Subspace and Hypothesis based Effective Segmentation of Co-articulated Basic-units for Concatenative Speech Synthesis

R. Muralishankar, R. Srikanth and A. G. Ramakrishnan
Department of Electrical Engineering
Indian Institute of Science
Bangalore, INDIA
080-293-2935
sripad, srikanth, ramkiag@ee.iisc.ernet.in

Abstract— In this paper, we present two new methods for Vowel-Consonant segmentation of a co-articulated basic-units employed in our Thirukkural Tamil Text-to-Speech synthesis system [1]. The basic-units considered in [1] are CV, VC, VCV, VCCC and VCCC, where C stands for a consonant and V for any vowel. In the first method, we use subspace-based approach for vowel-consonant segmentation. It uses oriented principal component analysis (OPCA) where the test feature vectors are projected onto the V and C subspaces. The crossover of the norm-contours obtained by projecting test basic-unit onto the V and C subspaces give the segmentation points which in turn helps in identifying the V and C durations of a test basic-unit. In the second method, we use probabilistic principal component analysis (PPCA) [2] to get probability models for V and C. We then use Neymen-Pearson (NP) test to segment the basic-unit into V and C. Finally, we show that the hypothesis testing turns out to be an energy detector for V-C segmentation which is similar to the first method.

1. INTRODUCTION

For the purpose of recognition or synthesis, speech often needs to be segmented into phonetic units. Manual segmentation is tedious, time consuming and error prone. Due to variability both in human visual and acoustic perceptual capability, it is almost impossible to reproduce the manual segmentation results. Hence manual segmentation is inherently inconsistent. Automatic segmentation is not faultless, but it is inherently consistent and results are reproducible. Ideally, one likes to have an automatic segmentation which can handle basic-units uttered by different speakers. There are two broad categories of speech segmentation [3] namely, implicit and explicit. Implicit methods split up the utterance without explicit information, such as the phonetic transcription, and are based on the definition of a segment as a spectrally stable part of a signal.

Motivation for Subspace based Segmentation

In our synthesis scheme [1], concatenation is always performed across identical vowels. Changes in duration, pitch and amplitude are obtained by processing the vowel parts of the basic-units only. Thus, the segmentation of basic-units into vowel and consonant parts is needed to keep the consonant portion of the waveform intact. Plosives, affricates and fricatives have a common property of low energy when compared with any of the vowels. Figure 1 shows the performance of energy based segmentation for plosive and co-articulated basic-units. As shown in Figs. 1(a) and (b), accurate segmentation can be obtained for non co-articulated units, and not for co-articulated basic-units. The true consonant part /y/ in the signal /ey/ is shown in Fig. 1(b) with the boundaries dotted.

We propose two methods for co-articulated basic-unit segmentation. In the first method, we use subspace approach using oriented principal component analysis (OPCA) for basic-unit segmentation. In the second method, we use probabilistic principal component analysis (PPCA) [2] to obtain probability models for vowel (V) and consonants (C). We employ Neymen-Pearson (NP) test using the probability models for basic-unit segmentation.

2. Subspace based Segmentation

When we consider an individual vowel or consonant, there exist techniques like LPC to model their statistical properties. While segmenting the vowel part of a basic-unit, we can consider the vowel information (VI) in the feature vectors as the signal and the consonant information (CI) as noise. Similarly when the segmentation of the consonant part is required, we can view CI as signal and VI as noise. We present a linear feature transformation that aims at finding a subspace, of the feature space, in which the Signal-to-Noise ratio (SNR) is maximum. Such a decomposition can be arrived at by representing VI and CI by training vectors obtained using manual segmentation. The directions in the feature space where the SNR is maximum can be obtained by the generalized eigenvalue decomposition of the covariance matrices of the above vectors. Consider a linear transformation matrix $W$ that maps the original feature vectors $x$ on to $\tilde{x}$:

$$\tilde{x} = W^T x$$  \hspace{1cm} (1)

where $x$ is an $n$-dimensional vector, $\tilde{x}$ is an $m$-dimensional vector, $\gamma_1 \leq \gamma_n$, and $W$ is an $\gamma_1 \times \gamma_n$ matrix with $m$ linearly independent columns. Let $d_v$ and $d_c$ represent the training vectors containing VI and CI, respectively. in the original feature space. The covariance matrices for these training feature vectors are:

$$\Sigma_v = \frac{1}{m_v} \sum_{i=1}^{m_v} (d_v - \mu_v)(d_v - \mu_v)^T$$  \hspace{1cm} (2)

$$\Sigma_c = \frac{1}{m_c} \sum_{i=1}^{m_c} (d_c - \mu_c)(d_c - \mu_c)^T$$  \hspace{1cm} (3)

where $d_v$ and $d_c$ are $\gamma_1 \times 1$ vectors, and $\mu_v$ and $\mu_c$ are $\gamma_1 \times 1$ vectors.

