Speaker Identification in Mobile Environment

Anil Kumar Vuppala

January 24, 2014

IIIT-HYD

Outline

Introduction

- The rapid growth of mobile users
- Exciting services of Speech systems for mobile devices
  1. Speech input interface
  2. Automatic dictation of documents
  3. Secured remote transactions
  4. Voice based person authentication
  5. Forensic investigation
- Major issues involved in wireless environment
  1. Effect of varying background conditions
  2. Degradations due to different speech coders
  3. Channel impairments

Introduction

- The rapid growth of mobile users.
- India had 851.70 million cellphone users at the end of June 2011.

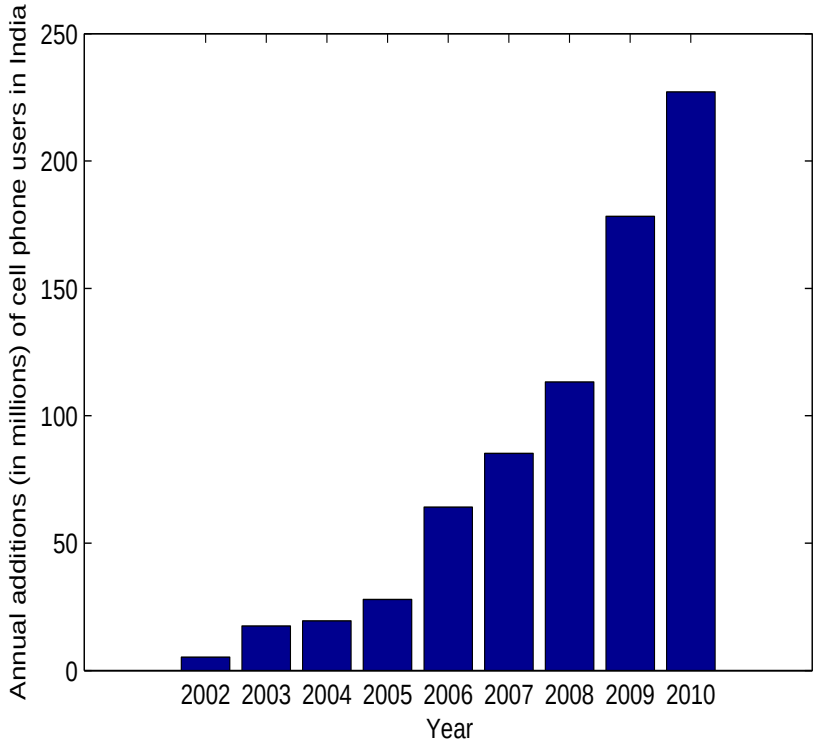

Figure: *Annual addition (in millions) of cellphone users in India* [Wikipedia Telecommunications Statistics in India]

Speech System Configurations in Wireless Environment:

- Embedded Speech systems (client-based)
- Network Speech systems (server-based)
- Distributed Speech systems (client-server-based)

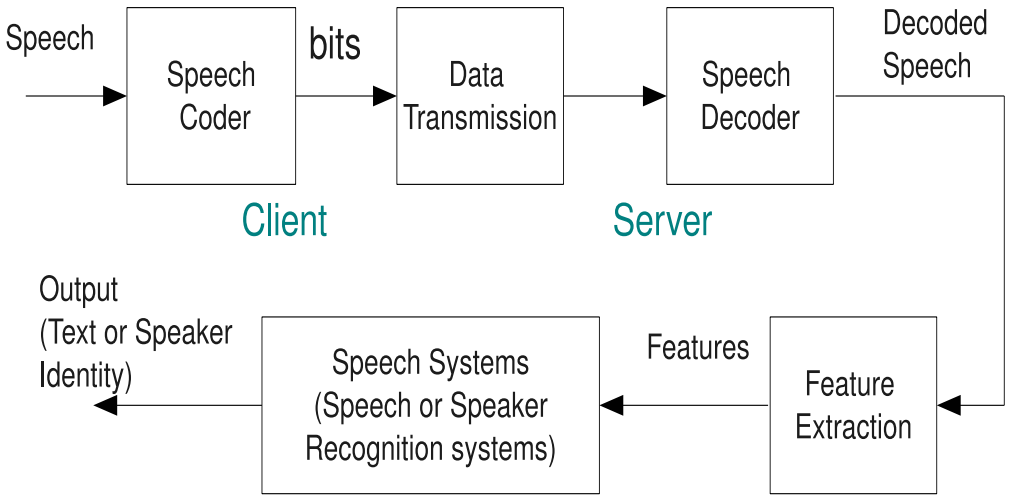

Figure: *Block diagram of speech systems in wireless network environment*

Speaker recognition

- Speaker Recognition: Recognizing speakers by extracting and modeling signal processing features from the speech signal.
- Classification
  - Speaker verification v/s Speaker identification
    - Verifying the identity claim of a speaker (I am so and so, please allow me to use)
    - Identifying the speaker of the speech signal(I will not tell who I am, bet you identify me)
  - Text-dependent v/s Text-independent
  - Closed set v/s Open set

