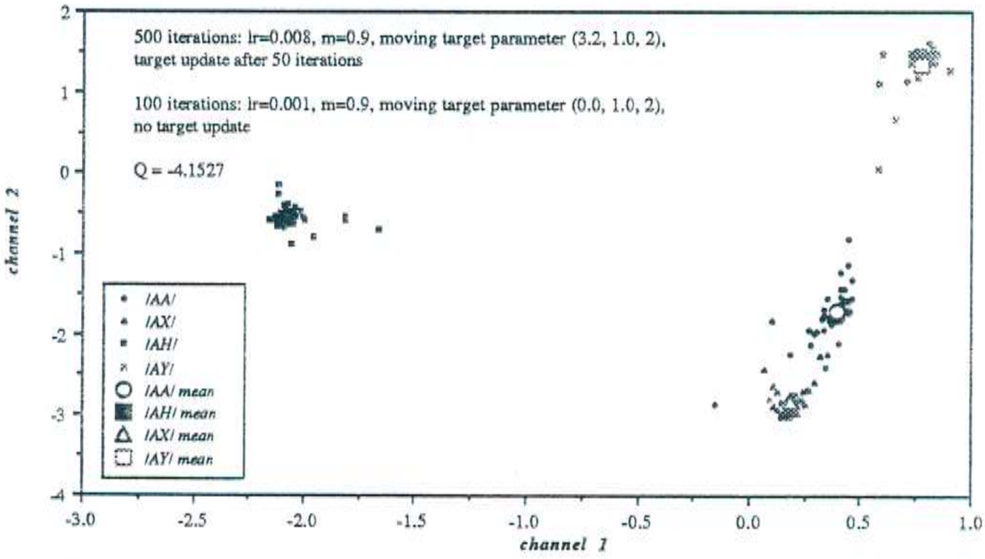


Figure 4.9: Data after NLDA $Y = A^T f(X)$ with 500 iterations, moving target parameter (3.2, 1.0, 2), centroid and drift vector update every 50 iterations, and additional 100 iterations with centroids as targets.

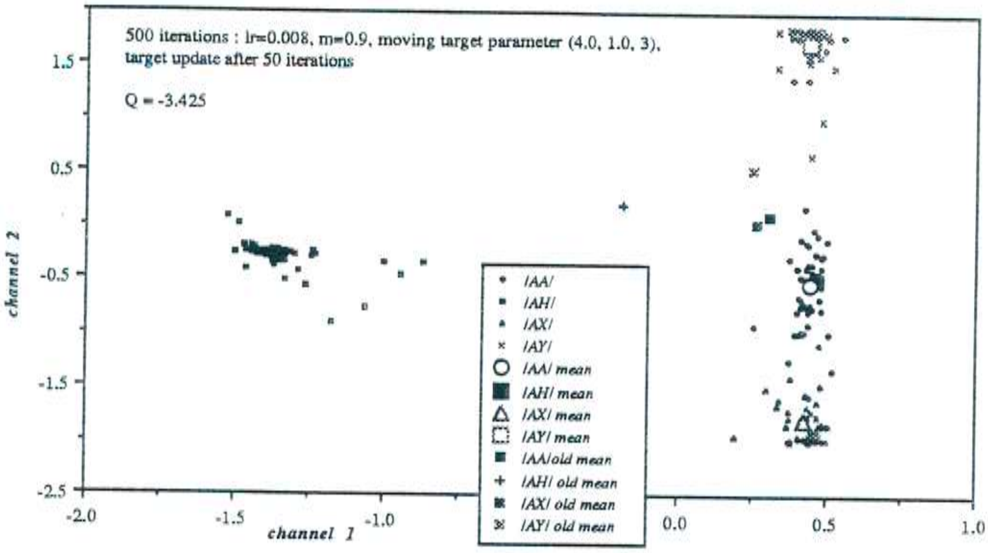


Figure 4.10: Data after NLDA $Y = A^T f(X)$ with 500 iterations, moving target parameter (4.0, 1.0, 3), and centroid and drift vector update every 50 iterations.

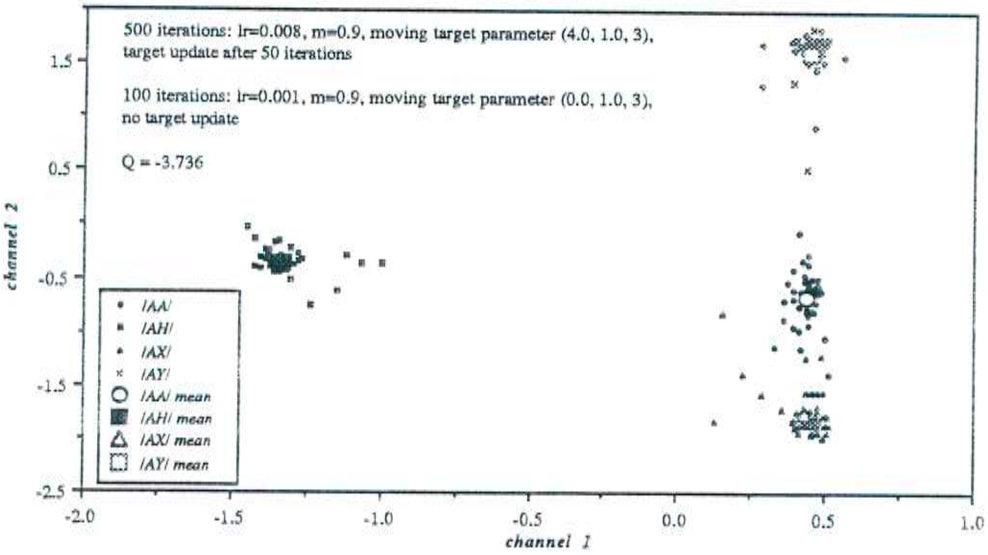


Figure 4.11: Data after NLDA $Y = A^T f(X)$ with 500 iterations, moving target parameter (4.0, 1.0, 3), centroid and drift vector update every 50 iterations, and additional 100 iterations with centroids as targets.

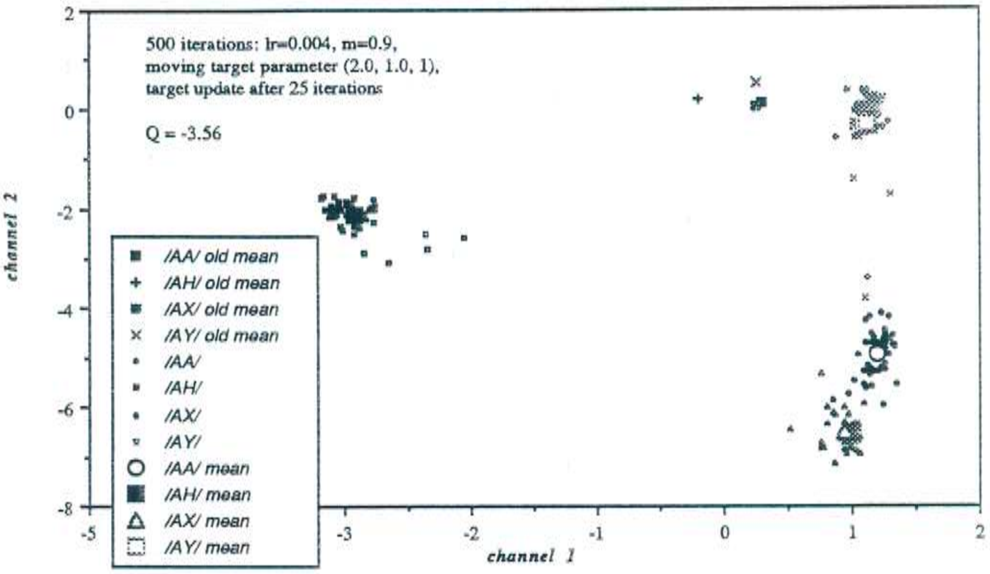


Figure 4.12: Data after NLDA $Y = A^T X + f(X)$ with 500 iterations, moving target parameter (2.0, 1.0, 1), and centroid and drift vector update every 25 iterations.

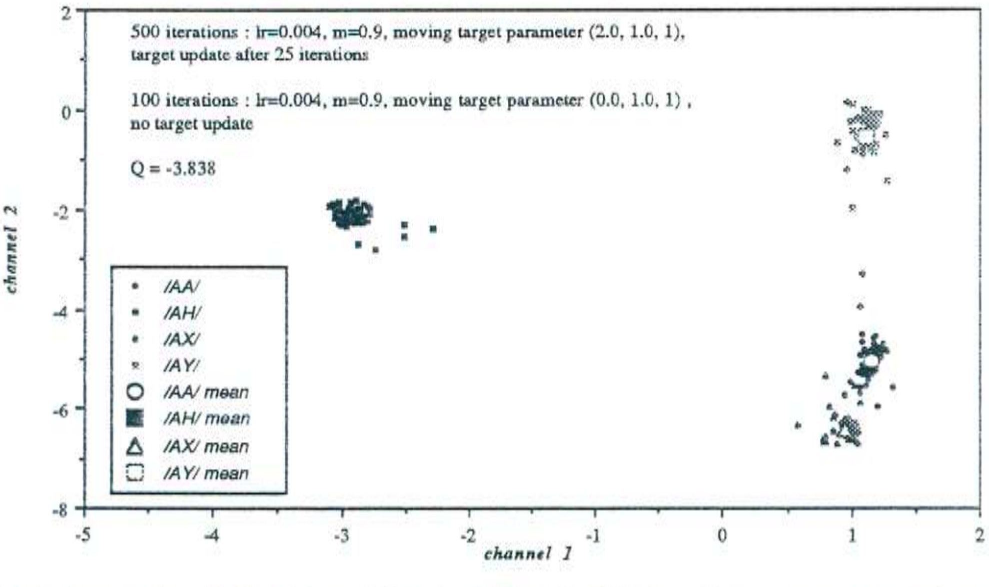


Figure 4.13: Data after NLDA $Y = A^T X + f(X)$ with 500 iterations, moving target parameter (2.0, 1.0, 1), centroid and drift vector update every 25 iterations, and additional 100 iterations with centroids as targets.

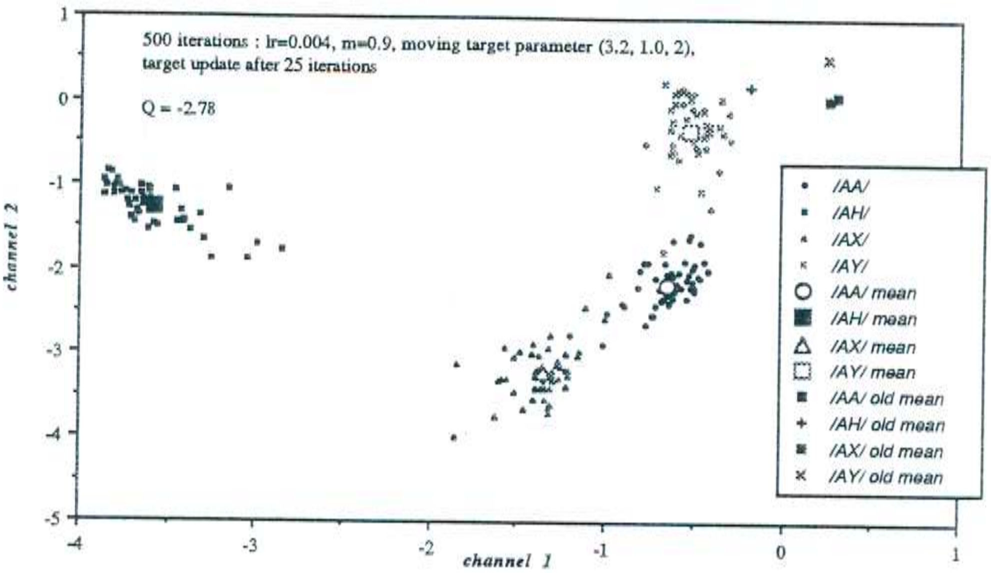


Figure 4.14: Data after NLDA $Y = A^T X + f(X)$ with 500 iterations, moving target parameter (3.2, 1.0, 2), and centroid and drift vector update every 25 iterations.

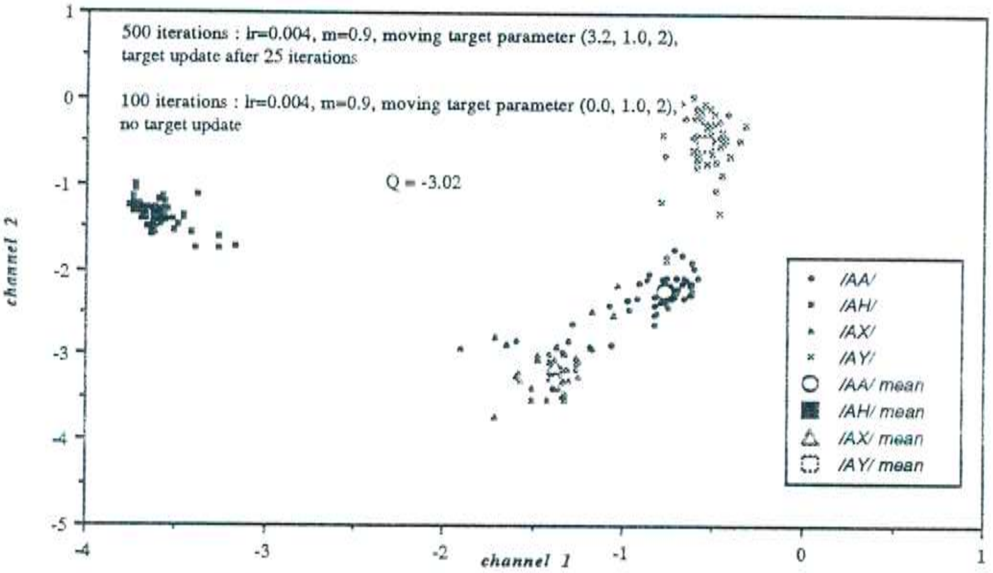


Figure 4.15: Data after NLDA $Y = A^T X + f(X)$ with 500 iterations, moving target parameter (3.2, 1.0, 2), centroid and drift vector update every 25 iterations, and additional 100 iterations with centroids as targets.

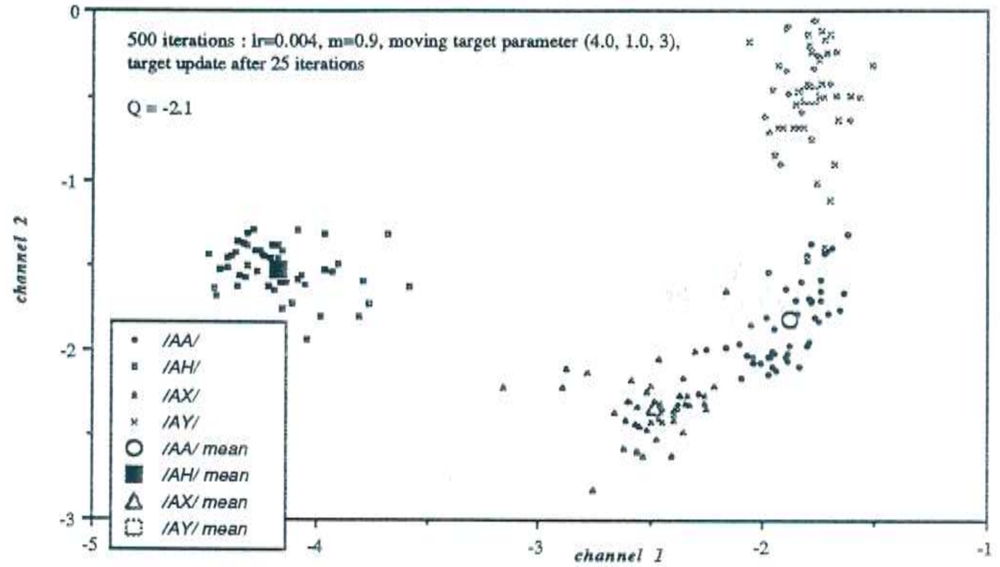


Figure 4.16: Data after NLDA $Y = A^T X + f(X)$ with 500 iterations, moving target parameter (4.0, 1.0, 3), and centroid and drift vector update every 25 iterations.

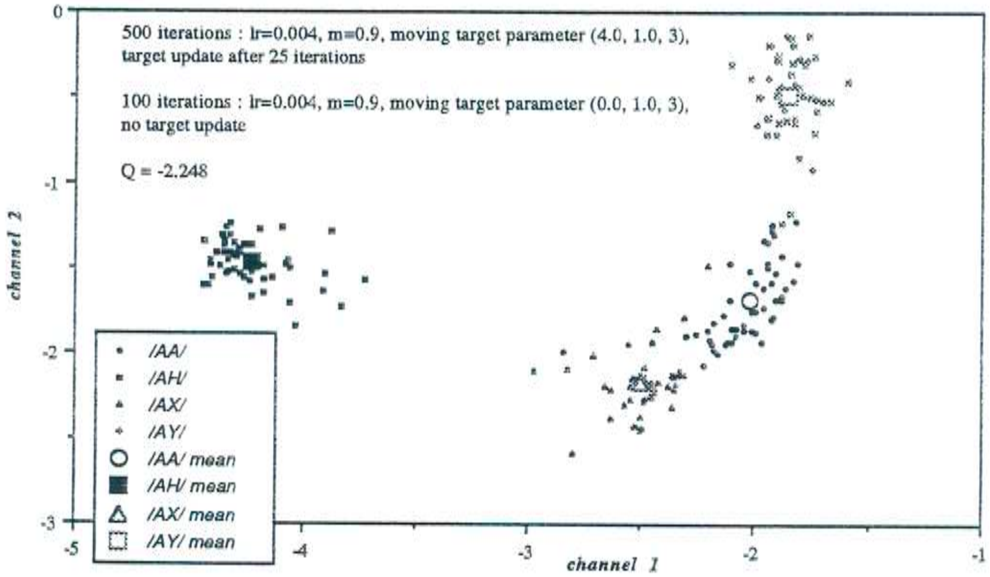


Figure 4.17: Data after NLDA $Y = A^T X + f(X)$ with 500 iterations, moving target parameter (4.0, 1.0, 3), centroid and drift vector update every 25 iterations, and additional 100 iterations with centroids as targets.

	β	α	γ	hidden units	iterations	Q	
LDA	-	-	-	-	-	-1.717	figure 3.1
NLDA	2.0	1.0	1	10	500	-3.56	figure 4.6
NLDA	3.2	1.0	2	10	500	-2.78	figure 4.8
NLDA	4.0	1.0	3	10	500	-2.10	figure 4.10

Here it is seen, that with increasing values of γ, the value of Q is not as strongly minimized. A look to the plots of the data in the figures 4.12, 4.14 and 4.16 shows that this is because concentration within the classes decreases as γ increases (while the class centroids are still far appart). Nevertheless, in all three examples the seperability of the classes is good. Again, especially the seperability between /AA/ and /AX/ has increased. These classes are now certainly seperable, which was not the case after LDA.

Again we made additional 100 iterations with the centroids as target to concentrate the classes further. Figures 4.13, 4.15 and 4.17 show the plots of the data in the output space. Due to the stronger concentration within the classes, Q decreases. The following table shows this in an overview:

moving target parameter (β, α, γ)	Q after 500 iterations	Q after additional 100 iterations with centroids as targets
(2.0, 1.0, 1)	-3.56	-3.84
(3.2, 1.0, 2)	-2.78	-3.02
(4.0, 1.0, 3)	-2.10	-2.25

Summarizing, it can be said that both approaches to the Nonlinear Discriminant Analysis are very promising. The class seperability in our example was strongly increased in all experiments. But it still needs to be shown that this way of minimizing the value of Q and increasing the class seperability also leads to an improvement in the actual recognition performance of a speech recognizer.

Chapter 5

Experiments on the Resource Management Database with a Continuous Speech Recognizer

In the previous chapter we have shown how to improve class seperability by applying a nonlinear transformation. But it still needs to be verify, that the improvement in seperability leads also to an improvement in recognition performance.

In this chapter, we want to describe some NLDA experiments we performed with a continuous speech recognizer on a speaker independent large vocabulary database.

5.1 The Database

The database we used for the experiments is the widely used *Resource Management Database* described in detail in [PWFP88]. This database is a speaker independent continuous speech database and ranges in the 1000 word domain. It consists of 72 male and female speakers for training, 37 male and female speakers for development (usually splitted into additional training material and test material) and 37 male and female speakers for evaluation and testing. The speakers are drawn from different american dialects. Each speaker is represented by 40 sentences.

The speakers were simultaneously recorded with low background noise and digitized at $20kHz$ sampling frequency. The digitized data was then downsampled to $16kHz$.

5.1.1 Primary Transformation

On the digitized speech data a 256-point fast fourier transform was performed (corresponding to a frame length of $16ms$). The frames have been overlapped by $6ms$ and weighted by a Hamming window. That leads to a effective frame rate of $10ms$. Only the amplitude spectrum was further used.

This whole procedure results in data vectors with 128 coefficients. To make a first dimensionality reduction of the data, the FFT vectors have been mapped onto 16 *melscale coefficients*. This is a heuristic dimensionality reduction which is justified by the analogy to the human auditory system. There, the bandwith of the cochlea filters is constant at low frequencies and increases after a certain point with frequency corresponding to a constant realtive bandwith. Table 5.1 shows the mapping rule from the spectral coefficients to the melscale coefficients after [WY81]. In the table, m_i corresponds to the i-th melscale coefficient and s_i corresponds to the i-th spectral coefficient

From the melscale values m_i, a dB scale representation l was calculated by applying

$$l_i = 10 \log m_i .$$

coefficient index	mapping rule	coresponding frequency interval/ [Hz]
1	$m_1 = s_1 + s_2 + s_3/2$	0 ... 187.5
2	$m_2 = s_3/2 + s_4 + \cdots + s_6 + s_7/2$	187.5 ... 437.5
3	$m_3 = s_7/2 + s_8 + \cdots + s_{10} + s_{11}/2$	437.5 ... 687.5
4	$m_4 = s_{11}/2 + s_{12} + \cdots + s_{14} + s_{15}/2$	687.5 ... 937.5
5	$m_5 = s_{15}/2 + s_{16} + \cdots + s_{18} + s_{19}/2$	937.5 ... 1187.5
6	$m_6 = s_{19}/2 + s_{20} + \cdots + s_{22} + s_{23}/2$	1187.5 ... 1437.5
7	$m_7 = s_{23}/2 + s_{24} + \cdots + s_{26} + s_{27}/2$	1437.5 ... 1687.5
8	$m_8 = s_{27}/2 + s_{28} + \cdots + s_{30} + s_{31}/2$	1687.5 ... 1937.5
9	$m_9 = s_{31}/2 + s_{32} + \cdots + s_{35} + s_{36}/2$	1937.5 ... 2250
10	$m_{10} = s_{36}/2 + s_{37} + \cdots + s_{41} + s_{42}/2$	2250 ... 2625
11	$m_{11} = s_{42}/2 + s_{43} + \cdots + s_{48} + s_{49}/2$	2625 ... 3062.5
12	$m_{12} = s_{49}/2 + s_{50} + \cdots + s_{57} + s_{58}/2$	3062.5 ... 3625
13	$m_{13} = s_{58}/2 + s_{59} + \cdots + s_{68} + s_{69}/2$	3625 ... 4312.5
14	$m_{14} = s_{69}/2 + s_{70} + \cdots + s_{81} + s_{82}/2$	4312.5 ... 5125
15	$m_{15} = s_{82}/2 + s_{83} + \cdots + s_{97} + s_{98}/2$	5125 ... 6125
16	$m_{16} = s_{98}/2 + s_{99} + \cdots + s_{116} + s_{117}$	6125 ... 7312

Table 5.1: Mapping from spectral coefficients to melscale coefficients.

This leads to preprimary features of dimensionality 16. Finally, the coefficients of the preprimary features for each sentence are lineary scaled to be in the intervall $[0 \ldots 1.0]$. This gives the possibility to store each in a 1-byte form by mapping the floating point values onto chars.

The primary features have been derived from the preprimary features by adding dynamic information in form of delta coefficients

$$d_i(k) = \begin{cases} l_i(1) - l_i(k+2) & 1 \leq k \leq 2 \\ l_i(k-2) - l_i(N) & N-1 \leq k \leq N \\ l_i(k-2) - l_i(k+2) & 2 < k < N-1 \end{cases} . \tag{5.1}$$

In (5.1) k means the actual frame number, N the total number of frames in the sentence. Thus, the primary feature vectors $x(k) = [l_1(k), \ldots, l_{16}(k), d_1(k), \ldots, d_{16}(k)]^T$ are of dimensionality 32. The delta coefficients are not stored but created online.

5.1.2 Class Assignment for training data

The database is not yet labeled, i.e., the assignment of a training data frame to a class is still unknown. Labels for all frames were created by handlabeling and an automatic postlabeling. For the postlabeling, the handlabeled data was used to train a HMM-recognizer. After the training the data was postlabeled by finding the Veterbi best path and reassigning the data. This procedure can be done iteratively.

5.2 The Recognizer

The recognizer used is part of the JANUS speech-to-specch translation system developed at the Carnegie Mellon University in Pittsburgh and the Karlsruhe University. It is a system designed for the recognition of continuous speech. In the following, the recognizer, a semicontinuous HMM-recognizer, is described in principle. A detailed description is given in [Wai91] and [Wai92].

5.2.1 Semicontinuous Hidden Markov Model

Hidden Markov Models (HMM) give one the possibility to find the probability that an observation was emitted by a source. Here, the source consists of a succession of states with different transition probabilities. (An observation is a succession of events.) Figure 5.1 shows a typical HMM. The

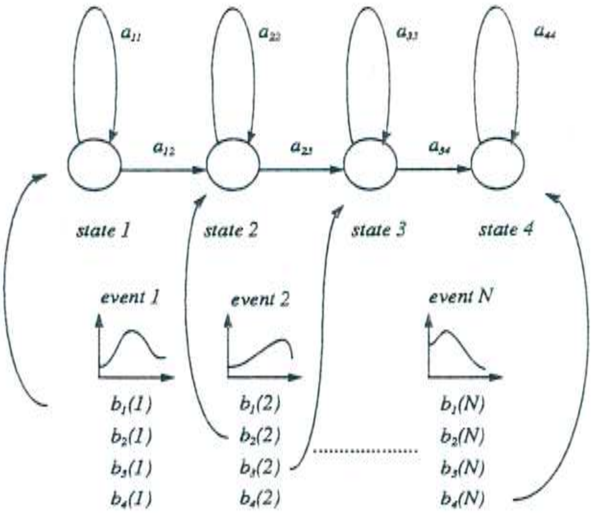


Figure 5.1: Hidden Markov Model with 4 states.

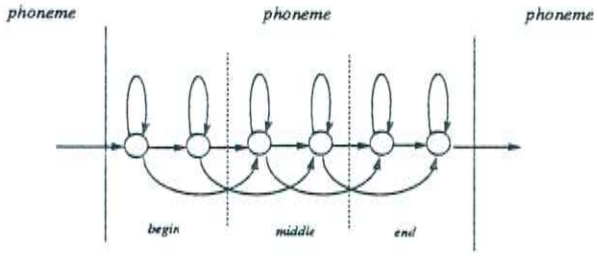


Figure 5.2: Phoneme HMM.

a_{ij} in figure 5.1 are the state transition probabilities from state i to state j. In order to find the probability for the emission of the observation by the model, one needs the probabilities for the emission of each event by each state. These are the emission probabilities $b_i(k)$, i.e., the probability that state i emits event k. Detailed description of HMM's and training algorithms to estimates the probabilities a_{ij} and $b_i(k)$ can be found in [Rab89] and [HJ89].

In speech recognition systems, one chooses the states to model the phonetic transcription of the words. Therefore, only transitions from left to right are possible. Since several subsequent events can belong to the same state, every state has a self loop.

In this particular recognizer, one further splits a phoneme into three subphonemes. This allows one to model the begining phase, the ending phase and the static phase seperately. Each subphoneme is modeled by 2 states. It is possible to jump over a subsequent state. Figure 5.2 shows the HMM for a phoneme. The transition probabilities at this level are chosen initially and remain fixed.

In a *discrete* HMM, the events are the vectors of a codebook C and the emission probabilities $b_i(k)$ are the discrete relative frequencies for the codebook vectors.

In a *semicontinuous* HMM [HJ89], one models the emission probabilities $b_i(k)$ by a discrete distribution $P(c_i \mid s)$ over a codebook $C = [c_1, \ldots, c_K]$ with K codebook vectors c_i and a continuous distribution $f(y \mid c_i)$ over each vector c_i. Therefore, the discrete emission probabilities $b_i(k)$ become continuous distributions

$$f(y \mid s) = \sum_{i=1}^{K} f(y \mid c_i) P(c_i \mid s). \tag{5.2}$$

In (5.2), s is the actual state of the HMM and y is a feature vector (for example, a primary feature vector $x(k)$). Given an observation $\mathcal{O} : x(1), \ldots, x(k), \ldots, x(N)$ the probability that a HMM M

created this observation is

$$P(\mathcal{O} \mid M) = \sum_{\mathcal{S}} P(\mathcal{O} \mid M, S_i) P(S_i \mid M), \tag{5.3}$$

where $\mathcal{S} = \{S_1, \ldots, S_\mu\}$ stands for the set of all possible state sequences. One is only interested in the largest component of this sum which corresponds to the most likely state sequence S_ν. One can find S_ν by applying the Veterbi algorithm [Rab89]. The probability that the HMM M with the state sequence S_ν has produced the observation $\mathcal{O}$ is

$$P(\mathcal{O} \mid M, S_\nu) P(S_i \mid M) = \prod_{k=1}^{N} a_{s_k s_{k+1}} f(y(k) \mid s). \tag{5.4}$$

In order to avoid numerical problems (due to the number of multiplications in (5.4)), one often applies a logarithm on (5.4). The resulting log probability is often called *score*

$$score(\mathcal{O} \mid M) = -\log\left(P(\mathcal{O} \mid M, S_\nu) P(S_i \mid M)\right) = -\sum_{k=1}^{N} \log a_{s_k s_{k+1}} - \sum_{k=1}^{N} \log f(y(k) \mid s). \tag{5.5}$$

The recognizer of the JANUS system takes the following simplifications:

- For the continuous distributions $f(y \mid c_i)$, normal distributions $N_y(c_i, \Sigma_i)$ are chosen. This is justified by the assumption that the class distributions are multimodal normal distributions and each codebook vector estimates the mean of a partial distribution.
- For the covaraince matrices Σ_i for the normal distributions $N_y(c_i, \Sigma_i)$ a unitary matrix is assumed. This assumption transforms the Mahalanobis distance in the exponent of (2.27) into the euclidean distance.
- Instead of the sum (5.2), only its largest component is considered. This is equivalent to searching the codebook vector with the smallest distance to the feature vector.
- All state transition probabilities are initially chosen and then remain fixed. This allows more flexibility in the bottom up construction of word and sentence HMM's.

With these simplifications, the score (5.5) becomes

$$score(\mathcal{O} \mid M) = -\sum_{k=1}^{N} \log a_{s_k s_{k+1}} - \sum_{k=1}^{N} \log P(c_\mu \mid s) + \sum_{k=1}^{N} h \cdot |y(k) - c_\mu| \tag{5.6}$$

with

$$\mu = \arg\min_i |y(k) - c_i|.$$

In (5.6) h is a scaling factor. It is not possible to derive h analytically; it has to be determined experimentally. It is chosen to maximize the perfomance of the whole sytem. Since this scaling factor may be different in training and testing and since it will be a parameter in later experiments, we refer to it as *TRcbfact* in a training session and *ac* in the test session.

In the training phase, the codebook $\mathcal{C}$ and the discrete distributions $P(c_i \mid s)$ are updated iteratively by using a Veterbi algorithm.

5.2.2 Dictonary and Grammar

In order to recognize sentences, one needs HMM's for both word and sentence level. Since it is impossible to have a HMM for each possible sentence and very expensive to have one for each possible word (which would mean 1000 HMM's), the HMM's are build up on line. To create a HMM for a word, one simply connects phoneme HMM's corresponding to the phonetic transcription of each word to each other. The phonetic transcription for each word is stored in the dictonary.

The transcription of the words is based on 48 different phonemes. Therefore, under consideration of 3 subphonemes per phoneme, the number of classes is 144.

In order to compose sentence HMM's out of word HMM's one needs to determine a grammatical structure. In the experiments here, we use a simple word pair grammar. This is a special bigramm grammar, in which the probabilities for each pair of words is either zero or one. The perplexity of this grammar is 60. At the beginning and the end of each sentence, a silence state is added. Thus, the number of different classes increases to 145. Furthermore, there is a additional state called STOP needed in the recognition process. Although not really a class, we will count the STOP state as class with zero members, thus increasing the number of classes to 146.

In the recognition phase, the number of valid sentence HMM's are succesively restricted by prunning and beam search techniques.

5.2.3 Versions for Different Features

For the experiments, two different settings for the recognizer are used. For the first experiment, the features are the primary features described above. Since melscale coefficients and delta coefficients are two completely different representations, they have seperate codebooks and seperate discrete distributions over these codebooks. The score therefore is modified to

$$\begin{aligned} score(\mathcal{O} \mid M) &= -\sum_{k=1}^{N} \log a_{s_k s_{k+1}} - \nu \left(\sum_{k=1}^{N} \log P(c_\mu \mid s) + \sum_{k=1}^{N} h \cdot |y(k) - c_\mu| \right)_{melscale} \\ &- (1-\nu) \left(\sum_{k=1}^{N} \log P(c_\mu \mid s) + \sum_{k=1}^{N} h \cdot |y(k) - c_\mu| \right)_{delta} . \end{aligned} \tag{5.7}$$

The factor ν is determined experimentally, and does not change.

The other experiments are carried out on features created on this primary features by LDA and NLDA. Because of the integrating abillity of LDA and NLDA, only one set of codebooks and distributions are necessary. The score is then normally calculated by (5.6).

Since the goal is to recognize continuous speech, two different accuracy measures on the word level are distinguished. The first measure is called *correct word rate (CWR)* and is simply calculated by

$$CWR = \frac{N_{correct\ words}}{N_{total\ words}}. \tag{5.8}$$

Since we deal here with continuous speech, it is still possible that between two subsequent words in the test utterance, recognized correctly by the recognizer, another word was inserted by the recognizer. Or, vise versa, a word in the utterance can be deleted in the recognition process. The *word error rate* WER is therefore

$$WER = \frac{N_{substitution} + N_{deletion} + N_{insertion}}{N_{correct\ words} + N_{deletions} + N_{substitutions}}. \tag{5.9}$$

The second accuracy measure is based on the *word error rate* and is called *word accuracy* WA. Its definition is

$$WA = 100\% - WER = \frac{N_{correct\ words} - N_{insertion}}{N_{correctwords} + N_{deletions} + N_{substitutions}}. \tag{5.10}$$

The *word accuracy* is the more important measure for continuous speech. When refering to word recognition rates we will use the *word accuracy* but give the *correct word rate* in brackets.

5.2.4 Training, Test and Cross Validation Set

From all the utterances in the database a training set, a test set, and a cross validation set have been chosen. All the sets consist of male speakers only. The training set contains 2830 sentences from 78 male speakers. The official test set contains 390 sentences from 13 male speakers.

Due to the concept of the recognizer, a two dimensional parameter space spanned by the recognizer parameters *TRcbfact* and *ac* has to be scanned in the LDA experiments. In the case of NLDA, we also have additional the moving target parameter, thus creating a five dimensional prameter space. This makes it expensive to use the full test set for the evaluation of good parameters. Therefore, a small cross validation set has been defined. It contains 48 sentences from 12 male speakers. All experiments in this work except the concluding experiments have been carried out on the cross validation set in the test process. The concluding experiments have been performed on the official test set with the best parameter combinations evaluated on the cross validation set.

5.3 Experiment on Primary Features

To get the possibility of comparism, an experiment on the primary features was carried out first. Since the primary features have usually been used with this recognizer, the value for ν in (5.7) was already experimentally investigated and hardcoded in the recognizer's program. Also the best values for the parameter *TRcbfact* and *ac* (cp. (5.6)) are known for this configuration. Those values also have been hardcoded as a origin for both parameters. Thus, *TRcbfact* and *ac* are virtually set to 1.0 in this configuration and every other configuration refers to this origin for the values of *TRcbfact* and *ac*.

The codebooks and distributions have been calculated on the 2830 training set. Then the training process was carried out, consisting of 6 iterations over the training set. With the trained codebooks and distributions, the cross validation set was evaluated. The following table shows the results:

correct	insertions	deletions	substitutions
75.2% (76.8%)	1.6%	8.0%	15.2%

(Again, in the correct column the value in brackets is the *correct word rate*). This values serve as first brenchmarks for the following experiments.

5.4 Experiment with LDA Derived Features

In [Mai94] the teamwork of the recognizer and features derived by LDA is investigated. There it is shown that a dimension reduction of up to 16 from the original 32 dimensions does not lead to any loss in performance worth mentioning. There it is also investigated how different LDA class definitions influence the recognition performance. It is shown there that the best results can be achieved by also using the 146 classes distinguished by the recognizer as classes for the LDA.

Thus, the LDA on the primary features was carried out here using 146 classes. Since, after LDA, the coefficients of the feature vectors do not necessarilly range in the interval [0..1], a scaling algorithm was needed to scale them back into this interval. This scaling procedure should not influence any of the LDA properties. The scaling was done by the following procedure :

1. Calculate the minimal and maximal value τ^i_{min} and τ^i_{max}, $i = 1, \ldots, 16$, for each of the 16 channels in the transformed space using the whole training data.

2. Scale the LDA kernel $K = [k_{ij}]$ using
$$k_{ij} \rightarrow \frac{0.8}{\tau^i_{max} - \tau^i_{min}} \cdot k_{ij}.$$

3. Calculate the offset ϕ^i for each channel by
$$\phi^i = 0.1 - \frac{0.8\tau^i_{min}}{\tau^i_{max} - \tau^i_{min}}.$$

4. Calculate the transformed within-class scatter matrix $\tilde{S}_w = K^T S_w K$. This is a diagonal matrix, but not a unity matrix because of the scaling of the kernel.

5. Find the smallest diagonal element σ of $\tilde{S}_w = [\tilde{s}_{ij}]$.

6. Rescale the LDA kernel K using

$$k_{ij} \rightarrow \sqrt{\frac{\sigma}{\tilde{s}_{ii}}} \cdot k_{ij}.$$

7. Rescale the channel offsets ϕ^i using

$$\phi^i \rightarrow \sqrt{\frac{\sigma}{\tilde{s}_{ii}}} \cdot \phi^i.$$

8. Calculate the feature vectors Y from the primary feature vectors X by applying

$$Y = K^T X + \Phi \tag{5.11}$$

with $\Phi = [\phi^1, \ldots, \phi^{16}]$.

Using this procedure, all the coefficients of the transformed training data range in the interval [0.1 .. 0.8]. The within-class scatter matrix in the feature space is not a unity matrix, but is derived from a unity matrix by scaling with the constant factor σ. Thus the scaling properties of LDA are preserved.

When transforming the test set and the cross validation set (applying (5.11)), it is still possible that some coefficients exceed the interval [0..1]. Those coefficients are set to 0.0 or 1.0 respectively. However, less than 0.0002% of the data fell outside [0..1].

Again, for later comparisons, the seperability criterion (3.5) for the LDA feature vectors is

$$Q_{146} = 7.329.$$

Here the index indicates the number of classes considered in the calculation of this value. Since it is also useful to know this value in the case that only phoneme classes instead of subphoneme classes are considered, it is given here

$$Q_{50} = 15.860.$$

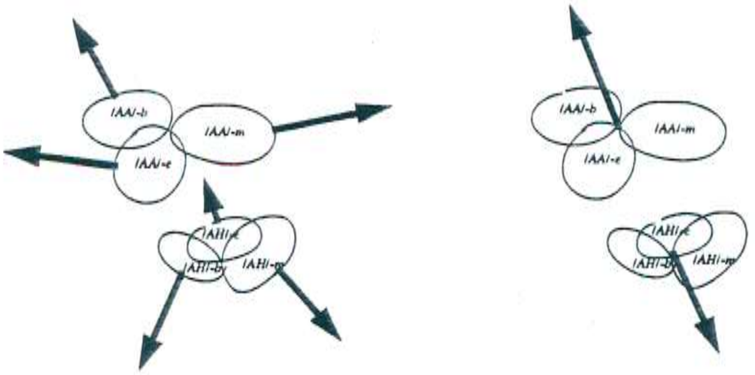
Here the classes are the 48 phonemes, the silence class and the STOP state.

The value of Q_{146} on the primary features is 36.46. Thus the value of Q is clearly minimized by LDA as was expected (although, due to the dimension reduction, those two values are not simply comparable).

On the so derived feature vectors, the recognizer was trained over 6 iterations. Since for these features the values for *TRcbfact* and *ac* are unknown, the parameter space spanned by these two parameters had to be scanned searching for the values which give the best recognition performance. Since a change in *TRcbfact* (where the whole training has to be repeated) is more expensive than a change in *ac*, the scan for *TRcbfact* is coarse. Also, experiments showed, that the recognition performance is more sensitive to changes in *ac* than to changes in *TRcbfact*. For *TRcbfact* and for *ac* 5 values and 14 values have been evaluated respectively. The detailed results can be found in Appendix D, section *Experiment with LDA feature vectors*. Here, only the best result is given. For this, the word recognition performance, insertions, deletions, and substitutions are:

correct	insertions	deletions	substitutions
84.5% (87.5%)	3.0%	3.9%	8.6%

The recognition performance is much higher than with the primary features. Especially the substitution rate has decreased. The sum of insertions and deletions has decreased slightly, substitutions and deletions have become more balanced. This is a strong improvement, especially if one considers that the feature vectors have only half the dimension of the primary features. The two main reasons for this improvement are the following: First, LDA is an integrating transformation in the sense that two different representations (melscale coefficients and delta coefficients) can be integrated statistically optimal into a homogenous representation. The analoge, at a heuristical level, for this integration in the case of the primary features is equation (5.7). The second reason is



Using one drift vector per subphoneme class accomodates the danger of disturbing neighbourhood relations

Using one drift vector for corresponding subphoneme classes keeps neighbourhood relations at subphoneme level

Figure 5.3: Illustration of the results of using one drift vector for each subphoneme class or one drift vector for corresponding subphoneme classes.

the scaling ability of the LDA, which means scaling each feature channel independently to have a unity variance within the classes [1] , or, in our case (due to the scaling algorithm), a unity matrix multiplied by a constant value. The decorrelation of the channels may be yet another aspect of importance.

5.5 Experiments with NLDA Derived Features

Experiments with NLDA derived features have been carried out with the NLDA approaches (4.10) and (4.12). Before describing the experiments in detail some general remarks will be given.

The recognizer distinguishes between the above mentioned 146 classes (mostly subphoneme classes). [Mai94] shows that LDA works best if the number of clases distinguished by LDA corresponds to number of classes used in the recognizer. Thus, seeing the problem as a classification problem with unrelated classes, one can use (4.16) to calculate for each subphoneme class its own drift vector. This assures that the value of Q_{146} is mimimized optimally.

However, the classes in speech recognition are related to each other, especially the subphoneme classes of one phoneme. It may not be good to let these corresponding subphoneme classes drift away from each other since there is the danger that subphoneme classes of different phonemes intermix, which would disturb neighbourhood relationships.

Therefore it seems justified to assign the same drift vector to coresponding subphoneme classes. Thus, three corresponding subphoneme classes would always stay close to each other and drift away together form other corresponding subphoneme classes which in turn stay together. Figure 5.3 illustrates this. This goal can be achieved by calculating the mean vectors of the centroids of the corresponding subphoneme classes and using them for the calculation of drift vectors. This results in 48 drift vectors for the phonemes plus one drift vector for the silence class and one for STOP, yielding 50 drift vectors.

Another possibility worth a try is using only phoneme classes in the NLDA training (i.e., 50 centroids and 50 drift vectors) but subphoneme classes in the the recognition process.

In all experiments the NLDA was trained with 6 iterations over the 2830 sentence training set. After the training, an additional LDA (always using 146 classes even in the case where the NLDA was trained using 50 classes) was calculated in order to return the scatter matrices to diagonal matrices. The data was scaled with the above described algorithm to range in the interval [0..1]. Since in all different approaches the magnitudes of the drift vectors equaled the magnitudes of the

[1] Which does not mean that each single class covariance matrix is a unity matrix, see chapter 3.

class centroids, the interval for the moving target parameter β was restricted to [0..1]. Experiments have been carried out with the values 0.0, 0.1, 0.2, 0.3, 0.4 and in some cases with 1.0. Since this procedure results in a good number of experiments, the values for α and γ have not been varied. In all experiments γ was 1. In almost all experiments the value 0.9 was chosen for α. This corresponds mostly to the attempt to concentrate the classes around their means. The centroids and drift vectors for the calculation of the backpropagation targets have been calculated in the beginning of each training and have then been kept for the 6 training iterations.

With every trained NLDA representation the recognizer was trained with 6 iterations over the set of training sentences. Here the parameter *TRcbfact* again was varied, mostly using 5 or 6 different values. Due to big number of experiments the test step was carried out in conjunction with the cross validation set. In the test stage, the parameter *ac* was varied finer, using between 10 and 25 different values.

For each experiment the detailed results are given in Appendix D.

5.5.1 Experiments with NLDA $Y = A^T f(X)$ Derived Features

Experiments with 146 Centroids and 146 Drift Vectors in Target Calculation

With this approach 5 experiments have been carried out. The value of γ has been always 1, the values for β have been incremented by 0.1 for each experiment starting with 0. The value for α has been always 0.9 except the experiment with $\beta = 0.0$, where $\alpha = 1.0$. Thus, the first experiment with the moving target parameter set (0.0, 1.0, 1) only concentrates the classes around their expected vectors without moving them actually away from each other.

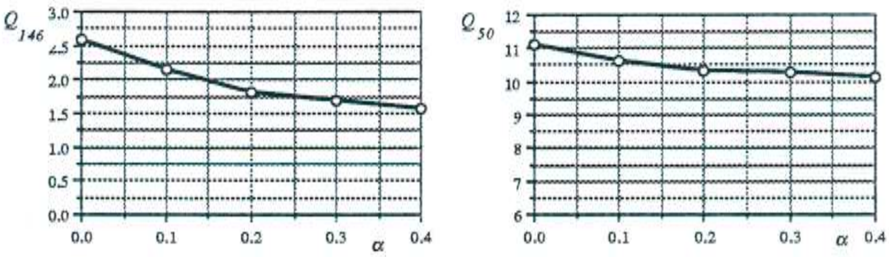


Figure 5.4: Seperation quality measures Q_{146}'s and Q_{50}'s dependence on the moving target parameter β on the transformed training set. It holds always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

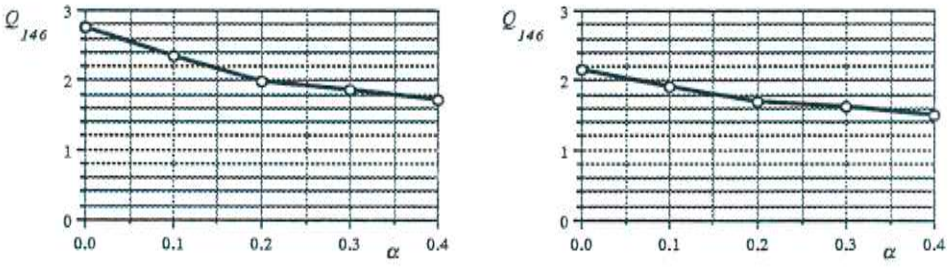


Figure 5.5: Seperation quality measure Q_{146}'s dependence on the moving target parameter β on the transformed test set (left) and cross validation set (right). Again it is always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

Figure 5.4 shows the developement of the class seperability measures' dependence on the parameter β. Both the measure with respect to 146 classes and the measure with respect to 50 classes decrease, as expected, with increasing value of β. However, even without moving targets ($\beta = 0$) the values for Q_{146} and Q_{50} are significant smaller than in the LDA case (cp. section 5.4). This points out that most of the minimization of the seperability measures comes through concentrating the classes, i.e., minimizing the within-class scatter.

To confirm that the transformation not only increases class seperability on the training set but also on the test and cross validation sets, it has been measured on these sets respectively. Figure 5.5 shows the result. Here it is good to mention that due to the small number of samples in the cross validation set, the values of Q_{146} for this set have to be considered with care. They may be only poor approximations of the true values.

parameter					word recognition	insertions	deletions	substitutions
β	α	γ	*TRcbfact*	*ac*	performance			
0.0	1.0	1	12.0	14.0	82.4% (84.7%)	2.3%	3.9%	11.4%
0.1	0.9	1	20.0	16.0	81.5% (82.7%)	1.2%	4.6%	12.7%
0.2	0.9	1	20.0	18.0	79.1% (81.1%)	2.7%	5.2%	13.0%
0.3	0.9	1	20.0	18.0	77.4% (80.2%)	2.9%	4.8%	15.0%
0.4	0.9	1	30.0	26.0	80.0% (83.2%)	3.2%	3.9%	12.8%

Table 5.2: Word recognition performance for the best values of *TRcbfact* and *ac*. The detailed evaluation of the experiments can be found in Appendix D, Experiment 1.1 - Experiment 1.5.

Table 5.2 shows the best word recognition results achieved on the cross validation set and gives the corresponding values for *TRcbfact* and *ac*, the deletions, the insertions, and the substitutions. In Appendix D, Experiments 1.1 - Experiments 1.5, the dependency of the recognition performance from the values of *TRcbfact* and *ac* is shown in more detail.

Surprisingly, although the class seperability has improved strongly for all experiments the word recognition performance has not. In all experiements the *word accuracy* is less than the value achieved with LDA. It is nevertheless still higher than the *word accuracy* on the primary features, in most cases significantly higher. The low performance is mostly due to an increment in the substitutions. The reason for this behaviour may be the above mentioned disturbance of neighbourhood relations when using one drift vector for each subphoneme class. But there are more reasons one could think of. They will be discussed further down after the evaluation of the experiments with 146 centroids/50 drift vectors and 50 centroids/50 drift vectors.

Experiments with 146 Centroids and 50 Drift Vectors in Target Calculation

If the disturbance of the neigbourhood relations is really the reason (or one reason) for the bad recognition performance, then using the approach of one drift vector for corresponding subphoneme classes should bring an improvement in the recognition performance.

There have been only 4 experiments necessary since for the parameter set (0.0, 1.0, 1) the number of drift vectors does not matter. (They are weighted by zero). Again the value of γ is 1 in each experiment and the value of α is 0.9.

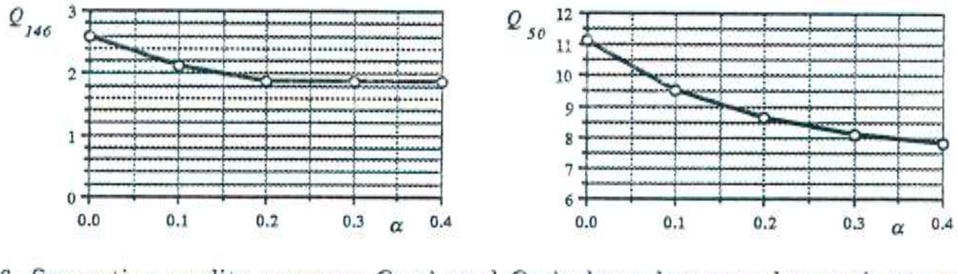


Figure 5.6: Seperation quality measures Q_{146}'s and Q_{50}'s dependence on the moving target parameter β on the transformed training set. It was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

Figure 5.6 again shows the developement of the seperability measure Q_{146}'s and Q_{50}'s dependence on the parameter β. The evolution of Q_{146} is as good as it was in the case of 146 drift vectors although

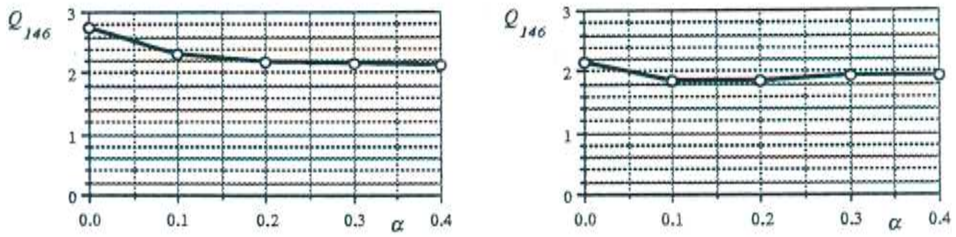


Figure 5.7: Sepereation quality measure Q_{146}'s dependence on the moving target parameter β on the transformed test set (left) and cross validation set (right). Again it was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

only 50 drift vectors have been applied here. However the value of Q_{50} has decreased much more than is has in the case of the 146 drift vectors. This was to be expected since corresponding subphoneme classes keep close to each other and drift apart from other subphoneme classes. Figure 5.7 shows Q_{146}'s dependence on β for the test and cross validation sets.

parameter					word recognition	insertions	deletions	substitutions
β	α	γ	*TRcbfact*	*ac*	performance			
0.0	1.0	1	12.0	14.0	82.4% (84.7%)	2.3%	3.9%	11.4%
0.1	0.9	1	20.0	26.0	83.4% (85.4%)	2.0%	3.9%	10.7%
0.2	0.9	1	40.0	36.0	81.5% (84.0%)	2.5%	3.2%	12.8%
0.3	0.9	1	10.0	32.0	79.0% (81.8%)	2.9%	3.9%	14.3%
0.4	0.9	1	40.0	26.0	79.3% (81.5%)	2.1%	4.5%	14.1%

Table 5.3: Word recognition performance for the best values of *TRcbfact* and *ac*. The detailed evaluation of the experiments can be found in Appendix D, Experiment 3.1 - Experiment 3.4.

The recognition performance on the NLDA representations trained in this way is displayed in table 5.3. The results are better when compared with the experiments with 146 drift vectors. The substitutions, as an average, have slightly decreased. This indicates that the disturbance of the neighbourhood relations was indeed a reason for the bad word recognition performance. However, since the word recognition performance is still worse than with LDA feature vectors, it is obviously not the only one.

Before discussing other possible reasons for the fact that the NLDA representation cannot, in terms of recognition performance, hold what it promised in terms of class seperability, the description of the experiments with 50 centroids and 50 drift vectors is given.

Experiments with 50 Centroids and 50 Drift Vectors in Target Calculation

Again 5 experiments have been carried out with this approach. The parameter γ has been always set equal to 1, α has been equal to 0.9 except for $\beta = 0.0$ were it was 1.0.

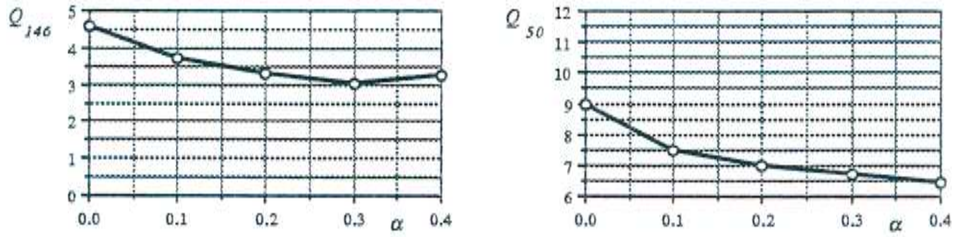


Figure 5.8: Seperation quality measures Q_{146}'s and Q_{50}'s dependence on the moving target parameter β on the transformed training set. It was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

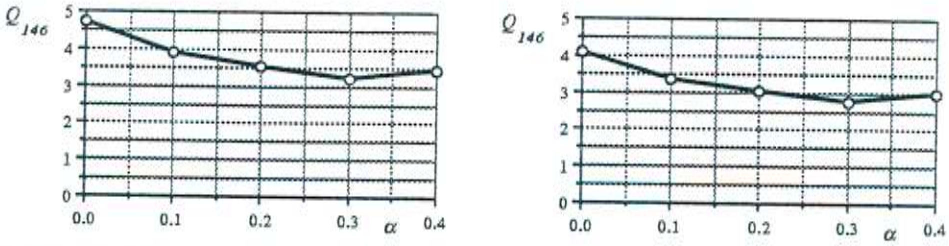


Figure 5.9: Sepereation quality measure Q_{146}'s dependence on the moving target parameter β on the transformed test set (left) and cross validation set (right). Again it was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

In figure 5.8 the seperability measures Q_{146}'s and Q_{50}'s dependence on the parameter β are displayed. The values for Q_{146} were in all experiments higher than the values in the experiments with 146 centroids and the same parameters. This is not a surprising fact since this approach merges corresponding subphoneme classes together. Figure 5.9 shows Q_{146} on the test and cross validation sets.

parameter					word recognition	insertions	deletions	substitutions
β	α	γ	*TRcbfact*	*ac*	performance			
0.0	1.0	1	20.0	38.0	78.1% (80.9%)	2.9%	3.6%	15.5%
0.1	0.9	1	20.0	36.0	81.3% (84.1%)	2.9%	3.9%	11.9%
0.2	0.9	1	20.0	30.0	79.3% (82.5%)	3.2%	4.5%	13.0%
0.3	0.9	1	30.0	20.0	81.1% (82.5%)	1.4%	5.9%	11.6%
0.4	0.9	1	10.0	18.0	79.0% (80.7%)	1.8%	5.7%	13.5%

Table 5.4: Word recognition performance for the best values of *TRcbfact* and *ac*. The detailed evaluation of the experiments can be found in Appendix D, Experiment 2.1 - Experiment 2.5.

The recognition performance achieved with the representations based on this experiments can be found in table 5.4. It is not as good as in the experiments with 146 centroids. That is explainable by the fact that during the training with 50 centroids, corresponding subphoneme classes are merged. Thus, classes distinguished by the recognizer are put together which decreases the recognition performance. Still, the *word accuracy* in all experiments is significantly higher than on the primary features.

Possible Reasons why It Does not Work

In none of the NLDA experiments done up to now an improvement in the word recognition performance compared with LDA could be achieved. This seems surprising since the class seperability could be improved significantly in all experiments. Why is this? It has been already pointed out that the optimization of the seperability measure (3.5) does not take into account the classifier which has to use the features derived by optimization of this criterion.

The recognizer used here is able to classify between classes with multimodal distributions. Moreover, it preassumes that the features have multimodal distributions within the classes. This preassumption is surely fulfilled for the primary vectors and therefore, since LDA is a linear operation, also for the feature vectors derived by LDA. However, in the NLDA case, since one trains the network to move the members of a class close to their centroid, this preassumption may not be fulfilled. Since there have been some strong simplifications made for the recognizer (cp. section 5.2, especially the assumption that the continuous distributions over the codebook vectors are normal distributions with unity covariance matrix) this may result in the following problem: How well the classifier works depends on the possibility to model the distributions within the classes with unimodal distributions having unity covariance matrices over the codebook vectors. The recognition performance indicates that this seems to work well for the primary features and therefore also for LDA derived features. For a NLDA representation however, the classes are highly concentrated around their expected vectors. Thus, most of the codebook vectors are very close to each other and the assumption of unimodal normal distributions with unity covariance matrix may model the real distribution poorly. Thus, the consideration of the real covaraince matrices over the codebook vectors could improve the recognition performance. However, to change this is beyond the scope of this thesis (but it is being worked on). Another way to handle this is to make a multimodal approach to NLDA as will be shown later.

Another possible reason is a more systematic one. In speech recognition we try to classify utterances produced by humans. But humans do make mistakes in their production of speech. There are more than two very similar sounding phoneme classes. Since similar sounding phoneme classes are also very close and partially overlaped in the primary feature space (this is due to the similarities between Fourier Transformation and the transformation done by the human auditory system) it is justified to assume that the human producer of speech also confuses acoustically close phoneme classes. Assuming this, a NLDA representation would move such a feature vector, intended to belong to one phoneme class but uttered as would belong to another, away form the class which it should belong in. In a LDA representation, however, it would stay close to the class it should belong in. This behaviour would trigger a sure misclassification in the NLDA case while the intended class assignment could possibly be restored in the LDA case. An indication of this argument is the increased rate of word substitutions in the NLDA experiments (compared to the LDA experiment).

Last but not least the goal of speech recognition is to achieve, as well as possible, a recognition performance on the word or even sentence level. But this is not what NLDA optimizes. The assumption of a correlation between the seperability of the subphoneme classes and the *word accuracy* may simply not be true or may be connected with some boundary conditions our appoaches to NLDA do not consider.

However, it is still to be seen how the second approach to NLDA works in terms of recognition performance.

5.5.2 Experiments with NLDA $Y = A^T X + f(X)$ Derived Features

This approach to NLDA, corresponding to equation (4.12), has an additional parameter than the approach just discussed, namely, the number of nonlinear units in the hidden layer. This is a very important parameter since with the number of hidden units increases the potential abillity of the neural net to learn the transformation. However, too large a number of hidden units may also result in bad performance. Therefore this parameter is, in any case, worth experimenting with. Since this increases the number of necessary experiments considerably, we have done experiments only with the combination of 146 centroids and 50 drift vectors in the target calculation. This is because this approach worked best in terms of recognition performance in the case of NLDA (4.10) and also makes the most sense from an acoustic standpoint (since it considers the affinity of subphoneme classes).

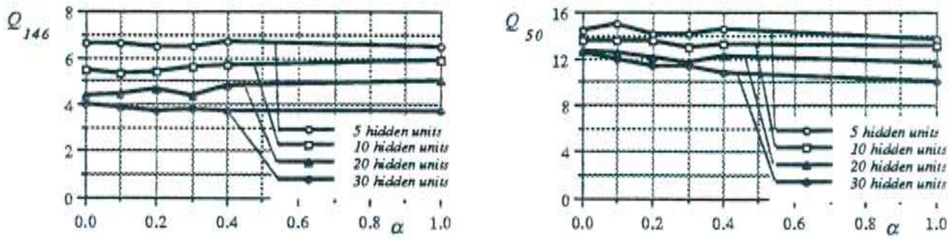


Figure 5.10: Seperation quality measures Q_{146}'s and Q_{50}'s dependence on the moving target parameter β on the transformed training set. It was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

Experiments have been carried out with 5, 10, 20 and 30 hidden units. The moving target parameter sets are the same as used in the experiments with NLDA (4.10). Additional experiments with the set (1.0, 0.9, 1) have been made in order to explore the behaviour for the case where

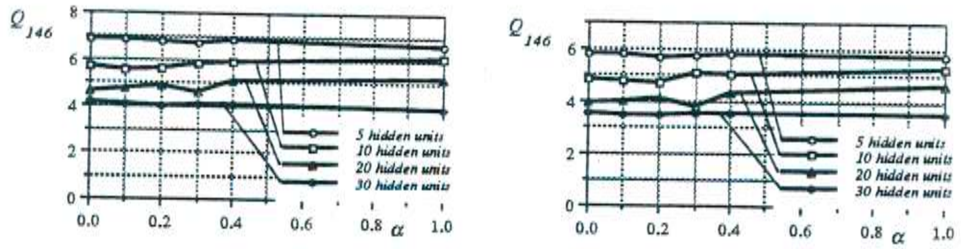


Figure 5.11: Sepereation quality measure Q_{146}'s dependence on the moving target parameter β on the transformed test set (left) and cross validation set (right). Again it was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

the order of magnitude of the drift vectors and the centroids is the same. Figure 5.10 shows the evolution of the values for the seperability measures Q_{146} and Q_{50} as a function of the moving target parameter β. Here, in contrast to the experiments with the NLDA approach (4.10) (where the seperability measures have decreased continuously with increasing of β), a strong dependency of these values from β can not be discovered. However, both, Q_{146} and Q_{50}, are decreasing with the increasing of the number of hidden units. Still, even with 30 hidden units, the values of Q_{146} and Q_{50} are a good deal higher than with the NLDA (4.10) (but, on the other hand, significantly lower than the values for the LDA case). Since the drift vectors have the same order of magnitude for both approaches, one concludes that this approach does not concentrate the classes as well as approach (4.10).

hidden units	parameter					word recognition performance	insertions	deletions	substitutions
	β	α	γ	*TRcbfact*	*ac*				
5	0.0	1.0	1	15.0	8.0	84.1% (85.9%)	1.8%	4.5%	9.6%
	0.1	0.9	1	15.0	10.0	84.8% (87.3%)	2.5%	4.1%	8.6%
	0.2	0.9	1	15.0	8.0	85.0% (87.0%)	2.0%	4.3%	8.7%
	0.3	0.9	1	10.0	8.0	85.6% (87.3%)	1.8%	4.3%	8.4%
	0.4	0.9	1	20.0	10.0	84.8% (87.2%)	2.3%	3.6%	9.3%
	1.0	0.9	1	10.0	8.0	84.5% (86.8%)	2.3%	4.3%	8.9%
10	0.0	1.0	1	25.0	10.0	84.5% (85.7%)	1.2%	4.3%	10.0%
	0.1	0.9	1	10.0	12.0	85.0% (87.7%)	2.7%	3.6%	8.7%
	0.2	0.9	1	10.0	8.0	84.1% (86.5%)	2.3%	3.9%	9.6%
	0.3	0.9	1	5.0	10.0	85.2% (87.0%)	1.8%	3.9%	9.1%
	0.4	0.9	1	2.0	14.0	85.6% (88.2%)	2.7%	3.0%	8.7%
	1.0	0.9	1	15.0	10.0	83.6% (85.2%)	1.6%	4.8%	10.0%
20	0.0	1.0	1	5.0	16.0	86.1% (88.1%)	2.0%	3.7%	8.0%
	0.1	0.9	1	20.0	12.0	84.5% (86.5%)	2.0%	4.3%	9.3%
	0.2	0.9	1	10.0	14.0	84.0% (85.9%)	2.0%	4.8%	9.3%
	0.3	0.9	1	15.0	18.0	86.3% (88.4%)	3.1%	3.9%	7.7%
	0.4	0.9	1	30.0	20.0	85.0% (88.2%)	3.2%	2.9%	8.9%
	1.0	0.9	1	10.0	16.0	82.4% (83.8%)	1.4%	5.7%	10.5%
30	0.0	1.0	1	20.0	12.0	84.8% (87.0%)	2.1%	4.3%	8.7%
	0.1	0.9	1	5.0	14.0	82.7% (83.8%)	1.1%	5.0%	11.2%
	0.2	0.9	1	10.0	12.0	84.3% (85.0%)	0.7%	4.8%	11.2%
	0.3	0.9	1	10.0	22.0	81.1% (83.2%)	2.1%	4.3%	12.5%
	0.4	0.9	1	15.0	20.0	83.6% (85.4%)	1.8%	3.4%	11.2%
	1.0	0.9	1	5.0	22.0	80.9% (84.3%)	3.4%	3.6%	12.1%

Table 5.5: Word recognition performance for the best values of *TRcbfact* and *ac*. The detailed evaluation of the experiments can be found in Appendix D, Experiment 4.1 - Experiment 7.6.