The relationship between $\Sigma_v$ and $\Sigma_c$ can be obtained from the following relationship:

$$\Sigma_v = \frac{1}{m_v} \sum_{i=1}^{m_v} (d_v - \mu_v)(d_v - \mu_v)^T$$  \hspace{1cm} (4)

$$\Sigma_c = \frac{1}{m_c} \sum_{i=1}^{m_c} (d_c - \mu_c)(d_c - \mu_c)^T$$  \hspace{1cm} (5)

Subspace and Hypothesis based Effective Segmentation of Co-articulated Basic-units for Concatenative Speech Synthesis

R. Muralishankar, R. Srikanth and A. G. Ramakrishnan
Department of Electrical Engineering
Indian Institute of Science
Bangalore, INDIA
080-293-2935
sripad, srikanth, ramkiag@ee.iisc.ernet.in

Abstract— In this paper, we present two new methods for Vowel-Consonant segmentation of a co-articulated basic-units employed in our Thirukkural Tamil Text-to-Speech synthesis system [1]. The basic-units considered in [1] are CV, VC, VCV, VCCC and VCCC, where C stands for a consonant and V for any vowel. In the first method, we use subspace-based approach for vowel-consonant segmentation. It uses oriented principal component analysis (OPCA) where the test feature vectors are projected onto the V and C subspaces. The crossover of the norm-contours obtained by projecting test basic-unit onto the V and C subspaces give the segmentation points which in turn helps in identifying the V and C durations of a test basic-unit. In the second method, we use probabilistic principal component analysis (PPCA) [2] to get probability models for V and C. We then use Neymen-Pearson (NP) test to segment the basic-unit into V and C. Finally, we show that the hypothesis testing turns out to be an energy detector for V-C segmentation which is similar to the first method.

1. INTRODUCTION

For the purpose of recognition or synthesis, speech often needs to be segmented into phonetic units. Manual segmentation is tedious, time consuming and error prone. Due to variability both in human visual and acoustic perceptual capability, it is almost impossible to reproduce the manual segmentation results. Hence manual segmentation is inherently inconsistent. Automatic segmentation is not faultless, but it is inherently consistent and results are reproducible. Ideally, one likes to have an automatic segmentation which can handle basic-units uttered by different speakers. There are two broad categories of speech segmentation [3] namely, implicit and explicit. Implicit methods split up the utterance without explicit information, such as the phonetic transcription, and are based on the definition of a segment as a spectrally stable part of a signal.

Motivation for Subspace based Segmentation

In our synthesis scheme [1], concatenation is always performed across identical vowels. Changes in duration, pitch and amplitude are obtained by processing the vowel parts of the basic-units only. Thus, the segmentation of basic-units into vowel and consonant parts is needed to keep the consonant portion of the waveform intact. Plosives, affricates and fricatives have a common property of low energy when compared with any of the vowels. Figure 1 shows the performance of energy based segmentation for plosive and co-articulated basic-units. As shown in Figs. 1(a) and (b), accurate segmentation can be obtained for non co-articulated units, and not for co-articulated basic-units. The true consonant part /y/ in the signal /ey/ is shown in Fig. 1(b) with the boundaries dotted.

We propose two methods for co-articulated basic-unit segmentation. In the first method, we use subspace approach using oriented principal component analysis (OPCA) for basic-unit segmentation. In the second method, we use probabilistic principal component analysis (PPCA) [2] to obtain probability models for vowel (V) and consonants (C). We employ Neymen-Pearson (NP) test using the probability models for basic-unit segmentation.

