Speaker-Specific Excitation Source Information from Linear Prediction Residual

S. R. M. Prasanna

Department of Electronics and Communication Engineering

Indian Institute of Technology Guwahati

prasanna@iitg.ernet.in

Acknowledgements

- Prof. B. Yegnanarayana
- S. P. Kishore
- K. Sharat Reddy
- C. S. Gupta
- Jinu M. Zachariah
- Debadatta Pati

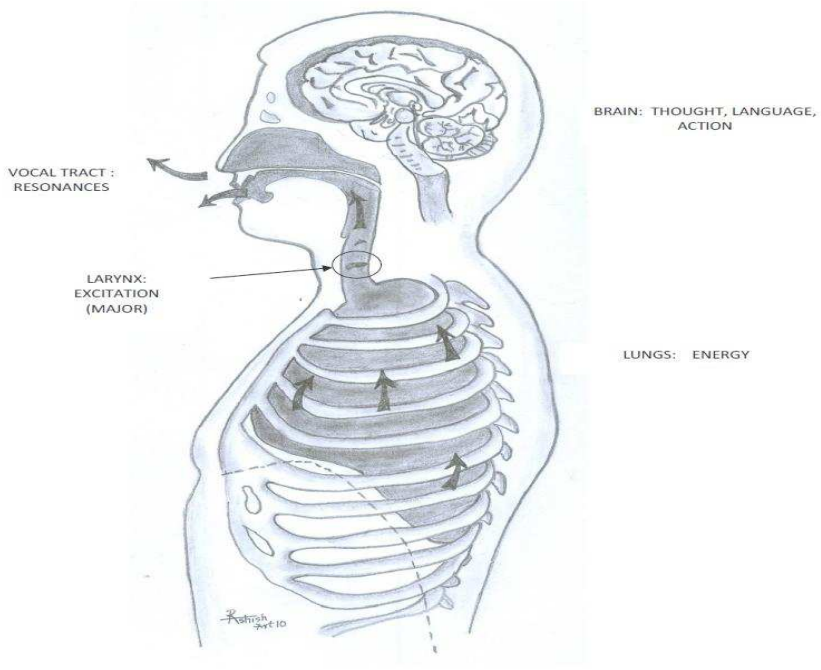
Organization

- Terminologies in speaker recognition task
- Speaker information in speech
  - speaker-specific vocal tract information
  - speaker-specific excitation source information
- Significance of speaker-specific excitation source information
- LP residual as the excitation source information
- Subsegmental, segmental and suprasegmental analysis of LP residual
- Implicit and explicit modeling of speaker information from LP residual
- Summary

Terminologies Speaker Recognition Task

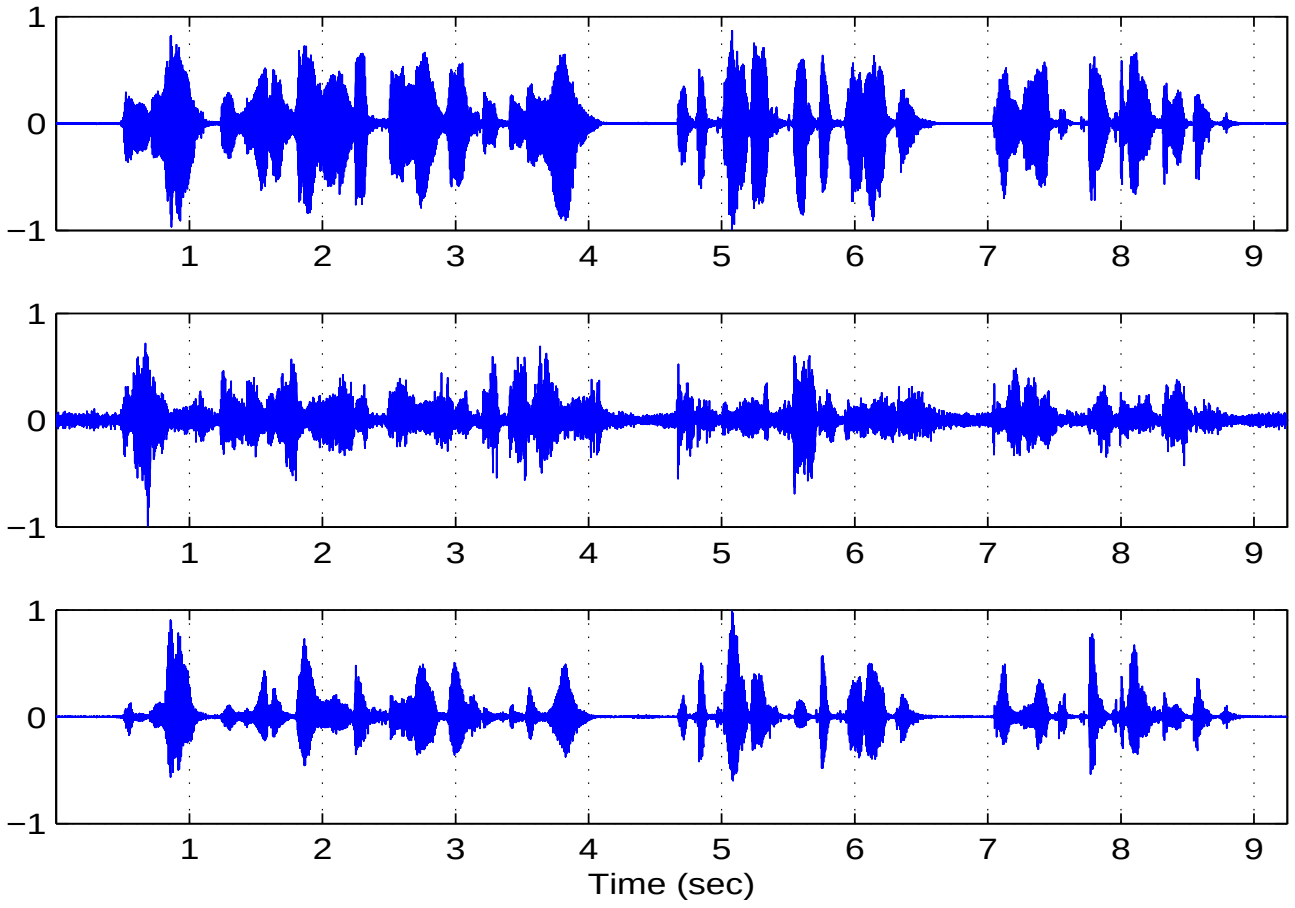
- Task of recognizing speaker of the speech signal
- Speaker verification vs speaker identification
- Text independent vs text dependent
- Automatic speaker recognition involves extracting, modeling and testing speaker information

