Statistical identification of critical, dependent and redundant articulators

Philip Jackson & Veena Singampalli

Centre for Vision, Speech and Signal Processing (CVSSP)
University of Surrey, UK.

July 2, 2008

Acoustics'08 Paris

Outline

- What do we know about articulation?
- Determining articulatory responsibilities
  - inter-articulator correlations
  - critical articulator identification
- Evaluation of results
  - comparison with phonetics (IPA)
  - benchmark against exhaustive search
- Investigation of modes by PCA
- Conclusion

About speech articulation

What we know:

1. abstraction of phones as phonemes
2. physiology of muscle groups
3. physical dynamics of articulators
4. critical articulators' ($1^y/2^y$) task dynamics
5. coarticulation causes
   - asymmetric smoothing and target undershoot
   - spreading of distinctive articulatory features
6. asynchrony of articulatory events (landmarks)
7. speech is efficient, adaptive, versatile, robust

About speech articulation

What we don't know:

1. how to specify phoneme-to-phone conversion
2. how motor commands are organized
3. how dynamical properties vary during utterance
4. precisely which articulator is responsible for what
5. how to incorporate within a coarticulation model
   - learning of actual articulatory targets
   - effects of feature spreading
6. how to detect and parameterize events
7. how to build a comprehensive model of speech articulation within the communicative loop

Critical articulator identification

Constraints on gestures determined by rôle:

- critical, dependent or redundant

Our algorithm [Singampalli and Jackson, 2007]:

- is entirely statistical and data driven
- captures constraints from
  - correlated movement of every articulator in space
  - correlations amongst articulators
- identifies articulatory responsibilities
  - i.e., changes in mean and variance
- provides information about the structure underlying gestural dynamics of speech articulation

Articulatory data

- Electro-magnetic articulography (EMA) data for two speakers (1 M, 1 F)) from MOCHA database [Wrench, 2001], low-pass filtered and resampled at 10 ms frame rate

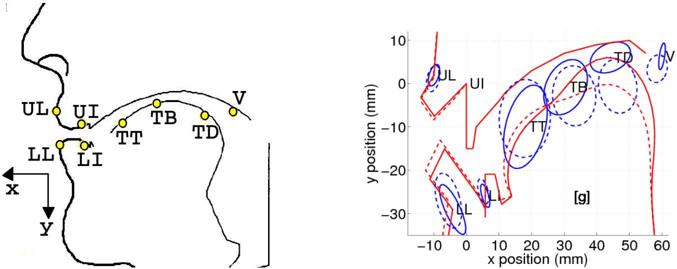


Figure: Midsagittal views of human speech production system [Saltzman and Munhall, 1989].