2. Subspace based Segmentation

When we consider an individual vowel or consonant, there exist techniques like LPC to model their statistical properties. While segmenting the vowel part of a basic-unit, we can consider the vowel information (VI) in the feature vectors as the signal and the consonant information (CI) as noise. Similarly when the segmentation of the consonant part is required, we can view CI as signal and VI as noise. We present a linear feature transformation that aims at finding a subspace, of the feature space, in which the Signal-to-Noise ratio (SNR) is maximum. Such a decomposition can be arrived at by representing VI and CI by training vectors obtained using manual segmentation. The directions in the feature space where the SNR is maximum can be obtained by the generalized eigenvalue decomposition of the covariance matrices of the above vectors. Consider a linear transformation matrix $W$ that maps the original feature vectors $x$ on to $\tilde{x}$:

$$\tilde{x} = W^T x$$  \hspace{1cm} (1)

where $x$ is an $n$-dimensional vector, $\tilde{x}$ is an $m$-dimensional vector, $\gamma_1 \leq \gamma_n$, and $W$ is an $\gamma_1 \times \gamma_n$ matrix with $m$ linearly independent columns. Let $d_v$ and $d_c$ represent the training vectors containing VI and CI, respectively. in the original feature space. The covariance matrices for these training feature vectors are:

$$\Sigma_v = \frac{1}{m_v} \sum_{i=1}^{m_v} (d_v - \mu_v)(d_v - \mu_v)^T$$  \hspace{1cm} (2)

$$\Sigma_c = \frac{1}{m_c} \sum_{i=1}^{m_c} (d_c - \mu_c)(d_c - \mu_c)^T$$  \hspace{1cm} (3)

where $d_v$ and $d_c$ are $\gamma_1 \times 1$ vectors, and $\mu_v$ and $\mu_c$ are $\gamma_1 \times 1$ vectors.
Speech Recognition and Synthesis

Figure 1. Basic-unit segmentation using energy based method. (a) Speech signal /aka/. (b) Co-articulated signal feyl (continuous vertical line: segmentation using energy based method).

vectors can be written as

$$C_v = E[(d_v - \bar{d}_v)(d_v - \bar{d}_v)^T]$$

$$C_c = E[(d_c - \bar{d}_c)(d_c - \bar{d}_c)^T]$$

where $\bar{d}_v$ and $\bar{d}_c$ represent the means of $d_v$ and $d_c$ respectively. We collect an ensemble of feature vectors of length N corresponding to different vowels to estimate the $N \times N$ vowel covariance matrix $C_v$. Similarly we estimate the consonant covariance matrix $C_c$. We wish to find a $W$ that maximizes the ratio of the variance of $V_1$ to that of $C_1$ after the transformation. If the density functions of $d_v$ and $d_c$ are assumed to be normally distributed, then their covariance matrices after transformation are given by

$$\tilde{C}_v = W^T C_v W$$

$$\tilde{C}_c = W^T C_c W$$

A simple measure of the variance or the 'scatter' is the determinant of the covariance matrix [4]. Thus, the criterion function to be maximized is given by

$$J(W) = \frac{|\tilde{C}_v|}{|\tilde{C}_c|} = \frac{|W^T C_v W|}{|W^T C_c W|}$$

The columns of the optimum $W$ are obtained as generalized eigenvectors for vowels (GEVV), corresponding to the largest eigenvalues in

$$C_v w_v^{(v)} = \lambda_v C_v w_v^{(v)}$$

Similarly, we obtain generalized eigenvectors for consonants (GEVC) as

$$C_c w_c^{(c)} = \lambda_c C_c w_c^{(c)}$$

Evaluation of Norm-Contours

The GEVV and GEVCs are obtained by solving equations 5 and 6. The test signal is divided into overlapping frames and the feature vector $z_k$ corresponding to the $k^{th}$ frame is obtained using LP-Cepstral coefficients (LPCC) or Mel-Cepstral coefficients (MCC). We evaluate the norm-contours as follows.

$$N_v(k) = \sum_{i=1}^{M} (w_v^{(v)})^T z_k \& N_c(k) = \sum_{i=1}^{M} (w_c^{(c)})^T z_k$$

$N_v$ and $N_c$ give the norm-contours from $V$ and $C$ subspaces. Norm of the projection of the feature vectors derived from the test basic-unit on GEVV and GEVC give the norm-contours. One of them represents the vowel information and the other, the consonant information. The resulting norm contours obtained for a test signal cross each other at the beginning and at the end of consonant region of a given test basic-unit. The segmentation points are the ones where $N_v(k) = N_c(k)$. We found that optimum results were obtained when $M = 3$.

