Workshop on Emerging Research Topics in Engineering,
DAIICT, Gandhinagarm 24-25 July 2009

Vocal Tract Shape Estimation for Visual Speech Training Aids

P.C. Pandey

EE Dept, IIT Bombay

pcpandey AT ee.iitb.ac.in >

http:www.ee.iitb.ac.in/~pcpandey

Signal Processing & Instrumentation Lab

EE Dept, IIT Bombay

http://www.ee.iitb.ac.in/~spilab

Impedance Cardiography

- Development of impedance cardiograph
- Artifact suppression in impedance cardiography

Speech & Hearing

- Low cost diagnostic audiometer & noise cancelling headphones
- Impedance glottography
- Enhancement of electrolaryngeal speech
- Speech synthesis and voice transformation
- Speech processing for hearing aids for sensorineural loss
- Speech training aids for the hearing impaired

P.C. Pandey, "Vocal tract shape estimation for speech training aids",
EE Dept, IIT Bombay. 31/Jan/09,
< pcpandey@ee.iitb.ac.in >, http:www.ee.iitb.ac.in/~pcpandey

Abstract **-- Children with prelingual profound hearing impairment lack auditory feedback and have great difficulty in acquiring speech. Most of them do not learn to speak properly despite a fully functional speech production system. Speech-training systems providing visual feedback of vocal-tract shape are found to be useful for improving vowel articulation. Vocal-tract shape estimation, based on LPC and other analysis techniques, generally fails during stop closures, and this restricts its effectiveness in speech training for production of consonants not having visible articulatory efforts.**

A technique based on two-dimensional surface modeling of the area values, estimated by LPC analysis, during the vowel-consonant and consonant-vowel transitions preceding and following the stop closure, has been investigated for interpolating the area values during the stop closures. Surface modeling was based on least-squares bivariate polynomials and Delaunay triangulation methods. Using the technique, the place of closure could be estimated consistently for various stop consonants.

Based on this research and work by others, a visual speech-training system is being developed to facilitate various aspects of speech learning by the hearing impaired children.

References

P.C. Pandey & M.S. Shah, "Estimation of place of articulation during stop closures of vowel-consonant-vowel utterances", IEEE Trans. Audio, Speech, and Language Proc., vol 17(2), pp 277-286, Feb. 2009.

M.S. Shah, "Estimation of place of articulation during stop closures of vovel-consonant-vowel syllables", Ph.D. dissertation, EE Dept., IIT Bombay, 2008.

J. F. Curtis (Ed.), Processes and Disorders of Human Communication. New York: Harper and Row, 1978.

R. S. Nikerson, "Characteristics of the speech of deaf persons," Volta Rev., vol. 77, pp. 342–362, 1975; reprinted in: Sensory Aids for the Hearing Impaired, pp. 540–545, H. Levitt, J. M. Pickett, and R. A. Houde (Eds.), New York: IEEE Press, 1980.

H. Levitt, J. M. Pickett, and R. A. Houde, (Eds.), Sensory Aids for the Hearing Impaired. New York: IEEE Press, 1980.

R. G. Crichton and F. Fallside, "Linear prediction model of speech production with applications to deaf speech training," Proc. IEE Control and Sci., vol. 121, pp. 865–873, 1974.

J. M. Pardo, "Vocal tract shape analysis for children," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 763–766, 1982.

S. Aguilera, A. Borrajo, J. M. Pardo, and E. Munoz, "Speech-analysis-based devices for diagnosis and education of speech and hearing impaired people," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 641–644, 1986.

M. Shigenaga and H. Kubo, "Speech training system for handicapped children using vocal tract lateral shapes," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 637–640, 1986.

N. D. Black, "Application of vocal tract shapes to vowel production," in Proc. 10th Int. Conf. IEEE Engg. Med. Biol. Soc., pp. 1535–1536, 1988.

S. H. Park, D. J. Kim, J. H. Lee, and T. S. Yoon, "Integrated speech training system for hearing impaired," IEEE Trans. Rehab. Engg., vol. 2, no. 4, pp. 189–196, 1994.

P. M. T. de Oliveira and M. N. Souza, "Speech aid for the deaf based on a representation of the vocal tract: the vowel module," in Proc. 19th Int. Conf. IEEE Engg. in Med. and Biol. Soc., pp. 1757–1759, 1997.

H. Wakita, "Direct estimation of the vocal tract shape by inverse filtering of acoustic speech waveforms," IEEE Trans. Audio Electroacoust., vol. 21, no. 5, pp. 417–427, 1973.

P. Ladefoged, R. Harshman, L. Goldstein, and L. Rice, "Generating vocal tract shapes from formant frequencies," J. Acoust. Soc. Am., vol. 64, no. 4, pp. 1027–1035, 1978.

D. Rossiter, D. M. Howard, and M. Downes, "A real-time LPC-based vocal tract area display for voice development," J. of Voice, vol. 8, no. 4, pp. 314–319, 1994.

Z. Yu and P. C. Ching, "Determination of vocal-tract shapes from formant frequencies based on perturbation theory and interpolation method," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., pp. 369–372, 1996.

J. Schroeter and M. M. Sondhi, "Techniques for estimating vocal-tract shapes from the speech signal," IEEE Trans. Speech Audio Process., vol. 2, no. 1, pt. 2, pp. 133–150, 1994.

L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals. Englewood Cliffs, NJ: Prentice-Hall, 1978.

D. O'Shaughnessy, Speech Communications: Human and Machines. Reading, Massachusetts: Addison-Wesley, 1987.

M. S. Shah and P. C. Pandey, "Estimation of vocal tract shape for VCV syllables for a speech training aid," in Proc. 27th Int. Conf. IEEE Engg. Med. Biol. Soc., pp. 6642–6645, 2005.

M.S. Shah and P.C. Pandey, "Estimation of place of articulation in stop consonants for visual feedback", in Proc. of Interspeech 2007, Paper No. FrB.O2-4, pp 2477-2480.

Presentation Outline

1. Introduction
2. Visual Speech-training Aids
3. LPC Based Vocal Tract Shape Estimation
4. Estimation of Vocal Tract Shape during Stop Closures
5. Results & Discussion
6. Summary & Conclusions

P.C. Pandey, EE Dept, IIT Bombay

1. Introduction

2. Visual Speech-training Aids

3. LPC Based Vocal Tract Shape Estimation

4. Estimation of Vocal Tract Shape during Stop Closures

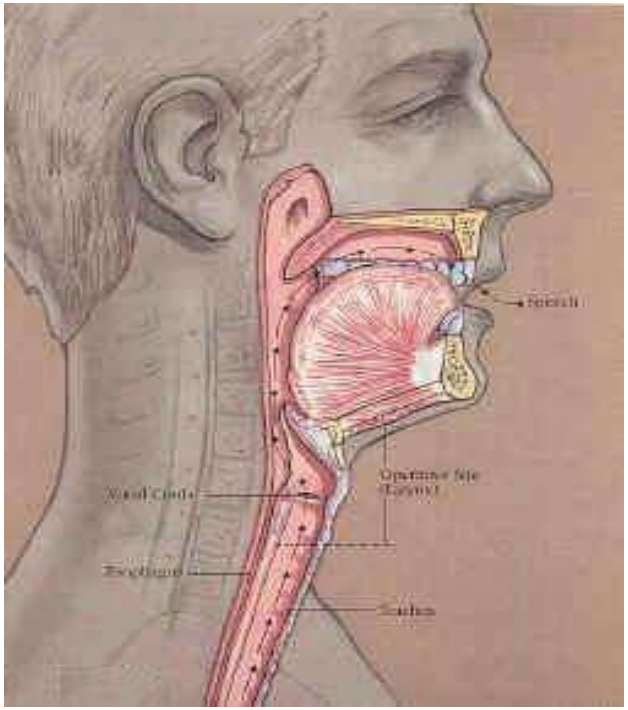
5. Results & Discussion

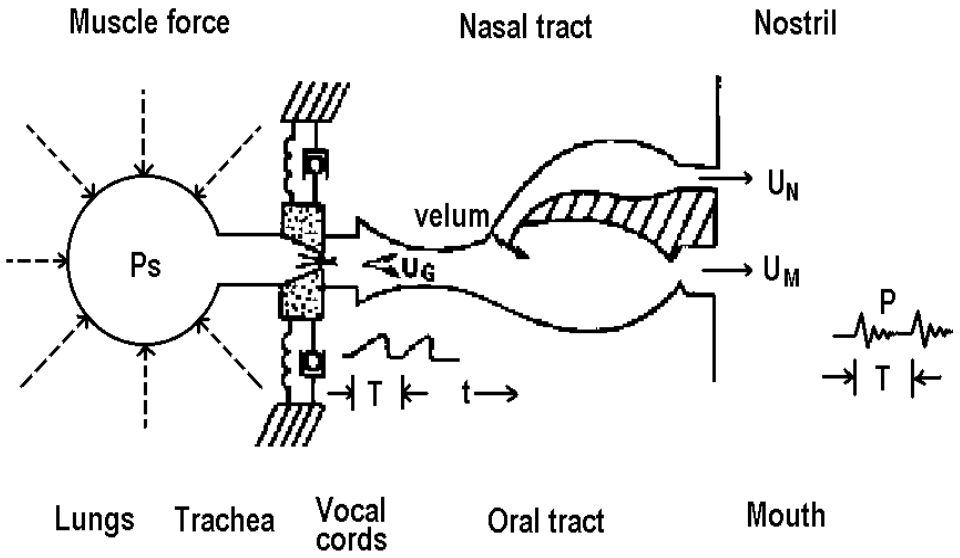
6. Summary & Conclusions

P.C. Pandey, EE Dept, IIT Bombay

Speech Production

(voicing, place, manner)





Basic speech sounds (phonemes)

Vowels **: Pure vowels, Diphthongs**

Consonants**: Semivowels, Fricatives, Oral stops, Affricates, Nasals**

P.C. Pandey, EE Dept, IIT Bombay

Speech Acquisition Process

Children with normal hearing

Acquisition of ability to control various articulators aided by auditory feedback.

Children with hearing impairment

- **Lack of auditory feedback during speech production.**
- **Articulation accuracy, stress, & intonation patterns affected.**
- **Vowels & consonants with tongue movement hidden in the mouth not distinguishable.**
- **Speech impairment, despite proper speech production mechanism.**

P.C. Pandey, EE Dept, IIT Bombay

Speech Training Aids

- Visual feedback
- Tactile feedback

Importance of visual feedback of articulatory gestures

- Only 20% of the English phonemes have cues visible on lips.
- Labial consonants by deaf are more intelligible than lingual consonants.
- Speech-training systems based on visual feedback of vocal tract shape are useful for improvement in vowel articulation.

P.C. Pandey, EE Dept, IIT Bombay

Techniques for Vocal Tract Shape Estimation

Direct methods

- ♦ Geometrical measurement
  - ▫ X-ray imaging, ▫ MRI
  - ▫ Electromagnetic articulography (EMA)
  - ▫ Optopalatography (EPG)
  - ▫ Ultrasonic imaging

Indirect techniques

- ♦ Acoustic measurement at lips
  - ▫ Impedance
  - ▫ Impulse response
- ♦ Processing of speech signal
  - ▫ Linear Predictive Coding (LPC)
  - ▫ Articulatory analysis by synthesis
  - ▫ Formant analysis

P.C. Pandey, EE Dept, IIT Bombay

LPC Based Estimation of Vocal Tract Shape

- Automated tracking of formants not required.
- Real time processing feasible.
- Transformation of LPC coefficients into other parameter sets for interpolation and smoothening of estimated shapes.
- Estimation satisfactory for vowels.
- Failure of shape estimation during stop closure due to very low signal energy & unavailability of relevant spectral information.
- Indication of place of constriction during consonants critical for a speech training aid.

P.C. Pandey, EE Dept, IIT Bombay

Research Objective

- To develop speech training aids with visual feedback of the articulatory efforts

- To develop techniques for estimation of place of constriction during oral stop closures of vowel-consonant-vowel syllables, for use in the speech training aids.

P.C. Pandey, EE Dept, IIT Bombay

Proposed Techniques

♦ Production of VCV syllables with oral stop consonants : movement of articulators from the articulatory position of the vowel towards that of the stop closure to that of the vowel.

→ *Dynamic variation in vocal tract shape and formants during VC and CV transitions related to movement of articulators.*

→ Surface modeling of the time varying vocal tract shape during the the transitions preceding and following the stop closure for estimating the place of constriction, during the closure duration.

♦ Surface modeling of time varying vocal tract shape using

- Bivariate polynomials
- Delaunay triangulation

P.C. Pandey, EE Dept, IIT Bombay

Investigations

- Surface modeling of time varying vocal tract shape during VC and CV transition segments by bivariate second & third degree polynomials & Delaunay triangulation.

- Estimation of vocal tract shape and place of constriction during the stop closure using 2D interpolation of the modeled surface.

P.C. Pandey, EE Dept, IIT Bombay

1. Introduction

2. Visual Speech-training Aids

3. LPC Based Vocal Tract Shape Estimation

4. Estimation of Vocal Tract Shape during Stop Closures

5. Results & Discussion

6. Summary & Conclusions

P.C. Pandey, EE Dept, IIT Bombay

Speech-training Systems

- Feedback of *Acoustic Parameters*

  speech intensity

  fundamental frequency

  spectral features

- Feedback of *Articulatory Parameters*

  voicing

  nasality

  lip & vocal tract movement

- Simultaneous display of the desired & estimated patterns

  for minimizing the mismatch.

P.C. Pandey, EE Dept, IIT Bombay

- Coyne (1938) Gruenz & Schott (1949): Feedback of pitch
- Risberg (1968): Visual feedback of acoustic / articulatory parameters [indicators for frication, intonation, rhythm, nasalization, spectrum]
- Flecher (1982): PC based system called *Dynamic Orometer* [feedback of movement of tongue, pattern of tongue contact against teeth & roof of mouth, movement of lips & jaw, spectrum, F0, intensity]
- Bernstein *et. al.* (1986): PC based system for sustained voicing & intensity control
- Zahorian & Venkat (1990): PC based vowel articulation system

P.C. Pandey, EE Dept, IIT Bombay

Systems for Vocal Tract Shape Visualization

- **LPC / formants based speech analysis** [e.g., Wakita (1973), Ladefoged *et al.* (1978), Yu & Ching (1996), Kshirsagar (1998), Mahdi (2003), Deng *et al.* (2005)]
- **Type of displays, games, motivation, etc for speech training** [e.g., Crichton & Fallside (1974), Pardo (1982), Bernstein *et al.* (1986), Aguilera *et al.* (1986), Shigenaga & Kubo (1986), Javkin *et al.* (1993), Park *et al.* (1994), Oliveira & Souza (1997), Watanabe *et al.* (2000)]
- **Commercially available PC based training systems**
  - **real time estimation & display of vocal tract shape** [e.g., Language Vision Inc. (2003)]
  - **games, motivation, etc.** [Dr. Speech Software Group (2003), Video Voice Speech Training System (2003)]

P.C. Pandey, EE Dept, IIT Bombay

1. Introduction

2. Visual Speech-training Aids

3. LPC Based Vocal Tract Shape Estimation

4. Estimation of Vocal Tract Shape during Stop Closures

5. Results & Discussion

6. Summary & Conclusions

P.C. Pandey, EE Dept, IIT Bombay

- **Vocal tract shape estimation**

  based on LPC analysis & Wakita's model (Wakita, 1973)

- **Speech processing and display package**

  'VTAG-1' developed in Matlab®

- **Analysis for shape estimation**

  using Areagram, 2D display of square-root of vocal tract area with time & glottis-to-lips distance

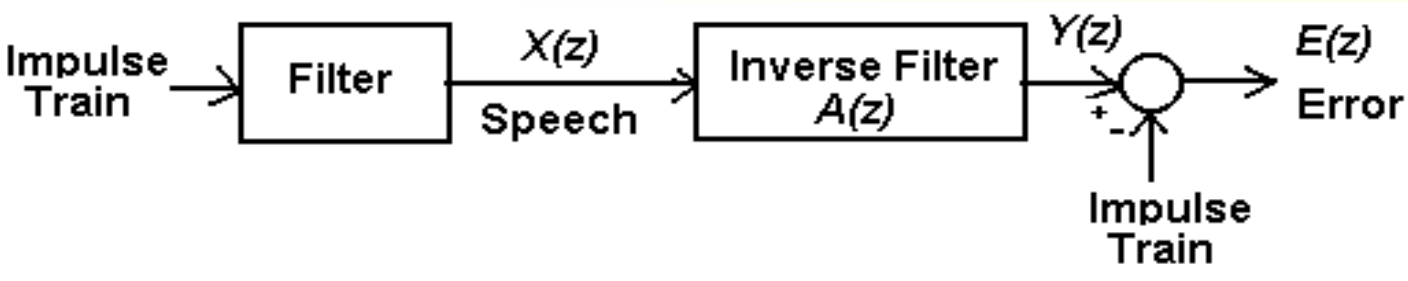
- **Optimum parameter values investigated**

  analysis window size, LPC order, sampling rate

- **Vowels & VCV syllables analyzed**

P.C. Pandey, EE Dept, IIT Bombay

Wakita's Speech Analysis Model (Wakita-1973)



