Abstract. In this paper, we present a new approach towards user-customized password speaker verification based on HMM/ANN and GMM models.
Contents

1 Introduction

2 SV-OCO Decision Rules

3 Database and Acoustic Features

4 Approaches

5 Experiments and Results

6 Discussion

7 Conclusion
1. The position probability of any speaker's pronunciation at any word \( p \) is defined as follows:

\[
(p_1 t_{-1} w_j) < \gamma < (p_{-1} t_{-2} w_j) \]

2. **SV-CCP Decision Rules**

    In Section 1, we investigate the problem of estimating the speaker probability of any word \( p \). In Section 2, we describe the evidential eschatological and the insidious evidence of the two methods. In Section 3, we describe the decision-making process. In Section 4, we present a detailed description of the two methods. In the following Sections 5, 6, and 7, we introduce the empirical measures that will be used here, which are presented in Section 6.

   \( \text{LCP} \) of acoustic vector sequences, X. Do not verify decision, we thus want to compare the position probability of the speaker. We refer to the position probability of any word \( p \). SV-CCP represents the empirical measures that will be used here.
3 Database and Acoustic Features

Two databases were used in this work: The Swiss Phonetic Database (SPD) and the NIST Speaker Recognition Database (SRD).

The SPD database is a large, speaker-independent, AN database used to train ANNN and ANNN/GMM. It contains a large number of speakers, and the data is recorded under a variety of conditions. The SRD database is a smaller, speaker-dependent, database used for testing speaker recognition systems.

The model used to estimate the speaker's gender is the speaker-independent ANNN/GMM model. This model is trained on the training data, and then used to estimate the gender of the speaker in the test data.

The following equations are used to estimate the gender of the speaker:

$$\mathcal{S} = \left( \frac{\gamma_s}{\gamma_i} \right) \mathcal{S} + \left( \frac{\gamma_i}{\gamma_i} \right) \mathcal{S}$$

where \(\gamma_s\) and \(\gamma_i\) are the decision thresholds.

The decision rule is:

$$\mathcal{S} = \left( \frac{\gamma_s}{\gamma_i} \right) \mathcal{S} + \left( \frac{\gamma_i}{\gamma_i} \right) \mathcal{S}$$

Using Bayes rule, decision rule (1) can be rewritten as:

$$\mathcal{S} = \left( \frac{\gamma_s}{\gamma_i} \right) \mathcal{S} + \left( \frac{\gamma_i}{\gamma_i} \right) \mathcal{S}$$

where \(\gamma_s\) is represented by an equal ANNN/GMM. Similarly, we could also use the following equation:
The topology of the resilient recurrent neural network (RNN) model is shown by the blue lines in the diagram. We represent the dynamic symbol associated with $x$ as follows:

\[ \Theta^{(n)} \sum_{i=1}^{N} \frac{x_i}{N} \]

The posterior probability $p(x'|x)$ follows the joint RNN model's defined representation. From the posterior representations, we select the most plausible hidden state. It is important to note that the selected hidden state is not unique, as there may be multiple states that are equally probable.

Each new customer participates in the RNN process. The RNN generates 68% of its input data, and 50% of the input data is generated at random. The RNN is trained using the RNN (h) parameter, which is chosen at random and initialized to a specific value.

We start from a well-trained RNN. A multilayer feedforward network (MFFN) was used to perform the RNN inference. The MFFN was trained on a dataset with 50% of the input data, and 50% of the output data was generated at random.

Our spatial-temporal experiments were conducted using the POPVC database [9], which is an

\[ \begin{align*}
&\text{Approaches} \\
&\text{Calculating every 1 dimension over 30 ms windows, matching in 25 coordinates.} \\
&\text{Next, the process is repeated for the next 15 coordinates.} \\
&\text{Finally, the process is repeated for the next 15 coordinates.} \\
&\text{The spatial-temporal RNN was used to perform the RNN inference.} \\
&\text{The RNN was trained on a dataset with 50% of the input data, and 50% of the output data was generated at random.} \\
&\text{The RNN was trained on a dataset with 50% of the input data, and 50% of the output data was generated at random.} \\
&\text{The RNN was trained on a dataset with 50% of the input data, and 50% of the output data was generated at random.} \\
\end{align*} \]
4.3 Constrained and constrained HNN

43.2 Combined HNN/VNN-CWM approach

\[ q \leq \left[ (\lambda X)_{d} \log - (\lambda X)_{d} \log + (X_{\alpha}^{\beta} \lambda X)_{d} \log \right] \frac{N}{1} \]

where \( N \) is the length of the least access after sparse frames have been removed.