3. STATISTICAL TESTING FOR SEGMENTATION

Here, we view the basic-unit segmentation as classifying frames of a given test basic-unit into one of two classes: vowels and consonants. This indeed is a hypothesis testing (two-class) problem and it requires probability models for vowel & consonant and a robust threshold as well. Tipping and Bishop [2] proposed PPCA, to emphasize the advantages of associating a probability model with principal component analysis (PCA), rather than considering the algorithmic perspective of
determining principal components. Using PPCA approach [2], we generate probability model for V and C.

Motivation for Hypothesis based Segmentation

In the section 2, we saw the performance of oriented principal component analysis based basic-unit segmentation. Although the performance of OPCA is good, the crossovers (threshold) are not robust. So, segmentation becomes difficult when there are multiple crossovers. Figure 4(a) shows a basic-unit /au/mai/, and Fig. 4(b) shows the norm-contours obtained after projecting the basic-unit in Fig. 4(a) on V and C subspaces. One can see from Fig. 4(b) that there are multiple crossovers, making it difficult to choose an appropriate threshold for segmentation. Here, classical decision theory [5] can be used to find the robust threshold. Along with the crossover information, decision regarding vowel or consonant class is used to obtain correct segmentation points by eliminating the false crossovers. The class decision is shown in Fig. 4(d).

Probability Models using PPCA

Unlike in PCA, PPCA gives a probability density for the data through a latent variable model. A latent variable model relates a d-dimensional observed data vector t to a q-dimensional (q < d) latent vector x by defining a noise model and a prior on the distribution of the latent variables in the form,

$$ t = U x + \mu + \epsilon, \quad (8) $$

where \( \epsilon \) is an \( x \)-independent noise process and \( U \) is the \( (d \times q) \) generative matrix. It is common that the prior distribution of the latent variables is a sample Gaussian distribution \( x \sim N(0, \Sigma) \).

![Figure 3](image1.png)

**Figure 3.** (a) Speech signal /o/ vs. (b) Its segmentation into vowel (/o/) and consonant (/y/ and /m/) regions, using both the vowel and consonant norm-contours. (c) Spectrogram of /oymo/.

![Figure 4](image2.png)

**Figure 4.** Segmentation of basic-unit using OPCA and PPCA. (a) Test basic-unit /au/mai/, (b) Vowel and consonant norm contours after projecting the test basic-unit into V & C subspaces. (c) Test statistics \( T(y) \), along with the threshold for a given \( P_{f A} \equiv 1 - \alpha \). (d) V & C decision after the class testing. (e) Spectrogram of /au/mai/.

If \( \psi = \sigma^2 I \), the latent variable model is called PPCA [2]. In this terminology, conventional PCA is recovered when \( \sigma^2 \rightarrow 0 \). One of the ways of estimating the parameters of these latent variable models is by the Expectation-Maximization algorithm. Tipping and Bishop [6] have recently formulated PCA within a maximum-likelihood framework based on a specific form of Gaussian latent variable model. They also showed that with \( C = \sigma^2 I + UU^T \), the only stationary points occur when

$$ U = W_q (\Lambda_q - \sigma^2 I)^{-1} R. \quad (10) $$

Here \( W_q \) has q eigenvectors of S (sample covariance matrix, given by \( S = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)(x_n - \mu)^T \)) as column vectors, \( \Lambda_q \) is a diagonal matrix with the corresponding eigenvalues, and \( R \) is an arbitrary \( q \times q \) orthogonal rotation matrix. Tipping and Bishop prove that when \( W_q \) contains the principal eigenvectors of \( S \), global maximum of the likelihood occurs. So with \( U \) in Equation 10, the latent variable model defines...
a mapping from the latent space into the principal subspace of the observed data. The maximum likelihood estimator for noise variance $\sigma^2$ is given by the average variance lost for each discarded dimension, and can be formulated as

$$\sigma^2_{\text{ML}} = \frac{1}{d-q} \sum_{j=q+1}^{d} \lambda_j.$$  

(11)

To sum up, a Probabilistic PCA is obtained by finding the principal eigenvectors and eigenvalues of the sample covariance matrix $S$, while Gaussian distribution with sample mean $\mu$ and covariance matrix $\sigma^2 I + UU^T$, gives the density model.