Assumptions

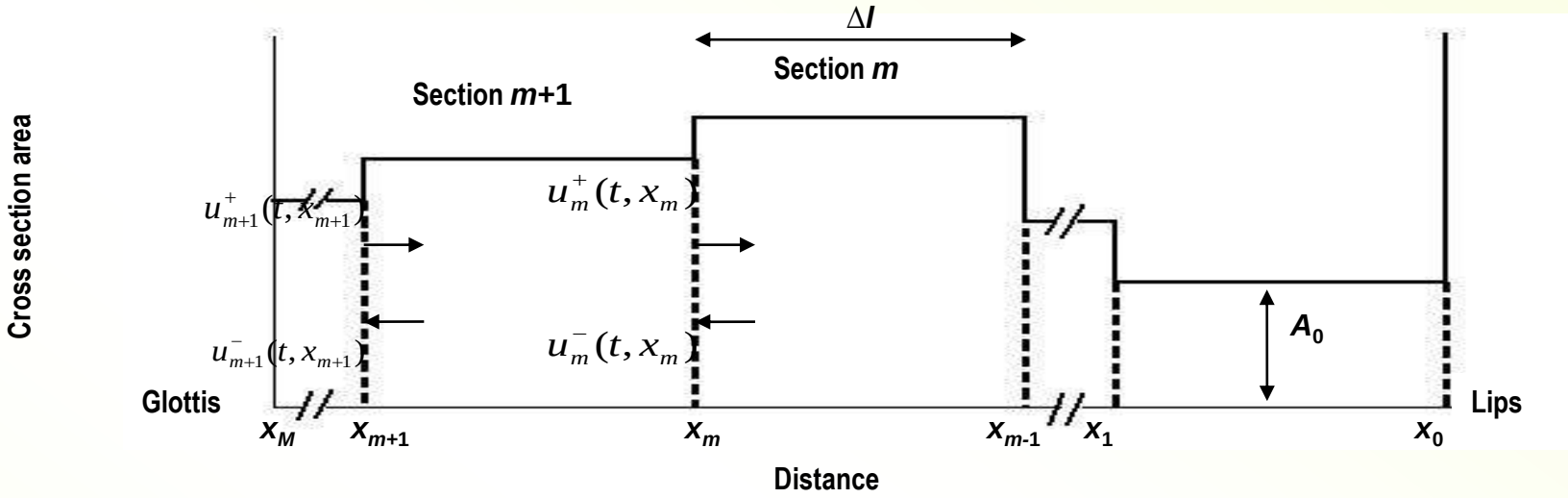
- Contributions of *glottal wave, vocal tract, & radiation impedance at lips included in the filter.*
- Speech to be analyzed: *periodic, non-nasalized, & voiced sound.*
- Power spectral envelope of speech signal: *approximated by poles only.*

Inverse filter coefficients: for min. the mean squared error.

Vocal tract: non-uniform acoustic tube filter, equivalent to the inverse filter.

P.C. Pandey, EE Dept, IIT Bombay

Acoustic Tube Model of the Vocal Tract



- **Assumptions:** plane wave propagation & no losses (viscosity & heat conduction)
- **At the m^{th} section:**

volume velocity: $$u_m(x,t) = u_m^+(x,t) - u_m^-(x,t)$$

pressure: $$p_m(x,t) = \frac{\rho c}{A_m}\left[u_m^+(x,t) + u_m^-(x,t)\right]$$

reflection coefficients: $$r_m = \frac{A_m - A_{m+1}}{A_m + A_{m+1}}$$

- **Reflection coefficients obtained from LPC analysis of speech signal.**

P.C. Pandey, EE Dept, IIT Bombay

Implementation for LPC Based Vocal Tract Shape Estimation

- F_s = 11.025 kHz
- LPC order = 12
- Analysis frame duration: twice the average pitch period
- Analysis window: Hamming
- Window shift: 5 ms
- Pre-emphasis for 6 dB/octave equalization

P.C. Pandey, EE Dept, IIT Bombay

Analysis of Vowels & VCV Syllables

- **Natural & synthesized vowels analyzed for shape estimation for**
  - checking consistency
  - studying effect of pitch and amplitude variation
- **VCV syllables involving semivowels (representing low energy, non-continuant, voiced sounds) for checking**
  - shape tracking during VC & CV transitions
- **VCV syllables involving stop consonants**

  for shape estimation during transition and closure segments
- **Use of VCV syllables for speech training:**

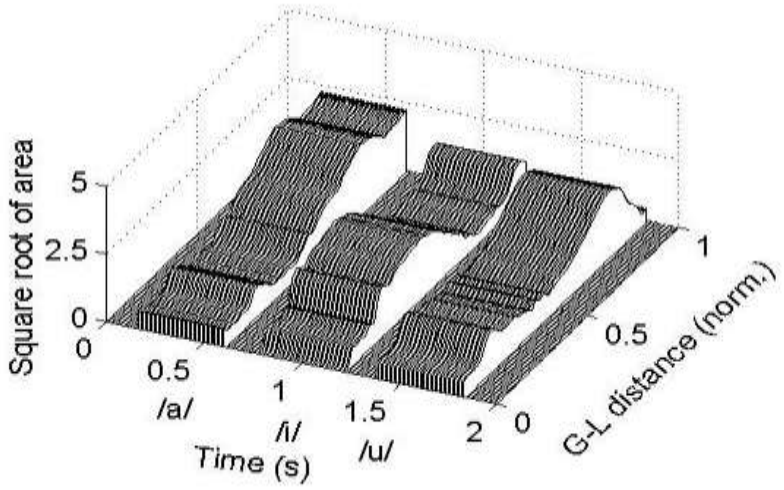
  short duration of dynamic shape display

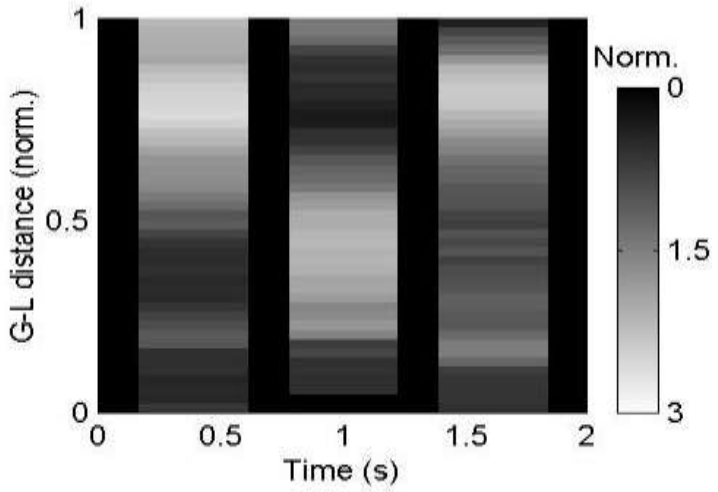
  → easier for a hearing impaired child to monitor & mimic

P.C. Pandey, EE Dept, IIT Bombay

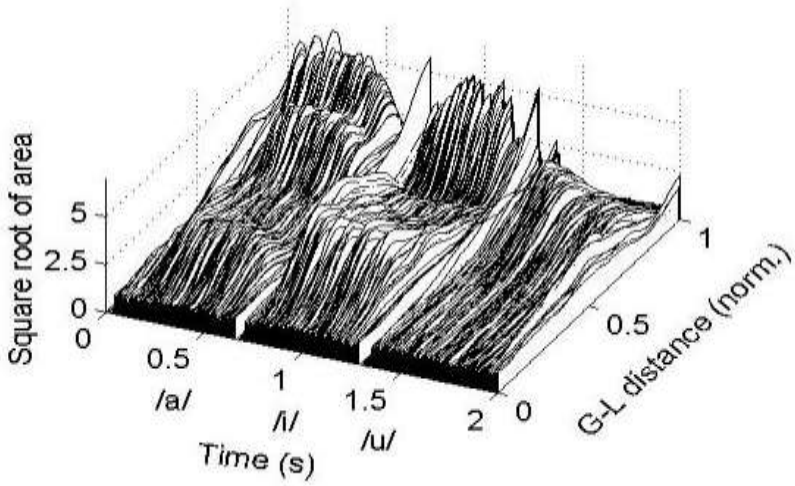
Comparison of shapes based on MRI & LPC

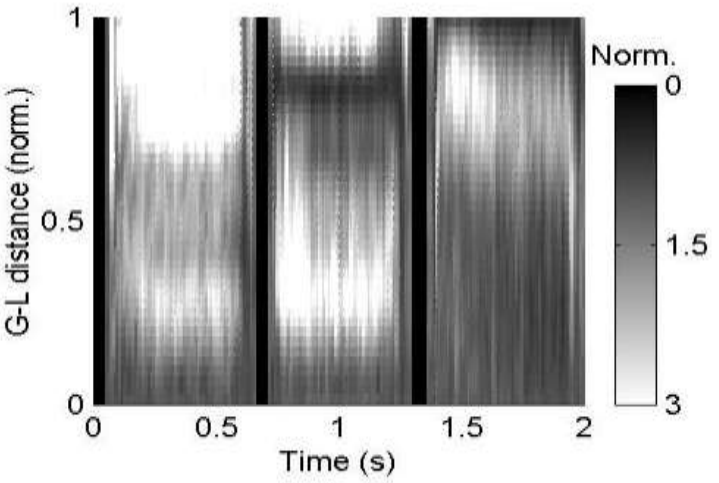
Based on MRI values





Based on LPC analysis





P.C. Pandey, EE Dept, IIT Bombay

Observations

LPC based estimated vocal tract shapes for vowels

- show proper tongue elevation
- compare well with shapes based on MRI data
- not affected by step/ramp pitch variation
- proper over an attenuation range of 0–40 dB.

P.C. Pandey, EE Dept, IIT Bombay

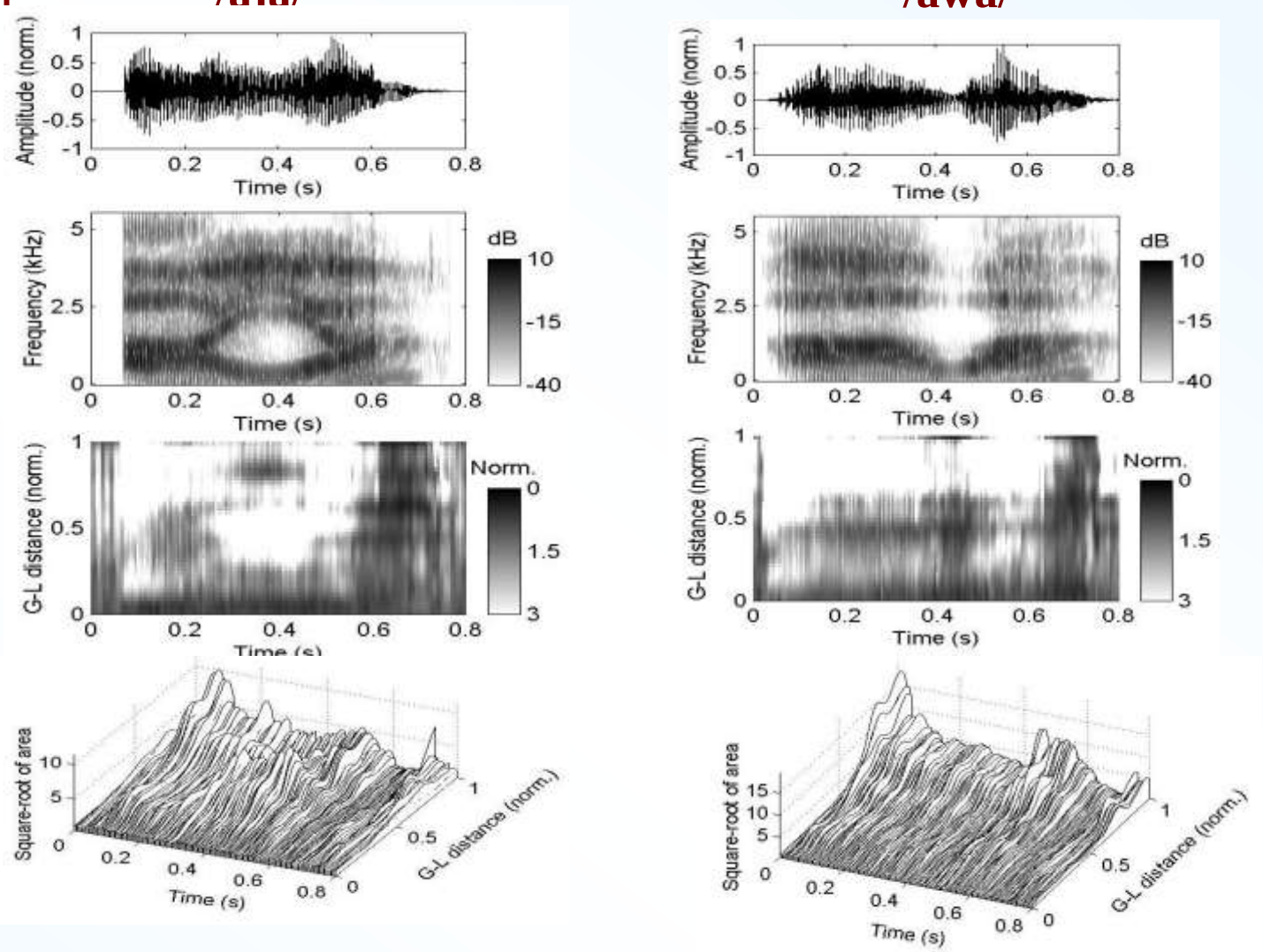
Analysis of /aia/ /awa/

(a)

(b)

(c)

(d)



