Password-Based Voice Verification Using SpeakerKey

A. Higgins, L. Bahler

ITT Industries
Goal

• Assess performance of SpeakerKey using single-word passwords over the telephone.
Speech Data

- Polyvar/Picasso
- 17 French passwords
- Enrollment: 5 repetitions, Test: 1 repetition
- Possible multiple telephones
- 38 client speakers (24m, 14f)
- 56 world-model speakers (28m, 28f)
System Description

• Features derived from 21-channel filterbank.
• Stochastic matching relative to speaker-independent HMM word models.
• Channel-compensation de-weighting factor: $0 \leq \alpha \leq 1$.
• Speaker likelihoods estimated using nearest-neighbor distance measure:

$$L(U \mid E) \cong \frac{1}{|U|} \sum_{u \in U} \min_{e \in E} (d_{ue})$$
System Description (continued)

- Euclidean distance $d_{ue}$ between unknown frame $u$ and enrollment frame $e$.
- Each test frame compared only to enrollment frames aligned with the same HMM node.
- World-model score is multiplied by $0 \leq \beta \leq 1$ before forming log likelihood ratio.
Equal-Error Rates

• Using $\alpha = \beta = 1.0$, and applying the same threshold to all words, EER = 5.2%.

• With some modifications, this was improved to EER = 3.1%.
Algorithm Modifications

• Reduced the effect of channel deconvolution by setting $\alpha = 0.3$.
• Modified the log likelihood ratio score by setting $\beta = 0.7$.
• Replaced Euclidean inter-frame distance with node-dependent full-covariance distance*.

* Covariances trained on client data.
Error-Rate Improvements

- Original: 5.2%
- Optimized Channel Deconvolution: 4.1%
- Optimized Scoring: 3.6%
- Covariance Modeling: 3.1%
Equal-Error Rates by Word

- suivant: 3.7
- quitter: 5.5
- precedent: 2.7
- musee: 5.5
- mode d'emploi: 2.7
- message: 2.4
- manifestation: 1.3
- Louis Moret: 3
- guide: 5.1
- gianadda: 3.1
- galerie du Manoir: 1.3
- exposition: 2.4
- corso: 2.4
- concert: 3
- cinema: 2.3
- casino: 2.1
- annulation: 1.7
EERs Sorted by Word Length
Conclusions

• Error rates from 1.7% to 5.5%, inversely correlated with number of syllables in password.
• Median password length = 3 syllables.
• Median error rate = 2.7%.
• The steps taken to optimize performance are appropriate for relatively benign conditions.