Vowel-Consonant Segmentation as Hypothesis Testing

To bring the segmentation problem into a Hypothesis testing framework, we need to have a probability model for vowels and consonants. This can be accomplished by calculating $U$, using equation (10) and plugging $U$ into the covariance model given by $C = \sigma^2 I + UU^T$. Let us call the probability model for vowels and consonants as $C_v$ and $C_c$ respectively. We have,

$$C_{pv} = \sigma^2 I + U_v U_v^T$$

$$C_{pc} = \sigma^2 I + U_c U_c^T$$

(12)

We use GEVV ($W_y^{(v)}$, GEVC ($W_y^{(c)}$) and the corresponding eigenvalue matrices ($\Lambda_{y}^{(v)}$ & $\Lambda_{y}^{(c)}$) to obtain the parameters $U_v, U_c$, and $\sigma$ using equations (10) and (11). These parameters are in turn used to generate probability model $C_{pv}$ and $C_{pc}$ using equation 12. Assuming the $V$ and $C$ features to be zero mean Gaussian random processes with covariance matrices $C_{pv}$ and $C_{pc}$, the segmentation problem is to distinguish between the hypotheses

$$x \sim \left\{ \begin{array}{ll} N(0, C_{pv}) & \text{under } H_0 \\ N(0, C_{pc}) & \text{under } H_1 \end{array} \right.$$  

Here, $x$ is a zero mean test basic unit. A NP detector decides $H_1$ if the likelihood ratio exceeds a threshold or if

$$Lh(x) = \frac{p(x; H_1)}{p(x; H_0)} > \gamma$$

where,

$$p(x; H_1) = \frac{1}{\sqrt{2 \pi} C_{pv}} \exp\left[-\frac{1}{2} x^T C_{pv}^{-1} x\right]$$

$$p(x; H_0) = \frac{1}{\sqrt{2 \pi} C_{pc}} \exp\left[-\frac{1}{2} x^T C_{pc}^{-1} x\right]$$

$$Lh(x) = \left\{ \left( x^T C_{pv}^{-1} x - (x^T C_{pc}^{-1} x) \right) \right\} > 2\lambda \left[ \frac{\sqrt{C_{pv}}}{\sqrt{C_{pc}}} \right]$$

the log-likelihood ratio (LLR) becomes

$$Lh(x) = \left\{ (x^T C_{pv}^{-1} x - (x^T C_{pc}^{-1} x)) \right\} > 2\lambda \left[ \frac{\sqrt{C_{pv}}}{\sqrt{C_{pc}}} \right]$$

Hence, we decide $H_1$ if

$$T(x) = x^T (C_{pc}^{-1} - C_{pv}^{-1}) x > \gamma \left[ \frac{\sqrt{C_{pv}}}{\sqrt{C_{pc}}} \right]$$

(13)

Evaluation of the Threshold

By simplifying the test statistics $T(x)$ (eq. 13) into a standard distribution form, then the threshold ($\gamma$) can be evaluated [5]. Let $A = C_{pc}^{-1}$ and $B = C_{pv}^{-1}$, equation (13) becomes

$$T(x) = x^T (A - B) x > \gamma.$$  

(14)

Define $G = A^{-1} B$. Let the eigenvalue decomposition of $G$, be

$$GV = \Lambda V,$$

such that

$$V^T A V = \Lambda_1$$

$$V^T B V = \Lambda_2$$

(15)

where, $\Lambda$, $\Lambda_1$, and $\Lambda_2$ are diagonal matrices. Writing equation 15 in terms of $A$ and $B$, we have

$$A = V \Lambda_1 V^T$$

$$B = V \Lambda_2 V^T$$

(16)

Substituting eq. 16 into 14, we have

$$T(x) = x^T V [\Lambda_1 - \Lambda_2] V^T x.$$  

(17)

Let $y = V^T x$. Then equation 17 becomes

$$T(y) = y^T \tilde{A} y$$

(18)

where, $\tilde{A} = \Lambda_1 - \Lambda_2$. We know that $x$ is zero mean. Therefore $y$ is also zero mean, i.e., $E[y] = 0$ and the covariance of $y$ is

$$E[y y^T] = \left\{ \begin{array}{ll} V^T E[x x^T] V & \text{if } x \text{ is a } \text{vowel} \\ V^T C_{pc} V & \text{if } x \text{ is a } \text{consonant} \end{array} \right.$$  