(a) waveforms; (b) spectrograms; (c) areagrams; (d) waterfall diagram

P.C. Pandey, EE Dept, IIT Bombay

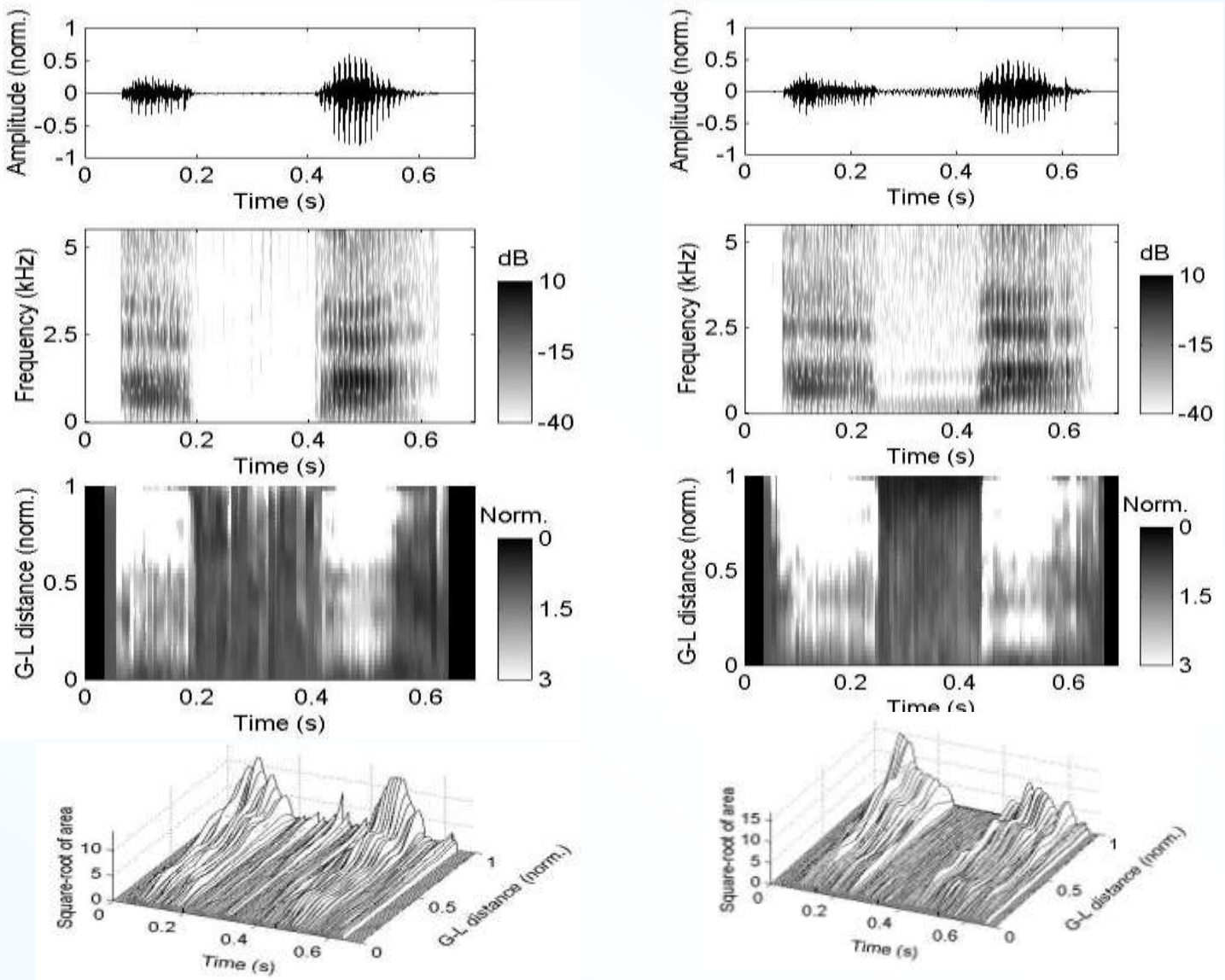
Analysis of /apa/ and /aba/

Wave-forms

Spectro-grams

Area-grams

Water-fall dia-grams



P.C. Pandey, EE Dept, IIT Bombay

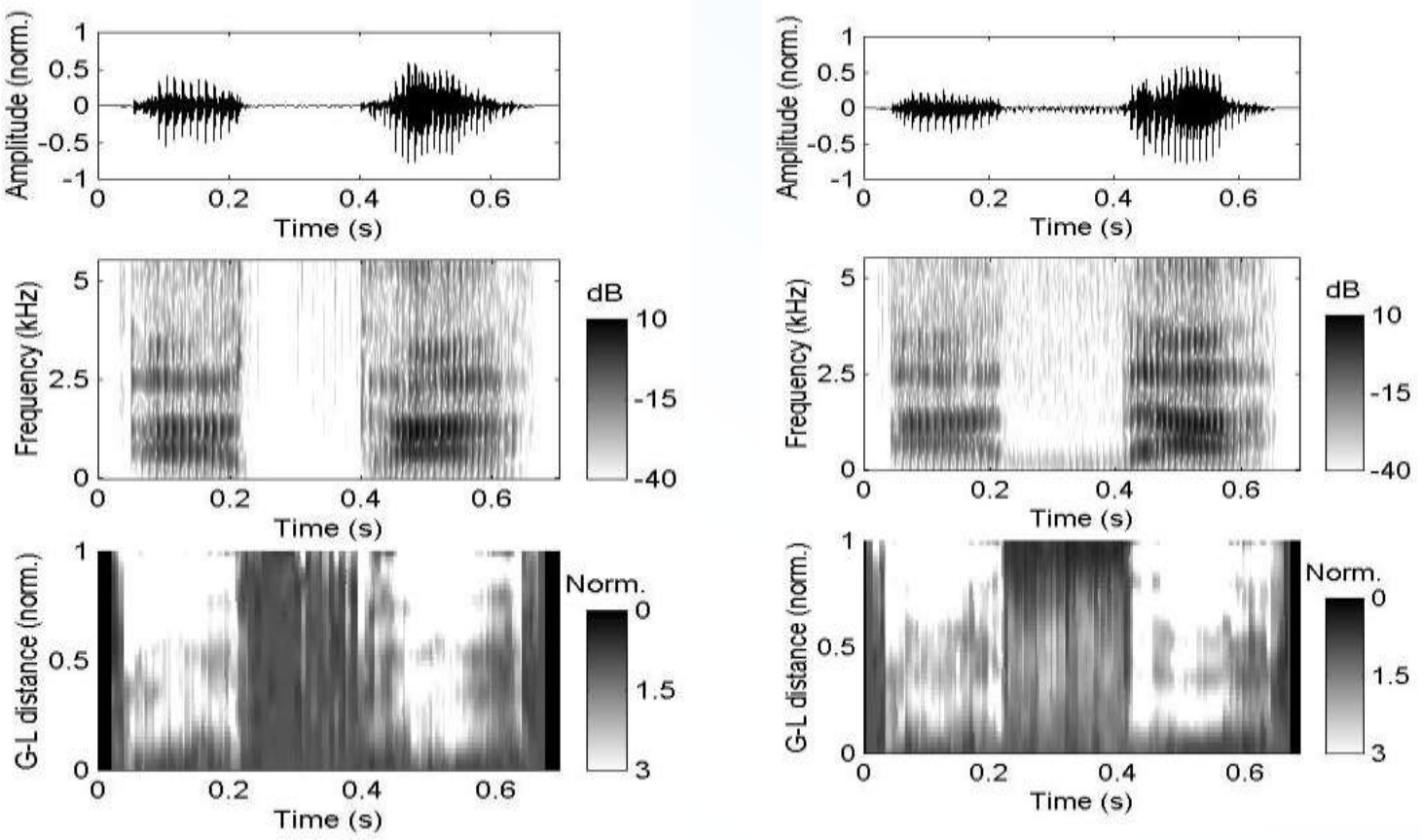
Analysis of /ata/ and /ada/

Wave-forms

Spectro-grams

Area-grams

Water-fall dia-grams



P.C. Pandey, EE Dept, IIT Bombay

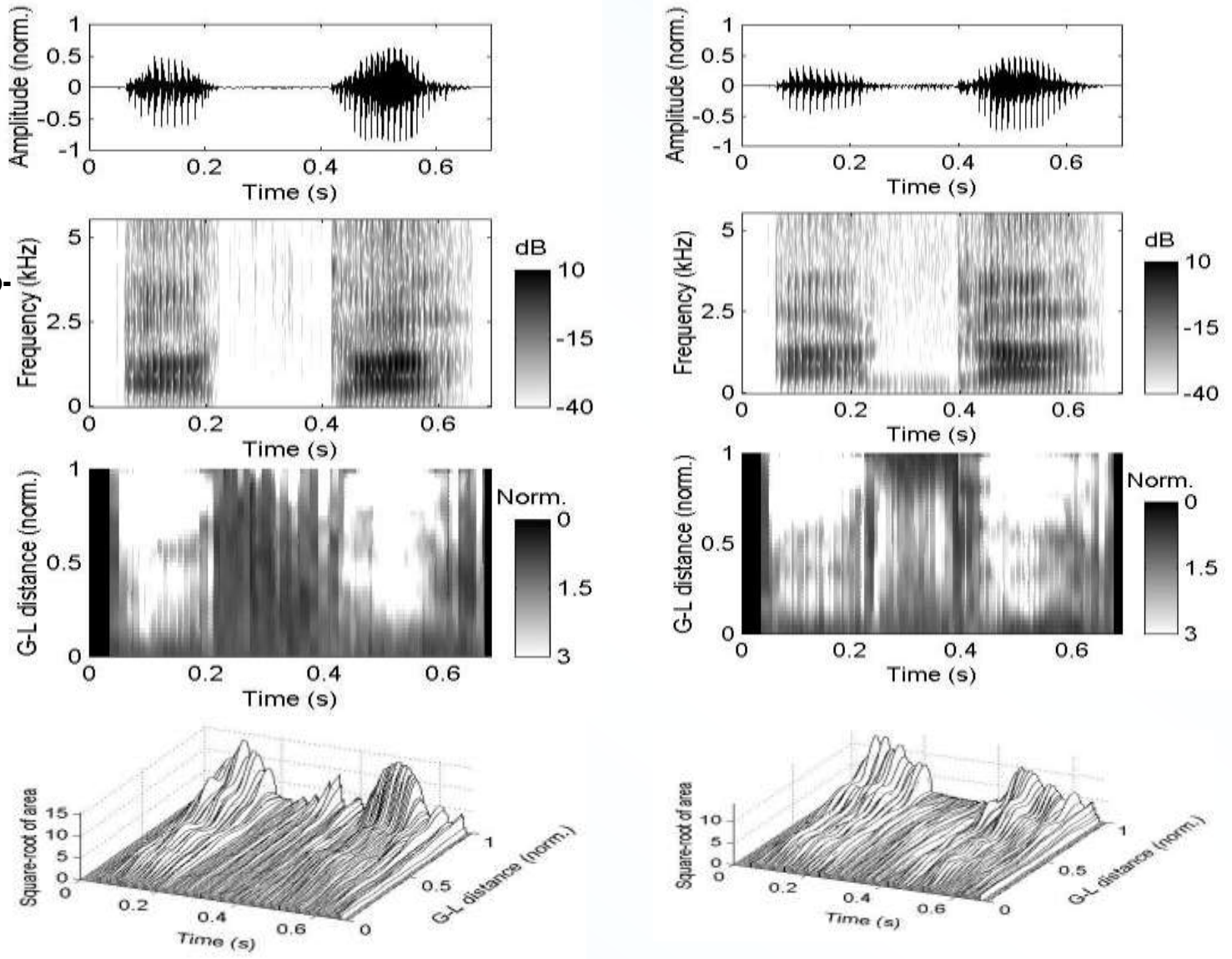
Analysis of /aka/ and /aga/

Wave-forms

Spectro-grams

Area-grams

Water-fall dia-grams



P.C. Pandey, EE Dept, IIT Bombay

Observations

- ***For semivowels***
  - place of constriction properly reflected in areagrams
- ***For stop consonants, estimated area values***
  - inconsistent during stop closures
  - distinctly different for different places of stop closures
  - related to movement of articulators during VC & CV transition segments and hence may contain information about the place of closure.

Further Investigations

Use of bivariate surfaces, representing values related to vocal tract shape (area values or coefficients of transfer function) over the VC & CV transition segments, for estimating the place of closure.

1. Introduction

2. Visual Speech-training Aids

3. LPC Based Vocal Tract Shape Estimation

4. Estimation of Vocal Tract Shape during Stop Closures

5. Results & Discussion

6. Summary & Conclusions

P.C. Pandey, EE Dept, IIT Bombay

Investigations

- 2D surface modeling of area values & line spectrum frequencies (LSFs) during VC & CV transition regions

  (LSFs are reported to behave well when interpolated)

- Modeling based on least-squares second & third degree bivariate polynomials & Delaunay triangulation based surfaces

  (articulatory dynamics may be accurately modeled by one of these surfaces)

- 2D surface interpolation during closure duration

  (for estimation of vocal tract shape and/or place of constriction)

P.C. Pandey, EE Dept, IIT Bombay

Least-squares polynomial approximation

- Find $f(x)$ that matches q data points g_n within a small error r_n i.e.,

$f(x_n) = g_n + r_n$ such that $E = \sum_{n=0}^{q-1} r_n^2$ is minimized

- In general, $f(x) = \sum_{k=0}^{p-1} c_k \Phi_k(x)$

where c_k: set of p parameters to be determined & Φ_k: set of a priori known functions

- In matrix notation, $\mathbf{Az} = \mathbf{b} + \mathbf{r}$

where,

$$\mathbf{A} = \begin{bmatrix} \Phi_0(x_0) & \Phi_1(x_0) & \dots & \Phi_{p-1}(x_0) \\ \Phi_0(x_1) & \Phi_1(x_1) & \dots & \Phi_{p-1}(x_1) \\ \vdots & \vdots & \ddots & \vdots \\ \Phi_0(x_{q-1}) & \Phi_1(x_{q-1}) & \dots & \Phi_{p-1}(x_{q-1}) \end{bmatrix}$$

$$\mathbf{z}^T = \begin{bmatrix} c_0 & c_1 & \cdots & c_{p-1} \end{bmatrix}$$

$$\mathbf{b}^T = \begin{bmatrix} g_0 & g_1 & \cdots & g_{q-1} \end{bmatrix}$$

$$\mathbf{r}^T = \begin{bmatrix} r_0 & r_1 & \cdots & r_{q-1} \end{bmatrix}$$

- To reduce interpolation errors: usually $p < q$
- Least-squares solution by pseudo-inverse of $\mathbf{A}$

$$\mathbf{z} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$$

P.C. Pandey, EE Dept, IIT Bombay

Surface modeling of variation in vocal tract shape with time

Least-squares approximation applied to bivariate data consisting of area values & LSFs, as a function of G-L distance and *t*, by

- 2nd and 3rd degree bivariate polynomial approximation
- Surface modeling by Delaunay triangulation

P.C. Pandey, EE Dept, IIT Bombay

Bivariate Polynomial Approximation

Area values and LSFs (represented by $g(x, y)$) during VC & CV transition regions approximated by second and third degree bivariate polynomial surfaces

2nd degree

$$f(x,y) = c_0 + c_1 x + c_2 y + c_3 xy + c_4 x^2 + c_5 y^2$$

3rd degree

$$f(x,y) = d_0 + d_1 x + d_2 x^2 + d_3 x^3 + d_4 y + d_5 y^2 + d_6 y^3 + d_7 xy + d_8 x^2 y + d_9 xy^2$$

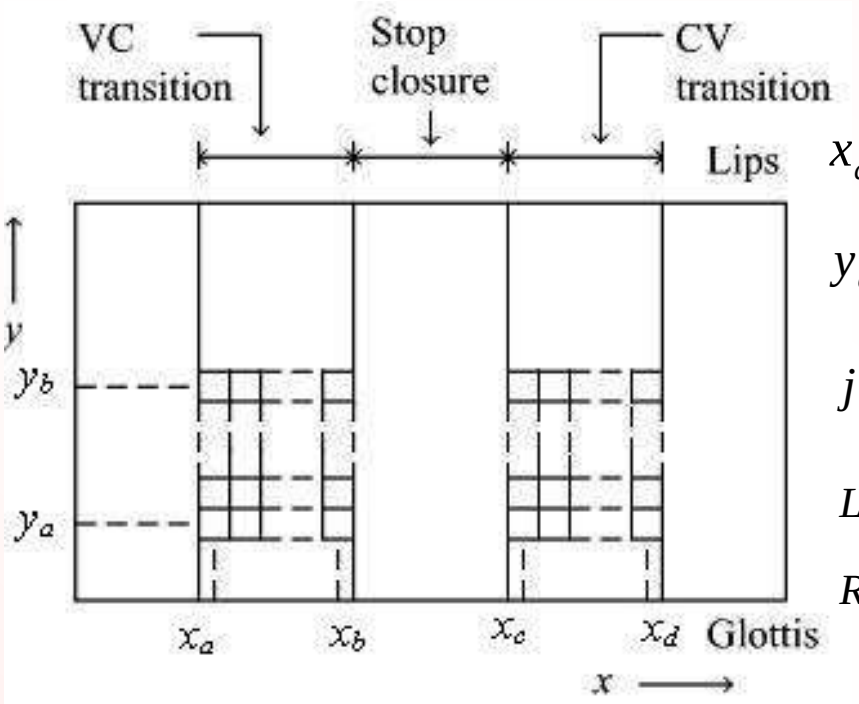
where $f(x, y)$ models $g(x, y)$ within a small error $r(x, y)$,

and c_0-c_5 & d_0-d_9 to be chosen to approximate $\{g(x, y)\}$

in the least-squares sense.

P.C. Pandey, EE Dept, IIT Bombay

Selection of area values (or LSFs) for surface approximation



$x_a \le x \le x_b$ and $x_c \le x \le x_d$

$y_a \le y \le y_b$ where $j = y_b - y_a \ge 2$ or 3

$j \le y_b \le 12$

$L_{col} = x_b - x_a + 1$

$R_{col} = x_d - x_c + 1$

- Overdetermined system of simultaneous linear eqns. for q > 6 for second degree & q >10 for third degree polynomial.
- Least-squares solution → approximated second or third degree surfaces.