Embedding Speaker Information into Speech

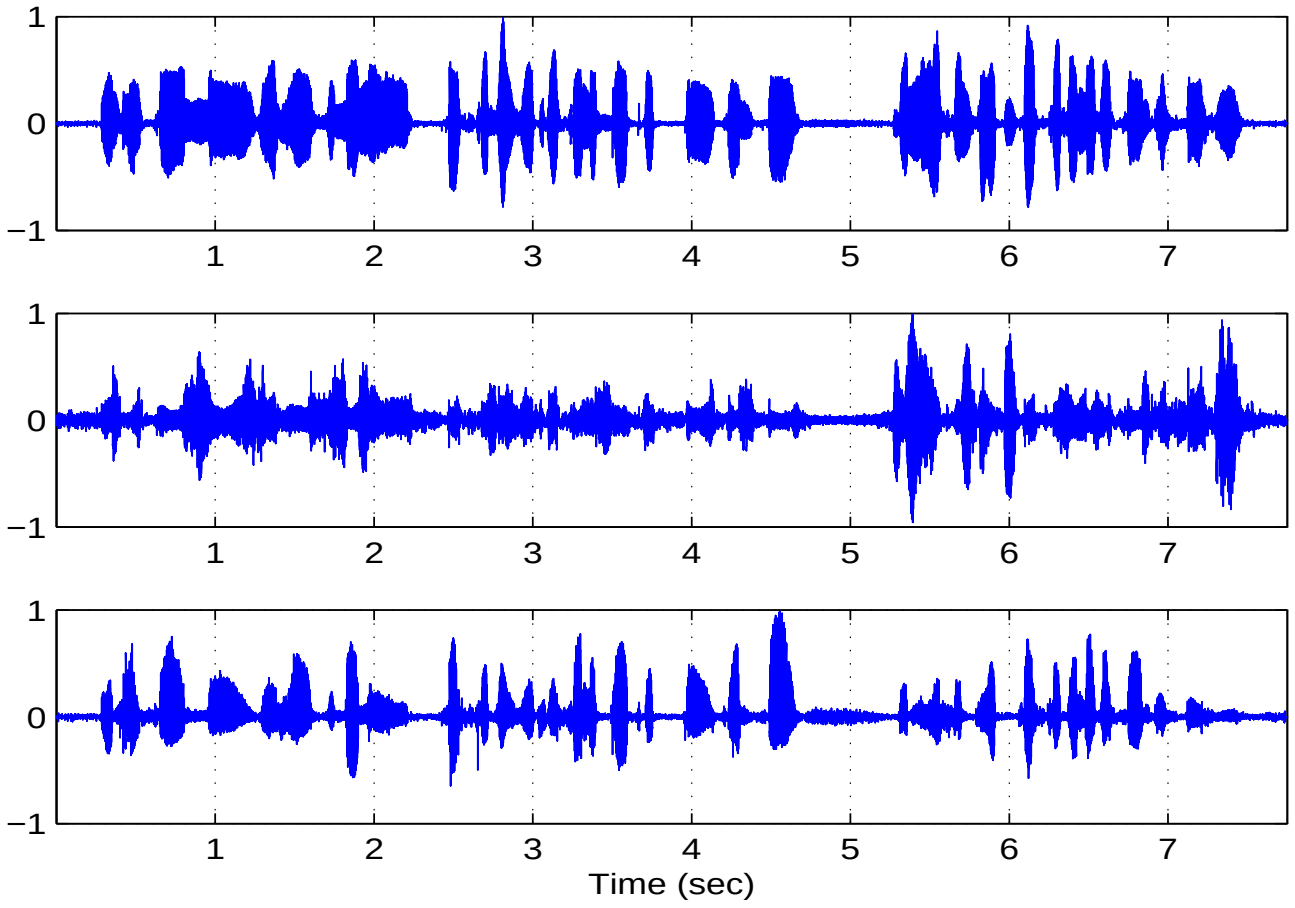


- Larynx major excitation source
- Vocal tract major resonant structure
- Speaker information is due to particular shape, size and dynamics of vocal tract and also excitation source

Significance of Speaker-Specific Source Information



Significance of Speaker-Specific Source Information



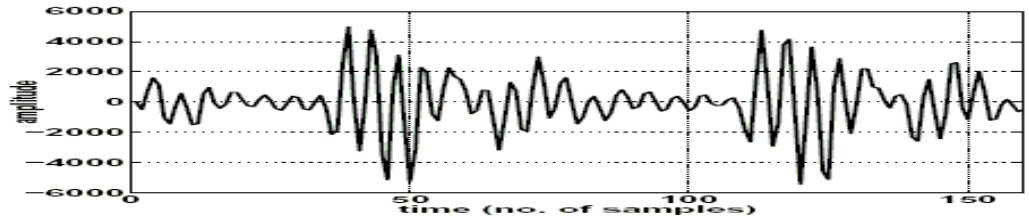
Observations based on Listening

- Vocal tract system component contributes to speaker information
- Excitation source component also contributes equally to speaker information
- Most speaker recognition studies exploit only vocal tract component
- How much will be the potential of excitation source component?
- Will it aid in providing robustness to automatic speaker recognition?

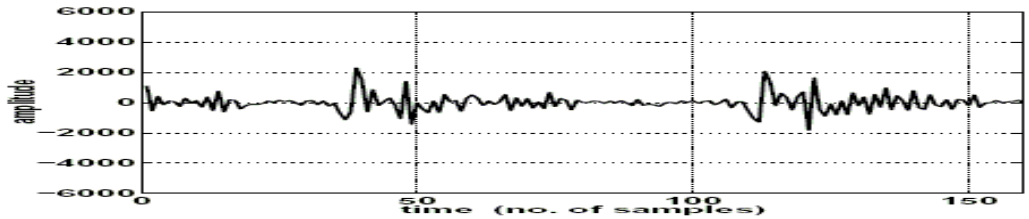
Source and System Separation by LP Analysis

- All pole modeling of speech using suitable LP order
- LPCs: Model vocal tract system information
- LP Residual: Inverse filtering using estimated LPCs
- Since LPCs model VT system information, LP residual mostly contains excitation source information

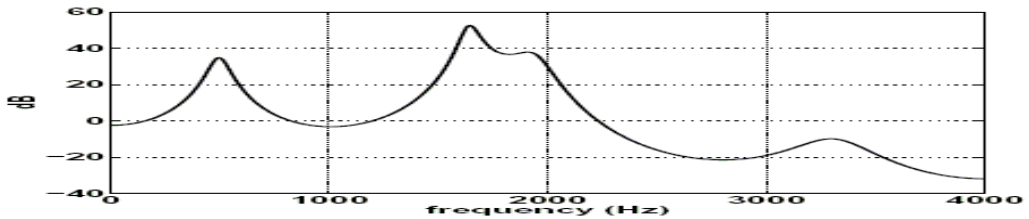
Source and System Separation by LP Analysis



(a)