Articulatory correlations

■ $|r| < 0.1$ ■ $0.1 \leq |r| < 0.45$ ■ $0.5 \leq |r| < 1$ ■ $r = 1$

	ULy	ULx	LLy	LLx	LIy	LIx	TTy	TBy	TDy	TTx	TBx	TDx	Vy	Vx
ULy	1.00	0.10	-0.28	-0.23	0.20	-0.03	0.32	0.18	-0.02	0.05	-0.07	-0.13	-0.02	-0.10
ULx	0.53	1.00	-0.34	0.56	-0.14	0.00	-0.10	0.02	0.03	-0.23	-0.28	-0.21	-0.05	-0.03
LLy	-0.31	-0.15	1.00	-0.31	0.53	-0.22	0.27	-0.04	-0.14	0.03	0.10	0.06	0.02	0.00
LLx	0.27	0.34	-0.70	1.00	-0.43	0.51	-0.29	-0.04	0.10	-0.09	-0.04	0.10	-0.06	-0.04
LIy	0.05	-0.17	0.65	-0.55	1.00	-0.34	0.56	0.32	0.03	-0.13	-0.14	-0.18	0.02	0.00
LIx	-0.04	0.03	-0.49	0.61	-0.71	1.00	-0.30	-0.05	0.07	0.13	0.32	0.37	-0.15	-0.34
TTy	0.29	0.02	0.32	-0.31	0.60	-0.43	1.00	0.41	0.00	0.16	0.00	-0.11	0.14	0.14
TBy	0.19	-0.04	0.06	-0.08	0.42	-0.36	0.53	1.00	0.73	0.05	-0.10	-0.05	0.07	0.16
TDy	0.08	-0.08	-0.10	0.14	0.09	-0.02	0.08	0.75	1.00	0.04	-0.08	0.10	0.24	0.30
TTx	-0.10	-0.18	0.01	-0.06	-0.01	0.03	0.06	0.06	-0.01	1.00	0.84	0.74	0.19	0.14
TBx	-0.08	-0.16	0.14	-0.19	0.12	-0.05	0.18	-0.03	-0.24	0.90	1.00	0.91	-0.04	-0.08
TDx	-0.15	-0.22	0.10	-0.17	0.04	0.03	0.11	-0.02	-0.21	0.82	0.92	1.00	-0.03	-0.08
Vy	0.02	-0.08	-0.09	0.02	-0.05	0.05	-0.06	0.08	0.23	0.19	0.06	0.01	1.00	0.70
Vx	0.13	0.05	-0.18	0.11	-0.12	0.12	-0.05	-0.04	0.06	0.24	0.14	0.05	0.81	1.00

Articulatory correlations

- Statistically significant ($\alpha = 0.05$) correlation matrix

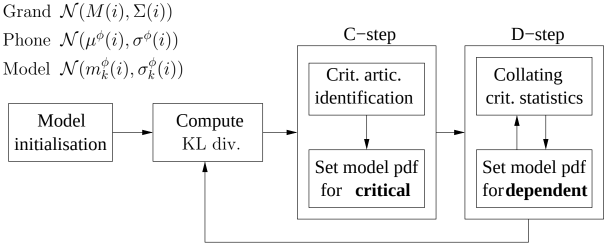
	ULy	ULx	LLy	LLx	LIy	LIx	TTy	TBy	TDy	TTx	TBx	TDx	Vy	Vx
ULy	1.00	0.10	-0.28	-0.23	0.20	-0.03	0.32	0.18	-0.02	0.05	-0.07	-0.13	-0.02	-0.10
ULx	0.53	1.00	-0.34	0.56	-0.14	0.00	-0.10	0.02	0.03	-0.23	-0.28	-0.21	-0.05	-0.03
LLy	-0.31	-0.15	1.00	-0.31	0.53	-0.22	0.27	-0.04	-0.14	0.03	0.10	0.06	0.02	0.00
LLx	0.27	0.34	-0.70	1.00	-0.43	0.51	-0.29	-0.04	0.10	-0.09	-0.04	0.10	-0.06	-0.04
LIy	0.05	-0.17	0.65	-0.55	1.00	-0.34	0.56	0.32	0.03	-0.13	-0.14	-0.18	0.02	0.00
LIx	-0.04	0.03	-0.49	0.61	-0.71	1.00	-0.30	-0.05	0.07	0.13	0.32	0.37	-0.15	-0.34
TTy	0.29	0.02	0.32	-0.31	0.60	-0.43	1.00	0.41	0.00	0.16	0.00	-0.11	0.14	0.14
TBy	0.19	-0.04	0.06	-0.08	0.42	-0.36	0.53	1.00	0.73	0.05	-0.10	-0.05	0.07	0.16
TDy	0.08	-0.08	-0.10	0.14	0.09	-0.02	0.08	0.75	1.00	0.04	-0.08	0.10	0.24	0.30
TTx	-0.10	-0.18	0.00	-0.06	0.00	0.03	0.06	0.06	0.00	1.00	0.84	0.74	0.19	0.14
TBx	-0.08	-0.16	0.14	-0.19	0.12	-0.05	0.18	-0.03	-0.24	0.90	1.00	0.91	-0.04	-0.08
TDx	-0.15	-0.22	0.10	-0.17	0.04	0.03	0.11	-0.02	-0.21	0.82	0.92	1.00	-0.03	-0.08
Vy	0.02	-0.08	-0.09	0.00	-0.05	0.05	-0.06	0.08	0.23	0.19	0.06	0.00	1.00	0.70
Vx	0.13	0.05	-0.18	0.11	-0.12	0.12	-0.05	-0.04	0.06	0.24	0.14	0.05	0.81	1.00

