# LIST OF TABLES

Table 1. A comparison of classification error rates is shown for a GMM system using static+delta features and a MixAR system operating only on the static features is shown on synthetic data. The number of parameters for each system is shown in parentheses. The GMM system, which uses static and delta features, performs significantly worse than the MixAR system as the nonlinearity in the data increases.

Table 2. Speaker verification EERs are shown for MixAR and GMM for a variety of feature vector combinations. MixAR does not need delta features since the model itself encodes temporal dynamics.

Table 3. EERs are shown as a function of the number of mixtures. MixAR performs slightly better with almost half the number of parameters.

Table 4. EERs are shown for a variety of noise conditions.

Table 5. EERs are shown for TIMIT (clean data) and NTIMIT (noisy data).

Table 6. Performance is analyzed as a function of the duration of the training data utterances. The evaluation utterance durations were held constant and varied between 20 and 40 seconds.

Table 7. Performance is analyzed as a function of the duration of the evaluation data is shown. The training utterance duration was fixed and averaged around 120s.

# Tables

|  |  |  |
| --- | --- | --- |
| ***α*** | **GMM (8 mix.) Static+∆** | **MixAR (4-mix.)**  **Static** |
| 0.00 | 1.50 (288) | 1.50 (240) |
| 0.25 | 3.25 (576) | 3.50 (240) |
| 0.50 | 10.25 (576) | 6.25 (240) |
| 0.75 | 24.75 (576) | 9.75 (240) |
| 1.00 | 26.75 (576) | 13.75 (240) |

Table . A comparison of classification error rates is shown for a GMM system using static+delta features and a MixAR system operating only on the static features is shown on synthetic data. The number of parameters for each system is shown in parentheses. The GMM system, which uses static and delta features, performs significantly worse than the MixAR system as the nonlinearity in the data increases.

|  |  |  |
| --- | --- | --- |
| **No. Mixtures** | **GMM**  **Static+∆+∆∆** | **MixAR**  **Static Only** |
| 2 | 23.1 (216) | 24.1(120) |
| 4 | 21.7 (432) | 19.2(240) |
| 8 | 20.5 (864) | 19.1(480) |
| 16 | 20.5 (1728) | 19.2(960) |

Table . EERs are shown as a function of the number of mixtures. MixAR performs slightly better with almost half the number of parameters.

|  |  |  |
| --- | --- | --- |
| **Features** | **GMM 16-mix.** | **MixAR**  **8-mix.** |
| Static(12) | 22.1 | 19.1 |
| Static+E(13) | 33.1 | 41.1 |
| Static+Δ(24) | 20.6 | 20.4 |
| Static+Δ+ΔΔ(36) | 20.5 | 20.5 |

Table . Speaker verification EERs are shown for MixAR and GMM for a variety of feature vector combinations. MixAR does not need delta features since the model itself encodes temporal dynamics.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **GMM (1168)** | **SNR (dB)** | **Clean** | **Car Noise** | **White Noise** | **Babble Noise** |
|  | **2.4** |  | | |
| 10 dB |  | 19.7 | 48.7 | 40.6 |
| 5 dB |  | 31.2 | 50.0 | 44.7 |
| 0 dB |  | 39.3 | 49.8 | 48.2 |
| **MixAR (480)** |  | **1.8** |  | | |
| 10 dB |  | 13.7 | 47.0 | 36.9 |
| 5 dB |  | 23.2 | 47.6 | 42.8 |
| 0dB |  | 33.9 | 48.5 | 47.6 |

Table . EERs are shown for a variety of noise conditions.

|  |  |  |
| --- | --- | --- |
|  | **Training Utterance Duration** | **EER** |
| **GMM**  **(864)** | 120 | 20.5 |
| 90 | 20.4 |
| 60 | 20.4 |
| 30 | 24.4 |
| 15 | 29.5 |
| **MixAR**  **(480)** | 120 | 19.2 |
| 90 | 21.5 |
| 60 | 21.8 |
| 30 | 21.8 |
| 15 | 24.3 |

Table . Performance is analyzed as a function of the duration of the training data utterances. The evaluation utterance durations were held constant and varied between 20 and 40 seconds.

|  |  |  |
| --- | --- | --- |
| **Database** | **GMM  Static+∆+∆∆ (1728)** | **MixAR  Static Only (480)** |
| TIMIT | 2.4 | 1.8 |
| NTIMIT | 21.0 | 20.9 |

Table . EERs are shown for TIMIT (clean data) and NTIMIT (noisy data).

|  |  |  |
| --- | --- | --- |
|  | **Evaluation Utterance Duration** | **EER** |
| **GMM**  **(864)** | 30 | 20.5 |
| 15 | 21.8 |
| 10 | 21.5 |
| 5 | 24.4 |
| 3 | 26.9 |
| **MixAR**  **(480)** | 30 | 19.2 |
| 15 | 23.4 |
| 10 | 23.1 |
| 5 | 25.6 |
| 3 | 25.6 |

Table . Performance is analyzed as a function of the duration of the evaluation data is shown. The training utterance duration was fixed and averaged around 120s.