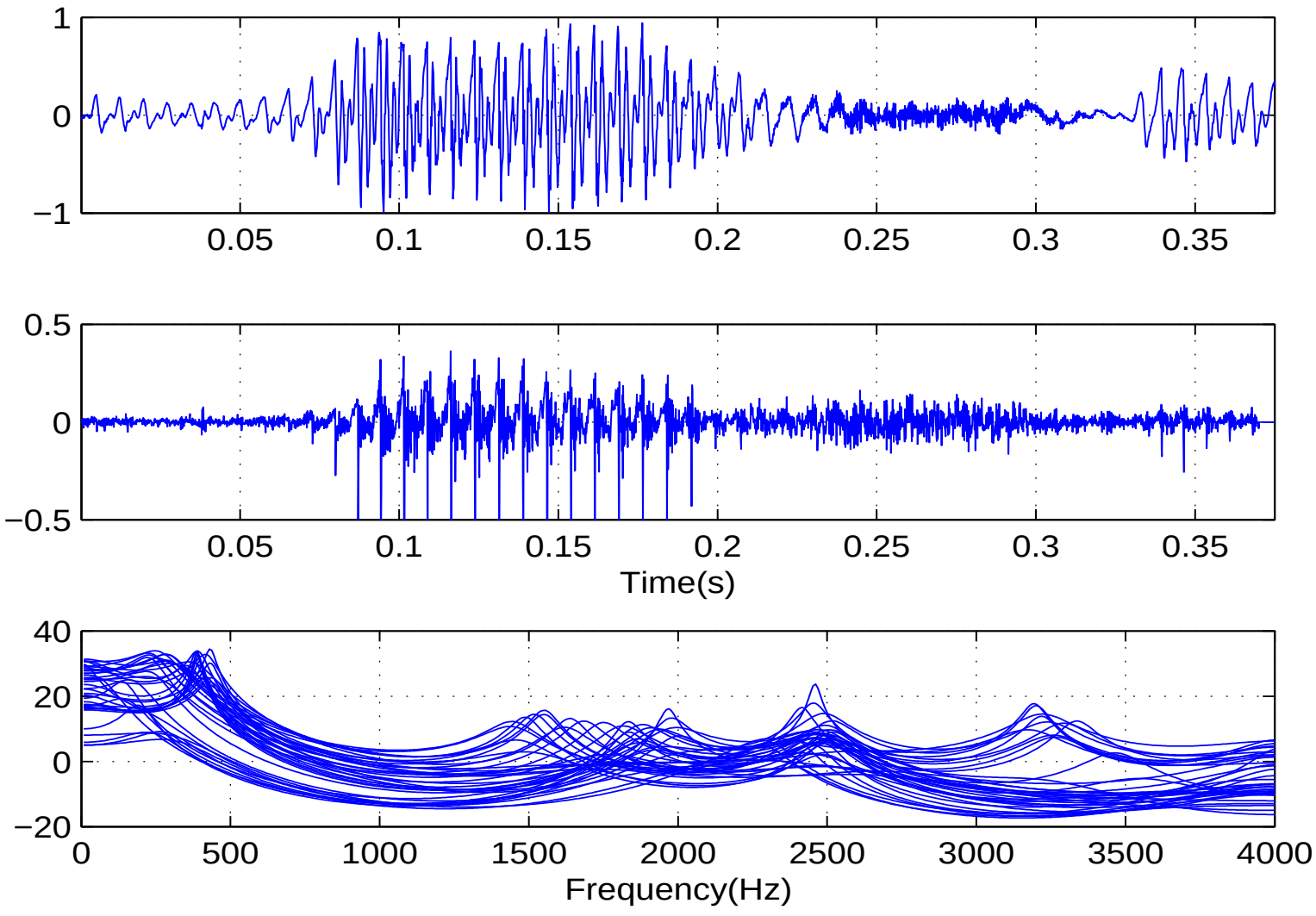


(b)



(c)

Source and System Separation by LP Analysis



Speaker information from Source and System

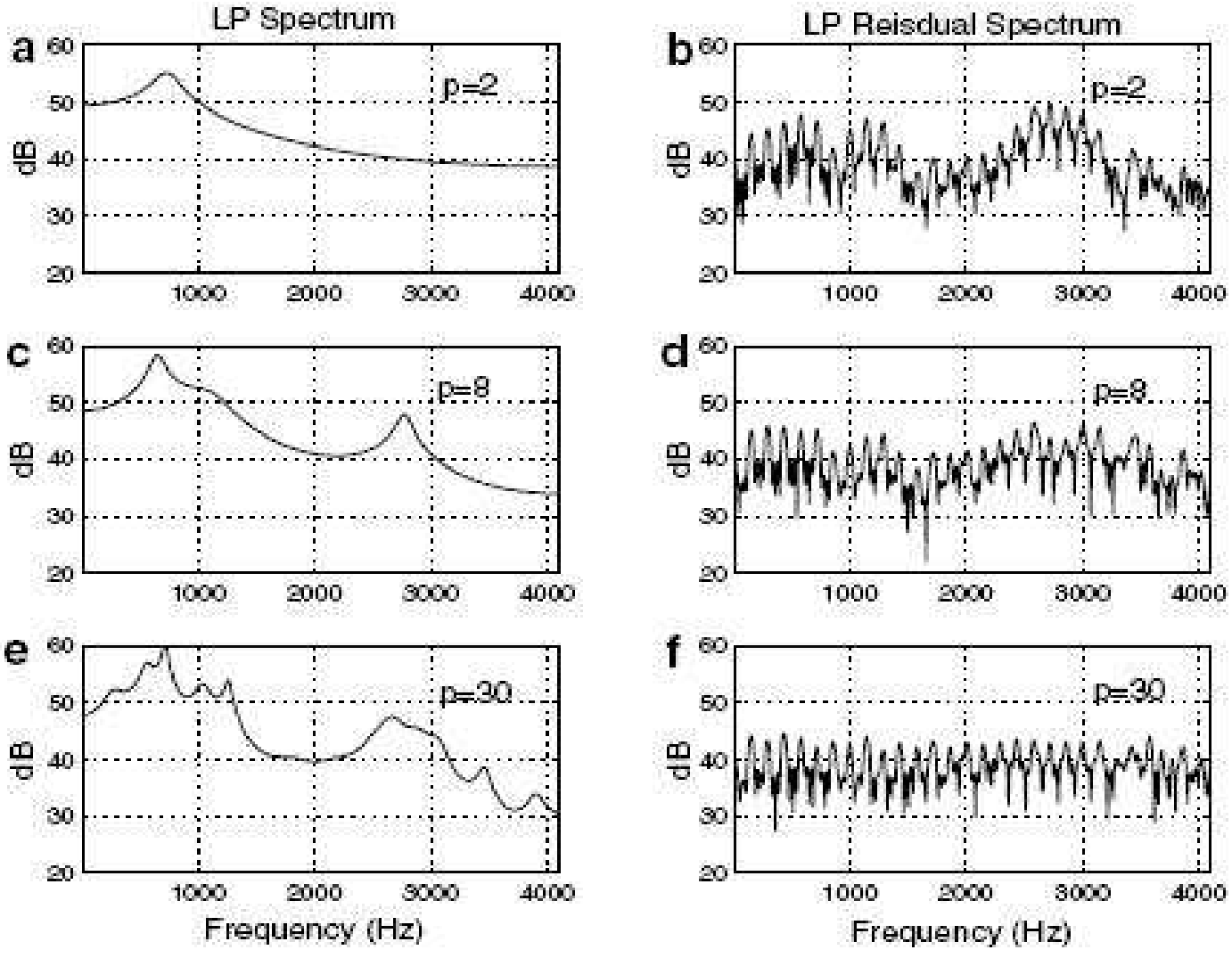
Table 1. Performance of Speaker Recognition using source and system features. The table shows the rank of the speaker obtained by matching with 20 speakers.

	Speaker No.→	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Set I	Rank of Model 1 (system features)	1	1	1	1	1	2	1	1	1	8	1	1	1	1	1	1	1	1	1	1
	Rank of Model 2 (source features)	2	1	1	1	1	1	4	1	1	1	1	2	1	1	1	13	1	1	1	1
	Rank of Combined Model	1	1	1	1	1	1	1	1	1	4	1	1	1	1	1	1	1	1	1	1