Articulatory correlations

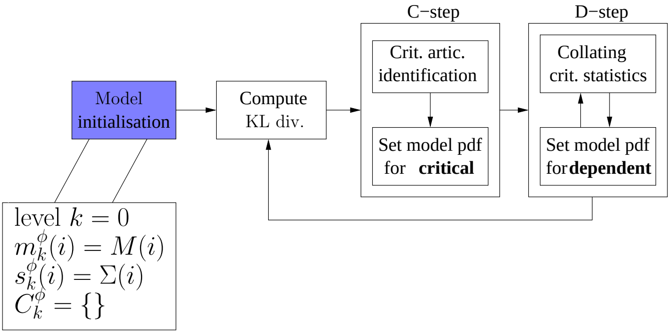
- Significant and strong ($|r_{ij}| > 0.1$) correlation matrix, R^*

	ULy	ULx	LLy	LLx	Lly	Llx	TTy	TBy	TDy	TTx	TBx	TDx	Vy	Vx
ULy	1.00	0.10	-0.28	-0.23	0.20	0.00	0.32	0.18	0.00	0.00	0.00	-0.13	0.00	-0.10
ULx	0.53	1.00	-0.34	0.56	-0.14	0.00	0.00	0.00	0.00	-0.23	-0.28	-0.21	0.00	0.00
LLy	-0.31	-0.15	1.00	-0.31	0.53	-0.22	0.27	0.00	-0.14	0.00	0.10	0.00	0.00	0.00
LLx	0.27	0.34	-0.70	1.00	-0.43	0.51	-0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Lly	0.00	-0.17	0.65	-0.55	1.00	-0.34	0.56	0.32	0.00	-0.13	-0.14	-0.18	0.00	0.00
Llx	0.00	0.00	-0.49	0.61	-0.71	1.00	-0.30	-0.05	0.00	0.13	0.32	0.37	-0.15	-0.34
TTy	0.29	0.00	0.32	-0.31	0.60	-0.43	1.00	0.41	0.00	0.16	0.00	-0.11	0.14	0.14
TBy	0.19	0.00	0.00	0.00	0.42	-0.36	0.53	1.00	0.73	0.00	0.00	0.00	0.00	0.16
TDy	0.00	0.00	-0.10	0.14	0.00	0.00	0.00	0.75	1.00	0.00	0.00	0.00	0.24	0.30
TTx	0.00	-0.18	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.84	0.74	0.19	0.14
TBx	0.00	-0.16	0.14	-0.19	0.12	0.00	0.18	0.00	-0.24	0.90	1.00	0.91	0.00	0.00
TDx	-0.15	-0.22	0.10	-0.17	0.00	0.00	0.11	0.00	-0.21	0.82	0.92	1.00	0.00	0.00
Vy	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.23	0.19	0.06	0.00	1.00	0.70
Vx	0.13	0.00	-0.18	0.11	-0.12	0.12	0.00	0.00	0.00	0.24	0.14	0.00	0.81	1.00

Algorithm (1)


Grand $\mathcal{N}(M(i), \Sigma(i))$
Phone $\mathcal{N}(\mu^{\phi}(i), \sigma^{\phi}(i))$
Model $\mathcal{N}(m_k^{\phi}(i), \sigma_k^{\phi}(i))$
Model initialisation
Compute KL div.
C−step
Crit. artic. identification
Set model pdf for **critical**
D−step
Collating crit. statistics
Set model pdf for **dependent**

Algorithm (2)


Model initialisation
Compute KL div.
C−step
Crit. artic. identification
Set model pdf for **critical**
D−step
Collating crit. statistics
Set model pdf for **dependent**
level $k = 0$
$m_k^{\phi}(i) = M(i)$
$s_k^{\phi}(i) = \Sigma(i)$
$C_k^{\phi} = \{\}$