Table 5.5 shows the word recognition performance, insertions, deletions, and substitutions for the best values for *TRcbfact* and *ac*. Again, the detailed evaluation can be found in appendix D, Experiment 4.1 - Experiment 7.6. Obviously, this approach works much better in terms of recognition performance. Most experiments at least reach the brenchmark set by LDA. The best experiment is in word recognition performance almost 2% better than LDA. That corresponds to a relative improvement of 11.6% in the word error rate. It is to mention, that especially the number of insertions has decreased (compared with LDA).

This results can be seen as yet another indication of the correctness of one of the above mentioned reasons for the rather disappointing results of NLDA approach (4.10) in terms of recognition performance. Since the classes are not as highly concentrated in approach (4.12), their distributions are better modeled by the unimodal normal distributions with unity covariance matrix over the codebook vectors than with approach (4.10). This leads to the question whether there is anything one could do to modify the NLDA in a way that solves this problem. One possibility may be to preserve the multimodality of the class distributions as described in the next section.

5.6 A Multimodal Approach to NLDA

The NLDA approaches that have been discussed here have one important feature: Due to the training algorithm the samples of a class are moved closer to their centroid. As a result of this, multimodal class distributions are possibly converted into unimodal distributions (or at least the tendency goes in this direction). The right hand side of figure 5.12 demonstrates this. As pointed out, this may result in problems for the following recognizer.

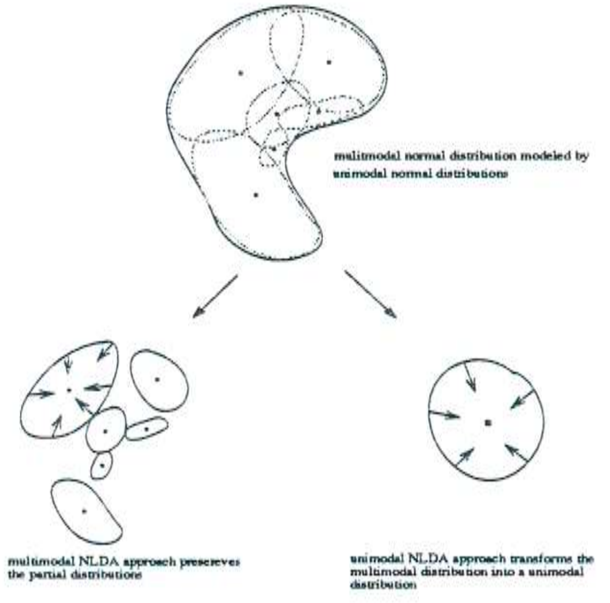

malitmodal normal distribution modeled by unimodal normal distributions
multimodal NLDA approach presereves the partial distributions
unimodal NLDA approach transforms the multimodal distribution into a unimodal distribution

Figure 5.12: A multimodal approach to NLDA would have to preserve the partial distributions.

The concept of the moving target should include the possibility of preserving multimodal distributions by choosing a small value for the parameter α. However, experiments showed that this does not help (Appendix D, Experiment 8.2 and 8.3). The problem here lies in the big number of targets (each sample has its own distinct target!) which become more and more scattered with the decreasing of the parameter α. Thus, the network has problems to learn a moderate transformation.

The question is now whether there is another possibility to preserve the multimodality of the class distributions. This can be done in the following way: First one models the multimodal distribution with unimodal partial distributions. Then, instead of taking the expected vector (i.e., the class centroid) as target for the training, one takes the expected vectors of the partial distributions as targets and moves the samples belonging to a partial distribution to its expected vector. Thus, the partial distributions are concentrated but the multimodal character of the class distribution is preserved. The left hand side of figure 5.12 displays this.

Unfortunately neither the partial distributions nor their expected vectors are known. Thus they have to be modeled somehow. One way to do this is to estimate the expected vectors by a vector quantization and assign the samples to the codebook vector with the smallest euclidean distance. This is only a simple method but should work well enough for first experiments. Of course, this method can be combined with the idea of the moving targets. This results in the following modified NLDA algorithm:

1. Choose a neural network corresponding to the different types of NLDA (4.10) or (4.12).
2. Choose the number of input units of the network to be n and the number of output units to be m.
3. If the NLDA approach is (4.12), choose the number of hidden units.
4. Choose a parameter triple (β, α, γ).
5. Initialize the network to achieve an initial LDA state.
6. Pass all training data through the network and calculate the class centroids in the output space, the scatter matrices in the output space and the criterion (3.5).
7. If there is no significant change in the value of (3.5) compared to the previous iteration or if this is the k-th iteration go to step 11.
8. If the number of iterations modulo l is zero:
    (a) Do a vector quantization in the output space for each class.
    (b) Calculate the drift vectors for each class by using (4.16) using the class centroids (not the codebook vectors!).
9. For each sample:
    (a) Find the codebook vector closest to the sample (out of the codebook vectors for this class).
    (b) Calculate target for error backpropagation using (4.17) and the closest codebook vector instead of the centroid.
    (c) Perform error backpropagation using mean squared error as error measure.
    (d) Update the network weights.
10. Go to step 6.
11. Calculate an additional LDA without dimension reduction on the output data to return the scatter matrices in diagonal matrices (this step can be skipped if diagonal matrices are not needed).
12. STOP.

Again the parameter l stands for the update rate of the codebooks and centroids and k for the maximum number of iterations.

With this approach some experiments have again been carried out. Again the training was made over 6 iterations. The drift vectors and codebook vectors have been calculated in the beginning and kept for all iterations. For the vector quantization, the routines of the recognizer were used (k-nearest neighbour). For each phoneme class, 50 codebook vectors have been calculated. Thus, the number of 'centroids' in the target calculation is 2451. Since the codebook vectors belonging to one phoneme class should stay together and move away from codebook vectors of another class, they have the same drift vector. This gives again 50 drift vectors.

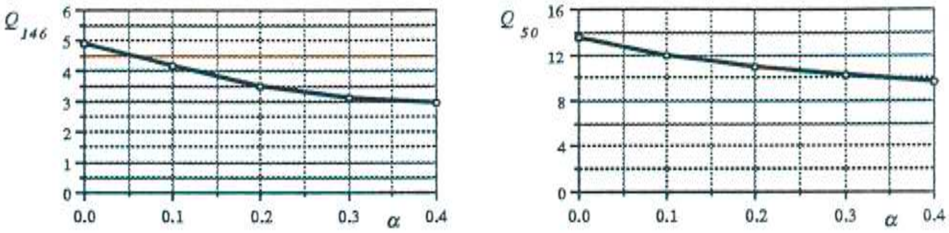


Figure 5.13: Seperation quality measures Q_{146}'s and Q_{50}'s dependence on the moving target parameter β on the transformed training set. It was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

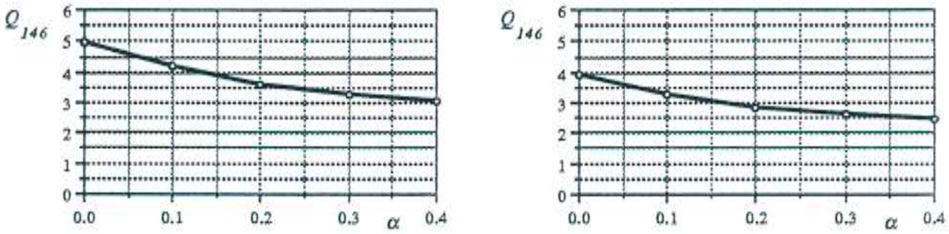


Figure 5.14: Sepereation quality measure Q_{146}'s dependence on the moving target parameter β on the transformed test set (left) and cross validation set (right). Again it was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

5.6.1 Experiments with Multimodal NLDA $Y = A^T f(X)$

Again the experiments are here reviewed only in a summary. Detailed information for each experiment can be found in appendix D, Experiment 9.1 - Experiment 9.5.

Figure 5.13 shows the seperability measure Q_{146} and Q_{50} on the trainings set as functions of the parameter β. The values for both measures are not as small as for the corresponding unimodal experiments. This is not surprising since, due to the training with the codebook vectors in the target calculation, the classes do not become as concentrated as in the unimodal approach. Figure 5.14 shows Q_{146} on the test and cross reference sets.

parameter					word recognition	insertions	deletions	substitutions
β	α	γ	*TRcbfact*	*ac*	performance			
0.0	1.0	1	5.0	18.0	83.6% (86.1%)	2.5%	3.9%	10.0%
0.1	0.9	1	10.0	12.0	84.5% (87.3%)	2.9%	3.2%	9.4%
0.2	0.9	1	20.0	12.0	82.2% (84.5%)	2.3%	4.3%	11.2%
0.3	0.9	1	15.0	36.0	81.6% (85.6%)	3.9%	2.9%	11.6%
0.4	0.9	1	30.0	18.0	80.6% (82.9%)	2.3%	4.1%	13.7%

Table 5.6: Word recognition performance for the best values of *TRcbfact* and *ac*. The detailed evaluation of the experiments can be found in Appendix D, Experiment 9.1 - Experiment 9.5.

Table 5.6 gives again the best recognition performance, insertions, deletions, and substitutions for each experiment with the corresponding parameters. The recognition performance improved in comparison to the unimodal approach (table 5.3). However, the best experiment is only as good as LDA. Again, especially the insertion rate decreased compared with the LDA experiment. The substitution rate is always higher than in the LDA case.

The problem this time lies again (probably) in the teamwork of feature extraction and recognizer. Although we now have concentrations around distinct codebook vectors the distributions over the codebook vectors are probably not normal with unity covaraince matrix as assumed by the recognizer. Here again only a change in the recognizer (consideration of the covariance matrices over the codebook vectors) can bring further improvement.

5.6.2 Experiments with Multimodal NLDA $Y = A^T X + f(X)$

Experiments with the multimodal approach have also been carried out with the other NLDA approach, (4.12). This time 10, 20 and 30 hidden units were used. More information about those experiments can be found in appendix D, Experiment 10.1 - Experiment 12.6.

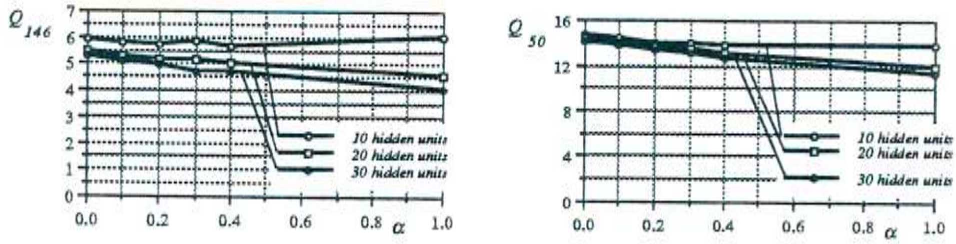


Figure 5.15: Seperation quality measures Q_{146}'s and Q_{50}'s dependence on the moving target parameter β on the transformed training set. It was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

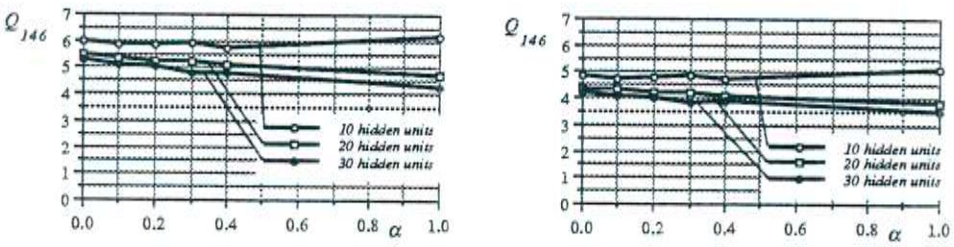


Figure 5.16: Sepereation quality measure Q_{146}'s dependence on the moving target parameter β on the transformed test set (left) and cross validation set (right). Again it was always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

number of	parameter					word recognition	insertions	deletions	substitutions
hidden units	β	α	γ	*TRcbfact*	*ac*	performance			
10	0.0	1.0	1	10.0	8.0	85.9% (88.1%)	2.1%	3.0%	8.9%
	0.1	0.9	1	2.0	8.0	84.7% (87.9%)	3.2%	3.0%	9.1%
	0.2	0.9	1	20.0	8.0	83.6% (87.3%)	3.7%	3.2%	9.4%
	0.3	0.9	1	20.0	6.0	85.6% (87.7%)	2.1%	2.9%	9.4%
	0.4	0.9	1	5.0	8.0	85.0% (87.9%)	2.9%	3.2%	8.9%
	1.0	0.9	1	15.0	8.0	85.0% (87.2%)	2.1%	3.9%	8.9%
20	0.0	1.0	1	5.0	8.0	83.8% (86.8%)	3.0%	3.0%	10.2%
	0.1	0.9	1	20.0	6.0	83.2% (85.6%)	2.3%	4.3%	10.2%
	0.2	0.9	1	15.0	6.0	84.8% (87.3%)	2.5%	3.4%	9.3%
	0.3	0.9	1	10.0	8.0	83.4% (87.3%)	3.9%	3.2%	9.4%
	0.4	0.9	1	20.0	8.0	83.1% (85.0%)	2.0%	4.3%	10.7%
	1.0	0.9	1	10.0	12.0	84.0% (85.2%)	1.2%	3.0%	11.8%
30	0.0	1.0	1	7.5	6.0	84.0% (86.1%)	2.1%	3.6%	10.3%
	0.1	0.9	1	20.0	8.0	84.3% (87.3%)	3.0%	3.0%	9.6%
	0.2	0.9	1	15.0	6.0	84.7% (86.6%)	2.0%	4.3%	9.1%
	0.3	0.9	1	5.0	6.0	84.0% (85.7%)	1.8%	3.9%	10.3%
	0.4	0.9	1	5.0	8.0	85.4% (88.4%)	2.1%	3.0%	9.4%
	1.0	0.9	1	20.0	14.0	83.2% (85.2%)	2.0%	4.1%	10.7%

Table 5.7: Word recognition performance for the best values of *TRcbfact* and *ac*. The detailed evaluation of the experiments can be found in Appendix D, Experiment 10.1 - Experiment 12.6.

Figures 5.15 and 5.16 show the seperation quality measures Q_{146} on the training set and Q_{146}

on the test and cross validation sets. Again, the values are decreased if the number of hidden units is increased.

Table 5.7 shows the best recognition performances, insertions, deletions, and substitutions for each experiment together with the corresponding parameter set. Again, this NLDA approach works better than the other, although this time the recognition performance could not be improved in comparison to the unimodal approach. The reason for this is the mentioned above. However, again this approach leads to a better recognition performance than LDA, the relative improvement in the word error rate is 9% for the best experiment.

5.7 Summary

At this stage a brief summary of all the experiments done should be made. Out of all the experiments made, the best one for each approach has been taken to represent that approach. The results with all corresponding parameters are in the following table:

feature extraction algorithm	parameter								word recognition performance
	hidden units	centroids in training	drift vectors	β	α	γ	*TRcbfact*	*ac*	
MSC + Delta	-	-	-	-	-	-	1.0	1.0	75.2% (76.8%)
LDA	-	146	-	-	-	-	8.0	7.0	84.5% (87.5%)
NLDA $Y = A^T f(X)$	32	146	50	0.1	0.9	1	20.0	26.0	83.4% (85.4%)
NLDA $Y = A^T X + f(X)$	20	146	50	0.3	0.9	1	15.0	18.0	86.3% (88.4%)
NLDA $Y = A^T f(X)$ multimodal approach	32	2451	50	0.1	0.9	1	10.0	12.0	84.5% (87.3%)
NLDA $Y = A^T X + f(X)$ multimodal approach	10	2451	50	0.0	1.0	1	10.0	8.0	85.9% (88.1%)

With this parameter sets, achieved on the cross validation set, there will be carried out the experiments on the official 390 sentences test set (chapter 7).

However, the recognition results achieved with NLDA features are a little disappointing compared to the ability of NLDA to optimize the class seperability. Possible reasons why the approaches to NLDA, although working well in the optimization of the seperation quality measure, do not work as well in terms of recognition performance have been discussed.

At this stage NLDA seems still to be too immature to be applied in the front end of a recognizer. There are some things that could be done to improve the suitability of NLDA as a feature representation in general and especially for speech recognition. Those ideas are discussed in chapter 8.

Chapter 6

Experiments with Context Dependent Phoneme Models

The experiments described in the previous chapter have been carried out in an enviroment which uses monophones as acoustic models, i.e, the recognizer distinguishes between the subphonems of 48 phoneme classes. Here the phoneme classes are isolated; the neighbours of the actual phoneme are not considered. Therefore this approach is called context independent.

On the other hand there is the context dependent approach. This approach considers the actual neibhbours (left and right) of a phoneme in a word. Thus, the acoustic models are triphones. This approach is justified by the assumption that much of the information in speech lies in the transitions between two subsequent phonemes. Thus it makes sense to preserve this information in the class structure instead of discarding it by merging the triphones together into a monophone. Of course, this procedure results in a considerable increment of classes.

In the context dependent environment here, we use the recognizer described in chapter 5. The number of distinguished triphones here is 2374. Additionally there is again a silence class and a STOP. Each triphone is again split into 3 sub(tri)phones. Thus, the number of different classes is 7124. Fortunately, due to the concept of semicontinous HMM's, all triphones belonging to the same monophone class share the same codebook vectors. Thus, the computational effort remains manageable although there are now 7124 distributions over the codebook instead of 146 in the context independent case. Training and testing take only slightly longer than in the context independent case.

Again, first an experiment on the primary features has been carried out. Then an experiment with LDA derived features was done. Finally again several experiments with different NLDA approaches have been carried out. Here only the unimodal approaches were investigated. That was partly due to time reasons but also because of the much finer class structure. This gives a unimodal approach much more justification than it had had in the monophone environment.

The recognizer was trained in every experiment with 6 iterations over the trainings set. In the test stage the cross validation set was always used. For the best parameter sets on the cross validation set, experiments on the official trainings set are described in chapter 7.

6.1 Experiment on the Primary Features

On the primary features, the values for *TRcbfact* and *ac* have both been 1.0. The word recognition performance, insertions, deletions and substitutions on the cross validation set are

correct	insertions	deletions	substitutions
86.8% (89.3%)	2.5%	2.7%	8.0%

(The value in brackets in the correct column is the correct word rate). This result is not only much better than the one achieved with monophones and primary features but also better than the word recognition performances with both LDA and NLDA derived features and monophone acoustic models.

6.2 Experiment with LDA Derived Features

The LDA was calculated using the 7124 subtriphone classes. Again the dimension was reduced from original 32 to 16. The seperation quality measure under consideration of those 7124 classes after the transformation is

$$Q_{7124} = 3.017$$

on the training set and

$$Q_{7124} = 1.817$$

on the test set. That the value for the test set is so much smaller than for the training set is due to the finer class structure. Because of this, there are many triphone classes without representative in the test set. This is still worse for the case of the cross validation set. Therefore, the measure's value on the cross validation set is not given here.

Under consideration of the 146 submonophone classes only the value for the class seperability measure on the training set is

$$Q_{146} = 7.734.$$

Thus, Q_{146} is in the same order of magnitude as in the LDA carried out using monophone acoustic modeling. If one considers only the monophone classes, the value of the seperability measure on the training set is

$$Q_{50} = 16.1534.$$

The best word recognition performance using the LDA derived features was:

correct	insertions	deletions	substitutions
90.9% (93.8%)	2.9%	1.1%	5.2%

Here the value for *TRcbfact* was 4.0, for *ac* it was 8.0. The table also shows insertions, deletions, and substitutions. Again, there is a strong improvement compared to the monophon experiments (due to the finer class structure) but also compared to the experiments with the primary features and the triphone acoustic models. Reason for this improvement are again the scaling abilities of LDA and its ability to integrate melscale and delta-melscale representation into a homogeneous feature space. A detailed evaluation of the LDA experiment can be found in appendix E, section Experiment with LDA Feature Vectors.

6.3 Experiments with NLDA Derived Features

For the experiments with NLDA derived features, again the two approaches (4.10) and (4.12) were employed. In each experiment, the NLDA network was trained with 6 iterations over the training set. Experiments have been carried out with the values 0.0, 0.1, 0.2, 0.3, 0.4, and 1.0 for the parameter β. The parameter α was always 0.9 except for the case that $\beta = 0.0$ where $\alpha = 1.0$. This case corresponds to a concentration of the feature clouds around their expected vectors without applying the concept of moving targets. In all experiments, the parameter γ was 1.0. The learning rate for the error backpropagation was 0.008 and the backpropagation momentum was 0.9.

Since the classes discriminated by the recognizer are the subtriphone classes, the different approaches for the drift vector calculation introduced in chapter 5 are applicable: each subtriphone has its own distinct drift vector (7124 drift vectors), the subtriphones corresponding to a triphone share one drift vector (2376 drift vectors), or the subtriphones corresponding to a monophone share the same drift vector. Due to time reasons not every concept could be applied. Since in the context independent experiments the concept of shared drift vectors worked best, it will be applied here, too.

The recognizer was trained with the NLDA features, again using different values for the parameter *TRcbfact*. In the testing the cross validation set was used and the parameter *ac* was varied. For each experiment there is a detailed evaluation in appendix E.

6.3.1 Experiments with NLDA $Y = A^T f(X)$ Derived Features

The experiments described first in this section have been carried out with 7124 centroids and 2376 drift vectors in the target calculation. Thus, the subtriphone classes are concentrated around their expected vectors and the three subtriphones which belong to a triphone are together moved away from the other subtriphones. In the second part of this section the experiments with 7124 centroids and 50 drift vectors in the target calculation are evaluated. In this approach the subtriphone classes are also concentrated around their expected vectors but all subtriphone classes belonging to the same monophone class stay close together while moving away from the other subtriphone classes.

Experiments with 7124 Centroids and 2376 Drift Vectors in the Target Calculation

Figure 6.1 shows the developement of the class seperability measure Q_{7124} as a function of the moving target parameter β on both the training set and the test set. As in the monophone case, the values for the NLDA features are strongly minimized compared with the values for the LDA features, thus pointing out a better class seperability. Figure 6.2 shows Q_{146}'s and Q_{50}'s dependence on the parameter β. The values for these measures are also smaller than the values for the LDA derived features.

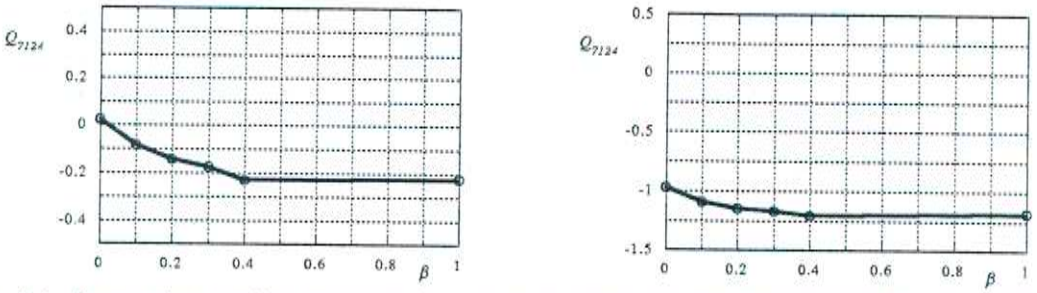


Figure 6.1: Seperation quality measure Q_{7124}'s dependence on the moving target parameter β on the training set (left) and test set (right). It holds always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

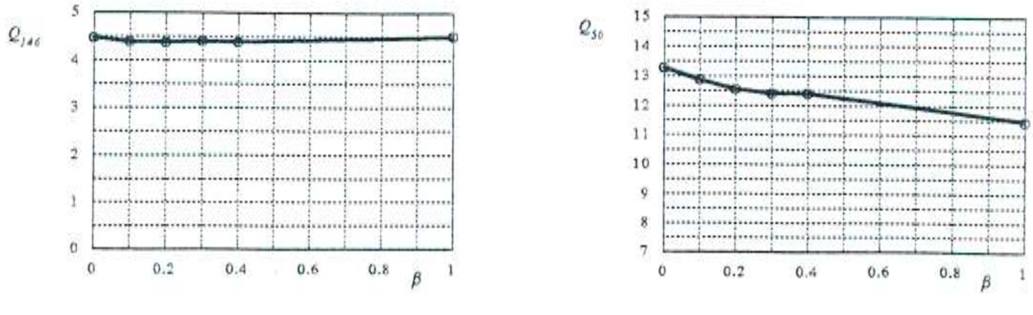


Figure 6.2: Seperation quality measures Q_{146}'s and Q_{50}'s dependence on the moving target parameter β on the training set. Again it is always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

Table 6.1 shows the best word recognition performance, insertions, deletions and substitutions for each experiment together with the corresponding parameters. As in the monophone case, there is no improvement in word recognition performance compared with LDA with this approach. However, the LDA performance is reached. As reasons for the discrepancy between the improvement in the

parameter					word recognition performance	insertions	deletions	substitutions
β	α	γ	*TRcbfact*	*ac*				
0.0	1.0	1	7.5	6.0	90.2% (92.2%)	2.0%	2.1%	5.7%
0.1	0.9	1	2.0	8.0	90.9% (92.3%)	1.4%	1.6%	6.1%
0.2	0.9	1	15.0	8.0	90.6% (92.3%)	1.8%	2.1%	5.5%
0.3	0.9	1	10.0	10.0	89.5% (92.2%)	2.7%	2.0%	5.9%
0.4	0.9	1	7.5	12.0	89.5% (92.2%)	2.7%	1.8%	6.1%
1.0	0.9	1	10.0	12.0	89.5% (92.2%)	2.7%	1.6%	6.2%

Table 6.1: Word recognition performance for the best values of *TRcbfact* and *ac*.

seperation quality measure and the nonimprovement in word recognition performance, the in chapter 5 mentioned arguments can be assumed.

Detailed evaluation of the experiments can be found in appendix E, experiments CD.1.1 - CD.1.6.

Experiments with 7124 Centroids and 50 Drift Vectors in the Target Calculation

Figure 6.3 shows the dependence of Q_{7124} fro the training set and test set in the case of 50 drift vectors. Again the values are much smaller than the values for the LDA derived features, corresponding to a better class seperability in all NLDA experiments. However, for $\beta = 1.0$ the value is increasing stronlgy (although it is still a good deal below the value in the LDA case). That might be because the order of magnitude of the weighted drift vectors and the class centroids is the same. Thus, the targets for the training come to lay far away and the network might not be able to perform as well.

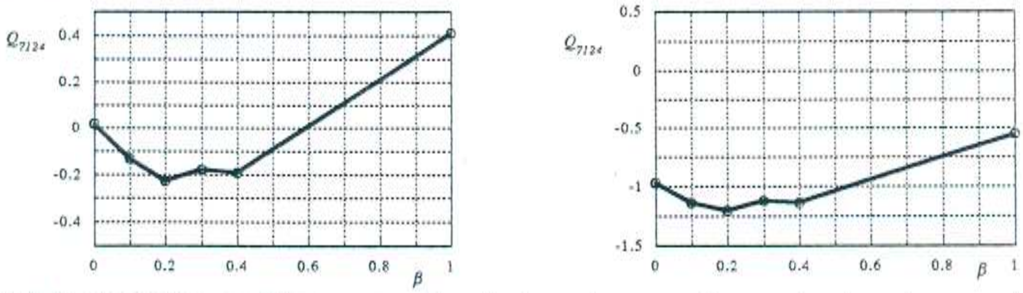


Figure 6.3: Seperation quality measure Q_{7124}'s dependence on the moving target parameter β on the training set (left) and test set (right). It holds always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

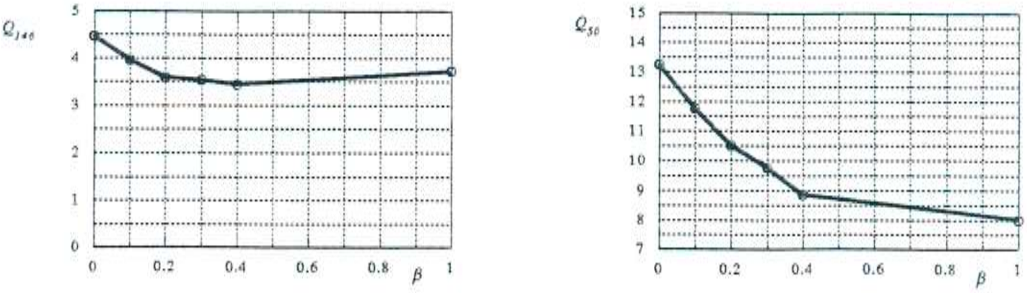


Figure 6.4: Seperation quality measures Q_{146}'s and Q_{50}'s dependence on the moving target parameter β on the training set. Again it is always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

In figure 6.4, the developement of the seperation quality measures Q_{146} and Q_{50} are shown as functions of the parameter β. Due to the consideration of the relationships between the subtriphone classes, the values of these measures are smaller than in the experiments with 2376 drift vectors.

parameter					word recognition performance	insertions	deletions	substitutions
β	α	γ	*TRcbfact*	*ac*				
0.0	1.0	1	2.0	6.0	90.6% (92.2%)	1.6%	1.4%	6.4%
0.1	0.9	1	10.0	6.0	91.1% (92.7%)	1.6%	2.1%	5.2%
0.2	0.9	1	5.0	8.0	90.7% (92.2%)	1.5%	2.1%	5.7%
0.3	0.9	1	20.0	8.0	89.7% (91.8%)	2.1%	2.5%	5.7%
0.4	0.9	1	20.0	10.0	89.5% (92.3%)	2.9%	2.1%	5.5%
1.0	0.9	1	15.0	18.8	86.6% (89.1%)	2.5%	2.3%	8.6%

Table 6.2: Word recognition performance for the best values of *TRcbfact* and *ac*.

The best word recognition performance, insertions, deletions and substitutions for each experiments together with the corresponding parameters are shown in table 6.2. With this approach, there is a slight improvement in word recognition performance with the moving target parameter set (0.1, 0.9, 1). However, this improvement is only 0.2% (or 2.2% relative improvement in the word error rate).

A detailed review of the experiments can be found in appendix E, experiments CD.2.1 - CD.2.6.

6.3.2 Experiments with NLDA $Y = A^T X + f(X)$ Derived Features

With this approach, there is again the number of hidden units as additional parameter. To keep the number of experiments managable, only experiments with 2376 drift vectors and 5, 10, and 20 hidden units were done.

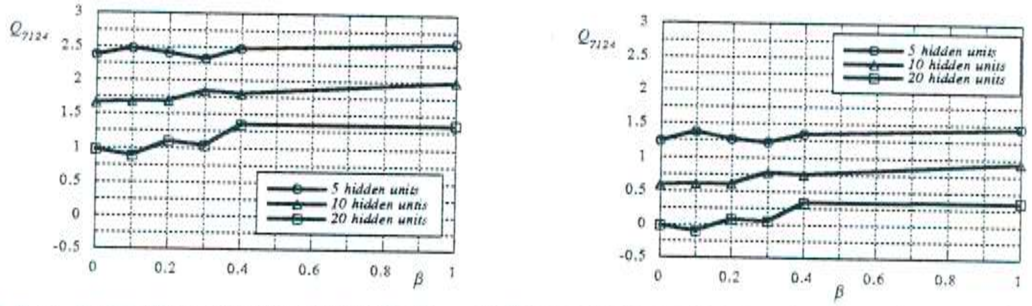


Figure 6.5: Seperation quality measure Q_{7124}'s dependence on the moving target parameter β on the training set (left) and test set (right). It holds always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

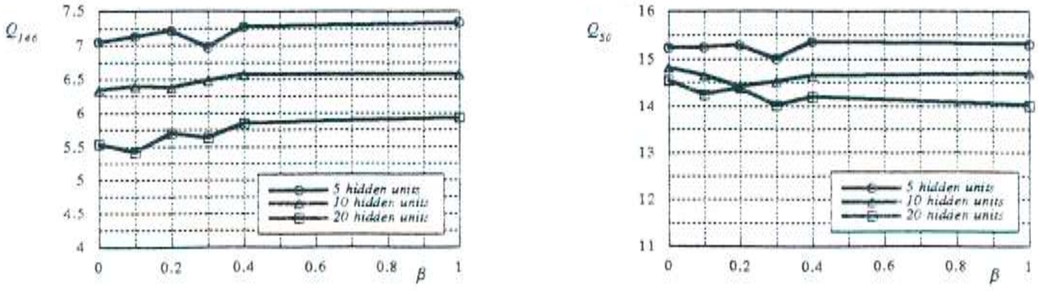


Figure 6.6: Seperation quality measures Q_{146}'s and Q_{50}'s dependence on the moving target parameter β on the training set. Again it is always $\gamma = 1$ and $\alpha = 0.9$ (except for $\beta = 0$ where $\alpha = 1.0$).

Figure 6.5 shows the evolution of the seperability measure Q_{7124} as function of the moving target parameter β on both the training set and the test set. As in the monophone case, there is not a strong dependence of Q_{7124} from β. However, as in the monophone case, Q_{7124} decreases with the number of hidden units in the network. In all experiments, the values for Q_{7124} are below the value in the LDA case.

Figure 6.6 shows Q_{146} and Q_{50} as functions of β on training set. The values for these measure are also below the values in the LDA case.

number of	parameter					word recognition	insertions	deletions	substitutions
hidden units	β	α	γ	*TRcbfact*	*ac*	performance			
5	0.0	1.0	1	2.0	4.0	89.8% (91.8%)	2.0%	1.6%	6.6%
	0.1	0.9	1	7.5	6.0	90.0% (92.3%)	2.3%	2.3%	5.3%
	0.2	0.9	1	10.0	8.0	90.2% (92.5%)	2.3%	1.2%	6.2%
	0.3	0.9	1	15.0	6.0	91.3% (93.6%)	2.3%	1.8%	4.6%
	0.4	0.9	1	2.0	6.0	89.8% (92.2%)	2.3%	1.6%	6.2%
	1.0	0.9	1	2.0	10.0	90.7% (93.4%)	2.7%	1.1%	5.5%
10	0.0	1.0	1	20.0	4.0	90.2% (91.8%)	1.6%	2.1%	6.1%
	0.1	0.9	1	10.0	8.0	89.7% (92.5%)	2.9%	1.8%	5.7%
	0.2	0.9	1	15.0	8.0	91.1% (93.6%)	2.5%	1.4%	5.0%
	0.3	0.9	1	7.5	8.0	91.1% (92.7%)	1.6%	2.1%	5.2%
	0.4	0.9	1	5.0	10.0	89.5% (91.6%)	2.1%	1.8%	6.6%
	1.0	0.9	1	2.0	6.0	90.2% (92.9%)	2.7%	1.8%	5.3%
20	0.0	1.0	1	2.0	4.0	89.3% (91.4%)	2.1%	2.1%	6.4%
	0.1	0.9	1	2.0	4.0	90.4% (92.2%)	2.5%	1.6%	5.5%
	0.2	0.9	1	20.0	8.0	89.8% (92.2%)	2.4%	2.0%	5.9%
	0.3	0.9	1	20.0	8.0	91.1% (93.0%)	1.9%	1.8%	5.2%
	0.4	0.9	1	20.0	4.0	90.4% (91.6%)	1.2%	2.7%	5.7%
	1.0	0.9	1	2.0	6.0	89.5% (91.4%)	2.0%	2.3%	6.2%

Table 6.3: Word recognition performance for the best values of *TRcbfact* and *ac*.

Table 6.3 shows the best word recognition performance, insertions, deletions and substitutions for all experiments. Again, for the best experiment the word recognition performance is only slightly better than with LDA derived features. The improvement in word recognition performance here is 0.4% or 4.4% relative improvement in the word error rate.

Each experiment is described in detail in appendix E, experiments CD.3.1 - CD.5.6.

6.4 Summary

The following table shows the best experiments together with the correpsonding parameter set for each approach.

feature extraction algorithm	parameter								word recognition performance
	hidden units	centroids in training	drift vectors	β	α	γ	*TRcbfact*	*ac*	
MSC + Delta	-	-	-	-	-	-	1.0	1.0	86.8% (89.3%)
LDA	-	7124	-	-	-	-	4.0	8.0	90.9% (93.8%)
NLDA $Y = A^T f(X)$	32	7124	50	0.1	0.9	1	10.0	6.0	91.1% (92.7%)
NLDA $Y = A^T X + f(X)$	5	7124	2376	0.3	0.9	1	15.0	6.0	91.3% (93.6%)

As in the context independent case, the achieved improvements in the word recognition performance are a little disappointing. One would have hoped for a little more since the class seperability was optimezed well by NLDA. As reasons for this, the in chapter 5 discussed ones can be seen.

Some ideas for improving the NLDA concept in a way that results also in a stronger improvement in word recognition performance are disscussed in chapter 8.

Chapter 7

Final Experiments on the Official Test Set

In the chapters 5 and 6, experiments with the cross validation set in the testing have been carried out. Therefore, it was possible to do many experiments in a managable time. With the parameter sets of the best experiments on the cross validation set, experiments on the 390 sentence test set have been made. There have been made 6 experiments in the context independent environment, 1 with the primary features, one with LDA derived features, and 4 with different NLDA approaches. In the context dependent case 4 experiments were made, 1 with the primary features, one with LDA, and 2 with NLDA derived features.

7.1 Experiments on the Official Test Set in Context Independent Environment

The following table shows the results of the epxeriments with context independent acoustic modeling.

feature extraction algorithm	parameter								word recognition performance
	hidden units	centroids in training	drift vectors	β	α	γ	$TRcbfact$	ac	
MSC + Delta	-	-	-	-	-	-	1.0	1.0	73.1% (74.4%)
LDA	-	146	-	-	-	-	8.0	7.0	85.4% (87.6%)
NLDA $Y = Af(X)$	32	146	50	0.1	0.9	1	20.0	26.0	81.7% (84.0%)
NLDA $Y = AX + f(X)$	20	146	50	0.3	0.9	1	15.0	18.0	84.2% (86.2%)
NLDA $Y = Af(X)$ multimodal approach	32	2451	50	0.1	0.9	1	10.0	12.0	84.6% (86.7%)
NLDA $Y = AX + f(X)$ multimodal approach	10	2451	50	0.0	1.0	1	10.0	8.0	85.7% (87.4%)

On this set the improvements with NLDA on the cross validation set are not verified. Only one experiment is slightly better than the LDA experiment. However, it is to be kept in mind, that the parameter for the experiments were searched on the cross validation set and do not have to be optimal on the test set. There are probably parameter combinations for which the word recognition performance on the test set is higher than with the parameter combinations used here.

Again, the experiments with multimodal NLDA work better than with unimodal NLDA.

7.2 Experiments on the Official Test Set in Context Dependent Environment

The following table shows the results of the experiments on the 390 sentence test set with the best parameter sets investigated on the cross validation set.

feature extraction algorithm	parameter								word recognition performance
	hidden units	centroids in training	drift vectors	β	α	γ	*TRcbfact*	*ac*	
MSC + Delta	-	-	-	-	-	-	1.0	1.0	87.9% (89.9%)
LDA	-	7124	-	-	-	-	4.0	8.0	92.8% (94.4%)
NLDA $Y = Af(X)$	32	7124	50	0.1	0.9	1	10.0	6.0	91.3% (92.7%)
NLDA $Y = AX + f(X)$	5	7124	2376	0.3	0.9	1	15.0	6.0	92.2% (93.1%)

As in the context independent case, the improvements achieved with NLDA on the cross validation set could not be verified on the test set. Again, it is to say, that the parameters investigated on the cross validation set do not have to be optimal for the test set as well.

Chapter 8

Summary and Conclusion

In this thesis it was shown how to derive a statistically based nonlinear signal transformation. The transformation reduces the dimensionality of feature vectors and improves in the same step the class seperability in the feature space.

In order to derive this transformation, it was shown how to find a measure for class seperabiblity (chapter 3). Based on this measure, it was made a linear optimization first. The result of this was the well known Linear Discriminant Analysis (LDA).

In the next step it was shown, how to derive a nonlinear transformation by a simultaneous two stage nonlinear optimization of this measure (chapter 4). Since the derivation of the transformation is based on a measure for the class seperability, the transformation is called NonLinear Discriminant Analysis (NLDA).

The transformation's ability to improve class seperability was graphically shown on a small example and compared to LDA's ability.

In order to investigate whether the new transformation is also able to improve word recognition performance, experiments with a continuous speech recognizer and the large vocabulary database have been carried out. Different NLDA networks and approches have been applied in this experiments. The experiments were made in a phoneme context independent environment (chapter 5) and in a phoneme context dependent environment (chapter 6). The results of the experiments have shown, that the class seperability could be improved also on a large database. However, this did not hold for the word recognition performance. There, only a slight improvement in some experiments could be achieved.

Reasons for this have been searched in the teamwork between feature extraction and recognizer but also in the NLDA concept. These reasons were discussed (chapter 5).

Summarizingly can be said, that the NLDA as feature extraction gives rather disappointing results compared with LDA. There have only been slight improvements in the best cases but due to the higher dimensional parameter space NLDA is much more expensive. Altogether, NLDA is still too immature to be applied as a front end feature extraction for a recognizer.

Future Work

There are two major things which could be tried to improve the performance of NLDA as feature extraction also on the word recognition level. In the formula (4.17) for the backpropagations's target calculation, the parameter β was introduced. Its function is a weighting of the drift vector in the sum. Since it is only a single factor, all coefficients of the drift vector are weighted by the same value. However, due to the initial LDA state of the NLDA network, the channels are already sortet by their importance for classification. Thus, it would make sense to have a small weighting factor for the channels in which the class seperability is already good and a large weighting factor for the channels in which the class seperability is not as good. Since the seperability in a channel is measurable by the corresponding eigenvalue, the drift vector weighting could be adapted considering the ratio of the eigenvalue in a channel and the eigenvalue in the first channel. Thus, (4.17) would be modified to

$$
\begin{aligned}
T^{\nu,\mu} &= [t_1^{\nu,\mu}, \ldots, t_m^{\nu,\mu}] \\
t_i^{\nu,\mu} &= \alpha m_i^\nu + (1-\alpha) y_i^\mu + \beta \frac{\lambda_1}{\lambda_i} d_i^\nu \quad i = 1, \ldots, m.
\end{aligned}
\tag{8.1}
$$

In (8.1), m_i^ν is the i-th component of the ν-th class centroid M^ν, y_i^μ is the i-th component of the network output Y^μ for the μ-th sample vector X^μ, and d_i^ν is the i-th component of the drift vector D^ν for the class ν. λ_i is the eigenvalue for the i-th channel and λ_1 is the eigenvalue for the first channel.

This procedure gives more drift in channels with bad seperability (small eigenvalue) and less drift in channels with already good seperability.

The second major thing is to introduce psychoacoustical knowledge into NLDA. It might be useful to give different classes different weighted drift vectors, depending on the psychoacoustical affinity of these classes. For example one could introduce superclasses like vowels, nasals, etc. and let them drift apart from each other. That would preserve relationships within these superclasses. Another possibility to derive an affinity knowledge is to create a frame by frame class substitution statistic. Once one has derived such a statistic, one could apply drift vectors only for often substituted classes.

Both these approaches have to be investigated in a further work.

Appendix A

Proof of Theorem 2.1

1. With $\tilde{S}_b = T^T S_b T$ and $\tilde{S}_w = T^T S_w T$ one gets

$$|\tilde{S}_w^{-1}\tilde{S}_b| = |\tilde{S}_w^{-1}||\tilde{S}_b| = \frac{|\tilde{S}_b|}{|\tilde{S}_w|} = \frac{|T^T S_b T|}{|T^T S_w T|} = \frac{|T|^2|S_b|}{|T|^2|S_w|} = \frac{|S_b|}{|S_w|} = |S_w^{-1}S_b|.$$

   This proves part 1 of the theorem.

2. With

$$Y = f(X) = [c_1 x_1, \ldots, c_n x_n]^T \qquad c_i \in \mathbf{R}, c_i \neq 0,$$

   one gets for the class means in the transformed space

$$\tilde{M}_i = [(\tilde{m}_1)_i, \ldots, (\tilde{m}_n)_i] = [c_1(m_1)_i, \ldots, c_n(m_n)_i] \qquad (i = 0, \ldots, L).$$

   Then the elements of the within-class scatter matrix are

$$\begin{aligned}
(\tilde{s}_{ij})_w &= \frac{1}{N}\sum_{k=1}^{L}\left(\sum_{l=1}^{N_i}(y_i)_l^k(y_j)_l^k - N_i(\tilde{m}_i)_k(\tilde{m}_j)_k\right)\\
&= c_i c_j \frac{1}{N}\sum_{k=1}^{L}\left(\sum_{l=1}^{N_i}(x_i)_l^k(x_j)_l^k - N_i(m_i)_k(m_j)_k\right)\\
&= c_i c_j (s_{ij})_w .
\end{aligned}$$

   For the elements of the between-class scatter one finds

$$\begin{aligned}
(\tilde{s}_{ij})_b &= \sum_{k=1}^{L}\frac{N_i}{N}(\tilde{m}_i)_k(\tilde{m}_j)_k - \frac{N_i}{N}(\tilde{m}_i)_0(\tilde{m}_0)_k\\
&= c_i c_j\left(\sum_{k=1}^{L}\frac{N_i}{N}(m_i)_k(m_j)_k - \frac{N_i}{N}(m_i)_0(m_0)_k\right)\\
&= c_i c_j (s_{ij})_b .
\end{aligned}$$

   For the determinant of $\tilde{S}_b$ follows then

$$\begin{aligned}
|\tilde{S}_b| &= \begin{vmatrix} (\tilde{s}_{11})_b & (\tilde{s}_{12})_b & \cdots & (\tilde{s}_{1n})_b \\ (\tilde{s}_{21})_b & (\tilde{s}_{22})_b & \cdots & (\tilde{s}_{2n})_b \\ \vdots & & & \vdots \\ (\tilde{s}_{n1})_b & (\tilde{s}_{n2})_b & \cdots & (\tilde{s}_{nn})_b \end{vmatrix}\\
&= \begin{vmatrix} c_1^2(s_{11})_b & c_1 c_2(s_{12})_b & \cdots & c_1 c_n(s_{1n})_b \\ c_2 c_1(s_{21})_b & c_2^2(s_{22})_b & \cdots & c_2 c_n(s_{2n})_b \\ \vdots & & & \vdots \\ c_n c_1(s_{n1})_b & c_n c_2(s_{n2})_b & \cdots & c_n^2(s_{nn})_b \end{vmatrix}
\end{aligned}$$

$$
\begin{aligned}
&= \begin{vmatrix} c_1(s_{11})_b & c_2(s_{12})_b & \cdots & c_n(s_{1n})_b \\ c_1(s_{21})_b & c_2(s_{22})_b & \cdots & c_n(s_{2n})_b \\ \vdots & & & \vdots \\ c_1(s_{n1})_b & c_2(s_{n2})_b & \cdots & c_n(s_{nn})_b \end{vmatrix} \cdot \prod_{i=1}^{n} c_i \\
&= \begin{vmatrix} (s_{11})_b & (s_{12})_b & \cdots & (s_{1n})_b \\ (s_{21})_b & (s_{22})_b & \cdots & (s_{2n})_b \\ \vdots & & & \vdots \\ (s_{n1})_b & (s_{n2})_b & \cdots & (s_{nn})_b \end{vmatrix} \cdot \left(\prod_{i=1}^{n} c_i\right)^2 \\
&= |S_b| \cdot \left(\prod_{i=1}^{n} c_i\right)^2
\end{aligned}
$$

and for the determinant of $\tilde{S}_w$ one gets in the same way

$$|\tilde{S}_w| = |S_w| \cdot \left(\prod_{i=1}^{n} c_i\right)^2 .$$

Therefore,

$$|\tilde{S}_w^{-1}\tilde{S}_b| = \frac{|\tilde{S}_b|}{|\tilde{S}_w|} = \frac{\left(\prod_{i=1}^{n} c_i\right)^2 \cdot |S_b|}{\left(\prod_{i=1}^{n} c_i\right)^2 \cdot |S_w|} = |S_w^{-1}S_b|.$$

This proves part 2 of the theorem.

3. The proof of part 3 follows immediatly from the invariance of covariance matrices against coordinate shifts.

Thus, theorem 2.1 is proved.

Appendix B

Test Data for the Examples

- class /AA/:

channel															
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.627	0.784	0.945	0.996	0.961	0.875	0.616	0.569	0.651	0.667	0.537	0.631	0.482	0.486	0.514	0.502
0.620	0.773	0.937	1.000	0.937	0.867	0.604	0.545	0.537	0.643	0.518	0.620	0.482	0.478	0.565	0.494
0.620	0.757	0.929	0.992	0.918	0.859	0.600	0.529	0.565	0.627	0.518	0.639	0.486	0.498	0.588	0.514
0.608	0.749	0.925	0.984	0.910	0.843	0.576	0.506	0.525	0.647	0.525	0.635	0.482	0.475	0.588	0.522
0.608	0.741	0.922	0.976	0.902	0.835	0.569	0.518	0.529	0.655	0.514	0.616	0.467	0.514	0.569	0.482
0.604	0.753	0.910	0.898	0.855	0.769	0.514	0.439	0.482	0.608	0.431	0.490	0.435	0.412	0.467	0.475
0.600	0.753	0.914	0.925	0.863	0.722	0.502	0.431	0.475	0.596	0.416	0.467	0.400	0.392	0.451	0.471
0.604	0.761	0.918	0.941	0.890	0.725	0.545	0.463	0.475	0.624	0.459	0.431	0.380	0.388	0.447	0.478
0.616	0.773	0.914	0.945	0.914	0.718	0.541	0.455	0.459	0.569	0.467	0.475	0.427	0.439	0.522	0.467
0.631	0.780	0.914	0.957	0.914	0.749	0.604	0.522	0.506	0.643	0.494	0.533	0.416	0.490	0.522	0.424
0.635	0.784	0.906	0.949	0.914	0.737	0.596	0.518	0.514	0.608	0.494	0.522	0.420	0.467	0.533	0.475
0.635	0.780	0.894	0.957	0.910	0.745	0.616	0.529	0.537	0.631	0.510	0.522	0.447	0.565	0.565	0.494
0.639	0.804	0.973	0.925	0.922	0.875	0.753	0.776	0.663	0.525	0.502	0.557	0.388	0.349	0.498	0.549
0.631	0.780	0.961	0.937	0.906	0.855	0.698	0.718	0.647	0.467	0.475	0.529	0.404	0.373	0.529	0.541
0.620	0.784	0.933	0.929	0.898	0.851	0.694	0.725	0.596	0.431	0.424	0.482	0.369	0.294	0.522	0.537
0.612	0.784	0.933	0.937	0.925	0.886	0.710	0.733	0.576	0.447	0.463	0.525	0.392	0.251	0.486	0.522
0.522	0.549	0.482	0.541	0.475	0.482	0.349	0.380	0.349	0.169	0.200	0.216	0.220	0.212	0.392	0.357
0.620	0.859	0.918	0.773	0.773	0.847	0.749	0.722	0.592	0.541	0.576	0.349	0.443	0.314	0.310	0.443
0.588	0.769	0.957	1.000	0.894	0.902	0.682	0.616	0.647	0.675	0.604	0.655	0.608	0.475	0.565	0.565
0.604	0.776	0.941	0.992	0.929	0.929	0.686	0.627	0.675	0.702	0.612	0.659	0.600	0.478	0.576	0.576
0.620	0.780	0.929	0.984	0.922	0.910	0.686	0.635	0.686	0.714	0.608	0.675	0.576	0.482	0.561	0.580
0.627	0.780	0.910	0.957	0.922	0.902	0.725	0.651	0.706	0.686	0.604	0.616	0.525	0.424	0.537	0.502
0.557	0.702	0.769	0.733	0.784	0.776	0.588	0.514	0.525	0.506	0.490	0.502	0.345	0.322	0.357	0.392
0.565	0.686	0.745	0.729	0.788	0.741	0.580	0.514	0.518	0.506	0.467	0.478	0.361	0.365	0.408	0.404
0.663	0.796	0.898	0.867	0.890	0.914	0.800	0.667	0.706	0.694	0.561	0.561	0.439	0.412	0.529	0.494
0.655	0.804	0.918	0.875	0.914	0.929	0.765	0.620	0.627	0.624	0.533	0.561	0.416	0.400	0.537	0.478
0.651	0.804	0.925	0.875	0.925	0.941	0.796	0.635	0.647	0.604	0.514	0.561	0.478	0.478	0.557	0.455
0.647	0.796	0.933	0.863	0.910	0.937	0.804	0.635	0.639	0.600	0.514	0.580	0.455	0.439	0.541	0.447
0.639	0.804	0.937	0.835	0.894	0.925	0.835	0.655	0.631	0.576	0.549	0.612	0.506	0.447	0.533	0.467
0.643	0.820	0.949	0.847	0.898	0.925	0.867	0.686	0.667	0.612	0.576	0.651	0.553	0.518	0.545	0.537
0.655	0.808	0.949	0.827	0.875	0.918	0.867	0.663	0.647	0.600	0.569	0.647	0.553	0.502	0.537	0.525
0.655	0.824	0.941	0.800	0.843	0.898	0.867	0.667	0.624	0.612	0.545	0.639	0.573	0.514	0.525	0.529
0.624	0.761	0.933	0.984	0.839	0.898	0.757	0.627	0.722	0.631	0.549	0.592	0.502	0.388	0.522	0.459
0.624	0.769	0.922	0.988	0.851	0.925	0.698	0.639	0.741	0.710	0.569	0.596	0.533	0.439	0.549	0.490
0.631	0.776	0.918	0.992	0.851	0.910	0.706	0.647	0.722	0.718	0.576	0.596	0.549	0.435	0.498	0.475
0.635	0.788	0.918	0.996	0.875	0.914	0.741	0.651	0.702	0.706	0.576	0.592	0.529	0.443	0.478	0.541
0.639	0.796	0.925	0.996	0.890	0.914	0.729	0.655	0.702	0.722	0.584	0.616	0.553	0.502	0.588	0.553
0.624	0.757	0.922	1.000	0.933	0.929	0.667	0.596	0.706	0.663	0.545	0.612	0.522	0.455	0.486	0.514
0.631	0.749	0.910	0.996	0.929	0.922	0.655	0.573	0.667	0.671	0.522	0.604	0.522	0.455	0.549	0.447
0.631	0.765	0.902	0.988	0.922	0.898	0.655	0.549	0.655	0.698	0.529	0.600	0.522	0.482	0.549	0.447
0.639	0.784	0.914	0.988	0.933	0.886	0.667	0.557	0.643	0.714	0.541	0.604	0.522	0.490	0.580	0.522
0.643	0.788	0.937	0.992	0.937	0.867	0.659	0.584	0.663	0.722	0.553	0.620	0.549	0.482	0.557	0.522
0.651	0.780	0.922	0.976	0.933	0.886	0.659	0.549	0.620	0.671	0.541	0.600	0.494	0.557	0.533	0.475
0.655	0.788	0.914	0.973	0.953	0.890	0.663	0.557	0.604	0.678	0.549	0.616	0.498	0.533	0.573	0.490
0.647	0.792	0.922	0.976	0.945	0.859	0.655	0.557	0.600	0.702	0.553	0.631	0.529	0.525	0.580	0.522
0.639	0.792	0.929	0.988	0.918	0.855	0.647	0.576	0.620	0.690	0.565	0.643	0.553	0.592	0.612	0.533
0.631	0.788	0.929	1.000	0.929	0.894	0.651	0.592	0.627	0.706	0.561	0.635	0.553	0.471	0.627	0.498
0.624	0.784	0.918	0.996	0.953	0.906	0.678	0.620	0.659	0.718	0.565	0.635	0.569	0.510	0.573	0.506
0.624	0.780	0.902	0.992	0.941	0.867	0.667	0.620	0.659	0.729	0.553	0.624	0.569	0.522	0.569	0.541
0.624	0.776	0.906	0.996	0.902	0.867	0.682	0.616	0.663	0.725	0.557	0.624	0.557	0.553	0.584	0.529

- class /AH/:

channel															
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.643	0.820	0.984	0.769	0.694	0.765	0.835	0.655	0.620	0.769	0.518	0.541	0.663	0.576	0.624	0.576
0.651	0.824	0.965	0.784	0.714	0.776	0.831	0.663	0.647	0.737	0.514	0.541	0.651	0.533	0.624	0.584
0.616	0.749	0.886	0.820	0.678	0.651	0.773	0.729	0.604	0.698	0.482	0.490	0.482	0.451	0.510	0.475
0.616	0.753	0.902	0.816	0.682	0.631	0.745	0.675	0.549	0.635	0.467	0.486	0.541	0.533	0.576	0.525
0.631	0.784	0.914	0.820	0.702	0.667	0.757	0.737	0.690	0.647	0.502	0.557	0.518	0.447	0.506	0.427
0.624	0.788	0.906	0.843	0.729	0.718	0.792	0.737	0.710	0.604	0.522	0.569	0.541	0.494	0.537	0.463
0.616	0.776	0.922	0.859	0.737	0.741	0.831	0.698	0.690	0.584	0.502	0.569	0.553	0.494	0.537	0.463
0.600	0.757	0.898	0.839	0.710	0.714	0.780	0.710	0.651	0.671	0.549	0.557	0.541	0.400	0.471	0.412
0.612	0.757	0.890	0.867	0.714	0.725	0.816	0.706	0.663	0.710	0.580	0.584	0.561	0.494	0.584	0.545
0.616	0.749	0.882	0.867	0.706	0.702	0.780	0.694	0.655	0.702	0.565	0.565	0.545	0.494	0.588	0.537
0.651	0.784	0.816	0.753	0.702	0.647	0.714	0.761	0.690	0.710	0.612	0.616	0.510	0.475	0.494	0.494
0.627	0.776	0.831	0.753	0.702	0.667	0.710	0.776	0.702	0.729	0.616	0.616	0.494	0.443	0.494	0.455
0.541	0.667	0.863	0.773	0.639	0.616	0.749	0.733	0.588	0.702	0.565	0.557	0.541	0.447	0.510	0.376
0.569	0.710	0.867	0.796	0.690	0.671	0.796	0.773	0.659	0.741	0.604	0.596	0.588	0.486	0.510	0.373
0.596	0.804	0.906	0.757	0.682	0.745	0.820	0.682	0.702	0.714	0.588	0.643	0.596	0.451	0.451	0.420

0.573	0.780	0.922	0.761	0.686	0.776	0.824	0.682	0.694	0.710	0.596	0.655	0.643	0.510	0.502	0.467
0.604	0.769	0.882	0.745	0.745	0.686	0.804	0.749	0.592	0.576	0.584	0.588	0.502	0.412	0.490	0.475
0.604	0.765	0.878	0.733	0.729	0.659	0.765	0.769	0.608	0.596	0.616	0.600	0.525	0.459	0.475	0.475
0.624	0.753	0.855	0.851	0.710	0.698	0.784	0.784	0.710	0.733	0.584	0.584	0.608	0.471	0.576	0.475
0.612	0.733	0.863	0.843	0.706	0.694	0.808	0.757	0.702	0.737	0.576	0.588	0.596	0.475	0.576	0.463
0.584	0.702	0.898	0.835	0.675	0.667	0.820	0.686	0.671	0.718	0.510	0.557	0.569	0.459	0.522	0.420
0.678	0.812	0.922	0.894	0.788	0.722	0.796	0.820	0.682	0.698	0.580	0.580	0.604	0.506	0.596	0.529
0.663	0.804	0.945	0.910	0.796	0.725	0.776	0.773	0.686	0.745	0.643	0.624	0.627	0.459	0.557	0.541
0.655	0.788	0.941	0.925	0.769	0.694	0.741	0.796	0.686	0.776	0.667	0.631	0.639	0.514	0.604	0.549
0.639	0.769	0.937	0.933	0.745	0.678	0.749	0.816	0.702	0.788	0.667	0.620	0.620	0.510	0.592	0.518
0.635	0.753	0.929	0.937	0.718	0.659	0.737	0.824	0.698	0.788	0.663	0.600	0.608	0.510	0.576	0.541
0.639	0.761	0.918	0.933	0.714	0.663	0.729	0.820	0.702	0.784	0.643	0.608	0.608	0.482	0.580	0.514
0.643	0.773	0.906	0.929	0.737	0.690	0.753	0.812	0.722	0.773	0.643	0.620	0.600	0.494	0.596	0.518
0.651	0.784	0.898	0.922	0.761	0.710	0.769	0.792	0.725	0.761	0.631	0.627	0.596	0.502	0.616	0.569
0.659	0.784	0.882	0.882	0.761	0.722	0.776	0.804	0.753	0.753	0.576	0.565	0.537	0.486	0.573	0.545
0.678	0.804	0.973	0.910	0.745	0.686	0.725	0.851	0.812	0.773	0.686	0.671	0.753	0.600	0.624	0.616
0.682	0.820	0.969	0.910	0.765	0.714	0.749	0.847	0.827	0.792	0.678	0.659	0.741	0.514	0.675	0.627
0.682	0.816	0.969	0.918	0.776	0.729	0.773	0.890	0.831	0.812	0.686	0.639	0.694	0.525	0.678	0.576
0.678	0.800	0.976	0.925	0.765	0.714	0.765	0.894	0.800	0.816	0.651	0.616	0.890	0.608	0.678	0.580
0.678	0.788	0.973	0.933	0.761	0.698	0.757	0.886	0.749	0.773	0.627	0.608	0.667	0.651	0.682	0.576
0.682	0.800	0.965	0.929	0.780	0.718	0.780	0.855	0.745	0.765	0.592	0.565	0.643	0.569	0.647	0.592
0.682	0.816	0.961	0.929	0.800	0.745	0.831	0.855	0.741	0.765	0.612	0.561	0.643	0.604	0.671	0.569
0.678	0.812	0.969	0.933	0.800	0.745	0.847	0.839	0.706	0.722	0.569	0.537	0.635	0.635	0.706	0.580
0.675	0.792	0.973	0.925	0.780	0.733	0.847	0.784	0.663	0.694	0.545	0.541	0.627	0.643	0.678	0.533
0.675	0.784	0.965	0.910	0.765	0.718	0.867	0.729	0.616	0.702	0.506	0.486	0.592	0.663	0.639	0.537
0.631	0.773	0.925	0.820	0.690	0.714	0.820	0.678	0.624	0.714	0.624	0.576	0.667	0.533	0.627	0.561
0.620	0.749	0.910	0.784	0.682	0.706	0.831	0.647	0.608	0.725	0.604	0.588	0.639	0.467	0.557	0.514
0.616	0.800	0.906	0.753	0.729	0.671	0.710	0.788	0.678	0.675	0.612	0.635	0.627	0.451	0.490	0.455
0.624	0.780	0.894	0.741	0.722	0.659	0.725	0.753	0.620	0.627	0.565	0.592	0.580	0.388	0.490	0.478
0.580	0.725	0.843	0.808	0.678	0.655	0.773	0.796	0.675	0.706	0.529	0.557	0.537	0.522	0.443	0.412
0.604	0.710	0.784	0.761	0.655	0.627	0.710	0.725	0.631	0.639	0.510	0.525	0.486	0.451	0.416	0.373
0.706	0.851	0.933	0.922	0.851	0.796	0.859	0.871	0.796	0.804	0.690	0.694	0.667	0.698	0.698	0.624
0.690	0.839	0.969	0.957	0.863	0.831	0.906	0.898	0.831	0.847	0.710	0.729	0.702	0.698	0.741	0.706
0.663	0.788	0.886	0.875	0.765	0.741	0.776	0.678	0.647	0.706	0.584	0.604	0.612	0.463	0.557	0.478
0.655	0.784	0.894	0.875	0.773	0.776	0.796	0.729	0.682	0.733	0.631	0.643	0.651	0.471	0.586	0.537