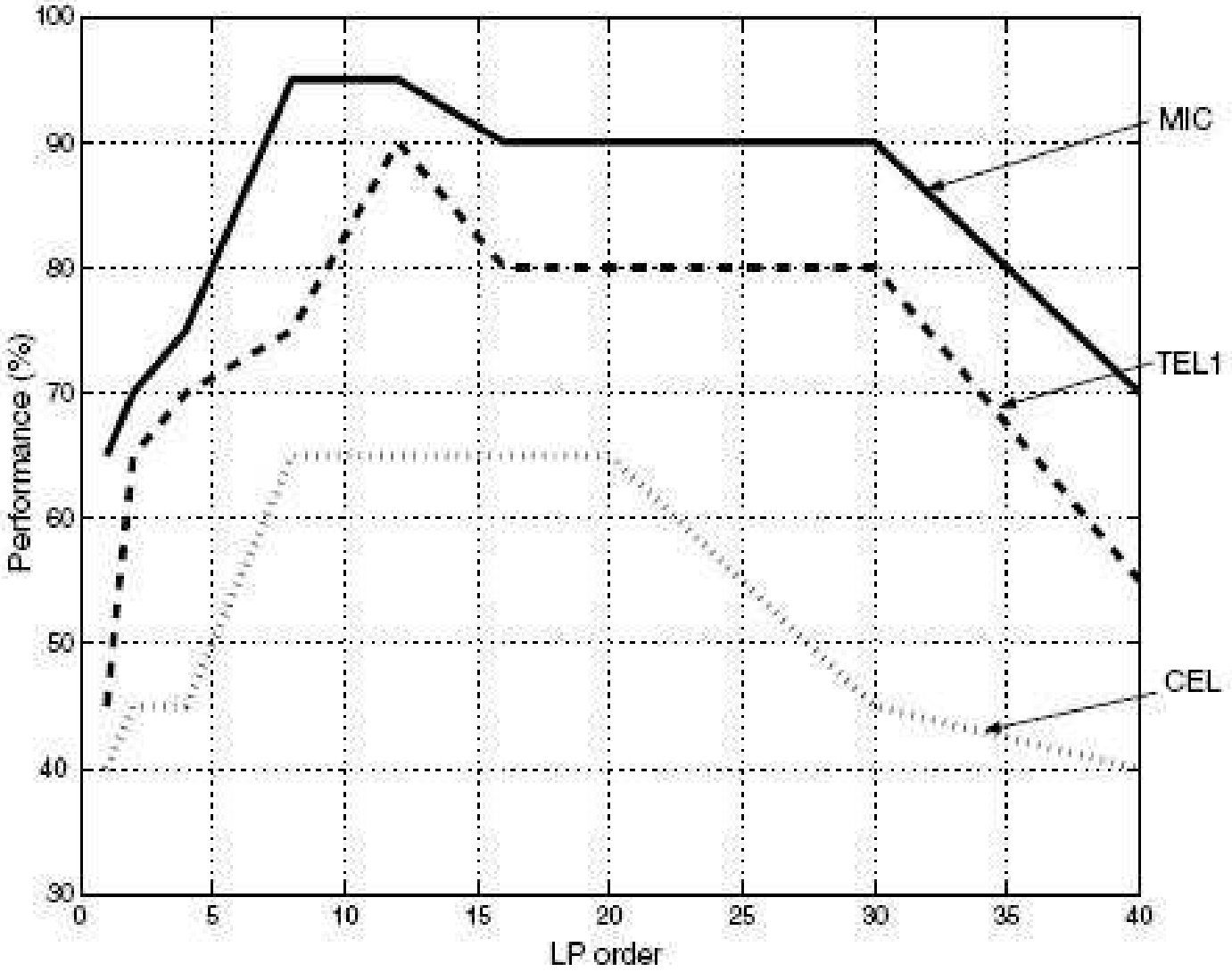
	Speaker No.→	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Set II	Rank of Model 1 (system features)	1	1	1	4	1	1	1	1	1	1	2	1	1	1	1	1	5	1	1	1
	Rank of Model 2 (source features)	1	1	1	1	1	10	1	1	1	1	10	1	1	1	1	3	2	2	1	1
	Rank of Combined Model	1	1	1	2	1	2	1	1	1	1	2	1	1	1	1	1	2	1	1	1

B. Yegnanarayana, K. S. Reddy and S. P. Kishore "Source and system features for speaker recognition using AANN models", in Proc. ICASSP, pp. 109-112, 2001

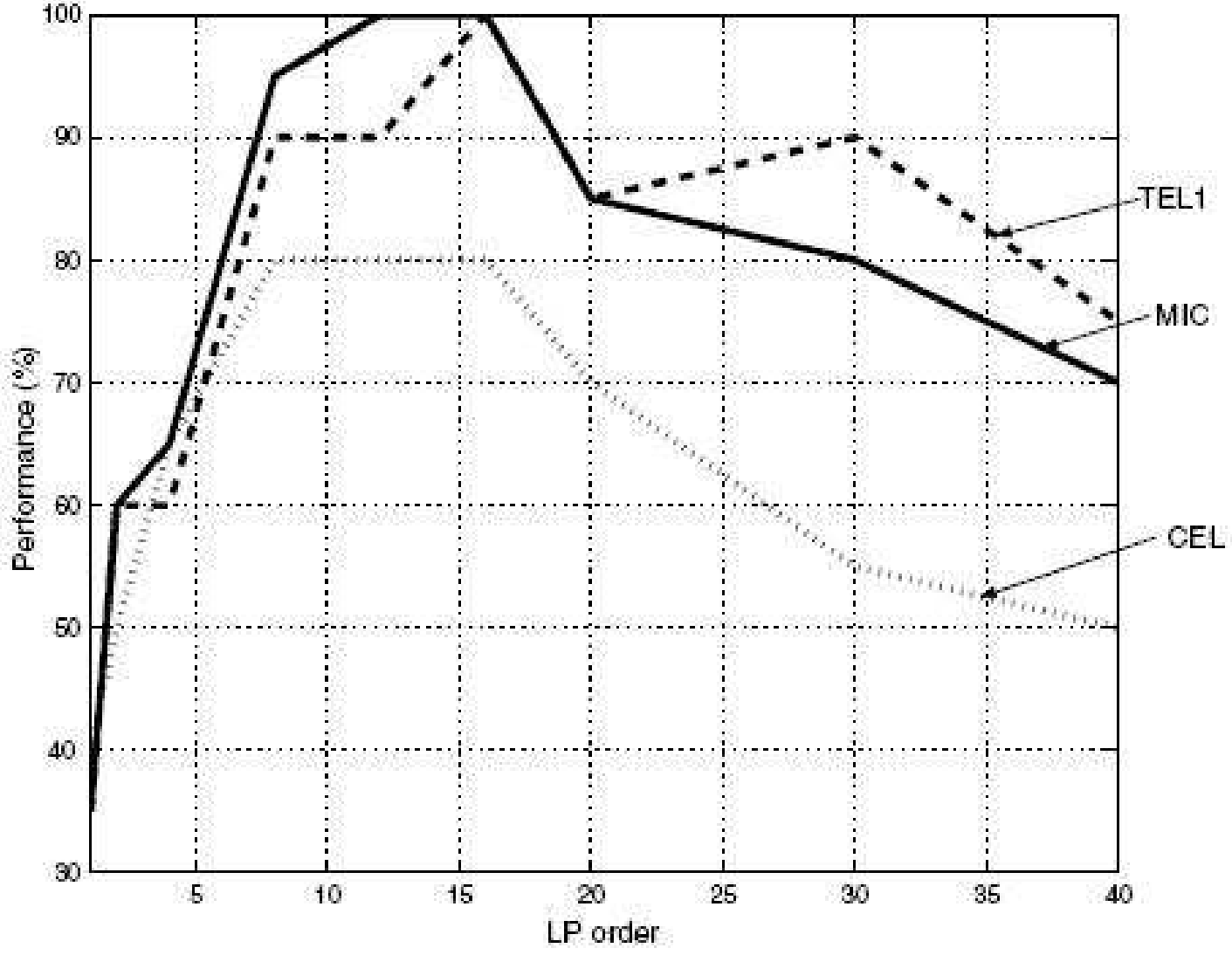
Effect LP order on Source and System Features