Literature survey

Table: Speaker recognition under coding

Reference	Major outcomes	Comments
[QSD+99]	Studied the effect of coding on speaker recognition.	Study
[RTDJ99]	Effect of speech coding on speaker recognition when training and testing conditions are matched and mismatched	Study
[RTDJ01]	Improving speaker recognition performance under coding using score normalization	Scope for improvement
[PG05]	Effect of speech coding on speaker identification in Indian languages	Study

Comparison of coders considered

Table: Comparison of coders

Algorithm	Bit-rate (kbps)	MOS	Complexity (MIPS)	Frame size (ms)
PCM	64	4.3	0.01	0
GSM FR	13	3.5-3.9	5-6	20
CELP	4.8	3.2	16	30
MELP	2.4	3.2	40	22.5

Degradation Measures of coders:

Table: Average quality measures comparison of coded utterances from different speakers

Algorithm	LLR	WSS	LSD
GSM FR (ETSI 06.10)	0.22	16.86	0.92
CELP (FS1016)	0.57	48.1	1.43
MELP (TI 2.4kbps)	0.28	35.09	1.12

- LLR–LogLikelihood Ratio
- WSS–Weighted Spectral Slope Measure
- LSD–LogSpectral Distance

Speaker Identification under coding

Text-independent speaker Identification system.

- **Database:** TIMIT (50 Male & 50 Female speakers considered)
- **Training:** 8 utterances per speaker
- **Testing:** 2 utterances per speaker
- **Features:** Mel-Frequency Cepstral Coefficients (MFCC), Linear Predictive Cepstral Coefficients (LPCC), Real Cepstral Coefficients (RCC) from residual. [for every 20 ms frame size with 10 ms frame shift]
- **Acoustic models:** Gaussian Mixture Models (GMM)
- **Coders:** GSM (FR-13 kbps), CELP (4.8 kbps), MELP (2.4 kbps)

Different features

- **LPCC:**
  - Linear prediction (LP) analysis of order 26 is performed on the speech signal frame.
  - LP coefficients are converted into cepstral domain.
  - 13-dimensional LPCC considered
- **MFCC:**
  - The short term Fourier transform (STFT) analysis is performed on the speech signal frame.
  - For each frame the STFT magnitude spectrum is computed and is further processed by using the 24 triangular shaped mel-filter banks to find out the filter bank energies.
  - Then discrete cosine transform (DCT) is taken on the spectral energies to obtain MFCCs.
  - 13-dimensional MFCC considered (excluding c_0)
- **RCC:**
  - LP analysis of order 8 is performed on the speech signal frame.
  - For each frame LP residual is calculated by using inverse filtering.
  - Cepstral analysis is performed on residual to obtain RCCs.
  - 16-dimensional RCC considered

Results

Table: speaker recognition performance using MFCC features under different speech coders

Coders	Recognition performance (%)	
	PCM training	Matched training
PCM (Clean)	88.78	-
GSM	70	82.63
CELP	51.04	77.08
MELP	75.79	86.32

Results

Table: speaker recognition performance using LPCC features under different speech coders

Coders	Recognition performance (%)	
	PCM training	Matched training
PCM (Clean)	89.3878	-
GSM	47.3684	84.21
CELP	43.75	75
MELP	50	80

Results

Table: speaker recognition performance using RCC features from LP residual under different speech coders

Coders	Recognition performance (%)	
	PCM training	Matched training
PCM (Clean)	85.8586	-
GSM	14.6465	53.5354
CELP	10.6061	45.4545
MELP	13.6364	63.1313

Summary

- Performance of the system is **decreasing** as coding rate **decreases**.
- Matched condition training is **improving** the performance.
- From the experimental results it is evident that **source features are more degraded** compared to system features due to coding.
- Interesting observation in this work is MELP coders are giving less degradation and high speaker identification performance compared to CELP coder, even though MELP works at less bit rate compared to CELP.
- Experimental results shows that coding has significant effect on speaker recognition performance.

Basu T.K. Porwal G., Patil H.A.

Effect of speech coding on text-independent speaker identification.
In *Proceedings of IEEE, ICISIP*, pages 415–420, 2005.

F. Quatieri, E. Singer, R.B. Dunn, D.A. Reynolds, and J.P. Campbell.

Speaker and language recognition using speech codec parameters.
In *Proceedings of Eurospeech*, pages 787–790, Sep. 1999.

Dunn R.B., Quatieri T.F., Reynolds D.A., and Campbell J.P.

Speaker recognition from coded speech in matched and mismatched condition.
pages 115–120. Speaker Recognition Workshop01, Crete, Greece, June 1999.

Dunn R.B., Quatieri T.F., Reynolds D.A., and Campbell J.P.

Speaker recognition from coded speech and the effects of score normalization.
In *Conference Record of the Thirty-Fifth Asilomar Conference on Signals, Systems and Computers*, pages 1562–1567. IEEE, Nov. 2001.

Thanking You