- class /AX/:

channel															
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.639	0.816	0.992	0.890	0.859	0.780	0.624	0.561	0.580	0.663	0.510	0.659	0.580	0.471	0.467	0.510
0.635	0.800	1.000	0.875	0.808	0.878	0.671	0.584	0.612	0.675	0.529	0.659	0.576	0.467	0.518	0.580
0.671	0.831	0.953	0.894	0.843	0.882	0.765	0.631	0.647	0.741	0.643	0.667	0.714	0.580	0.635	0.655
0.671	0.827	0.949	0.902	0.851	0.894	0.773	0.627	0.651	0.733	0.655	0.675	0.694	0.573	0.612	0.631
0.537	0.702	0.792	0.694	0.741	0.737	0.522	0.455	0.455	0.443	0.412	0.412	0.322	0.329	0.365	0.322
0.545	0.718	0.800	0.718	0.753	0.761	0.576	0.494	0.482	0.478	0.420	0.439	0.294	0.290	0.306	0.337
0.655	0.824	0.941	0.808	0.875	0.847	0.616	0.537	0.620	0.702	0.541	0.655	0.596	0.482	0.529	0.549
0.667	0.839	0.937	0.816	0.808	0.824	0.627	0.569	0.596	0.635	0.549	0.663	0.604	0.498	0.537	0.529
0.647	0.796	0.922	0.875	0.812	0.804	0.584	0.463	0.482	0.525	0.482	0.604	0.494	0.349	0.459	0.478
0.651	0.784	0.937	0.871	0.784	0.831	0.620	0.502	0.510	0.545	0.478	0.604	0.522	0.431	0.478	0.525
0.643	0.800	0.925	0.839	0.773	0.788	0.635	0.525	0.486	0.620	0.604	0.588	0.604	0.502	0.522	0.451
0.639	0.784	0.910	0.847	0.824	0.820	0.596	0.522	0.486	0.596	0.596	0.584	0.576	0.490	0.486	0.471
0.667	0.847	0.953	0.827	0.765	0.812	0.749	0.596	0.522	0.686	0.714	0.659	0.773	0.494	0.576	0.616
0.663	0.835	0.992	0.878	0.808	0.847	0.761	0.643	0.596	0.655	0.663	0.678	0.773	0.541	0.600	0.643
0.659	0.804	0.988	0.906	0.792	0.863	0.737	0.627	0.584	0.675	0.686	0.651	0.757	0.545	0.604	0.592
0.651	0.792	0.961	0.902	0.765	0.851	0.753	0.620	0.576	0.659	0.690	0.635	0.718	0.518	0.565	0.624
0.647	0.808	0.949	0.902	0.784	0.843	0.757	0.639	0.592	0.639	0.647	0.655	0.714	0.537	0.588	0.608
0.651	0.816	0.957	0.902	0.800	0.878	0.749	0.651	0.608	0.655	0.675	0.667	0.718	0.553	0.600	0.627
0.639	0.808	0.969	0.898	0.804	0.878	0.749	0.655	0.624	0.671	0.686	0.667	0.706	0.561	0.612	0.608
0.627	0.792	0.976	0.894	0.788	0.859	0.757	0.643	0.608	0.675	0.698	0.663	0.702	0.541	0.596	0.616
0.620	0.776	0.973	0.898	0.773	0.851	0.733	0.616	0.596	0.655	0.686	0.667	0.686	0.514	0.565	0.620
0.608	0.761	0.965	0.894	0.757	0.843	0.682	0.608	0.604	0.671	0.694	0.659	0.678	0.510	0.553	0.600
0.608	0.753	0.957	0.890	0.749	0.855	0.651	0.604	0.604	0.702	0.710	0.678	0.686	0.533	0.553	0.573
0.620	0.749	0.945	0.882	0.745	0.859	0.643	0.600	0.604	0.694	0.694	0.690	0.686	0.537	0.541	0.565
0.624	0.761	0.937	0.871	0.749	0.855	0.635	0.592	0.600	0.706	0.682	0.694	0.675	0.545	0.545	0.561
0.588	0.769	0.859	0.824	0.886	0.671	0.557	0.506	0.510	0.522	0.533	0.522	0.435	0.404	0.349	0.416
0.643	0.800	0.953	0.824	0.769	0.863	0.820	0.631	0.592	0.722	0.694	0.620	0.671	0.533	0.624	0.624
0.639	0.804	0.953	0.871	0.776	0.831	0.804	0.643	0.608	0.725	0.659	0.651	0.682	0.506	0.604	0.627
0.643	0.800	0.949	0.886	0.780	0.855	0.792	0.655	0.627	0.733	0.647	0.671	0.675	0.561	0.631	0.592
0.620	0.824	0.890	0.714	0.663	0.714	0.710	0.565	0.514	0.667	0.569	0.569	0.643	0.541	0.518	0.514
0.678	0.851	0.886	0.749	0.733	0.769	0.549	0.427	0.376	0.569	0.463	0.675	0.565	0.412	0.447	0.482
0.690	0.847	0.863	0.737	0.694	0.753	0.592	0.451	0.408	0.498	0.451	0.624	0.569	0.502	0.459	0.471
0.639	0.827	0.941	0.843	0.953	0.843	0.596	0.506	0.557	0.565	0.529	0.529	0.404	0.475	0.482	0.486
0.647	0.816	0.937	0.804	0.937	0.867	0.635	0.529	0.573	0.569	0.522	0.537	0.408	0.451	0.475	0.475
0.647	0.816	0.996	0.902	0.859	0.894	0.737	0.651	0.694	0.769	0.588	0.710	0.659	0.549	0.569	0.596
0.631	0.808	1.000	0.886	0.835	0.922	0.820	0.698	0.753	0.816	0.624	0.702	0.659	0.541	0.635	0.651
0.663	0.827	0.984	0.886	0.831	0.898	0.722	0.639	0.612	0.663	0.659	0.659	0.753	0.561	0.651	0.620
0.663	0.827	1.000	0.910	0.843	0.910	0.780	0.675	0.616	0.659	0.694	0.675	0.749	0.573	0.643	0.588
0.671	0.808	0.996	0.922	0.820	0.886	0.769	0.647	0.588	0.667	0.671	0.663	0.722	0.565	0.643	0.624
0.612	0.749	0.929	0.898	0.890	0.741	0.549	0.510	0.557	0.675	0.553	0.612	0.537	0.502	0.412	0.478
0.627	0.780	0.925	0.910	0.914	0.847	0.604	0.569	0.608	0.698	0.588	0.651	0.569	0.514	0.447	0.478
0.659	0.827	0.918	0.871	0.871	0.863	0.729	0.596	0.592	0.573	0.655	0.659	0.518	0.541	0.525	0.475
0.659	0.812	0.902	0.835	0.816	0.839	0.690	0.584	0.557	0.522	0.584	0.616	0.431	0.553	0.498	0.455
0.651	0.820	0.890	0.812	0.788	0.867	0.745	0.624	0.592	0.569	0.627	0.659	0.537	0.600	0.553	0.435
0.678	0.847	0.992	0.847	0.824	0.847	0.596	0.486	0.588	0.686	0.506	0.643	0.569	0.455	0.400	0.486
0.686	0.839	1.000	0.855	0.784	0.890	0.690	0.514	0.592	0.725	0.506	0.678	0.600	0.447	0.506	0.498
0.643	0.808	0.953	0.831	0.847	0.737	0.541	0.482	0.525	0.624	0.482	0.655	0.498	0.447	0.459	0.455
0.651	0.827	0.957	0.855	0.835	0.824	0.627	0.545	0.580	0.655	0.510	0.651	0.518	0.506	0.498	0.494
0.671	0.863	0.996	0.863	0.773	0.824	0.776	0.580	0.533	0.667	0.706	0.671	0.710	0.396	0.510	0.573
0.667	0.839	1.000	0.898	0.769	0.843	0.776	0.624	0.553	0.682	0.678	0.714	0.749	0.478	0.525	0.569

- class /AY/:

channel															
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0.608	0.788	1.000	0.976	0.945	0.855	0.639	0.604	0.682	0.620	0.537	0.616	0.498	0.459	0.514	0.502
0.592	0.765	0.996	0.980	0.945	0.855	0.616	0.569	0.682	0.651	0.518	0.600	0.475	0.447	0.459	0.514
0.561	0.761	0.992	0.941	0.945	0.824	0.655	0.710	0.749	0.533	0.545	0.569	0.537	0.447	0.506	0.569
0.580	0.741	0.973	0.929	0.953	0.788	0.620	0.671	0.753	0.525	0.541	0.561	0.522	0.459	0.431	0.486
0.647	0.890	0.945	0.898	0.651	0.522	0.498	0.471	0.476	0.459	0.451	0.482	0.451	0.435	0.439	0.443
0.643	0.886	0.945	0.914	0.627	0.467	0.396	0.388	0.408	0.337	0.318	0.420	0.384	0.322	0.357	0.388
0.561	0.757	0.827	0.820	0.639	0.490	0.514	0.467	0.565	0.565	0.451	0.337	0.349	0.325	0.510	0.502
0.557	0.749	0.816	0.804	0.655	0.467	0.471	0.424	0.467	0.451	0.298	0.314	0.200	0.204	0.310	0.373
0.576	0.737	0.773	0.784	0.588	0.424	0.361	0.357	0.380	0.263	0.263	0.310	0.263	0.227	0.322	0.353

0.569	0.725	0.796	0.796	0.663	0.471	0.420	0.416	0.416	0.306	0.267	0.357	0.263	0.263	0.357	0.396
0.631	0.839	0.902	0.898	0.686	0.494	0.471	0.416	0.408	0.510	0.455	0.459	0.431	0.353	0.388	0.482
0.627	0.831	0.902	0.898	0.694	0.494	0.443	0.376	0.396	0.525	0.431	0.459	0.404	0.298	0.373	0.518
0.627	0.847	0.925	0.906	0.698	0.522	0.475	0.427	0.451	0.557	0.431	0.478	0.416	0.376	0.420	0.545
0.639	0.843	0.925	0.898	0.690	0.498	0.427	0.333	0.412	0.557	0.380	0.478	0.357	0.314	0.380	0.549
0.639	0.867	0.929	0.867	0.710	0.514	0.431	0.400	0.471	0.522	0.443	0.514	0.408	0.341	0.392	0.557
0.647	0.882	0.969	0.878	0.733	0.529	0.443	0.408	0.518	0.522	0.424	0.455	0.431	0.325	0.396	0.561
0.655	0.875	0.980	0.894	0.804	0.573	0.494	0.420	0.576	0.435	0.482	0.424	0.435	0.302	0.365	0.533
0.659	0.847	0.933	0.925	0.773	0.545	0.455	0.443	0.580	0.439	0.400	0.443	0.392	0.318	0.294	0.302
0.663	0.855	0.949	0.929	0.824	0.569	0.502	0.494	0.616	0.447	0.424	0.482	0.427	0.353	0.345	0.341
0.655	0.808	0.898	0.875	0.741	0.557	0.518	0.482	0.576	0.612	0.310	0.475	0.439	0.318	0.255	0.416
0.671	0.816	0.890	0.890	0.737	0.573	0.525	0.478	0.569	0.627	0.333	0.478	0.467	0.306	0.290	0.416
0.686	0.812	0.878	0.890	0.725	0.573	0.522	0.471	0.569	0.616	0.318	0.431	0.467	0.337	0.278	0.412
0.690	0.808	0.871	0.886	0.710	0.576	0.518	0.490	0.565	0.600	0.329	0.416	0.471	0.286	0.306	0.447
0.694	0.788	0.859	0.878	0.686	0.569	0.522	0.451	0.533	0.565	0.373	0.427	0.459	0.337	0.345	0.424
0.690	0.776	0.855	0.843	0.718	0.565	0.486	0.435	0.525	0.565	0.345	0.400	0.435	0.298	0.294	0.361
0.678	0.765	0.859	0.824	0.757	0.557	0.459	0.420	0.510	0.569	0.349	0.396	0.435	0.294	0.290	0.361
0.663	0.757	0.875	0.827	0.824	0.545	0.471	0.424	0.561	0.588	0.365	0.424	0.471	0.259	0.271	0.349
0.635	0.784	0.886	0.855	0.796	0.569	0.490	0.471	0.518	0.388	0.357	0.392	0.341	0.314	0.314	0.322
0.635	0.780	0.886	0.851	0.812	0.580	0.490	0.471	0.518	0.329	0.282	0.361	0.243	0.255	0.239	0.231
0.639	0.765	0.898	0.839	0.812	0.592	0.498	0.498	0.545	0.384	0.337	0.420	0.353	0.263	0.231	0.220
0.635	0.745	0.902	0.804	0.843	0.616	0.518	0.514	0.576	0.392	0.357	0.447	0.392	0.302	0.290	0.294
0.631	0.780	0.878	0.792	0.765	0.537	0.463	0.439	0.486	0.439	0.384	0.443	0.392	0.306	0.314	0.306
0.561	0.729	0.945	0.867	0.855	0.439	0.396	0.357	0.588	0.451	0.329	0.522	0.420	0.302	0.294	0.290
0.655	0.831	0.859	0.827	0.800	0.600	0.553	0.549	0.671	0.588	0.380	0.439	0.494	0.294	0.298	0.345
0.714	0.847	0.867	0.918	0.753	0.624	0.533	0.498	0.490	0.490	0.463	0.404	0.447	0.314	0.416	0.435
0.722	0.863	0.902	0.937	0.788	0.671	0.588	0.557	0.561	0.557	0.518	0.522	0.529	0.314	0.424	0.475
0.718	0.886	0.929	0.973	0.820	0.690	0.631	0.588	0.592	0.592	0.561	0.616	0.612	0.388	0.498	0.533
0.651	0.827	0.925	0.894	0.729	0.514	0.408	0.396	0.463	0.376	0.357	0.431	0.475	0.333	0.341	0.345
0.651	0.824	0.922	0.886	0.737	0.498	0.400	0.392	0.486	0.353	0.298	0.416	0.478	0.267	0.263	0.302
0.682	0.839	0.910	0.851	0.878	0.686	0.584	0.624	0.533	0.388	0.365	0.467	0.396	0.337	0.267	0.278
0.651	0.851	0.925	0.906	0.769	0.565	0.498	0.498	0.600	0.408	0.357	0.478	0.424	0.243	0.227	0.231
0.659	0.855	0.929	0.898	0.839	0.580	0.514	0.502	0.616	0.412	0.376	0.502	0.439	0.247	0.247	0.239
0.624	0.788	0.886	0.878	0.820	0.820	0.655	0.608	0.702	0.533	0.447	0.533	0.463	0.275	0.267	0.278
0.635	0.792	0.894	0.886	0.863	0.784	0.651	0.620	0.702	0.533	0.463	0.518	0.455	0.212	0.192	0.208
0.573	0.827	0.953	0.906	0.820	0.529	0.475	0.439	0.584	0.396	0.361	0.424	0.349	0.306	0.310	0.306
0.608	0.827	0.957	0.906	0.804	0.522	0.475	0.443	0.580	0.400	0.349	0.412	0.365	0.294	0.286	0.282
0.616	0.725	0.824	0.796	0.765	0.533	0.471	0.478	0.494	0.369	0.349	0.345	0.333	0.294	0.298	0.306
0.608	0.714	0.847	0.804	0.792	0.537	0.482	0.482	0.522	0.380	0.380	0.380	0.365	0.353	0.329	0.329
0.627	0.878	0.969	0.929	0.773	0.533	0.467	0.443	0.624	0.420	0.365	0.553	0.498	0.420	0.306	0.286
0.631	0.871	0.961	0.933	0.804	0.533	0.455	0.459	0.631	0.388	0.357	0.545	0.506	0.439	0.306	0.271

- within-class scatter matrix of this data (all elements multiplied by 1000)

$$\begin{bmatrix}
1.134 & 0.971 & 0.780 & 0.937 & 0.643 & 0.806 & 0.840 & 0.753 & 0.648 & 1.121 & 0.671 & 0.793 & 1.364 & 0.870 & 0.903 & 0.918 \\
0.971 & 1.650 & 1.425 & 1.057 & 0.740 & 0.893 & 1.095 & 0.750 & 0.744 & 1.007 & 1.002 & 1.196 & 1.494 & 0.998 & 0.973 & 1.204 \\
0.780 & 1.425 & 2.984 & 2.622 & 2.154 & 2.298 & 1.993 & 1.751 & 2.122 & 2.301 & 2.065 & 2.329 & 2.857 & 2.055 & 2.113 & 2.117 \\
0.937 & 1.057 & 2.622 & 4.582 & 2.579 & 2.288 & 1.225 & 1.926 & 2.599 & 3.451 & 2.641 & 2.700 & 3.246 & 2.713 & 3.000 & 2.576 \\
0.643 & 0.740 & 2.154 & 2.579 & 4.384 & 3.138 & 1.631 & 2.174 & 2.956 & 1.993 & 1.730 & 2.278 & 1.428 & 1.897 & 1.338 & 0.688 \\
0.806 & 0.893 & 2.298 & 2.288 & 3.138 & 5.652 & 4.476 & 3.806 & 4.225 & 3.220 & 3.243 & 3.281 & 3.416 & 2.213 & 2.400 & 2.010 \\
0.840 & 1.095 & 1.993 & 1.225 & 1.631 & 4.476 & 5.955 & 4.205 & 3.586 & 2.656 & 3.281 & 2.447 & 3.618 & 2.054 & 2.507 & 2.280 \\
0.753 & 0.750 & 1.751 & 1.926 & 2.174 & 3.806 & 4.205 & 5.634 & 4.372 & 2.725 & 3.481 & 2.344 & 3.161 & 1.779 & 2.325 & 2.163 \\
0.648 & 0.744 & 2.122 & 2.599 & 2.956 & 4.225 & 3.586 & 4.372 & 5.776 & 3.625 & 3.243 & 3.211 & 3.566 & 2.193 & 2.065 & 1.679 \\
1.121 & 1.007 & 2.301 & 3.451 & 1.993 & 3.220 & 2.656 & 2.254 & 3.625 & 7.205 & 3.842 & 3.630 & 4.916 & 3.260 & 3.346 & 3.766 \\
0.671 & 1.002 & 2.065 & 2.641 & 1.730 & 3.243 & 3.281 & 3.481 & 3.243 & 3.842 & 5.239 & 3.412 & 4.601 & 2.735 & 3.351 & 3.405 \\
0.793 & 1.196 & 2.329 & 2.700 & 2.278 & 3.281 & 2.447 & 2.344 & 3.211 & 3.630 & 3.412 & 4.575 & 4.203 & 2.680 & 2.495 & 2.289 \\
1.364 & 1.494 & 2.857 & 3.246 & 1.428 & 3.416 & 3.618 & 3.161 & 3.566 & 4.916 & 4.601 & 4.203 & 7.343 & 3.698 & 4.022 & 3.990 \\
0.870 & 0.998 & 2.055 & 2.713 & 1.897 & 2.213 & 2.054 & 1.779 & 2.193 & 3.260 & 2.735 & 2.680 & 3.698 & 4.879 & 3.868 & 2.892 \\
0.903 & 0.973 & 2.113 & 3.000 & 1.338 & 2.400 & 2.507 & 2.325 & 2.065 & 3.346 & 3.351 & 2.495 & 4.022 & 3.868 & 5.483 & 4.661 \\
0.918 & 1.204 & 2.117 & 2.576 & 0.688 & 2.010 & 2.280 & 2.163 & 1.679 & 3.766 & 3.405 & 2.289 & 3.990 & 2.892 & 4.661 & 6.069
\end{bmatrix}$$

- between-class scatter matrix of this data (all elements multiplied by 1000)

$$\begin{bmatrix}
0.045 & 0.075 & 0.066 & -0.194 & -0.277 & -0.230 & -0.018 & -0.017 & -0.074 & -0.000 & 0.103 & 0.071 & 0.258 & 0.057 & -0.069 & 0.040 \\
0.075 & 0.248 & 0.097 & -0.209 & -0.264 & -0.757 & -1.112 & -1.188 & -0.680 & -0.923 & -0.561 & -0.360 & -0.242 & -0.612 & -0.998 & -0.392 \\
0.066 & 0.097 & 0.241 & -0.237 & 0.012 & 0.899 & 0.436 & 0.013 & -0.181 & 0.443 & 0.815 & 0.758 & 0.890 & 0.635 & 0.522 & 0.615 \\
-0.194 & -0.209 & -0.237 & 0.981 & 1.594 & 1.184 & -0.799 & -1.079 & -0.272 & -0.728 & -0.888 & -0.494 & -1.571 & -0.711 & -0.308 & -0.384 \\
-0.277 & -0.264 & 0.012 & 1.594 & 3.451 & 4.587 & -0.601 & -2.110 & -0.887 & -0.458 & -0.031 & 0.658 & -1.359 & -0.042 & 0.620 & 0.581 \\
-0.230 & -0.757 & 0.899 & 1.184 & 4.587 & 12.439 & 6.229 & 2.786 & 0.914 & 5.644 & 6.576 & 6.145 & 4.353 & 5.814 & 7.447 & 5.438 \\
-0.018 & -1.112 & 0.436 & -0.799 & -0.601 & 6.229 & 10.361 & 10.025 & 4.619 & 8.953 & 7.823 & 5.624 & 6.888 & 7.343 & 9.068 & 5.237 \\
-0.017 & -1.188 & 0.013 & -1.079 & -2.110 & 2.786 & 10.025 & 11.074 & 5.396 & 8.520 & 6.533 & 4.103 & 6.019 & 6.367 & 8.036 & 3.993 \\
-0.074 & -0.680 & -0.181 & -0.272 & -0.887 & 0.914 & 4.619 & 5.396 & 2.782 & 3.872 & 2.614 & 1.483 & 2.233 & 2.671 & 3.623 & 1.537 \\
-0.000 & -0.923 & 0.443 & -0.728 & -0.458 & 5.644 & 8.953 & 8.520 & 3.872 & 7.758 & 6.910 & 5.029 & 6.116 & 6.447 & 7.895 & 4.655 \\
0.103 & -0.561 & 0.815 & -0.888 & -0.031 & 6.576 & 7.823 & 6.533 & 2.614 & 6.910 & 6.999 & 5.473 & 6.404 & 6.287 & 7.266 & 4.895 \\
0.071 & -0.360 & 0.758 & -0.494 & 0.658 & 6.145 & 5.624 & 4.103 & 1.483 & 5.029 & 5.473 & 4.509 & 4.866 & 4.835 & 5.547 & 3.981 \\
0.258 & -0.242 & 0.890 & -1.571 & -1.359 & 4.353 & 6.888 & 6.019 & 2.233 & 6.116 & 6.404 & 4.866 & 6.525 & 5.658 & 6.060 & 4.290 \\
0.057 & -0.612 & 0.635 & -0.711 & -0.042 & 5.814 & 7.343 & 6.367 & 2.671 & 6.447 & 6.287 & 4.835 & 5.658 & 5.712 & 6.741 & 4.367 \\
-0.069 & -0.998 & 0.522 & -0.308 & 0.620 & 7.447 & 9.068 & 8.036 & 3.623 & 7.895 & 7.266 & 5.547 & 6.060 & 6.741 & 8.395 & 5.102 \\
0.040 & -0.392 & 0.615 & -0.384 & 0.581 & 5.438 & 5.237 & 3.993 & 1.537 & 4.655 & 4.895 & 3.981 & 4.290 & 4.367 & 5.102 & 3.541
\end{bmatrix}$$

Appendix C

LDA on Test Data

- matrix $S_w^{-1} S_b$:

$$\begin{bmatrix}
0.173 & -0.134 & 0.223 & -0.145 & -0.096 & -0.090 & -0.034 & 0.019 & 0.029 & -0.012 & 0.073 & 0.043 & 0.008 & 0.059 & -0.067 & -0.003 \\
0.375 & -0.071 & 0.247 & 0.023 & -0.044 & 0.087 & -0.164 & -0.427 & 0.270 & -0.344 & 0.228 & -0.142 & 0.148 & 0.055 & -0.469 & 0.287 \\
0.122 & -0.164 & 0.227 & -0.379 & -0.137 & 0.535 & -0.314 & -0.001 & -0.305 & -0.011 & 0.215 & 0.086 & 0.048 & 0.173 & -0.058 & 0.022 \\
-0.725 & 0.750 & -1.142 & 0.824 & 0.534 & 0.931 & -0.079 & -0.560 & -0.061 & -0.270 & -0.165 & -0.393 & 0.115 & -0.261 & -0.046 & 0.316 \\
-1.324 & 1.244 & -1.963 & 0.920 & 0.841 & 3.141 & -0.836 & -1.138 & -0.835 & -0.553 & 0.091 & -0.629 & 0.331 & -0.184 & -0.163 & 0.683 \\
-2.203 & 0.662 & -1.763 & -1.184 & 0.293 & 5.633 & -1.838 & 0.589 & -3.555 & 0.785 & 0.310 & 0.402 & -0.017 & 0.511 & 1.435 & -0.366 \\
-1.121 & -1.177 & 0.750 & -2.612 & -0.947 & -0.400 & 0.301 & 3.852 & -2.738 & 2.788 & -0.670 & 1.867 & -1.069 & 0.556 & 3.169 & -2.430 \\
-0.788 & -1.440 & 1.204 & -2.384 & -1.082 & -2.694 & 1.217 & 4.406 & -1.889 & 3.091 & -1.099 & 2.017 & -1.313 & 0.339 & 3.394 & -2.802 \\
-0.553 & -0.527 & 0.324 & -0.845 & -0.390 & -1.602 & 0.809 & 2.137 & -0.743 & 1.519 & -0.711 & 0.906 & -0.674 & 0.024 & 1.726 & -1.377 \\
-0.940 & -1.033 & 0.678 & -2.326 & -0.835 & -0.136 & 0.151 & 3.267 & -2.437 & 2.366 & -0.506 & 1.608 & -0.892 & 0.522 & 2.683 & -2.055 \\
-0.620 & -1.018 & 0.797 & -2.477 & -0.846 & 1.204 & -0.561 & 2.464 & -2.578 & 1.789 & 0.026 & 1.373 & -0.579 & 0.728 & 1.982 & -1.512 \\
-0.586 & -0.592 & 0.351 & -1.833 & -0.530 & 1.871 & -0.802 & 1.476 & -2.184 & 1.116 & 0.209 & 0.893 & -0.296 & 0.606 & 1.288 & -0.892 \\
0.150 & -1.462 & 1.632 & -2.714 & -1.141 & 0.282 & -0.443 & 2.397 & -2.034 & 1.623 & 0.207 & 1.422 & -0.539 & 0.828 & 1.577 & -1.443 \\
-0.685 & -0.886 & 0.628 & -2.166 & -0.736 & 0.816 & -0.338 & 2.406 & -2.311 & 1.753 & -0.118 & 1.283 & -0.597 & 0.596 & 1.972 & -1.490 \\
-1.341 & -0.709 & 0.138 & -2.189 & -0.626 & 0.994 & -0.195 & 2.976 & -2.833 & 2.238 & -0.429 & 1.467 & -0.793 & 0.523 & 2.671 & -1.877 \\
-0.608 & -0.510 & 0.256 & -1.601 & -0.458 & 1.502 & -0.610 & 1.445 & -1.951 & 1.094 & 0.095 & 0.831 & -0.315 & 0.503 & 1.277 & -0.883
\end{bmatrix}$$

- nonvanishing eigenvalues of this matrix:

$$\lambda_1 = 8.701, \quad \lambda_2 = 6.002 \quad \text{and} \quad \lambda_3 = 2.018$$

- corresponding eigenvectors:

$$e_1 = \begin{bmatrix} -0.054 \\ 0.220 \\ -0.266 \\ 0.234 \\ 0.154 \\ 0.609 \\ -0.226 \\ -0.470 \\ 0.026 \\ -0.306 \\ 0.127 \\ -0.215 \\ 0.146 \\ -0.022 \\ -0.298 \\ 0.298 \end{bmatrix} \quad e_2 = \begin{bmatrix} 0.227 \\ 0.024 \\ 0.083 \\ 0.290 \\ 0.041 \\ -0.570 \\ 0.201 \\ -0.232 \\ 0.494 \\ -0.198 \\ -0.032 \\ -0.137 \\ 0.043 \\ -0.100 \\ -0.266 \\ 0.142 \end{bmatrix} e_3 = \begin{bmatrix} 0.602 \\ -0.419 \\ 0.723 \\ -0.506 \\ -0.306 \\ 0.045 \\ -0.275 \\ -0.105 \\ 0.033 \\ -0.154 \\ 0.350 \\ 0.091 \\ 0.093 \\ 0.237 \\ -0.352 \\ 0.104 \end{bmatrix}$$

- transformation kernel for projection onto 2 dimensions:

$$K = \begin{bmatrix} -0.054 & 0.227 \\ 0.220 & 0.024 \\ -0.266 & 0.083 \\ 0.234 & 0.290 \\ 0.154 & 0.041 \\ 0.609 & -0.570 \\ -0.226 & 0.201 \\ -0.470 & -0.232 \\ 0.026 & 0.494 \\ -0.306 & -0.198 \\ 0.127 & -0.032 \\ -0.215 & -0.137 \\ 0.146 & 0.043 \\ -0.022 & -0.100 \\ -0.298 & -0.266 \\ 0.298 & 0.142 \end{bmatrix}$$

Appendix D

Experiments on Resource Management Database (Context Independent)

Resource Management Database :

- speakerindependent database
- 109 male and female speakers of different american dialects
- 4360 sentences for training
- data sampled at $16kHz$ with $16bit$ quantisation

primary transformation :

- 256-point FFT with Hamming Window, window length $16ms$
- $6ms$ window overlapp
- dimension reduction to 16 melscale coefficients

phonetic modeling :

- 48 monophones, each splited in 3 subphonemes
- 1 silence class

training :

- training set: 2830 sentences form 78 male speakers
- 6 iterations over the training set, codebook and distribution updates after each iteration

test set :

- 48 sentences form 12 male speakers
- neither the speakers nor the sentences are part of the training material
- word pair grammar, perplexity 60

D.1 Experiment with LDA Feature Vectors

genaral parameters

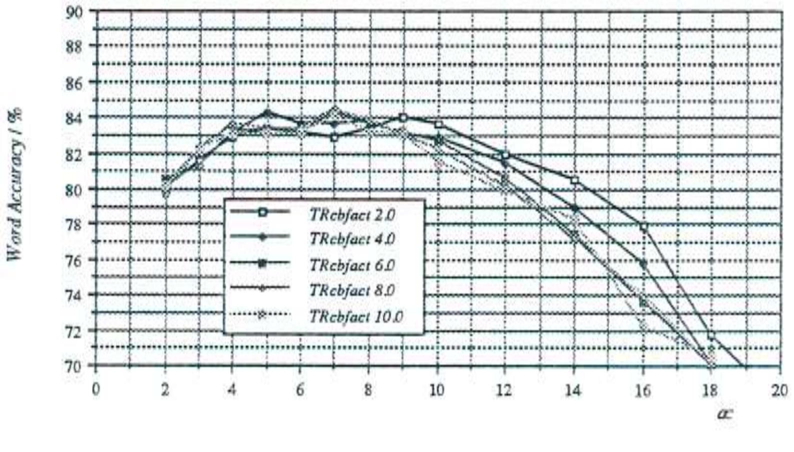
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in LDA :	146

developement of class separability

	primary feature vectors (32 cofficients)	LDA derived feature vectors (16 coefficients)
Q_{146}	36.461	7.329
Q_{50}	58.341	15.860

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	2.0	4.0	6.0	8.0	10.0
2.0	80.0% (80.7%)	80.0% (81.1%)	80.6% (81.5%)	79.7% (80.6%)	80.2% (81.1%)
3.0	81.6% (83.1%)	82.2% (84.0%)	81.3% (83.2%)	82.2% (84.1%)	82.2% (84.1%)
4.0	82.9% (84.7%)	83.1% (85.0%)	83.2% (85.2%)	83.6% (85.7%)	83.1% (85.0%)
5.0	83.4% (85.7%)	84.3% (86.5%)	83.4% (85.6%)	83.4% (85.7%)	83.1% (85.4%)
6.0	83.2% (86.6%)	83.6% (86.5%)	83.1% (85.7%)	83.4% (86.1%)	83.1% (85.7%)
7.0	82.9% (86.5%)	83.6% (86.6%)	84.3% (87.3%)	84.5% (87.5%)	84.1% (87.3%)
8.0	83.4% (87.3%)	83.8% (87.7%)	83.6% (87.3%)	83.8% (87.5%)	83.2% (87.0%)
9.0	84.0% (88.2%)	83.1% (87.0%)	83.2% (87.0%)	83.1% (87.0%)	83.1% (86.8%)
10.0	83.6% (87.9%)	82.9% (87.0%)	82.7% (86.8%)	82.2% (86.5%)	81.6% (86.5%)
12.0	82.0% (86.3%)	81.5% (85.6%)	80.7% (85.9%)	80.2% (85.4%)	79.9% (85.0%)
14.0	80.6% (85.0%)	79.0% (84.0%)	77.5% (82.5%)	77.2% (82.7%)	78.3% (84.0%)
16.0	77.9% (82.2%)	75.8% (81.1%)	73.6% (79.7%)	74.0% (80.4%)	72.2% (79.9%)
18.0	71.8% (78.3%)	70.2% (76.6%)	70.1% (76.3%)	70.1% (76.5%)	70.6% (77.9%)
20.0	67.9% (75.4%)	66.7% (74.2%)	66.7% (74.3%)	64.5% (72.0%)	66.0% (74.2%)



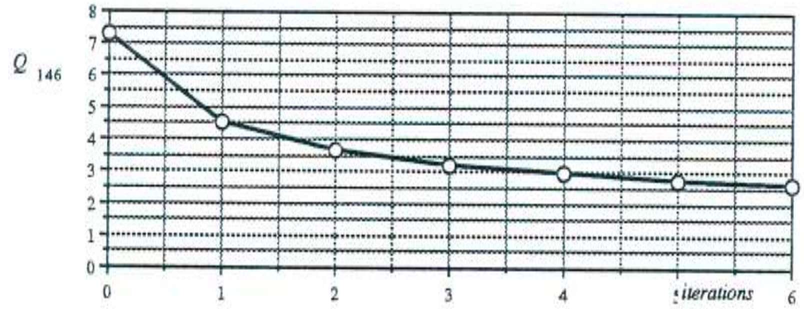
D.2 Experiment E.1.1

NLDA approach : $Y = Af(X)$

```
genaral parameters                                    backpropagation parameters
   dimension of orignial space :          32             number of discriminated targets :   146
   dimension of image space :             16             learning rate :                     0.008
   discriminated classes in on top LDA :  146            momentum :                          0.9
target drift parameters                                  number of hidden units :            32
   α :                                    1.0            iterations :                        6
   β :                                    0.0            target update after :               6
   m :                                    1              sample selection :                  random
   number of drift vectors :              146 (one per subphonem class)
```

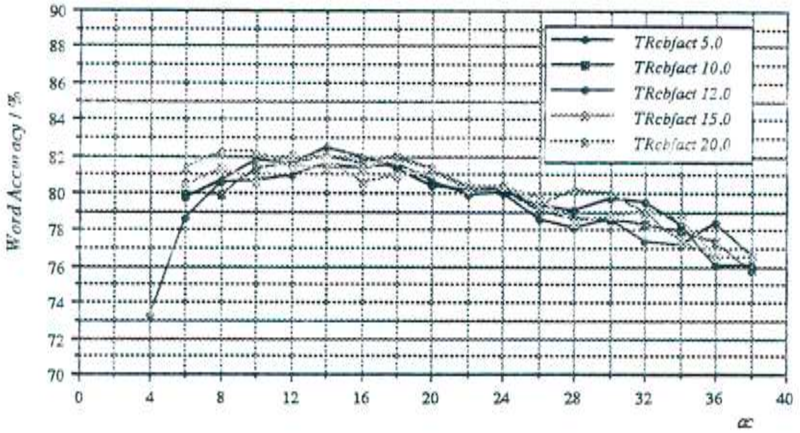
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	4.506	3.643	3.215	2.932	2.741	2.600
Q_{sn}	15.860	-	-	-	-	-	11.105



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	12.0	15.0	20.0
4.0	73.3% (74.3%)	-% (-%)	-% (-%)	-% (-%)	-% (-%)
6.0	78.6% (80.2%)	80.0% (81.8%)	79.7% (80.9%)	80.4% (81.6%)	81.3% (82.5%)
8.0	80.6% (82.4%)	79.9% (82.0%)	80.7% (82.2%)	81.3% (82.7%)	82.2% (84.7%)
10.0	80.7% (82.9%)	81.3% (83.2%)	81.8% (83.6%)	80.6% (82.7%)	82.2% (84.5%)
12.0	80.9% (83.2%)	81.6% (84.0%)	81.6% (84.0%)	82.0% (84.0%)	81.5% (83.6%)
14.0	81.6% (83.8%)	82.0% (84.1%)	82.4% (84.7%)	81.8% (84.0%)	81.3% (84.0%)
16.0	81.3% (83.6%)	81.6% (84.3%)	82.0% (84.5%)	80.6% (83.4%)	81.3% (84.1%)
18.0	81.6% (84.0%)	82.0% (84.7%)	81.3% (84.1%)	80.9% (83.8%)	81.8% (84.7%)
20.0	80.7% (83.8%)	81.3% (84.3%)	80.4% (84.0%)	81.1% (84.0%)	81.3% (84.3%)
22.0	79.9% (83.6%)	80.2% (84.0%)	80.2% (84.0%)	80.2% (84.1%)	80.4% (84.0%)
24.0	80.0% (83.8%)	80.2% (84.1%)	80.2% (84.3%)	80.4% (84.3%)	80.4% (84.0%)
26.0	78.6% (82.9%)	79.1% (84.3%)	79.3% (84.3%)	79.7% (84.0%)	79.3% (83.6%)
28.0	78.1% (82.9%)	78.6% (84.0%)	79.1% (84.3%)	78.4% (83.6%)	80.2% (84.8%)
30.0	78.6% (83.4%)	78.6% (84.1%)	79.7% (84.7%)	78.8% (84.1%)	80.0% (84.8%)
32.0	77.4% (82.9%)	78.3% (83.8%)	79.5% (84.8%)	79.0% (84.1%)	79.0% (84.8%)
34.0	77.2% (82.9%)	77.9% (84.0%)	78.3% (84.3%)	78.7% (83.8%)	77.4% (84.0%)
36.0	78.4% (83.4%)	77.4% (83.4%)	76.1% (82.4%)	76.5% (83.2%)	76.5% (83.4%)
38.0	76.6% (82.0%)	75.8% (82.0%)	76.1% (82.4%)	76.5% (83.1%)	76.5% (83.8%)



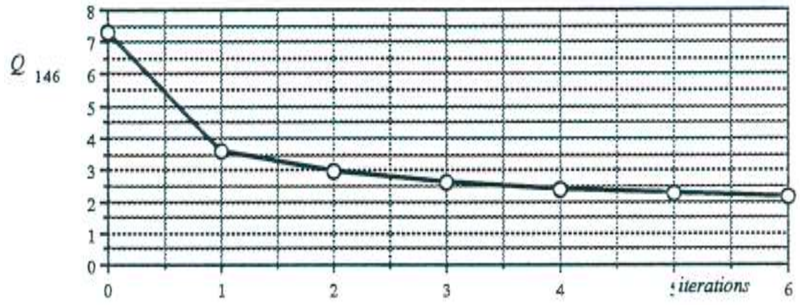
D.3 Experiment E.1.2

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	146 (one per subphonem class)		

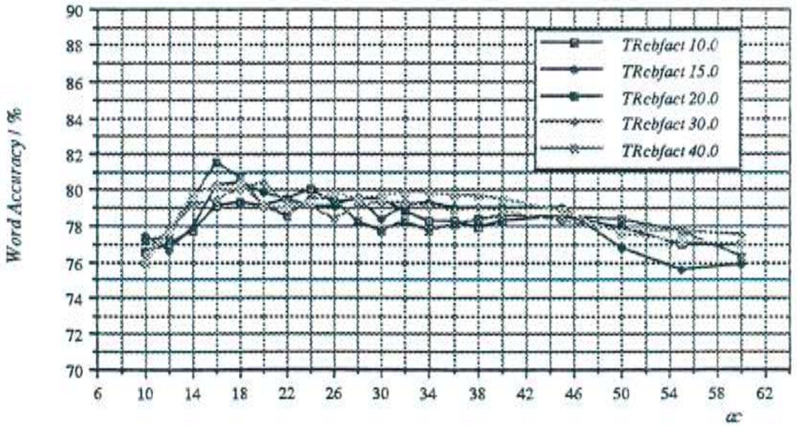
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	3.599	2.933	2.593	2.382	2.260	2.165
Q_{kn}	15.860	-	-	-	-	-	10.619



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	10.0	15.0	20.0	30.0	40.0
10	76.5% (77.7%)	77.4% (78.4%)	77.2% (78.8%)	76.3% (77.7%)	75.9% (77.4%)
12	77.0% (78.4%)	76.6% (77.9%)	77.4% (78.6%)	77.7% (79.3%)	77.7% (79.3%)
14	77.7% (79.1%)	77.9% (79.1%)	79.5% (80.7%)	79.1% (80.9%)	79.7% (81.5%)
16	79.1% (81.3%)	80.2% (81.6%)	81.5% (82.7%)	79.5% (81.3%)	80.2% (82.0%)
18	79.3% (81.5%)	80.4% (82.4%)	80.7% (82.7%)	80.0% (81.8%)	80.2% (82.5%)
20	79.1% (81.8%)	79.9% (82.4%)	79.1% (81.6%)	80.4% (82.7%)	79.1% (82.0%)
22	79.5% (82.0%)	79.5% (82.0%)	78.6% (81.8%)	79.3% (82.5%)	79.0% (82.0%)
24	80.0% (82.5%)	79.1% (82.0%)	79.5% (82.7%)	79.1% (82.4%)	79.3% (82.5%)
26	79.3% (82.7%)	79.1% (82.4%)	79.5% (82.5%)	78.4% (81.8%)	79.9% (82.9%)
28	79.5% (83.2%)	79.6% (82.2%)	78.3% (82.2%)	79.1% (82.2%)	79.5% (82.7%)
30	79.5% (83.6%)	78.4% (82.2%)	77.7% (81.8%)	79.3% (82.7%)	79.7% (82.9%)
32	78.8% (83.2%)	79.1% (83.4%)	78.3% (82.4%)	79.3% (83.1%)	79.9% (83.6%)
34	78.3% (82.9%)	79.3% (83.6%)	77.7% (82.2%)	79.1% (83.2%)	79.9% (83.6%)
36	78.3% (82.9%)	79.0% (83.6%)	78.1% (82.2%)	78.8% (83.1%)	79.7% (83.4%)
38	77.9% (82.9%)	79.0% (83.6%)	78.4% (82.5%)	79.0% (83.2%)	79.7% (83.4%)
40	78.3% (83.4%)	79.0% (83.6%)	78.6% (83.1%)	78.8% (83.6%)	79.5% (84.3%)
45	78.6% (84.0%)	79.0% (83.6%)	78.6% (83.1%)	78.3% (83.4%)	78.8% (84.1%)
50	77.9% (83.4%)	76.8% (83.1%)	78.4% (84.0%)	77.7% (83.8%)	77.5% (83.6%)
55	77.0% (82.5%)	75.6% (82.2%)	77.7% (83.4%)	77.7% (83.8%)	77.2% (83.2%)
60	77.0% (82.2%)	75.9% (81.8%)	76.3% (82.4%)	77.5% (83.2%)	77.0% (82.7%)



D.4 Experiment E.1.3

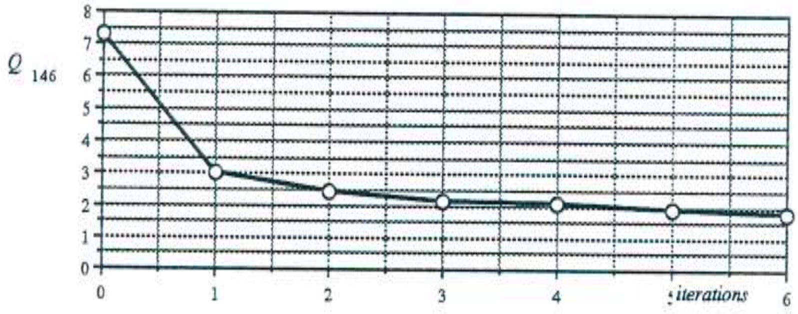
NLDA approach : $Y = Af(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.2
m :	1
number of drift vectors :	146 (one per subphonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	32
iterations :	6
target update after :	6
sample selection :	random

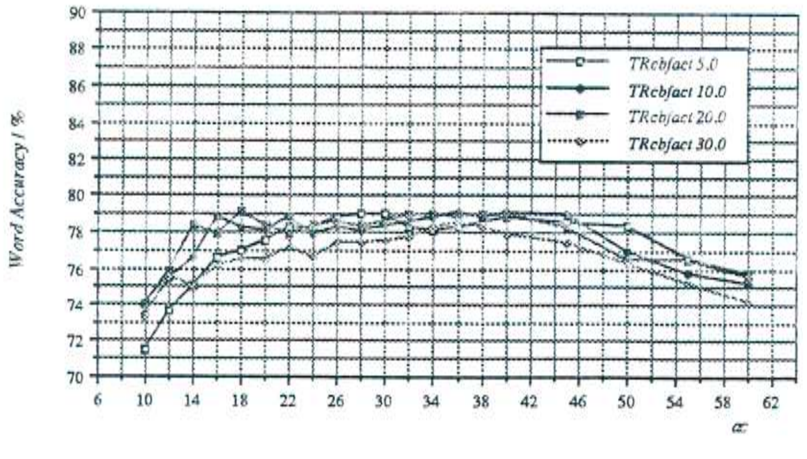
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	3.018	2.444	2.163	2.074	1.925	1.822
Q_{KN}	15.860	-	-	-	-	-	10.347



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	20.0	30.0
10	71.5% (72.5%)	73.4% (75.2%)	74.0% (75.6%)	73.3% (75.2%)
12	73.6% (75.4%)	75.4% (77.4%)	75.9% (76.1%)	75.6% (76.1%)
14	75.0% (77.2%)	76.6% (78.6%)	78.4% (80.0%)	75.0% (77.9%)
16	76.6% (78.6%)	78.8% (80.6%)	77.9% (80.4%)	76.1% (79.0%)
18	77.0% (79.3%)	78.3% (80.4%)	79.1% (81.1%)	76.5% (79.5%)
20	77.5% (80.4%)	78.1% (80.6%)	78.4% (81.5%)	76.5% (79.9%)
22	78.3% (81.1%)	78.8% (81.5%)	77.9% (81.1%)	77.2% (80.7%)
24	78.3% (81.1%)	77.9% (81.1%)	78.4% (81.6%)	76.6% (80.6%)
26	78.8% (81.8%)	78.3% (81.5%)	78.6% (81.8%)	77.4% (81.1%)
28	79.0% (82.5%)	78.1% (81.3%)	78.4% (82.0%)	77.4% (81.3%)
30	79.0% (82.5%)	78.4% (82.0%)	78.6% (82.5%)	77.5% (81.8%)
32	78.3% (81.8%)	78.6% (82.2%)	79.0% (82.9%)	77.7% (82.0%)
34	78.1% (81.6%)	78.8% (82.5%)	79.0% (82.9%)	78.3% (82.5%)
36	78.3% (82.0%)	79.1% (82.9%)	78.8% (82.9%)	78.6% (82.9%)
38	78.6% (82.4%)	78.8% (82.9%)	79.0% (83.1%)	78.3% (82.9%)
40	78.8% (82.5%)	79.1% (83.2%)	79.0% (83.1%)	77.9% (82.5%)
45	78.6% (83.1%)	79.0% (83.2%)	78.3% (82.5%)	77.4% (82.2%)
50	78.4% (83.1%)	77.0% (82.7%)	76.5% (81.5%)	76.3% (81.8%)
55	76.6% (82.5%)	75.8% (82.5%)	76.5% (81.8%)	75.2% (81.6%)
60	75.8% (81.8%)	75.2% (82.2%)	75.6% (81.8%)	74.2% (81.1%)



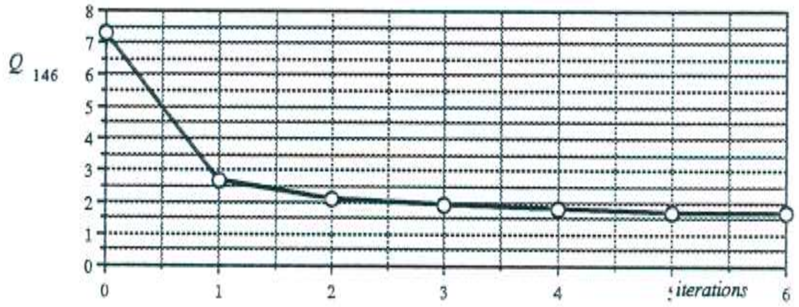
D.5 Experiment E.1.4

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	146 (one per subphonem class)		

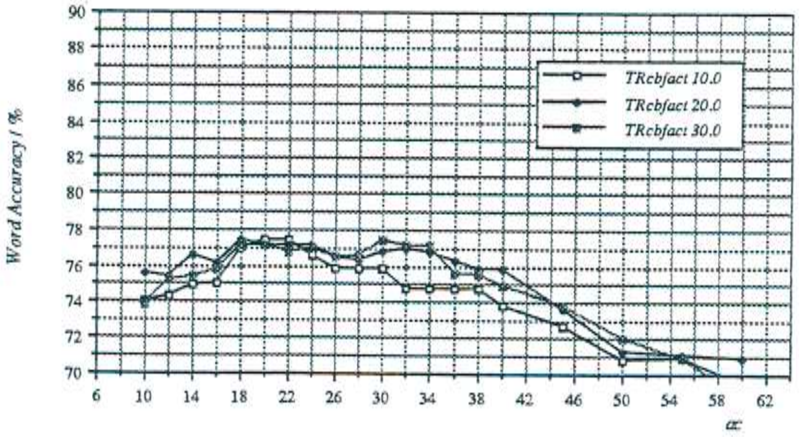
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	2.646	2.075	1.903	1.817	1.708	1.687
Q_{kn}	15.860	-	-	-	-	-	10.258



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	10.0	20.0	30.0	40.0
10	74.0% (75.8%)	75.6% (77.5%)	73.8% (75.6%)	61.5% (62.7%)
12	74.3% (76.1%)	75.4% (77.4%)	75.2% (76.8%)	63.3% (64.3%)
14	74.9% (76.6%)	76.6% (78.6%)	75.4% (77.5%)	63.3% (64.9%)
16	75.0% (77.4%)	76.1% (78.4%)	75.8% (78.3%)	62.2% (64.7%)
18	77.0% (79.5%)	77.4% (80.2%)	77.2% (79.9%)	63.3% (66.1%)
20	77.4% (80.4%)	77.2% (80.4%)	77.2% (80.0%)	63.8% (66.7%)
22	77.4% (80.2%)	77.2% (80.6%)	76.8% (79.7%)	64.2% (67.2%)
24	76.6% (80.0%)	77.2% (80.6%)	77.0% (79.9%)	64.0% (67.9%)
26	75.9% (80.0%)	76.5% (80.4%)	76.5% (78.7%)	62.2% (67.4%)
28	75.9% (80.4%)	76.3% (80.2%)	76.6% (79.9%)	61.7% (67.6%)
30	75.9% (80.4%)	76.8% (80.7%)	77.4% (80.7%)	61.3% (67.6%)
32	74.7% (79.9%)	77.0% (81.1%)	77.2% (81.3%)	61.0% (67.4%)
34	74.7% (79.9%)	76.8% (81.1%)	77.2% (81.3%)	60.4% (67.2%)
36	74.7% (80.0%)	76.3% (80.7%)	75.6% (80.2%)	61.0% (68.1%)
38	74.7% (80.0%)	75.9% (80.6%)	75.6% (80.2%)	60.4% (68.4%)
40	73.8% (79.3%)	75.9% (80.6%)	74.9% (79.9%)	60.4% (68.4%)
45	72.7% (79.0%)	73.6% (79.3%)	73.8% (79.7%)	58.8% (67.7%)
50	70.8% (77.4%)	71.3% (77.5%)	72.0% (78.3%)	57.4% (67.0%)
55	70.9% (77.7%)	71.1% (78.3%)	70.9% (77.5%)	54.7% (65.4%)
60	69.5% (77.4%)	70.9% (78.4%)	68.8% (76.1%)	52.0% (64.3%)



D.6 Experiment E.1.5

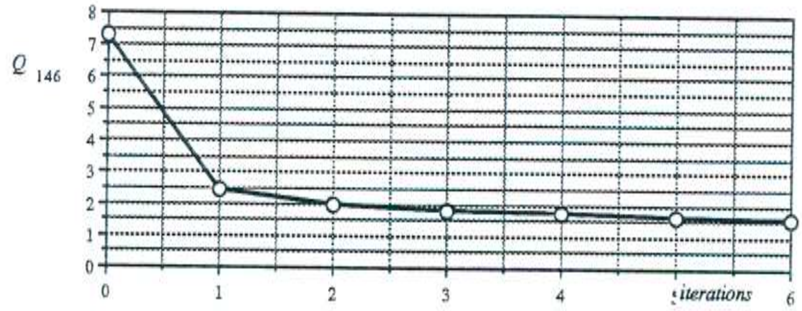
NLDA approach : $Y = Af(X)$

genaral parameters	
dimension of origníal space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.4
m :	1
number of drift vectors :	146 (one per subphonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	32
iterations :	6
target update after :	6
sample selection :	random

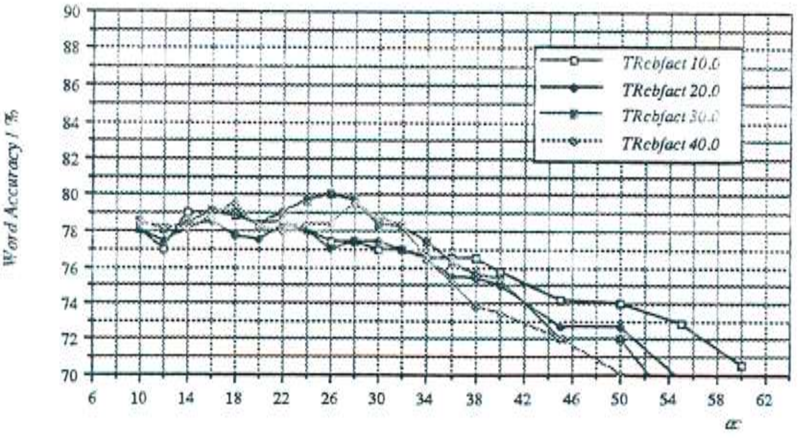
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	2.441	1.975	1.781	1.751	1.605	1.573
Q_{Kn}	15.860	-	-	-	-	-	10.122



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	10.0	20.0	30.0	40.0
10	78.1% (79.5%)	78.1% (79.3%)	78.6% (79.9%)	78.4% (80.0%)
12	77.0% (79.0%)	77.4% (79.5%)	78.1% (80.0%)	78.3% (80.2%)
14	79.0% (80.9%)	78.1% (80.4%)	78.4% (80.4%)	78.1% (80.2%)
16	79.0% (80.9%)	78.6% (81.3%)	79.1% (81.3%)	78.8% (80.9%)
18	78.8% (81.1%)	77.7% (80.9%)	79.0% (81.3%)	79.5% (81.6%)
20	78.4% (81.1%)	77.5% (80.9%)	78.4% (80.9%)	78.3% (80.7%)
22	78.6% (81.5%)	78.3% (81.5%)	79.0% (82.0%)	77.9% (80.7%)
24	78.1% (80.9%)	78.1% (81.5%)	79.7% (82.9%)	78.4% (81.5%)
26	77.4% (81.1%)	77.0% (81.6%)	80.0% (83.2%)	78.4% (81.5%)
28	77.4% (81.1%)	77.4% (82.0%)	79.7% (83.1%)	79.5% (82.7%)
30	77.0% (81.1%)	77.4% (82.0%)	78.4% (82.7%)	78.8% (82.5%)
32	77.0% (81.1%)	77.0% (81.8%)	78.3% (82.5%)	78.3% (82.4%)
34	76.5% (80.9%)	76.6% (81.8%)	77.4% (82.5%)	76.5% (81.5%)
36	76.5% (81.6%)	75.5% (81.5%)	76.3% (82.0%)	75.2% (81.1%)
38	76.5% (81.8%)	75.4% (81.5%)	75.6% (81.6%)	73.8% (80.2%)
40	75.8% (81.5%)	75.0% (81.6%)	75.4% (82.0%)	73.4% (80.2%)
45	74.2% (81.6%)	72.7% (79.9%)	72.0% (80.4%)	72.0% (80.6%)
50	74.0% (81.8%)	72.7% (80.0%)	72.0% (80.2%)	70.1% (78.4%)
55	72.9% (80.9%)	69.7% (78.4%)	67.9% (77.0%)	66.3% (76.5%)
60	70.6% (79.3%)	66.5% (75.9%)	66.7% (75.9%)	65.4% (75.2%)



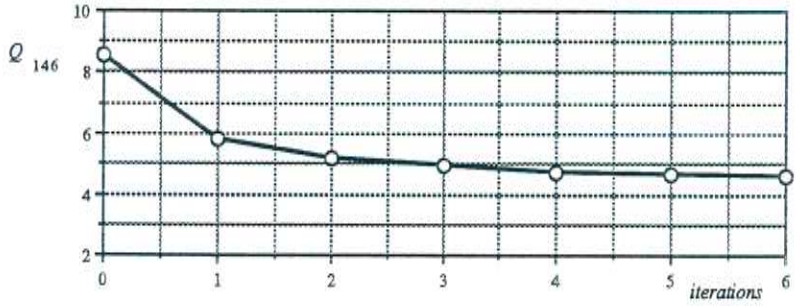
D.7 Experiment E.2.1

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	50
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

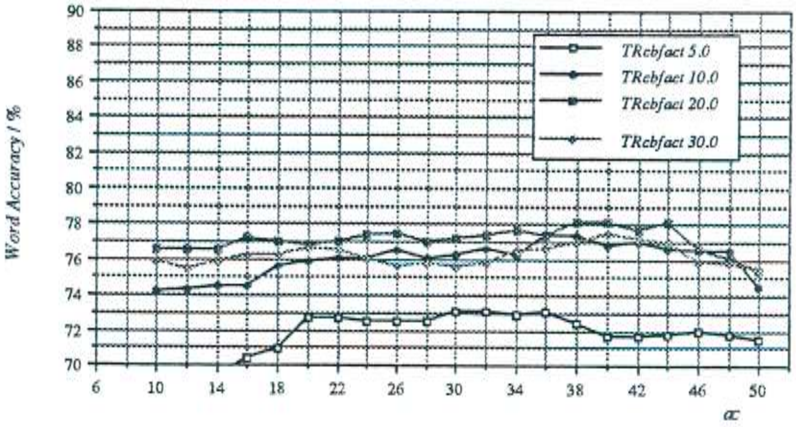
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	8.560	5.834	5.134	4.934	4.734	4.665	4.596
Q_{50}	13.796	-	-	-	-	-	8.954



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	20.0	30.0
10.0	64.5% (65.2%)	74.2% (75.0%)	76.5% (76.8%)	75.9% (77.0%)
12.0	67.4% (68.1%)	74.3% (75.2%)	76.5% (77.7%)	75.4% (76.6%)
14.0	69.3% (70.1%)	74.5% (75.6%)	76.5% (78.4%)	75.9% (76.8%)
16.0	70.4% (72.0%)	74.5% (76.1%)	77.2% (78.8%)	76.3% (77.4%)
18.0	70.9% (74.2%)	75.6% (77.5%)	77.0% (77.7%)	76.3% (77.9%)
20.0	72.7% (74.2%)	75.9% (77.9%)	76.8% (78.8%)	76.6% (78.3%)
22.0	72.7% (74.3%)	76.1% (77.9%)	77.0% (79.0%)	76.5% (78.3%)
24.0	72.5% (74.5%)	76.1% (77.9%)	77.4% (79.9%)	76.1% (78.4%)
26.0	72.5% (75.0%)	76.5% (78.4%)	77.5% (80.2%)	75.6% (78.4%)
28.0	72.5% (75.0%)	76.1% (78.3%)	77.0% (79.7%)	75.8% (79.0%)
30.0	73.1% (75.6%)	76.3% (78.4%)	77.2% (80.6%)	75.6% (78.8%)
32.0	73.1% (75.6%)	76.6% (79.1%)	77.4% (80.7%)	75.8% (79.3%)
34.0	72.9% (75.6%)	76.3% (78.8%)	77.7% (81.1%)	76.5% (80.0%)
36.0	73.1% (75.9%)	77.4% (80.2%)	77.4% (80.9%)	76.6% (80.2%)
38.0	72.4% (75.4%)	77.4% (79.7%)	78.1% (80.9%)	77.0% (80.4%)
40.0	71.7% (74.9%)	76.8% (79.5%)	78.1% (80.9%)	77.5% (80.6%)
42.0	71.7% (74.9%)	77.0% (79.7%)	77.7% (80.6%)	77.2% (80.4%)
44.0	71.8% (75.2%)	76.6% (79.5%)	78.1% (80.9%)	77.0% (80.6%)
46.0	72.0% (75.4%)	76.5% (79.5%)	76.6% (80.6%)	75.9% (79.9%)
48.0	71.8% (75.2%)	76.5% (79.7%)	76.1% (80.6%)	75.8% (79.9%)
50.0	71.5% (75.0%)	74.5% (78.6%)	75.4% (79.7%)	75.4% (79.9%)



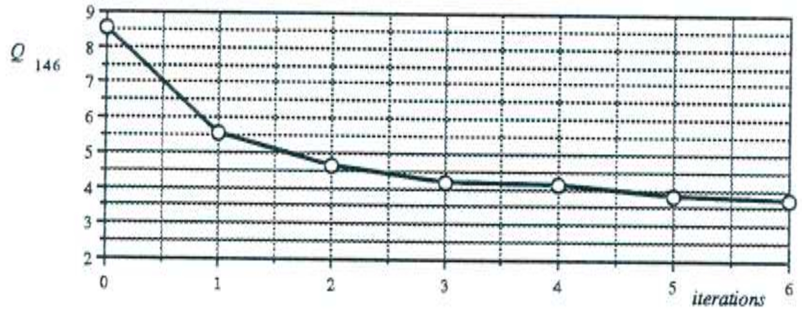
D.8 Experiment E.2.2

NLDA approach : $Y = Af(X)$

```
genaral parameters                                        backpropagation parameters
   dimension of orignial space :             32              number of discriminated targets :     50
   dimension of image space :                16              learning rate :                       0.008
   discriminated classes in on top LDA :    146              momentum :                            0.9
target drift parameters                                      number of hidden units :              32
   α :                                       0.9             iterations :                          6
   β :                                       0.1             target update after :                 6
   m :                                       1               sample selection :                    random
   number of drift vectors :                 50 (one per phonem class)
```