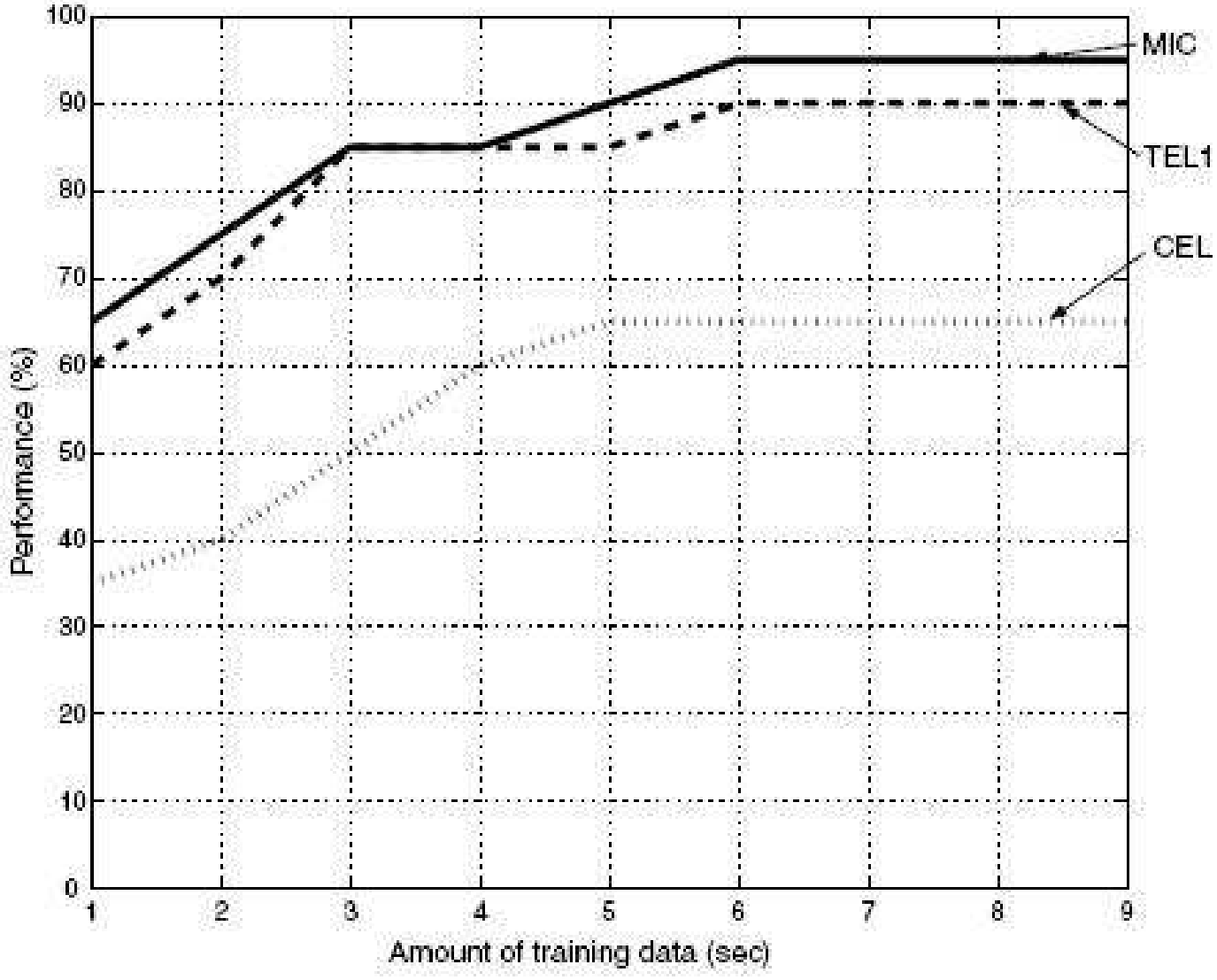
Effect LP order on Source Information



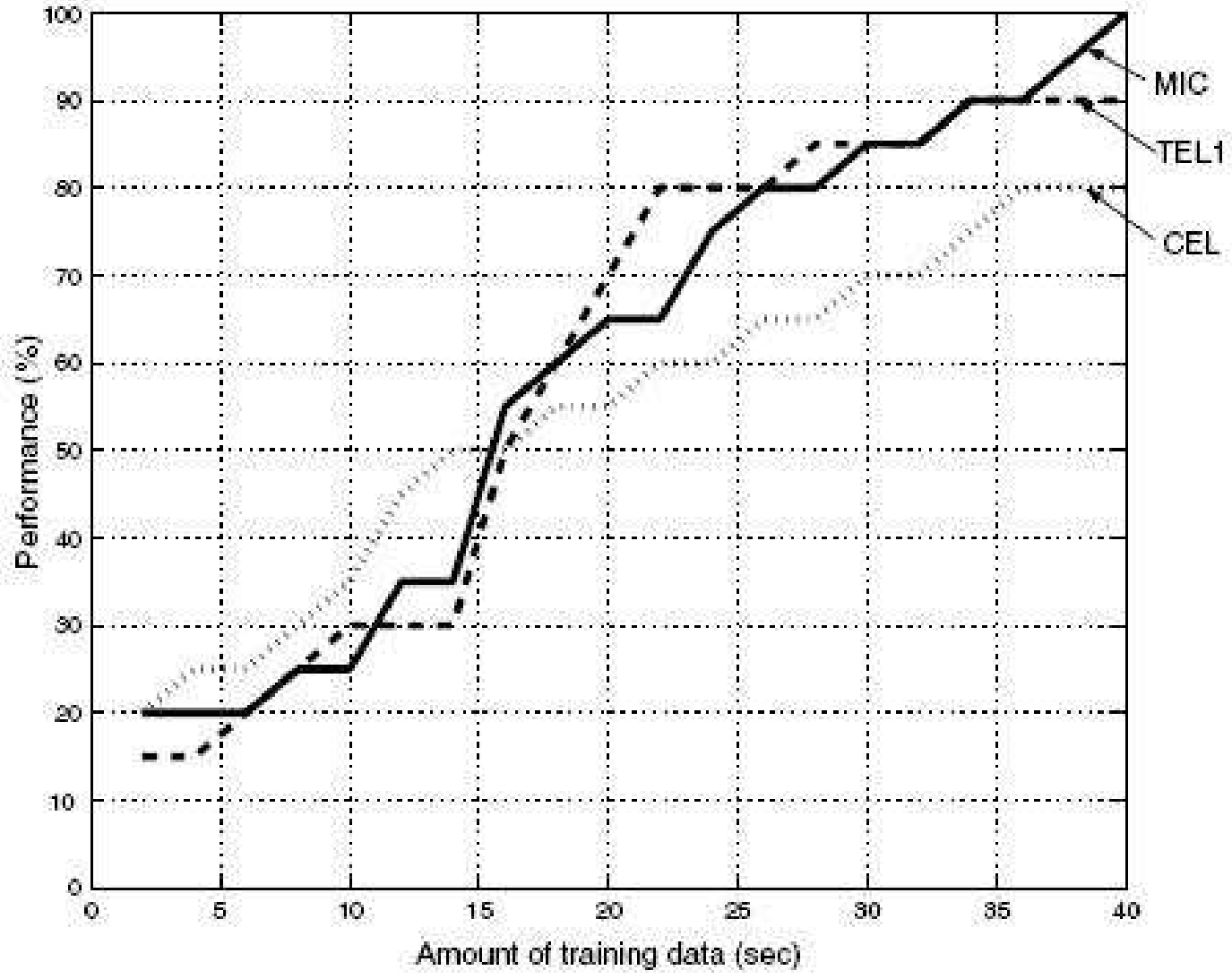
Effect LP order on System Information



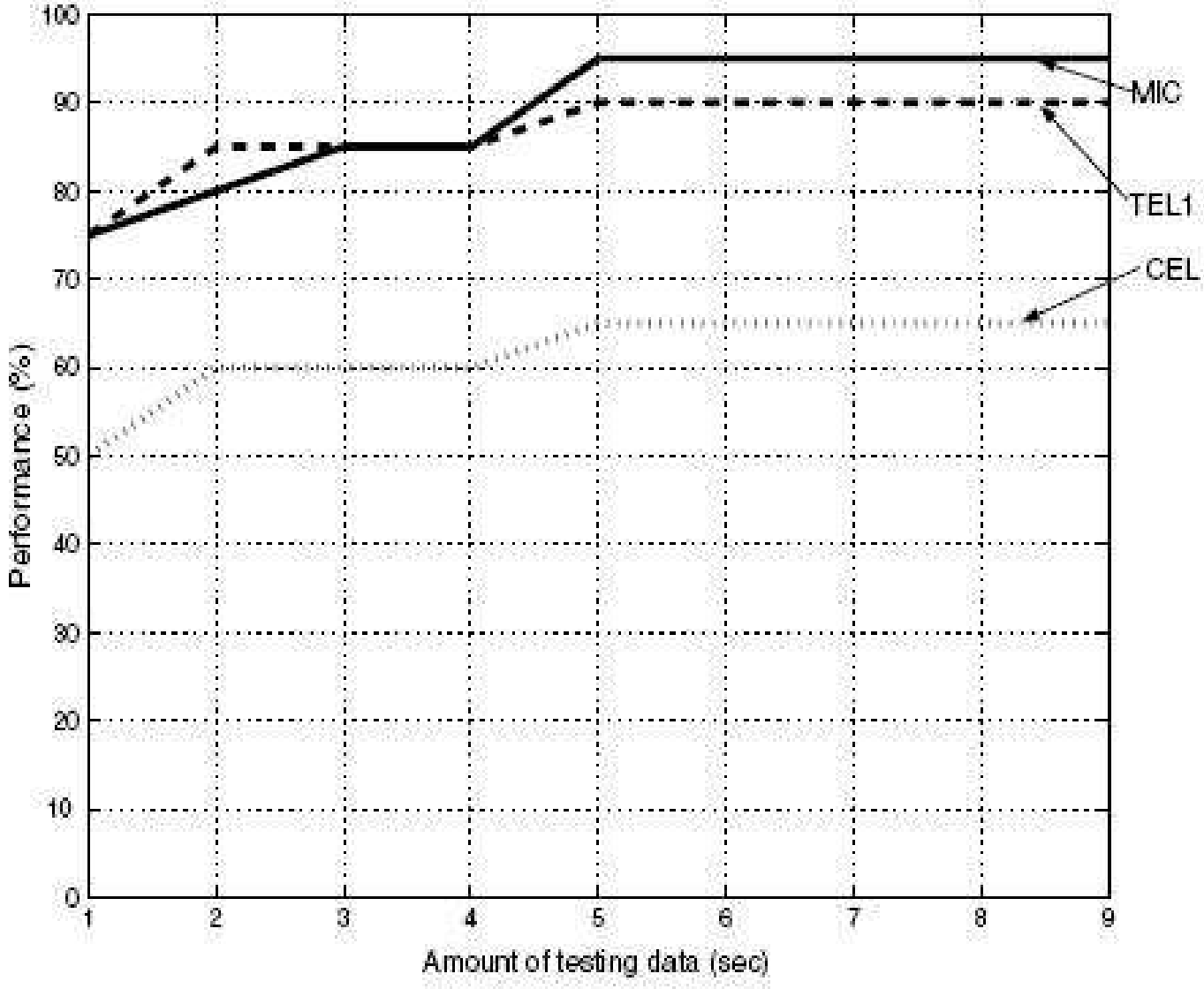
Training Data for Speaker-Specific Source Information



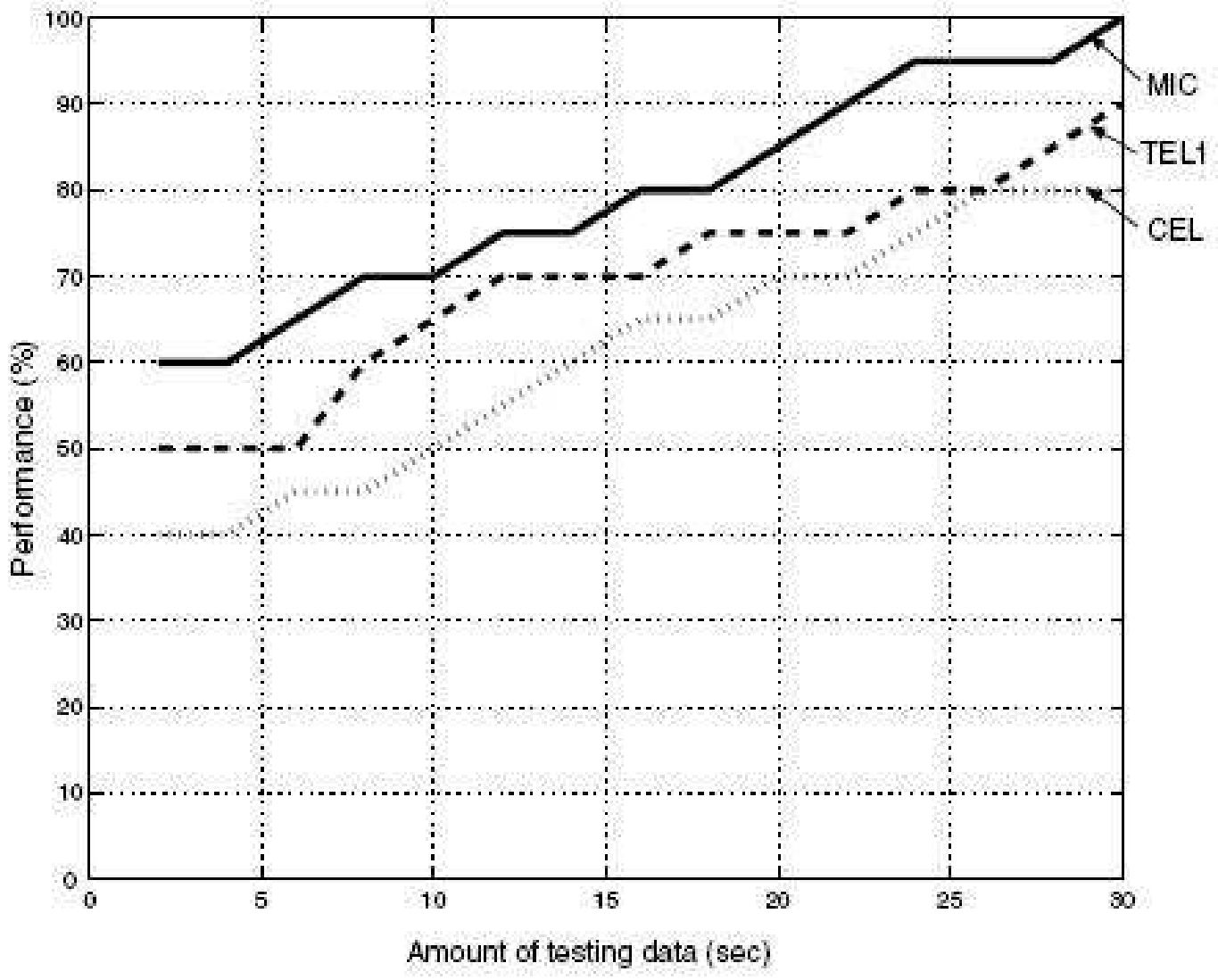
Training Data for Speaker-Specific System Information



Testing Data for Speaker-Specific Source Information



Testing Data for Speaker-Specific System Information



Speaker Identification Performance

Feature	Rank 1	Rank 1 & 2
Source	80	88
System	83	93

Speaker Verification Performance

Feature	EER
Source	23.8
System	17.2
Source + System	15.2
System2	8.6
System2 + Source	7.8
System2 + System	8.0
Source + System + System2	7.1