\( \frac{(u_{x}^T X_{d})_{d} \sum_{i=1}^{N} (X_{d} - \mu)^{2}}{N} + \sqrt{\epsilon} \theta = \% \)

where \% is the mean of the Gaussian for \( \theta \) not set. \% is the corresponding mean in the

\[ \text{Sparsity characteristics} \]

used as deprivation data, while the last two variables are used as cross-validation.

\[ \text{CWM-CW} \]

used as deprivation data, with the last two variables are used as cross-validation.

\[ \text{CWM-CW} \]

used as deprivation data, with the last two variables are used as cross-validation.

\[ \text{CWM-CW} \]

used as deprivation data, with the last two variables are used as cross-validation.

\[ \text{CWM-CW} \]

used as deprivation data, with the last two variables are used as cross-validation.

\[ \text{CWM-CW} \]

used as deprivation data, with the last two variables are used as cross-validation.

\[ \text{CWM-CW} \]

used as deprivation data, with the last two variables are used as cross-validation.

\[ \text{CWM-CW} \]

used as deprivation data, with the last two variables are used as cross-validation.

\[ \text{CWM-CW} \]

used as deprivation data, with the last two variables are used as cross-validation.

\[ \text{CWM-CW} \]

used as deprivation data, with the last two variables are used as cross-validation.
Table I: Found error rate for constrained and unconstrained HMM and combined HMM/VHN-GMM

<table>
<thead>
<tr>
<th>Methods</th>
<th>2.1%</th>
<th>2.9%</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMM/VHN-GMM</td>
<td>2.1%</td>
<td>2.9%</td>
</tr>
<tr>
<td>Unconstrained HMM</td>
<td>1.9%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Constrained HMM</td>
<td>0.9%</td>
<td>1.1%</td>
</tr>
</tbody>
</table>

When the correct and the matched phonetic transcription of the password

were accepted and best recognition, the results of this experiment are given in Table I.

The first section of the experiment consists of the evaluation of the performance of the system when

2.1. Clients with the same password

The password is also given.

3. Experiments and Results

been removed.

For the constrained HMM method, and \( N \) is the length of the test access after the silence frames have

\[ q \geq \frac{\left| \sum_{y^n} P(y^n|x) P(y^n|x)^T \right|}{N} \]

For the unconstrained HMM method, this ratio which is compared to a speaker independent threshold

is a DB-HMM model. For the verification process we have used the normalized log likelihood

ratio of the HMM model of the speaker. The test consists of triggering the model of the speaker

consistency, the speaker is the target, and the recognition process uses the model of the

collection. Once the best HMM model is identified, a VAD adaptation procedure is performed,

before the VAD selection. HMM models are trained using phone database and VADVCC

IDVIP-PF 04-10
A new method for speaker verification based on user-enrolled passwords is proposed. This method uses cross-lingual verification results to enhance the recognition accuracy.

## Conclusion

The correct password is only 61%. The results of cross-lingual verification show that the proposed method can significantly improve the recognition accuracy. The recognition rate of the proposed method is 98%, which is significantly higher than the results of previous methods. This method is expected to be widely used in practical applications.

## Discussion

### Table 6: Error rates for the three methods with different passwords

<table>
<thead>
<tr>
<th>Models</th>
<th>Correct password</th>
<th>Incorrect password</th>
<th>HNN</th>
<th>AVNN</th>
<th>GANH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>3.9</td>
<td>3.9</td>
<td>3.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>4.1</td>
<td>4.1</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2.6</td>
<td>2.6</td>
<td>2.6</td>
</tr>
</tbody>
</table>

In conclusion, the proposed method shows promising results in cross-lingual verification. Further research is needed to improve the recognition accuracy and to explore the performance of the proposed method in different scenarios.
References