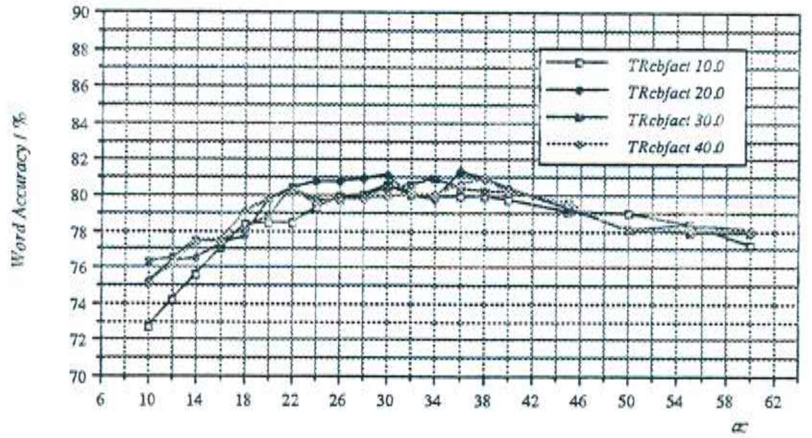
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	8.560	5.570	4.660	4.201	4.114	3.816	3.723
Q_{50}	13.798	-	-	-	-	-	7.497



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	10.0	20.0	30.0	40.0
10.0	72.7% (73.8%)	75.2% (76.3%)	76.3% (77.7%)	75.0% (76.3%)
12.0	74.2% (75.2%)	76.3% (77.5%)	76.5% (77.9%)	76.1% (77.2%)
14.0	75.6% (76.8%)	76.5% (77.7%)	77.4% (78.8%)	77.4% (78.4%)
16.0	77.0% (78.3%)	77.2% (78.4%)	77.4% (78.6%)	77.5% (78.6%)
18.0	78.3% (79.5%)	77.7% (79.1%)	78.1% (79.3%)	79.1% (80.6%)
20.0	78.4% (80.4%)	79.7% (81.1%)	78.8% (80.4%)	79.7% (81.1%)
22.0	78.4% (80.6%)	80.4% (82.0%)	80.4% (81.8%)	80.2% (81.8%)
24.0	79.3% (81.6%)	80.7% (82.4%)	79.7% (82.2%)	79.5% (82.0%)
26.0	79.9% (82.2%)	80.7% (82.5%)	79.7% (82.2%)	79.7% (82.2%)
28.0	80.0% (82.4%)	80.9% (82.7%)	79.9% (82.4%)	79.7% (82.2%)
30.0	80.8% (83.1%)	81.1% (82.9%)	80.4% (83.1%)	79.9% (82.4%)
32.0	80.0% (82.5%)	80.0% (82.5%)	80.6% (83.6%)	80.0% (82.9%)
34.0	79.9% (82.7%)	79.7% (82.5%)	80.9% (84.0%)	80.0% (82.9%)
36.0	79.9% (82.9%)	81.3% (84.1%)	80.4% (83.6%)	80.6% (83.4%)
38.0	79.9% (82.9%)	80.9% (83.8%)	80.2% (83.4%)	80.9% (84.1%)
40.0	79.7% (82.9%)	80.4% (83.6%)	80.2% (83.4%)	80.2% (83.4%)
45.0	79.1% (82.0%)	79.3% (82.7%)	79.5% (83.4%)	79.5% (83.4%)
50.0	79.0% (82.4%)	78.1% (82.7%)	78.1% (82.5%)	78.1% (82.5%)
55.0	78.3% (81.8%)	78.3% (82.7%)	77.9% (81.8%)	78.1% (82.4%)
60.0	77.2% (80.9%)	78.1% (82.2%)	77.9% (81.8%)	78.1% (82.4%)



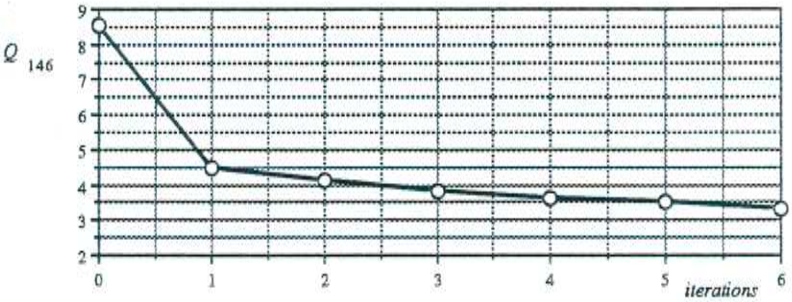
D.9 Experiment E.2.3

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	50
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

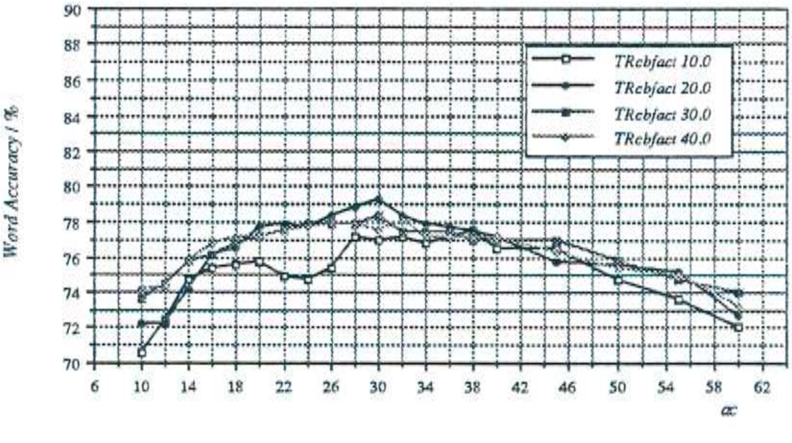
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	8.560	4.510	4.115	3.842	3.608	3.524	3.313
Q_{50}	13.796	-	-	-	-	-	7.009



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRebfact			
ac	10.0	20	30	40
10.0	70.6% (71.8%)	72.2% (74.5%)	73.6% (75.2%)	74.2% (75.6%)
12.0	72.4% (73.4%)	72.7% (74.9%)	74.5% (76.5%)	74.3% (75.8%)
14.0	74.7% (76.5%)	74.2% (76.1%)	75.8% (77.7%)	75.8% (77.9%)
16.0	75.4% (77.2%)	76.1% (78.1%)	76.1% (78.4%)	76.8% (79.3%)
18.0	75.6% (77.7%)	76.5% (78.4%)	76.8% (79.3%)	77.2% (79.7%)
20.0	75.8% (77.9%)	77.7% (80.2%)	77.2% (79.7%)	77.2% (79.9%)
22.0	74.9% (78.3%)	77.9% (80.4%)	77.5% (80.6%)	77.5% (80.7%)
24.0	74.7% (78.1%)	77.7% (80.6%)	77.9% (80.9%)	77.7% (80.7%)
26.0	75.4% (79.0%)	78.4% (81.6%)	77.9% (80.9%)	77.9% (81.3%)
28.0	77.2% (80.7%)	78.8% (82.0%)	77.9% (81.3%)	77.9% (81.3%)
30.0	77.0% (80.7%)	79.3% (82.5%)	78.4% (81.8%)	77.5% (81.3%)
32.0	77.2% (80.9%)	78.4% (82.0%)	77.4% (81.6%)	78.1% (82.0%)
34.0	76.8% (80.6%)	77.9% (81.5%)	77.4% (81.6%)	77.2% (81.5%)
36.0	77.2% (80.9%)	77.7% (81.5%)	77.4% (81.6%)	77.2% (81.5%)
38.0	77.5% (81.3%)	77.5% (81.1%)	77.0% (81.3%)	77.2% (81.5%)
40.0	76.5% (80.6%)	77.2% (81.1%)	77.0% (81.3%)	77.2% (81.5%)
45.0	76.5% (80.9%)	75.8% (80.4%)	77.0% (81.5%)	76.3% (81.5%)
50.0	74.7% (80.0%)	75.6% (80.2%)	75.9% (80.7%)	75.4% (80.7%)
55.0	73.6% (79.0%)	75.2% (80.2%)	74.7% (80.2%)	75.0% (80.6%)
60.0	72.0% (78.4%)	72.7% (79.0%)	74.0% (79.9%)	73.3% (79.3%)



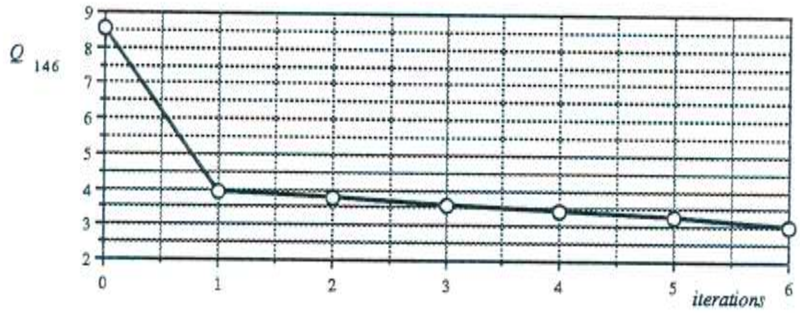
D.10 Experiment E.2.4

NLDA approach : $Y = Af(X)$

```
genaral parameters                                           backpropagation parameters
    dimension of orignial space :              32                number of discriminated targets :    50
    dimension of image space :                 16                learning rate :                      0.008
    discriminated classes in on top LDA :      146               momentum :                           0.9
target drift parameters                                          number of hidden units :             32
    α :                                        0.9               iterations :                         6
    β :                                        0.3               target update after :                6
    m :                                        1                 sample selection :                   random
    number of drift vectors :                  50 (one per phonem class)
```

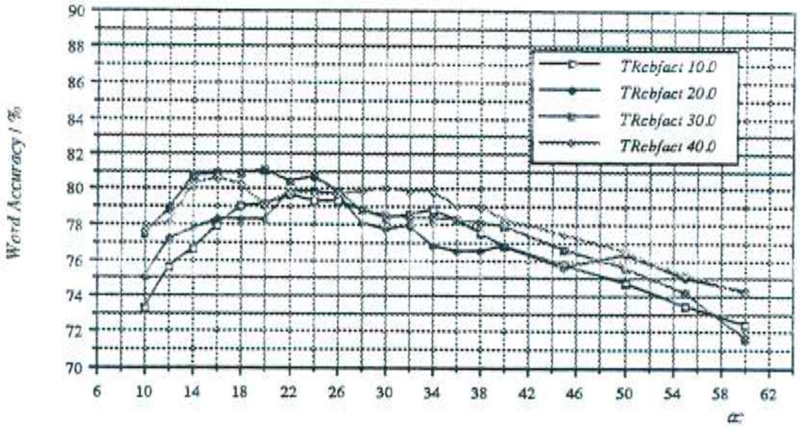
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	8.560	3.903	3.800	3.586	3.399	3.271	3.020
Q_{50}	13.796	-	-	-	-	-	6.724



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	10.0	20.0	30.0	40.0
10.0	73.3% (74.0%)	75.0% (76.3%)	77.4% (78.3%)	77.7% (78.4%)
12.0	75.6% (76.3%)	77.2% (78.3%)	78.8% (79.9%)	78.1% (79.0%)
14.0	76.6% (77.9%)	77.7% (79.0%)	80.7% (81.8%)	80.2% (81.3%)
16.0	77.9% (78.8%)	78.3% (79.7%)	80.9% (82.2%)	80.6% (81.6%)
18.0	79.0% (80.4%)	78.3% (79.7%)	80.9% (82.5%)	80.2% (81.6%)
20.0	79.1% (80.6%)	78.3% (79.9%)	81.1% (81.8%)	79.0% (80.0%)
22.0	79.7% (81.8%)	79.9% (81.6%)	80.4% (82.5%)	79.9% (81.5%)
24.0	79.3% (81.3%)	79.9% (81.6%)	80.7% (82.2%)	79.7% (81.6%)
26.0	79.3% (81.3%)	79.7% (81.6%)	79.9% (81.6%)	79.7% (82.2%)
28.0	78.8% (80.9%)	78.1% (80.9%)	78.8% (81.5%)	79.9% (82.7%)
30.0	78.4% (80.7%)	77.7% (80.7%)	78.6% (81.5%)	80.0% (82.9%)
32.0	78.6% (80.9%)	77.9% (81.6%)	78.4% (81.5%)	79.9% (82.9%)
34.0	78.8% (81.6%)	76.8% (81.1%)	78.4% (82.0%)	79.9% (83.1%)
36.0	78.3% (81.1%)	76.5% (81.1%)	78.3% (81.8%)	79.0% (82.7%)
38.0	77.5% (80.8%)	76.5% (81.1%)	78.1% (82.0%)	79.0% (82.9%)
40.0	76.8% (79.9%)	76.8% (81.5%)	77.9% (81.8%)	78.3% (82.7%)
45.0	75.8% (79.0%)	75.6% (80.7%)	76.6% (81.6%)	77.4% (82.4%)
50.0	74.7% (78.6%)	76.3% (81.6%)	75.6% (81.1%)	76.5% (82.0%)
55.0	73.4% (78.1%)	75.0% (81.5%)	74.2% (80.2%)	75.2% (81.8%)
60.0	72.4% (77.4%)	74.3% (81.1%)	71.7% (79.0%)	74.2% (80.9%)



D.11 Experiment E.2.5

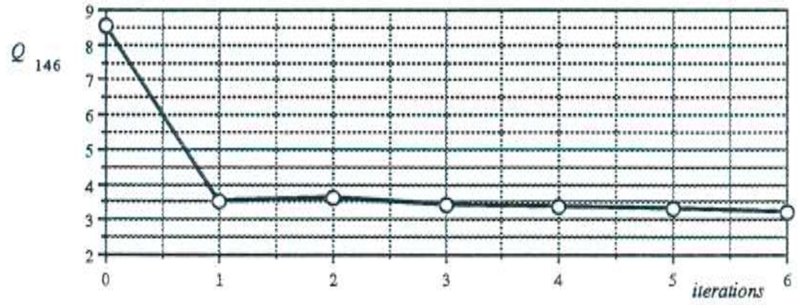
NLDA approach : $Y = Af(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.4
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	50
learning rate :	0.008
momentum :	0.9
number of hidden units :	32
iterations :	6
target update after :	6
sample selection :	random

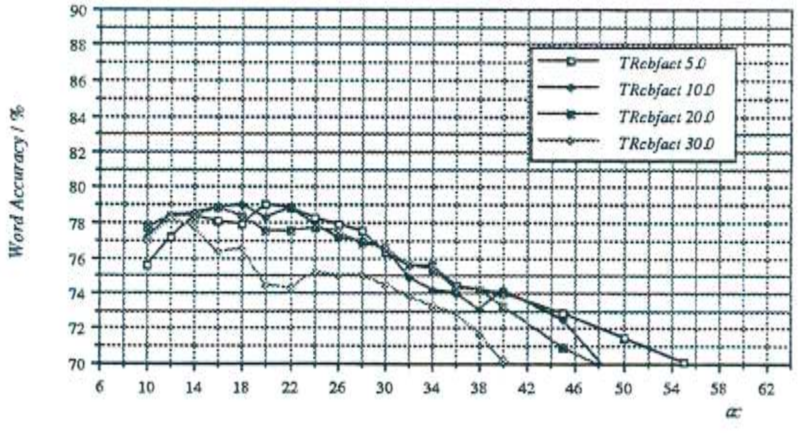
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	8.560	3.541	3.609	3.429	3.387	3.295	3.234
Q_{50}	13.796	-	-	-	-	-	6.444



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRebfact			
ac	5.0	10.0	20.0	30.0
10.0	75.6% (76.5%)	77.2% (78.6%)	77.7% (79.1%)	77.0% (78.8%)
12.0	77.2% (77.7%)	78.4% (79.9%)	78.4% (80.0%)	78.3% (80.0%)
14.0	78.4% (79.3%)	78.6% (80.2%)	78.4% (80.7%)	77.7% (79.5%)
16.0	78.1% (79.5%)	78.8% (80.6%)	78.8% (81.1%)	76.3% (79.0%)
18.0	77.9% (79.1%)	79.0% (80.7%)	78.4% (80.7%)	76.5% (79.5%)
20.0	79.0% (80.7%)	78.3% (80.9%)	77.5% (79.9%)	74.5% (78.1%)
22.0	78.8% (80.6%)	78.8% (81.6%)	77.5% (80.9%)	74.3% (78.4%)
24.0	78.3% (80.2%)	77.9% (81.1%)	77.7% (81.5%)	75.2% (79.5%)
26.0	77.9% (80.2%)	77.2% (80.9%)	77.5% (81.3%)	75.0% (79.3%)
28.0	77.5% (80.0%)	77.0% (80.9%)	76.8% (81.5%)	75.0% (80.0%)
30.0	76.3% (79.5%)	76.5% (80.7%)	76.6% (81.5%)	74.5% (79.7%)
32.0	75.6% (79.1%)	74.9% (79.1%)	75.6% (80.7%)	73.8% (79.1%)
34.0	75.6% (79.3%)	74.2% (78.8%)	75.4% (80.6%)	73.3% (79.0%)
36.0	74.5% (79.0%)	74.0% (79.1%)	74.3% (79.9%)	72.9% (78.8%)
38.0	74.2% (78.8%)	73.1% (78.8%)	74.2% (80.0%)	71.7% (78.1%)
40.0	74.0% (78.6%)	74.2% (79.7%)	73.3% (79.0%)	70.2% (77.0%)
45.0	72.9% (79.1%)	72.5% (78.6%)	70.9% (77.5%)	68.6% (76.1%)
50.0	71.5% (78.6%)	68.6% (76.8%)	69.3% (77.2%)	68.6% (76.1%)
55.0	70.1% (77.2%)	67.0% (75.6%)	66.1% (74.9%)	62.9% (72.2%)
60.0	68.1% (76.5%)	64.5% (74.5%)	64.3% (72.9%)	61.7% (71.7%)



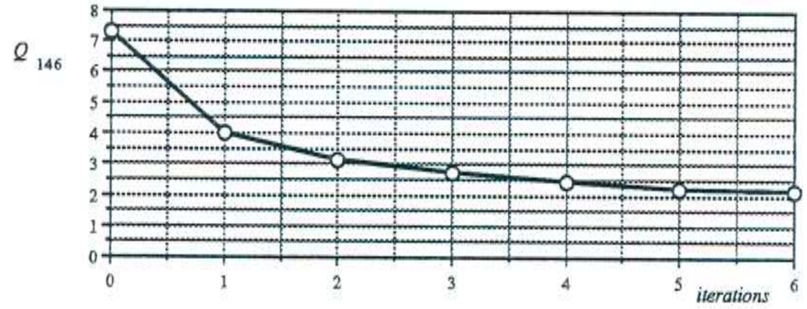
D.12 Experiment E.3.1

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

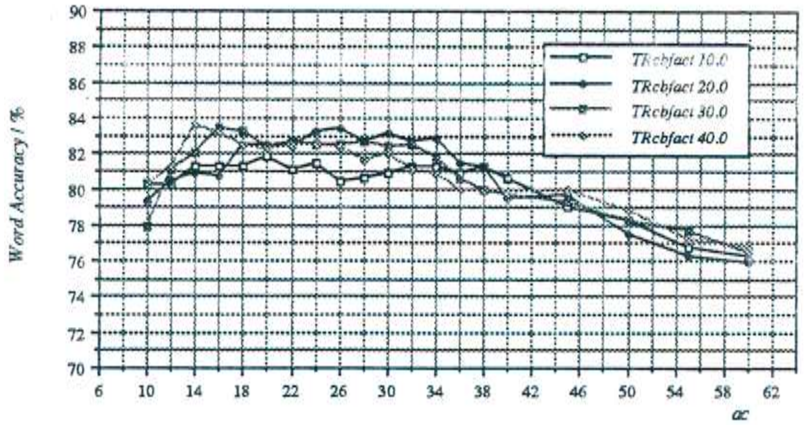
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	4.026	3.128	2.725	2.411	2.230	2.119
Q_{50}	15.860	-	-	-	-	-	9.501



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	10.0	20.0	30.0	40.0
10	80.2% (81.3%)	79.3% (80.6%)	77.9% (79.1%)	80.2% (81.5%)
12	80.2% (81.8%)	80.4% (81.1%)	81.1% (82.5%)	81.3% (82.7%)
14	81.3% (82.5%)	80.9% (82.4%)	82.0% (83.1%)	83.6% (84.7%)
16	81.3% (82.5%)	80.7% (82.2%)	83.4% (84.8%)	83.1% (84.7%)
18	81.3% (83.1%)	82.5% (84.3%)	83.2% (84.8%)	82.5% (84.5%)
20	81.8% (83.6%)	82.5% (84.5%)	82.4% (84.3%)	82.2% (84.1%)
22	81.1% (83.2%)	82.4% (84.3%)	82.7% (84.7%)	82.5% (84.5%)
24	81.5% (83.6%)	83.2% (85.2%)	82.5% (84.5%)	82.5% (84.5%)
26	80.4% (83.2%)	83.4% (85.4%)	82.5% (84.7%)	82.4% (84.5%)
28	80.6% (83.4%)	82.7% (85.2%)	82.7% (84.8%)	81.6% (84.0%)
30	80.9% (83.6%)	83.1% (85.6%)	82.4% (84.8%)	82.0% (84.1%)
32	81.3% (84.0%)	82.7% (85.2%)	82.5% (85.0%)	81.1% (83.6%)
34	81.3% (84.0%)	82.9% (85.6%)	81.8% (84.5%)	80.9% (83.4%)
36	80.9% (84.0%)	81.5% (84.3%)	80.6% (83.6%)	80.0% (82.7%)
38	81.3% (84.5%)	81.3% (84.1%)	80.0% (83.2%)	79.9% (83.1%)
40	80.6% (84.0%)	79.5% (83.2%)	79.7% (83.4%)	79.9% (83.1%)
45	79.0% (83.1%)	79.7% (83.8%)	79.3% (83.4%)	80.0% (83.6%)
50	78.3% (82.4%)	77.5% (82.7%)	76.3% (82.9%)	78.8% (83.6%)
55	76.8% (82.2%)	76.3% (81.5%)	77.7% (82.7%)	77.2% (82.5%)
60	76.3% (81.6%)	75.9% (80.9%)	76.5% (82.4%)	76.8% (82.9%)



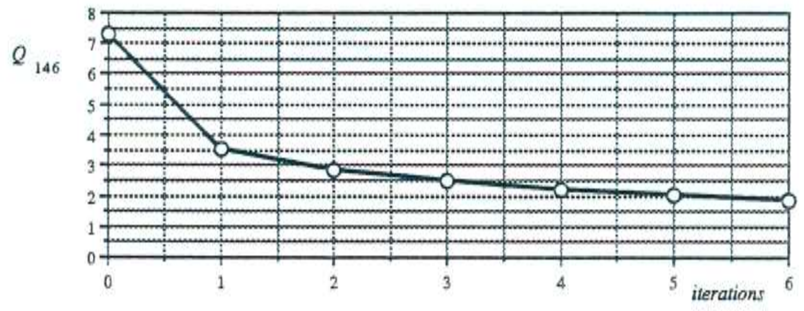
D.13 Experiment E.3.2

NLDA approach : $Y = Af(X)$

```
genaral parameters                                    backpropagation parameters
   dimension of orignial space :           32            number of discriminated targets :    146
   dimension of image space :              16            learning rate :                      0.008
   discriminated classes in on top LDA : 146             momentum :                           0.9
target drift parameters                                  number of hidden units :             32
   α :                                     0.9           iterations :                         6
   β :                                     0.2           target update after :                6
   m :                                     1             sample selection :                   random
   number of drift vectors :               50 (one per phonem class)
```

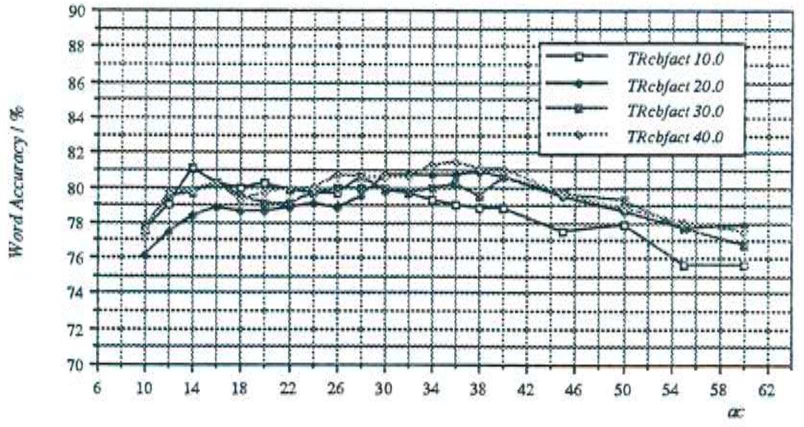
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	3.512	2.816	2.476	2.207	2.025	1.884
Q_{Kn}	15.860	-	-	-	-	-	8.619



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	10.0	20.0	30.0	40.0
10	77.5% (79.0%)	76.1% (77.5%)	77.7% (79.1%)	77.5% (79.0%)
12	79.0% (80.4%)	77.4% (78.8%)	79.5% (80.6%)	79.7% (80.7%)
14	81.1% (82.5%)	78.4% (79.9%)	79.7% (80.9%)	79.9% (80.9%)
16	80.2% (82.2%)	78.8% (80.2%)	80.2% (81.5%)	80.2% (81.5%)
18	80.0% (82.4%)	78.6% (80.4%)	79.5% (81.6%)	79.3% (81.1%)
20	80.2% (82.5%)	78.6% (80.4%)	79.1% (81.3%)	79.7% (82.0%)
22	79.9% (82.4%)	78.8% (80.9%)	79.1% (81.5%)	80.0% (82.5%)
24	79.7% (82.5%)	79.1% (81.5%)	79.7% (82.0%)	80.0% (82.5%)
26	79.7% (82.7%)	78.8% (81.3%)	80.0% (82.4%)	80.7% (83.2%)
28	80.6% (83.6%)	79.5% (82.0%)	80.0% (82.5%)	80.6% (83.1%)
30	79.9% (83.4%)	80.7% (83.2%)	80.0% (82.5%)	80.6% (83.1%)
32	79.7% (83.2%)	80.7% (83.2%)	79.7% (82.2%)	80.6% (83.1%)
34	79.3% (82.9%)	80.7% (83.2%)	80.0% (82.5%)	81.3% (83.8%)
36	79.0% (82.7%)	80.7% (83.6%)	80.2% (82.7%)	81.5% (84.0%)
38	78.8% (82.5%)	80.9% (83.8%)	79.5% (82.4%)	81.1% (83.6%)
40	78.8% (82.4%)	80.6% (83.6%)	80.6% (83.1%)	81.1% (83.4%)
45	77.5% (82.0%)	79.5% (83.4%)	79.7% (83.4%)	79.7% (83.4%)
50	77.9% (82.7%)	78.6% (82.5%)	79.3% (83.1%)	78.8% (82.9%)
55	75.6% (81.6%)	77.7% (82.2%)	77.7% (82.0%)	78.1% (82.4%)
60	75.6% (81.5%)	77.9% (82.2%)	76.8% (81.6%)	77.5% (82.0%)



D.14 Experiment E.3.3

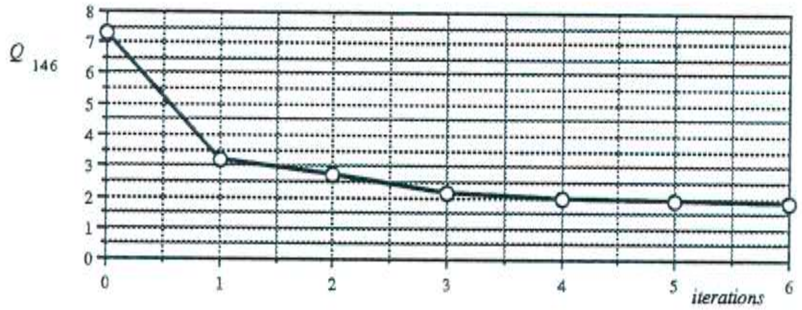
NLDA approach : $Y = Af(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.3
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	32
iterations :	6
target update after :	6
sample selection :	random

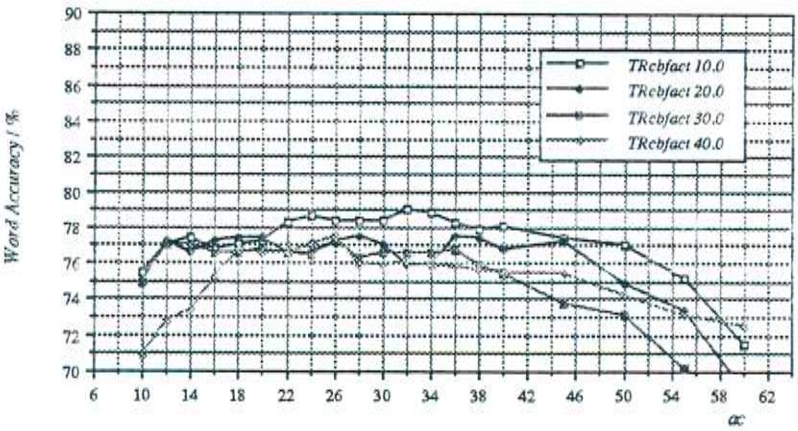
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	3.203	2.696	2.167	1.972	1.898	1.874
Q_{50}	15.860	-	-	-	-	-	8.089



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	10.0	20.0	30.0	40.0
10	75.4% (77.4%)	75.0% (77.0%)	74.9% (76.8%)	70.8% (72.7%)
12	77.0% (78.8%)	77.2% (78.8%)	77.0% (78.6%)	72.7% (74.3%)
14	77.4% (79.3%)	76.6% (79.0%)	77.0% (78.8%)	73.4% (75.0%)
16	76.8% (79.3%)	77.2% (79.5%)	76.6% (79.0%)	75.2% (76.8%)
18	77.0% (79.3%)	77.4% (79.5%)	76.6% (78.6%)	76.8% (78.4%)
20	77.2% (79.7%)	77.4% (79.5%)	76.8% (79.0%)	76.6% (78.6%)
22	78.3% (80.7%)	76.6% (79.5%)	76.6% (79.1%)	76.6% (78.8%)
24	78.6% (81.1%)	76.6% (79.5%)	76.5% (79.0%)	77.0% (79.1%)
26	78.4% (80.9%)	77.2% (80.0%)	77.2% (79.5%)	77.5% (79.3%)
28	78.4% (80.9%)	77.5% (80.4%)	76.3% (79.0%)	75.9% (78.3%)
30	78.4% (80.9%)	77.0% (80.0%)	76.6% (79.3%)	75.9% (78.3%)
32	79.0% (81.8%)	75.9% (79.5%)	76.6% (79.5%)	76.1% (78.6%)
34	78.8% (81.8%)	75.9% (79.5%)	76.6% (79.5%)	75.8% (78.8%)
36	78.3% (81.6%)	77.5% (81.1%)	76.8% (79.9%)	75.8% (79.0%)
38	77.9% (81.3%)	77.4% (81.1%)	75.9% (79.1%)	75.6% (79.1%)
40	78.1% (81.5%)	76.8% (80.9%)	75.4% (79.3%)	75.4% (78.8%)
45	77.4% (81.6%)	77.2% (81.3%)	73.8% (78.8%)	75.4% (79.5%)
50	77.0% (81.5%)	74.9% (80.2%)	73.1% (78.6%)	74.2% (78.8%)
55	75.2% (80.4%)	73.4% (79.9%)	70.2% (77.5%)	73.1% (78.8%)
60	71.5% (78.6%)	69.0% (77.0%)	67.6% (75.8%)	72.5% (78.6%)



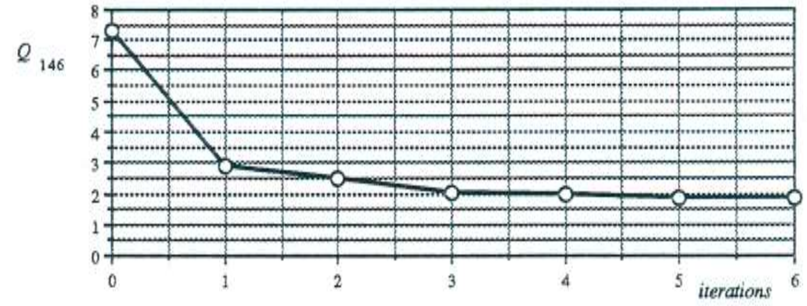
D.15 Experiment E.3.4

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.4	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

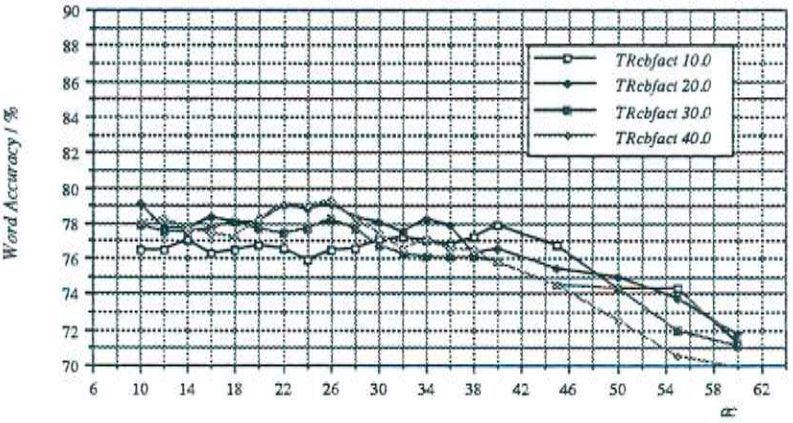
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	2.918	2.470	2.019	1.943	1.880	1.868
Q_{50}	15.860	-	-	-	-	-	7.829



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	10.0	20.0	30.0	40.0
10	76.5% (78.3%)	79.1% (80.2%)	77.9% (79.0%)	78.1% (79.7%)
12	76.5% (78.3%)	77.7% (79.5%)	77.5% (79.1%)	78.3% (79.9%)
14	77.0% (79.0%)	77.7% (79.7%)	77.5% (79.5%)	77.7% (79.5%)
16	76.3% (78.4%)	78.4% (80.4%)	77.7% (79.3%)	77.4% (79.1%)
18	76.5% (78.6%)	78.1% (80.2%)	78.1% (80.0%)	77.2% (79.1%)
20	76.8% (79.1%)	78.1% (80.2%)	77.7% (79.7%)	78.3% (80.4%)
22	76.6% (79.0%)	79.0% (81.3%)	77.4% (79.5%)	79.1% (81.3%)
24	75.9% (78.6%)	78.8% (81.1%)	77.7% (79.9%)	79.0% (81.3%)
26	76.5% (79.1%)	79.1% (81.5%)	78.3% (80.4%)	79.3% (81.5%)
28	76.6% (79.5%)	78.4% (81.1%)	77.7% (80.4%)	78.3% (80.9%)
30	77.0% (79.9%)	78.1% (81.1%)	76.8% (79.7%)	77.4% (80.2%)
32	77.2% (80.2%)	77.5% (80.6%)	76.3% (79.3%)	76.6% (79.9%)
34	77.0% (80.0%)	78.3% (81.5%)	76.1% (79.1%)	77.0% (80.2%)
36	76.8% (80.0%)	77.9% (81.5%)	76.1% (79.5%)	76.6% (80.2%)
38	77.2% (80.7%)	76.3% (80.2%)	76.1% (80.0%)	76.5% (80.2%)
40	77.9% (81.8%)	76.6% (80.9%)	75.8% (80.0%)	75.8% (80.2%)
45	76.8% (81.1%)	75.4% (80.7%)	74.5% (80.2%)	74.5% (80.4%)
50	74.3% (79.5%)	75.0% (80.6%)	74.3% (80.6%)	72.5% (79.9%)
55	74.3% (79.9%)	73.8% (79.9%)	72.0% (79.1%)	70.6% (78.3%)
60	71.3% (78.8%)	71.8% (79.5%)	71.1% (77.9%)	69.9% (77.4%)



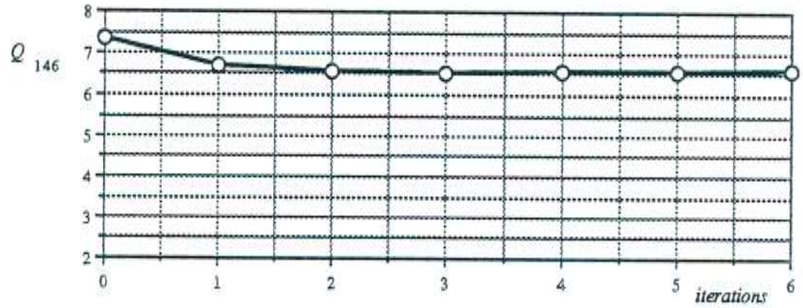
D.16 Experiment E.4.1

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	5
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

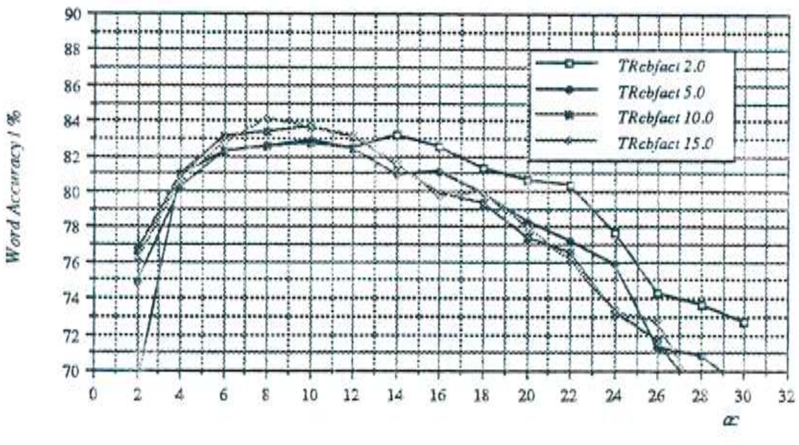
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.898	6.572	6.539	6.577	6.555	6.617
Q_{50}	15.860	-	-	-	-	-	14.352



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRebfact			
ac	2.0	5.0	10.0	15.0
2.0	69.2% (69.7%)	74.9% (75.8%)	76.6% (77.4%)	76.1% (76.8%)
4.0	80.9% (81.8%)	80.2% (81.1%)	80.9% (81.6%)	80.7% (81.8%)
6.0	82.2% (83.4%)	82.2% (83.6%)	83.1% (84.5%)	82.7% (84.0%)
8.0	82.5% (84.3%)	82.5% (84.5%)	83.4% (85.2%)	84.1% (85.9%)
10.0	82.7% (85.2%)	82.9% (85.0%)	83.6% (85.6%)	83.6% (85.6%)
12.0	82.4% (85.0%)	82.4% (85.0%)	83.1% (85.7%)	83.2% (85.4%)
14.0	83.2% (85.9%)	80.9% (84.7%)	81.5% (84.8%)	81.6% (85.2%)
16.0	82.5% (85.7%)	81.1% (85.0%)	79.9% (84.1%)	79.7% (83.8%)
18.0	81.3% (85.7%)	79.9% (84.0%)	79.3% (83.8%)	79.9% (84.0%)
20.0	80.7% (85.6%)	78.3% (82.7%)	77.4% (82.7%)	77.9% (83.1%)
22.0	80.4% (85.9%)	77.2% (81.6%)	76.5% (82.4%)	76.1% (81.1%)
24.0	77.7% (82.9%)	75.9% (80.6%)	73.3% (79.3%)	73.3% (79.0%)
26.0	74.3% (79.5%)	71.3% (76.8%)	71.7% (78.1%)	72.7% (78.4%)
28.0	73.6% (79.0%)	70.8% (76.5%)	68.3% (75.8%)	68.4% (76.3%)
30.0	72.7% (78.1%)	69.2% (75.8%)	68.6% (75.6%)	69.0% (75.6%)



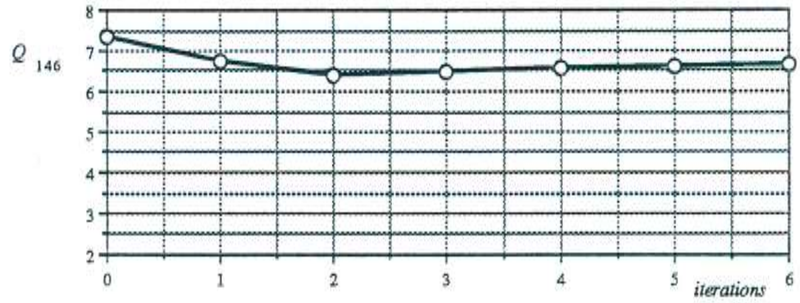
D.17 Experiment E.4.2

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	5
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

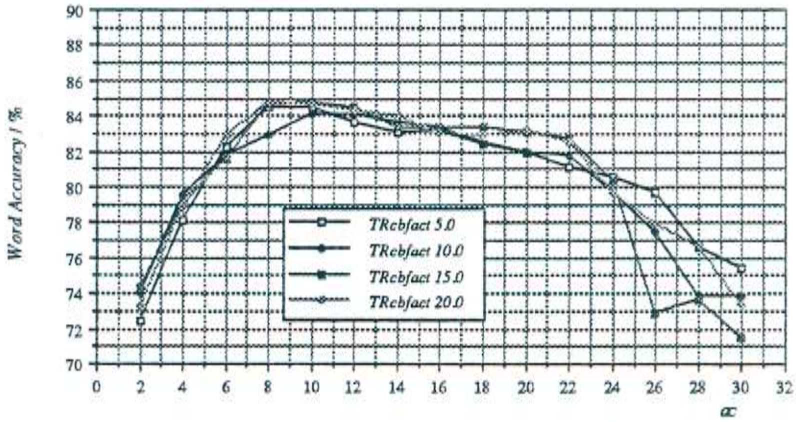
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.755	6.390	6.466	6.560	6.626	6.652
Q_{50}	15.860	-	-	-	-	-	14.974



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	15.0	20.0
2.0	72.4% (73.4%)	74.0% (75.0%)	74.3% (75.2%)	73.3% (74.3%)
4.0	78.1% (79.0%)	79.5% (80.4%)	79.1% (80.2%)	79.0% (80.0%)
6.0	82.2% (83.2%)	81.8% (83.1%)	81.6% (82.7%)	82.9% (84.1%)
8.0	84.5% (86.8%)	82.9% (85.2%)	84.7% (87.0%)	84.8% (87.0%)
10.0	84.5% (87.0%)	84.1% (86.5%)	84.8% (87.3%)	84.7% (87.0%)
12.0	83.6% (87.2%)	84.1% (87.3%)	84.5% (87.7%)	84.3% (87.3%)
14.0	83.1% (86.6%)	83.6% (87.2%)	83.4% (87.3%)	84.0% (87.5%)
16.0	83.2% (87.2%)	83.4% (87.5%)	83.4% (87.5%)	83.1% (87.2%)
18.0	82.4% (86.1%)	82.5% (87.0%)	83.4% (87.7%)	82.9% (87.0%)
20.0	82.0% (85.9%)	82.0% (85.9%)	83.1% (87.0%)	83.2% (87.0%)
22.0	81.1% (84.5%)	81.8% (85.6%)	82.7% (86.1%)	82.5% (86.3%)
24.0	80.6% (84.3%)	79.7% (84.5%)	80.4% (85.0%)	79.7% (85.4%)
26.0	79.7% (84.1%)	77.5% (83.2%)	72.9% (83.8%)	77.9% (83.8%)
28.0	76.6% (82.5%)	73.8% (80.4%)	73.6% (80.6%)	76.6% (83.2%)
30.0	75.4% (81.5%)	73.8% (80.7%)	71.5% (79.1%)	73.4% (81.1%)



D.18 Experiment E.4.3

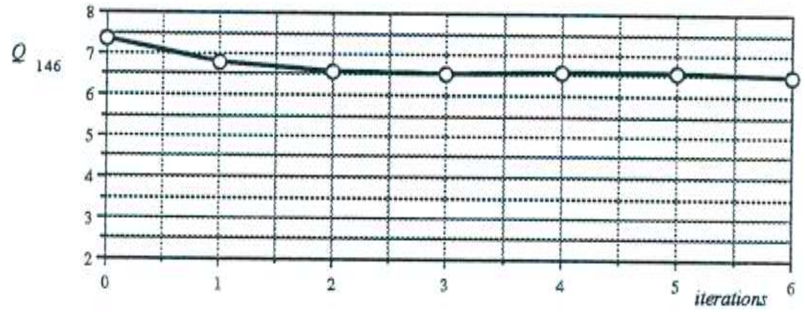
NLDA approach : $Y = A(X) + f(x)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.2
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	5
iterations :	6
target update after :	6
sample selection :	random

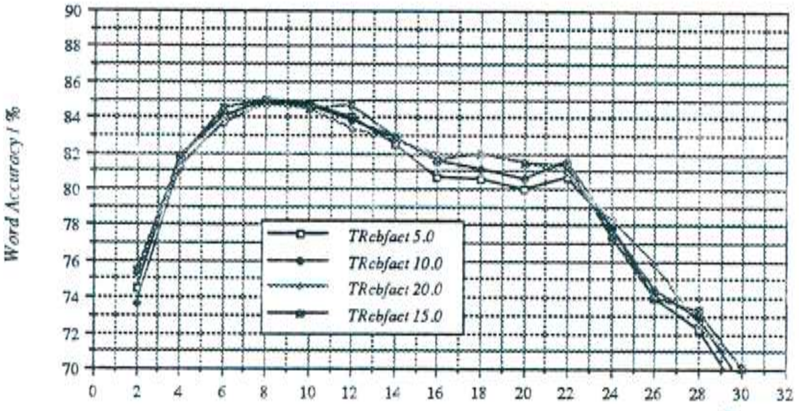
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.798	6.544	6.529	6.581	6.545	6.494
Q_{50}	15.860	-	-	-	-	-	14.124



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	15.0	20.0
2.0	74.5% (75.4%)	73.6% (74.7%)	75.4% (75.0%)	75.2% (76.5%)
4.0	81.8% (82.7%)	81.1% (82.2%)	81.6% (82.5%)	81.3% (82.5%)
6.0	84.0% (85.2%)	83.6% (85.0%)	84.5% (85.7%)	83.8% (84.8%)
8.0	85.0% (86.6%)	84.8% (87.0%)	85.0% (87.0%)	84.7% (86.5%)
10.0	84.8% (87.2%)	84.7% (87.3%)	84.5% (87.5%)	84.5% (87.5%)
12.0	84.0% (86.8%)	83.8% (87.7%)	84.7% (87.9%)	83.4% (87.2%)
14.0	82.5% (86.5%)	82.9% (87.2%)	82.9% (87.3%)	82.7% (87.2%)
16.0	80.7% (84.8%)	81.6% (86.6%)	81.6% (86.5%)	82.0% (86.5%)
18.0	80.6% (84.5%)	81.1% (85.6%)	82.0% (86.1%)	82.0% (86.5%)
20.0	80.0% (84.1%)	80.6% (85.2%)	81.5% (85.9%)	81.5% (86.1%)
22.0	80.7% (84.8%)	81.6% (85.6%)	81.3% (85.4%)	80.9% (85.7%)
24.0	77.9% (83.4%)	77.9% (83.8%)	77.4% (83.1%)	78.3% (83.8%)
26.0	74.0% (80.4%)	74.5% (80.9%)	74.0% (81.5%)	75.9% (82.4%)
28.0	72.2% (79.1%)	72.9% (79.9%)	73.3% (80.6%)	72.5% (80.0%)
30.0	68.4% (76.5%)	69.3% (77.0%)	70.1% (78.1%)	70.2% (78.1%)



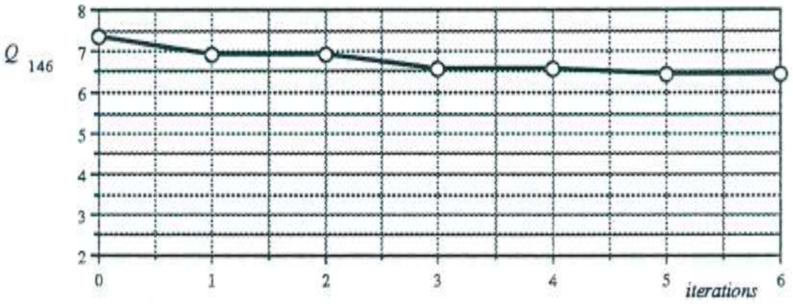
D.19 Experiment E.4.4

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	8
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

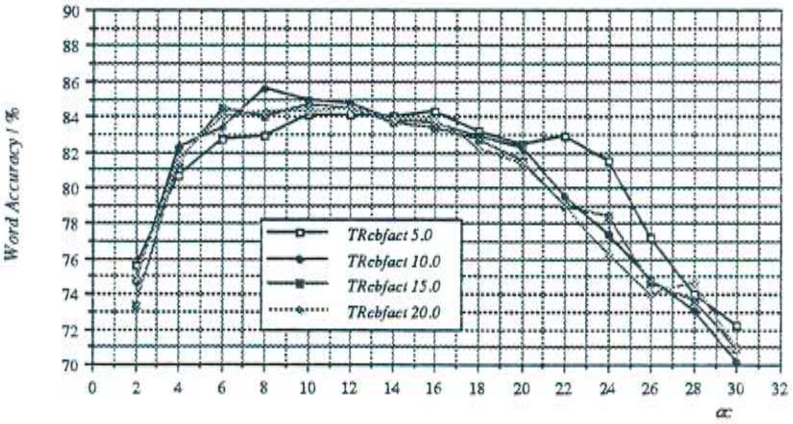
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.899	6.910	6.583	6.571	6.424	6.442
Q_{50}	15.860	-	-	-	-	-	14.039



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	15.0	20.0
2.0	75.6% (76.8%)	74.7% (75.2%)	73.4% (74.3%)	74.9% (75.6%)
4.0	80.7% (81.6%)	82.2% (83.2%)	81.3% (82.2%)	81.6% (82.5%)
6.0	82.7% (84.0%)	83.4% (85.2%)	84.5% (86.3%)	84.0% (85.2%)
8.0	82.9% (84.8%)	85.6% (87.3%)	84.0% (86.3%)	84.3% (86.5%)
10.0	84.1% (86.8%)	85.0% (87.7%)	84.7% (87.0%)	84.3% (86.8%)
12.0	84.1% (87.0%)	84.8% (87.7%)	84.5% (87.5%)	84.5% (87.3%)
14.0	84.0% (87.2%)	83.8% (87.2%)	83.6% (87.0%)	83.8% (87.2%)
16.0	84.3% (87.9%)	83.6% (87.3%)	83.4% (87.0%)	83.8% (87.2%)
18.0	83.2% (86.6%)	82.9% (87.0%)	82.7% (86.5%)	82.2% (86.6%)
20.0	82.4% (85.7%)	82.2% (85.9%)	81.5% (85.2%)	81.3% (84.8%)
22.0	82.9% (86.3%)	79.5% (84.1%)	79.0% (83.8%)	79.0% (83.4%)
24.0	81.5% (85.9%)	77.4% (82.7%)	78.4% (83.6%)	76.3% (82.4%)
26.0	77.2% (83.6%)	74.9% (80.6%)	74.7% (80.9%)	74.0% (80.4%)
28.0	74.0% (80.9%)	73.1% (79.3%)	73.6% (80.0%)	74.7% (81.1%)
30.0	72.2% (79.7%)	70.2% (77.5%)	70.9% (78.3%)	70.9% (78.3%)



D.20 Experiment E.4.5

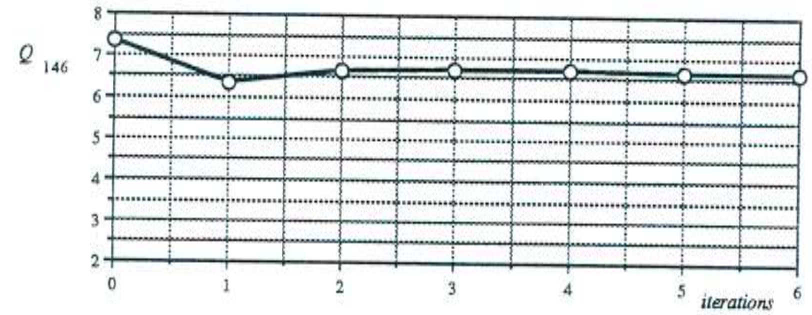
NLDA approach : $Y = A(X) + f(x)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.4
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	5
iterations :	6
target update after :	6
sample selection :	random

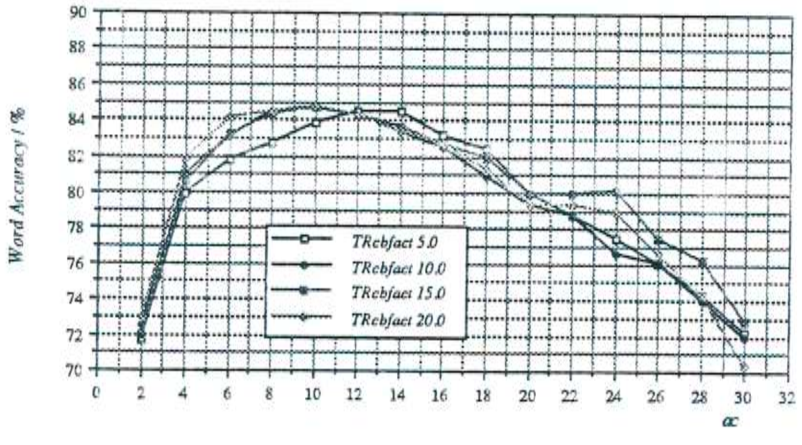
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.365	6.501	6.682	6.681	6.670	6.662
Q_{50}	15.860	-	-	-	-	-	14.555



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	15.0	20.0
2.0	71.7% (72.9%)	72.5% (73.8%)	72.0% (72.7%)	72.9% (73.4%)
4.0	79.9% (81.3%)	80.6% (81.6%)	80.9% (81.8%)	81.8% (82.9%)
6.0	81.8% (83.4%)	83.1% (84.8%)	83.2% (84.7%)	84.1% (85.7%)
8.0	82.7% (84.5%)	84.5% (86.5%)	84.3% (86.5%)	84.5% (86.8%)
10.0	83.8% (85.9%)	84.7% (87.0%)	84.8% (87.2%)	84.8% (87.2%)
12.0	84.5% (87.0%)	84.3% (87.2%)	84.3% (87.0%)	84.3% (87.0%)
14.0	84.5% (87.2%)	83.4% (86.5%)	83.6% (87.0%)	83.4% (86.8%)
16.0	83.2% (86.6%)	82.5% (86.5%)	82.7% (86.6%)	82.7% (86.5%)
18.0	82.4% (86.5%)	80.9% (85.7%)	82.0% (85.9%)	81.3% (85.6%)
20.0	80.0% (84.0%)	79.3% (84.7%)	80.0% (85.2%)	79.3% (84.7%)
22.0	78.8% (83.2%)	78.8% (84.0%)	80.0% (84.7%)	79.3% (84.1%)
24.0	77.5% (81.6%)	76.6% (82.9%)	80.2% (85.2%)	79.0% (84.3%)
26.0	76.1% (80.7%)	76.1% (81.5%)	77.5% (83.1%)	76.6% (82.4%)
28.0	74.3% (78.8%)	74.0% (79.7%)	76.3% (82.4%)	74.3% (80.6%)
30.0	72.2% (77.0%)	72.0% (78.8%)	72.9% (79.1%)	70.4% (77.5%)



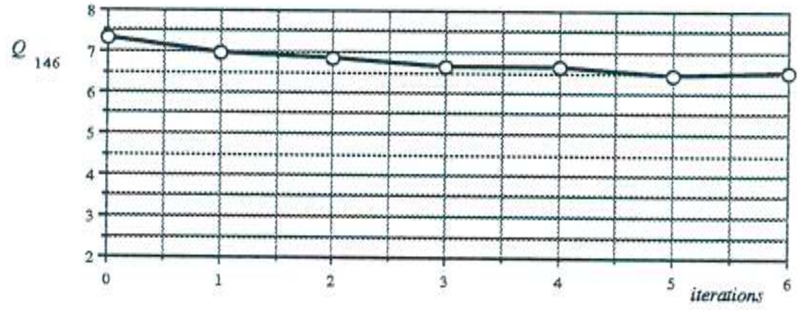
D.21 Experiment E.4.6

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	5
α :	0.9	iterations :	6
β :	1.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

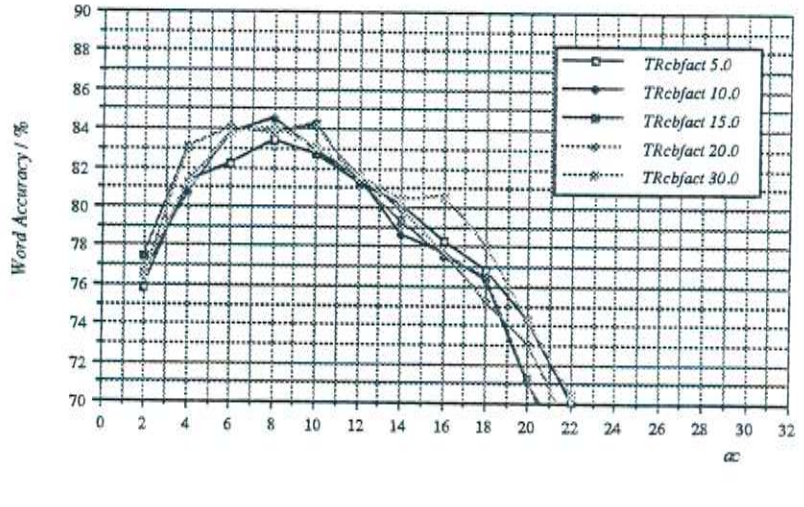
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.967	6.826	6.658	6.641	6.428	6.506
Q_{50}	15.860	-	-	-	-	-	13.727



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	75.8% (76.1%)	76.6% (77.4%)	77.4% (78.3%)	76.3% (77.0%)	76.6% (77.2%)
4.0	81.3% (82.4%)	80.7% (81.8%)	83.1% (84.1%)	83.1% (84.0%)	81.3% (82.2%)
6.0	82.2% (84.0%)	83.8% (85.2%)	84.0% (85.6%)	84.0% (85.6%)	83.8% (85.6%)
8.0	83.4% (85.7%)	84.5% (86.8%)	83.8% (86.3%)	83.8% (86.1%)	84.1% (87.0%)
10.0	82.7% (85.6%)	82.9% (85.7%)	84.3% (87.3%)	84.1% (87.2%)	83.1% (86.1%)
12.0	81.3% (84.5%)	81.5% (84.8%)	81.3% (84.8%)	81.5% (84.8%)	81.6% (85.0%)
14.0	80.2% (83.6%)	78.6% (82.9%)	79.3% (82.7%)	79.9% (82.9%)	80.4% (83.4%)
16.0	78.3% (82.9%)	77.7% (82.7%)	77.5% (82.4%)	77.7% (82.4%)	80.6% (84.0%)
18.0	76.8% (82.7%)	76.3% (81.8%)	76.5% (81.3%)	75.2% (80.6%)	78.1% (82.9%)
20.0	74.3% (80.2%)	71.1% (77.7%)	71.3% (77.5%)	73.1% (79.3%)	74.3% (80.4%)
22.0	70.1% (76.8%)	67.6% (75.2%)	66.1% (74.5%)	68.6% (76.5%)	70.4% (77.4%)
24.0	60.1% (69.9%)	63.3% (72.0%)	62.6% (73.1%)	60.6% (71.1%)	61.7% (70.8%)
26.0	57.0% (67.2%)	58.8% (69.3%)	57.4% (68.8%)	59.0% (69.7%)	61.7% (70.4%)
28.0	57.2% (66.1%)	55.4% (65.5%)	56.9% (66.0%)	58.6% (67.7%)	54.4% (62.4%)



D.22 Experiment E.5.1

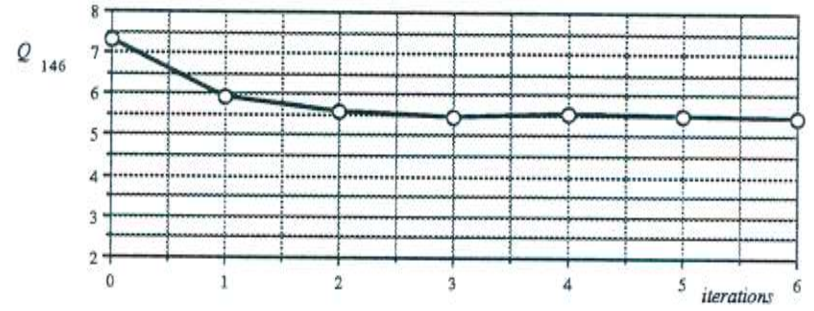
NLDA approach : $Y = A(X) + f(x)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	1.0
β :	0.0
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	10
iterations :	6
target update after :	6
sample selection :	random

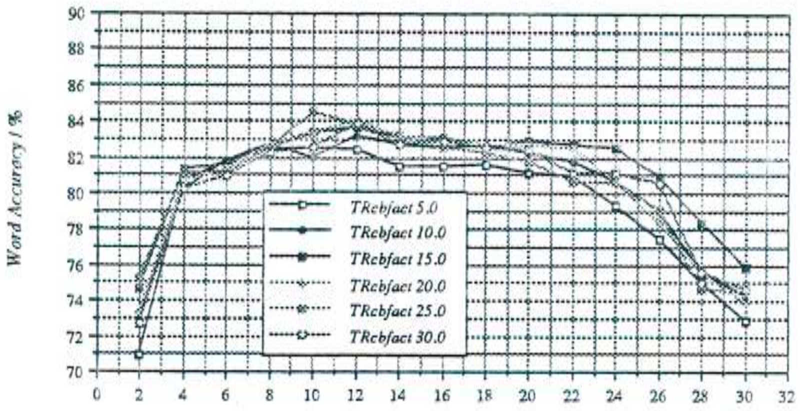
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.931	5.583	5.474	5.544	5.512	5.463
Q_{50}	15.860	-	-	-	-	-	13.577



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	5.0	10.0	15.0	20.0	25.0	30.0
2.0	70.9% (71.7%)	72.9% (74.0%)	74.7% (75.4%)	73.4% (74.2%)	75.2% (75.6%)	72.7% (73.1%)
4.0	80.6% (81.5%)	80.2% (81.1%)	81.3% (82.0%)	80.7% (81.5%)	81.1% (81.8%)	80.2% (80.9%)
6.0	81.8% (82.0%)	81.8% (82.7%)	81.6% (82.5%)	81.3% (82.0%)	81.1% (82.0%)	80.9% (81.8%)
8.0	82.4% (83.2%)	82.7% (83.6%)	82.5% (83.4%)	82.7% (83.8%)	82.5% (83.6%)	82.2% (83.2%)
10.0	82.5% (83.2%)	82.0% (83.6%)	83.4% (85.2%)	83.4% (85.4%)	84.5% (85.7%)	82.5% (85.2%)
12.0	82.4% (84.5%)	83.2% (85.2%)	83.6% (85.6%)	84.0% (86.3%)	83.8% (86.1%)	83.8% (86.1%)
14.0	81.5% (84.7%)	82.7% (85.4%)	83.1% (85.7%)	83.2% (85.6%)	83.2% (85.6%)	82.7% (85.6%)
16.0	81.5% (85.2%)	82.5% (85.7%)	83.1% (85.7%)	82.9% (86.1%)	82.5% (85.7%)	82.9% (86.1%)
18.0	81.6% (85.0%)	82.7% (85.7%)	82.5% (85.7%)	82.7% (85.7%)	82.2% (85.2%)	82.5% (85.6%)
20.0	81.1% (84.8%)	82.2% (85.6%)	82.9% (86.1%)	82.7% (85.7%)	81.8% (84.8%)	82.4% (85.4%)
22.0	80.9% (84.8%)	81.8% (85.6%)	82.7% (86.1%)	82.5% (85.6%)	80.7% (83.8%)	81.1% (84.0%)
24.0	79.3% (83.4%)	80.7% (84.3%)	82.5% (86.3%)	80.7% (84.5%)	80.6% (84.3%)	81.1% (84.5%)
26.0	77.5% (82.5%)	79.1% (83.4%)	80.9% (84.7%)	78.4% (82.9%)	79.1% (82.9%)	80.6% (84.1%)
28.0	75.0% (80.7%)	75.6% (80.9%)	78.4% (83.1%)	75.6% (80.7%)	74.7% (80.7%)	75.6% (81.6%)
30.0	72.9% (78.6%)	74.3% (80.0%)	75.9% (80.6%)	73.8% (80.0%)	74.5% (80.6%)	74.7% (79.7%)