Observations from Initial Studies

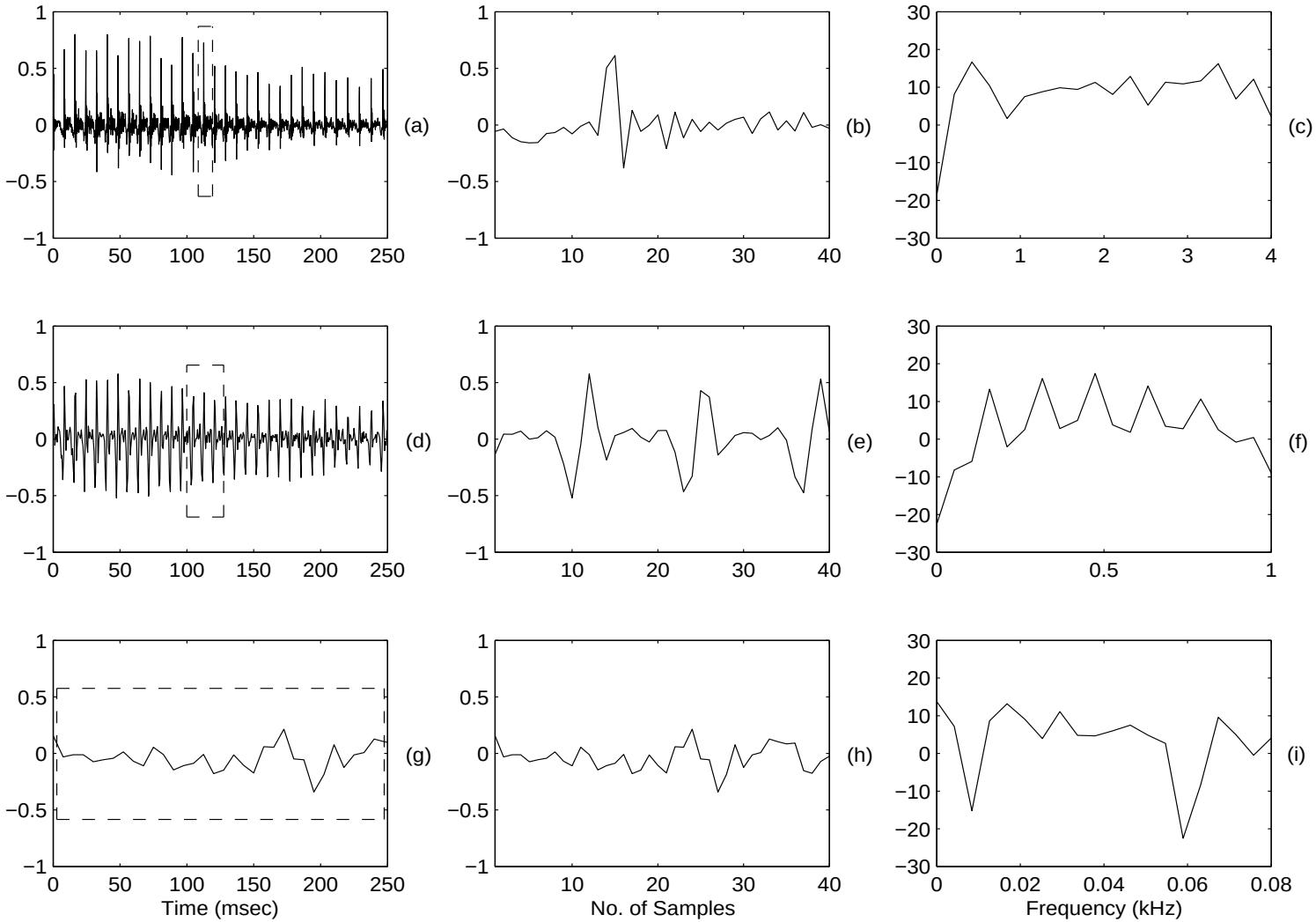
- For small population, source provides comparable performance to system
- LP order in the range 8-16 for 8 kHz sampled signal
- Source needs less data for training and testing
- System needs more data for training and testing
- Source combines well with system to further improve the performance

S. R. M. Prasanna, C. S. Gupta and B. Yegnanarayana, "Extraction of speaker specific information from linear prediction residual of speech", Speech Communication, vol. 28, pp. 1243 - 1261, 2006.

Subsegmental, Segmental and Suprasegmental Processing of LP Residual

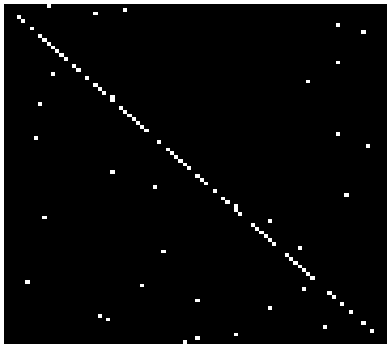
- LP residual only at subsegmental level
- Speaker information in LP residual may be viewed at different levels
- Subsegmental: Each glottal cycle or pitch period (3-5 msec)
- Segmental: Across 2-3 pitch periods (10-30 msec)
- Suprasegmental: Across 20-30 pitch periods (100-300 msec)
- How much speaker information is present at each level?
- Is the speaker information different at the three levels?

Speaker Information at Subsegmental, Segmental and Suprasegmental Levels of LP Residual

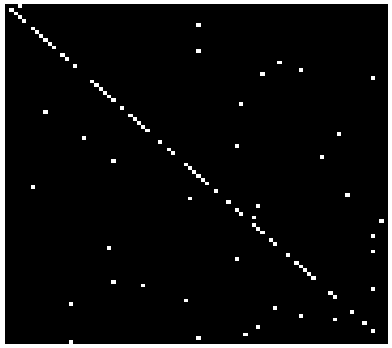


Confusion Patterns at Subsegmental, Segmental and Suprasegmental Levels of LP Residual

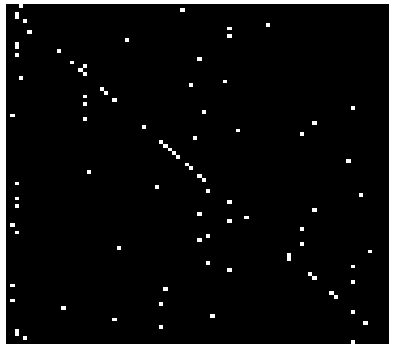
Subsegmental, 64%



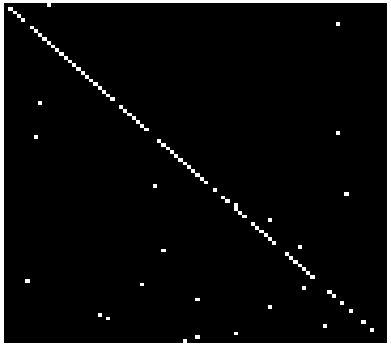
Segmental, 60%



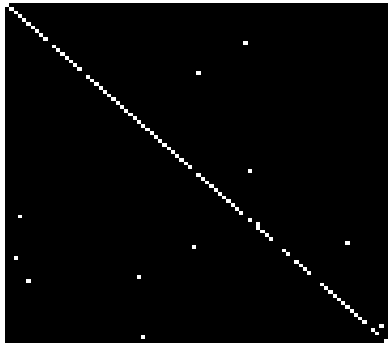
Suprasegmental, 31%



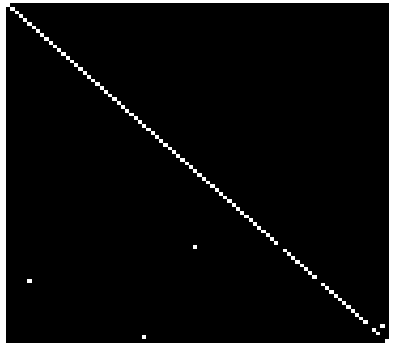
Src−2, 76%



MFCC, 87%



Src−2+MFCC, 96%



Speaker Identification and Verification Results

Sl. No.	Sub	Seg	Supra	Src-1	Src-2	MFCC	Src-2 + MFCC
1	64	60	31	71	76	87	96
2	57	58	13	67	67	66	79
3	11	3	58	6	12	24	18
4	41	27	44	23	21	22	17

- 1. Spkr. Identification using 90 spkrs. from NIST 99
- 2. Spkr. Identification using 90 spkrs. from NIST 03
- 3. Relative degradation in identification performance
- 4. Spkr. Verification using whole NIST 03 database

Observations from Subsegmental, Segmental and Suprasegmental Processing of LP Residual

- Both subsegmental and segmental levels seem to contain significant speaker information
- Suprasegmental level seems to have lowest speaker information, intra-speaker variability and text-independent mode
- Confusion patterns across all three levels are different indicating different aspect of speaker information
- Relatively more robust under degraded condition
- Combine well with vocal tract information to further improve performance

D. Pati and S. R. M. Prasanna, "Subsegmental, segmental and suprasegmental processing of linear prediction residual for speaker information", Int. J. Speech Technology, (accepted Dec 2010).