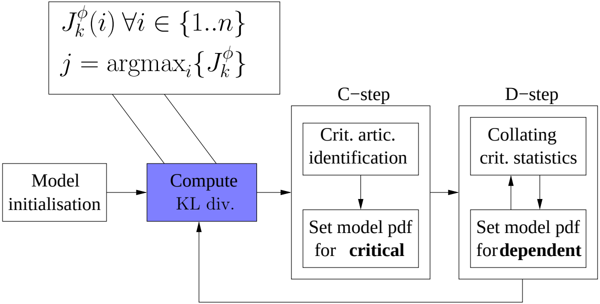
Identification divergence

Assuming the articulatory distributions are normal, we have

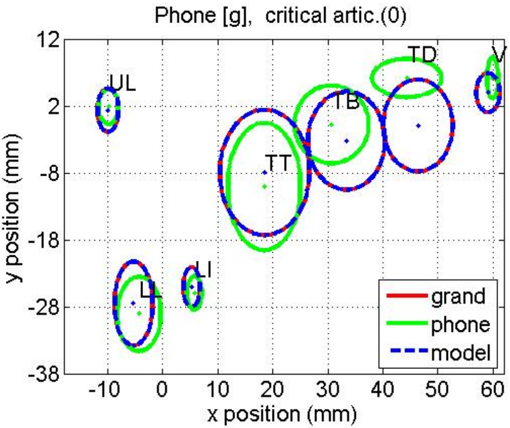
$$J_k^{\phi} = \frac{1}{2}\Big[\mathrm{tr}\,(s_k - \sigma_\phi)\left(\sigma_\phi^{-1} - s_k^{-1}\right) + \mathrm{tr}\left(s_k^{-1} + \sigma_\phi^{-1}\right)(m_k - \mu_\phi)(m_k - \mu_\phi)^{\mathsf{T}}\Big] \quad (1)$$

where m_k and s_k are the mean and variance of model at level k, μ_ϕ and σ_ϕ are the mean and variance of the phone pdf

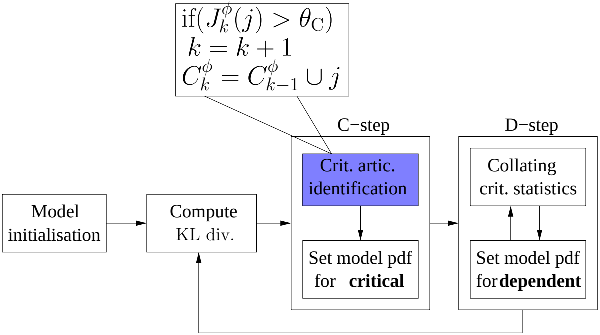
Algorithm (3)


$J_k^\phi(i)\ \forall i \in \{1..n\}$
$j = \mathrm{argmax}_i\{J_k^\phi\}$
Model initialisation
Compute KL div.
C−step
Crit. artic. identification
Set model pdf for **critical**
D−step
Collating crit. statistics
Set model pdf for **dependent**

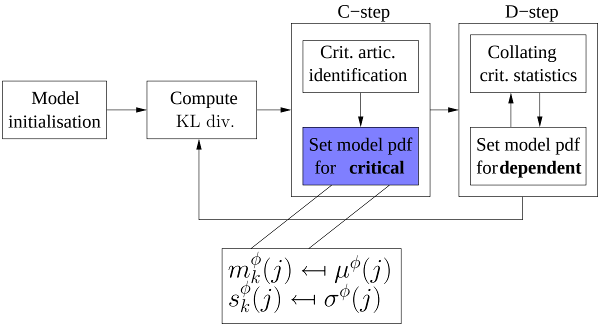
Velar consonant: initialisation



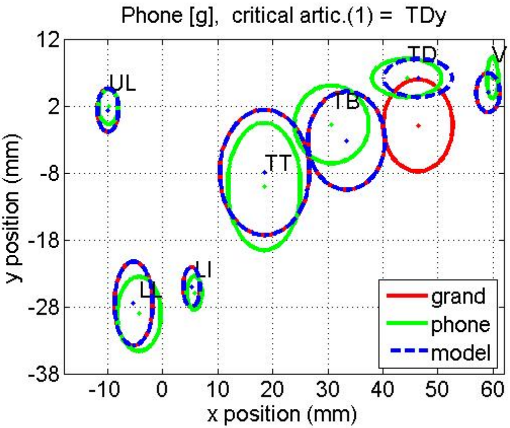
Algorithm (4)