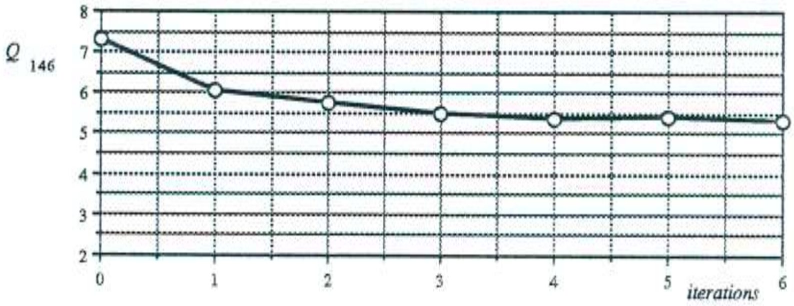
D.23 Experiment E.5.2

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

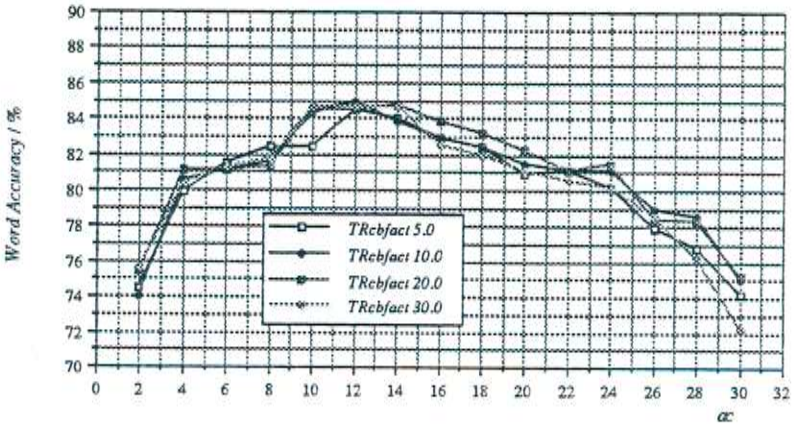
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.0380	5.755	5.479	5.353	5.403	5.332
Q_{50}	15.860	-	-	-	-	-	13.519



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	20.0	30.0
2.0	74.5% (75.2%)	74.0% (75.4%)	75.4% (76.1%)	75.6% (76.3%)
4.0	79.9% (80.7%)	80.6% (81.6%)	81.1% (82.0%)	80.4% (81.3%)
6.0	81.5% (83.2%)	81.1% (82.4%)	81.1% (82.4%)	81.3% (82.2%)
8.0	82.4% (85.2%)	81.6% (83.2%)	81.3% (82.9%)	81.8% (83.1%)
10.0	82.4% (85.2%)	84.3% (86.8%)	84.5% (87.0%)	84.7% (87.0%)
12.0	84.5% (87.3%)	85.0% (87.7%)	84.5% (87.3%)	84.7% (87.5%)
14.0	84.0% (87.2%)	83.8% (87.2%)	84.8% (87.9%)	84.7% (87.7%)
16.0	82.9% (86.1%)	82.9% (85.7%)	83.8% (86.6%)	82.5% (86.1%)
18.0	82.4% (85.7%)	82.4% (86.3%)	83.2% (87.2%)	82.0% (86.5%)
20.0	80.9% (85.9%)	81.5% (85.9%)	82.2% (86.8%)	81.1% (86.1%)
22.0	81.1% (86.1%)	81.1% (85.6%)	81.1% (86.1%)	80.6% (85.6%)
24.0	80.2% (84.8%)	81.1% (85.7%)	81.5% (86.8%)	80.2% (85.6%)
26.0	77.9% (84.1%)	79.0% (85.0%)	78.4% (85.2%)	78.6% (84.7%)
28.0	76.8% (83.1%)	78.6% (84.3%)	78.3% (85.2%)	76.3% (83.6%)
30.0	74.2% (81.1%)	75.0% (81.8%)	75.2% (82.5%)	72.2% (80.4%)



D.24 Experiment E.5.3

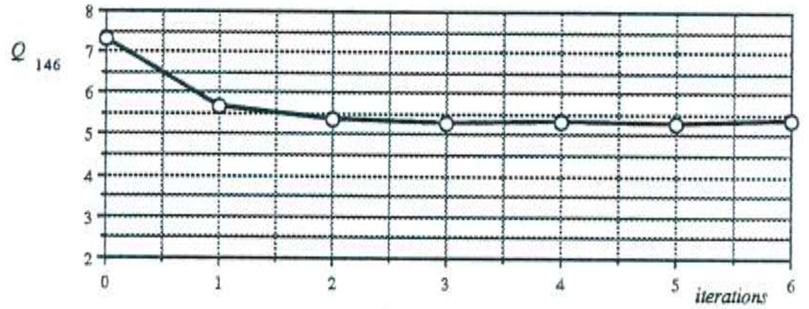
NLDA approach : $Y = A(X) + f(x)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.2
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	10
iterations :	6
target update after :	6
sample selection :	random

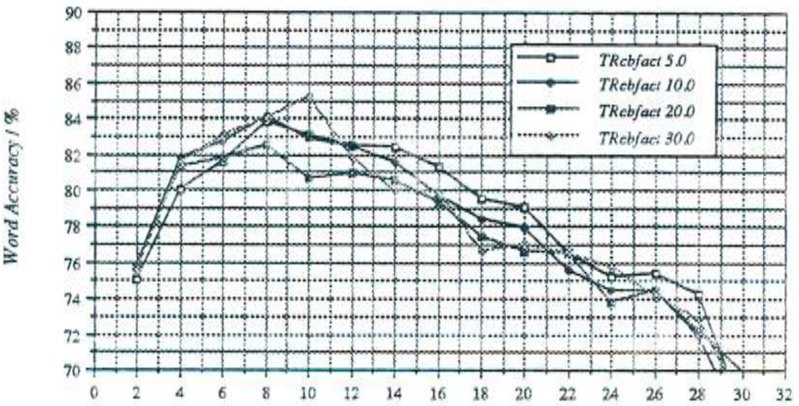
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.683	5.383	5.297	5.303	5.293	5.381
Q_{50}	15.860	-	-	-	-	-	14.483



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	20.0	30.0
2.0	75.0% (76.1%)	75.8% (76.5%)	75.8% (76.3%)	75.6% (76.7%)
4.0	80.0% (80.9%)	81.8% (82.9%)	81.3% (82.0%)	81.5% (82.2%)
6.0	81.6% (82.9%)	82.7% (84.7%)	81.8% (82.7%)	83.1% (84.0%)
8.0	83.8% (85.9%)	84.1% (86.5%)	82.5% (84.7%)	84.0% (85.9%)
10.0	83.1% (86.1%)	82.9% (86.1%)	80.7% (84.1%)	85.2% (85.9%)
12.0	82.5% (85.7%)	82.4% (86.3%)	80.9% (84.7%)	81.8% (85.4%)
14.0	82.4% (85.7%)	81.6% (85.7%)	80.6% (84.3%)	80.0% (84.3%)
16.0	81.3% (84.8%)	79.7% (83.2%)	79.3% (83.1%)	79.9% (84.1%)
18.0	79.5% (83.2%)	78.4% (82.5%)	77.5% (82.4%)	76.6% (81.5%)
20.0	79.1% (83.6%)	77.9% (82.9%)	76.6% (81.6%)	77.0% (82.0%)
22.0	76.5% (82.0%)	75.6% (80.6%)	76.5% (81.5%)	76.5% (80.6%)
24.0	75.2% (80.6%)	74.5% (79.9%)	73.8% (80.0%)	75.8% (79.9%)
26.0	75.4% (80.9%)	74.5% (80.2%)	74.5% (80.4%)	74.2% (80.0%)
28.0	74.3% (78.8%)	72.0% (78.4%)	72.2% (78.8%)	72.9% (79.9%)
30.0	67.7% (73.6%)	66.7% (74.7%)	69.9% (76.6%)	68.3% (76.6%)



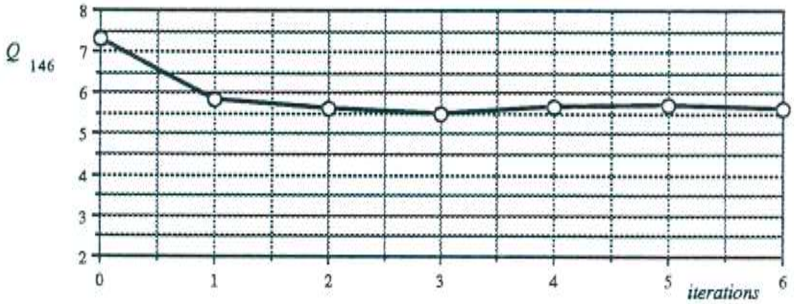
D.25 Experiment E.5.4

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

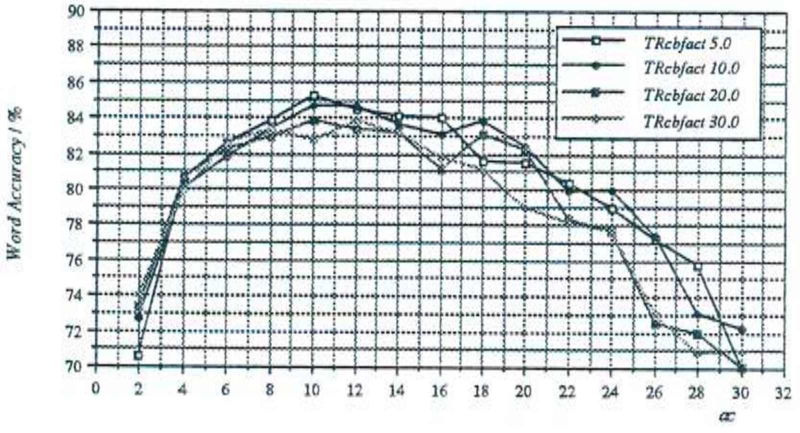
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.856	5.633	5.516	5.655	5.699	5.629
Q_{50}	15.860	-	-	-	-	-	13.066



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	20.0	30.0
2.0	70.6% (72.0%)	72.7% (73.6%)	73.4% (74.2%)	74.0% (74.7%)
4.0	80.6% (81.5%)	80.0% (80.9%)	80.7% (81.8%)	79.9% (80.7%)
6.0	82.5% (83.6%)	81.8% (82.9%)	82.2% (83.2%)	82.5% (83.6%)
8.0	83.8% (85.0%)	83.4% (85.2%)	82.9% (84.7%)	83.4% (85.0%)
10.0	85.2% (87.0%)	84.7% (86.5%)	83.8% (85.9%)	82.7% (84.8%)
12.0	84.5% (86.8%)	84.7% (86.8%)	83.4% (85.7%)	83.8% (85.9%)
14.0	84.1% (86.5%)	83.6% (86.3%)	83.2% (86.3%)	83.2% (86.1%)
16.0	84.0% (86.8%)	83.1% (85.7%)	81.1% (84.7%)	81.8% (85.6%)
18.0	81.6% (85.4%)	83.8% (85.6%)	83.1% (85.7%)	81.1% (84.8%)
20.0	81.5% (85.2%)	82.4% (85.9%)	82.2% (85.6%)	79.0% (83.8%)
22.0	80.4% (84.5%)	79.9% (84.7%)	78.3% (83.1%)	78.1% (83.1%)
24.0	79.0% (83.6%)	80.0% (84.1%)	77.7% (82.0%)	77.7% (82.5%)
26.0	77.2% (81.8%)	77.4% (82.0%)	72.5% (78.8%)	73.1% (79.9%)
28.0	75.8% (81.3%)	73.1% (78.6%)	72.0% (77.2%)	70.8% (77.2%)
30.0	70.1% (76.8%)	72.2% (77.4%)	70.1% (75.8%)	70.9% (76.3%)



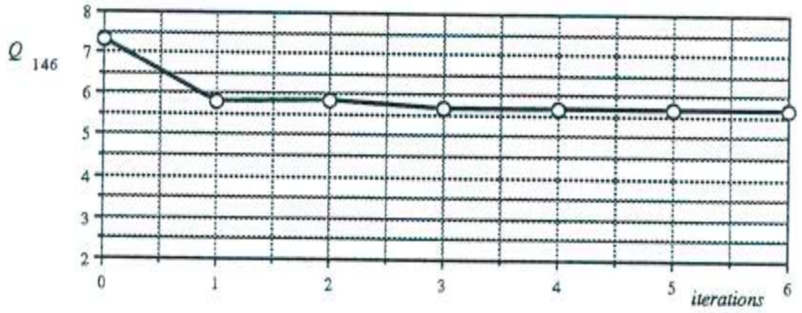
D.26 Experiment E.5.5

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	0.9	iterations :	6
β :	0.4	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	80 (one per phonem class)		

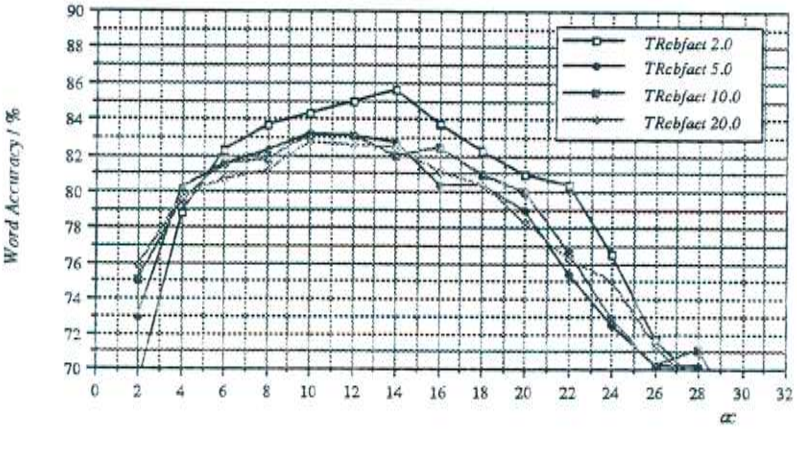
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.810	5.824	5.685	5.677	5.561	5.648
Q_{80}	15.860	-	-	-	-	-	13.239



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	2.0	5.0	10.0	20.0
2.0	68.8% (69.7%)	72.9% (74.5%)	75.0% (75.9%)	75.8% (76.5%)
4.0	78.8% (80.4%)	80.2% (81.3%)	79.5% (80.4%)	79.7% (80.6%)
6.0	82.2% (83.4%)	81.5% (82.7%)	81.5% (83.1%)	80.7% (82.0%)
8.0	83.6% (85.4%)	82.2% (84.3%)	81.8% (83.6%)	81.1% (82.9%)
10.0	84.3% (86.5%)	83.2% (85.6%)	83.1% (85.4%)	82.7% (85.2%)
12.0	85.0% (87.5%)	83.1% (85.7%)	83.1% (85.6%)	82.5% (85.9%)
14.0	85.6% (88.2%)	82.7% (86.1%)	82.0% (86.1%)	82.4% (86.5%)
16.0	83.7% (86.8%)	80.4% (84.7%)	82.4% (86.5%)	81.1% (85.7%)
18.0	82.2% (85.4%)	80.4% (84.1%)	80.9% (85.2%)	80.4% (84.8%)
20.0	80.9% (84.8%)	79.0% (83.6%)	80.0% (84.1%)	78.3% (84.1%)
22.0	80.4% (84.0%)	75.4% (81.3%)	76.6% (82.0%)	76.3% (81.5%)
24.0	76.5% (82.9%)	72.5% (78.8%)	72.9% (78.8%)	75.0% (80.6%)
26.0	71.7% (78.3%)	70.2% (77.0%)	70.2% (76.6%)	71.3% (77.4%)
28.0	69.2% (76.3%)	70.2% (77.0%)	71.1% (77.4%)	68.6% (75.8%)
30.0	67.0% (74.7%)	65.2% (73.3%)	67.0% (73.4%)	66.7% (73.8%)



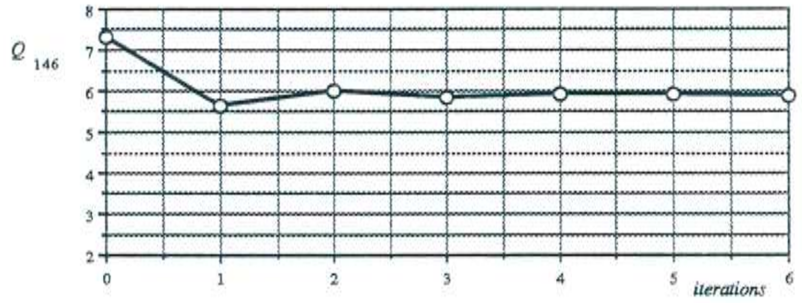
D.27 Experiment E.5.6

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	0.9	iterations :	6
β :	1.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

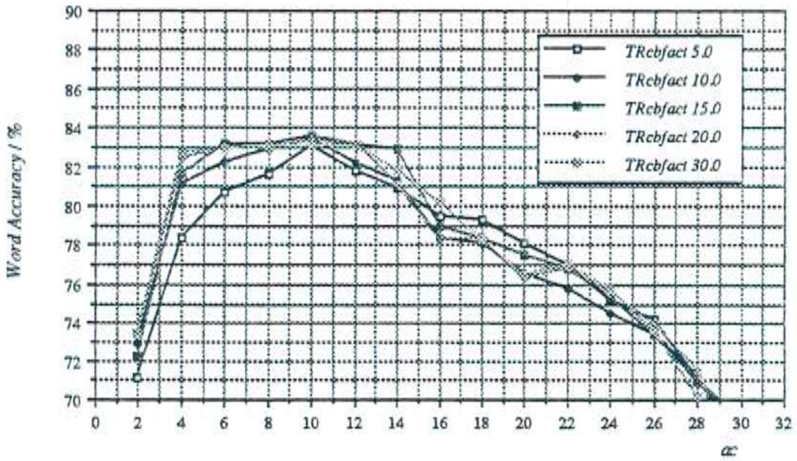
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.638	6.002	5.858	5.917	5.940	5.878
Q_{50}	15.860	-	-	-	-	-	13.075



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	71.1% (72.0%)	72.9% (73.4%)	72.2% (72.7%)	73.3% (73.6%)	73.4% (73.8%)
4.0	78.4% (79.5%)	81.1% (82.2%)	81.6% (82.7%)	82.4% (83.4%)	82.7% (83.6%)
6.0	80.7% (81.8%)	82.2% (83.4%)	83.2% (84.3%)	83.1% (84.0%)	82.9% (83.8%)
8.0	81.6% (82.7%)	82.9% (84.3%)	83.2% (84.5%)	82.9% (84.3%)	83.2% (84.5%)
10.0	83.1% (84.5%)	83.4% (84.8%)	83.6% (85.2%)	83.4% (85.4%)	83.1% (85.2%)
12.0	81.8% (84.7%)	82.2% (84.7%)	83.2% (85.4%)	83.2% (85.6%)	83.2% (85.7%)
14.0	80.9% (84.3%)	81.3% (84.7%)	82.9% (86.1%)	81.1% (84.3%)	81.8% (84.7%)
16.0	79.5% (83.4%)	78.4% (83.4%)	79.0% (83.4%)	78.3% (83.1%)	80.2% (85.2%)
18.0	79.3% (82.9%)	78.1% (82.9%)	78.4% (83.6%)	78.4% (83.2%)	78.3% (83.6%)
20.0	78.1% (82.0%)	76.8% (81.5%)	77.5% (82.2%)	76.3% (81.5%)	76.5% (81.8%)
22.0	77.0% (80.6%)	75.8% (80.4%)	76.8% (81.8%)	76.8% (81.8%)	77.0% (81.6%)
24.0	75.2% (79.7%)	74.5% (79.9%)	75.4% (80.9%)	75.6% (80.6%)	75.8% (81.1%)
26.0	74.2% (79.0%)	73.4% (79.0%)	73.4% (79.0%)	74.0% (79.5%)	73.6% (78.8%)
28.0	71.1% (77.4%)	71.3% (77.4%)	70.9% (77.7%)	71.5% (78.1%)	70.2% (76.6%)
30.0	68.8% (75.2%)	67.7% (74.9%)	68.8% (75.0%)	65.4% (72.9%)	64.9% (72.4%)



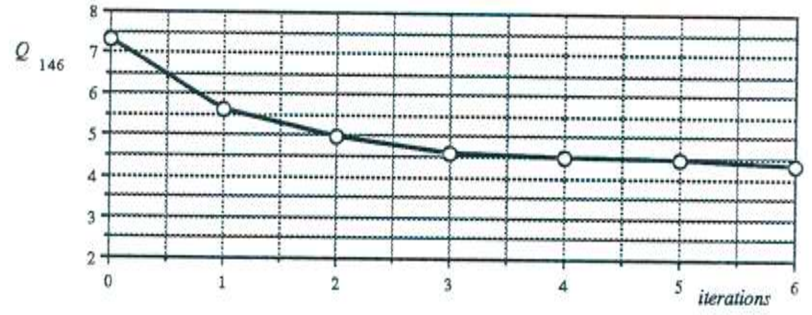
D.28 Experiment E.6.1

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

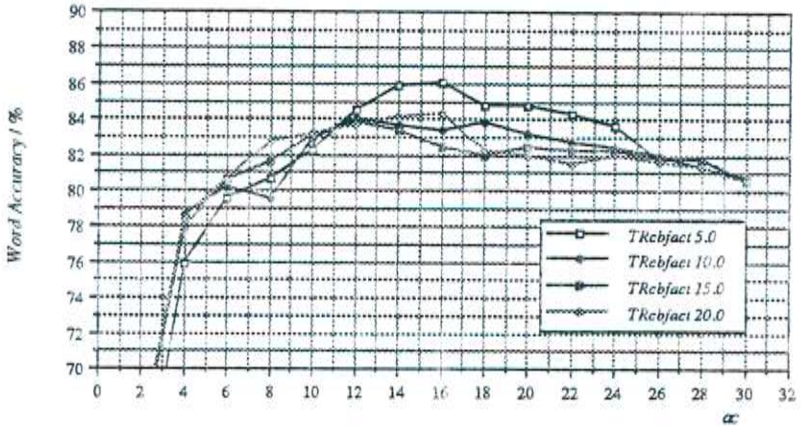
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.643	4.977	4.595	4.512	4.448	4.343
Q_{50}	15.860	-	-	-	-	-	12.614



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	15.0	20.0
2.0	60.2% (60.2%)	64.5% (65.4%)	65.2% (66.3%)	65.4% (66.0%)
4.0	75.9% (76.6%)	78.6% (79.7%)	77.9% (78.8%)	78.4% (79.5%)
6.0	79.5% (80.6%)	80.2% (80.9%)	80.6% (81.5%)	80.7% (81.6%)
8.0	80.7% (82.7%)	79.5% (80.9%)	81.6% (83.4%)	82.7% (83.6%)
10.0	82.2% (84.1%)	82.9% (84.3%)	83.1% (84.8%)	83.2% (84.8%)
12.0	84.5% (86.3%)	84.0% (85.4%)	83.8% (85.6%)	83.6% (85.7%)
14.0	85.9% (87.9%)	83.6% (85.4%)	83.4% (85.9%)	84.1% (86.5%)
16.0	86.1% (88.1%)	83.4% (85.9%)	82.4% (85.6%)	84.3% (86.6%)
18.0	84.8% (87.3%)	83.8% (86.5%)	82.0% (85.4%)	82.2% (85.6%)
20.0	84.8% (87.5%)	83.2% (86.3%)	82.4% (85.7%)	82.0% (85.6%)
22.0	84.3% (87.5%)	82.7% (86.6%)	82.2% (85.9%)	81.5% (85.6%)
24.0	83.6% (82.2%)	82.4% (86.1%)	82.2% (86.1%)	82.0% (86.1%)
26.0	81.8% (86.8%)	82.0% (85.9%)	81.8% (85.9%)	81.6% (86.1%)
28.0	81.8% (86.5%)	81.3% (85.6%)	81.8% (85.9%)	81.3% (85.6%)
30.0	80.6% (85.4%)	80.7% (85.0%)	80.7% (84.8%)	80.6% (84.8%)



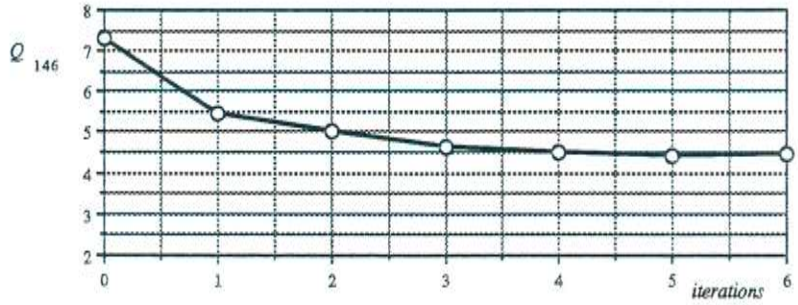
D.29 Experiment E.6.2

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

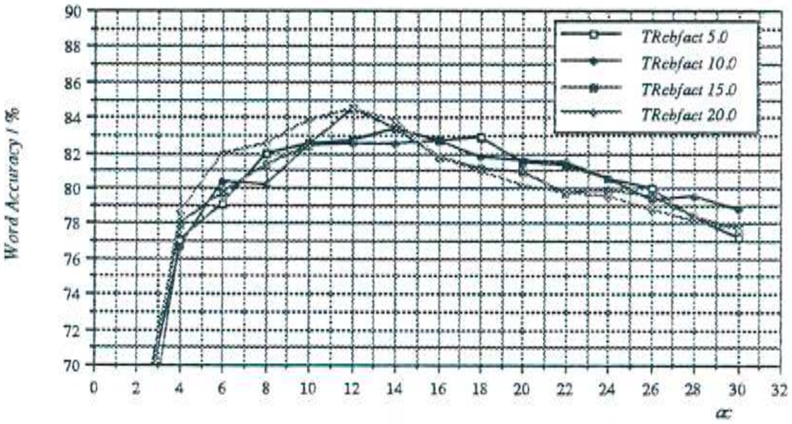
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.462	5.036	4.634	4.506	4.428	4.458
Q_{50}	15.860	-	-	-	-	-	12.570



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	61.0% (62.4%)	64.9% (65.6%)	65.2% (66.3%)	63.8% (64.9%)	67.0% (68.1%)
4.0	77.0% (78.3%)	76.6% (78.1%)	77.9% (79.3%)	78.6% (79.9%)	78.8% (79.7%)
6.0	79.1% (80.7%)	80.4% (81.8%)	79.7% (81.1%)	82.0% (83.1%)	81.5% (82.5%)
8.0	82.0% (84.0%)	80.2% (82.4%)	81.3% (83.1%)	82.5% (84.0%)	83.2% (84.8%)
10.0	82.5% (84.7%)	82.4% (84.1%)	82.4% (84.1%)	83.8% (85.7%)	83.8% (85.7%)
12.0	82.7% (84.7%)	82.5% (84.3%)	84.5% (86.5%)	84.5% (86.5%)	84.5% (86.5%)
14.0	83.4% (85.7%)	82.5% (85.0%)	83.4% (85.9%)	83.8% (86.1%)	84.1% (86.5%)
16.0	82.7% (85.6%)	82.7% (85.4%)	81.8% (84.8%)	81.8% (85.9%)	83.1% (85.7%)
18.0	82.9% (86.1%)	81.8% (85.4%)	81.1% (84.3%)	80.9% (84.8%)	81.5% (84.8%)
20.0	81.5% (85.4%)	81.6% (85.2%)	80.9% (84.5%)	80.2% (84.5%)	80.9% (84.8%)
22.0	81.3% (85.4%)	81.5% (85.2%)	79.7% (84.5%)	79.7% (84.2%)	80.9% (85.0%)
24.0	80.6% (84.5%)	80.6% (85.2%)	79.9% (84.7%)	79.5% (84.1%)	80.6% (85.7%)
26.0	80.0% (84.3%)	79.3% (83.4%)	79.5% (83.8%)	78.8% (83.2%)	78.8% (83.6%)
28.0	78.3% (83.8%)	79.5% (83.8%)	78.4% (83.1%)	78.1% (82.9%)	78.4% (83.2%)
30.0	77.2% (82.2%)	78.8% (83.1%)	77.7% (82.4%)	77.7% (82.5%)	77.7% (82.9%)



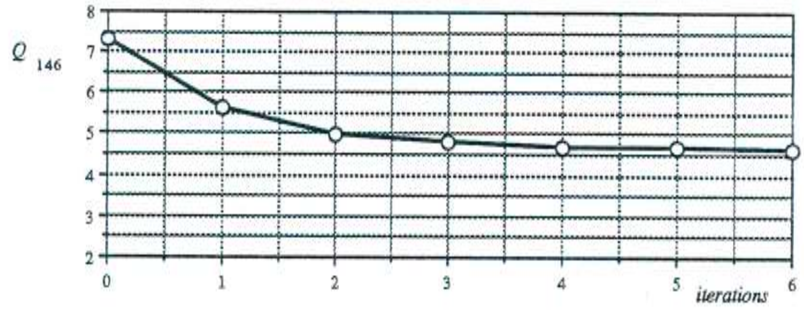
D.30 Experiment E.6.3

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

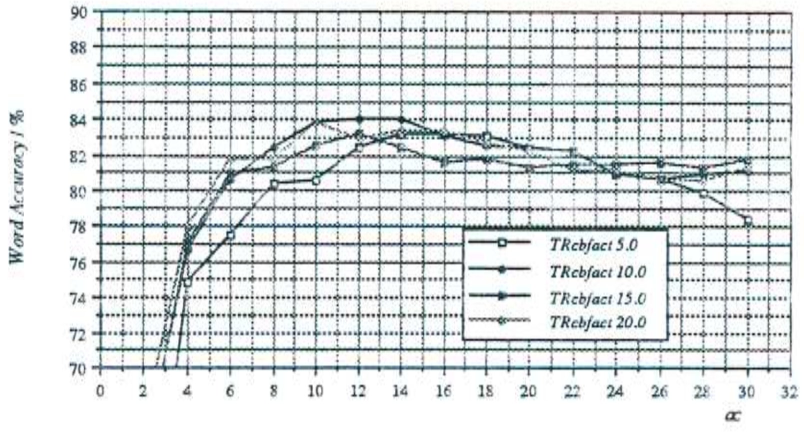
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.613	4.992	4.797	4.670	4.673	4.632
Q_{50}	15.860	-	-	-	-	-	12.059



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	54.9% (56.0%)	64.2% (65.2%)	64.3% (65.6%)	66.3% (67.6%)	62.9% (64.0%)
4.0	74.9% (76.3%)	76.6% (77.7%)	77.0% (78.3%)	77.9% (79.1%)	76.5% (77.4%)
6.0	77.5% (79.0%)	80.6% (81.5%)	80.9% (81.6%)	81.8% (82.5%)	81.3% (82.2%)
8.0	80.4% (81.8%)	82.4% (83.2%)	81.3% (82.0%)	81.8% (83.1%)	82.2% (83.6%)
10.0	80.6% (82.4%)	83.8% (85.0%)	82.5% (83.8%)	83.8% (85.2%)	81.5% (83.2%)
12.0	82.4% (84.3%)	84.0% (85.7%)	83.2% (85.0%)	82.9% (85.0%)	83.6% (85.4%)
14.0	83.2% (85.2%)	84.0% (85.9%)	82.4% (85.4%)	83.4% (85.7%)	83.1% (85.2%)
16.0	83.2% (85.2%)	83.1% (85.4%)	81.6% (84.0%)	83.4% (85.7%)	83.1% (85.4%)
18.0	83.1% (85.4%)	82.5% (85.4%)	81.8% (84.3%)	82.7% (85.6%)	83.4% (86.5%)
20.0	82.4% (85.0%)	82.4% (85.4%)	81.3% (84.5%)	82.2% (85.9%)	81.8% (85.4%)
22.0	82.2% (85.2%)	82.2% (85.4%)	81.5% (84.8%)	81.1% (84.8%)	80.6% (84.5%)
24.0	80.9% (84.7%)	80.9% (84.7%)	81.5% (85.2%)	81.1% (84.8%)	80.6% (84.5%)
26.0	80.7% (84.5%)	80.7% (84.5%)	81.6% (85.4%)	80.7% (84.7%)	80.7% (84.7%)
28.0	79.9% (83.4%)	80.9% (84.7%)	81.3% (85.2%)	80.6% (84.5%)	80.6% (84.1%)
30.0	78.4% (82.7%)	81.1% (84.7%)	81.8% (84.8%)	81.3% (84.7%)	79.3% (82.7%)



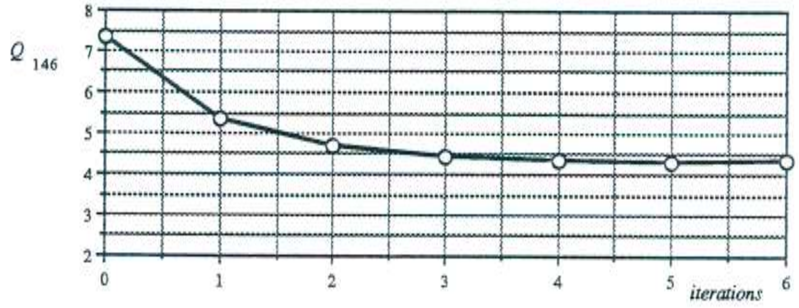
D.31 Experiment E.6.4

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

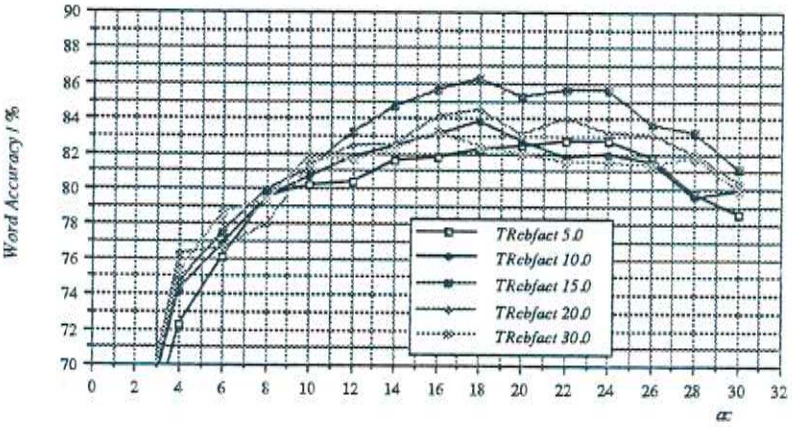
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.353	4.679	4.416	4.339	4.299	4.348
Q_{50}	15.860	-	-	-	-	-	11.679



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	61.9% (62.6%)	64.7% (66.0%)	65.2% (66.5%)	63.6% (64.3%)	64.9% (66.1%)
4.0	72.2% (73.4%)	74.2% (75.4%)	74.7% (76.1%)	76.3% (77.7%)	75.2% (76.8%)
6.0	76.1% (77.7%)	76.8% (78.4%)	77.5% (79.1%)	76.6% (78.4%)	78.6% (80.2%)
8.0	79.5% (81.3%)	79.5% (81.1%)	79.9% (81.6%)	77.9% (79.7%)	79.3% (80.7%)
10.0	80.2% (81.8%)	80.7% (82.4%)	81.1% (83.1%)	81.3% (82.9%)	81.8% (83.2%)
12.0	80.4% (82.4%)	81.8% (83.1%)	83.2% (84.8%)	82.4% (83.6%)	81.6% (83.2%)
14.0	81.6% (84.0%)	82.4% (84.0%)	84.7% (86.3%)	82.5% (84.0%)	82.2% (84.1%)
16.0	81.8% (84.1%)	83.1% (85.0%)	85.6% (87.7%)	84.1% (85.9%)	83.2% (85.2%)
18.0	82.2% (84.5%)	83.8% (85.7%)	86.3% (88.4%)	84.5% (86.8%)	82.4% (85.0%)
20.0	82.4% (85.2%)	82.7% (85.2%)	85.2% (87.7%)	83.1% (86.3%)	82.0% (85.2%)
22.0	82.7% (85.7%)	81.8% (84.7%)	85.6% (88.4%)	84.0% (86.8%)	81.5% (85.6%)
24.0	82.7% (86.1%)	82.0% (84.8%)	85.6% (88.4%)	83.2% (86.8%)	81.5% (85.6%)
26.0	81.8% (85.6%)	81.5% (84.1%)	83.6% (86.6%)	83.1% (86.3%)	81.3% (85.4%)
28.0	79.7% (84.0%)	79.5% (83.2%)	83.2% (86.5%)	82.0% (85.2%)	81.8% (85.9%)
30.0	78.6% (83.1%)	79.9% (83.1%)	81.1% (84.0%)	80.4% (83.8%)	79.9% (84.7%)



D.32 Experiment E.6.5

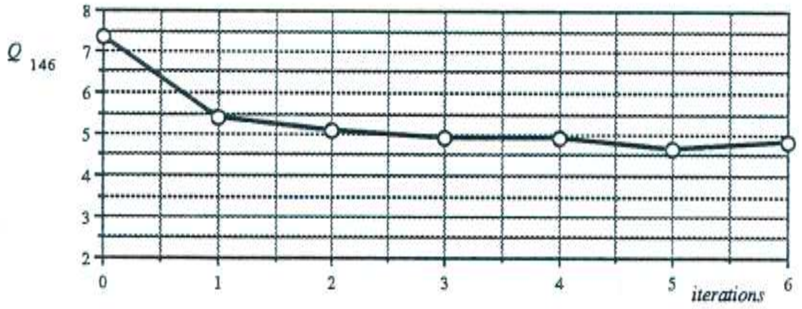
NLDA approach : $Y = A(X) + f(x)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.4
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	20
iterations :	6
target update after :	6
sample selection :	random

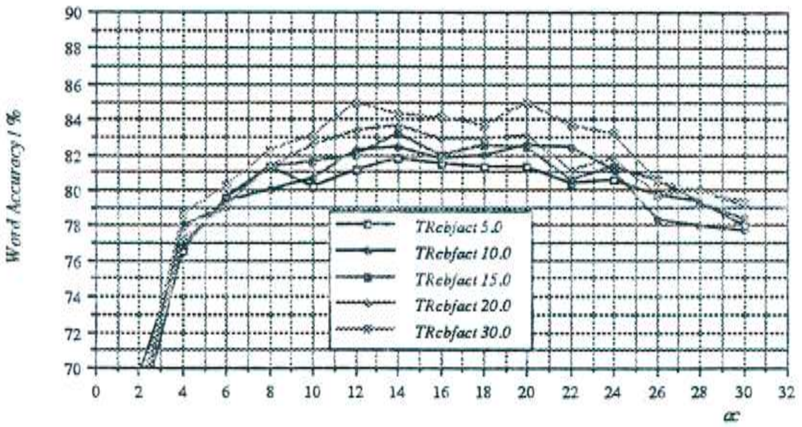
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.385	5.075	4.912	4.921	4.635	4.805
Q_{50}	15.860	-	-	-	-	-	12.184



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	66.3% (67.6%)	69.3% (70.4%)	67.0% (67.9%)	67.9% (68.8%)	67.9% (68.3%)
4.0	76.5% (77.7%)	76.8% (78.4%)	77.9% (79.3%)	77.2% (78.3%)	78.6% (79.5%)
6.0	79.5% (80.4%)	79.3% (80.7%)	79.1% (80.4%)	79.0% (80.2%)	80.2% (80.5%)
8.0	81.3% (82.5%)	80.0% (81.6%)	81.3% (82.7%)	81.3% (82.9%)	82.2% (83.2%)
10.0	80.2% (82.4%)	80.7% (82.9%)	81.6% (84.1%)	82.5% (85.0%)	83.1% (84.7%)
12.0	81.1% (83.8%)	82.2% (85.0%)	82.0% (84.7%)	83.4% (86.3%)	85.0% (87.2%)
14.0	81.8% (84.8%)	82.4% (85.6%)	83.2% (86.5%)	83.6% (86.6%)	84.3% (86.8%)
16.0	81.5% (84.5%)	81.8% (85.2%)	82.0% (85.4%)	82.9% (86.3%)	84.1% (87.0%)
18.0	81.3% (84.3%)	82.0% (85.4%)	82.5% (86.1%)	82.9% (85.9%)	83.6% (86.6%)
20.0	81.3% (84.3%)	82.5% (85.7%)	82.4% (86.1%)	83.1% (86.6%)	85.0% (88.2%)
22.0	80.4% (83.2%)	82.4% (84.8%)	80.6% (84.3%)	81.1% (84.8%)	83.6% (87.0%)
24.0	80.6% (83.1%)	81.1% (84.3%)	81.3% (84.8%)	81.8% (85.2%)	83.2% (86.6%)
26.0	79.7% (82.5%)	78.3% (82.9%)	80.6% (84.3%)	79.7% (84.1%)	80.5% (85.0%)
28.0	79.3% (82.4%)	77.9% (82.0%)	79.3% (83.2%)	80.0% (84.0%)	79.9% (83.6%)
30.0	77.9% (82.0%)	77.7% (82.0%)	78.3% (82.5%)	79.1% (83.2%)	79.3% (83.4%)



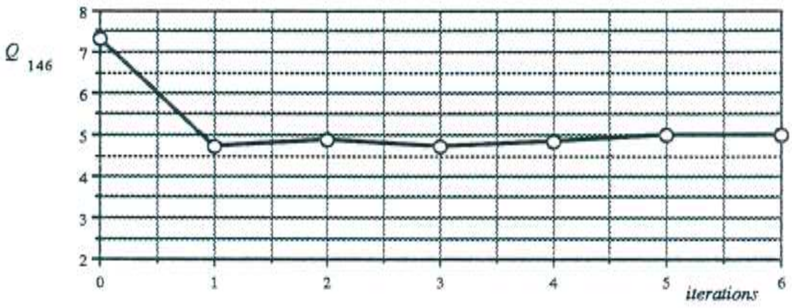
D.33 Experiment E.6.6

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	1.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

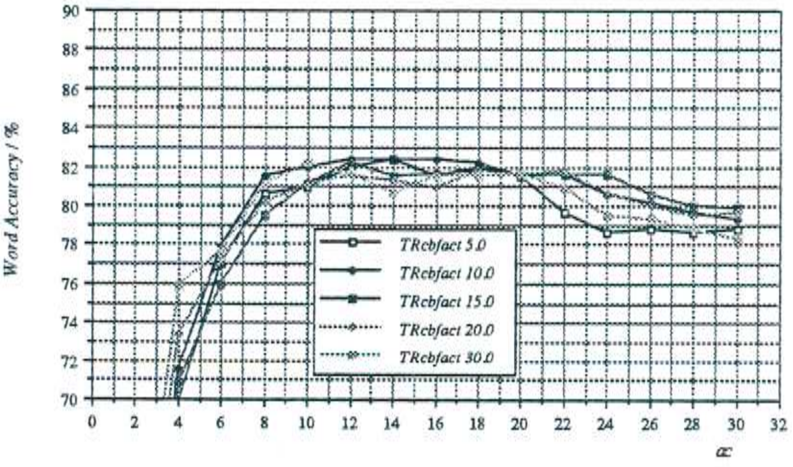
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	4.726	4.898	4.727	4.843	5.000	4.985
Q_{50}	15.860	-	-	-	-	-	11.691



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	46.2% (46.9%)	50.4% (51.3%)	54.7% (55.6%)	57.4% (58.3%)	56.1% (57.0%)
4.0	69.7% (70.6%)	71.5% (72.9%)	70.8% (71.8%)	73.3% (74.5%)	75.8% (76.5%)
6.0	77.0% (77.9%)	77.9% (79.1%)	75.9% (77.5%)	77.2% (78.3%)	77.7% (78.8%)
8.0	80.6% (81.3%)	81.5% (82.5%)	79.5% (80.4%)	80.2% (81.1%)	81.1% (82.0%)
10.0	80.9% (81.6%)	82.0% (82.9%)	81.1% (82.0%)	81.1% (82.4%)	82.2% (83.1%)
12.0	82.0% (82.7%)	82.4% (83.4%)	82.2% (83.2%)	81.5% (82.9%)	81.6% (82.5%)
14.0	82.4% (83.2%)	82.4% (83.2%)	81.5% (83.1%)	81.3% (83.1%)	80.7% (82.4%)
16.0	81.5% (83.2%)	82.4% (83.8%)	81.6% (83.4%)	80.9% (82.9%)	81.5% (83.2%)
18.0	82.0% (84.0%)	82.2% (83.8%)	81.8% (83.8%)	81.8% (83.6%)	81.5% (83.4%)
20.0	81.5% (83.8%)	81.5% (83.6%)	81.6% (83.6%)	81.6% (83.8%)	81.6% (83.6%)
22.0	79.7% (83.4%)	81.6% (84.0%)	81.6% (84.3%)	81.8% (84.1%)	80.9% (83.1%)
24.0	78.6% (82.4%)	80.6% (83.2%)	81.6% (84.5%)	80.7% (83.2%)	79.5% (82.5%)
26.0	78.8% (82.5%)	80.2% (83.6%)	80.6% (83.8%)	80.0% (83.1%)	79.3% (82.2%)
28.0	78.6% (82.2%)	79.7% (83.1%)	80.0% (83.2%)	79.5% (82.9%)	78.8% (82.2%)
30.0	78.8% (82.9%)	79.3% (83.4%)	79.9% (83.8%)	79.7% (83.4%)	78.3% (82.4%)



D.34 Experiment E.7.1

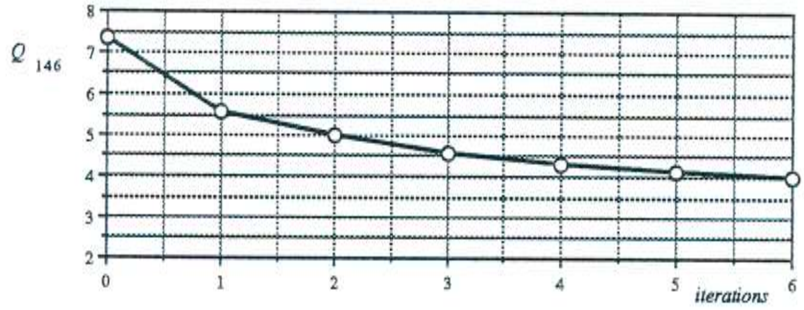
NLDA approach : $Y = A(X) + f(x)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	1.0
β :	0.0
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	30
iterations :	6
target update after :	6
sample selection :	random

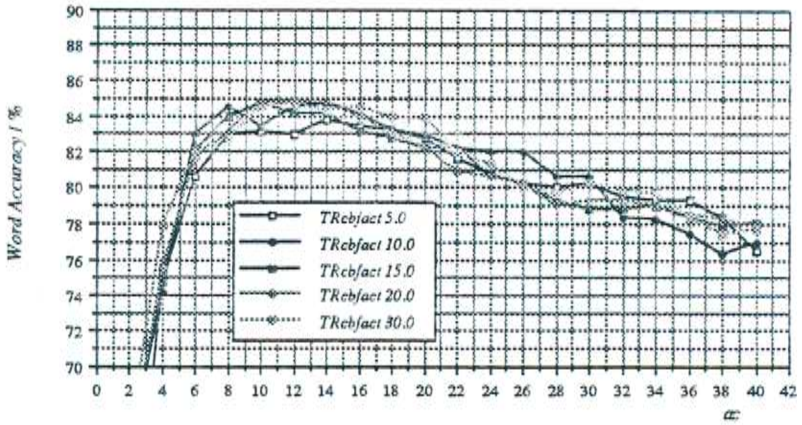
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.551	5.008	4.567	4.295	4.139	4.000
Q_{50}	15.860	-	-	-	-	-	12.518



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	55.1% (55.4%)	65.2% (66.1%)	64.3% (65.4%)	64.7% (65.8%)	66.0% (66.8%)
4.0	74.7% (76.1%)	74.2% (75.4%)	75.2% (76.5%)	75.6% (76.8%)	77.9% (79.0%)
6.0	80.6% (81.8%)	82.9% (84.1%)	82.0% (83.2%)	81.6% (82.7%)	82.0% (82.7%)
8.0	82.9% (83.8%)	84.5% (85.7%)	84.0% (84.7%)	83.1% (83.8%)	82.7% (83.4%)
10.0	83.1% (84.5%)	83.4% (85.2%)	84.7% (85.9%)	84.8% (86.1%)	84.5% (85.7%)
12.0	82.9% (85.6%)	84.5% (86.5%)	84.1% (86.5%)	84.8% (87.0%)	84.7% (86.5%)
14.0	83.8% (86.6%)	84.7% (87.0%)	84.1% (86.6%)	84.7% (87.0%)	84.3% (86.6%)
16.0	83.4% (86.5%)	84.1% (87.0%)	83.1% (86.5%)	84.0% (86.8%)	84.5% (87.2%)
18.0	83.2% (86.3%)	83.2% (86.6%)	82.7% (86.1%)	83.1% (86.3%)	84.0% (87.2%)
20.0	82.7% (85.7%)	82.9% (86.6%)	82.2% (85.4%)	82.2% (85.6%)	84.0% (87.2%)
22.0	81.6% (84.7%)	82.2% (86.1%)	82.2% (86.1%)	80.9% (84.7%)	82.0% (86.3%)
24.0	80.7% (84.3%)	82.0% (86.3%)	80.7% (84.7%)	80.6% (84.3%)	81.3% (85.6%)
26.0	80.2% (83.6%)	82.0% (86.5%)	80.2% (84.3%)	80.2% (84.3%)	80.0% (84.0%)
28.0	80.0% (83.4%)	80.6% (85.0%)	79.3% (83.4%)	79.0% (83.2%)	79.7% (84.1%)
30.0	80.2% (83.0%)	80.6% (85.0%)	78.8% (83.2%)	79.3% (83.8%)	80.2% (84.7%)
32.0	79.5% (83.2%)	78.4% (83.4%)	78.8% (83.2%)	79.3% (84.1%)	79.9% (84.3%)
34.0	79.3% (83.2%)	78.3% (83.4%)	79.0% (83.2%)	79.0% (83.8%)	79.7% (84.7%)
36.0	79.3% (83.2%)	77.4% (82.7%)	78.4% (82.9%)	78.6% (83.4%)	78.4% (83.4%)
38.0	78.4% (83.1%)	76.3% (82.4%)	77.9% (82.7%)	78.3% (83.2%)	77.4% (82.7%)
40.0	76.5% (81.3%)	77.0% (82.5%)	78.1% (82.4%)	77.9% (82.9%)	77.7% (83.1%)



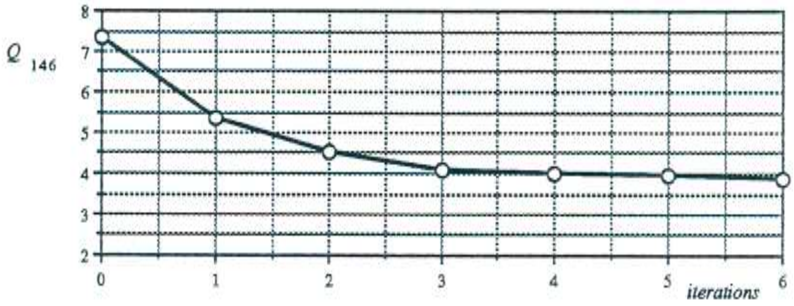
D.35 Experiment E.7.2

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	30
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

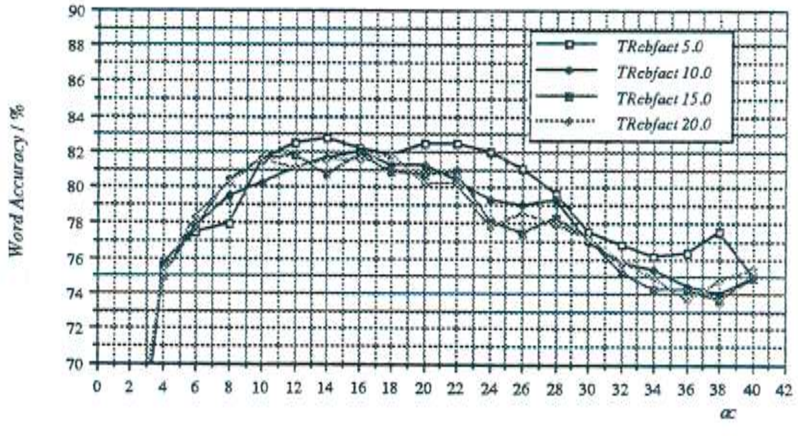
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.361	4.516	4.075	3.984	3.935	3.862
Q_{50}	15.860	-	-	-	-	-	11.912



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	15.0	20.0
2.0	56.7% (57.2%)	60.1% (60.8%)	61.9% (62.4%)	62.2% (62.9%)
4.0	75.6% (76.3%)	75.6% (77.0%)	74.9% (76.3%)	75.2% (76.5%)
6.0	77.4% (78.4%)	77.7% (79.5%)	77.7% (79.3%)	78.3% (79.9%)
8.0	77.9% (79.0%)	79.5% (81.3%)	80.4% (82.0%)	80.2% (81.8%)
10.0	81.5% (82.4%)	80.2% (82.0%)	81.5% (83.1%)	81.5% (83.1%)
12.0	82.5% (83.2%)	81.1% (82.9%)	81.8% (83.4%)	81.1% (83.4%)
14.0	82.7% (83.8%)	81.6% (84.1%)	80.7% (83.6%)	81.5% (83.8%)
16.0	82.2% (84.3%)	82.0% (85.0%)	81.8% (83.8%)	81.5% (84.3%)
18.0	81.8% (84.3%)	81.3% (84.5%)	80.9% (84.0%)	81.8% (84.7%)
20.0	82.5% (85.6%)	81.3% (84.7%)	80.7% (84.5%)	80.2% (84.1%)
22.0	82.5% (85.9%)	80.4% (84.3%)	80.9% (84.1%)	80.2% (84.3%)
24.0	82.0% (85.6%)	79.3% (83.6%)	78.1% (83.2%)	77.7% (83.1%)
26.0	81.1% (85.0%)	79.0% (83.4%)	77.4% (82.9%)	78.6% (84.1%)
28.0	79.7% (83.8%)	79.3% (84.5%)	78.3% (84.0%)	77.9% (83.6%)
30.0	77.5% (82.7%)	77.0% (83.2%)	77.2% (83.4%)	77.2% (83.1%)
32.0	76.8% (82.5%)	75.8% (81.8%)	75.2% (82.0%)	75.9% (82.4%)
34.0	76.1% (82.2%)	75.4% (81.6%)	74.3% (81.1%)	74.9% (81.6%)
36.0	76.3% (82.5%)	74.5% (80.7%)	74.3% (80.9%)	73.6% (81.3%)
38.0	77.5% (84.0%)	74.0% (80.4%)	73.6% (80.7%)	74.7% (81.6%)
40.0	75.0% (82.0%)	74.9% (80.7%)	74.9% (81.5%)	75.4% (80.9%)



D.36 Experiment E.7.3

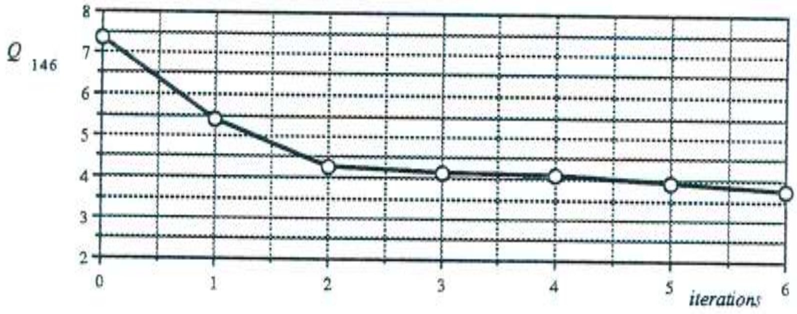
NLDA approach : $Y = A(X) + f(x)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.2
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	146
learning rate :	0.008
momentum :	0.9
number of hidden units :	30
iterations :	6
target update after :	6
sample selection :	random

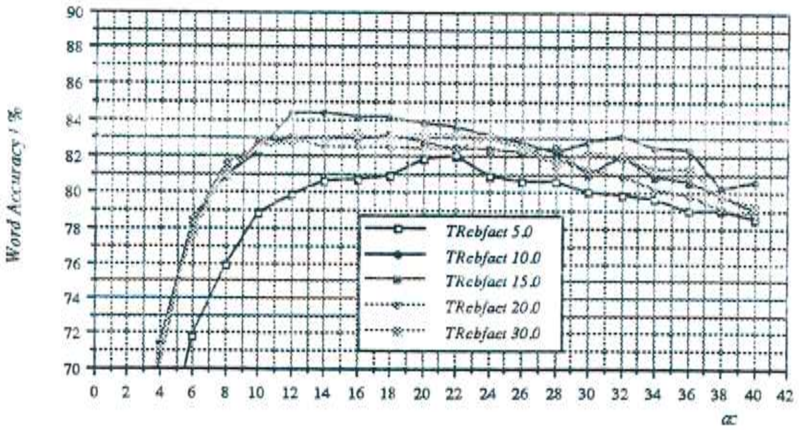
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.382	4.271	4.113	4.074	3.923	3.740
Q_{50}	15.860	-	-	-	-	-	11.408



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	38.3% (38.7%)	50.8% (51.3%)	54.4% (54.9%)	56.0% (56.7%)	57.2% (57.9%)
4.0	62.9% (64.3%)	71.3% (72.0%)	71.3% (72.0%)	69.9% (70.6%)	71.1% (72.7%)
6.0	71.8% (73.3%)	78.3% (79.3%)	78.4% (79.5%)	78.4% (79.3%)	77.7% (79.0%)
8.0	75.9% (76.8%)	80.9% (82.0%)	80.9% (82.2%)	81.5% (82.7%)	81.1% (82.7%)
10.0	78.8% (79.5%)	82.2% (82.9%)	82.7% (83.6%)	82.4% (83.8%)	82.4% (83.8%)
12.0	79.9% (80.7%)	84.3% (85.0%)	82.9% (83.8%)	82.9% (84.0%)	82.7% (83.8%)
14.0	80.6% (81.5%)	84.3% (85.0%)	82.9% (84.1%)	82.5% (84.0%)	82.9% (84.1%)
16.0	80.7% (81.6%)	84.1% (85.0%)	82.9% (84.3%)	82.5% (84.0%)	83.2% (84.7%)
18.0	80.9% (81.8%)	84.1% (85.0%)	83.1% (84.5%)	82.5% (84.0%)	82.9% (84.3%)
20.0	81.8% (82.9%)	83.8% (84.8%)	82.7% (84.7%)	82.4% (84.1%)	83.2% (84.7%)
22.0	82.0% (83.1%)	83.6% (85.2%)	82.4% (84.7%)	82.4% (84.3%)	83.1% (84.8%)
24.0	80.9% (82.9%)	83.1% (85.4%)	82.4% (84.7%)	82.9% (85.0%)	83.1% (85.2%)
26.0	80.6% (82.9%)	82.7% (85.2%)	82.2% (84.7%)	82.5% (85.2%)	82.5% (85.2%)
28.0	80.6% (82.9%)	82.2% (85.2%)	82.5% (85.0%)	81.8% (85.0%)	81.5% (84.7%)
30.0	80.0% (83.1%)	82.7% (85.7%)	81.0% (84.7%)	81.3% (84.8%)	82.2% (86.1%)
32.0	79.9% (83.4%)	83.1% (85.9%)	82.0% (85.7%)	80.9% (84.5%)	82.0% (86.1%)
34.0	79.7% (83.6%)	82.5% (85.4%)	80.9% (84.7%)	80.2% (84.1%)	81.3% (85.7%)
36.0	79.0% (82.7%)	82.4% (85.9%)	80.6% (84.7%)	79.9% (84.0%)	81.3% (85.9%)
38.0	79.0% (83.1%)	80.2% (83.4%)	79.9% (84.0%)	79.1% (83.6%)	79.9% (84.5%)
40.0	78.6% (83.8%)	80.6% (83.6%)	79.0% (83.1%)	78.8% (83.2%)	79.3% (84.0%)



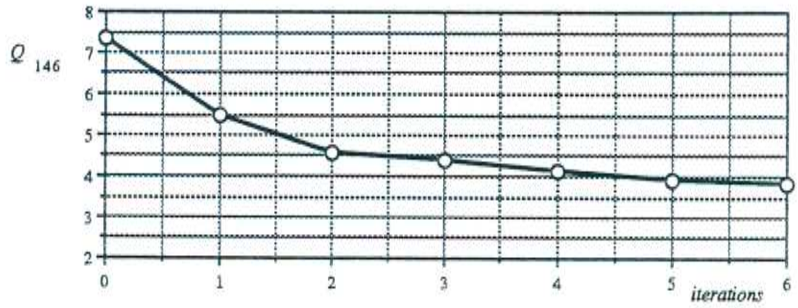
D.37 Experiment E.7.4

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	30
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

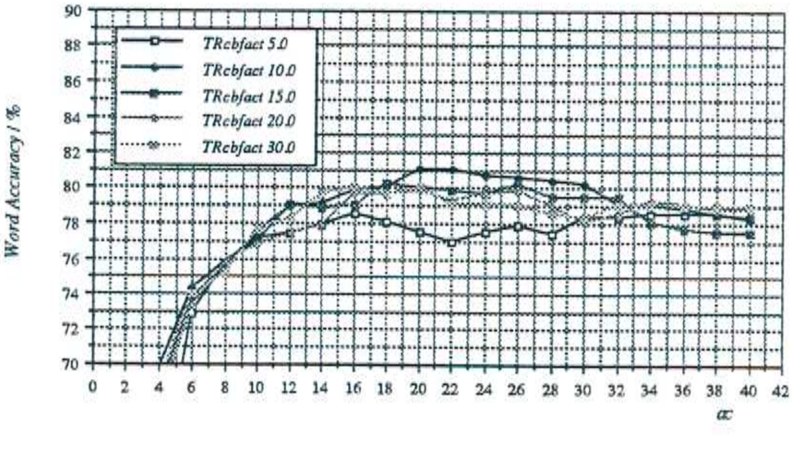
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.496	4.573	4.394	4.117	3.921	3.811
Q_{sn}	15.860	-	-	-	-	-	11.527



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	36.5% (37.3%)	53.5% (54.0%)	49.7% (50.4%)	48.3% (49.0%)	51.9% (52.2%)
4.0	62.6% (64.3%)	69.7% (70.8%)	68.4% (69.3%)	67.9% (69.0%)	69.0% (69.9%)
6.0	72.9% (74.5%)	74.3% (76.1%)	73.8% (75.4%)	73.3% (74.7%)	74.0% (75.6%)
8.0	75.8% (77.5%)	75.9% (77.9%)	75.6% (77.5%)	75.9% (77.4%)	75.2% (76.8%)
10.0	77.2% (78.8%)	77.0% (78.4%)	77.4% (78.8%)	76.8% (78.1%)	77.7% (79.9%)
12.0	77.4% (79.0%)	79.0% (80.4%)	79.0% (80.4%)	77.4% (78.8%)	78.3% (79.7%)
14.0	77.9% (79.5%)	79.1% (80.6%)	78.8% (80.6%)	77.9% (79.1%)	79.7% (80.7%)
16.0	78.6% (80.2%)	79.9% (81.6%)	79.0% (80.7%)	79.7% (81.5%)	80.0% (81.5%)
18.0	78.1% (80.2%)	80.0% (81.8%)	80.2% (82.0%)	79.7% (81.5%)	79.7% (81.5%)
20.0	77.5% (80.0%)	81.1% (82.9%)	80.0% (81.8%)	79.9% (81.8%)	80.0% (81.8%)
22.0	77.0% (79.9%)	81.1% (83.2%)	79.9% (81.8%)	79.3% (81.6%)	79.1% (81.6%)
24.0	77.5% (80.4%)	80.7% (83.1%)	79.7% (82.0%)	79.7% (82.2%)	79.1% (81.6%)
26.0	77.9% (80.4%)	80.6% (83.1%)	80.2% (82.4%)	79.9% (82.5%)	79.0% (81.5%)
28.0	77.4% (80.0%)	80.4% (82.7%)	79.5% (82.2%)	79.0% (82.0%)	78.6% (81.5%)
30.0	78.4% (81.5%)	80.2% (82.5%)	79.5% (82.5%)	78.1% (82.0%)	78.4% (82.2%)
32.0	78.4% (82.0%)	79.1% (82.4%)	79.5% (82.7%)	78.6% (82.5%)	79.1% (83.1%)
34.0	78.6% (82.2%)	79.1% (83.1%)	78.1% (82.0%)	79.3% (83.4%)	79.1% (83.1%)
36.0	78.6% (82.2%)	78.8% (82.7%)	77.7% (81.8%)	79.1% (83.4%)	79.0% (83.2%)
38.0	78.6% (82.2%)	78.6% (82.7%)	77.5% (81.6%)	78.8% (83.4%)	79.0% (83.2%)
40.0	78.4% (82.2%)	78.3% (82.4%)	77.5% (81.6%)	78.8% (83.4%)	79.0% (84.0%)



