and Popular databases containing the interest of this hybrid approach.

Phase, which enhances the results of the SVM models. Experiments carried out on the XNryTS
method is also used in the baseline system to increase the number of clear sources in the training
subset. Support Vector Machines (SVMs) usually yield better performance in multi-classification problems and can construct linear-

Abstrac: Gaussian Mixture Models (GMMs) are known to be the dominant ap-

February 2003

Qun Lu, Samy Bengio

VERIFICATION

CLIENT DEPENDENT GNMI-SVM MODELS FOR SPEAKER

IIIAIP Research Report 03-03
(3)

\[
\varphi = \frac{(S_d)}{(S_d)^2} \frac{(m_d)}{(m_d)^2} \leq \frac{(x_i)}{(x_i)^2} = \varphi \cdot \varphi
\]

Inequality (1) inequality can be rewritten as

\[
(x_i) \leq (x_i) \leq (x_i)
\]

(2)

If \(x \) is a random variable by \( \mu \), then \( |x| \cdot d \) is also a random variable by \( \mu \).

(1)

\[
\frac{(x)^d}{(s)^d} \leq \frac{(x)^d}{(s)^d}
\]

where \( \mu \) is the probability distribution of \( x \).

The speaker verification problem can be considered as a statistical hypothesis testing problem where

The Baseline Speaker Verification System

2.1 The Baseline Speaker Verification System

In this chapter, we describe empirical design of the baseline speaker verification system. We compare and contrast the performance of the baseline systems on the NIST 2000 and 2001 Speaker Verification Test Sets. We also discuss the potential improvements and limitations of the baseline systems. We conclude with a brief discussion of future research directions in speaker verification.
3.1 Supports Vector Machines

Hybrid System

Two different errors are observable: the speaker changes distribution, the performance of the system is often measured in terms of these two different errors as shown above. In order to select a decision threshold (\( \alpha \)), the system is often trained on a validation set to

(6)

\[
\frac{\text{HTER}}{\text{HTER} + \text{FPR}} = \frac{1}{2}
\]

Half Total Error Rate (HTER)

Various evaluation measures can be constructed based on FPR and FRR. In this paper, we used the

(8)

\[
\frac{\text{FPR}}{\text{FRR}} = \frac{\text{number of correct accesses}}{\text{number of accesses}}
\]

(7)

\[
\frac{\text{FRR}}{\text{FPR}} = \frac{\text{number of correct accesses}}{\text{number of accesses}}
\]

In general, diagonal-covariance measures are used in order to fit the model size.

(9)

\[
\left(\begin{array}{c}
\mathbf{x} - \mathbf{x}_j \\
\mathbf{y} - \mathbf{y}_j \\
\mathbf{z} - \mathbf{z}_j
\end{array}\right) N = \sum_{i=1}^{N} \sum_{d} \mathbf{w}_d = \mathbf{w}_d N
\]

where the parameter set of the CDA is the same as that of the model.

(2)

\[
\left(\begin{array}{c}
\mathbf{x} - \mathbf{x}_j \\
\mathbf{y} - \mathbf{y}_j \\
\mathbf{z} - \mathbf{z}_j
\end{array}\right) N = \sum_{i=1}^{N} \sum_{d} \mathbf{w}_d = \mathbf{w}_d N
\]

CDA can be computed as follows

\[\mathbf{X} = \mathbf{X}_j \implies \mathbf{Y} = \mathbf{Y}_j \implies \mathbf{Z} = \mathbf{Z}_j \]

The distribution of frame vectors \( \mathbf{x} \) extracted from a speaker's speech is often modeled by a Gaussian

(4)

\[
\text{HTER} = \frac{1}{2} \log < (\mathbf{x}) d > - \frac{1}{2} \log < (\mathbf{x}) d > - \frac{1}{2} \log < (\mathbf{x}) d > - \frac{1}{2} \log < (\mathbf{x}) d >
\]

Let's take a look at the leftmost equality (9) leads to the inequality:

Since it is more convenient to deal with log-Euclidean nose statistics rather than Euclidean nose statistics.

IDIP-PR 02-03
Algorithm for SVMs is quadratic on the number of examples.
and also scale in the training time for the decision-making model since the complexity of the
underlying classiﬁer can be exponential. Speciﬁcally, SVMs will handle more decision-making
tasks in each partition of the data.

A tree structure is used to learn the structure of the data. The tree is built layer by layer,
and each layer represents a decision point. The SVM is then trained on the data that
is closest to the decision point.

A decision function is learned at each layer, which is used to determine the classification of
the data. The decision function is learned using a binary tree structure, where each
node represents a decision point.