if$(J_k^\phi(j) > \theta_C)$
$k = k + 1$
$C_k^\phi = C_{k-1}^\phi \cup j$
C−step
D−step
Crit. artic. identification
Collating crit. statistics
Model initialisation
Compute KL div.
Set model pdf for **critical**
Set model pdf for **dependent**

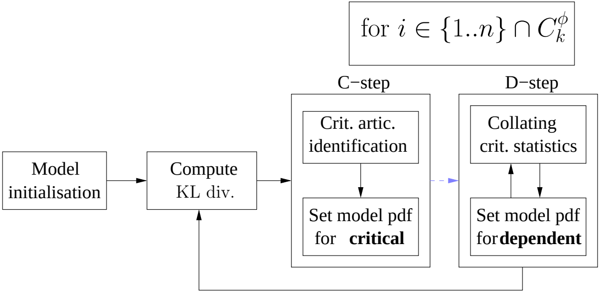
Algorithm (5)


C−step
D−step
Model initialisation
Compute KL div.
Crit. artic. identification
Set model pdf for **critical**
Collating crit. statistics
Set model pdf for **dependent**
$m_k^{\phi}(j) \leftarrow \mu^{\phi}(j)$
$s_k^{\phi}(j) \leftarrow \sigma^{\phi}(j)$

Velar consonant: first identification



Algorithm (6)

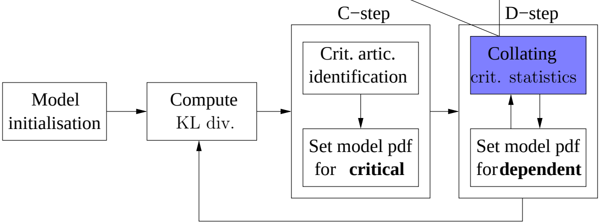

for $i \in \{1..n\} \cap C_k^{\phi}$
C−step
D−step
Model initialisation
Compute KL div.
Crit. artic. identification
Set model pdf for **critical**
Collating crit. statistics
Set model pdf for **dependent**

Algorithm (7)

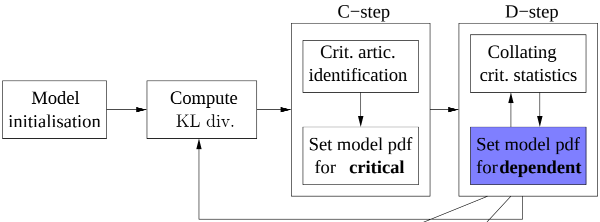
$$\text{if}(J_k^{\phi}(i) > \theta_{\text{dep}})$$

$$M_c = M_{\{C_k^{\phi}\}}, \Sigma_c = \Sigma_{\{C_k^{\phi}\}\{C_k^{\phi}\}}, \Sigma_d = \Sigma_{i\{C_k^{\phi}\}}$$

$$\mu_c = \mu^{\phi}_{\{C_k^{\phi}\}}, \sigma_c = \Sigma^{\phi}_{\{C_k^{\phi}\}\{C_k^{\phi}\}}$$



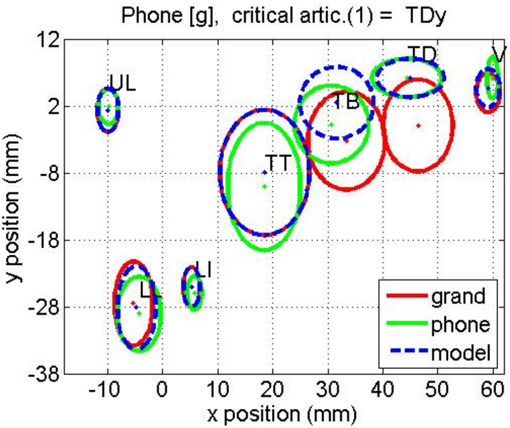
Algorithm (8)