P.C. Pandey, EE Dept, IIT Bombay

Simultaneous linear equations in matrix notation

$$\mathbf{Az = b+r}$$

For 2nd degree polynomial approximation,

$$\mathbf{b}^T = \begin{bmatrix} g(x_a,y_a) & g(x_a,y_{a+1}) & \cdots & g(x_a,y_b) & g(x_{a+1},y_a) & g(x_{a+1},y_{a+1}) & \cdots & g(x_{a+1},y_b) & \cdots & g(x_d,y_b) \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & x_a & y_a & x_a y_a & x_a^2 & y_a^2 \\ 1 & x_a & y_{a+1} & x_a y_{a+1} & x_a^2 & y_{a+1}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_a & y_b & x_a y_b & x_a^2 & y_b^2 \\ 1 & x_{a+1} & y_a & x_{a+1} y_a & x_{a+1}^2 & y_a^2 \\ 1 & x_{a+1} & y_{a+1} & x_{a+1} y_{a+1} & x_{a+1}^2 & y_{a+1}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{a+1} & y_b & x_{a+1} y_b & x_{a+1}^2 & y_b^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_d & y_b & x_d y_b & x_d^2 & y_b^2 \end{bmatrix}$$

$$\mathbf{z}^T = \begin{bmatrix} c_0 & c_1 & c_2 & c_3 & c_4 & c_5 \end{bmatrix}$$

P.C. Pandey, EE Dept, IIT Bombay

For 3rd degree polynomial approximation,

$$\mathbf{b}^T = \begin{bmatrix} g(x_a, y_a) & g(x_a, y_{a+1}) & \cdots & g(x_a, y_b) & g(x_{a+1}, y_a) & g(x_{a+1}, y_{a+1}) & \cdots & g(x_{a+1}, y_b) & \cdots & g(x_d, y_b) \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 1 & x_a & x_a^2 & x_a^3 & y_a & y_a^2 & y_a^3 & x_a y_a & x_a^2 y_a & x_a y_a^2 \\ 1 & x_a & x_a^2 & x_a^3 & y_{a+1} & y_{a+1}^2 & y_{a+1}^3 & x_a y_{a+1} & x_a^2 y_{a+1} & x_a y_{a+1}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_a & x_a^2 & x_a^3 & y_b & y_b^2 & y_b^3 & x_a y_b & x_a^2 y_b & x_a y_b^2 \\ 1 & x_{a+1} & x_{a+1}^2 & x_{a+1}^3 & y_a & y_a^2 & y_a^3 & x_{a+1} y_a & x_{a+1}^2 y_a & x_{a+1} y_a^2 \\ 1 & x_{a+1} & x_{a+1}^2 & x_{a+1}^3 & y_{a+1} & y_{a+1}^2 & y_{a+1}^3 & x_{a+1} y_{a+1} & x_{a+1}^2 y_{a+1} & x_{a+1} y_{a+1}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{a+1} & x_{a+1}^2 & x_{a+1}^3 & y_b & y_b^2 & y_b^3 & x_{a+1} y_b & x_{a+1}^2 y_b & x_{a+1} y_b^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_d & x_d^2 & x_d^3 & y_b & y_b^2 & y_b^3 & x_d y_b & x_d^2 y_b & x_d y_b^2 \end{bmatrix}$$

$$\mathbf{z}^T = \begin{bmatrix} d_0 & d_1 & d_2 & d_3 & d_4 & d_5 & d_6 & d_7 & d_8 & d_9 \end{bmatrix}$$

P.C. Pandey, EE Dept, IIT Bombay

Least-squares solution for simultaneous linear equations by pseudo- inverse operation gives $\mathbf{z} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$

2D interpolation of polynomial surfaces during stop closures for $x_b < x < x_c, \; y = y_b$ and $\mathbf{z}$ as obtained above

carried out using $\hat{\mathbf{b}} = \hat{\mathbf{A}}\,\mathbf{z}$

where

$$\hat{\mathbf{A}} = \begin{bmatrix} 1 & x_{b+1} & y_b & x_{b+1}y_b & x_{b+1}^2 & y_b^2 \\ 1 & x_{b+2} & y_b & x_{b+2}y_b & x_{b+2}^2 & y_b^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{c-1} & y_b & x_{c-1}y_b & x_{c-1}^2 & y_b^2 \end{bmatrix}$$

for 2nd degree polynomial

$$\hat{\mathbf{A}} = \begin{bmatrix} 1 & x_{b+1} & x_{b+1}^2 & x_{b+1}^3 & y_b & y_b^2 & y_b^3 & x_{b+1}y_b & x_{b+1}^2y_b & x_{b+1}y_b^2 \\ 1 & x_{b+2} & x_{b+2}^2 & x_{b+2}^3 & y_b & y_b^2 & y_b^3 & x_{b+2}y_b & x_{b+2}^2y_b & x_{b+2}y_b^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{c-1} & x_{c-1}^2 & x_{c-1}^3 & y_b & y_b^2 & y_b^3 & x_{c-1}y_b & x_{c-1}^2y_b & x_{c-1}y_b^2 \end{bmatrix}$$

for 3rd degree polynomial

P.C. Pandey, EE Dept, IIT Bombay

Delaunay Triangulation Based Surface Modeling

- **Triangulation involves**

  subdivision of an area (volume) into triangles (tetrahedrons)

- **Delaunay triangulation & its properties**
  - A set of lines connecting each point to its natural neighbor
  - No data points are contained within a circle circumscribing the triangles
  - Maximizes the smallest angle over all triangulation

- **Delaunay surface modeling of**

  area values & LSFs during VC & CV transition regions carried out (using Matlab® functions)

- **For estimation of vocal tract shape and/or place of constriction**

  2D Delaunay surface interpolation during stop closure carried out

P.C. Pandey, EE Dept, IIT Bombay

Estimation of Stop Closure Boundary Locations

within a VCV syllable, for polynomial & Delaunay surface generation & its interpolation

- **Estimation** (2-step process)
  - step 1: estimation of VCV syllable end-points
  - step 2: based on step 1, stop closure boundary locations estimated
  - *(using avg. short time magnitude & empirically selected thresholds)*
- **Estimated stop closure end location**
  - shifted beyond the fricative burst during CV transition
  - *(as LPC based area estimation during turbulent noise inconsistent & uncorrelated to place of articulation)*

P.C. Pandey, EE Dept, IIT Bombay

Validation of the Proposed Technique

- Vowels /a/, /i/, & /u/ *(static vocal tract shape & formants)*
- VCV syllables /aja/ & /awa/ *(dynamic vocal tract shape & formant transitions)*
- Vowels and VCV syllables /aja/ & /awa/ with artificially silenced middle segment for proper recovery of vocal tract shape and/or place of articulation during silence gap.

P.C. Pandey, EE Dept, IIT Bombay

- Validation carried out with artificially silenced segments of different length for
  - Vowels synthesized & recorded for a male speaker.
  - VCV syllables recorded for three male (SM1, SM2, & SM3) and two female (SF4 & SF5) speakers.

- Analysis of VCV syllables for estimation of
  - minimum transition segments required
  - typical surface generation & interpolation parameters required

    (no. of frames to the left L_{col} & right R_{col} of silence gap and no. of rows j)

for proper recovery of vocal tract shape and/or place of articulation.

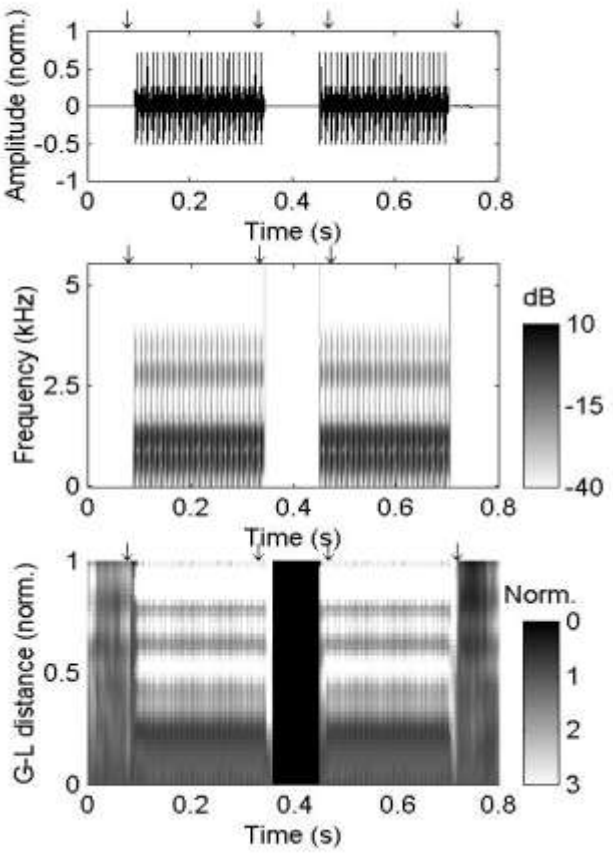
P.C. Pandey, EE Dept, IIT Bombay

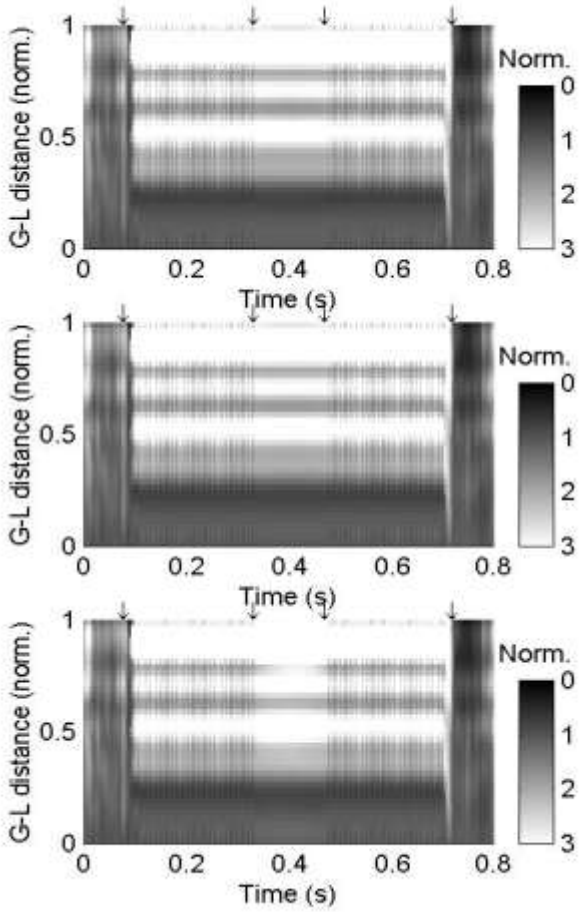
Result 1: Analysis of synthesized vowel /a/ (interpolation of area values)

(a)

(b)

(c)





(d) (d)

(e) (e)

(f) (f)

(a) waveform; (b) spectrogram (Δf = 300 Hz); (c) original areagram; (d), (e), and (f) areagrams obtained after 2D interpolation of second deg., third deg., & Delaunay surfaces respectively
(surface generation parameters j = 5, L_{col} = 2, and R_{col} = 2)

P.C. Pandey, EE Dept, IIT Bombay

Result 2: Analysis of /aja/ (speaker SM1)

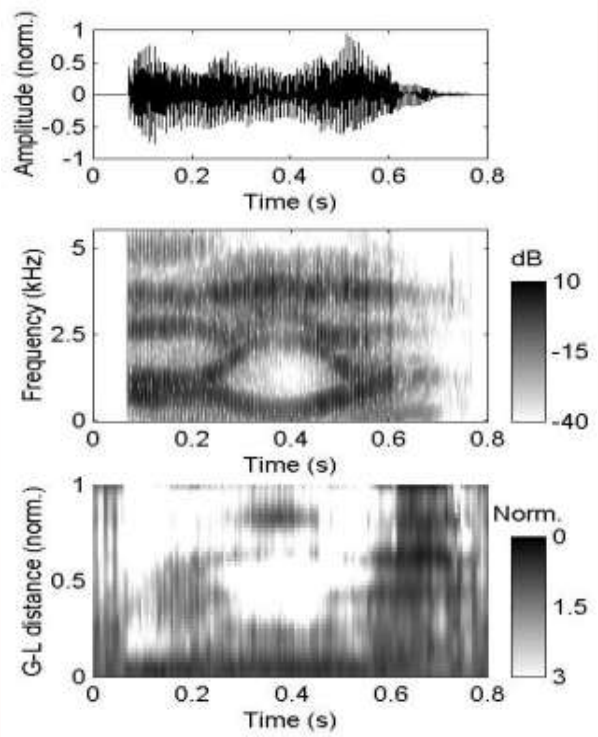
(VC & CV transition segments: 120 ms each, middle nearly steady state segment: 70 ms)

(a) waveform

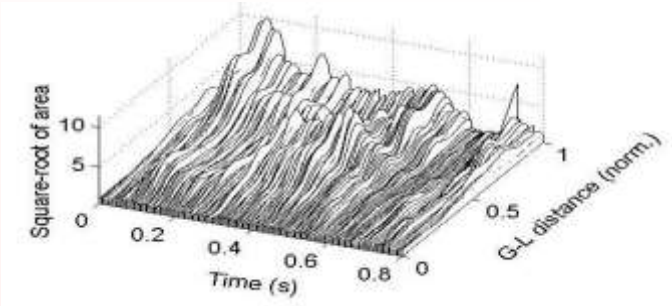
(b) Spectrogram

(Δf = 300 Hz)

(c) Original areagram



(d) Original waterfall diagram



P.C. Pandey, EE Dept, IIT Bombay

Result 3: 2D interpolation of area values for /aja/ (case 1, speaker SM1)

(a) waveform

(b) Spectrogram

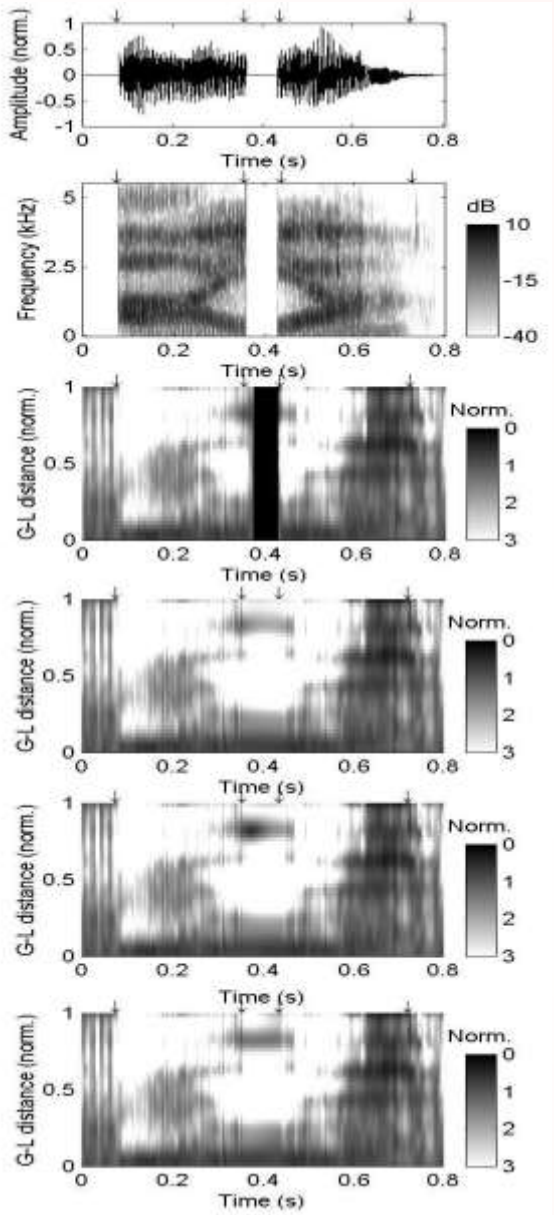
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial interpolation

(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Silence gap: 70 ms

Available VC & CV transition segments:

120 ms each

Surface generation parameters:

$j = 3$, $L_{col} = 3$, $R_{col} = 3$

P.C. Pandey, EE Dept, IIT Bombay

Result 4: 2D interpolation of area values for /aja/ (case 2, speaker SM1)

(a) waveform

(b) Spectrogram

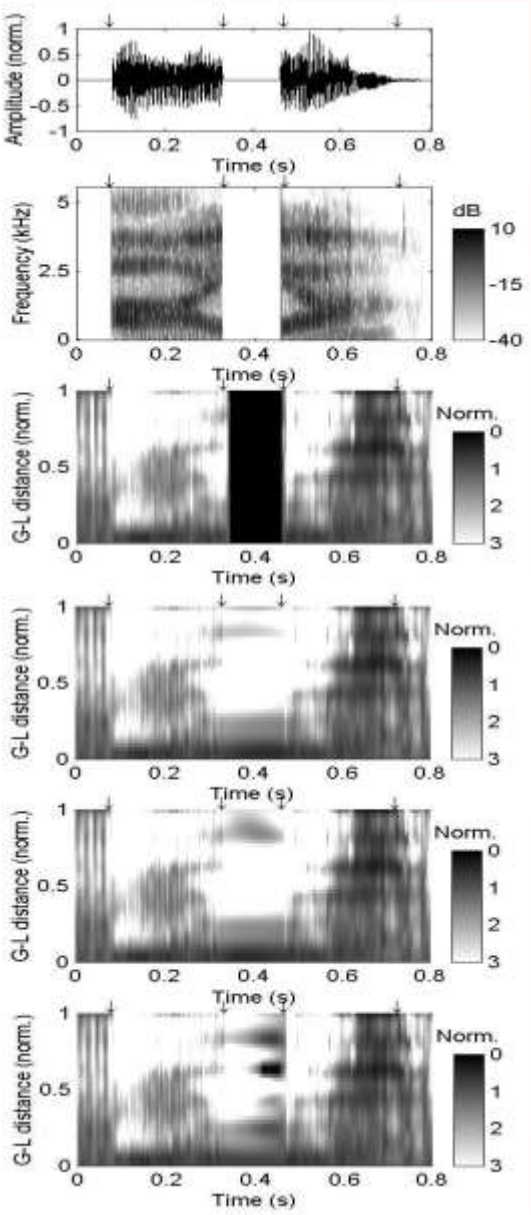
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial interpolation

(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Silence gap: 130 ms

Available VC & CV transition segments:

90 ms each

Surface generation parameters:

$j = 3$, $L_{col} = 6$, $R_{col} = 6$

P.C. Pandey, EE Dept, IIT Bombay

Result 5: 2D interpolation of area values for /aja/ (case 3, speaker SM1)

