**INTRODUCTION**

**What is it?**
- A database of utterances in Spanish of isolated digits recorded with two different smartphones.
- It is intended for research on biometrics and secure applications that integrate ASR and speaker recognition/verification.

**Motivation**
- The importance of speech-related tasks in smartphones and remote secure systems.

**DATA RECORDING**
- Single-channel utterances of isolated digits (0 to 9).
- Two smartphones: high-range (Sony Xperia S) and mid-range (HTC WildFire).
- Three sessions of ten minutes each. Far-talk conditions.
- An enrollment utterance (10 digits) and six verification utterances (4 digits) per session and smartphone.
- Speech recording in a rather silent office of 12 m². Average SNR of 32.9 dB.

A total of 169 speakers, most of them from Eastern Andalusia.

**STRUCTURE OF THE DATABASE**
- Corpus of 42 utterances/speaker (in all, 7098 utterances).
- Two different datasets: ENROLL and VERIF (with the enrollment and verification utterances, respectively).
- WAV files containing the speech utterances.
- XML annotation files: speaker annotation files (one per speaker) and utterance annotation files (one per WAV file).

**EVALUATION: SPEECH RECOGNITION**

**Experimental framework**
- We use the ETSI front-end to extract MFCC features.
- Speakers are divided in two subsets: A (100 speakers) and B (69 speakers).
- Subset A (both ENROLL and VERIF datasets) is used to train GMM-HMM acoustic models.

Three methods for evaluating the recognition accuracy:
1. VERIF dataset of subset B is used for testing.
2. Same as above, but CMVN is applied to both training and testing features.
3. Same as above, but ENROLL dataset of subset B is used to perform speaker adaptive training by means of MLLR.

**Results**

<table>
<thead>
<tr>
<th>Method</th>
<th>WAcc (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFCC</td>
<td>99.67</td>
</tr>
<tr>
<td>MFCC+CMVN</td>
<td>99.62</td>
</tr>
<tr>
<td>MFCC+CMVN+MLLR</td>
<td>99.84</td>
</tr>
</tbody>
</table>

**EVALUATION: SPEAKER VERIFICATION**

**Experimental framework**
- Front-end: VAD, pre-emphasis filtering, MFCC and CMVN.

**Results**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>EER (NIST 2008)</th>
<th>minDCF (NIST 2008)</th>
<th>minDCF (NIST 2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>0.69%</td>
<td>0.45%</td>
<td>0.12%</td>
</tr>
</tbody>
</table>

**CONCLUSION**

- The SecuVoice database has been described, and both speech recognition and speaker verification results are given as baseline. Speech researchers can evaluate and compare the performance of their own algorithms within this framework.
- SecuVoice is available through ELRA (European Language Resources Association).

**CONTACT INFORMATION**

Juan M. Martín-Doñas, I. López-Espejo, A. M. Gomez, A. M. Peinado
E-mail: mdjuanmart@correo.ugr.es, iloes@ugr.es, amg@ugr.es, amp@ugr.es
Dept. of Signal Theory, Telematics and Communications, University of Granada, Spain