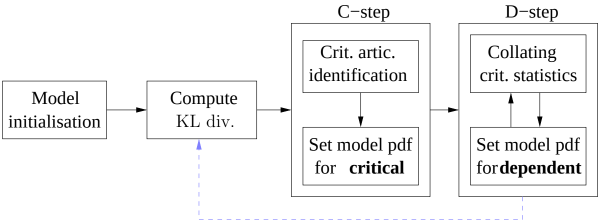
$$m_k^{\phi}(i) \leftarrow M(i) + \Sigma_d \Sigma_c^{-1}(\mu_c - M_c)$$

$$s_k^{\phi}(i) \leftarrow \Sigma_i + \Sigma_d \Sigma_c^{-1}(\sigma_c - \Sigma_c)\Sigma_c^{-1}\Sigma_d'$$

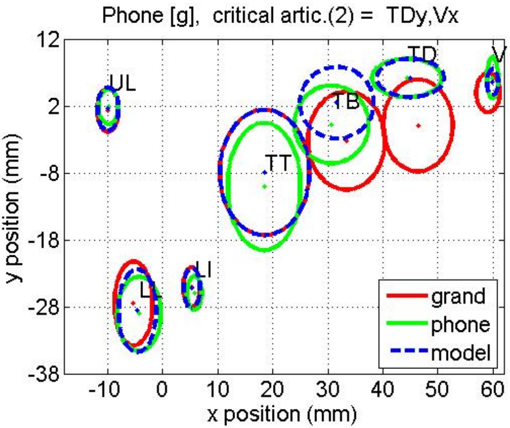
Velar consonant: dependencies



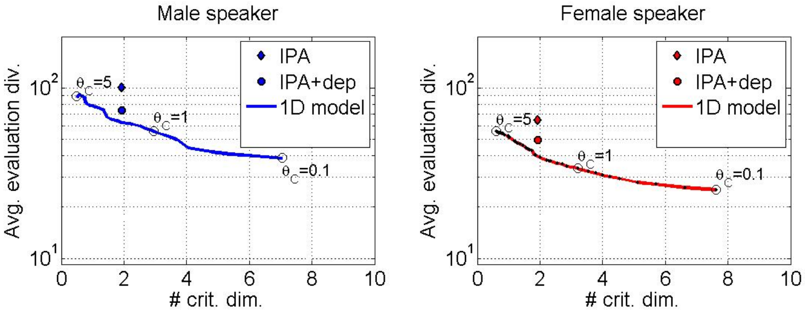
Algorithm (9)


C−step
D−step
Model initialisation
Compute KL div.
Crit. artic. identification
Set model pdf for **critical**
Collating crit. statistics
Set model pdf for **dependent**

Velar consonant: second identification



Fit in comparison with IPA



- Algorithm achieved more accurate fit to data than IPA
- θ_C ensures consistent precision across phone set
- Critical articulators similar to IPA (cf., velum, 3^y artic)