If $x$ is a vowel, then $C_{xx} = C_{pv}$ and $V^T C_{pv} V = \Lambda_1^{-1}$. Similarly, if $x$ is a consonant, then $C_{xx} = C_{pc}$ and $V^T C_{pc} V = \Lambda_2^{-1}$. So, we can write equation 18 as

$$T(y) = \sum_{n=1}^{N} \frac{y_n^2(n)}{\eta^2(n)}$$

(19)

where $\left\{ \frac{1}{\eta^2(1)}, \frac{1}{\eta^2(2)}, \ldots, \frac{1}{\eta^2(N)} \right\}$ are diagonal elements of $\tilde{A}$. Because $y$ is i.i.d., $T(y)$ is $\chi^2$ (Chi-square) distributed and the detector turns out to be an energy detector. The required threshold, $\gamma$, for segmentation is computed by using an optimization algorithm for the fixed probability of false alarm $P_{FA}[5]$.

$$P_{FA} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \prod_{n=1}^{N} \frac{1}{\sqrt{1 - 2 j \alpha n \omega}} \exp(-j \omega t) \frac{dw}{2\pi}$$

(20)

$$P_D = \frac{1}{2\pi} \int_{-\infty}^{\infty} \prod_{n=1}^{N} \exp(-j \omega t) \frac{dw}{2\pi} dt.$$  

(21)

where,

$$\alpha = \frac{1}{\eta^2(n)}.$$  

Now, the decision (segmentation) rule is
Speech segmentation experiments were conducted on a Kannada speech database spoken by a female volunteer. GEVV's and GEVC's were obtained from a Tamil speech database spoken by a male volunteer using the method discussed in section 2. Feature vectors were obtained for each frame of a test basic-unit. Duration of each frame of speech was 30 ms, with an overlap of 20 ms between successive frames. Each frame of speech was Hamming windowed and processed to yield a 13-dimensional feature vector. The feature vectors used were MCC and LPCC. We have seen in Fig. 1(b) that energy based segmentation fails to identify the co-articulated consonant region. On the other hand, in Fig. 2, the same consonant region has been correctly identified using subspace-based method. Figure 2(c) also displays the spectrogram of the basic-unit /eyol/. Here, spectrogram shows clear second formant transition from the front vowel /e/ to the back vowel /o/ and the transition region corresponds to the consonant /l/.

Basic-unit considered in Fig. 3 is VCCV and both C's are glide and nasal (/l/ and /m/).

Probability models for V and C were obtained using equations 10, 11 and 12. Here, the observed feature vector x, and its dimension is 13, the latent variable x has dimension of 3. i.e., d = 13 and q = 3. Dimension of the latent variable x has been chosen based on the dimensions of GEVV and GEVC. The multiple crossovers present some uncertainty in classifying the basic-units into vowel and consonant parts. By employing classical decision theory, we can remove this uncertainty by choosing a threshold on a statistical basis. In plot 4(d), the basic-unit is seen to be clearly segmented. In Fig. 4(b), there are multiple crossovers and using the class decision approach shown in Fig. 4(d), we can eliminate false crossovers. We used fixed \( P_{FA} = 10^{-4} \) to calculate the threshold, \( \gamma_{II} \). Figure 4(e) shows the spectrogram of /jumal/. Figure 5(a) shows the co-articulated basic-unit /llal/. In Fig. 5(b) we can see the multiple crossovers of the norm-contours. Figure 5(c) shows the test statistics along with the threshold and V & C decision is shown in Fig. 5(d). Figure 5(e) shows the spectrogram of /oolal/.

### 5. Conclusion

We have presented subspace and hypothesis based methods for co-articulated basic-unit segmentation. The first method uses crossovers of norm-contours for segmentation. Norm-contours give a measure of energy in the projected subspaces. There may be an ambiguity when there are multiple crossovers. In the second method this ambiguity is resolved by finding a statistical threshold. The test statistic using NP criterion turns out to be an energy detector as in the first method but without any ambiguity in segmentation. PPCA method can be further extended to classify within the vowel or consonant segments in a basic-unit using mixture-PPCA [6].

### REFERENCES