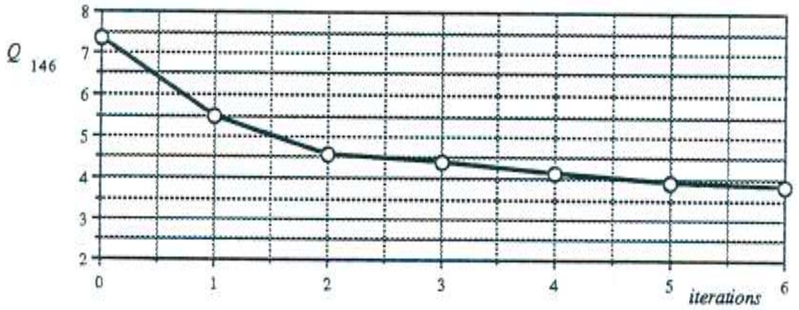
D.37 Experiment E.7.4

NLDA approach : $Y = A(X) + f(x)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	30
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

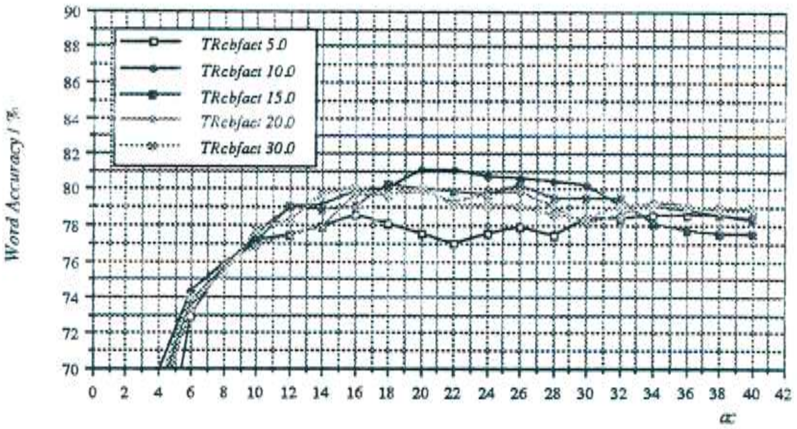
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.496	4.573	4.394	4.117	3.921	3.811
Q_{Kn}	15.860	-	-	-	-	-	11.527



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	36.5% (37.3%)	53.5% (54.0%)	49.7% (50.4%)	48.3% (49.0%)	51.9% (52.2%)
4.0	62.6% (64.3%)	69.7% (70.8%)	68.4% (69.3%)	67.9% (69.0%)	69.0% (69.9%)
6.0	72.9% (74.5%)	74.3% (76.1%)	73.8% (75.4%)	73.3% (74.7%)	74.0% (75.6%)
8.0	75.8% (77.5%)	75.9% (77.9%)	75.6% (77.5%)	75.9% (77.4%)	75.2% (76.8%)
10.0	77.2% (78.8%)	77.0% (78.4%)	77.4% (78.8%)	76.8% (78.1%)	77.7% (79.9%)
12.0	77.4% (79.0%)	79.0% (80.4%)	79.0% (80.4%)	77.4% (78.8%)	78.3% (79.7%)
14.0	77.9% (79.5%)	79.1% (80.6%)	78.8% (80.6%)	77.9% (79.1%)	79.7% (80.7%)
16.0	78.6% (80.2%)	79.9% (81.6%)	79.0% (80.7%)	79.7% (81.5%)	80.0% (81.5%)
18.0	78.1% (80.2%)	80.0% (81.8%)	80.2% (82.0%)	79.7% (81.5%)	79.7% (81.5%)
20.0	77.5% (80.0%)	81.1% (82.9%)	80.0% (81.8%)	79.9% (81.8%)	80.0% (81.8%)
22.0	77.0% (79.9%)	81.1% (83.2%)	79.9% (81.8%)	79.3% (81.6%)	79.1% (81.6%)
24.0	77.5% (80.4%)	80.7% (83.1%)	79.7% (82.0%)	79.7% (82.2%)	79.1% (81.6%)
26.0	77.9% (80.4%)	80.6% (83.1%)	80.2% (82.4%)	79.9% (82.5%)	79.0% (81.5%)
28.0	77.4% (80.0%)	80.4% (82.7%)	79.5% (82.2%)	79.0% (82.0%)	78.6% (81.5%)
30.0	78.4% (81.5%)	80.2% (82.5%)	79.5% (82.5%)	78.1% (82.0%)	78.4% (82.2%)
32.0	78.4% (82.0%)	79.1% (82.4%)	79.5% (82.7%)	78.6% (82.5%)	79.1% (83.1%)
34.0	78.6% (82.2%)	79.1% (83.1%)	78.1% (82.0%)	79.3% (83.4%)	79.1% (83.1%)
36.0	78.6% (82.2%)	78.8% (82.7%)	77.7% (81.8%)	79.1% (83.4%)	79.0% (83.2%)
38.0	78.6% (82.2%)	78.6% (82.7%)	77.5% (81.6%)	78.8% (83.4%)	79.0% (83.2%)
40.0	78.4% (82.2%)	78.3% (82.4%)	77.5% (81.6%)	78.8% (83.4%)	79.0% (84.0%)



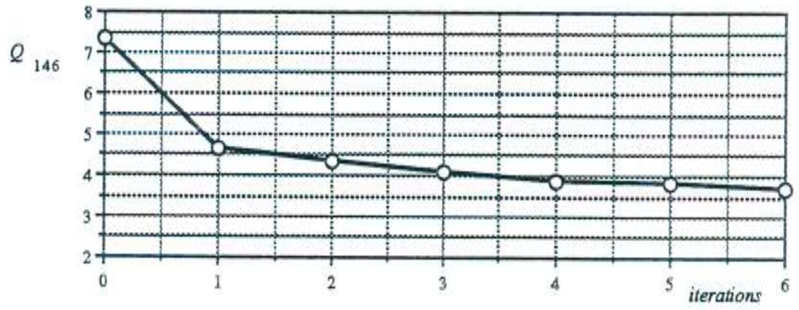
D.39 Experiment E.7.6

NLDA approach : $Y = A(X) + f(x)$

```
genaral parameters                                      backpropagation parameters
    dimension of orignial space :           32             number of discriminated targets :   146
    dimension of image space :              16             learning rate :                     0.008
    discriminated classes in on top LDA :   146            momentum :                          0.9
target drift parameters                                    number of hidden units :            30
    α :                                     0.9            iterations :                        6
    β :                                     1.0            target update after :               6
    m :                                     1              sample selection :                  random
    number of drift vectors :               50 (one per phonem class)
```

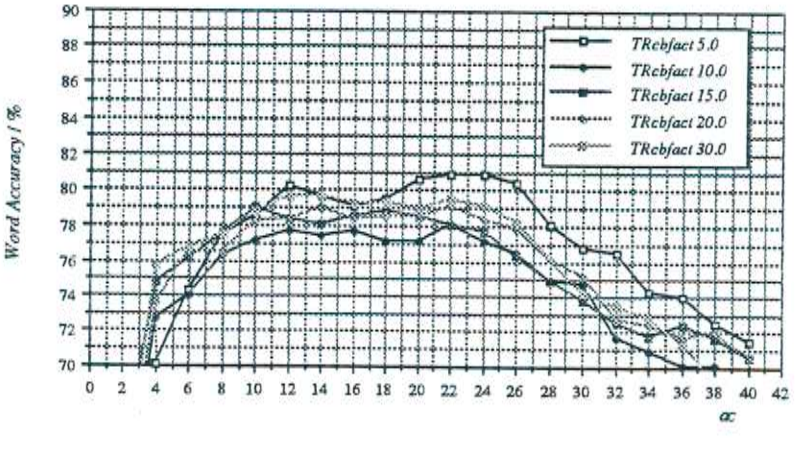
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	4.672	4.358	4.108	3.890	3.828	3.700
Q_{sn}	15.860	-	-	-	-	-	10.104



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	51.0% (51.5%)	59.5% (60.4%)	64.7% (65.8%)	61.0% (62.6%)	62.0% (62.9%)
4.0	70.1% (71.1%)	72.7% (74.0%)	74.7% (75.8%)	73.6% (75.0%)	75.8% (76.8%)
6.0	74.3% (75.6%)	74.0% (75.8%)	76.1% (77.4%)	76.5% (77.9%)	76.8% (78.3%)
8.0	77.5% (79.0%)	76.3% (78.3%)	77.5% (79.1%)	76.5% (78.6%)	77.5% (79.3%)
10.0	78.6% (80.2%)	77.2% (79.3%)	79.0% (80.7%)	78.3% (80.6%)	78.8% (80.6%)
12.0	80.2% (82.2%)	77.7% (80.4%)	78.4% (80.6%)	78.4% (80.6%)	79.7% (81.5%)
14.0	79.7% (81.8%)	77.4% (79.7%)	78.1% (80.0%)	79.0% (81.3%)	79.7% (81.8%)
16.0	79.1% (81.6%)	77.7% (80.0%)	78.6% (81.1%)	78.4% (81.1%)	79.3% (81.8%)
18.0	79.5% (82.5%)	77.2% (80.0%)	78.8% (81.3%)	78.6% (81.3%)	79.3% (81.8%)
20.0	80.6% (83.4%)	77.2% (80.4%)	78.6% (81.3%)	78.6% (81.3%)	79.0% (81.5%)
22.0	80.9% (84.3%)	78.1% (81.6%)	78.1% (81.5%)	79.1% (82.2%)	79.5% (82.4%)
24.0	80.9% (84.3%)	77.2% (80.6%)	77.7% (81.3%)	78.4% (81.8%)	79.1% (82.2%)
26.0	80.4% (84.0%)	76.5% (80.2%)	76.3% (80.7%)	77.9% (81.8%)	78.3% (82.4%)
28.0	78.1% (82.4%)	74.9% (79.7%)	74.9% (80.0%)	76.1% (81.5%)	76.1% (81.5%)
30.0	76.8% (81.3%)	74.7% (79.7%)	73.8% (79.1%)	75.2% (80.9%)	74.2% (79.9%)
32.0	76.5% (80.7%)	71.7% (77.5%)	72.5% (78.6%)	72.7% (80.0%)	73.4% (79.7%)
34.0	74.3% (80.2%)	70.9% (77.4%)	71.8% (78.3%)	72.4% (79.7%)	72.9% (79.0%)
36.0	74.0% (80.6%)	70.1% (77.2%)	72.4% (79.3%)	71.7% (79.7%)	71.3% (78.3%)
38.0	72.5% (80.6%)	70.1% (77.2%)	71.7% (79.1%)	72.0% (80.0%)	69.2% (77.9%)
40.0	71.5% (79.0%)	68.8% (76.5%)	70.6% (78.6%)	70.6% (79.1%)	69.0% (78.1%)



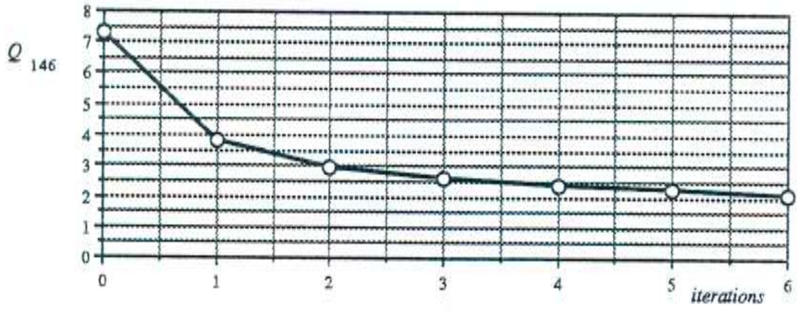
D.40 Experiment E.8.1

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.4	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	80 (one per phonem class)		

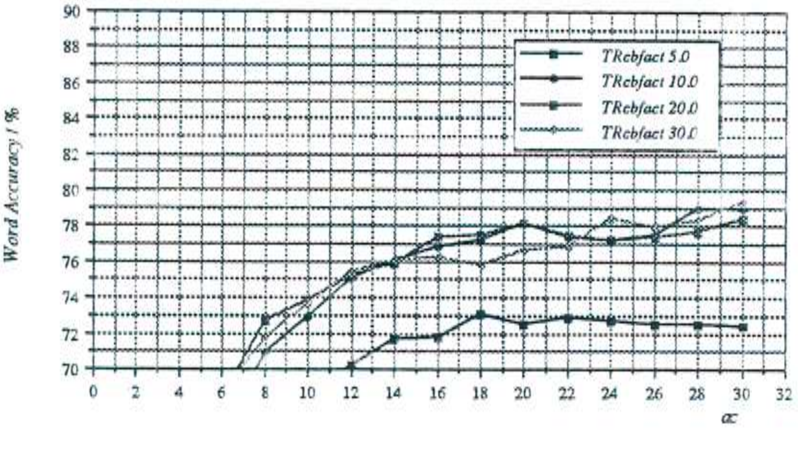
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	3.694	2.952	2.453	2.305	2.123	2.048
Q_{80}	15.860	-	-	-	-	-	8.054



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact			
ac	5.0	10.0	20.0	30.0
2.0	29.2% (31.6%)	46.7% (47.1%)	51.9% (52.4%)	53.7% (53.3%)
4.0	48.5% (49.2%)	58.6% (60.1%)	62.0% (62.9%)	63.5% (64.7%)
6.0	58.6% (59.9%)	66.3% (67.2%)	68.4% (69.5%)	68.4% (69.7%)
8.0	65.6% (67.4%)	70.9% (71.8%)	72.7% (74.2%)	71.8% (73.1%)
10.0	68.8% (70.8%)	72.9% (74.0%)	73.8% (75.4%)	73.6% (75.0%)
12.0	70.2% (72.0%)	75.0% (76.1%)	75.2% (76.8%)	75.4% (76.5%)
14.0	71.7% (73.4%)	76.1% (77.5%)	75.9% (77.0%)	76.1% (77.2%)
16.0	71.8% (74.2%)	76.8% (79.0%)	77.4% (78.4%)	76.3% (77.7%)
18.0	73.1% (75.2%)	77.2% (78.8%)	77.5% (79.1%)	75.8% (77.7%)
20.0	72.5% (74.7%)	78.1% (80.0%)	78.1% (79.9%)	76.6% (78.8%)
22.0	72.9% (75.2%)	77.4% (79.7%)	77.5% (80.0%)	76.8% (79.5%)
24.0	72.7% (75.6%)	77.2% (79.3%)	77.2% (79.9%)	78.4% (80.9%)
26.0	72.5% (75.4%)	77.5% (80.4%)	77.4% (80.2%)	77.9% (80.9%)
28.0	72.5% (75.4%)	79.0% (81.8%)	77.7% (80.9%)	78.3% (81.3%)
30.0	72.4% (75.4%)	79.0% (82.2%)	78.4% (81.6%)	79.3% (82.4%)



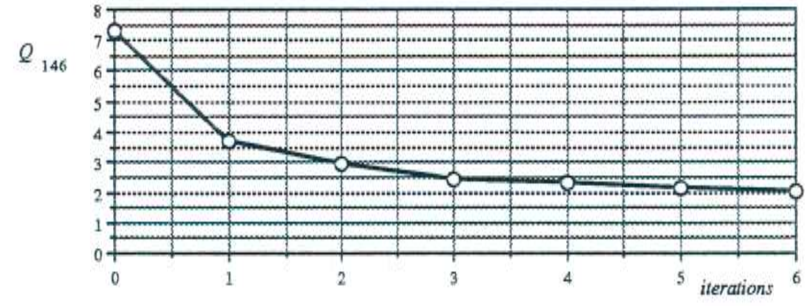
D.41 Experiment E.8.2

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	146
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.3	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

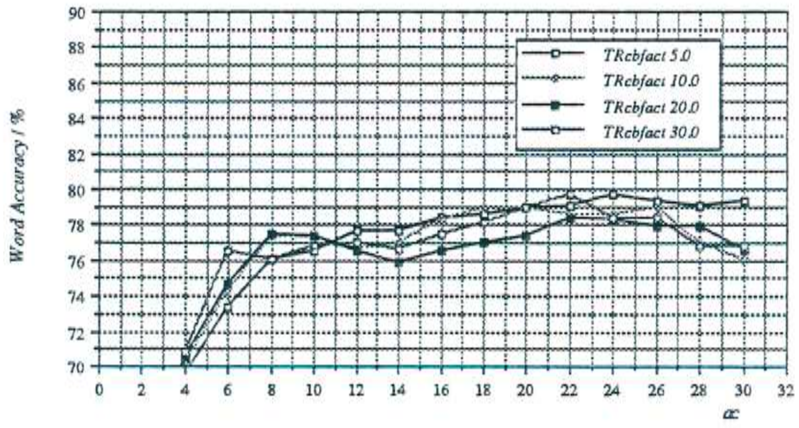
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	3.823	2.956	2.616	2.402	2.247	2.096
Q_{50}	15.860	-	-	-	-	-	7.915



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	2.0	5.0	10.0	20.0	30.0
2.0	3.4% (2.9%)	51.2% (51.2%)	58.8% (59.7%)	57.0% (57.6%)	57.9% (58.8%)
4.0	11.6% (12.5%)	69.5% (70.4%)	70.2% (71.1%)	70.6% (71.5%)	70.8% (71.5%)
6.0	17.5% (20.1%)	73.4% (74.3%)	74.3% (75.6%)	74.7% (75.8%)	76.5% (77.5%)
8.0	24.8% (27.3%)	76.1% (77.4%)	77.4% (79.1%)	77.5% (78.3%)	76.1% (77.5%)
10.0	32.8% (34.9%)	76.5% (78.8%)	77.4% (79.9%)	77.4% (79.5%)	76.8% (79.1%)
12.0	39.6% (42.6%)	77.7% (79.9%)	76.6% (79.9%)	76.5% (79.3%)	77.0% (79.5%)
14.0	41.5% (44.9%)	77.7% (80.0%)	77.0% (70.4%)	75.9% (79.3%)	76.6% (78.5%)
16.0	64.3% (48.3%)	78.4% (80.7%)	78.3% (81.3%)	76.5% (79.9%)	77.5% (80.4%)
18.0	46.3% (51.9%)	78.6% (80.9%)	79.0% (81.5%)	77.0% (80.9%)	78.1% (80.7%)
20.0	48.8% (55.4%)	79.0% (81.3%)	79.0% (81.8%)	77.4% (80.6%)	79.0% (81.6%)
22.0	50.4% (57.6%)	79.1% (81.6%)	78.6% (81.6%)	78.4% (81.6%)	79.7% (82.5%)
24.0	51.0% (58.6%)	79.7% (82.0%)	78.6% (82.0%)	78.3% (81.6%)	78.4% (81.6%)
26.0	52.9% (60.8%)	79.3% (81.6%)	79.0% (82.4%)	77.9% (81.8%)	78.4% (81.8%)
28.0	56.0% (63.5%)	79.1% (81.6%)	77.0% (81.6%)	77.9% (82.2%)	76.8% (81.6%)
30.0	56.9% (64.7%)	79.3% (82.0%)	76.1% (81.3%)	76.6% (81.8%)	76.8% (82.0%)



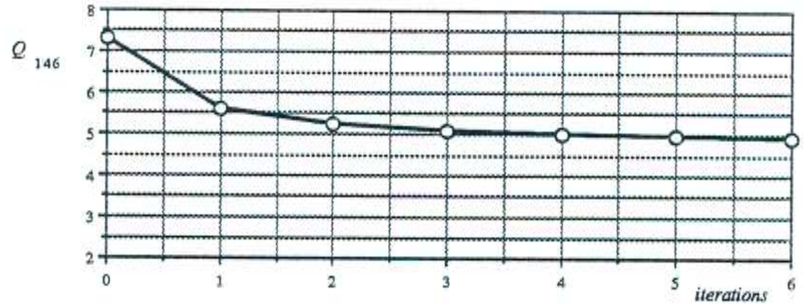
D.42 Experiment E.9.1

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of origmial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

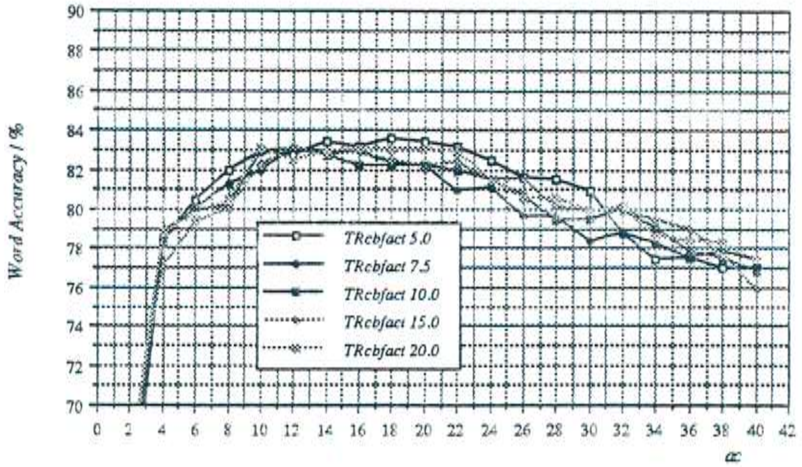
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.620	5.237	5.072	5.002	4.956	4.923
Q_{50}	15.860	-	-	-	-	-	13.477



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRebfact				
ac	5.0	7.5	10.0	15.0	20.0
2.0	61.3% (61.5%)	63.5% (63.8%)	64.3% (64.5%)	66.8% (67.4%)	66.5% (67.0%)
4.0	78.4% (79.1%)	78.4% (79.1%)	78.8% (79.7%)	77.0% (78.1%)	78.8% (79.7%)
6.0	80.4% (81.3%)	80.0% (81.3%)	79.9% (81.1%)	79.3% (80.4%)	79.7% (80.7%)
8.0	82.0% (82.9%)	81.3% (82.4%)	80.2% (81.5%)	80.0% (81.3%)	80.4% (81.8%)
10.0	82.9% (84.7%)	82.0% (83.6%)	82.9% (84.8%)	82.2% (84.0%)	83.1% (85.2%)
12.0	82.7% (84.5%)	83.1% (85.0%)	83.1% (85.0%)	83.2% (85.9%)	82.5% (85.2%)
14.0	83.4% (85.6%)	82.9% (85.0%)	82.7% (85.4%)	82.7% (85.6%)	82.9% (85.6%)
16.0	83.2% (85.7%)	82.9% (85.6%)	82.2% (85.0%)	82.9% (85.7%)	83.1% (85.7%)
18.0	83.6% (86.1%)	82.4% (85.2%)	82.2% (85.0%)	82.5% (85.6%)	83.1% (86.1%)
20.0	83.4% (86.1%)	82.2% (85.0%)	82.2% (85.2%)	82.2% (85.2%)	83.1% (86.1%)
22.0	83.2% (86.1%)	80.9% (84.1%)	82.0% (84.8%)	82.4% (85.4%)	82.9% (85.9%)
24.0	82.5% (85.4%)	81.1% (84.7%)	81.5% (84.7%)	81.5% (85.6%)	81.5% (85.6%)
26.0	81.6% (84.8%)	79.7% (84.1%)	80.7% (84.8%)	81.5% (85.6%)	80.6% (85.2%)
28.0	81.5% (84.8%)	79.7% (84.1%)	79.5% (84.8%)	80.2% (84.5%)	80.6% (85.2%)
30.0	80.9% (84.5%)	78.4% (83.6%)	79.5% (84.3%)	79.9% (84.3%)	79.9% (84.8%)
32.0	78.8% (83.2%)	78.8% (84.5%)	80.0% (84.5%)	79.9% (84.7%)	80.2% (84.7%)
34.0	77.4% (83.1%)	78.3% (84.3%)	79.1% (84.1%)	79.5% (84.0%)	78.6% (83.4%)
36.0	77.5% (83.2%)	77.5% (83.4%)	77.9% (82.9%)	79.0% (84.0%)	78.4% (83.4%)
38.0	77.0% (83.1%)	77.9% (83.4%)	77.5% (82.7%)	77.7% (82.9%)	78.3% (83.1%)
40.0	77.0% (83.1%)	77.4% (83.2%)	76.8% (82.7%)	75.9% (81.5%)	77.4% (82.7%)



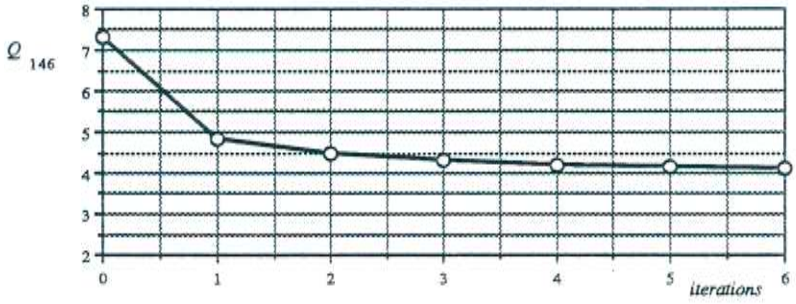
D.43 Experiment E.9.2

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

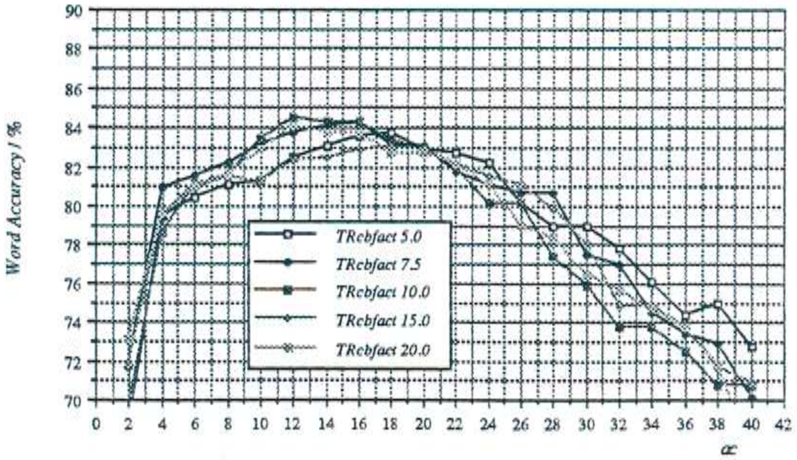
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	4.832	4.487	4.303	4.213	4.167	4.135
Q_{50}	15.860	-	-	-	-	-	11.970



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	7.5	10.0	15.0	20.0
2.0	68.4% (69.2%)	71.7% (72.7%)	69.5% (70.6%)	71.8% (72.7%)	73.1% (73.8%)
4.0	79.5% (80.6%)	80.9% (82.2%)	78.6% (80.0%)	79.7% (80.9%)	79.7% (80.4%)
6.0	80.4% (81.8%)	81.5% (83.2%)	81.1% (82.7%)	80.9% (82.9%)	81.1% (82.7%)
8.0	81.1% (83.8%)	82.2% (84.3%)	81.6% (83.8%)	81.5% (83.6%)	81.6% (83.8%)
10.0	81.3% (83.8%)	83.2% (85.6%)	83.4% (85.7%)	81.3% (83.6%)	82.9% (85.2%)
12.0	82.5% (84.8%)	83.8% (86.3%)	84.5% (87.3%)	82.4% (84.8%)	84.1% (86.8%)
14.0	83.1% (85.9%)	84.1% (86.8%)	84.3% (87.9%)	82.5% (86.1%)	83.8% (86.6%)
16.0	83.6% (86.6%)	84.3% (87.0%)	84.3% (87.9%)	82.9% (86.6%)	83.8% (86.6%)
18.0	83.8% (86.8%)	83.2% (86.8%)	83.4% (87.2%)	83.1% (86.8%)	82.7% (86.5%)
20.0	82.9% (86.6%)	83.1% (86.6%)	82.9% (86.8%)	82.9% (86.6%)	82.7% (86.5%)
22.0	82.7% (86.6%)	81.8% (85.7%)	82.0% (86.1%)	82.2% (85.9%)	82.2% (86.1%)
24.0	82.2% (86.3%)	81.1% (85.6%)	80.2% (85.2%)	81.5% (85.9%)	81.1% (85.4%)
26.0	80.2% (85.4%)	80.7% (85.6%)	80.2% (85.2%)	81.1% (85.9%)	79.1% (85.2%)
28.0	79.0% (84.5%)	80.7% (85.6%)	77.4% (83.1%)	79.9% (84.7%)	78.4% (84.5%)
30.0	79.0% (84.7%)	77.5% (83.4%)	75.9% (82.5%)	78.1% (83.8%)	76.5% (82.9%)
32.0	77.9% (84.0%)	77.0% (82.7%)	73.8% (80.7%)	74.9% (81.3%)	75.6% (81.8%)
34.0	76.1% (82.2%)	74.5% (81.3%)	73.8% (80.6%)	74.9% (81.5%)	74.7% (80.9%)
36.0	74.4% (80.9%)	73.4% (79.9%)	72.5% (78.4%)	73.8% (79.3%)	73.8% (80.2%)
38.0	75.0% (81.1%)	72.9% (79.1%)	70.8% (77.7%)	71.7% (78.3%)	71.8% (78.4%)
40.0	72.7% (79.3%)	70.2% (76.5%)	70.8% (78.3%)	70.9% (77.5%)	67.4% (74.5%)



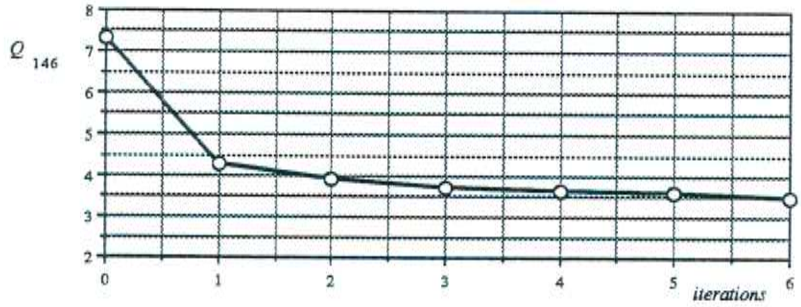
D.44 Experiment E.9.3

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	80 (one per phonem class)		

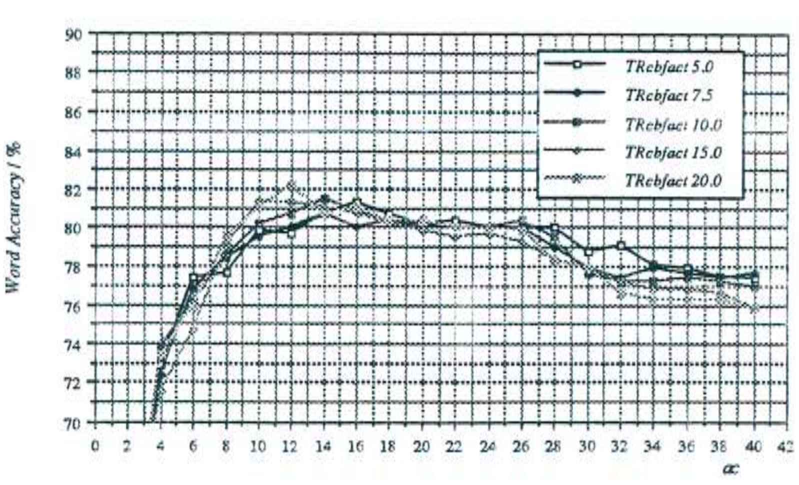
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	4.291	3.934	3.729	3.644	3.585	3.482
Q_{80}	15.860	-	-	-	-	-	10.956



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRebfact				
ac	5.0	7.5	10.0	15.0	20.0
2.0	57.2% (56.1%)	56.9% (57.4%)	59.7% (61.0%)	62.7% (64.0%)	62.6% (63.6%)
4.0	72.9% (74.0%)	72.4% (73.4%)	73.8% (75.2%)	71.5% (72.9%)	73.4% (74.7%)
6.0	77.4% (79.0%)	76.8% (78.6%)	76.1% (78.1%)	74.7% (76.5%)	75.9% (77.2%)
8.0	77.7% (79.5%)	78.4% (80.2%)	78.6% (80.4%)	79.1% (81.1%)	79.5% (81.1%)
10.0	79.9% (82.2%)	79.5% (82.0%)	80.2% (82.7%)	81.3% (83.2%)	81.1% (83.6%)
12.0	79.7% (82.0%)	80.0% (83.1%)	80.7% (83.6%)	81.3% (84.1%)	82.2% (84.5%)
14.0	80.7% (83.2%)	80.7% (83.8%)	81.5% (84.3%)	81.1% (84.3%)	80.7% (84.0%)
16.0	81.3% (83.8%)	80.0% (83.6%)	80.9% (84.1%)	80.7% (84.0%)	81.1% (84.3%)
18.0	80.7% (83.6%)	80.4% (83.8%)	80.4% (84.1%)	80.2% (83.8%)	80.4% (84.1%)
20.0	80.2% (83.4%)	80.2% (83.8%)	80.0% (83.8%)	79.9% (83.6%)	80.4% (84.3%)
22.0	80.4% (83.8%)	80.0% (83.6%)	80.0% (83.8%)	79.5% (83.2%)	80.0% (84.0%)
24.0	80.0% (83.6%)	80.0% (83.6%)	80.0% (83.6%)	79.7% (84.0%)	80.0% (84.0%)
26.0	80.0% (83.6%)	80.0% (83.6%)	80.4% (84.3%)	79.3% (84.0%)	79.9% (83.8%)
28.0	80.0% (83.6%)	79.0% (82.9%)	79.5% (83.8%)	78.3% (83.2%)	78.3% (83.1%)
30.0	78.8% (83.1%)	77.9% (83.1%)	77.7% (83.1%)	77.9% (83.2%)	77.9% (83.1%)
32.0	79.1% (83.6%)	77.4% (83.1%)	77.2% (83.1%)	77.2% (83.2%)	76.6% (82.9%)
34.0	78.1% (82.9%)	77.9% (83.8%)	77.2% (83.1%)	76.8% (83.2%)	76.3% (82.7%)
36.0	77.9% (82.7%)	77.7% (83.6%)	77.4% (83.6%)	76.8% (83.2%)	76.3% (82.7%)
38.0	77.5% (82.5%)	77.4% (83.6%)	77.2% (83.4%)	76.6% (83.1%)	76.3% (82.7%)
40.0	77.4% (82.5%)	77.7% (84.0%)	77.0% (83.6%)	75.9% (82.2%)	75.8% (81.8%)



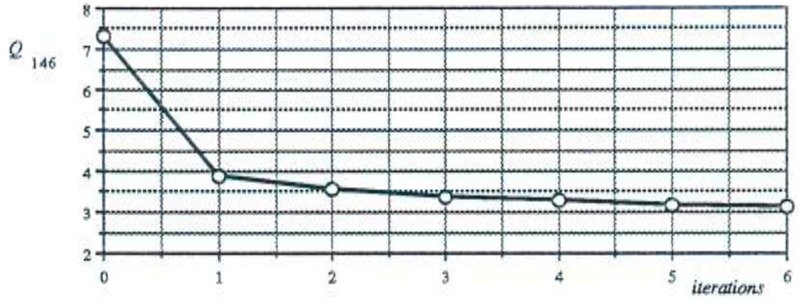
D.45 Experiment E.9.4

NLDA approach : $Y = Af(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

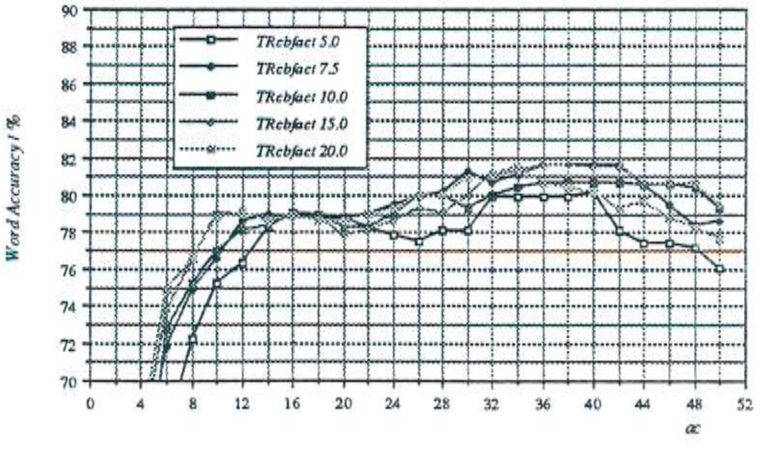
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	3.875	3.553	3.378	3.271	3.150	3.127
Q_{50}	15.860	-	-	-	-	-	10.180



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	7.5	10.0	15.0	20.0
6.0	67.4% (68.6%)	71.8% (72.5%)	72.7% (73.8%)	73.8% (74.9%)	74.9% (75.8%)
8.0	72.2% (74.0%)	74.9% (75.9%)	75.2% (76.5%)	76.5% (77.9%)	76.6% (78.1%)
10.0	75.2% (76.8%)	76.6% (78.1%)	77.0% (78.8%)	79.1% (80.7%)	78.8% (80.4%)
12.0	76.3% (77.5%)	78.6% (80.6%)	78.1% (80.4%)	79.0% (80.7%)	79.1% (80.9%)
14.0	78.3% (79.9%)	79.0% (80.9%)	78.4% (80.7%)	78.8% (81.1%)	78.6% (80.9%)
16.0	79.1% (80.7%)	78.8% (80.9%)	79.1% (81.1%)	78.8% (81.3%)	79.0% (81.1%)
18.0	78.6% (80.7%)	79.0% (81.5%)	79.0% (81.3%)	78.8% (81.1%)	78.6% (80.9%)
20.0	78.8% (81.1%)	78.8% (81.8%)	78.3% (81.1%)	77.9% (80.9%)	78.6% (81.3%)
22.0	78.3% (80.9%)	79.0% (82.0%)	78.3% (81.1%)	78.1% (81.1%)	79.0% (81.8%)
24.0	77.9% (80.9%)	79.5% (82.9%)	79.0% (81.8%)	78.6% (81.5%)	79.3% (82.4%)
26.0	77.5% (80.9%)	79.9% (83.4%)	80.0% (83.2%)	79.3% (82.2%)	80.0% (83.2%)
28.0	78.1% (81.6%)	80.2% (84.0%)	80.0% (83.8%)	79.1% (82.7%)	80.0% (83.8%)
30.0	78.1% (82.0%)	81.3% (85.0%)	79.3% (84.1%)	80.0% (83.8%)	80.9% (84.8%)
32.0	79.9% (83.8%)	80.7% (84.5%)	80.0% (83.8%)	81.1% (85.0%)	81.1% (85.2%)
34.0	79.9% (83.8%)	81.1% (85.0%)	80.4% (84.3%)	81.3% (85.2%)	81.5% (85.4%)
36.0	79.9% (84.0%)	81.6% (85.6%)	80.7% (84.8%)	81.6% (85.6%)	80.7% (84.8%)
38.0	79.9% (84.5%)	81.6% (85.6%)	80.7% (84.8%)	81.6% (85.6%)	80.4% (84.7%)
40.0	80.2% (85.0%)	81.6% (85.6%)	80.7% (84.8%)	81.5% (85.6%)	80.2% (84.7%)
42.0	78.1% (83.6%)	81.6% (85.6%)	80.7% (84.8%)	81.5% (85.6%)	79.3% (84.1%)
44.0	77.4% (83.2%)	80.6% (84.5%)	80.6% (84.7%)	80.7% (85.2%)	79.7% (84.3%)
46.0	77.4% (83.2%)	79.5% (84.1%)	80.6% (84.7%)	80.6% (85.0%)	78.8% (84.1%)
48.0	77.2% (83.1%)	78.4% (83.6%)	80.4% (85.0%)	80.7% (85.2%)	78.3% (83.8%)
50.0	76.1% (82.5%)	78.6% (84.0%)	79.3% (84.5%)	79.5% (84.7%)	77.7% (83.6%)



D.46 Experiment E.9.5

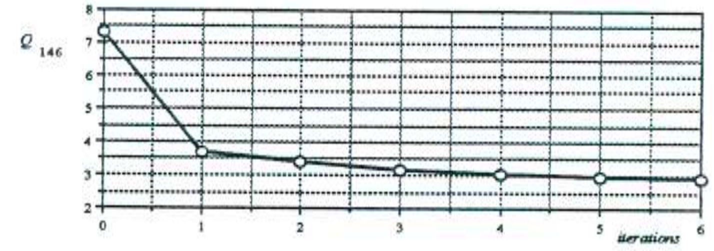
NLDA approach : $Y = Af(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.4
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	2451
learning rate :	0.008
momentum :	0.9
number of hidden units :	32
iterations :	6
target update after :	6
sample selection :	random

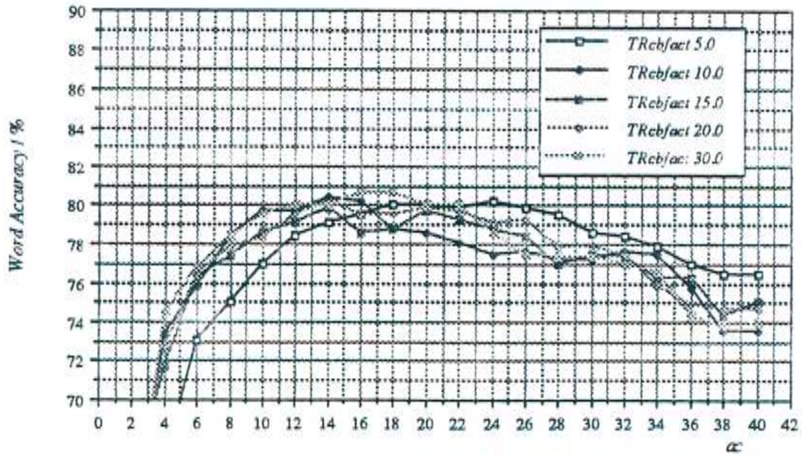
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	3.662	3.389	3.158	3.048	2.953	2.930
Q_{50}	15.860	-	-	-	-	-	9.610



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	10.0	15.0	20.0	30.0
2.0	55.8% (56.7%)	62.4% (63.5%)	64.7% (65.6%)	62.9% (64.0%)	62.4% (62.7%)
4.0	66.7% (68.3%)	73.4% (74.9%)	71.7% (72.7%)	72.7% (74.2%)	74.5% (75.4%)
6.0	73.1% (75.4%)	75.8% (76.8%)	76.3% (77.4%)	76.5% (78.1%)	76.8% (77.5%)
8.0	75.0% (77.0%)	78.4% (79.7%)	77.4% (78.4%)	77.9% (79.1%)	78.4% (79.1%)
10.0	77.0% (78.6%)	79.7% (81.3%)	78.6% (79.7%)	78.3% (79.3%)	79.5% (80.6%)
12.0	78.4% (80.0%)	79.7% (81.6%)	79.1% (80.9%)	79.5% (80.7%)	80.0% (81.1%)
14.0	79.1% (80.7%)	80.4% (82.4%)	79.9% (81.6%)	80.2% (82.5%)	80.0% (81.6%)
16.0	79.5% (81.5%)	80.2% (82.2%)	78.6% (80.9%)	79.5% (82.2%)	80.6% (82.9%)
18.0	80.0% (82.0%)	78.8% (81.3%)	78.8% (81.5%)	79.5% (82.2%)	80.6% (82.9%)
20.0	80.0% (82.0%)	78.6% (81.5%)	79.7% (82.5%)	79.9% (82.7%)	80.0% (82.9%)
22.0	79.9% (82.0%)	78.1% (81.3%)	79.3% (82.4%)	79.7% (82.7%)	80.0% (83.1%)
24.0	80.2% (82.5%)	77.5% (81.1%)	78.8% (82.2%)	79.1% (82.5%)	78.6% (82.7%)
26.0	79.9% (82.4%)	77.7% (81.5%)	78.4% (82.4%)	79.3% (83.1%)	77.5% (82.4%)
28.0	79.5% (82.2%)	77.2% (81.3%)	77.0% (82.2%)	77.9% (82.9%)	77.4% (82.5%)
30.0	78.6% (81.6%)	77.2% (81.8%)	77.5% (82.7%)	77.9% (82.9%)	77.4% (82.5%)
32.0	78.4% (81.5%)	77.7% (82.2%)	77.4% (82.5%)	77.5% (82.7%)	77.0% (82.4%)
34.0	77.9% (81.3%)	77.5% (82.0%)	76.3% (82.0%)	75.9% (82.0%)	76.6% (82.0%)
36.0	77.0% (80.6%)	75.8% (81.1%)	76.3% (81.6%)	74.9% (81.5%)	74.3% (81.3%)
38.0	76.5% (80.2%)	73.6% (79.3%)	74.3% (80.2%)	74.7% (81.3%)	73.8% (80.4%)
40.0	76.5% (80.6%)	73.6% (79.3%)	75.0% (80.7%)	74.7% (80.4%)	74.0% (80.2%)



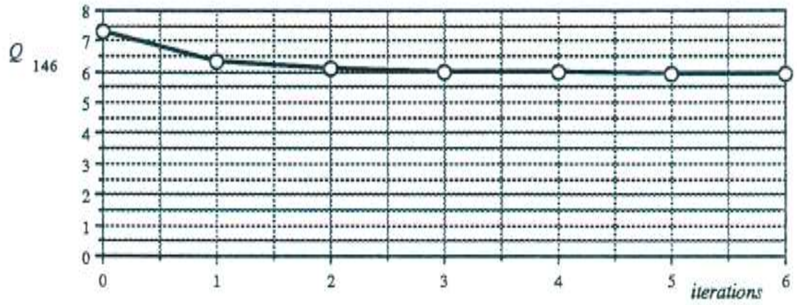
D.47 Experiment E.10.1

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	80 (one per phonem class)		

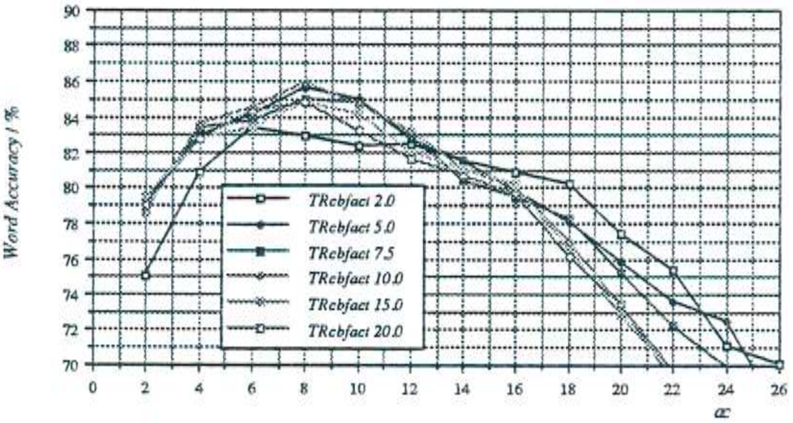
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.334	6.090	5.990	5.972	5.934	5.940
Q_{80}	15.860	-	-	-	-	-	14.691



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	75.0% (75.9%)	79.1% (79.9%)	78.8% (79.5%)	78.6% (79.3%)	79.5% (80.2%)	79.0% (79.7%)
4.0	80.9% (82.2%)	82.9% (83.8%)	83.4% (84.3%)	83.6% (84.5%)	82.7% (84.0%)	82.7% (84.0%)
6.0	83.4% (86.1%)	84.1% (85.6%)	83.8% (85.4%)	84.5% (86.3%)	83.4% (85.0%)	84.3% (85.9%)
8.0	82.9% (85.9%)	85.6% (87.3%)	85.0% (87.5%)	85.9% (88.1%)	84.8% (87.7%)	84.8% (87.5%)
10.0	82.4% (86.3%)	85.0% (87.9%)	84.8% (87.9%)	84.7% (87.3%)	84.1% (86.8%)	83.2% (86.6%)
12.0	82.5% (87.0%)	82.7% (86.1%)	82.9% (86.3%)	83.2% (87.5%)	82.0% (86.6%)	81.6% (86.3%)
14.0	81.5% (85.9%)	81.5% (86.3%)	80.4% (86.3%)	81.1% (86.6%)	81.1% (86.6%)	80.7% (86.3%)
16.0	80.9% (85.9%)	79.9% (84.8%)	79.5% (84.8%)	80.0% (85.4%)	80.2% (85.7%)	79.7% (85.4%)
18.0	80.2% (85.6%)	78.1% (82.7%)	78.3% (83.2%)	77.0% (81.5%)	76.5% (81.5%)	76.1% (80.9%)
20.0	77.4% (83.1%)	75.9% (81.5%)	75.2% (80.2%)	73.4% (79.0%)	72.9% (77.7%)	73.4% (79.3%)
22.0	75.4% (81.1%)	73.6% (78.4%)	72.2% (77.2%)	69.2% (75.9%)	69.5% (76.1%)	69.7% (76.6%)
24.0	71.1% (77.9%)	72.5% (77.2%)	69.9% (75.0%)	67.6% (72.7%)	63.8% (71.7%)	67.7% (73.4%)
26.0	70.1% (76.1%)	67.2% (74.2%)	67.4% (73.6%)	66.5% (73.3%)	64.2% (70.4%)	62.7% (70.1%)



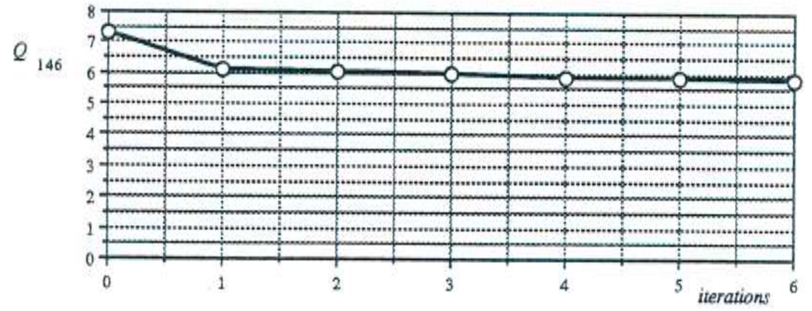
D.48 Experiment E.10.2

NLDA approach : $Y = AX + f(X)$

```
genaral parameters                                   backpropagation parameters
    dimension of orignial space :              32        number of discriminated targets :    2451
    dimension of image space :                 16        learning rate :                      0.008
    discriminated classes in on top LDA :      146       momentum :                           0.9
target drift parameters                                  number of hidden units :             10
    α :                                        0.9       iterations :                         6
    β :                                        0.1       target update after :                6
    m :                                        1         sample selection :                   random
    number of drift vectors :                  50 (one per phonem class)
```

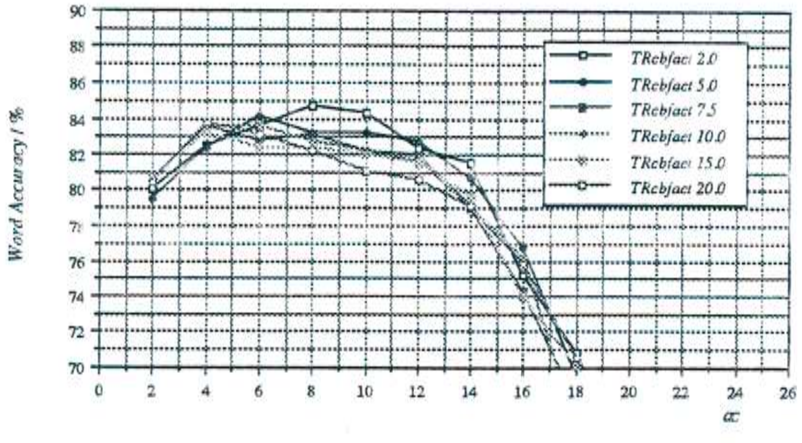
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.108	6.036	5.999	5.883	5.875	5.817
Q_{Kn}	15.860	-	-	-	-	-	14.392



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRebfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	80.0% (81.1%)	79.5% (80.4%)	80.4% (81.3%)	80.6% (81.5%)	80.7% (81.5%)	80.4% (81.3%)
4.0	82.5% (85.0%)	82.4% (84.3%)	83.6% (85.6%)	83.6% (85.6%)	83.2% (85.2%)	83.8% (85.9%)
6.0	83.6% (86.5%)	84.1% (87.0%)	82.7% (85.7%)	83.6% (85.6%)	82.4% (85.7%)	83.1% (86.5%)
8.0	84.7% (87.9%)	83.2% (86.8%)	83.1% (86.6%)	82.7% (86.6%)	82.4% (86.1%)	82.2% (86.3%)
10.0	84.3% (88.1%)	83.2% (87.7%)	82.2% (86.6%)	82.2% (86.3%)	82.0% (86.1%)	81.1% (85.9%)
12.0	82.4% (87.5%)	82.7% (86.5%)	82.0% (87.0%)	81.6% (86.1%)	81.5% (85.6%)	80.6% (85.0%)
14.0	81.5% (86.6%)	80.7% (85.2%)	79.0% (84.0%)	79.3% (83.6%)	79.7% (83.8%)	79.1% (84.1%)
16.0	75.2% (81.6%)	76.8% (82.5%)	74.3% (81.1%)	76.1% (82.2%)	74.0% (79.5%)	75.9% (81.8%)
18.0	70.8% (78.4%)	69.5% (76.1%)	68.1% (75.2%)	66.8% (74.0%)	70.1% (76.6%)	70.6% (78.1%)
20.0	64.2% (72.0%)	66.0% (73.1%)	64.7% (71.8%)	64.5% (72.7%)	64.5% (73.4%)	64.3% (72.9%)
22.0	60.2% (69.3%)	61.1% (70.4%)	63.5% (71.8%)	63.5% (71.8%)	61.3% (70.4%)	62.0% (70.9%)
24.0	54.4% (64.0%)	59.9% (68.3%)	61.7% (69.3%)	59.7% (67.7%)	58.8% (65.6%)	58.6% (67.4%)
26.0	54.4% (63.3%)	57.2% (65.4%)	53.5% (61.1%)	57.6% (65.1%)	56.3% (63.1%)	54.4% (63.1%)



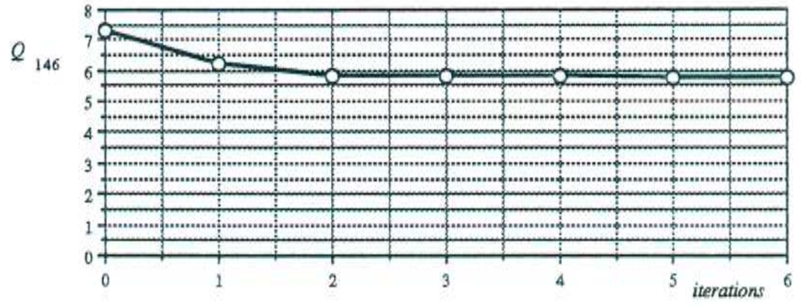
D.49 Experiment E.10.3

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	0.9	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

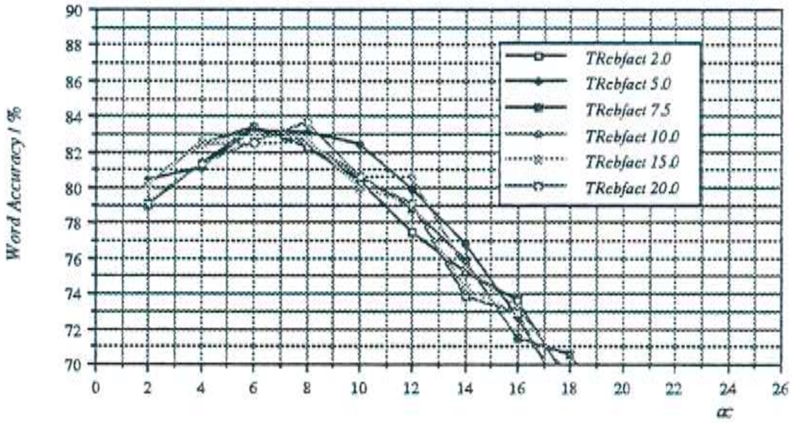
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.206	5.808	5.802	5.788	5.777	5.749
Q_{50}	15.860	-	-	-	-	-	14.034



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	79.0% (80.2%)	80.4% (80.9%)	80.2% (80.7%)	80.2% (80.7%)	80.2% (80.9%)	79.1% (79.9%)
4.0	81.3% (83.2%)	81.1% (84.3%)	82.4% (84.7%)	82.4% (84.7%)	82.5% (85.0%)	81.3% (83.4%)
6.0	83.4% (85.7%)	83.2% (85.9%)	83.4% (86.3%)	82.4% (85.4%)	82.9% (86.3%)	82.5% (85.9%)
8.0	82.2% (85.6%)	83.1% (86.3%)	82.5% (86.1%)	82.5% (86.5%)	82.7% (86.3%)	83.6% (87.3%)
10.0	80.4% (84.7%)	82.4% (86.5%)	80.7% (848.0%)	80.6% (85.0%)	80.0% (84.8%)	80.6% (85.4%)
12.0	77.5% (83.6%)	79.9% (84.7%)	78.8% (83.6%)	80.6% (84.5%)	79.1% (84.5%)	79.1% (84.1%)
14.0	75.2% (80.9%)	76.8% (81.6%)	75.9% (80.7%)	75.4% (80.6%)	74.3% (80.0%)	73.8% (79.5%)
16.0	73.6% (79.7%)	72.7% (78.4%)	71.5% (77.7%)	73.3% (78.3%)	72.9% (78.8%)	72.9% (78.8%)
18.0	69.0% (75.4%)	67.7% (74.3%)	70.6% (76.5%)	69.3% (76.3%)	67.6% (74.7%)	67.7% (74.3%)
20.0	64.2% (71.3%)	69.0% (75.0%)	67.2% (74.2%)	64.0% (71.5%)	64.9% (72.4%)	64.3% (72.4%)
22.0	64.2% (71.5%)	60.4% (68.1%)	60.8% (68.8%)	61.1% (69.0%)	58.8% (67.0%)	60.4% (68.4%)
24.0	56.5% (64.3%)	54.2% (62.7%)	55.6% (64.2%)	57.2% (65.8%)	56.5% (64.2%)	58.6% (66.7%)
26.0	57.2% (60.2%)	57.5% (60.4%)	54.4% (61.1%)	53.5% (60.4%)	50.3% (57.2%)	49.9% (57.6%)



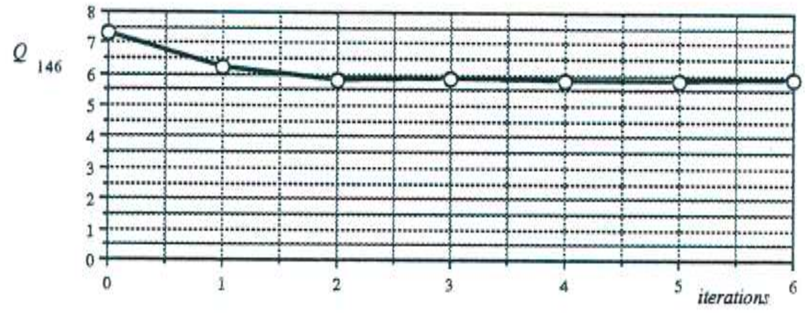
D.50 Experiment E.10.4

NLDA approach : $Y = AX + f(X)$

```
genaral parameters                                        backpropagation parameters
    dimension of orignial space :               32            number of discriminated targets :   2451
    dimension of image space :                  16            learning rate :                     0.008
    discriminated classes in on top LDA :       146           momentum :                          0.9
target drift parameters                                       number of hidden units :            10
    α :                                         0.9           iterations :                        6
    β :                                         0.3           target update after :               6
    m :                                         1             sample selection :                  random
    number of drift vectors :                   50 (one per phonem class)
```

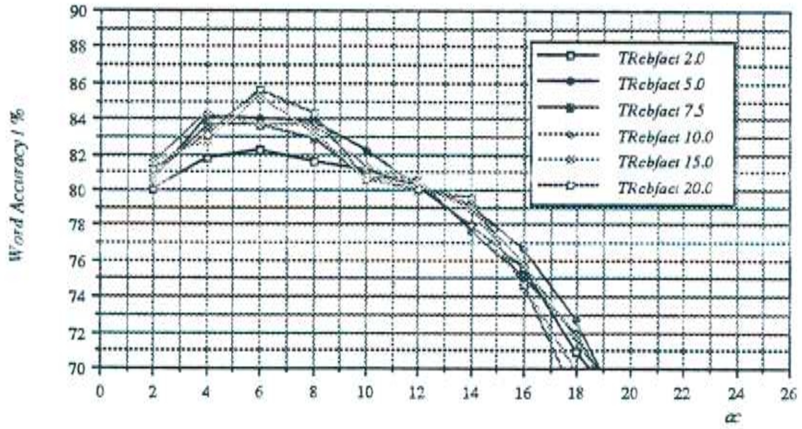
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.209	5.814	5.848	5.837	5.828	5.845
Q_{50}	15.860	-	-	-	-	-	13.959



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	80.0% (80.7%)	80.6% (81.3%)	80.7% (81.5%)	81.6% (82.4%)	81.6% (82.4%)	81.1% (81.1%)
4.0	81.8% (83.4%)	84.1% (85.7%)	83.6% (85.2%)	84.3% (85.7%)	82.7% (84.7%)	83.1% (84.5%)
6.0	82.2% (84.8%)	84.0% (86.6%)	83.6% (85.7%)	83.6% (85.9%)	85.2% (87.5%)	85.6% (87.7%)
8.0	81.6% (84.7%)	83.8% (87.7%)	82.9% (86.1%)	83.6% (86.6%)	83.2% (86.1%)	84.3% (87.0%)
10.0	81.1% (84.7%)	82.2% (85.7%)	80.7% (84.3%)	81.3% (84.8%)	80.7% (84.3%)	80.6% (84.3%)
12.0	80.2% (84.7%)	80.2% (84.5%)	80.6% (84.3%)	80.4% (84.3%)	80.2% (84.1%)	80.0% (84.5%)
14.0	77.9% (83.6%)	79.1% (83.6%)	77.7% (83.2%)	79.1% (83.6%)	79.0% (83.2%)	79.5% (83.8%)
16.0	75.6% (81.3%)	76.6% (81.8%)	75.0% (80.2%)	75.8% (81.1%)	74.7% (80.7%)	74.7% (80.7%)
18.0	70.9% (78.1%)	72.7% (79.0%)	72.0% (78.4%)	71.8% (78.3%)	69.9% (77.0%)	68.3% (75.8%)
20.0	67.4% (75.8%)	66.5% (74.2%)	66.7% (73.6%)	66.5% (73.3%)	66.0% (73.1%)	64.9% (73.4%)
22.0	65.2% (73.1%)	63.1% (70.8%)	65.6% (72.0%)	64.3% (70.9%)	62.7% (70.9%)	59.9% (69.3%)
24.0	61.3% (70.4%)	60.1% (68.1%)	61.5% (69.2%)	60.1% (68.3%)	63.5% (69.9%)	60.4% (68.3%)
26.0	58.3% (67.0%)	59.0% (67.2%)	59.0% (66.3%)	59.7% (67.7%)	58.1% (65.1%)	56.9% (64.3%)



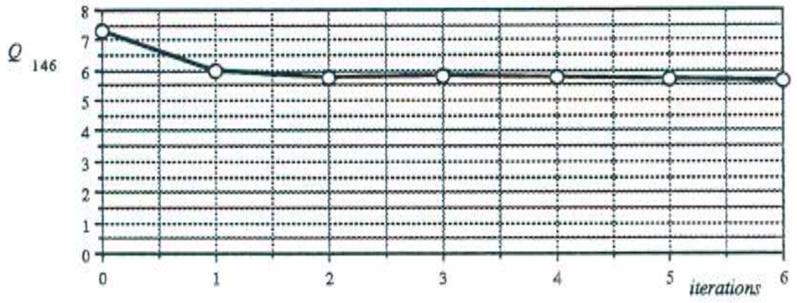
D.51 Experiment E.10.5

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	0.9	iterations :	6
β :	0.4	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