(a) waveform

(b) Spectrogram

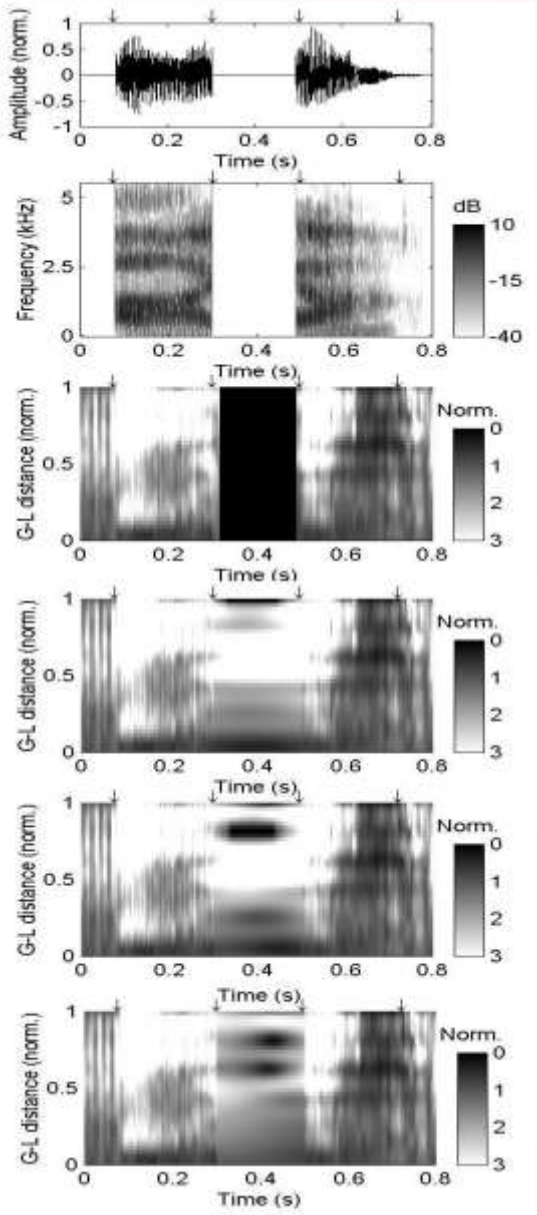
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree surface interpolation

(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Silence gap: 190 ms

Available VC & CV transition segments:

60 ms each

Surface generation parameters:

$j = 3$, $L_{col} = 8$, $R_{col} = 8$

P.C. Pandey, EE Dept, IIT Bombay

Result 6: 2D interpolation of area values for /aja/ (case 4, speaker SM1)

(a) waveform

(b) Spectrogram

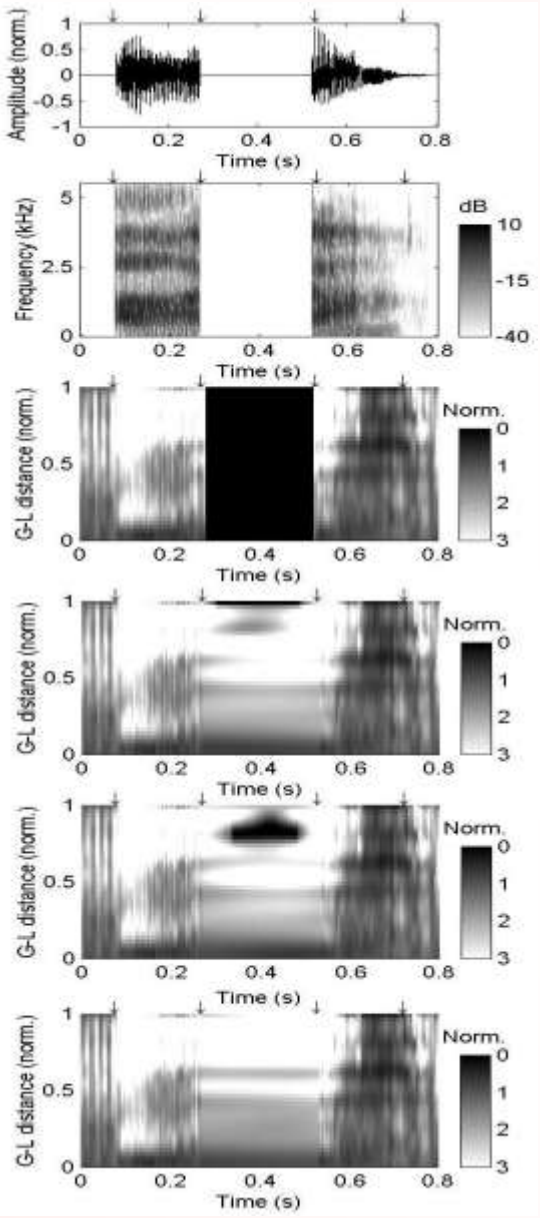
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree surface interpolation

(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



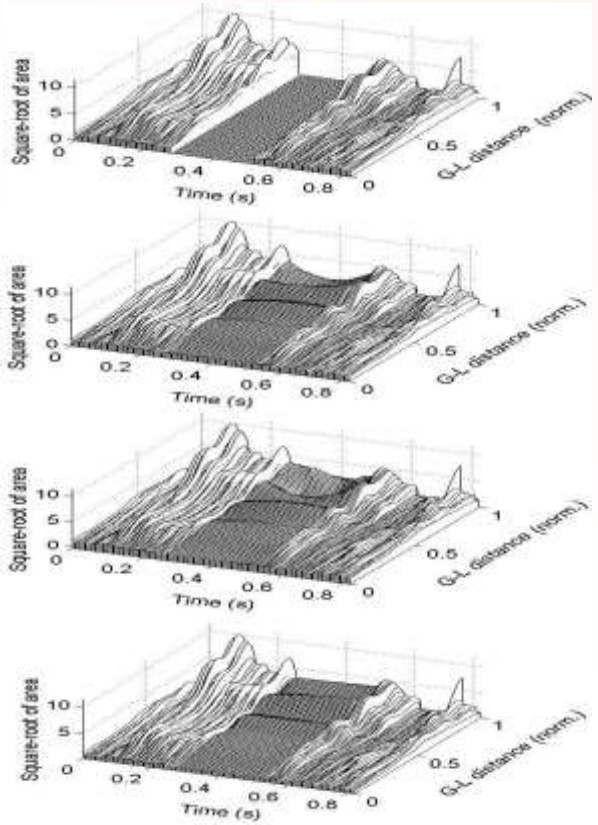
Silence gap: 250 ms

Available VC & CV transition segments:

30 ms each

Surface generation parameters:

$j = 3$, $L_{col} = 7$, $R_{col} = 7$



P.C. Pandey, EE Dept, IIT Bombay

Observations

- **Second & third degree polynomial surface interpolation of area values results in proper estimation of**
  - vocal tract shape for the first two cases (transition seg. = 120, 90 ms)
  - place of constriction for all the four cases (tr. seg. = 120, 90, 60, 30 ms)
- **Delaunay triangulation based surface interpolation of area values**
  - proper estimation of place of constriction for the first two cases
- ***Minimum 30 ms of VC & CV transition segments required for proper estimation of place of articulation.***

P.C. Pandey, EE Dept, IIT Bombay

Table 1 Summary of analysis results for /aja/

Sp.	Values used for surface modeling and 2D interpolation of									
	Area values					LSFs				
	Surface	Min. trans width (ms)		No. of frames		Surface	Min. trans width (ms)		No. of frames	
		VC	CV	L_{col}	R_{col}		VC	CV	L_{col}	R_{col}
SM1	2nd deg., 3rd deg.	30	30	7	7	2nd deg., Del.	50	50	3	3
SM2	2nd , 3rd deg., Del.	25	25	6	6	2nd deg., 3rd deg.	25	25	8	8
SM3	2nd , 3rd deg.	30	30	5	5	3rd.deg.	30	30	7	7
SF4	2nd , 3rd deg.	30	35	6	6	3rd deg., Del	30	35	4	9
SF5	2nd deg.	35	25	2	2	3rd deg.	35	25	9	9
	Mean	30	29	5.2	5.2	Mean	34	33	6.2	7.2

P.C. Pandey, EE Dept, IIT Bombay

Table 2 Summary of analysis results for /awa/

Sp.	Values used for surface modeling and 2D interpolation of									
	Area values					LSFs				
	Surface	Min. trans width (ms)		No. of frames		Surface	Min. trans width (ms)		No. of frames	
		VC	CV	L_{col}	R_{col}		VC	CV	L_{col}	R_{col}
SM1	2nd deg., 3rd deg.	30	30	7	7	2nd deg., Del.	30	30	8	8
SM2	2nd , 3rd deg., Del.	30	30	5	5	2nd , 3rd deg., Del.	30	30	4	4
SM3	2nd deg.	30	30	5	5	2nd , 3rd deg., Del.	70	70	3	3
SF4	2nd deg.	30	30	5	5	2nd , 3rd deg.	30	30	7	7
SF5	2nd , 3rd deg.	20	20	6	6	2nd , 3rd deg.	40	20	6	6
	Mean	28	28	5.6	5.6	Mean	40	30	5.6	5.6

P.C. Pandey, EE Dept, IIT Bombay

Observations

- Proper estimation of place of articulation dependent on
  - type of surface used for modeling of VC & CV transition values
  - number of frames used during VC & CV transition segments
- Minimum required transition width in a syllable is more in case of surface modeling of LSFs compared to surface modeling of area values.
- 2D interpolation based on second degree polynomial surface approximation of area values & LSFs found to be the most successful technique
  - required minimum mean VC & CV transition segments of 31.5 ms each

P.C. Pandey, EE Dept, IIT Bombay

P.C. Pandey, EE Dept, IIT Bombay

- 2D interpolation based on

  second degree, third degree, & Delaunay surfaces representing area values and LSFs

applied to VCV syllables of the type

/aCa/, /iCa/ (recorded for 3 male & 2 female speakers)

/aCi/, /iCi/, & /uCu/ (recorded for a male speaker)

involving stop consonants /p/, /b/, /t/, /d/, /k/, & /g/

for the estimation of place of closure.

- Estimated place of constriction compared with

  earlier reported articulation places based on MRI & X-ray images

- Typical range for the place of constriction, with normalized distance of 0 to 1 (0: glottis, 1: lips)

  Bilabial stops (p, b) : 1

  Alveolar stops (t, d) : 0.75 to 0.89

  Velar stops (k, g) : 0.47 to 0.7

P.C. Pandey, EE Dept, IIT Bombay

Result 1: 2D interpolation of area values for /apa/ (speaker SM1)

(a) waveform

(b) Spectrogram

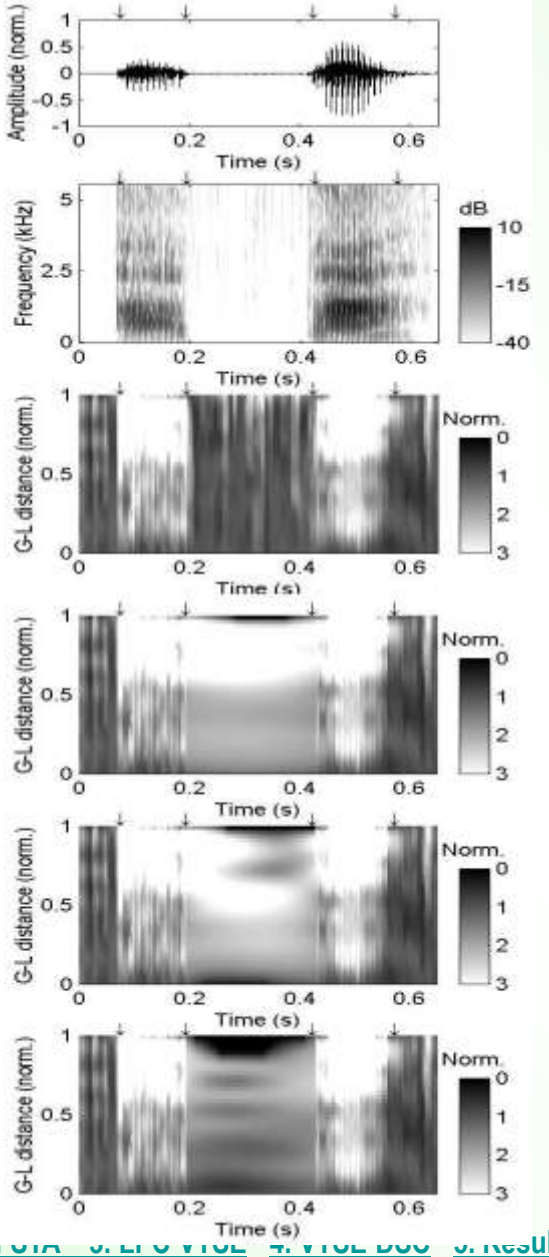
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

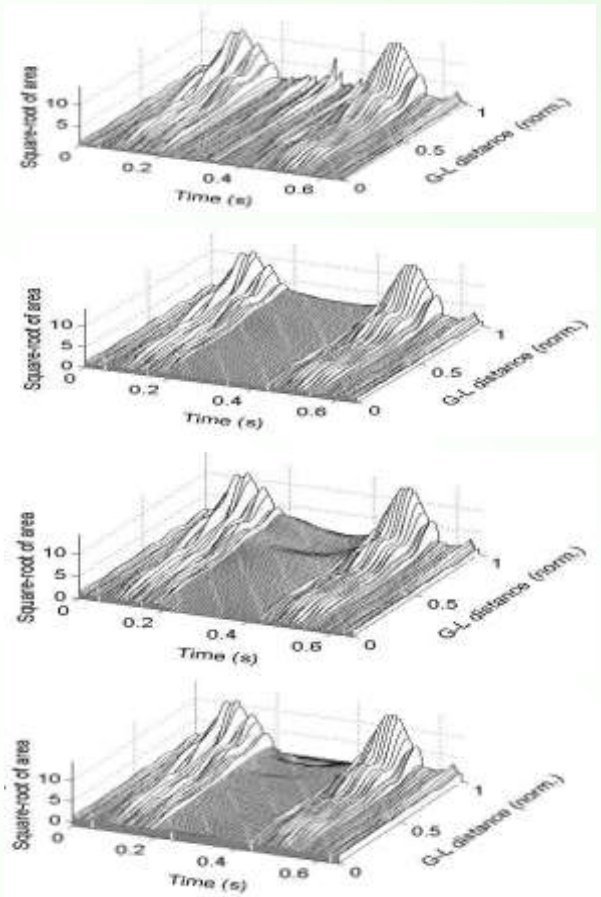
(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 5$, $L_{col} = 3$, $R_{col} = 3$



P.C. Pandey, EE Dept, IIT Bombay

Result 2: 2D interpolation of area values for /aba/ (speaker SM1)

(a) waveform

(b) Spectrogram

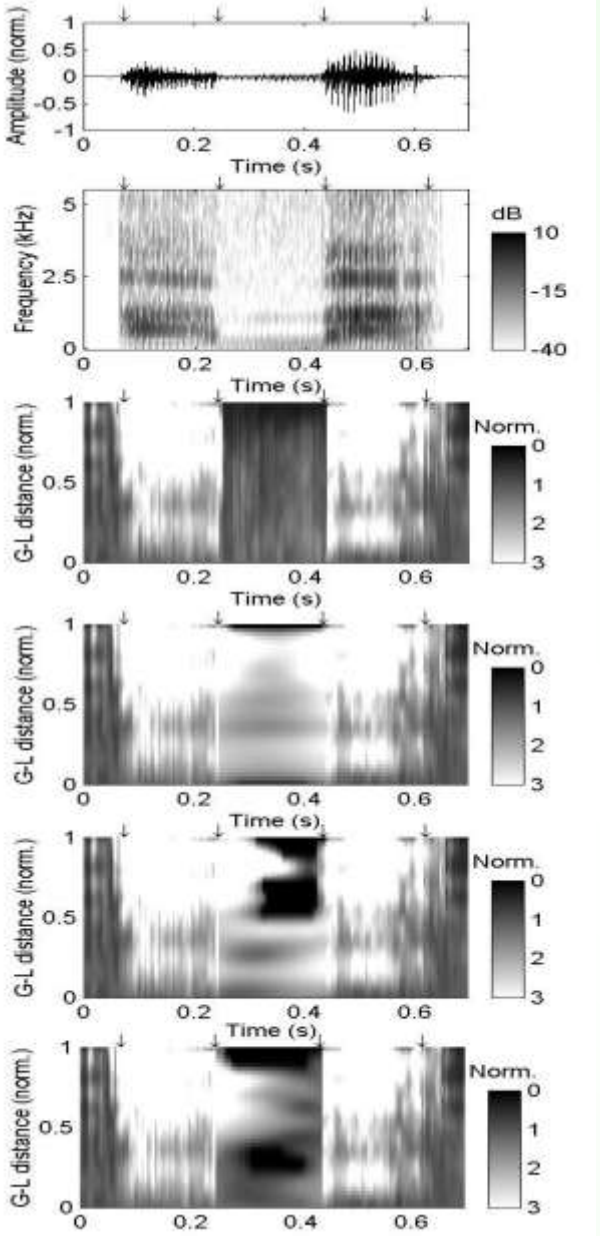
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