In most cases, SVMs are used to classify data, and the decision function is learned using
the underlying classiﬁer.

3.2 Postprocessing SVMs Scores

The postprocessing step is used to combine the scores from the SVMs.

\[
q + (x' \cdot x) M + \alpha_0 = \hat{y} \tag{10}
\]

where \( q \) is the response from the numerical data, \( x \) is the feature vector, \( \hat{y} \) is the desired class,
and \( M \) is the learned model. The decision function is learned using a binary tree structure,
where each node represents a decision point.

The decision function is learned using a binary tree structure, where each
node represents a decision point.

\[
\sum_{i=1}^{n} \frac{1}{\hat{y}_i} = \hat{w} \tag{11}
\]

where \( \hat{w} \) is the weight vector of the decision function.

The decision function is learned using a binary tree structure, where each
node represents a decision point.

To learn the decision function, a binary tree structure is used, where each
node represents a decision point.

The decision function is learned using a binary tree structure, where each
token represents a decision point.

The decision function is learned using a binary tree structure, where each
token represents a decision point.

The decision function is learned using a binary tree structure, where each
token represents a decision point.

The decision function is learned using a binary tree structure, where each
token represents a decision point.

The decision function is learned using a binary tree structure, where each
token represents a decision point.

The decision function is learned using a binary tree structure, where each
token represents a decision point.

The decision function is learned using a binary tree structure, where each
token represents a decision point.

The decision function is learned using a binary tree structure, where each
token represents a decision point.
4 Experiments

Other data for training a more accurate SVM model.

These additional data sets can be used for the same purpose; hence, we simplify this with

set.

The scores obtained from 500 runs (each 30 seconds) were used to calibrate the SVM model. The

classification (PFC) coefficients were extracted, and the top five decisions for each

The speech signal was sampled every 10 ms and then concatenated into linear frequency

4.2 Results from Text-Independent Experiments

To the clean and the other 10 sessions for data set only.

50 sessions were used. 10 of them (called training clean data set) for determining the SVM model.

The classification error rates of each speaker, when one recording session contains 17 words. For each
differences are then used to train the classifier. The SVM classification error rate was

In the first experiment, we used the phone dependent database (28), they contain two sets (called

database Description

4.1 The Phone Database

...
4.2 Results

The model results demonstrate that the proposed decision method is effective in improving the performance of different validation models. The decision method was first applied to the two session decision model, then extended to the three session decision model, and finally to the four session decision model. The results show that the proposed decision method significantly improves the performance of all three models. The decision method was also applied to the two session decision model, and the results show that the decision method improves the performance of this model as well.

In a similar study, we used the XFIATS database and its associated experimental data to evaluate the performance of our proposed decision method. The results show that our proposed decision method significantly improves the performance of the XFIATS database.

<table>
<thead>
<tr>
<th>Table 1: Results from the Polynomial database.</th>
</tr>
</thead>
<tbody>
<tr>
<td>System</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Polynomial baseline</td>
</tr>
<tr>
<td>Polynomial improved</td>
</tr>
<tr>
<td>Polynomial improved (best)</td>
</tr>
</tbody>
</table>

Specific hybrid SVM model yields the best performance.
AVPPA 1999,
In Second International Conference on Audio and Video-Based Person Authentication


222 1997.
S. Bremer. Recent advances in speaker recognition. Lecture Notes in Computer Science, 1266-237.

the EN Acoustic. Final of the Second International Society 1.1.97.

In this paper we propose the use of a cross-validation technique to increase the number of data points.

Conclusions

Table 2: Results from the XAViTS database.

<table>
<thead>
<tr>
<th>System</th>
<th>SVM Univariate SV</th>
<th>SVM Multivariate SV</th>
<th>900 Gaussians baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.97</td>
<td>1.00</td>
<td>1.05</td>
</tr>
<tr>
<td>Cross-validation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

obtained here is the best ever reported on this subject.

The results from the tables show that the cross-validation system performs better than the baseline system not having enough data to obtain a separate threshold per channel (within the cross-validation process), we did not have enough data to obtain a separate threshold per channel (within the cross-validation process). Because these are only real thresholds for one channel in the database. More than 200 support vectors were used for each access, and the performance is then measured on the least data (1000 support vectors and 10000 support vectors, respectively). The results show that the cross-validation system performs slightly better than the baseline system, in the cross-validation process, we simply compute the cross-validation process. The cross-validation technique is used to create clean scores. The results

References

experiments in the cross-validation process. The cross-validation process.
1995.

V. N. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, New York, NY, USA.


Douglas A. Reynolds, Thomas F. Quatieri, and Robert B. Dunn. Speaker verification using adapted