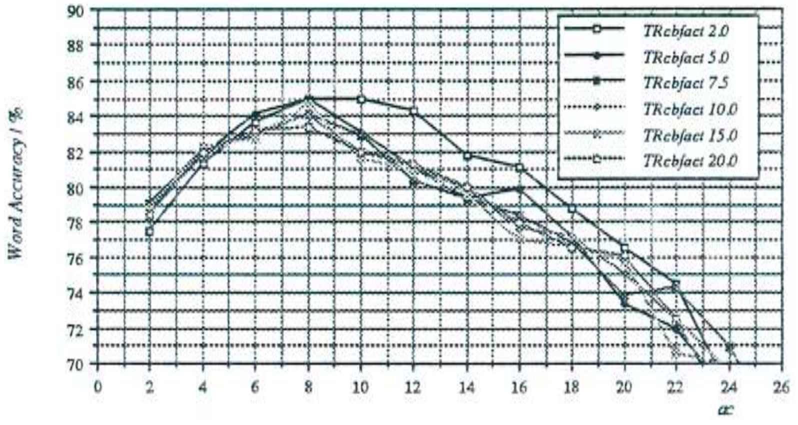
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.967	5.735	5.785	5.734	5.720	5.644
Q_{50}	15.860	-	-	-	-	-	13.825



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	77.5% (78.3%)	78.6% (79.3%)	79.1% (79.9%)	79.0% (79.7%)	78.8% (79.3%)	78.4% (79.1%)
4.0	81.3% (82.9%)	81.8% (83.1%)	82.0% (83.4%)	81.8% (83.2%)	82.2% (83.2%)	82.0% (83.1%)
6.0	83.6% (85.0%)	84.1% (85.9%)	82.9% (84.5%)	83.1% (84.7%)	82.7% (84.5%)	83.2% (85.6%)
8.0	85.0% (87.7%)	85.0% (87.9%)	84.1% (87.3%)	84.0% (86.6%)	84.7% (87.3%)	83.4% (86.1%)
10.0	85.0% (88.4%)	83.1% (86.6%)	82.9% (86.6%)	82.0% (85.7%)	81.6% (85.2%)	82.0% (85.7%)
12.0	84.3% (88.6%)	81.1% (84.3%)	80.4% (84.1%)	81.1% (84.3%)	80.9% (84.8%)	81.3% (85.6%)
14.0	81.8% (85.9%)	79.3% (82.7%)	79.3% (82.7%)	79.9% (83.2%)	79.5% (82.9%)	80.0% (84.0%)
16.0	81.1% (84.7%)	79.9% (84.0%)	78.4% (82.2%)	77.7% (81.8%)	77.0% (81.1%)	77.9% (82.7%)
18.0	78.8% (82.9%)	77.2% (81.5%)	76.8% (81.5%)	77.2% (81.6%)	76.5% (81.3%)	76.5% (81.6%)
20.0	76.5% (81.6%)	73.4% (79.5%)	73.8% (80.0%)	75.0% (80.9%)	75.8% (81.5%)	76.1% (81.8%)
22.0	74.5% (80.2%)	72.0% (76.1%)	74.3% (79.9%)	72.5% (79.1%)	70.6% (77.5%)	72.5% (79.1%)
24.0	68.6% (75.0%)	68.3% (75.0%)	70.9% (77.9%)	69.5% (76.8%)	69.9% (76.6%)	67.0% (73.1%)
26.0	65.6% (72.0%)	67.2% (74.5%)	66.0% (73.8%)	65.4% (73.6%)	66.7% (74.0%)	62.4% (70.1%)



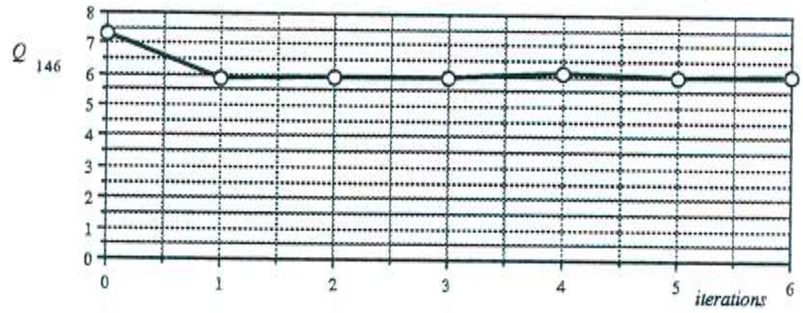
D.52 Experiment E.10.6

NLDA approach : $Y = AX + f(X)$

```
genaral parameters                                    backpropagation parameters
   dimension of orignial space :         32              number of discriminated targets :   2451
   dimension of image space :            16              learning rate :                     0.008
   discriminated classes in on top LDA : 146             momentum :                          0.9
target drift parameters                                  number of hidden units :            10
   α :                                   0.9             iterations :                        6
   β :                                   1.0             target update after :               6
   m :                                   1               sample selection :                  random
   number of drift vectors :             50 (one per phonem class)
```

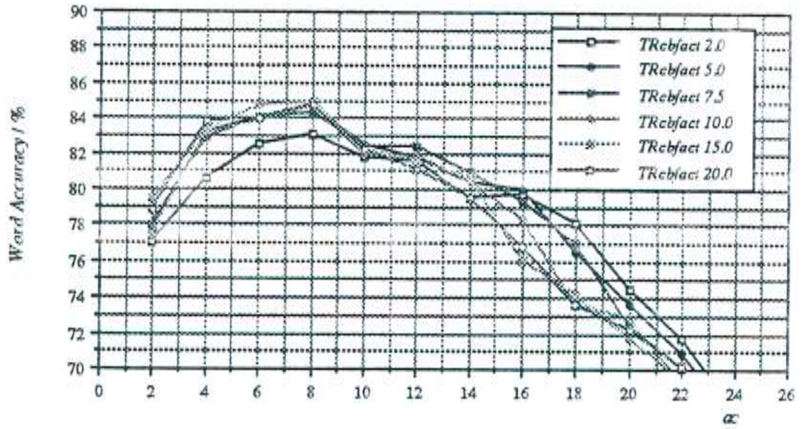
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.882	5.923	5.955	6.100	5.976	6.022
Q_{50}	15.860	-	-	-	-	-	13.783



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	77.0% (77.7%)	77.7% (78.4%)	77.9% (79.0%)	77.5% (78.4%)	79.3% (80.2%)	78.8% (79.5%)
4.0	80.6% (81.8%)	83.2% (84.1%)	82.9% (84.0%)	83.2% (84.1%)	83.4% (84.5%)	83.8% (84.8%)
6.0	82.5% (84.8%)	84.0% (85.6%)	84.0% (85.9%)	83.8% (85.6%)	84.7% (86.5%)	84.0% (85.7%)
8.0	83.1% (85.9%)	84.7% (86.8%)	84.3% (86.8%)	84.7% (86.8%)	85.0% (87.2%)	84.8% (87.2%)
10.0	81.8% (84.8%)	82.5% (86.5%)	82.4% (85.7%)	82.0% (85.4%)	82.2% (85.6%)	82.2% (85.7%)
12.0	81.6% (85.0%)	81.8% (85.7%)	82.4% (85.2%)	82.0% (84.8%)	81.6% (84.8%)	81.1% (84.5%)
14.0	79.5% (83.6%)	80.4% (84.0%)	80.9% (84.5%)	80.4% (84.7%)	80.7% (84.7%)	79.5% (83.8%)
16.0	79.7% (83.6%)	80.0% (84.3%)	79.3% (84.1%)	78.3% (83.4%)	76.1% (81.1%)	76.8% (82.2%)
18.0	78.1% (83.2%)	76.5% (82.0%)	77.0% (82.4%)	75.6% (81.1%)	74.2% (81.5%)	73.6% (80.9%)
20.0	74.5% (81.3%)	73.6% (80.4%)	72.4% (79.7%)	72.9% (80.4%)	71.8% (79.7%)	72.2% (79.3%)
22.0	71.8% (78.4%)	70.9% (79.1%)	70.1% (78.1%)	69.3% (78.6%)	69.3% (77.9%)	70.2% (77.5%)
24.0	67.6% (75.6%)	67.2% (75.8%)	67.7% (75.9%)	61.9% (71.5%)	63.6% (73.3%)	64.5% (73.1%)
26.0	63.3% (71.1%)	60.4% (70.8%)	58.6% (69.2%)	57.2% (67.7%)	57.0% (67.6%)	57.8% (66.8%)



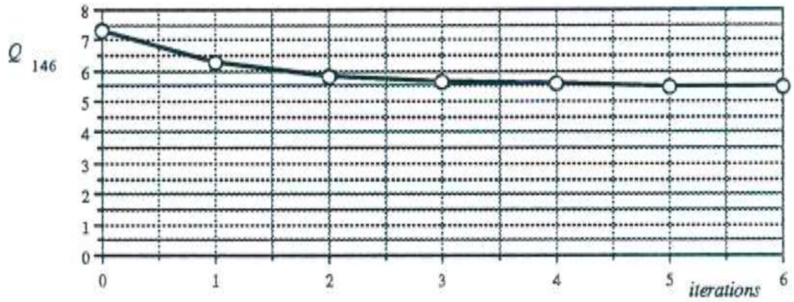
D.53 Experiment E.11.1

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

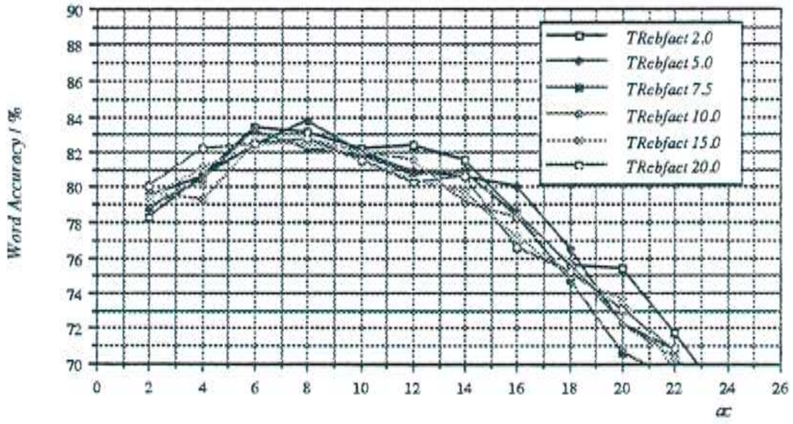
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.290	5.801	5.648	5.556	5.445	5.451
Q_{kn}	15.860	-	-	-	-	-	14.461



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	78.3% (79.5%)	79.5% (80.4%)	78.8% (79.7%)	79.7% (80.4%)	79.3% (79.7%)	80.0% (80.7%)
4.0	80.6% (82.2%)	80.6% (81.1%)	80.4% (81.6%)	79.3% (81.1%)	81.1% (82.4%)	82.2% (83.4%)
6.0	83.4% (85.7%)	82.4% (85.2%)	83.2% (85.9%)	82.4% (85.2%)	82.4% (85.2%)	82.5% (85.6%)
8.0	83.1% (86.6%)	83.8% (86.8%)	82.2% (85.6%)	82.7% (85.7%)	82.5% (85.9%)	82.7% (85.7%)
10.0	82.2% (86.1%)	82.0% (85.9%)	81.8% (85.6%)	82.0% (85.6%)	81.5% (85.7%)	81.5% (85.2%)
12.0	82.4% (86.5%)	80.9% (85.7%)	80.7% (85.4%)	81.5% (85.9%)	80.4% (85.9%)	80.2% (85.2%)
14.0	81.5% (85.4%)	80.6% (84.8%)	80.9% (85.2%)	79.1% (84.0%)	79.7% (84.8%)	80.6% (84.5%)
16.0	78.6% (82.9%)	80.0% (84.0%)	78.4% (83.8%)	78.3% (83.4%)	77.2% (83.4%)	76.6% (82.2%)
18.0	75.6% (81.6%)	76.5% (82.2%)	74.7% (80.6%)	75.8% (81.6%)	75.0% (81.3%)	75.2% (80.4%)
20.0	75.4% (81.5%)	72.2% (79.7%)	70.6% (77.2%)	72.2% (78.8%)	73.6% (79.7%)	73.1% (78.4%)
22.0	71.8% (78.6%)	70.8% (77.0%)	69.3% (75.9%)	70.4% (77.2%)	69.7% (76.1%)	70.8% (77.0%)
24.0	67.9% (74.5%)	64.5% (71.5%)	65.4% (72.0%)	67.2% (73.4%)	66.8% (73.3%)	64.7% (72.2%)
26.0	62.4% (69.5%)	60.1% (67.4%)	66.0% (72.4%)	65.1% (71.7%)	64.3% (70.9%)	62.2% (68.8%)



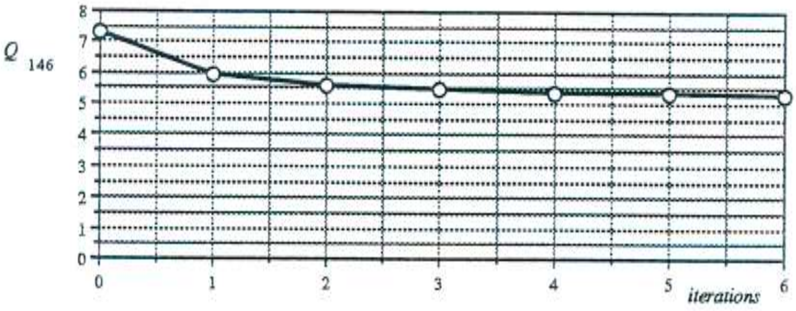
D.54 Experiment E.11.2

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

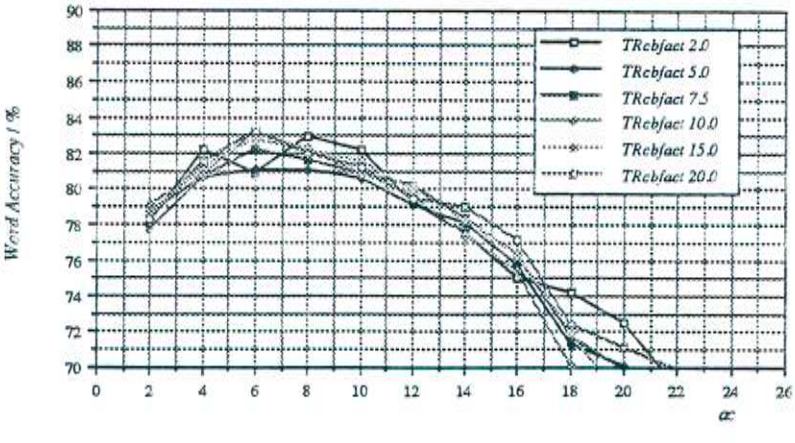
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.924	5.590	5.479	5.355	5.372	5.291
Q_{50}	15.860	-	-	-	-	-	13.972



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	78.1% (79.0%)	77.7% (79.1%)	78.6% (79.5%)	79.1% (79.9%)	78.6% (79.9%)	78.3% (79.1%)
4.0	82.2% (83.1%)	80.6% (82.2%)	80.7% (82.2%)	80.9% (82.5%)	81.3% (82.9%)	81.5% (83.1%)
6.0	80.9% (83.2%)	81.1% (83.6%)	82.2% (84.1%)	83.1% (85.4%)	82.7% (84.5%)	83.2% (85.6%)
8.0	82.9% (85.6%)	81.1% (84.5%)	81.6% (84.8%)	82.2% (85.4%)	82.2% (85.7%)	82.0% (85.2%)
10.0	82.2% (84.3%)	80.6% (83.6%)	80.9% (84.1%)	80.9% (84.1%)	81.5% (84.7%)	81.3% (84.5%)
12.0	79.3% (82.9%)	79.1% (83.4%)	80.2% (84.1%)	79.7% (83.8%)	80.0% (84.3%)	79.5% (84.3%)
14.0	77.5% (81.5%)	78.1% (83.2%)	78.4% (83.2%)	77.4% (82.7%)	78.3% (83.6%)	79.0% (84.3%)
16.0	75.0% (80.0%)	75.8% (82.0%)	76.3% (81.5%)	75.4% (80.7%)	76.3% (82.0%)	77.2% (82.5%)
18.0	74.2% (79.1%)	71.3% (78.1%)	71.8% (78.1%)	70.1% (76.8%)	72.0% (78.1%)	72.4% (79.0%)
20.0	72.5% (76.5%)	70.1% (77.2%)	69.9% (76.8%)	67.0% (74.0%)	67.7% (74.7%)	71.1% (77.2%)
22.0	69.0% (75.0%)	66.7% (73.4%)	67.0% (73.1%)	67.6% (73.6%)	67.2% (73.6%)	69.9% (75.4%)
24.0	66.1% (71.8%)	65.1% (71.1%)	67.7% (72.7%)	67.4% (72.0%)	64.7% (72.2%)	67.2% (72.9%)
26.0	62.6% (70.1%)	62.9% (67.9%)	61.0% (67.2%)	62.7% (67.9%)	63.1% (69.3%)	64.0% (69.9%)



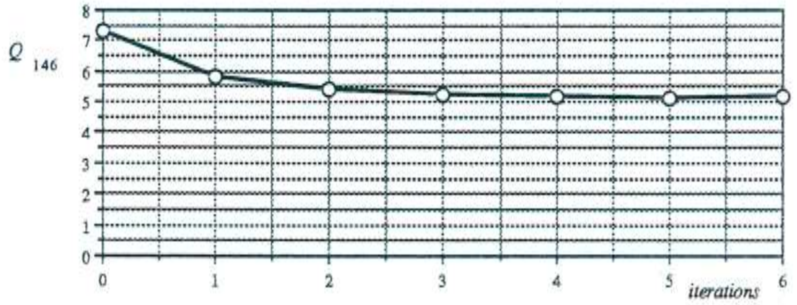
D.55 Experiment E.11.3

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

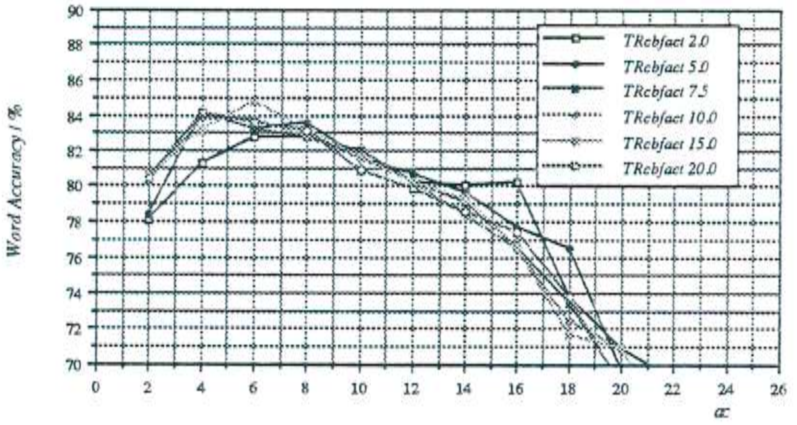
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.807	5.402	5.265	5.204	5.150	5.157
Q_{50}	15.860	-	-	-	-	-	13.652



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	76.1% (79.0%)	78.4% (79.7%)	80.4% (80.9%)	80.2% (80.7%)	80.6% (81.8%)	80.4% (81.1%)
4.0	81.3% (83.2%)	84.1% (85.2%)	84.1% (85.2%)	83.8% (85.0%)	83.1% (84.3%)	84.1% (85.6%)
6.0	82.7% (84.8%)	83.2% (85.9%)	83.2% (85.7%)	83.8% (85.9%)	84.8% (87.3%)	83.8% (86.1%)
8.0	82.7% (85.7%)	83.6% (87.3%)	82.7% (86.5%)	82.7% (86.5%)	83.1% (87.2%)	83.1% (87.0%)
10.0	81.6% (85.6%)	81.5% (85.7%)	82.0% (84.8%)	81.5% (85.4%)	81.6% (85.7%)	80.9% (85.7%)
12.0	80.2% (84.1%)	80.7% (85.4%)	80.2% (84.8%)	80.2% (84.8%)	80.2% (85.6%)	79.9% (85.0%)
14.0	80.0% (84.5%)	79.7% (84.8%)	79.1% (84.3%)	78.4% (84.1%)	79.3% (84.8%)	78.8% (84.1%)
16.0	80.2% (84.8%)	77.7% (83.1%)	76.6% (82.4%)	76.5% (82.2%)	76.6% (82.5%)	77.4% (83.1%)
18.0	73.8% (80.0%)	76.5% (82.4%)	73.4% (79.5%)	72.4% (78.6%)	71.7% (78.3%)	73.8% (80.0%)
20.0	70.9% (77.7%)	69.7% (76.5%)	69.0% (76.3%)	70.9% (77.4%)	70.9% (77.2%)	69.0% (76.6%)
22.0	69.3% (76.1%)	64.3% (72.5%)	66.5% (73.1%)	66.7% (73.8%)	67.0% (73.1%)	66.1% (73.3%)
24.0	61.7% (71.5%)	63.8% (69.9%)	63.3% (70.4%)	64.5% (71.7%)	62.9% (69.7%)	62.9% (70.6%)
26.0	61.3% (70.9%)	63.5% (69.3%)	64.0% (70.9%)	61.9% (69.9%)	62.0% (67.9%)	60.6% (67.4%)



D.56 Experiment E.11.4

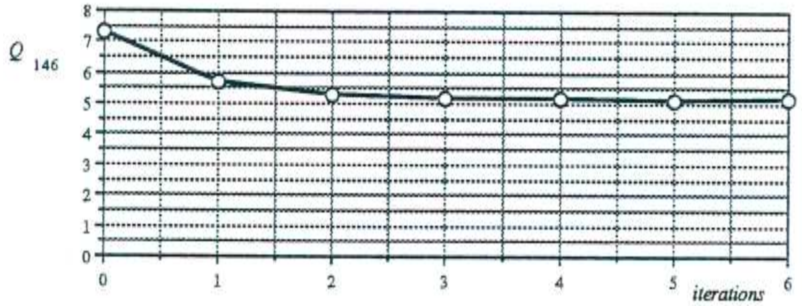
NLDA approach : $Y = AX + f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.3
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	2451
learning rate :	0.008
momentum :	0.9
number of hidden units :	20
iterations :	6
target update after :	6
sample selection :	random

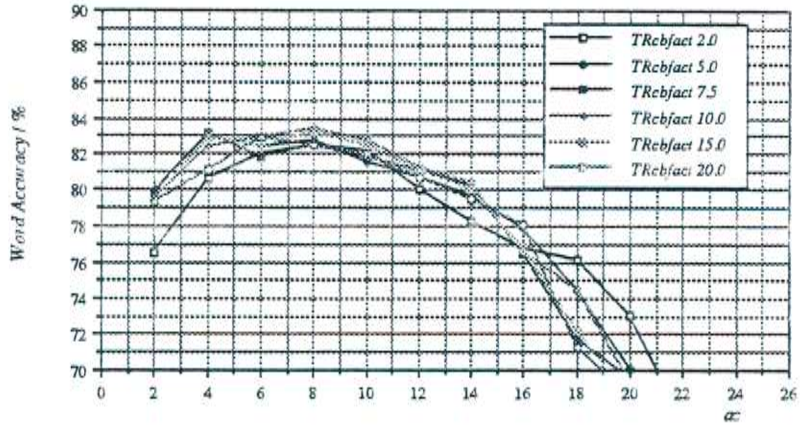
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.691	5.270	5.188	5.179	5.134	5.164
Q_{50}	15.860	-	-	-	-	-	13.494



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRebfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	76.5% (77.7%)	79.3% (80.2%)	79.9% (80.4%)	79.5% (80.0%)	79.3% (79.9%)	79.3% (79.9%)
4.0	80.6% (82.7%)	82.9% (84.7%)	83.1% (84.7%)	82.4% (84.3%)	82.9% (85.0%)	81.1% (83.2%)
6.0	82.0% (84.8%)	82.4% (85.7%)	81.8% (85.0%)	82.9% (86.1%)	82.4% (85.4%)	82.9% (85.7%)
8.0	82.5% (85.7%)	82.7% (87.0%)	82.5% (86.5%)	83.4% (87.3%)	83.2% (87.2%)	82.5% (86.3%)
10.0	82.2% (86.3%)	81.5% (87.0%)	81.8% (87.2%)	82.7% (87.3%)	82.5% (87.7%)	82.0% (87.0%)
12.0	80.0% (85.7%)	80.7% (85.6%)	81.1% (85.9%)	81.3% (85.9%)	81.1% (86.3%)	80.7% (85.7%)
14.0	78.3% (85.0%)	79.7% (85.0%)	80.2% (85.0%)	80.4% (85.2%)	80.0% (85.6%)	79.5% (84.5%)
16.0	76.8% (83.8%)	77.9% (83.6%)	76.5% (82.7%)	76.6% (82.9%)	76.8% (83.4%)	78.1% (83.8%)
18.0	76.1% (83.4%)	74.5% (81.3%)	71.8% (78.6%)	74.5% (81.3%)	72.2% (79.7%)	71.3% (79.0%)
20.0	73.1% (79.9%)	70.1% (76.5%)	69.5% (76.6%)	69.3% (75.9%)	68.4% (75.2%)	68.6% (75.2%)
22.0	67.0% (76.8%)	66.0% (73.6%)	61.3% (69.5%)	64.2% (71.3%)	63.8% (71.1%)	63.6% (71.7%)
24.0	63.1% (72.0%)	61.7% (70.4%)	59.9% (66.1%)	55.2% (67.7%)	59.5% (68.3%)	56.0% (64.7%)
26.0	56.5% (67.7%)	53.7% (64.0%)	52.9% (63.5%)	53.7% (63.8%)	55.4% (64.3%)	53.3% (63.5%)



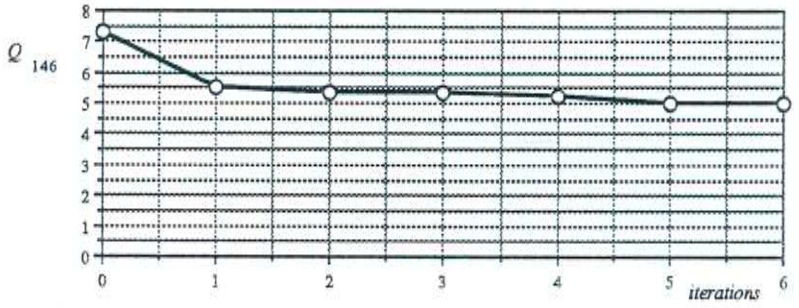
D.57 Experiment E.11.5

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	0.4	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

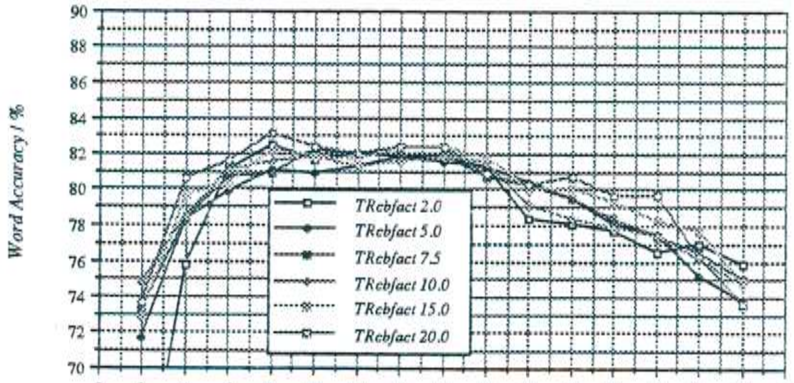
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.536	5.333	5.332	5.213	5.028	5.016
Q_{50}	15.860	-	-	-	-	-	13.127



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	60.1% (61.3%)	71.7% (72.7%)	73.3% (74.9%)	74.7% (75.8%)	72.9% (73.8%)	73.6% (74.3%)
4.0	75.8% (77.2%)	78.3% (79.3%)	78.4% (79.9%)	78.6% (80.0%)	79.7% (81.8%)	80.6% (81.6%)
6.0	81.1% (82.9%)	79.9% (81.6%)	80.7% (82.5%)	81.1% (82.9%)	80.2% (81.8%)	81.5% (83.2%)
8.0	82.5% (84.3%)	81.1% (82.9%)	80.9% (82.7%)	81.5% (83.4%)	82.2% (84.0%)	83.1% (85.0%)
10.0	81.6% (84.5%)	80.9% (83.6%)	82.2% (85.0%)	81.8% (85.0%)	81.8% (84.7%)	82.4% (85.0%)
12.0	82.0% (85.8%)	81.3% (84.1%)	82.0% (85.7%)	81.3% (85.0%)	81.6% (85.2%)	82.0% (85.4%)
14.0	81.8% (84.8%)	81.8% (85.4%)	82.0% (85.7%)	81.5% (85.2%)	82.4% (85.9%)	82.4% (85.9%)
16.0	81.8% (85.2%)	81.5% (85.4%)	82.0% (85.7%)	81.8% (85.7%)	82.4% (85.9%)	82.4% (86.1%)
18.0	81.3% (85.2%)	81.3% (85.7%)	80.7% (85.6%)	81.5% (85.7%)	81.8% (86.1%)	80.9% (85.4%)
20.0	78.4% (83.4%)	80.4% (85.2%)	80.2% (85.4%)	79.1% (84.7%)	80.0% (85.2%)	80.2% (85.4%)
22.0	78.1% (83.1%)	79.5% (84.7%)	79.5% (85.0%)	78.4% (83.8%)	80.0% (85.0%)	80.7% (85.6%)
24.0	77.7% (83.2%)	78.1% (83.2%)	78.4% (83.6%)	77.9% (84.0%)	79.3% (84.8%)	79.7% (84.5%)
26.0	76.5% (82.0%)	77.5% (83.1%)	77.4% (82.4%)	77.4% (82.9%)	78.3% (83.6%)	79.7% (84.7%)
28.0	77.0% (82.4%)	75.2% (81.8%)	76.5% (82.2%)	76.1% (82.5%)	77.7% (82.5%)	76.3% (81.6%)
30.0	75.9% (81.5%)	73.8% (80.2%)	75.0% (80.7%)	74.7% (80.9%)	74.9% (80.0%)	73.6% (79.7%)



D.58 Experiment E.11.6

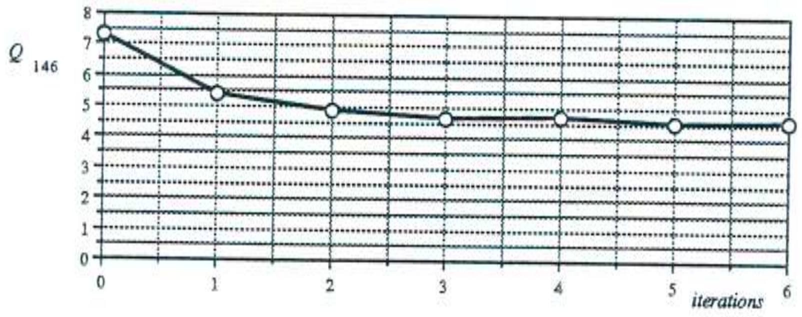
NLDA approach : $Y = AX + f(X)$

general parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	1.0
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	2451
learning rate :	0.008
momentum :	0.9
number of hidden units :	20
iterations :	6
target update after :	6
sample selection :	random

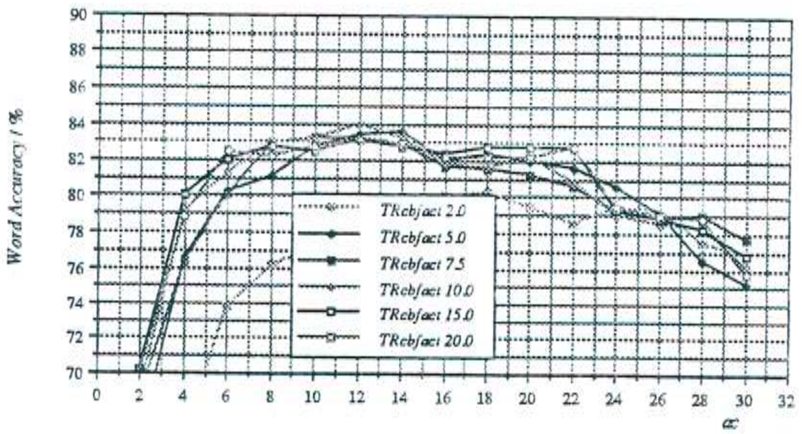
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.402	4.894	4.683	4.720	4.569	4.604
Q_{50}	15.860	-	-	-	-	-	11.941



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	47.7% (47.8%)	65.8% (66.3%)	68.6% (68.8%)	69.0% (69.9%)	70.2% (70.8%)	69.9% (70.2%)
4.0	65.8% (67.7%)	76.5% (77.5%)	76.3% (77.4%)	79.1% (79.9%)	80.0% (80.7%)	78.8% (79.5%)
6.0	73.8% (76.3%)	80.2% (81.5%)	80.0% (81.1%)	81.3% (82.5%)	82.0% (82.9%)	82.5% (83.4%)
8.0	76.1% (79.0%)	81.1% (82.5%)	82.9% (84.1%)	82.9% (84.1%)	82.7% (84.0%)	82.2% (83.6%)
10.0	77.2% (80.4%)	82.7% (84.1%)	83.2% (84.7%)	83.2% (84.5%)	82.5% (84.0%)	82.7% (84.3%)
12.0	77.7% (80.9%)	83.4% (84.8%)	83.2% (84.7%)	84.0% (85.2%)	83.1% (84.7%)	83.1% (84.8%)
14.0	79.0% (82.0%)	83.6% (85.2%)	82.9% (84.3%)	83.2% (84.8%)	82.7% (84.5%)	82.9% (84.8%)
16.0	80.0% (83.2%)	82.0% (85.0%)	81.8% (84.0%)	81.8% (84.5%)	82.4% (85.0%)	82.2% (84.8%)
18.0	80.2% (83.4%)	82.4% (85.4%)	81.5% (84.0%)	81.6% (84.7%)	82.7% (85.7%)	82.2% (85.0%)
20.0	79.5% (83.6%)	82.0% (85.4%)	81.3% (84.1%)	82.2% (85.2%)	82.7% (85.7%)	82.2% (85.2%)
22.0	78.6% (82.9%)	81.6% (84.7%)	80.6% (84.0%)	80.9% (84.5%)	82.7% (85.7%)	82.7% (85.9%)
24.0	79.5% (84.3%)	80.6% (84.0%)	79.3% (83.6%)	79.0% (83.8%)	79.3% (83.8%)	79.5% (84.0%)
26.0	79.3% (84.3%)	79.1% (83.4%)	78.8% (82.9%)	78.6% (83.8%)	78.8% (83.6%)	79.0% (83.8%)
28.0	77.5% (83.4%)	76.5% (81.6%)	79.0% (83.2%)	78.4% (83.2%)	78.3% (83.2%)	78.8% (83.8%)
30.0	76.8% (83.2%)	75.2% (80.4%)	77.7% (82.4%)	76.1% (81.5%)	76.8% (81.6%)	75.8% (80.6%)



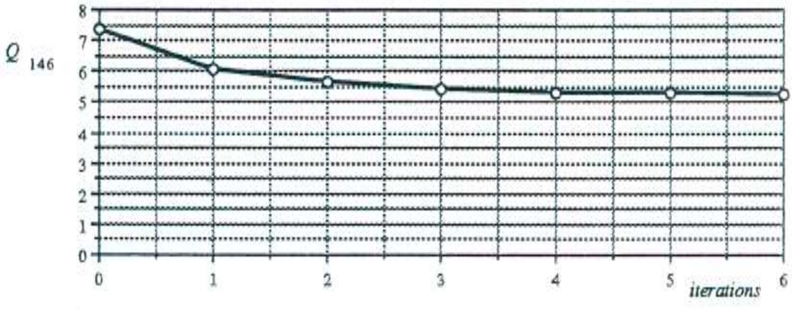
D.59 Experiment E.12.1

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	30
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

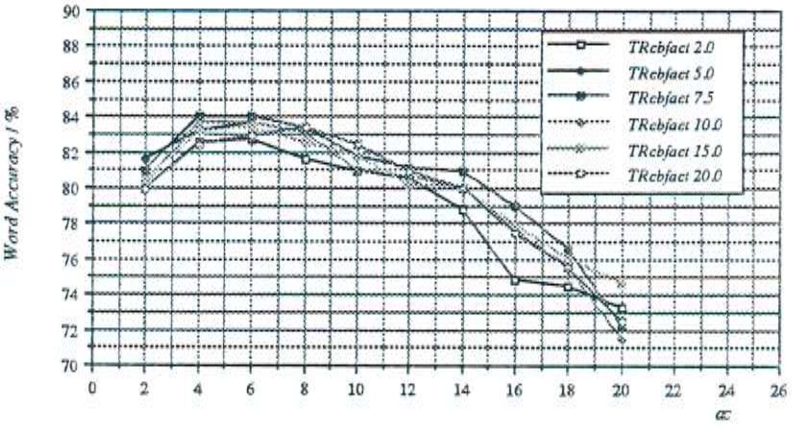
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	6.071	5.687	5.436	5.321	5.298	5.267
Q_{sn}	15.860	-	-	-	-	-	14.152



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	79.9% (80.7%)	81.6% (82.5%)	80.9% (82.0%)	80.2% (81.3%)	80.4% (81.3%)	79.9% (80.6%)
4.0	82.5% (83.8%)	83.2% (84.8%)	84.0% (85.4%)	83.6% (85.2%)	83.2% (85.0%)	82.4% (84.3%)
6.0	82.7% (85.2%)	83.6% (85.7%)	84.0% (86.1%)	83.6% (85.7%)	83.4% (85.4%)	82.9% (85.2%)
8.0	81.6% (85.2%)	83.1% (86.5%)	83.4% (87.0%)	83.1% (85.6%)	82.5% (85.6%)	83.4% (86.8%)
10.0	80.9% (84.3%)	80.9% (84.8%)	81.8% (85.7%)	80.9% (85.2%)	81.8% (84.8%)	82.4% (85.9%)
12.0	80.6% (84.7%)	80.7% (84.7%)	81.1% (85.0%)	80.7% (85.6%)	80.2% (84.2%)	80.9% (85.2%)
14.0	78.8% (83.4%)	80.0% (84.0%)	80.9% (84.7%)	80.0% (83.4%)	80.0% (84.0%)	80.0% (83.8%)
16.0	74.9% (80.9%)	77.7% (82.2%)	79.0% (83.4%)	77.7% (82.9%)	77.9% (82.4%)	77.5% (82.2%)
18.0	74.5% (80.2%)	75.6% (80.6%)	76.6% (81.3%)	75.6% (80.2%)	76.1% (81.5%)	75.6% (80.7%)
20.0	73.3% (78.6%)	71.5% (78.1%)	72.2% (78.4%)	71.5% (80.0%)	74.7% (80.0%)	72.5% (78.8%)



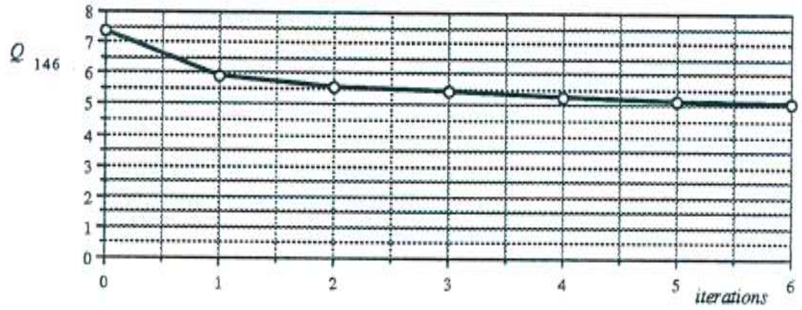
D.60 Experiment E.12.2

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	30
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

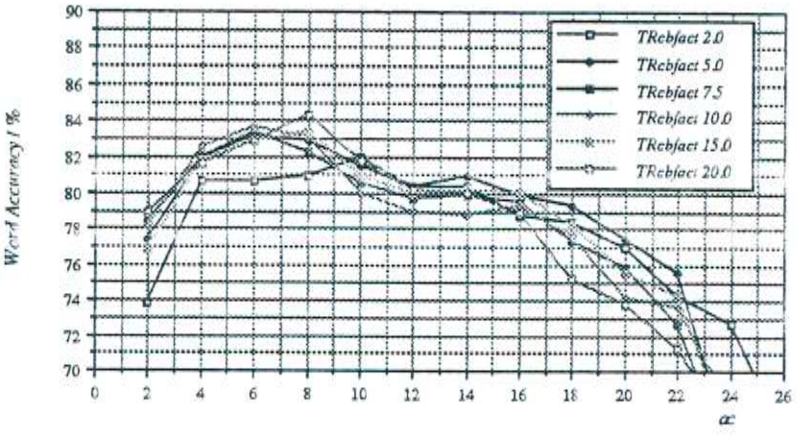
developement of class separability

iteration(s)	0	1	2	3	4	5	6
ϱ_{146}	7.329	5.881	5.550	5.409	5.255	5.128	5.068
ϱ_{en}	15.860	-	-	-	-	-	13.769



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	73.8% (75.6%)	77.4% (79.0%)	78.4% (79.5%)	78.1% (79.3%)	76.8% (78.4%)	79.0% (79.9%)
4.0	80.7% (82.5%)	82.0% (83.8%)	82.0% (83.8%)	82.5% (84.3%)	81.8% (84.0%)	81.6% (84.0%)
6.0	80.7% (84.0%)	83.2% (86.8%)	83.4% (86.3%)	83.6% (86.5%)	82.9% (85.7%)	82.9% (85.7%)
8.0	80.9% (84.7%)	82.9% (87.0%)	82.2% (85.9%)	83.1% (86.3%)	83.4% (86.8%)	84.3% (87.3%)
10.0	82.0% (86.5%)	81.5% (86.6%)	80.6% (85.7%)	80.0% (85.4%)	80.9% (85.6%)	81.8% (85.9%)
12.0	80.4% (85.7%)	80.4% (85.4%)	79.7% (84.3%)	79.0% (84.1%)	80.0% (85.0%)	80.2% (84.7%)
14.0	80.4% (85.6%)	80.9% (85.9%)	79.9% (84.7%)	78.8% (84.1%)	80.2% (85.0%)	80.0% (84.5%)
16.0	78.8% (84.1%)	79.9% (85.2%)	79.5% (84.8%)	79.3% (84.5%)	80.0% (85.6%)	79.0% (84.7%)
18.0	78.4% (83.2%)	79.3% (84.5%)	77.4% (82.9%)	77.7% (83.2%)	78.3% (83.6%)	75.2% (81.3%)
20.0	77.0% (82.9%)	77.5% (83.6%)	75.9% (82.5%)	74.2% (81.3%)	75.6% (81.6%)	73.8% (80.2%)
22.0	74.3% (80.2%)	75.6% (82.4%)	72.7% (78.1%)	73.6% (80.4%)	74.3% (80.7%)	71.3% (78.6%)
24.0	72.7% (78.4%)	65.8% (73.4%)	64.7% (72.2%)	68.1% (75.4%)	66.3% (73.3%)	66.5% (74.0%)
26.0	66.3% (74.9%)	63.5% (71.3%)	65.6% (71.8%)	68.1% (72.5%)	64.7% (71.8%)	64.5% (71.7%)



D.61 Experiment E.12.3

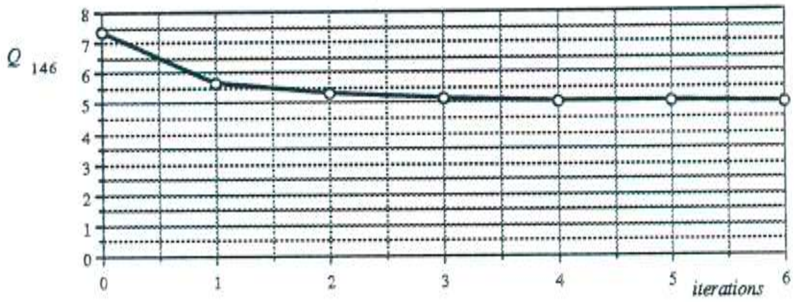
NLDA approach : $Y = AX + f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	0.2
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	2451
learning rate :	0.008
momentum :	0.9
number of hidden units :	30
iterations :	6
target update after :	6
sample selection :	random

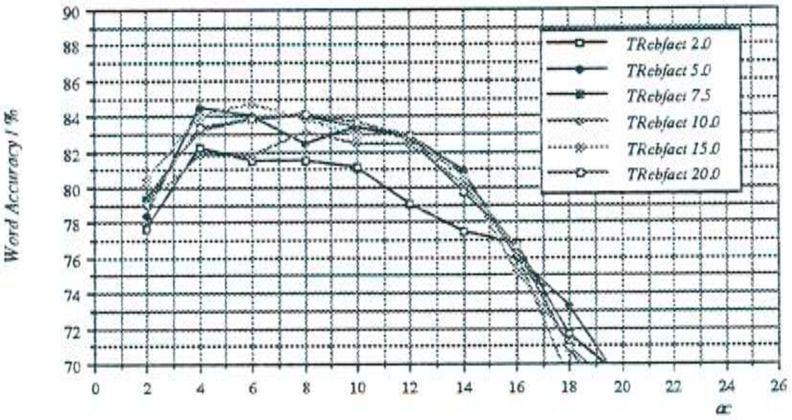
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.669	5.286	5.114	5.007	4.994	4.946
Q_{50}	15.860	-	-	-	-	-	13.856



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	77.7% (78.4%)	78.4% (79.7%)	79.3% (80.4%)	79.1% (80.2%)	80.4% (81.5%)	79.1% (80.2%)
4.0	82.2% (83.4%)	84.5% (85.7%)	83.2% (84.8%)	81.8% (83.6%)	84.0% (85.6%)	83.4% (84.7%)
6.0	81.5% (83.6%)	84.0% (86.3%)	83.8% (86.1%)	81.8% (84.3%)	84.7% (86.6%)	83.8% (86.3%)
8.0	81.5% (84.1%)	82.4% (85.7%)	84.0% (87.5%)	83.1% (87.0%)	83.8% (86.8%)	84.1% (87.3%)
10.0	81.1% (84.1%)	83.4% (87.5%)	83.4% (87.7%)	82.4% (86.6%)	82.9% (87.2%)	83.6% (88.2%)
12.0	79.1% (83.6%)	82.9% (87.5%)	82.7% (86.8%)	82.4% (86.6%)	82.9% (86.6%)	82.9% (87.3%)
14.0	77.5% (82.4%)	80.9% (85.4%)	80.7% (85.2%)	80.2% (85.2%)	80.6% (85.6%)	79.7% (84.8%)
16.0	76.8% (81.8%)	76.1% (82.0%)	76.5% (81.8%)	76.3% (82.0%)	75.2% (81.6%)	76.6% (82.2%)
18.0	71.7% (77.9%)	73.3% (79.5%)	70.9% (78.1%)	69.2% (76.8%)	71.1% (79.0%)	70.9% (78.3%)
20.0	69.2% (77.0%)	68.8% (76.5%)	66.3% (74.0%)	64.5% (72.7%)	66.1% (74.9%)	68.3% (77.9%)
22.0	69.2% (76.5%)	63.5% (71.7%)	63.8% (71.8%)	63.1% (71.1%)	62.6% (71.3%)	62.0% (69.9%)
24.0	66.8% (72.9%)	60.6% (68.4%)	62.2% (69.2%)	55.8% (64.7%)	60.8% (68.4%)	61.9% (69.5%)
26.0	60.8% (68.6%)	59.5% (66.7%)	57.9% (63.8%)	57.2% (63.5%)	56.5% (63.6%)	57.9% (64.9%)



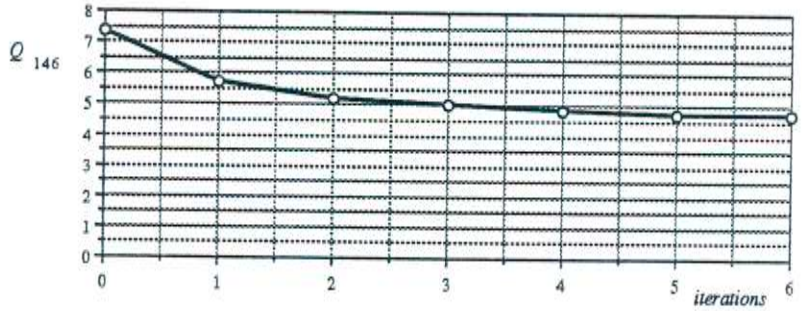
D.62 Experiment E.12.4

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	30
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	80 (one per phonem class)		

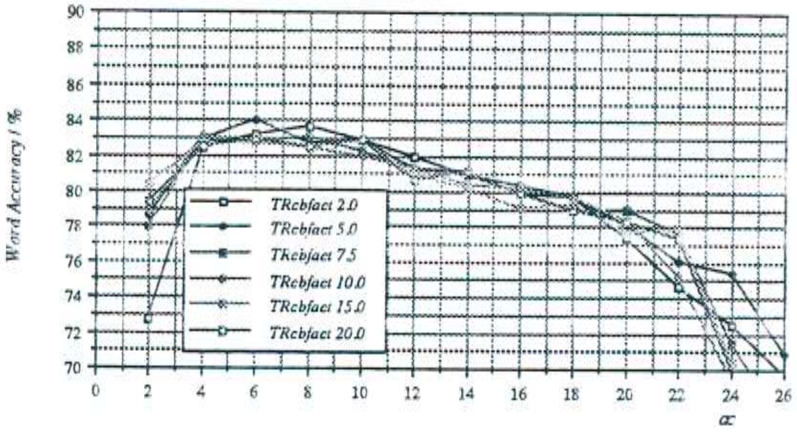
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.721	5.206	5.008	4.819	4.730	4.726
Q_{80}	15.860	-	-	-	-	-	13.106



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	72.7% (77.9%)	76.6% (79.7%)	77.9% (79.0%)	78.8% (79.3%)	80.4% (80.7%)	79.3% (79.7%)
4.0	82.4% (83.2%)	82.9% (83.8%)	82.9% (83.6%)	83.1% (83.6%)	82.7% (83.2%)	82.5% (83.1%)
6.0	83.2% (85.0%)	84.0% (85.7%)	82.7% (84.8%)	82.9% (84.8%)	82.7% (84.7%)	82.9% (84.8%)
8.0	83.6% (86.3%)	82.7% (85.4%)	83.1% (85.6%)	82.4% (84.8%)	82.4% (85.4%)	82.5% (85.0%)
10.0	82.9% (85.7%)	82.9% (86.1%)	82.2% (85.2%)	82.0% (85.0%)	82.9% (85.9%)	82.9% (85.9%)
12.0	82.0% (85.0%)	81.3% (84.8%)	81.3% (85.0%)	81.3% (85.2%)	80.9% (84.7%)	80.7% (84.5%)
14.0	80.9% (84.7%)	81.1% (85.7%)	80.9% (85.6%)	80.4% (85.2%)	80.2% (85.4%)	81.1% (85.4%)
16.0	80.4% (85.4%)	79.9% (85.0%)	80.0% (85.2%)	80.2% (85.4%)	79.1% (84.1%)	79.9% (84.8%)
18.0	79.7% (84.7%)	79.7% (84.7%)	79.0% (84.1%)	79.7% (84.5%)	79.3% (84.1%)	79.0% (83.8%)
20.0	77.4% (83.6%)	78.1% (84.3%)	79.0% (84.5%)	78.4% (83.8%)	77.7% (83.4%)	78.3% (83.2%)
22.0	74.7% (82.2%)	76.1% (83.1%)	77.7% (84.1%)	75.2% (81.8%)	77.9% (83.8%)	77.4% (83.2%)
24.0	72.5% (80.6%)	75.4% (82.4%)	71.5% (79.5%)	69.5% (77.4%)	70.4% (78.1%)	69.9% (77.7%)
26.0	69.7% (78.3%)	70.9% (78.3%)	66.8% (74.7%)	66.8% (75.0%)	64.5% (72.9%)	66.1% (73.4%)



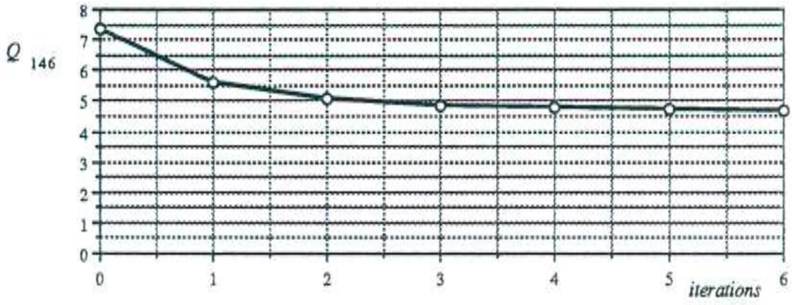
D.63 Experiment E.12.5

NLDA approach : $Y = AX + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	2451
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	146	momentum :	0.9
target drift parameters		number of hidden units :	30
α :	0.9	iterations :	6
β :	0.4	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per phonem class)		

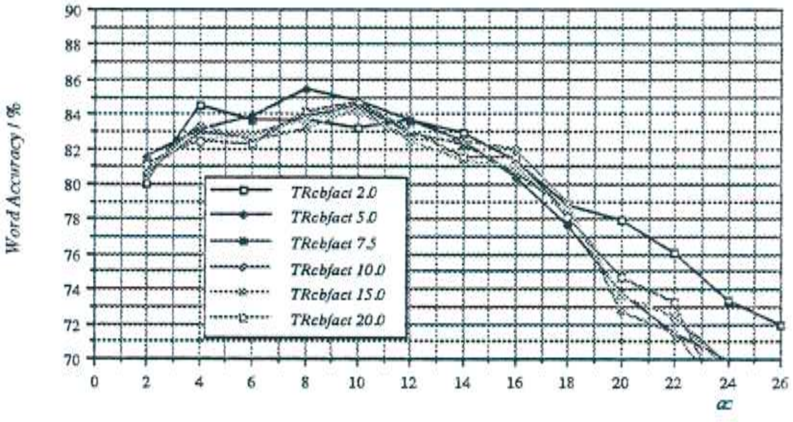
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.588	5.104	4.823	4.786	4.757	4.685
Q_{kn}	15.860	-	-	-	-	-	12.713



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	80.0% (80.0%)	81.5% (82.0%)	80.6% (81.3%)	80.7% (81.6%)	81.1% (82.0%)	81.1% (82.0%)
4.0	84.5% (85.6%)	83.1% (84.1%)	82.9% (84.1%)	83.4% (84.7%)	82.5% (83.8%)	82.4% (83.6%)
6.0	83.6% (85.7%)	83.8% (85.7%)	82.7% (85.0%)	82.4% (84.7%)	82.7% (84.8%)	82.2% (84.5%)
8.0	83.6% (86.3%)	85.4% (88.4%)	83.8% (87.0%)	83.1% (86.1%)	83.6% (86.5%)	84.1% (87.5%)
10.0	83.2% (86.6%)	84.8% (88.4%)	84.5% (88.1%)	84.3% (87.5%)	84.1% (87.5%)	84.7% (88.2%)
12.0	83.6% (87.5%)	83.6% (88.4%)	82.9% (88.4%)	82.5% (87.7%)	82.4% (87.5%)	83.1% (88.2%)
14.0	82.9% (87.2%)	82.4% (87.5%)	82.2% (87.3%)	82.4% (87.2%)	81.3% (86.5%)	81.5% (86.3%)
16.0	81.3% (85.9%)	80.4% (85.2%)	80.7% (85.7%)	82.0% (86.8%)	80.9% (86.3%)	81.5% (86.8%)
18.0	78.8% (84.7%)	77.7% (83.2%)	78.6% (83.6%)	79.0% (84.0%)	78.6% (83.6%)	78.1% (83.6%)
20.0	77.9% (83.4%)	73.8% (80.7%)	73.6% (79.5%)	72.7% (79.5%)	73.6% (79.9%)	74.7% (81.1%)
22.0	76.1% (82.2%)	71.5% (77.5%)	72.4% (78.6%)	71.5% (77.7%)	72.4% (77.9%)	73.3% (78.6%)
24.0	73.4% (79.5%)	69.9% (74.7%)	69.7% (76.8%)	67.9% (74.3%)	69.2% (75.0%)	66.8% (72.7%)
26.0	72.0% (77.9%)	67.0% (73.3%)	67.4% (73.3%)	65.4% (72.7%)	62.2% (69.7%)	65.8% (71.8%)



D.64 Experiment E.12.6

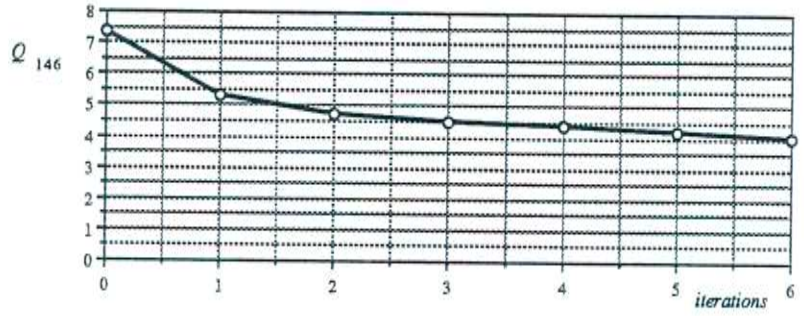
NLDA approach : $Y = AX + f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	146
target drift parameters	
α :	0.9
β :	1.0
m :	1
number of drift vectors :	50 (one per phonem class)

backpropagation parameters	
number of discriminated targets :	2451
learning rate :	0.008
momentum :	0.9
number of hidden units :	30
iterations :	6
target update after :	6
sample selection :	random

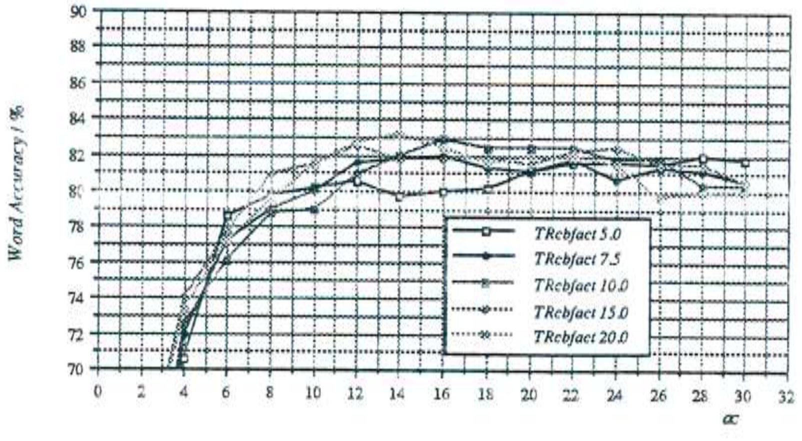
developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{146}	7.329	5.317	4.718	4.479	4.366	4.255	4.050
Q_{50}	15.860	-	-	-	-	-	11.416



word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	5.0	7.5	10.0	15.0	20.0
2.0	49.7% (50.4%)	57.2% (57.9%)	58.1% (59.0%)	62.7% (63.6%)	62.4% (63.3%)
4.0	70.6% (71.3%)	72.0% (72.9%)	72.5% (73.6%)	74.0% (75.0%)	73.4% (74.5%)
6.0	78.6% (79.1%)	77.2% (78.1%)	76.1% (77.2%)	77.7% (78.4%)	77.2% (77.9%)
8.0	79.7% (80.0%)	79.0% (80.0%)	78.8% (80.2%)	80.9% (81.6%)	79.5% (80.4%)
10.0	80.2% (81.3%)	80.0% (81.1%)	79.0% (80.0%)	81.5% (82.5%)	81.5% (82.7%)
12.0	80.6% (81.8%)	81.6% (82.7%)	80.9% (82.2%)	82.5% (84.0%)	82.7% (84.3%)
14.0	79.7% (81.5%)	81.8% (84.0%)	82.0% (84.0%)	81.8% (83.8%)	83.2% (85.2%)
16.0	80.0% (82.0%)	82.0% (84.7%)	82.9% (85.2%)	81.8% (84.0%)	82.4% (84.8%)
18.0	80.2% (82.7%)	81.3% (84.0%)	82.4% (85.0%)	82.0% (84.7%)	81.6% (84.7%)
20.0	81.1% (84.1%)	81.1% (84.1%)	82.4% (85.2%)	81.5% (84.5%)	82.0% (85.0%)
22.0	81.6% (84.5%)	81.8% (84.8%)	82.4% (85.4%)	82.2% (85.0%)	82.0% (85.0%)
24.0	81.6% (84.5%)	80.7% (84.5%)	81.8% (85.4%)	82.4% (85.9%)	81.5% (85.2%)
26.0	81.5% (84.3%)	81.3% (85.0%)	81.8% (85.4%)	81.5% (85.7%)	79.7% (83.8%)
28.0	82.0% (84.8%)	81.1% (84.8%)	80.4% (84.3%)	81.5% (85.7%)	79.9% (84.1%)
30.0	81.8% (84.7%)	80.6% (84.7%)	80.4% (84.3%)	80.6% (84.8%)	79.9% (84.1%)



Appendix E

Experiments on Resource Management Database (Context Dependent)

Resource Management Database :

- speakerindependent database
- 109 male and female speakers of different american dialects
- 4360 sentences for training
- data sampled at $16kHz$ with $16bit$ quantisation

primary transformation :

- 256-point FFT with Hamming Window, window length $16ms$
- $6ms$ window overlapp
- dimension reduction to 16 melscale coefficients

phonetic modeling :

- 2374 triphones, each splited in 3 subphonemes
- 1 silence class

training :

- training set: 2830 sentences form 78 male speakers
- 6 iterations over the training set, codebook and distribution updates after each iteration

test set :

- 48 sentences form 12 male speakers
- neither the speakers nor the sentences are part of the training material
- word pair grammar, perplexity 60

E.1 Experiment with LDA Feature Vectors

genaral parameters

dimension of orignial space :	32
dimension of image space :	16
discriminated classes in LDA :	7124 (subtriphones)

developement of class separability

	primary feature vectors (32 cofficients)	LDA derived feature vectors (16 coefficients)
Q_{7124}	22.807	3.017
Q_{146}	36.461	7.734
Q_{50}	58.341	16.153

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	1.0	2.0	4.0	6.0	8.0
2.0	88.8% (90.9%)	89.7% (91.3%)	88.6% (90.7%)	88.6% (90.6%)	88.6% (90.9%)
4.0	90.0% (92.3%)	90.4% (92.5%)	89.5% (92.3%)	89.5% (92.2%)	90.0% (92.3%)
6.0	90.2% (92.9%)	90.9% (93.0%)	90.0% (92.7%)	90.2% (92.5%)	90.4% (92.7%)
8.0	89.7% (92.3%)	90.9% (93.6%)	90.9% (93.8%)	90.0% (92.9%)	89.8% (92.5%)
10.0	89.8% (92.9%)	90.7% (93.4%)	90.9% (93.8%)	90.0% (93.2%)	90.2% (93.4%)
12.0	88.4% (92.0%)	89.8% (92.9%)	90.2% (93.2%)	90.0% (93.0%)	89.7% (92.2%)
14.0	87.5% (91.4%)	88.8% (92.5%)	88.9% (91.8%)	87.7% (90.9%)	88.6% (91.6%)
16.0	87.2% (91.6%)	88.6% (92.5%)	88.4% (91.8%)	87.3% (91.1%)	87.3% (90.4%)
18.0	86.5% (91.4%)	85.9% (90.2%)	87.2% (90.6%)	87.2% (90.6%)	86.3% (89.8%)
20.0	82.5% (88.4%)	85.4% (89.7%)	86.6% (90.0%)	83.4% (88.2%)	83.1% (87.7%)

E.2 Experiment CD.1.1

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	1.381	0.791	0.489	0.277	0.126	0.017
Q_{146}	7.734	-	-	-	-	-	4.463
Q_{50}	16.153	-	-	-	-	-	13.25

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	86.3% (88.4%)	86.1% (88.8%)	86.3% (88.1%)	86.8% (88.4%)	86.6% (88.6%)	86.1% (87.7%)
4.0	88.8% (91.1%)	89.1% (91.6%)	88.4% (90.9%)	88.4% (90.9%)	87.9% (90.6%)	87.3% (90.0%)
6.0	88.9% (91.8%)	89.1% (91.6%)	90.2% (92.2%)	89.3% (91.4%)	89.1% (91.6%)	89.3% (91.6%)
8.0	88.9% (92.0%)	89.7% (92.0%)	89.5% (91.8%)	89.7% (92.0%)	89.5% (92.0%)	89.5% (92.0%)
10.0	89.1% (92.2%)	89.7% (92.2%)	89.3% (91.8%)	89.3% (91.6%)	89.7% (92.2%)	89.7% (92.2%)
12.0	88.4% (91.8%)	90.2% (92.7%)	89.3% (91.8%)	89.7% (92.0%)	89.7% (92.0%)	89.3% (92.8%)
14.0	88.8% (91.6%)	89.5% (92.0%)	89.1% (91.8%)	89.1% (92.0%)	89.5% (92.2%)	89.3% (92.2%)
16.0	88.9% (92.0%)	88.9% (92.2%)	88.6% (91.4%)	88.4% (91.4%)	88.4% (91.6%)	88.8% (91.6%)
18.0	88.8% (91.8%)	88.4% (92.2%)	88.2% (91.3%)	87.9% (90.7%)	88.4% (91.1%)	88.2% (90.9%)
20.0	88.8% (92.0%)	87.2% (91.1%)	87.2% (90.9%)	87.2% (90.9%)	87.9% (91.1%)	88.1% (91.1%)