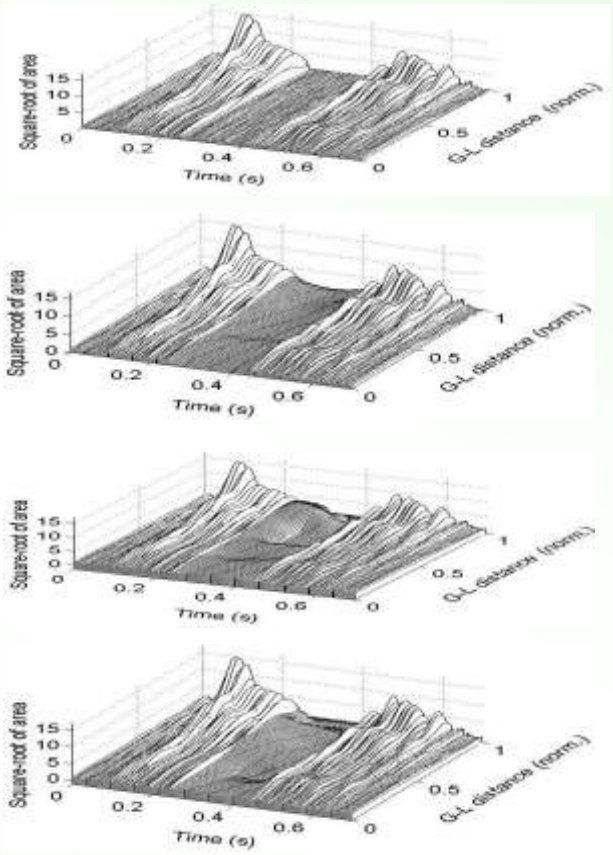
(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 5$, $L_{col} = 2$, $R_{col} = 2$



P.C. Pandey, EE Dept, IIT Bombay

Result 3: 2D interpolation of area values for /ata/ (speaker SM1)

(a) waveform

(b) Spectrogram

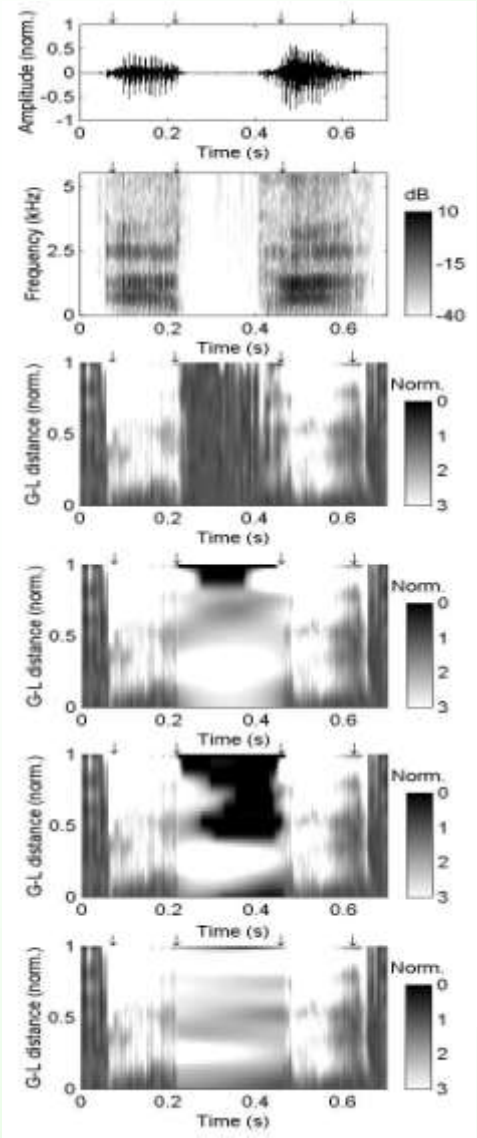
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

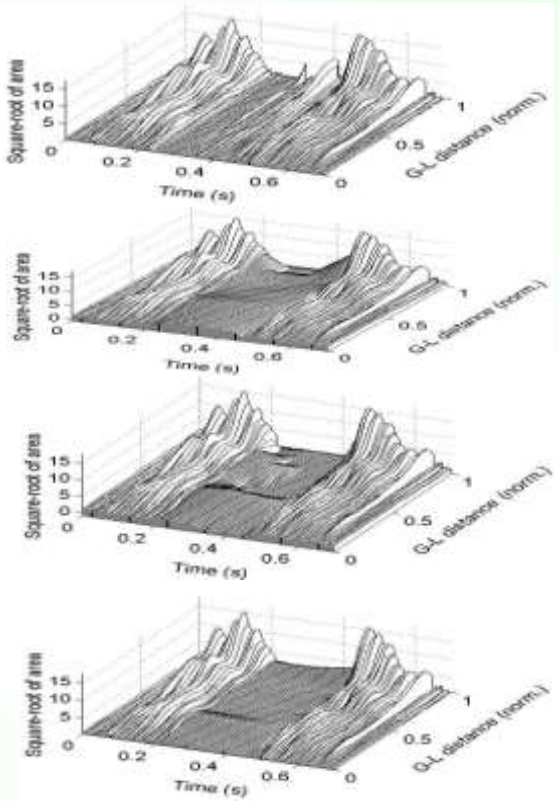
(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 4$, $L_{col} = 5$, $R_{col} = 4$



P.C. Pandey, EE Dept, IIT Bombay

Result 4: 2D interpolation of area values for /ada/ (speaker SM1)

(a) waveform

(b) Spectrogram

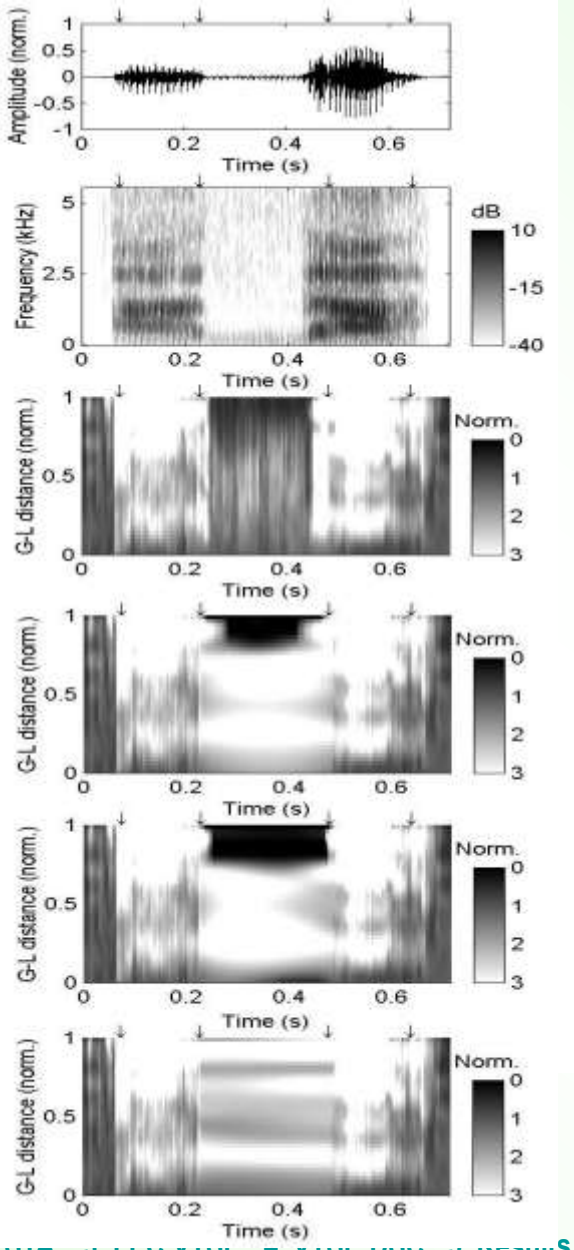
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

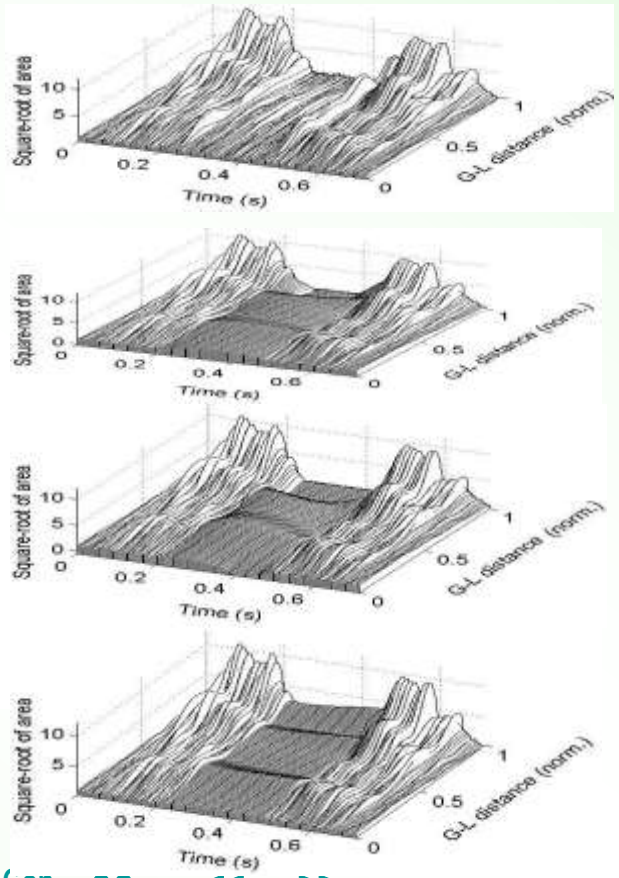
(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 4$, $L_{col} = 4$, $R_{col} = 4$



P.C. Pandey, EE Dept, IIT Bombay

Result 5: 2D interpolation of area values for /aka/ (speaker SM1)

(a) waveform

(b) Spectrogram

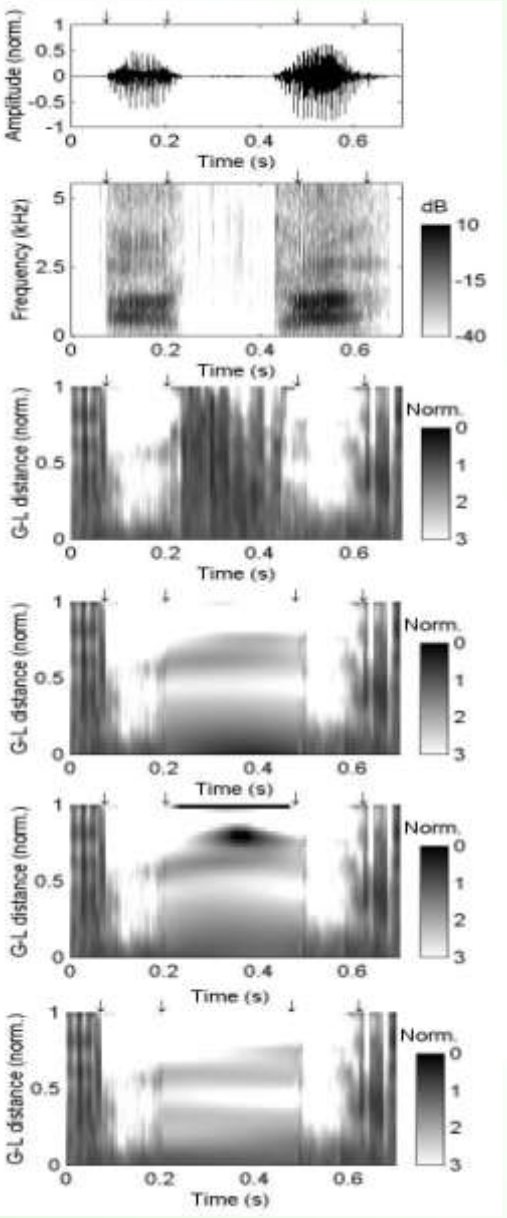
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 6$, $L_{col} = 4$, $R_{col} = 4$

P.C. Pandey, EE Dept, IIT Bombay

Result 6: 2D interpolation of area values for /aga/ (speaker SM1)

(a) waveform

(b) Spectrogram

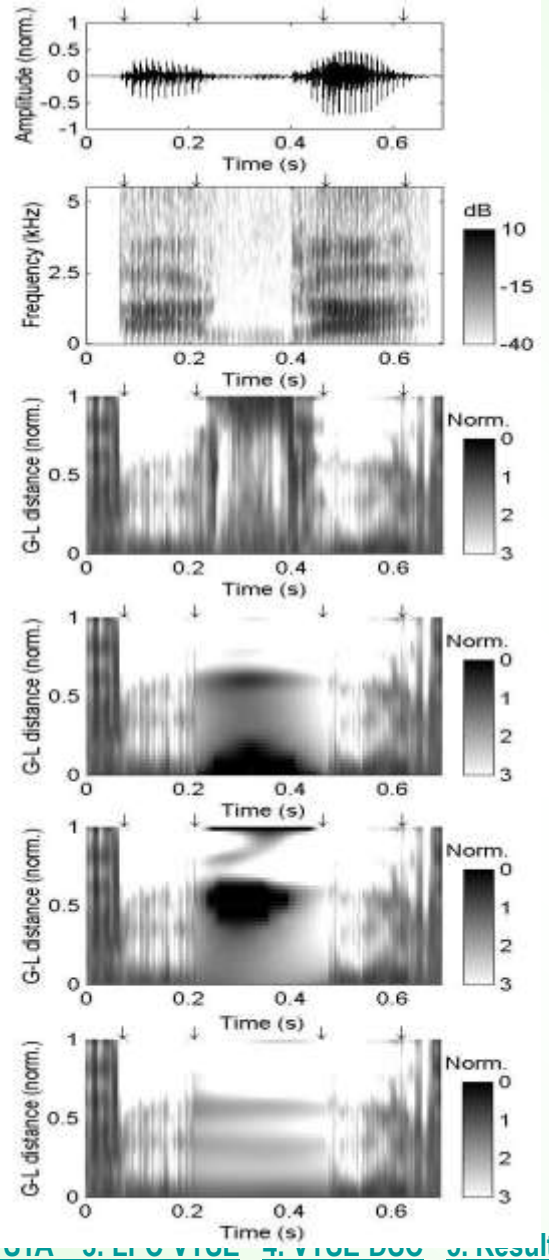
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

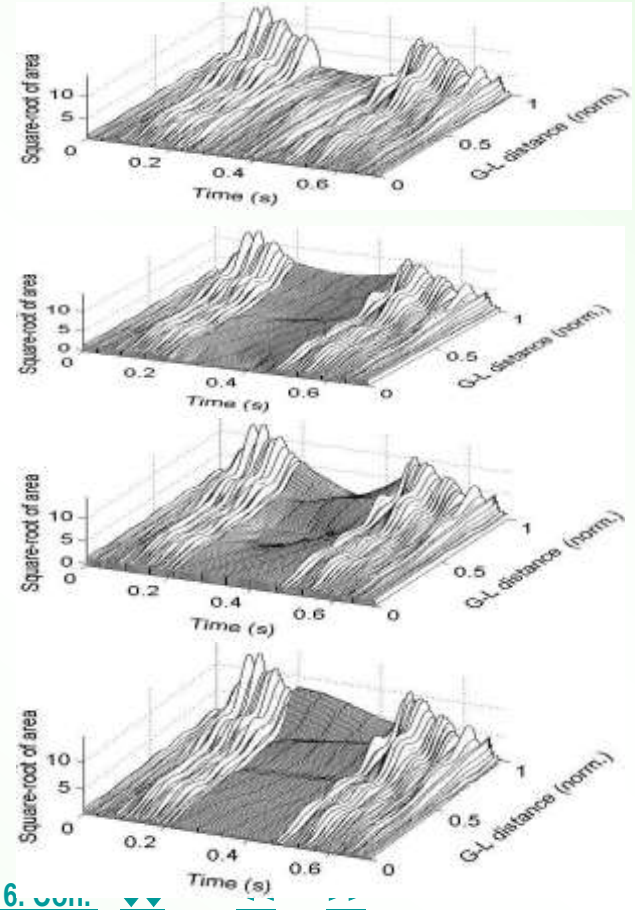
(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 7$, $L_{col} = 5$, $R_{col} = 3$



P.C. Pandey, EE Dept, IIT Bombay

Result 7: 2D interpolation of LSFs for /apa/ (speaker SM1)

(a) waveform

(b) Spectrogram

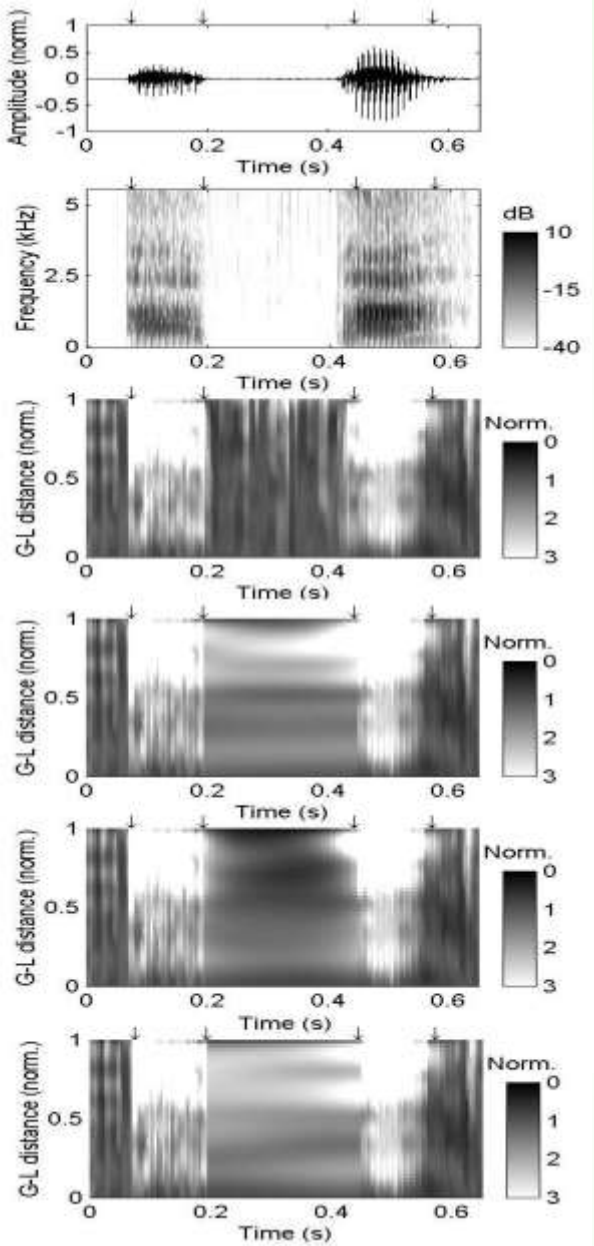
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