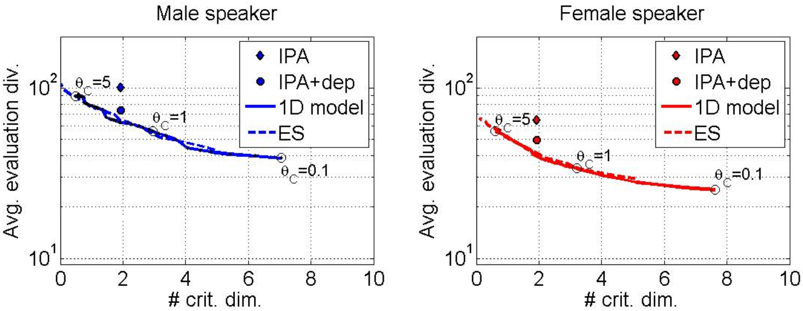
Exhaustive search trials

Proposed method ES method

$\circ\ J^{\phi}_{k-1}$ $\bullet\ J^{\phi}_{k}$ $\square\ \hat{J}^{\phi}_{k}$

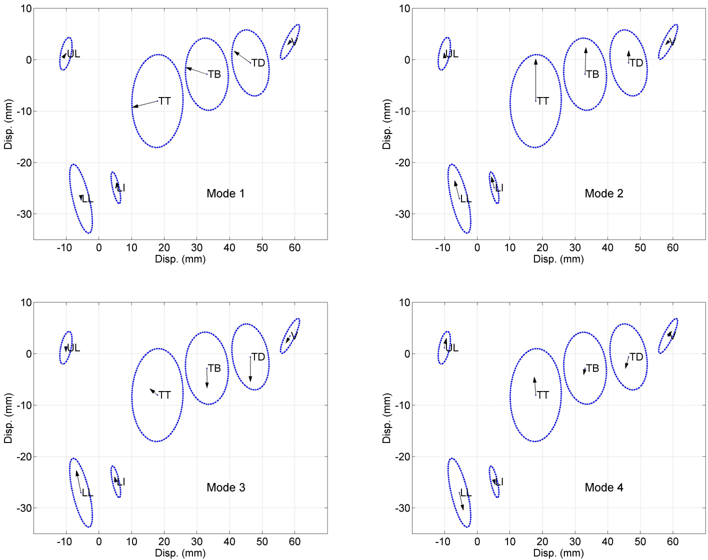
- Can we improve fit?
- Will it lead to different sets of critical articulators?
- Does the improvement justify extra computation?

Exhaustive search results

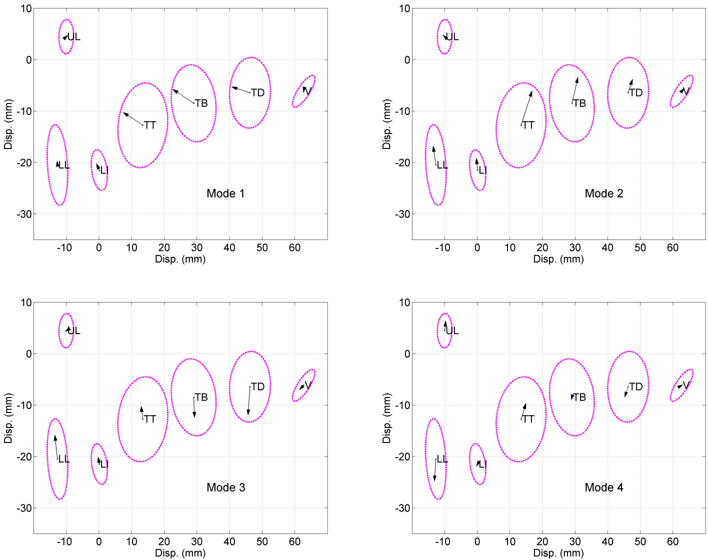


- Although ES reduced the identification divergence, did not lower evaluation divergence
- Most changes were correlated articulator substitutions
- Computation time $10^6\times$ slower at level $k = 6$

PCA modes – male



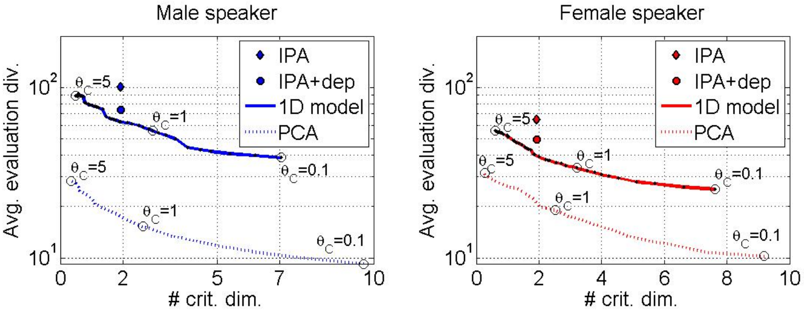
PCA modes – female



Statistical identification of critical, dependent and redundant articulators

Philip Jackson & Veena Singampalli

Fit with PCA



- Evaluation divergence reduced from 60/40 to ∼20:
  - because PCA optimized to fit data in LS sense
  - because different muscle grous act separately (e.g., intrinsic and extrinsic tongue muscles as superposed modes)