E.3 Experiment CD.1.2

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	1.119	0.595	0.333	0.150	0.026	-0.091
Q_{146}	7.734	-	-	-	-	-	4.375
Q_{50}	16.153	-	-	-	-	-	12.863

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact				
ac	2.0	5.0	7.5	10.0	15.0
2.0	81.5% (84.5%)	83.6% (86.1%)	83.8% (86.3%)	84.1% (86.8%)	84.3% (86.8%)
4.0	88.5% (90.2%)	87.0% (89.5%)	87.0% (89.7%)	87.9% (90.2%)	87.0% (89.5%)
6.0	90.0% (91.6%)	88.8% (91.3%)	88.9% (90.7%)	88.8% (90.4%)	89.5% (91.3%)
8.0	90.9% (92.2%)	88.8% (91.3%)	88.8% (90.9%)	88.8% (90.9%)	89.5% (91.8%)
10.0	90.6% (92.2%)	89.1% (91.8%)	88.6% (91.3%)	88.8% (91.3%)	89.3% (91.8%)
12.0	89.5% (92.0%)	88.9% (91.8%)	88.4% (91.4%)	89.1% (91.8%)	89.1% (91.6%)
14.0	90.0% (92.9%)	88.4% (91.6%)	88.4% (91.6%)	88.2% (91.4%)	88.6% (91.4%)
16.0	89.1% (92.9%)	88.2% (91.6%)	88.2% (91.6%)	88.1% (91.3%)	87.9% (90.9%)
18.0	88.1% (92.3%)	88.4% (91.8%)	88.2% (92.0%)	87.1% (91.1%)	87.5% (91.3%)
20.0	86.8% (91.4%)	88.4% (91.8%)	88.2% (92.0%)	87.2% (91.3%)	87.5% (91.3%)

E.4 Experiment CD.1.3

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	0.878	0.430	0.193	0.028	-0.072	-0.152
Q_{146}	7.734	-	-	-	-	-	4.362
Q_{50}	16.153	-	-	-	-	-	12.545

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	80.0% (84.0%)	85.6% (86.6%)	85.4% (86.8%)	85.4% (86.6%)	85.4% (86.6%)	85.7% (87.2%)
4.0	87.3% (90.0%)	88.1% (89.1%)	88.2% (89.7%)	88.2% (89.7%)	89.1% (90.6%)	87.7% (89.3%)
6.0	88.2% (91.1%)	89.3% (90.6%)	89.7% (91.1%)	89.1% (90.7%)	89.7% (91.4%)	88.9% (90.7%)
8.0	89.3% (92.5%)	90.0% (91.6%)	89.7% (91.6%)	89.3% (91.3%)	90.6% (92.3%)	89.5% (91.4%)
10.0	88.6% (91.6%)	89.5% (91.4%)	89.7% (91.8%)	89.1% (91.3%)	89.5% (91.6%)	88.8% (91.1%)
12.0	88.9% (92.0%)	89.1% (91.3%)	89.3% (91.8%)	88.9% (91.4%)	88.8% (91.4%)	88.4% (91.3%)
14.0	88.6% (92.0%)	88.9% (91.6%)	88.8% (91.3%)	88.8% (91.4%)	88.2% (91.1%)	88.1% (90.9%)
16.0	88.8% (92.2%)	88.9% (91.6%)	87.9% (90.9%)	86.5% (90.4%)	85.7% (90.0%)	86.5% (90.2%)
18.0	88.6% (92.3%)	87.5% (91.3%)	86.3% (90.4%)	86.3% (90.4%)	85.7% (90.0%)	85.9% (90.0%)
20.0	87.7% (92.0%)	86.8% (91.1%)	85.4% (90.2%)	85.7% (90.0%)	85.0% (89.8%)	86.1% (90.2%)

E.5 Experiment CD.1.4

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	0.696	0.280	0.080	-0.050	-0.115	-0.184
Q_{146}	7.734	-	-	-	-	-	4.391
Q_{50}	16.153	-	-	-	-	-	12.376

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	79.0% (82.5%)	84.5% (87.5%)	85.9% (88.1%)	85.9% (87.9%)	84.5% (86.6%)	84.3% (86.5%)
4.0	82.5% (86.3%)	86.8% (89.7%)	87.3% (89.3%)	87.3% (89.3%)	87.3% (89.7%)	87.5% (90.0%)
6.0	86.6% (89.5%)	88.4% (91.6%)	88.9% (91.3%)	89.1% (91.6%)	88.6% (90.9%)	88.4% (91.1%)
8.0	86.8% (89.8%)	88.6% (91.6%)	89.3% (91.6%)	89.5% (92.2%)	88.9% (91.6%)	89.1% (92.0%)
10.0	86.8% (89.5%)	88.4% (91.8%)	88.2% (91.3%)	89.5% (92.2%)	88.6% (91.6%)	88.6% (91.8%)
12.0	87.5% (90.2%)	87.5% (91.1%)	87.9% (91.3%)	89.1% (92.2%)	88.6% (91.8%)	88.1% (91.4%)
14.0	87.2% (90.4%)	86.8% (90.7%)	87.9% (91.3%)	88.6% (92.2%)	87.9% (91.4%)	88.2% (91.6%)
16.0	86.6% (90.2%)	87.0% (90.9%)	87.7% (91.4%)	88.2% (92.0%)	87.2% (91.1%)	86.8% (90.7%)
18.0	86.3% (90.2%)	84.5% (89.8%)	85.6% (90.4%)	87.7% (91.6%)	87.0% (90.7%)	86.8% (90.9%)
20.0	85.7% (90.6%)	84.3% (89.7%)	84.8% (90.2%)	86.6% (91.6%)	87.2% (91.1%)	86.8% (91.1%)

E.6 Experiment CD.1.5

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.4	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	0.547	0.198	0.028	-0.098	-0.170	-0.239
Q_{146}	7.734	-	-	-	-	-	4.374
Q_{50}	16.153	-	-	-	-	-	12.390

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	81.3% (84.5%)	85.9% (88.6%)	86.3% (88.4%)	85.9% (81.4%)	85.4% (87.9%)	84.8% (86.8%)
4.0	87.2% (89.1%)	88.1% (90.4%)	87.9% (90.2%)	87.0% (89.7%)	87.2% (90.2%)	87.3% (90.4%)
6.0	89.3% (91.6%)	88.1% (90.6%)	88.8% (91.1%)	87.2% (89.8%)	88.4% (91.3%)	88.1% (90.7%)
8.0	88.2% (91.1%)	88.4% (91.1%)	88.8% (90.7%)	88.4% (90.7%)	88.8% (91.3%)	89.1% (91.4%)
10.0	88.4% (91.1%)	88.2% (91.4%)	88.4% (90.7%)	88.9% (91.3%)	88.4% (91.1%)	88.4% (91.1%)
12.0	88.1% (91.3%)	88.2% (91.6%)	89.5% (92.2%)	88.6% (91.4%)	88.2% (91.1%)	88.2% (91.1%)
14.0	87.9% (91.3%)	89.5% (92.2%)	88.8% (92.0%)	88.6% (91.8%)	88.1% (91.6%)	87.0% (90.7%)
16.0	86.6% (90.7%)	87.9% (92.0%)	86.5% (90.6%)	87.3% (91.3%)	87.9% (91.4%)	87.0% (90.9%)
18.0	87.0% (90.9%)	87.0% (91.3%)	85.7% (90.2%)	85.2% (90.2%)	85.6% (90.2%)	85.6% (90.6%)
20.0	85.6% (90.4%)	86.6% (91.1%)	85.2% (90.4%)	85.0% (90.2%)	85.0% (90.2%)	85.0% (90.4%)

E.7 Experiment CD.1.6

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	1.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	0.138	-0.121	-0.208	-0.212	-0.235	-0.232
Q_{146}	7.734	-	-	-	-	-	4.494
Q_{50}	16.153	-	-	-	-	-	11.443

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	66.8% (69.9%)	77.4% (81.1%)	77.9% (80.4%)	77.7% (80.4%)	81.1% (83.8%)	79.3% (81.8%)
4.0	78.6% (82.2%)	83.1% (86.6%)	86.5% (88.8%)	85.7% (88.2%)	85.9% (88.1%)	85.7% (88.1%)
6.0	81.6% (84.8%)	87.5% (90.2%)	87.7% (90.0%)	86.8% (89.1%)	87.3% (89.7%)	86.3% (88.9%)
8.0	83.6% (86.5%)	88.2% (90.7%)	87.5% (89.8%)	88.4% (90.4%)	88.4% (91.3%)	87.9% (90.7%)
10.0	84.8% (87.7%)	88.1% (90.6%)	87.2% (89.5%)	88.8% (91.4%)	89.3% (91.3%)	88.9% (91.8%)
12.0	85.6% (88.2%)	88.2% (90.7%)	88.2% (90.7%)	89.5% (92.2%)	89.3% (92.2%)	88.9% (91.8%)
14.0	84.1% (87.7%)	86.4% (90.7%)	88.1% (90.7%)	89.1% (92.2%)	89.3% (92.3%)	88.6% (91.6%)
16.0	83.8% (87.7%)	88.1% (90.2%)	88.1% (90.7%)	89.1% (92.0%)	89.3% (92.3%)	87.9% (91.3%)
18.0	83.4% (87.7%)	88.2% (90.4%)	87.7% (90.4%)	88.9% (91.8%)	89.1% (91.8%)	88.1% (91.4%)
20.0	82.2% (87.0%)	88.1% (90.6%)	87.2% (90.4%)	89.1% (92.0%)	88.1% (91.3%)	87.7% (91.4%)

E.8 Experiment CD.2.1

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per monophone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	1.375	0.785	0.479	0.269	0.130	0.011
Q_{146}	7.734	-	-	-	-	-	4.457
Q_{50}	16.153	-	-	-	-	-	13.255

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	83.2% (84.8%)	84.5% (87.0%)	84.0% (86.5%)	84.7% (87.3%)	85.2% (87.3%)	83.8% (86.3%)
4.0	88.9% (90.9%)	87.0% (89.1%)	88.1% (90.7%)	87.9% (90.7%)	87.3% (90.2%)	87.7% (90.0%)
6.0	90.6% (92.2%)	88.4% (90.7%)	88.2% (90.9%)	88.1% (90.7%)	88.2% (90.7%)	88.1% (90.7%)
8.0	90.0% (92.0%)	88.2% (90.7%)	89.1% (91.6%)	87.7% (90.6%)	88.9% (91.6%)	88.2% (91.1%)
10.0	90.0% (92.5%)	88.6% (91.3%)	88.8% (91.6%)	88.1% (91.3%)	88.6% (91.3%)	89.3% (91.6%)
12.0	88.4% (91.3%)	88.2% (91.4%)	87.9% (91.1%)	88.6% (91.6%)	89.3% (92.0%)	89.3% (92.0%)
14.0	87.7% (91.3%)	88.6% (91.6%)	87.7% (91.3%)	88.4% (91.6%)	88.9% (92.0%)	88.9% (92.0%)
16.0	87.7% (91.1%)	88.2% (91.6%)	87.7% (91.4%)	88.1% (91.3%)	88.8% (91.8%)	88.1% (91.6%)
18.0	87.7% (91.3%)	88.2% (91.4%)	87.9% (91.4%)	87.9% (91.1%)	88.4% (92.0%)	88.1% (91.6%)
20.0	85.4% (90.0%)	87.3% (91.4%)	87.5% (91.4%)	87.3% (91.1%)	88.2% (91.8%)	87.5% (91.4%)

E.9 Experiment CD.2.2

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per monophone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	0.916	0.445	0.201	0.051	-0.046	-0.132
Q_{146}	7.734	-	-	-	-	-	3.951
Q_{50}	16.153	-	-	-	-	-	11.756

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	84.3% (86.3%)	87.5% (89.3%)	87.0% (89.1%)	86.6% (88.1%)	85.9% (86.1%)	86.5% (87.9%)
4.0	88.8% (90.4%)	89.8% (90.9%)	89.1% (91.3%)	89.7% (91.4%)	89.8% (91.3%)	89.7% (90.9%)
6.0	90.2% (92.0%)	89.7% (92.0%)	90.7% (92.5%)	91.1% (92.7%)	90.6% (92.3%)	90.7% (92.3%)
8.0	90.0% (92.0%)	90.9% (92.7%)	90.4% (92.2%)	90.7% (92.7%)	90.2% (92.3%)	90.2% (92.2%)
10.0	90.4% (92.7%)	90.4% (92.7%)	90.9% (93.0%)	90.7% (92.9%)	90.4% (92.5%)	90.2% (92.5%)
12.0	89.8% (92.3%)	90.4% (92.9%)	90.7% (93.0%)	90.2% (92.3%)	89.3% (92.0%)	89.5% (92.0%)
14.0	89.5% (92.0%)	89.1% (92.0%)	89.7% (92.3%)	98.1% (91.8%)	88.8% (91.4%)	89.5% (92.2%)
16.0	89.1% (92.0%)	88.9% (92.0%)	89.3% (92.3%)	89.1% (92.0%)	88.1% (91.3%)	88.8% (92.0%)
18.0	87.9% (91.1%)	88.1% (91.4%)	88.6% (91.8%)	88.4% (91.8%)	87.7% (91.4%)	87.9% (91.4%)
20.0	87.0% (91.1%)	87.0% (91.1%)	87.7% (91.8%)	88.2% (92.0%)	87.2% (91.6%)	87.5% (91.4%)

E.10 Experiment CD.2.3

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per monophone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	0.759	0.316	0.114	-0.039	-0.165	-0.237
Q_{146}	7.734	-	-	-	-	-	3.597
Q_{50}	16.153	-	-	-	-	-	10.514

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	72.0% (75.8%)	84.7% (86.5%)	84.1% (86.1%)	84.7% (86.5%)	83.1% (85.4%)	84.0% (86.1%)
4.0	83.2% (85.9%)	89.5% (90.7%)	87.2% (89.1%)	86.8% (88.9%)	87.5% (89.5%)	86.5% (88.6%)
6.0	85.6% (88.4%)	90.2% (91.8%)	88.4% (90.2%)	88.4% (90.6%)	88.6% (90.6%)	88.4% (90.7%)
8.0	85.0% (88.6%)	90.7% (92.2%)	89.3% (91.3%)	88.6% (90.9%)	89.1% (91.6%)	88.2% (91.1%)
10.0	84.3% (88.1%)	89.7% (91.6%)	88.2% (90.4%)	88.1% (90.4%)	88.9% (91.6%)	88.9% (91.4%)
12.0	84.7% (88.1%)	88.6% (91.1%)	89.3% (91.4%)	89.1% (91.1%)	88.9% (91.8%)	88.4% (91.4%)
14.0	85.2% (88.6%)	88.8% (91.4%)	89.5% (91.6%)	88.6% (90.9%)	89.1% (92.0%)	88.4% (91.6%)
16.0	84.5% (88.1%)	88.4% (91.3%)	89.3% (91.6%)	88.6% (91.1%)	88.9% (91.8%)	88.4% (91.4%)
18.0	84.5% (88.2%)	88.1% (90.9%)	89.3% (91.6%)	89.1% (91.6%)	88.9% (92.0%)	88.8% (91.8%)
20.0	84.0% (88.1%)	88.2% (91.1%)	87.9% (90.7%)	89.1% (91.6%)	89.5% (92.3%)	88.6% (92.0%)

E.11 Experiment CD.2.4

NLDA approach : $Y = A^T f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	32
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	50 (one per monophone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	0.603	0.251	-0.006	-0.142	-0.196	-0.189
Q_{146}	7.734	-	-	-	-	-	3.526
Q_{50}	16.153	-	-	-	-	-	9.768

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	57.4% (61.0%)	79.0% (81.1%)	79.9% (82.9%)	82.5% (84.5%)	83.4% (85.2%)	82.2% (84.0%)
4.0	67.7% (73.1%)	85.0% (87.3%)	86.1% (88.2%)	86.8% (88.6%)	86.3% (88.2%)	85.7% (87.7%)
6.0	74.3% (80.0%)	87.3% (89.7%)	87.9% (90.2%)	87.7% (89.7%)	88.4% (90.6%)	87.9% (89.7%)
8.0	75.4% (81.6%)	88.6% (91.4%)	88.4% (90.6%)	89.1% (91.3%)	88.6% (90.9%)	89.7% (91.8%)
10.0	77.0% (82.5%)	88.6% (91.4%)	88.8% (91.3%)	89.1% (91.4%)	89.3% (91.6%)	88.8% (91.3%)
12.0	77.0% (83.2%)	88.9% (92.0%)	88.8% (91.3%)	89.1% (91.4%)	88.8% (91.4%)	88.8% (91.6%)
14.0	78.8% (84.7%)	88.9% (91.8%)	88.9% (91.3%)	89.1% (91.4%)	88.8% (91.4%)	88.9% (91.8%)
16.0	79.0% (84.8%)	88.8% (91.8%)	88.9% (91.4%)	88.9% (91.3%)	88.8% (91.6%)	89.1% (92.2%)
18.0	77.5% (84.5%)	88.6% (91.8%)	89.1% (91.6%)	88.8% (91.4%)	88.8% (91.6%)	89.3% (92.2%)
20.0	77.9% (84.7%)	88.8% (91.6%)	88.9% (91.4%)	88.4% (91.6%)	88.2% (91.6%)	88.6% (92.2%)

E.12 Experiment CD.2.5

NLDA approach : $Y = A^T f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	7124
target drift parameters	
α :	0.9
β :	0.4
m :	1
number of drift vectors :	50 (one per monophone)

backpropagation parameters	
number of discriminated targets :	7124
learning rate :	0.008
momentum :	0.9
number of hidden units :	32
iterations :	6
target update after :	6
sample selection :	random

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	0.516	0.064	0.005	-0.060	-0.184	-0.200
Q_{146}	7.734	-	-	-	-	-	3.431
Q_{50}	16.153	-	-	-	-	-	8.834

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	75.0% (77.7%)	81.1% (84.5%)	80.7% (83.2%)	82.2% (84.3%)	83.1% (85.0%)	82.7% (85.0%)
4.0	81.1% (84.8%)	86.6% (89.7%)	86.5% (88.6%)	87.7% (89.7%)	87.7% (89.5%)	86.6% (88.6%)
6.0	82.5% (86.6%)	87.9% (90.7%)	88.8% (91.4%)	88.2% (91.1%)	88.6% (91.1%)	87.7% (90.4%)
8.0	84.3% (88.1%)	88.6% (91.6%)	88.8% (91.6%)	88.4% (91.8%)	88.4% (91.6%)	89.1% (92.2%)
10.0	84.8% (88.4%)	88.1% (91.1%)	88.2% (91.1%)	88.9% (92.3%)	88.8% (92.2%)	89.5% (92.3%)
12.0	84.8% (88.6%)	88.1% (91.1%)	88.8% (91.6%)	89.3% (92.5%)	88.2% (91.4%)	88.8% (91.8%)
14.0	83.6% (88.1%)	87.9% (91.3%)	88.6% (91.4%)	88.2% (91.4%)	87.2% (91.1%)	88.8% (92.0%)
16.0	83.4% (88.6%)	87.7% (91.4%)	87.3% (91.1%)	87.5% (91.6%)	87.2% (91.3%)	87.7% (91.8%)
18.0	81.3% (88.1%)	87.0% (91.1%)	87.2% (91.1%)	87.5% (91.6%)	87.2% (91.3%)	87.5% (91.4%)
20.0	81.5% (87.7%)	86.8% (91.1%)	87.2% (91.3%)	87.5% (91.6%)	86.8% (91.3%)	87.5% (92.0%)

E.13 Experiment CD.2.6

NLDA approach : $Y = A^T f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	7124
target drift parameters	
α :	0.9
β :	1.0
m :	1
number of drift vectors :	50 (one per monophone)

backpropagation parameters	
number of discriminated targets :	7124
learning rate :	0.008
momentum :	0.9
number of hidden units :	32
iterations :	6
target update after :	6
sample selection :	random

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	0.518	0.483	0.428	0.443	0.441	0.393
Q_{146}	7.734	-	-	-	-	-	3.726
Q_{50}	16.153	-	-	-	-	-	7.995

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	27.8% (33.9%)	53.7% (56.1%)	61.7% (64.7%)	63.5% (65.8%)	68.8% (71.7%)	68.8% (71.5%)
4.0	40.5% (46.2%)	70.1% (72.9%)	77.5% (80.0%)	75.4% (78.8%)	78.3% (81.6%)	79.7% (82.2%)
6.0	44.2% (50.4%)	75.2% (77.7%)	81.3% (84.0%)	81.5% (84.5%)	82.9% (85.6%)	81.6% (84.8%)
8.0	49.2% (55.3%)	80.0% (82.7%)	82.5% (85.6%)	83.1% (86.5%)	84.0% (87.0%)	83.2% (86.3%)
10.0	50.1% (56.3%)	82.7% (85.7%)	83.1% (86.5%)	84.0% (87.3%)	85.4% (88.1%)	84.5% (87.7%)
12.0	52.4% (60.6%)	83.2% (85.9%)	84.1% (87.3%)	85.7% (88.9%)	85.9% (88.6%)	85.4% (88.4%)
14.0	54.2% (62.7%)	83.2% (86.1%)	84.3% (87.9%)	85.7% (89.1%)	85.4% (88.6%)	86.3% (88.9%)
16.0	52.9% (62.7%)	84.5% (87.7%)	85.4% (88.8%)	85.4% (88.8%)	85.6% (88.4%)	86.1% (88.6%)
18.0	53.1% (63.5%)	84.1% (87.3%)	85.9% (88.9%)	86.1% (89.3%)	86.6% (89.1%)	85.7% (88.2%)
20.0	54.2% (64.0%)	83.6% (87.0%)	86.3% (89.3%)	85.8% (88.8%)	86.5% (88.9%)	85.7% (88.2%)
22.0	54.0% (64.2%)	83.6% (87.0%)	86.5% (89.5%)	85.4% (88.6%)	86.3% (88.9%)	85.7% (88.2%)
24.0	52.6% (63.8%)	82.9% (86.8%)	86.6% (89.3%)	85.4% (88.6%)	86.1% (88.9%)	85.7% (88.4%)
26.0	51.0% (62.9%)	83.1% (87.5%)	86.1% (88.8%)	85.2% (88.2%)	86.1% (88.9%)	85.7% (88.2%)
28.0	51.0% (63.1%)	83.6% (87.2%)	85.6% (88.2%)	85.2% (88.2%)	86.1% (88.8%)	85.7% (88.1%)
30.0	51.9% (63.8%)	83.1% (86.6%)	85.0% (88.1%)	85.4% (88.2%)	86.5% (88.1%)	85.7% (88.2%)

E.14 Experiment CD.3.1

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	5
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.592	2.346	2.317	2.342	2.289	2.365
Q_{146}	7.734	-	-	-	-	-	7.036
Q_{50}	16.153	-	-	-	-	-	15.230

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	1.5	2.0	5.0	7.5	10.0	15.0
2.0	87.0% (88.6%)	87.0% (89.3%)	86.1% (88.6%)	85.9% (88.4%)	85.9% (88.2%)	86.1% (87.9%)
4.0	88.6% (90.6%)	89.8% (91.8%)	87.9% (90.6%)	87.7% (90.0%)	87.9% (90.2%)	89.1% (91.1%)
6.0	89.7% (92.0%)	89.5% (91.8%)	88.8% (91.1%)	88.9% (91.3%)	89.1% (91.4%)	89.3% (91.6%)
8.0	88.9% (92.2%)	88.6% (91.8%)	88.9% (91.6%)	89.1% (91.8%)	89.5% (92.0%)	88.8% (91.8%)
10.0	86.6% (90.0%)	87.5% (91.3%)	87.0% (90.4%)	87.7% (90.9%)	87.7% (90.9%)	87.0% (90.4%)
12.0	85.6% (89.7%)	86.1% (90.9%)	86.1% (90.2%)	86.6% (90.2%)	86.8% (90.4%)	86.1% (89.7%)
14.0	85.0% (89.3%)	85.7% (90.4%)	85.6% (90.0%)	86.3% (90.2%)	85.6% (89.7%)	84.8% (88.9%)
16.0	82.5% (87.9%)	84.8% (90.7%)	85.2% (89.8%)	84.8% (89.8%)	83.8% (88.6%)	84.1% (88.9%)
18.0	80.7% (87.2%)	83.4% (90.2%)	82.9% (88.9%)	81.8% (88.2%)	82.0% (87.3%)	82.5% (87.7%)
20.0	79.5% (86.3%)	81.3% (88.6%)	81.1% (88.1%)	80.9% (87.9%)	80.0% (87.2%)	78.4% (85.9%)

E.15 Experiment CD.3.2

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	5
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.621	2.472	2.460	2.268	2.313	2.456
Q_{146}	7.734	-	-	-	-	-	7.120
Q_{50}	16.153	-	-	-	-	-	15.241

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	1.5	2.0	5.0	7.5	10.0	15.0
2.0	87.2% (89.5%)	86.8% (89.5%)	86.1% (88.8%)	88.1% (90.0%)	87.9% (90.0%)	87.0% (89.5%)
4.0	88.9% (91.4%)	89.1% (91.8%)	89.5% (91.8%)	89.3% (91.8%)	88.9% (91.8%)	88.8% (91.3%)
6.0	89.8% (92.7%)	89.5% (92.3%)	89.5% (92.0%)	90.0% (92.3%)	89.7% (92.0%)	88.9% (91.6%)
8.0	89.7% (92.7%)	89.5% (92.5%)	88.9% (92.0%)	88.9% (92.2%)	88.9% (92.2%)	89.5% (92.7%)
10.0	88.8% (92.3%)	89.5% (92.5%)	88.9% (92.5%)	88.9% (92.5%)	88.8% (92.3%)	88.9% (92.5%)
12.0	87.3% (92.0%)	88.1% (91.8%)	88.1% (92.2%)	88.1% (92.0%)	87.2% (91.6%)	88.1% (91.8%)
14.0	86.5% (90.9%)	87.7% (91.6%)	87.2% (91.8%)	86.8% (91.3%)	86.5% (90.9%)	86.1% (90.7%)
16.0	85.4% (90.2%)	86.3% (90.7%)	87.0% (91.6%)	86.8% (91.4%)	85.7% (90.7%)	85.4% (90.2%)
18.0	84.0% (89.1%)	84.5% (90.0%)	86.3% (90.9%)	85.7% (90.4%)	84.7% (90.0%)	85.2% (89.8%)
20.0	82.0% (88.4%)	83.6% (89.7%)	82.2% (88.4%)	81.6% (87.3%)	82.2% (88.1%)	82.2% (87.7%)

E.16 Experiment CD.3.3

NLDA approach : $Y = A^T X + f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	7124
target drift parameters	
α :	0.9
β :	0.2
m :	1
number of drift vectors :	2376 (one per triphone)

backpropagation parameters	
number of discriminated targets :	7124
learning rate :	0.008
momentum :	0.9
number of hidden units :	5
iterations :	6
target update after :	6
sample selection :	random

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.678	2.522	2.475	2.523	2.459	2.404
Q_{146}	7.734	-	-	-	-	-	7.207
Q_{50}	16.153	-	-	-	-	-	15.280

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	1.5	2.0	5.0	7.5	10.0	15.0
2.0	86.6% (88.8%)	89.3% (90.7%)	86.6% (88.9%)	86.6% (89.1%)	87.2% (89.3%)	87.7% (89.7%)
4.0	89.7% (91.4%)	88.9% (91.6%)	88.8% (90.9%)	88.4% (90.9%)	89.1% (91.3%)	88.9% (90.9%)
6.0	89.8% (92.0%)	88.6% (91.4%)	89.5% (91.8%)	89.1% (91.6%)	90.0% (92.0%)	89.5% (91.4%)
8.0	89.8% (92.2%)	88.6% (91.4%)	89.3% (91.8%)	89.1% (91.6%)	90.2% (92.5%)	89.1% (91.4%)
10.0	89.7% (92.3%)	88.1% (91.8%)	87.5% (90.9%)	89.5% (92.3%)	89.7% (92.7%)	88.8% (92.0%)
12.0	88.1% (91.4%)	87.0% (91.1%)	87.5% (90.4%)	88.8% (91.4%)	88.2% (91.1%)	87.9% (90.9%)
14.0	86.5% (90.2%)	86.5% (90.6%)	86.6% (90.2%)	87.5% (90.7%)	87.2% (90.6%)	86.6% (89.8%)
16.0	85.6% (89.5%)	86.8% (90.7%)	85.6% (89.3%)	85.6% (90.0%)	85.6% (89.8%)	84.8% (89.1%)
18.0	84.1% (89.1%)	84.8% (90.0%)	85.9% (90.2%)	85.9% (90.2%)	84.3% (89.1%)	84.1% (88.9%)
20.0	84.5% (88.8%)	84.0% (89.1%)	85.7% (90.4%)	85.4% (89.1%)	85.2% (89.5%)	85.0% (89.3%)

E.17 Experiment CD.3.4

NLDA approach : $Y = A^T X + f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	7124
target drift parameters	
α :	0.9
β :	0.3
m :	1
number of drift vectors :	2376 (one per triphone)

backpropagation parameters	
number of discriminated targets :	7124
learning rate :	0.008
momentum :	0.9
number of hidden units :	5
iterations :	6
target update after :	6
sample selection :	random

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.782	2.643	2.655	2.602	2.368	2.314
Q_{146}	7.734	-	-	-	-	-	6.965
Q_{50}	16.153	-	-	-	-	-	14.998

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	1.5	2.0	5.0	7.5	10.0	15.0
2.0	89.1% (90.2%)	88.9% (90.9%)	86.8% (89.1%)	87.2% (89.7%)	88.6% (90.7%)	87.7% (89.5%)
4.0	90.0% (91.4%)	90.6% (92.3%)	89.1% (91.6%)	89.1% (91.4%)	90.2% (92.3%)	90.0% (91.6%)
6.0	89.7% (92.0%)	89.7% (92.2%)	90.9% (93.4%)	90.7% (93.2%)	90.9% (93.2%)	91.3% (93.6%)
8.0	89.3% (92.3%)	89.5% (92.9%)	90.0% (93.4%)	90.8% (93.4%)	90.0% (92.3%)	91.1% (93.6%)
10.0	88.2% (91.8%)	87.3% (90.9%)	88.8% (92.5%)	87.5% (91.6%)	88.2% (92.0%)	88.4% (91.8%)
12.0	87.9% (91.6%)	86.1% (90.4%)	87.0% (91.6%)	87.0% (90.9%)	87.5% (91.4%)	87.5% (91.1%)
14.0	86.3% (91.3%)	85.6% (90.2%)	86.1% (90.7%)	85.7% (90.2%)	86.5% (90.9%)	85.9% (90.9%)
16.0	84.5% (90.4%)	84.8% (90.6%)	84.7% (89.7%)	84.1% (89.7%)	84.8% (89.8%)	85.4% (90.6%)
18.0	82.9% (88.9%)	84.1% (89.8%)	82.2% (87.9%)	83.4% (89.3%)	82.9% (88.6%)	84.1% (89.1%)
20.0	80.7% (87.7%)	82.5% (89.5%)	79.0% (86.1%)	80.6% (87.0%)	79.5% (86.1%)	81.8% (87.7%)

E.18 Experiment CD.3.5

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	5
α :	0.9	iterations :	6
β :	0.4	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.679	2.503	2.489	2.456	2.447	2.472
Q_{146}	7.734	-	-	-	-	-	7.272
Q_{50}	16.153	-	-	-	-	-	15.349

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	1.5	2.0	5.0	7.5	10.0	15.0
2.0	88.4% (90.6%)	87.5% (89.3%)	86.8% (89.8%)	87.2% (90.0%)	87.5% (90.2%)	87.2% (89.7%)
4.0	89.7% (91.6%)	89.7% (91.8%)	88.2% (91.3%)	89.5% (91.6%)	89.7% (91.6%)	89.5% (91.4%)
6.0	89.3% (92.2%)	89.8% (92.2%)	88.9% (91.8%)	89.7% (92.5%)	89.7% (92.5%)	89.3% (92.0%)
8.0	88.1% (90.9%)	89.3% (92.3%)	88.8% (92.2%)	88.6% (91.8%)	89.5% (92.7%)	88.9% (91.4%)
10.0	87.2% (91.1%)	87.5% (91.4%)	87.7% (91.1%)	87.5% (91.3%)	87.2% (91.1%)	87.5% (91.1%)
12.0	87.3% (91.3%)	87.2% (91.4%)	86.6% (91.3%)	87.0% (91.3%)	86.8% (91.1%)	86.6% (90.7%)
14.0	85.6% (91.1%)	84.5% (90.2%)	86.3% (91.3%)	85.9% (91.3%)	85.7% (91.3%)	85.7% (91.1%)
16.0	84.8% (90.4%)	83.8% (89.3%)	84.1% (90.2%)	84.1% (89.8%)	83.6% (89.5%)	84.1% (89.8%)
18.0	83.1% (89.1%)	81.6% (87.9%)	83.4% (89.3%)	84.0% (89.3%)	82.9% (88.1%)	83.4% (88.9%)
20.0	80.9% (87.2%)	79.9% (87.3%)	81.5% (86.8%)	81.3% (87.7%)	80.2% (87.0%)	81.3% (87.7%)

E.19 Experiment CD.3.6

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	5
α :	0.9	iterations :	6
β :	1.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.711	2.649	2.546	2.544	2.529	2.577
Q_{146}	7.734	-	-	-	-	-	7.335
Q_{50}	16.153	-	-	-	-	-	15.293

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	1.5	2.0	5.0	7.5	10.0	15.0
2.0	86.8% (88.9%)	88.6% (90.2%)	86.8% (88.4%)	86.8% (88.8%)	87.5% (88.9%)	86.8% (88.8%)
4.0	89.3% (91.6%)	90.7% (92.9%)	88.8% (91.1%)	89.7% (91.3%)	89.3% (91.3%)	89.7% (91.8%)
6.0	89.5% (92.0%)	90.6% (92.9%)	89.1% (91.6%)	89.1% (91.4%)	90.2% (92.7%)	89.8% (92.2%)
8.0	90.0% (92.5%)	90.6% (93.0%)	89.1% (91.8%)	88.8% (91.4%)	90.0% (92.5%)	89.8% (92.3%)
10.0	89.5% (92.0%)	90.7% (93.4%)	88.9% (91.8%)	88.6% (91.4%)	90.2% (92.7%)	89.5% (92.0%)
12.0	89.1% (92.2%)	89.1% (92.5%)	89.1% (92.3%)	88.6% (92.0%)	88.9% (92.3%)	88.8% (92.2%)
14.0	88.6% (92.0%)	87.9% (92.2%)	87.9% (91.6%)	88.6% (91.6%)	89.1% (92.3%)	88.1% (91.6%)
16.0	87.2% (91.1%)	87.5% (92.2%)	87.0% (91.3%)	87.5% (91.3%)	86.1% (89.8%)	85.7% (90.2%)
18.0	84.7% (90.0%)	85.6% (91.3%)	84.8% (90.4%)	85.4% (90.2%)	85.7% (90.6%)	86.1% (91.1%)
20.0	83.8% (88.8%)	83.8% (90.6%)	82.9% (89.7%)	82.0% (89.1%)	83.6% (89.8%)	84.5% (89.8%)

E.20 Experiment CD.4.1

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
$\mathcal{Q}_{7124}$	3.010	2.458	2.076	1.898	1.775	1.665	1.656
$\mathcal{Q}_{146}$	7.734	-	-	-	-	-	6.340
$\mathcal{Q}_{50}$	16.153	-	-	-	-	-	14.813

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	87.9% (90.0%)	86.8% (88.9%)	88.4% (90.0%)	87.3% (89.7%)	88.6% (90.4%)	88.4% (90.2%)
4.0	90.0% (91.3%)	88.9% (90.7%)	89.8% (91.4%)	89.1% (90.9%)	89.7% (91.6%)	90.2% (91.8%)
6.0	89.7% (91.6%)	89.8% (91.8%)	89.1% (91.4%)	89.1% (91.6%)	89.3% (92.0%)	89.5% (91.8%)
8.0	88.4% (91.8%)	88.4% (91.4%)	87.7% (91.3%)	87.9% (91.3%)	88.8% (92.2%)	89.3% (92.5%)
10.0	86.8% (91.8%)	88.1% (91.3%)	86.8% (91.1%)	85.9% (90.9%)	85.0% (90.2%)	87.3% (92.0%)
12.0	86.5% (91.4%)	85.9% (90.0%)	84.7% (89.8%)	84.3% (89.3%)	84.8% (89.8%)	86.1% (90.9%)
14.0	84.8% (91.1%)	84.3% (90.0%)	84.7% (89.8%)	84.5% (89.5%)	84.5% (89.5%)	85.0% (90.0%)
16.0	84.8% (91.3%)	84.5% (90.7%)	84.1% (89.8%)	83.8% (89.7%)	83.2% (89.3%)	84.1% (89.5%)
18.0	84.3% (91.1%)	84.1% (90.0%)	82.9% (87.7%)	82.7% (87.7%)	82.7% (87.7%)	82.7% (87.7%)
20.0	83.2% (89.8%)	82.5% (88.4%)	82.4% (87.5%)	81.5% (87.2%)	80.7% (87.0%)	80.2% (86.8%)

E.21 Experiment CD.4.2

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
$\mathcal{Q}_{7124}$	3.010	2.183	1.910	1.780	1.764	1.690	1.672
$\mathcal{Q}_{146}$	7.734	-	-	-	-	-	6.396
$\mathcal{Q}_{50}$	16.153	-	-	-	-	-	14.651

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	86.6% (89.3%)	86.1% (88.8%)	87.0% (89.1%)	86.5% (88.6%)	87.2% (89.3%)	86.3% (88.8%)
4.0	87.7% (90.4%)	89.1% (91.8%)	88.6% (90.9%)	88.9% (91.3%)	89.1% (91.4%)	89.3% (91.4%)
6.0	88.4% (91.4%)	89.5% (92.2%)	89.7% (92.2%)	89.7% (92.2%)	89.3% (92.0%)	88.9% (92.0%)
8.0	88.8% (92.0%)	88.9% (92.5%)	88.9% (92.3%)	89.7% (92.5%)	89.5% (92.5%)	88.6% (92.2%)
10.0	88.2% (91.6%)	88.2% (92.2%)	88.4% (92.0%)	87.9% (91.6%)	88.4% (92.0%)	88.1% (91.8%)
12.0	87.7% (91.8%)	88.2% (92.3%)	88.8% (92.5%)	88.4% (92.2%)	88.1% (92.2%)	88.4% (92.2%)
14.0	87.0% (91.4%)	88.1% (92.3%)	88.4% (92.3%)	88.1% (92.0%)	87.7% (91.6%)	87.0% (91.4%)
16.0	85.7% (90.9%)	88.1% (92.3%)	87.7% (91.6%)	87.0% (91.4%)	86.6% (90.7%)	87.0% (91.3%)
18.0	85.8% (90.7%)	85.7% (91.1%)	86.5% (90.6%)	86.3% (90.6%)	85.9% (90.2%)	85.9% (91.1%)
20.0	83.4% (89.7%)	82.9% (89.1%)	85.0% (90.0%)	83.8% (89.1%)	83.6% (88.2%)	84.5% (89.3%)

E.22 Experiment CD.4.3

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	0.9	iterations :	6
β :	0.2	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.089	1.901	1.766	1.686	1.703	1.680
Q_{146}	7.734	-	-	-	-	-	6.372
Q_{50}	16.153	-	-	-	-	-	14.399

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	87.0% (89.1%)	87.3% (89.7%)	88.4% (90.9%)	89.1% (91.3%)	89.7% (91.4%)	88.9% (91.1%)
4.0	87.5% (90.2%)	89.3% (91.4%)	89.8% (92.0%)	89.7% (91.6%)	90.9% (92.7%)	90.9% (92.5%)
6.0	87.9% (91.1%)	90.2% (92.3%)	90.4% (93.0%)	91.1% (93.2%)	90.9% (93.0%)	90.7% (92.9%)
8.0	89.1% (92.5%)	90.2% (93.4%)	90.2% (93.2%)	90.7% (93.6%)	91.1% (93.6%)	90.9% (93.4%)
10.0	88.6% (92.9%)	89.5% (92.7%)	89.3% (92.7%)	90.2% (93.0%)	89.8% (92.7%)	90.4% (93.0%)
12.0	87.0% (92.2%)	87.9% (92.7%)	88.1% (92.5%)	88.6% (93.2%)	88.4% (92.9%)	88.1% (93.0%)
14.0	85.0% (91.8%)	87.2% (92.2%)	86.6% (91.4%)	87.0% (91.4%)	86.3% (90.7%)	86.6% (91.1%)
16.0	84.0% (90.7%)	85.9% (90.9%)	85.9% (90.7%)	87.0% (91.3%)	86.1% (90.9%)	86.6% (91.1%)
18.0	83.2% (90.2%)	83.8% (90.7%)	83.8% (90.4%)	85.6% (91.1%)	85.0% (90.7%)	85.7% (91.1%)
20.0	82.5% (90.0%)	82.9% (90.9%)	82.5% (89.7%)	82.4% (89.8%)	82.7% (89.8%)	82.9% (89.5%)

E.23 Experiment CD.4.4

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	10
α :	0.9	iterations :	6
β :	0.3	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.318	2.050	1.909	1.867	1.851	1.832
Q_{146}	7.734	-	-	-	-	-	6.481
Q_{50}	16.153	-	-	-	-	-	14.522

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	82.4% (85.6%)	87.2% (88.9%)	87.3% (89.1%)	86.8% (88.6%)	87.0% (88.9%)	87.5% (89.3%)
4.0	87.0% (90.0%)	88.1% (90.4%)	88.2% (90.4%)	88.9% (90.9%)	89.1% (91.1%)	88.9% (91.1%)
6.0	88.8% (90.9%)	89.7% (91.4%)	90.0% (92.2%)	90.0% (92.3%)	90.2% (92.2%)	89.8% (92.2%)
8.0	88.8% (90.9%)	89.3% (91.4%)	91.1% (92.7%)	89.8% (92.0%)	89.7% (91.8%)	90.4% (92.3%)
10.0	88.9% (91.8%)	89.3% (91.6%)	89.3% (91.4%)	89.1% (91.8%)	89.8% (92.3%)	89.7% (92.3%)
12.0	88.8% (92.2%)	88.1% (90.9%)	88.4% (90.7%)	88.4% (91.4%)	89.1% (91.8%)	89.5% (92.2%)
14.0	88.2% (91.8%)	87.5% (90.9%)	87.3% (90.9%)	86.6% (90.7%)	88.6% (92.0%)	88.9% (92.3%)
16.0	88.2% (92.0%)	87.3% (90.9%)	87.3% (90.9%)	86.5% (90.9%)	87.3% (91.3%)	87.0% (92.3%)
18.0	88.2% (92.0%)	86.8% (90.7%)	85.9% (90.4%)	86.1% (90.9%)	87.5% (91.4%)	87.5% (91.4%)
20.0	87.2% (91.6%)	85.6% (90.0%)	86.1% (90.7%)	86.5% (91.1%)	86.6% (90.9%)	86.6% (91.1%)

E.24 Experiment CD.4.5

NLDA approach : $Y = A^T X + f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	7124
target drift parameters	
α :	0.9
β :	0.4
m :	1
number of drift vectors :	2376 (one per triphone)

backpropagation parameters	
number of discriminated targets :	7124
learning rate :	0.008
momentum :	0.9
number of hidden units :	10
iterations :	6
target update after :	6
sample selection :	random

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.232	1.912	1.815	1.806	1.744	1.803
Q_{146}	7.734	-	-	-	-	-	6.512
Q_{50}	16.153	-	-	-	-	-	14.634

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	87.0% (88.9%)	86.6% (88.8%)	84.8% (87.2%)	86.1% (88.1%)	86.3% (88.4%)	86.1% (88.2%)
4.0	88.4% (90.4%)	88.4% (90.6%)	87.5% (90.0%)	88.2% (90.6%)	87.5% (90.2%)	87.5% (90.9%)
6.0	88.1% (91.1%)	88.1% (91.1%)	88.8% (91.1%)	88.8% (91.3%)	88.4% (90.9%)	88.6% (91.1%)
8.0	87.3% (90.6%)	89.5% (91.6%)	88.8% (91.1%)	88.8% (91.3%)	88.9% (91.4%)	89.1% (91.3%)
10.0	87.3% (90.2%)	89.5% (91.6%)	89.1% (91.4%)	88.6% (91.1%)	88.6% (91.1%)	88.8% (91.1%)
12.0	88.1% (90.9%)	89.3% (91.8%)	89.1% (91.8%)	88.9% (91.6%)	88.9% (91.4%)	89.3% (91.8%)
14.0	87.9% (90.9%)	88.6% (91.3%)	89.1% (91.6%)	88.9% (92.0%)	88.4% (91.6%)	88.2% (91.6%)
16.0	86.8% (90.7%)	87.9% (91.3%)	87.9% (91.3%)	87.9% (91.1%)	87.7% (90.9%)	87.7% (91.1%)
18.0	86.3% (90.6%)	88.1% (91.4%)	87.7% (91.3%)	87.0% (90.7%)	87.5% (90.9%)	87.7% (91.1%)
20.0	85.9% (90.9%)	87.9% (91.4%)	87.3% (91.1%)	87.2% (90.7%)	87.9% (90.9%)	87.7% (91.1%)

E.25 Experiment CD.4.6

NLDA approach : $Y = A^T X + f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	7124
target drift parameters	
α :	0.9
β :	1.0
m :	1
number of drift vectors :	2376 (one per triphone)

backpropagation parameters	
number of discriminated targets :	7124
learning rate :	0.008
momentum :	0.9
number of hidden units :	10
iterations :	6
target update after :	6
sample selection :	random

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.549	2.083	2.026	1.994	1.942	2.000
Q_{146}	7.734	-	-	-	-	-	6.577
Q_{50}	16.153	-	-	-	-	-	14.710

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	87.9% (89.8%)	86.3% (88.9%)	86.6% (89.1%)	87.5% (89.8%)	87.3% (89.3%)	86.5% (88.6%)
4.0	89.7% (91.8%)	88.6% (91.1%)	89.1% (91.6%)	89.5% (92.2%)	88.9% (91.6%)	89.1% (91.4%)
6.0	90.2% (92.9%)	88.8% (91.8%)	88.8% (92.0%)	88.9% (91.8%)	89.3% (92.0%)	89.5% (92.0%)
8.0	88.6% (92.2%)	88.8% (91.8%)	88.4% (92.0%)	88.8% (92.0%)	88.6% (91.8%)	88.4% (91.8%)
10.0	87.3% (91.4%)	87.2% (90.7%)	88.2% (92.2%)	88.2% (91.8%)	88.1% (91.6%)	88.6% (92.2%)
12.0	86.3% (90.4%)	86.6% (90.4%)	87.2% (91.3%)	87.0% (90.9%)	87.2% (91.3%)	87.0% (91.3%)
14.0	86.1% (90.6%)	86.6% (90.7%)	87.2% (91.1%)	86.8% (90.9%)	87.2% (91.1%)	86.5% (90.7%)
16.0	85.2% (90.4%)	85.7% (90.4%)	85.7% (90.7%)	85.2% (90.6%)	85.4% (90.4%)	85.7% (90.2%)
18.0	85.0% (90.0%)	83.8% (89.5%)	83.8% (89.8%)	82.5% (88.2%)	84.0% (90.0%)	82.7% (88.9%)
20.0	82.7% (88.9%)	79.3% (86.6%)	80.9% (87.9%)	80.0% (86.8%)	81.8% (88.4%)	80.4% (87.5%)

E.26 Experiment CD.5.1

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	1.0	iterations :	6
β :	0.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.220	1.751	1.505	1.525	1.115	0.965
Q_{146}	7.734	-	-	-	-	-	5.530
Q_{50}	16.153	-	-	-	-	-	14.536

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	87.7% (90.0%)	85.7% (88.6%)	84.8% (87.2%)	85.4% (87.9%)	86.1% (88.2%)	86.3% (88.2%)
4.0	89.3% (91.4%)	88.2% (90.6%)	88.6% (90.4%)	88.4% (90.6%)	88.6% (90.6%)	88.2% (90.2%)
6.0	88.8% (91.3%)	88.1% (90.9%)	89.1% (91.1%)	89.1% (91.4%)	88.1% (90.9%)	87.9% (90.7%)
8.0	88.2% (91.4%)	87.5% (91.3%)	87.7% (90.9%)	88.8% (92.0%)	88.6% (91.8%)	88.1% (91.3%)
10.0	86.6% (91.6%)	87.2% (91.1%)	88.2% (91.8%)	88.2% (91.6%)	88.1% (91.8%)	88.4% (92.0%)
12.0	86.5% (91.8%)	88.1% (92.2%)	88.4% (92.2%)	87.5% (91.1%)	86.5% (91.3%)	85.4% (90.4%)
14.0	85.2% (91.1%)	87.3% (91.3%)	86.3% (91.3%)	85.9% (91.1%)	86.5% (91.3%)	84.5% (90.0%)
16.0	84.7% (91.1%)	85.7% (90.9%)	85.6% (90.6%)	85.2% (90.4%)	85.4% (90.6%)	85.7% (90.6%)
18.0	84.0% (90.7%)	85.6% (90.0%)	84.3% (89.5%)	84.8% (91.4%)	83.1% (89.5%)	83.2% (89.7%)
20.0	82.0% (88.9%)	83.6% (89.5%)	82.7% (88.6%)	81.3% (87.2%)	81.1% (87.7%)	79.9% (86.5%)

E.27 Experiment CD.5.2

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	0.1	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.065	1.552	1.282	1.132	0.901	0.883
Q_{146}	7.734	-	-	-	-	-	5.411
Q_{50}	16.153	-	-	-	-	-	14.248

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	88.9% (91.1%)	88.2% (90.4%)	86.6% (89.1%)	87.3% (89.5%)	87.2% (89.5%)	86.5% (88.9%)
4.0	90.4% (92.9%)	89.1% (91.6%)	88.4% (91.1%)	88.9% (91.6%)	89.8% (92.2%)	89.3% (91.8%)
6.0	89.7% (92.2%)	89.7% (92.0%)	88.2% (91.1%)	89.1% (91.8%)	89.3% (92.2%)	90.0% (92.3%)
8.0	88.9% (92.2%)	88.1% (91.3%)	87.9% (91.1%)	88.9% (92.0%)	88.1% (91.6%)	89.3% (92.3%)
10.0	88.4% (92.0%)	87.9% (91.3%)	87.7% (91.3%)	88.4% (92.0%)	85.9% (90.4%)	87.5% (91.4%)
12.0	88.2% (92.0%)	86.1% (90.6%)	86.3% (90.6%)	86.3% (90.6%)	85.9% (90.6%)	86.5% (91.1%)
14.0	86.8% (91.8%)	84.7% (90.0%)	86.1% (90.6%)	85.4% (90.0%)	85.9% (90.4%)	85.6% (89.5%)
16.0	84.7% (90.9%)	84.0% (89.7%)	84.5% (89.8%)	84.5% (88.9%)	85.4% (88.7%)	86.3% (89.8%)
18.0	84.0% (90.7%)	83.8% (89.7%)	84.8% (89.7%)	84.3% (89.3%)	85.7% (89.5%)	85.7% (90.2%)
20.0	81.8% (89.7%)	79.1% (87.3%)	79.3% (86.3%)	80.9% (87.7%)	79.0% (86.1%)	80.0% (86.3%)

E.28 Experiment CD.5.3

NLDA approach : $Y = A^T X + f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	7124
target drift parameters	
α :	0.9
β :	0.2
m :	1
number of drift vectors :	2376 (one per triphone)

backpropagation parameters	
number of discriminated targets :	7124
learning rate :	0.008
momentum :	0.9
number of hidden units :	20
iterations :	6
target update after :	6
sample selection :	random

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	1.995	1.382	1.187	1.060	1.029	1.096
Q_{146}	7.734	-	-	-	-	-	5.693
Q_{50}	16.153	-	-	-	-	-	14.379

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	85.9% (88.6%)	87.0% (89.1%)	85.4% (88.1%)	86.3% (88.6%)	87.0% (89.5%)	86.8% (88.8%)
4.0	88.8% (91.1%)	88.2% (90.9%)	87.9% (90.6%)	88.2% (90.7%)	88.9% (91.3%)	87.7% (90.2%)
6.0	89.5% (92.7%)	89.7% (92.2%)	88.8% (91.6%)	88.8% (91.6%)	89.5% (92.2%)	89.1% (91.8%)
8.0	89.5% (92.5%)	89.3% (92.0%)	89.3% (92.0%)	89.7% (92.0%)	89.7% (92.0%)	89.8% (92.2%)
10.0	89.7% (92.9%)	89.1% (92.5%)	88.8% (92.0%)	88.9% (92.0%)	88.6% (91.6%)	88.4% (91.3%)
12.0	88.4% (92.3%)	88.6% (91.8%)	88.4% (91.4%)	88.2% (91.3%)	88.4% (91.3%)	88.4% (91.1%)
14.0	88.6% (92.3%)	88.8% (92.0%)	87.2% (91.1%)	88.1% (91.6%)	88.2% (91.4%)	87.3% (90.9%)
16.0	87.7% (91.8%)	87.7% (91.4%)	87.5% (91.1%)	88.2% (91.4%)	87.9% (91.1%)	87.3% (90.6%)
18.0	85.6% (91.1%)	87.0% (91.3%)	86.5% (90.4%)	86.5% (90.7%)	87.0% (90.9%)	87.3% (90.6%)
20.0	84.5% (90.6%)	86.1% (90.9%)	85.9% (90.4%)	85.6% (90.2%)	86.8% (90.9%)	86.6% (90.6%)

E.29 Experiment CD.5.4

NLDA approach : $Y = A^T X + f(X)$

genaral parameters	
dimension of orignial space :	32
dimension of image space :	16
discriminated classes in on top LDA :	7124
target drift parameters	
α :	0.9
β :	0.3
m :	1
number of drift vectors :	2376 (one per triphone)

backpropagation parameters	
number of discriminated targets :	7124
learning rate :	0.008
momentum :	0.9
number of hidden units :	20
iterations :	6
target update after :	6
sample selection :	random

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.179	1.520	1.236	1.155	1.141	1.032
Q_{146}	7.734	-	-	-	-	-	5.630
Q_{50}	16.153	-	-	-	-	-	14.014

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	87.7% (89.7%)	85.0% (87.5%)	84.5% (87.3%)	85.2% (87.5%)	84.5% (86.5%)	85.4% (87.3%)
4.0	90.6% (92.5%)	87.7% (90.2%)	87.9% (90.0%)	86.4% (90.4%)	88.4% (90.2%)	88.8% (90.6%)
6.0	90.4% (92.5%)	89.1% (91.4%)	89.3% (91.6%)	88.6% (91.3%)	89.3% (91.8%)	89.8% (92.2%)
8.0	89.8% (92.2%)	90.2% (92.7%)	89.7% (92.0%)	89.5% (92.2%)	90.6% (92.7%)	91.1% (93.0%)
10.0	90.0% (92.7%)	90.2% (93.4%)	89.5% (92.5%)	90.0% (92.9%)	90.2% (93.0%)	90.9% (93.4%)
12.0	88.4% (92.0%)	89.8% (93.4%)	89.8% (92.9%)	89.5% (92.7%)	90.0% (93.4%)	90.0% (93.2%)
14.0	87.7% (91.8%)	89.1% (92.2%)	89.7% (93.0%)	89.8% (93.4%)	89.7% (93.2%)	89.8% (93.0%)
16.0	86.6% (91.1%)	87.7% (92.3%)	88.8% (92.7%)	88.1% (92.2%)	87.5% (92.9%)	88.8% (93.0%)
18.0	86.6% (91.6%)	86.1% (92.2%)	86.3% (92.3%)	87.5% (92.7%)	87.2% (93.0%)	87.9% (92.3%)
20.0	85.7% (91.8%)	85.7% (91.8%)	85.9% (92.0%)	85.7% (92.2%)	85.9% (91.8%)	85.7% (91.8%)

E.30 Experiment CD.5.5

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	0.4	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.130	1.515	1.305	1.318	1.283	1.337
Q_{146}	7.734	-	-	-	-	-	5.852
Q_{50}	16.153	-	-	-	-	-	14.192

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	89.3% (90.9%)	88.1% (90.2%)	87.5% (89.5%)	87.7% (89.5%)	87.5% (89.5%)	88.2% (89.7%)
4.0	90.2% (92.3%)	89.5% (91.3%)	88.6% (90.7%)	88.8% (90.7%)	88.9% (91.1%)	90.4% (91.6%)
6.0	89.7% (92.0%)	89.5% (91.8%)	88.9% (91.4%)	88.8% (91.4%)	89.3% (91.8%)	89.5% (92.2%)
8.0	89.1% (92.2%)	89.1% (92.0%)	88.9% (91.8%)	88.6% (91.8%)	89.3% (92.3%)	89.3% (92.3%)
10.0	88.2% (91.8%)	88.8% (91.8%)	88.4% (91.3%)	87.5% (91.3%)	88.4% (92.0%)	88.4% (91.8%)
12.0	87.5% (92.0%)	88.1% (92.0%)	86.6% (91.1%)	86.5% (91.1%)	88.4% (92.5%)	86.8% (91.6%)
14.0	86.3% (91.1%)	86.6% (92.3%)	86.6% (90.7%)	85.6% (90.7%)	86.6% (92.0%)	87.3% (92.3%)
16.0	85.2% (90.9%)	85.6% (91.3%)	85.7% (91.3%)	85.7% (91.3%)	85.7% (91.3%)	85.2% (91.1%)
18.0	84.1% (90.6%)	84.5% (90.4%)	85.6% (90.7%)	84.5% (90.7%)	85.0% (90.7%)	83.4% (90.2%)
20.0	84.1% (90.6%)	83.2% (89.7%)	82.4% (88.8%)	82.2% (88.8%)	84.3% (88.8%)	81.6% (89.1%)

E.31 Experiment CD.5.6

NLDA approach : $Y = A^T X + f(X)$

genaral parameters		backpropagation parameters	
dimension of orignial space :	32	number of discriminated targets :	7124
dimension of image space :	16	learning rate :	0.008
discriminated classes in on top LDA :	7124	momentum :	0.9
target drift parameters		number of hidden units :	20
α :	0.9	iterations :	6
β :	1.0	target update after :	6
m :	1	sample selection :	random
number of drift vectors :	2376 (one per triphone)		

developement of class separability

iteration(s)	0	1	2	3	4	5	6
Q_{7124}	3.010	2.144	1.621	1.644	1.420	1.378	1.363
Q_{146}	7.734	-	-	-	-	-	5.936
Q_{50}	16.153	-	-	-	-	-	14.013

word recognition perfomance on 12 speaker 48 sentence evaluation set

	TRcbfact					
ac	2.0	5.0	7.5	10.0	15.0	20.0
2.0	86.1% (88.4%)	84.3% (86.6%)	84.5% (86.5%)	58.6% (87.7%)	85.6% (87.7%)	85.2% (87.7%)
4.0	88.9% (91.1%)	87.0% (89.3%)	87.3% (89.5%)	87.0% (89.1%)	87.0% (89.1%)	87.7% (90.0%)
6.0	89.5% (91.4%)	88.2% (90.6%)	88.2% (90.7%)	88.4% (90.9%)	88.2% (90.7%)	88.1% (90.9%)
8.0	88.9% (91.6%)	88.1% (90.7%)	88.1% (90.7%)	87.9% (90.7%)	87.9% (90.9%)	88.1% (91.1%)
10.0	88.2% (91.4%)	88.2% (91.1%)	88.1% (91.1%)	88.2% (91.4%)	88.1% (91.1%)	87.5% (90.9%)
12.0	86.6% (91.1%)	87.9% (91.3%)	87.5% (90.9%)	87.2% (90.4%)	85.9% (89.8%)	86.6% (90.6%)
14.0	86.8% (91.3%)	85.7% (90.4%)	86.1% (90.6%)	85.6% (90.0%)	86.1% (90.4%)	86.5% (90.7%)
16.0	86.3% (91.3%)	85.0% (90.2%)	85.9% (90.4%)	85.7% (90.2%)	85.7% (90.4%)	86.3% (90.7%)
18.0	85.6% (90.7%)	84.7% (90.0%)	85.9% (90.4%)	85.7% (90.4%)	85.6% (90.4%)	85.6% (90.2%)
20.0	85.4% (90.4%)	84.1% (89.3%)	85.0% (89.5%)	85.4% (90.0%)	84.1% (89.3%)	84.1% (88.9%)

Bibliography

[BM53] G. Birkhoff and S. MacLane. *A Survey of Modern Algebra.* Macmillan, New York, 1953.

[Fis26] R.A. Fisher. *Contributions to Mathematical Statistics.* John Wiley & Sons, Inc. New York, 1926. Papers 32 and 33.

[Fuk90] K. Fukunaga. *Introduction to Statistical Pattern Recognition.* Academic Press, Boston, 1990.

[HJ89] X.D. Huang and M.A. Jack. *Semi-Continous Hidden Markov Models for Speech Recognition.* Computer Speech and Language 3, July 1989. pp. 239-252.

[HKP91] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computing.* Addison Wesley Publishers, 1991.

[HL89] M. Hunt and C. Lefevre. *A Comparison of Several Acoustic Representations for Speech Recognition with Degraded and Undegraded Speech.* Proc. ICASSP, New York, 1989.

[HRBA91] M. Hunt, S. Richardson, D. Bateman, and A.Piau. *An Investigation of PLP and IMELDA Acoustic Representation and of their Potential of Combination.* Proc. ICASSP, New York, 1991.

[HTB93] T. Hastie, R. Tibshirani, and A. Buja. *Flexible Discriminant Analysis by Optimal Scoring.* AT&T Bell Laboratories internal report, February 1993.

[KNF75] W.L.G. Koontz, P.M. Narendra, and K. Fukunaga. *A Branch and Bound Clustering Algorithm.* Trans. IEEE Computers C-24, 1975. pp. 908-915.

[Mai94] M. Maier. *Dimensionalitaetsreduktion von Sprachsignalen mit statistischen und neuronalen Methoden.* Master's thesis, University Karlsruhe, Germany, 1994.

[Piz62] S.M. Pizer. *Numerical Computing and Mathematical Analysis.* Science Research Associates, Chicago, 1962.

[PWFP88] P. Price, J. Bernstein W. Fischer, and D. Pallet. *A Database for Continous Speech Recognition in a 1000-Word Domain.* Proc. ICASSP, New York, 1988.

[Rab89] L.R. Rabiner. *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition.* Proceedings of IEEE, 1989.

[Ste73] G.W. Stewart. *Introduction to Matrix Computations.* Academic Press, New York, 1973.

[TSM85] D.M. Titterington, A.F.M. Smith, and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions.* John Wiley & Sons, New York, 1985.

[TW69] C.B. Tompkins and W.L. Wilson. *Elemeantary Numerical Analysis.* Prentice Hall, 1969.

[Wai91] A. Waibel *et al.* *JANUS: A Speech-to-Speech Translating System Using Connectionist and Symbolic Processing Strategies.* Proc. ICASSP, Toronto, Canada, 1991.

[Wai92] A. Waibel *et al.* *JANUS: Speech-to-Speech Translating Using Connectionist.* Proc. ICASSP, Toronto, Canada, 1992.

[Wil63] S. Wilks. *Mathematical Statistics.* John Wiley & Sons, New York-London, 1963.

[WL90] A. Webb and D. Lowe. *The Optimised Internal Representation of Multilayer Classifier Networks Performs Nonlinear Discriminant Analysis. Neural Networks*, 3:pp. 367–375, 1990.

[WY81] A. Waibel and B. Yegnanarayana. Comparative Study of Nonlinear Time Warping Techniques in Isolated Word Speech Recognition Systems. Technical Report, Carnegie Mellon University, 1981.