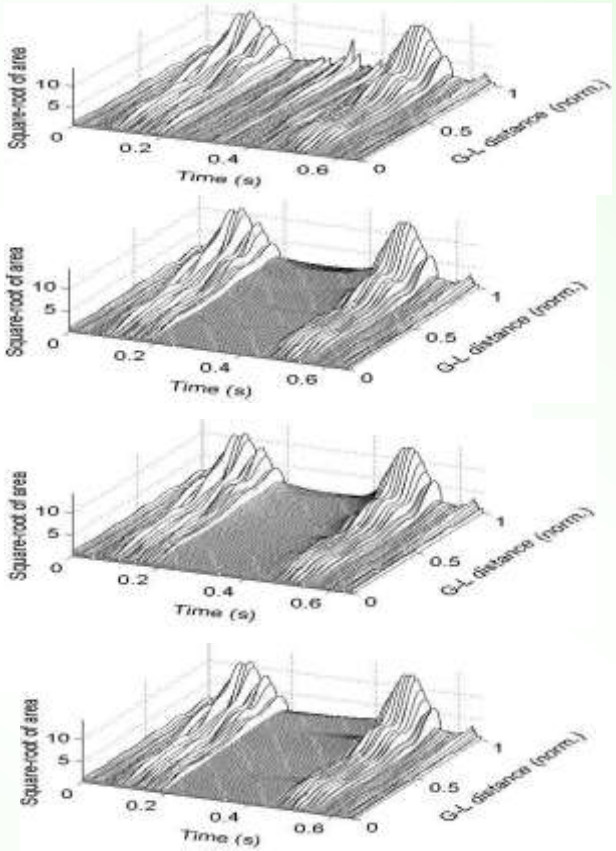
(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 4$, $L_{col} = 6$, $R_{col} = 6$



P.C. Pandey, EE Dept, IIT Bombay

Result 8: 2D interpolation of LSFs for /aba/ (speaker SM1)

(a) waveform

(b) Spectrogram

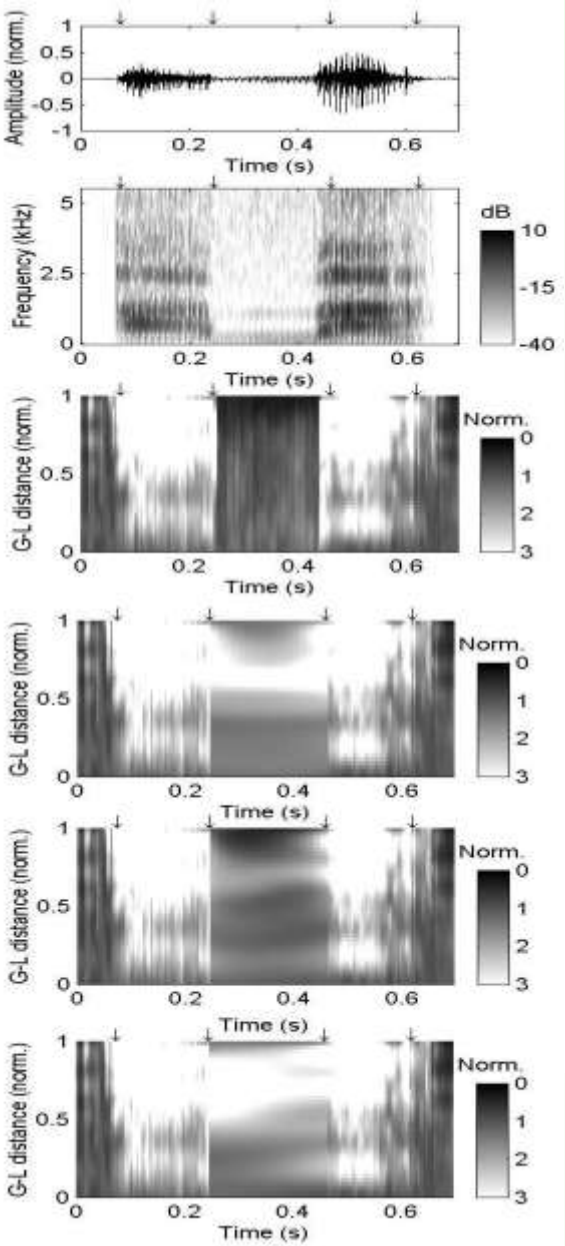
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

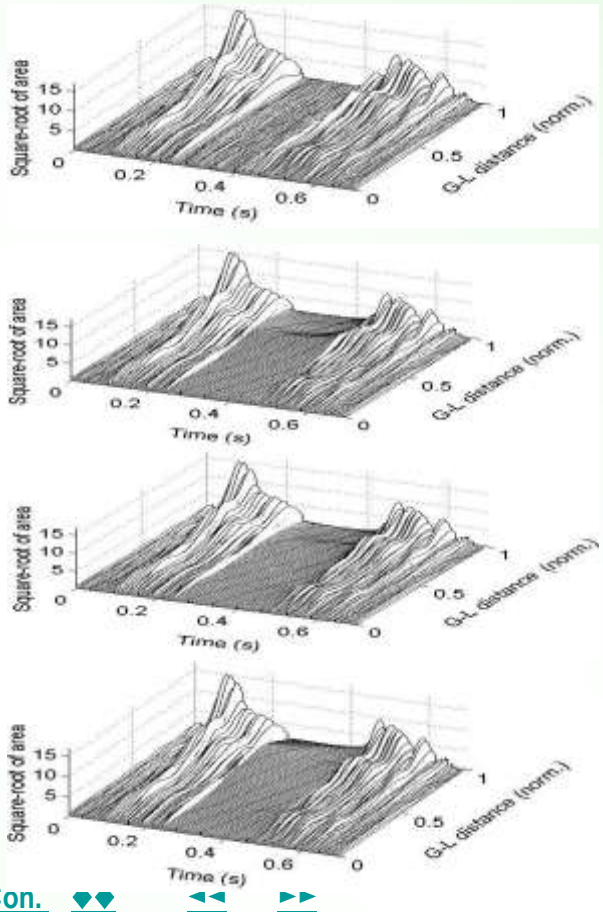
(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 5$, $L_{col} = 2$, $R_{col} = 2$



P.C. Pandey, EE Dept, IIT Bombay

Result 9: 2D interpolation of LSFs for /ata/ (speaker SM1)

(a) waveform

(b) Spectrogram

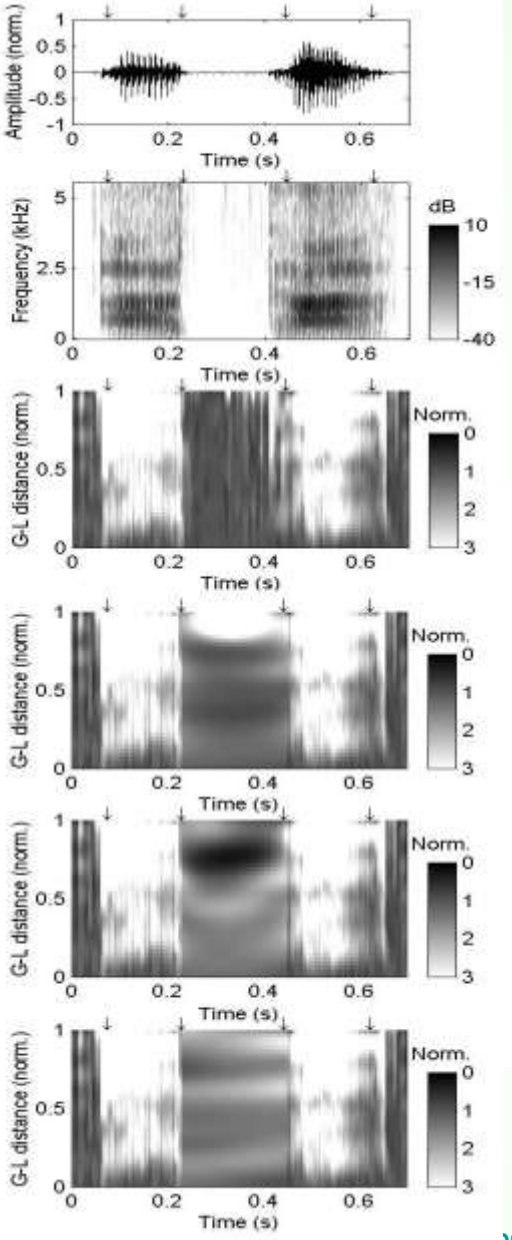
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 4$, $L_{col} = 7$, $R_{col} = 7$

P.C. Pandey, EE Dept, IIT Bombay

Result 10: 2D interpolation of LSFs for /ada/ (speaker SM1)

(a) waveform

(b) Spectrogram

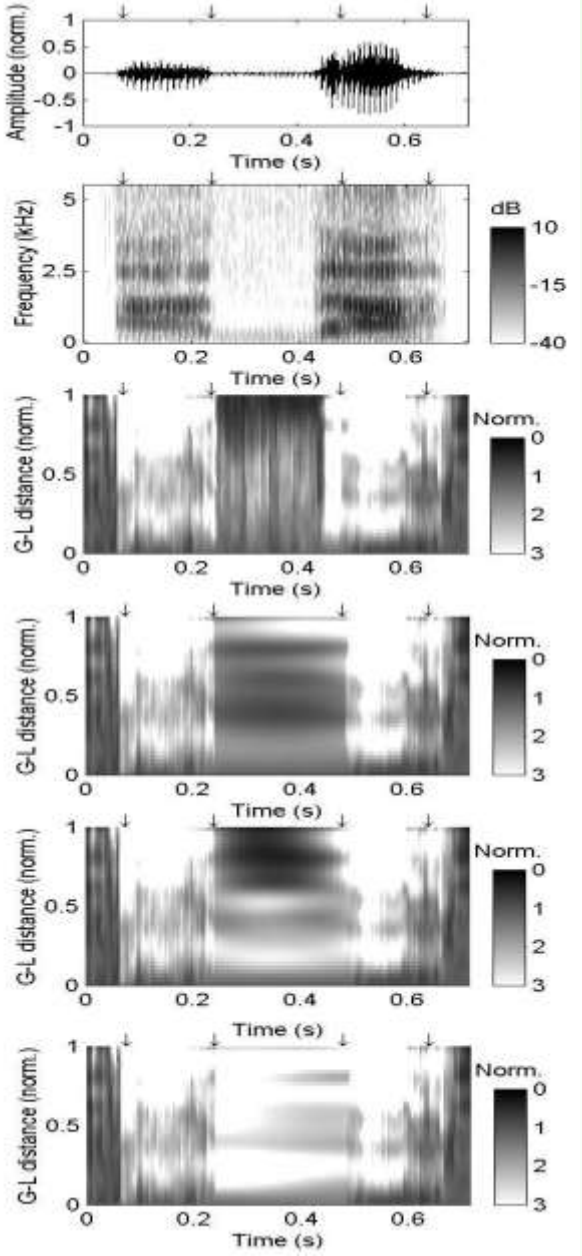
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

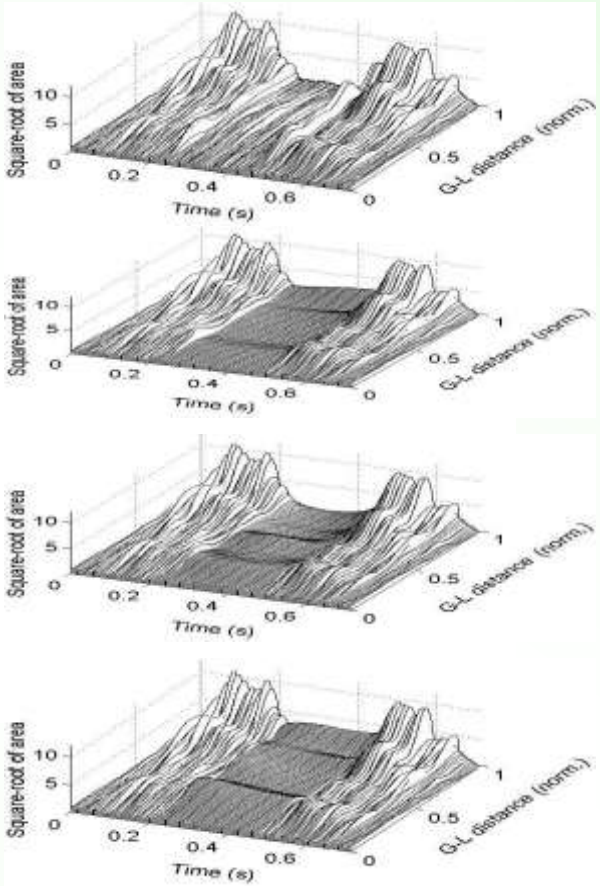
(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 4$, $L_{col} = 7$, $R_{col} = 7$



P.C. Pandey, EE Dept, IIT Bombay

Result 11: 2D interpolation of LSFs for /aka/ (speaker SM1)

(a) waveform

(b) Spectrogram

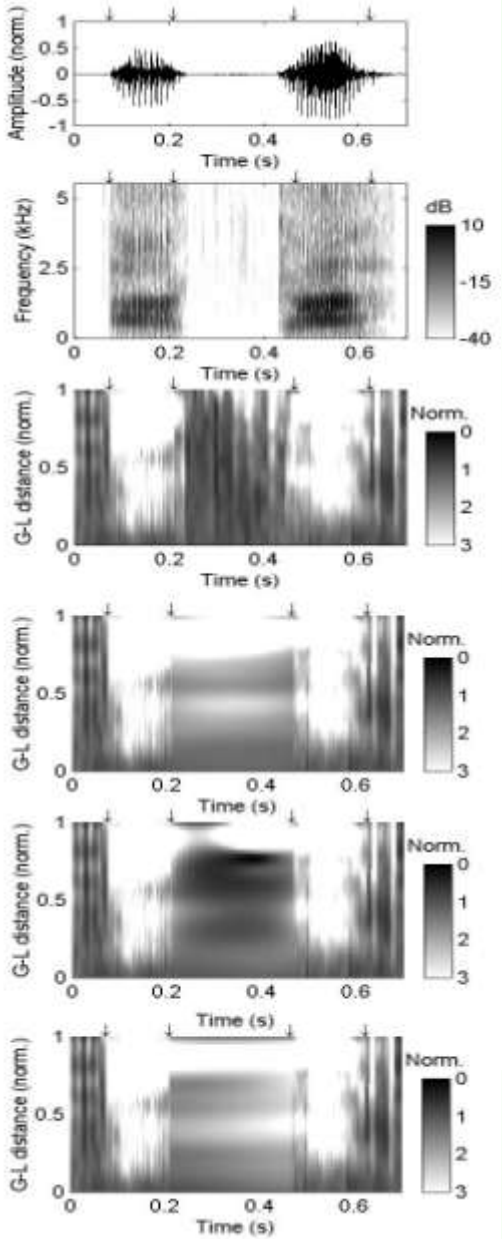
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation



Surface generation parameters:

$j = 5$, $L_{col} = 4$, $R_{col} = 4$

P.C. Pandey, EE Dept, IIT Bombay

Result 12: 2D interpolation of LSFs for /aga/ (speaker SM1)