Implicit vs Explicit Modeling of Speaker Information from LP Residual

- Implicit: Process LP residual directly without any parameters extraction and the model learns the speaker information
- Explicit: Processing LP residual to estimate some parameters and the parameters are used for modelling
- Which one is better, Implicit or Explicit?

Implicit vs Explicit Modeling of Subsegmental Speaker Information

- Implicit modeling by direct processing of LP residual
- Explicit modeling by estimating glottal wave and its derivative parameters
- More like an innovation seq and hence gain by implicit modeling

Modelling	Identfn 1	Identfn 2	Verification
Implicit	64	57	41
Explicit	30	25	39

Implicit vs Explicit Modeling of Segmental Speaker Information

- Implicit modeling by direct processing of LP residual
- Explicit modeling by M-PDSS and R-MFCC features
- Explicit modeling provides better performance

Modelling	Identfn 1	Identfn 2	Verification
Implicit	60	58	27
Explicit	88	61	27

Implicit vs Explicit Modeling of Suprasegmental Speaker Information

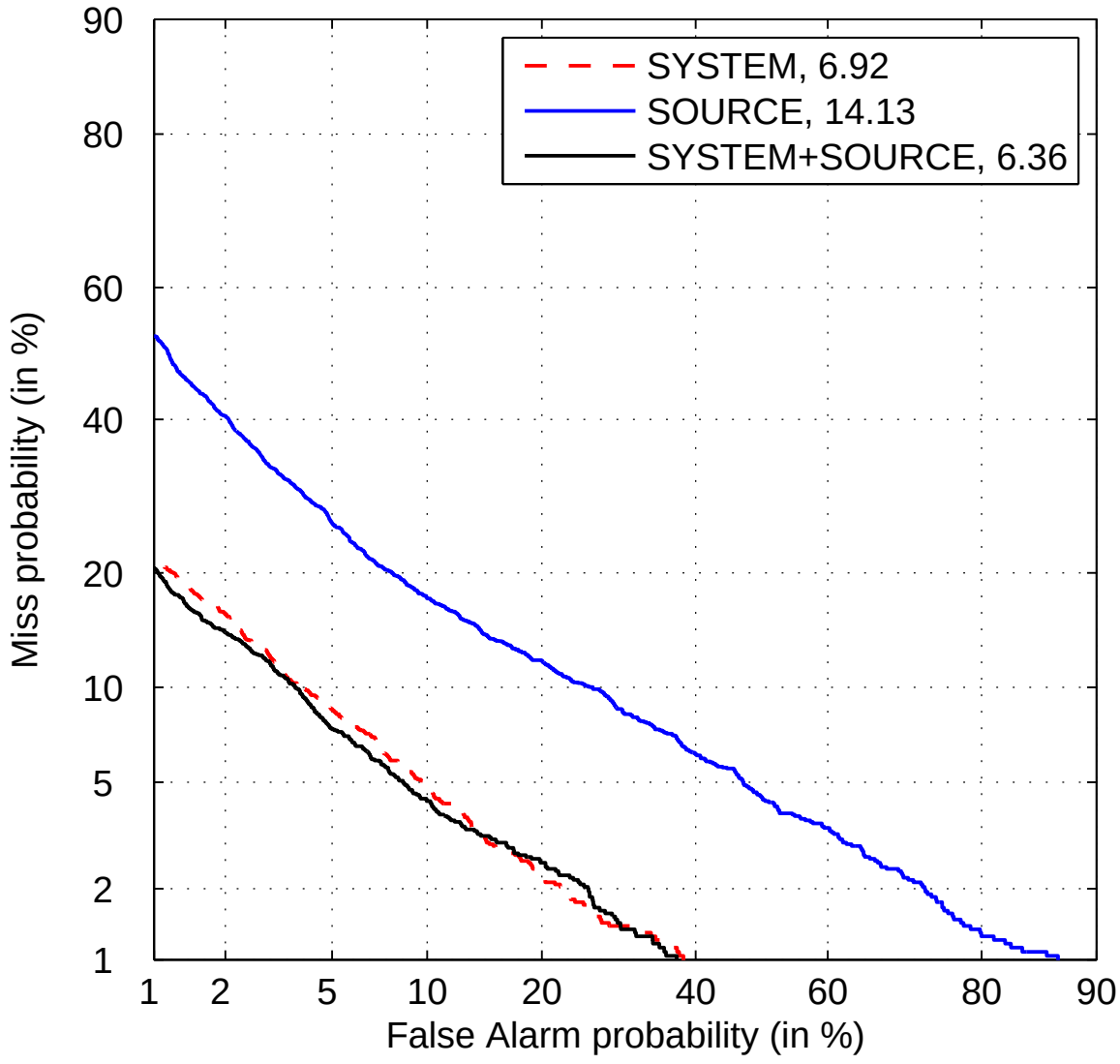
- Implicit modeling by direct processing of LP residual
- Explicit modeling by pitch and epoch strength contours
- Even though like innovation sequence, large variability prefers explicit modeling

Modelling	Identfn 1	Identfn 2	Verification
Implicit	31	17	33
Explicit	33	21	31

Summary of Implicit and Explicit Modeling of LP Residual

- Subsegmental: Implicit modeling
- Segmental: Explicit modeling
- Suprasegmental: Explicit seem to be better
- Verification: Explicit modeling
- Identification: Implicit for subsegmental, and explicit for segmental and suprasegmental processing

Speaker Verification Study on NIST 2003 Database



Summary

- Initial studies demonstrated presence of significant speaker information in the LP residual
- Different speaker information at subsegmental, segmental and suprasegmental levels
- Choice of implicit or explicit modeling depends on the level of processing
- For large population size, excitation source based system still lags behind the vocal tract based system
- Avenues for exploring new analysis, feature extraction and modeling approaches for the development of source based speaker recognition system

References

1. B. Yegnanarayana, K. S. Reddy and S. P. Kishore "Source and system features for speaker recognition using AANN models", in Proc. ICASSP, pp. 109-112, 2001
2. K. S. Reddy "Source and system features for speaker recognition", MS Thesis, Dept of CSE, IIT Madras, 2001.
3. J. M. Zachariah "Text dependent speaker verification using segmental, suprasegmental and source features", MS Thesis, Dept of CSE, IIT Madras, 2002.
4. S. C. Gupta "Significance of source features for speaker recognition", MS Thesis, Dept of CSE, IIT Madras, 2003.
5. S. R. M. Prasanna, C. S. Gupta and B. Yegnanarayana, "Extraction of speaker specific information from linear prediction residual of speech", Speech Communication, vol. 28, pp. 1243 - 1261, 2006.
6. D. Pati and S. R. M. Prasanna, "Subsegmental, segmental and suprasegmental processing of linear prediction residual for speaker information", Int. J. Speech Technology, (accepted Dec 2010).
7. D. Pati and S. R. M. Prasanna, "Processing linear prediction residual in spectral and cepstral domains for speaker information", Int. J. Speech Technology, (under review).
8. D. Pati and S. R. M. Prasanna, "Explicit and implicit modeling subsegmental speaker-specific excitation information" (to be communicated).