Speech Biometric

Sunil Kumar Kopparapu

`SunilKumar.Kopparapu@TCS.COM`

Speech and Natural Language Processing Group

TCS Innovation Labs - Mumbai

Advanced Technology Applications Group,

Yantra Park, Thane (West), Maharastra

May 2007

Information in Speech

- **non-linguistic**, (*who said it*)
  gender, emotional states, speaker name

- **linguistic** (*what he said*)
  Language name and what was said (written language)

- **paralinguistic** (*how well said* – manner, clarity or accent, aspects related to quality)
  deliberately added by the speaker, and not inferable from the written text.

Goal: Automatically extract information in speech signal

Speaker Recognition Vs Verification

Gallery	Who spoke?
/Edna/	
/Sunil/	
/Akhilesh/	/????/
/Dipti/	
/Devanuj/	

Response: Sunil

Speaker Recognition Vs Verification

/????/ | **Is this Sunil?**

/????/ | **Is this Sunil?**

Response: Yes or No

Will concentrate on Speaker Verification ....

Speaker Verification: Overview

- Is the process of verifying the claimed identity of a registered speaker using his voice characteristics.

- The speaker needs to enroll before using the system.

- During enrollment, the speaker speaks a given set of utterances, using which the systems builds statistical (or deterministic) models characterizing the speaker's voice.

- A user claims he is X. Speaks a pass-phrase. The system gives a binary output YES (accept claimed identity) | NO.

**Need for threshold to be able to say Yes or No.**

Types of Speaker Verification

1. **Fixed Phrase** – pre-determined phrase used for verification

2. **Fixed Vocabulary** – verification more flexible and practical; training and testing materials for a speaker are generated based on words of a fixed vocabulary

3. **Flexible Vocabulary** – a general set of sub-word phone models is created during speaker model training

4. **Text-Independent** – user is not constrained to say fixed or prompted phrases

Clearly, both complexity and security increases as we go from fixed phrase to text-independent.

Speaker Verification Terms

- FAR - False Acceptance Ratio (The percentage of **incorrect successful** verifications)

- FRR - False Rejection Ratio (The percentage of **incorrect failed** verifications)

- EER - Equal Error Rate (The value at which FAR equals FRR)
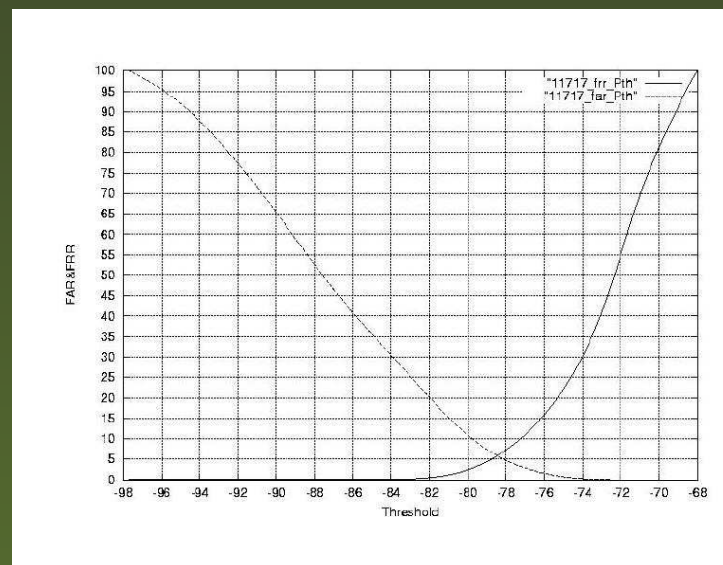
Our Objective

- Utilize the speech signal processing knowledge to research and build in-house a state of the art speech recognition, synthesis, verification engine

- Understand and explore aspects of speech recognition technology which can be used to enhance the performance in actual use

- Utilize in-house engine to custom build applications, rather than use an off the shelf speech recognition system from a third party vendor as a black-box.

Our Speaker Verification System [1]

- Uses the state-of-the-art speaker verification engine tuned for telephone speech

Our Speaker Verification System [1]

- Uses the state-of-the-art speaker verification engine tuned for telephone speech

- Easily customizable and configurable and can work even for a large company setup with thousands of employees. It can easily be integrated into any existing employee database of the company.

Our Speaker Verification System [1]

- Uses the state-of-the-art speaker verification engine tuned for telephone speech

- Easily customizable and configurable and can work even for a large company setup with thousands of employees. It can easily be integrated into any existing employee database of the company.

- Functional for attendance monitoring at several locations of TCS

Our Speaker Verification System [2]

■ System accessible over the telephone line (EPBAX) using telephony card interface.

Our Speaker Verification System [2]

- System accessible over the telephone line (EPBAX) using telephony card interface.

- Speaker verification engine, conceived and engineered in-house.

Our Speaker Verification System [2]

- System accessible over the telephone line (EPBAX) using telephony card interface.

- Speaker verification engine, conceived and engineered in-house.

- Select speech feature set which captures the identity of the speaker makes it very robust with very low FRR and FAR

Performance

	T_1	T_2
FAR	6.47%	1.75%
FRR	0.95%	10.90%

- Threshold (T_1) was chosen to be such that the FRR was close to 0% (pass all) and T_2 was chosen so that FRR was approximately 10%.

- Experiments carried out on a set of 15 speakers. Imposter's aware of the pass phrase (i.e. skilled forgery)

- 25 Parameters; Continuous HMM models used

- Ported and tested on BREW SDK2.0 simulator

Thank You