(a) waveform

(b) Spectrogram

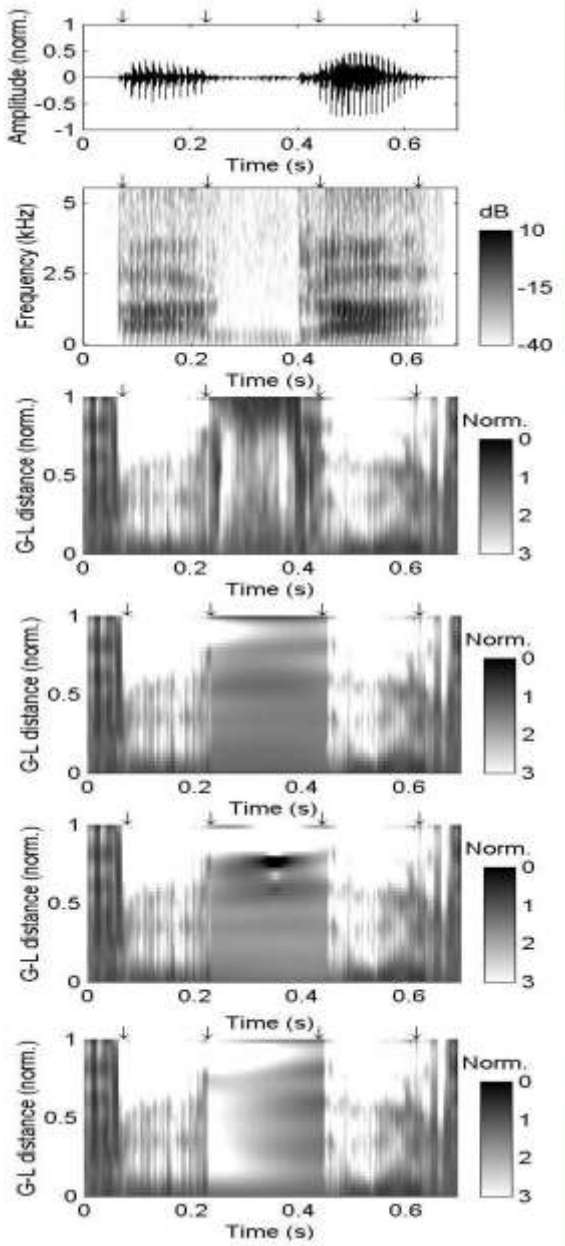
(Δf = 300 Hz)

(c) Original areagram & waterfall diagram

(d) areagram & waterfall diagram based on second degree polynomial surface interpolation

(e) areagram & waterfall diagram based on third degree polynomial surface interpolation

(f) areagram & waterfall diagram based on Delaunay surface interpolation

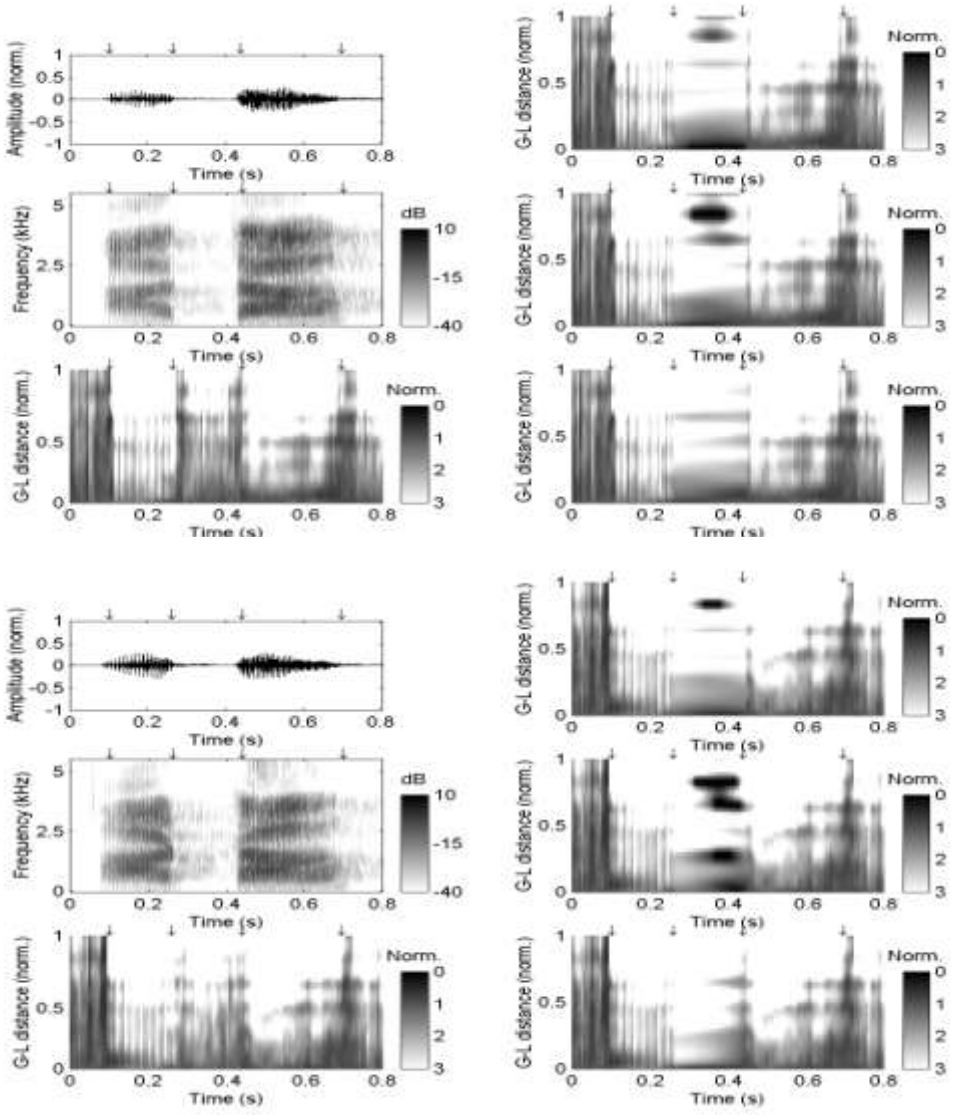


Surface generation parameters:

$j = 6$, $L_{col} = 7$, $R_{col} = 7$

P.C. Pandey, EE Dept, IIT Bombay

Result 13: 2D interpolation of area values for Marathi /aṭa/ and /ata/ (speaker SM2)



Surface generation parameters for (dental stop): /aṭa/
$j = 3$, $L_{col} = 7$, $R_{col} = 7$

Surface generation parameters for /ata/ (retroflex-alveolar):
$j = 3$, $Lcol = 7$, $Rcol = 7$

P.C. Pandey, EE Dept, IIT Bombay

Result Summary

- For /aCa/, estimation of place of constriction for bilabial, alveolar, & velar stops is most accurate with 2nd degree polynomial surface modeling of area values & LSFs

(in conformity with observations during initial validation of the technique with artificially introduced silence gaps in semivowels)

→ articulatory movement during production of /aCa/ modeled more appropriately by 2nd degree polynomials.

- For /iCa/, /aCi/, & /iCi/, estimation of place of constriction for velar stops is not consistent across speakers

→ the proposed technique less effective for articulatory movement involving transition of place of articulation from front (as for vowel /i/) to back (as for velar /k/ & /g/).

P.C. Pandey, EE Dept, IIT Bombay

- For /aCa/ involving bilabial, alveolar, & velar stops, average number of frames required for proper surface modeling based on area values (6.1, 6.8, and 5.9 frames resp.) are less compared to modeling of LSFs (7.6, 7.5, and 7.3 frames resp.)

(in conformity with observations during initial validation of the technique with artificially introduced silence gaps in semivowels).

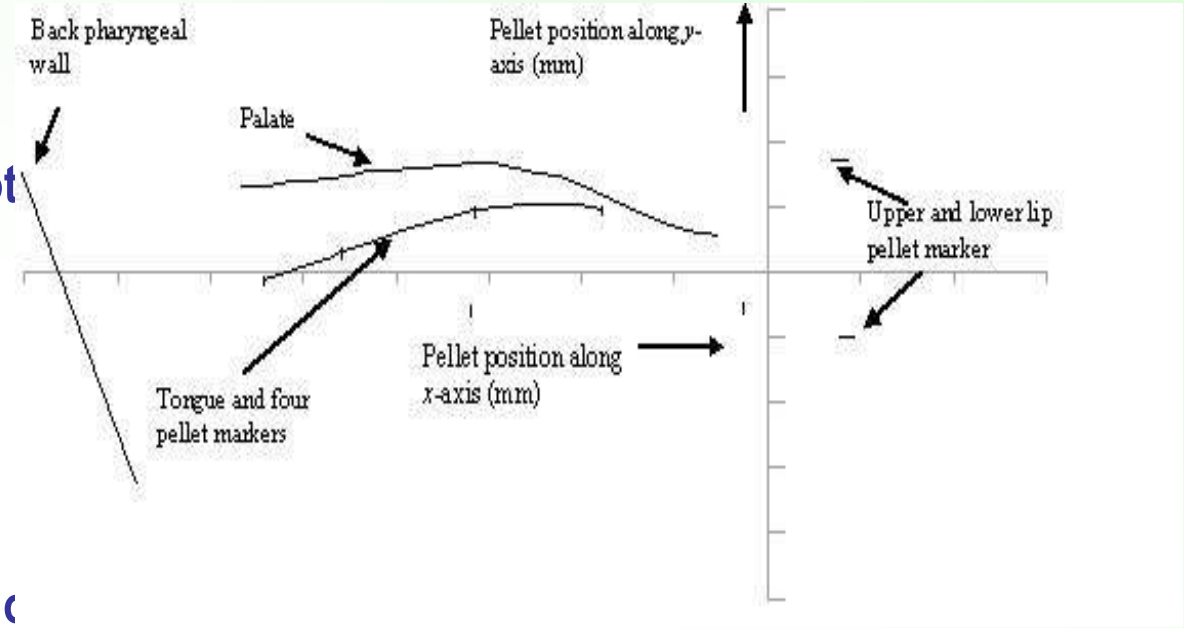
P.C. Pandey, EE Dept, IIT Bombay

Investigations for Direct Validation of the Technique

- For direct validation of the technique, acoustic signals that have been simultaneously acquired with articulatory data analyzed
  - Database from the University of Wisconsin.
  - Articulatory data acquired using X-ray microbeam (XRMB) system.
  - Articulatory plot shows position of pellets in the midsaggital plane.
  - Position of pellets gives a point-parameterized representation of lingual, labial, and mandibular movements.
  - Information about the lower part of the vocal tract not available.

P.C. Pandey, EE Dept, IIT Bombay

- Sample articulatory plot



- 2D interpolation based

second degree surfaces representing area values

applied to 120 VCV syllables of the type

/^Ca/ (from XRMB database)

involving stop consonants /b/, /d/, & /g/ for the estimation of place of closure.

- Estimated place of constriction compared with

actual constriction locations obtained from articulatory database.

P.C. Pandey, EE Dept, IIT Bombay

Scatter Plot

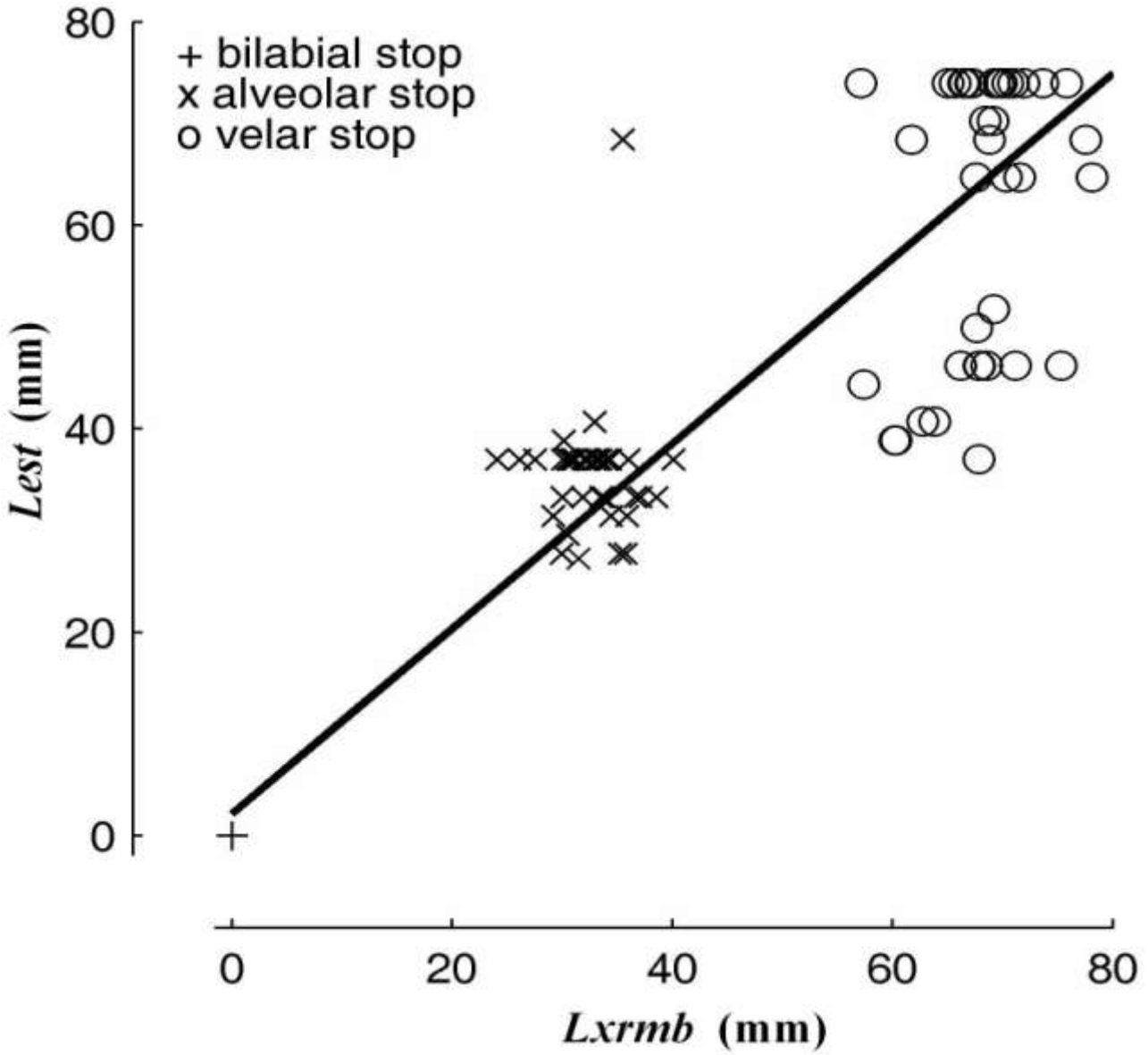
Estimated place L_{est} (IL-G distance, in mm) vs actual constriction locations (from the XRMB database) L_{xrmb} for the 120 /ACa/ utterances involving stop consonants /p/, /t/, and /k/.

Linear regression

$L_{est} = 2.179 + 0.909 L_{xrmb}$

Corr. coeff. = 0.928

($p < 0.0001$)



P.C. Pandey, EE Dept, IIT Bombay

1. Introduction

2. Visual Speech-training Aids

3. LPC Based Vocal Tract Shape Estimation

4. Estimation of Vocal Tract Shape during Stop Closures

5. Results & Discussion

6. Summary & Conclusions

P.C. Pandey, EE Dept, IIT Bombay

Summary of Investigations

- **Implementation of vocal tract shape estimation based on LPC analysis & Wakita's model and investigations for**
  - optimum value of analysis parameters (W, M, and F_s) for vowels
  - effect of pitch and amplitude variations on shape estimation for vowels
  - shape estimation for VCV syllables involving semivowels & stop consonants

- **Technique for estimation of place of closure in a VCV syllable:**

  surface approximation of values related to vocal tract shape during VC & CV transition segments (based on least-squares bivariate polynomials and Delaunay triangulation) for 2D interpolation during closure duration

P.C. Pandey, EE Dept, IIT Bombay

- **Estimation of stop closure boundary locations in VCV syllables,**

  using avg. short-time magnitude & empirically selected thresholds

- **Validation of the proposed techniques**

  for artificially introduced silence segments in vowels /a/, /i/, & /u/ and VCV syllables /aja/ & /awa/

  and estimation of the minimum transition segments required

- **Estimation of place of closure by 2D interpolation**

  - VCV syllables with English stops (3 places of articulation, 5 speakers)
  - VCV syllables with Marathi stops (5 places of articulation, 1 speaker)

P.C. Pandey, EE Dept, IIT Bombay

Conclusions

- Estimation of place of closure feasible using polynomial and Delaunay surface modeling of area values as well as LSFs for /aCa/ syllables across speakers.

Difficulties in place estimation for syllables involving front vowel /i/ and velar stops /k/ & /g/.

- As compared to LSF based estimation, area value based estimation required a smaller number of frames and resulted in more consistent estimates.

- For /aCa/ syllables, second degree polynomial surfaces gave most consistent place estimation.

P.C. Pandey, EE Dept, IIT Bombay

Future Work

- Application of the technique to recordings from a larger number of speakers with different age groups and language backgrounds.
- Application of the technique on recordings with vocal tract shapes simultaneously captured by imaging techniques.
- Investigations with shape estimation using other analysis techniques (e.g. formant tracking, articulatory analysis by synthesis).
- Development of speech training aid with dynamic display of vocal tract shape.
- Evaluation for speech training of hearing impaired children.

P.C. Pandey, EE Dept, IIT Bombay

Thank you

Prem C. Pandey

Prof. Pandey received the B.Tech. degree in electronics engineering from the Banaras Hindu University in 1979, the M.Tech. degree in electrical engineering from the Indian Institute of Technology Kanpur (India) in 1981, and the Ph.D. degree in biomedical engineering from the University of Toronto (Canada) in 1987.

In 1987, he joined the University of Wyoming (USA) as an Assistant Professor in electrical engineering and later joined the Indian Institute of Technology Bombay in 1989, where he is a Professor in electrical engineering, with a concurrent association with the biomedical engineering program.

His research interests include speech and signal processing; biomedical signal processing; embedded system design and electronic instrumentation.