Summary

- To model articulation and coarticulation, we need to determine responsibilities of articulator during speech
- Algorithm identified critical, dependent and redundant articulators from EMA data
- Greedy search was faster and as good as exhaustive search
- PCA used to diagonalised articulatory covariance
- Improvements in fit attributed to
  - alignment of objective function with data pdfs
  - independent control of articulatory modes
- Phonetic interpretation of algorithm's output at ISSP '08
- More: subjects, fleshpoints, languages, utterances, etc.

http://www.ee.surrey.ac.uk/Personal/P.Jackson/Dansa/

References

R. Bakis. Coarticulation modeling with continuous-state HMMs. *Proc. IEEE Workshop Automatic Speech Recognition, Harriman, New York*, 1991.

S.C. Blackburn and S. Young. A self-learning predictive model of articulator movements during speech production. *JASA*, 103(3):1659–70, March 2000.

R. Daniloff and R. Hammarberg. On defining coarticulation. *J. Phon.*, 1:239–248, 1973.

Ellen Eide. Distinctive features for use in an automatic speech recognition system. *Proc. Eurospeech '01,* Aalborg, Denmark, 2001.

Kevin Erler and George H. Freeman. An HMM-based speech recognizer using overlapping articulatory features. *JASA*, pages 2500–2513, 1996.

W. L. Henke. *Dynamic articulatory model of speech production using computer simulation*. PhD thesis, MIT, Cambridge, MA, 1965.

K. Kirchhoff. *Robust speech recognition using articulatory information*. PhD thesis, University of Bielefeld,, 1999.

D. Recasens, D.M. Pallarés, and J. Fontdevilla. A model of lingual coarticulation based on articulatory constraints. *JASA*, 102:544–561, 1997.

Matthew Richardson, Jeff Blimes, and Chris Diorio. Hidden-articulatory Markov models for speech recognition. *Proc. Int. Conf. on Spoken Lang. Proc.,* Beijing, 2000.

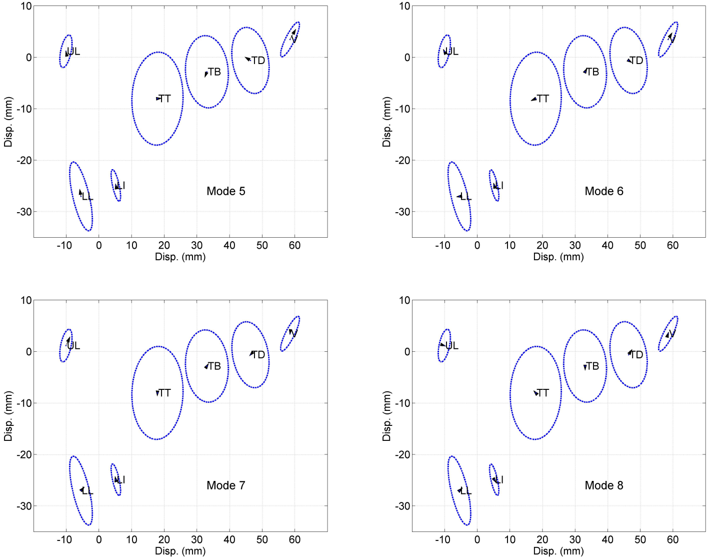
E.L. Saltzman and K. Munhall. A dynamic approach to gestural patterning in speech production. *Ecology Psychology*, 1:333–82, 1989.

V. D. Singampalli and P. J. B. Jackson. Statistical identification of critical, dependent and redundant articulators. In *Proc. Interspeech '07,* Antwerp, volume 1, pages 70–73, 2007.

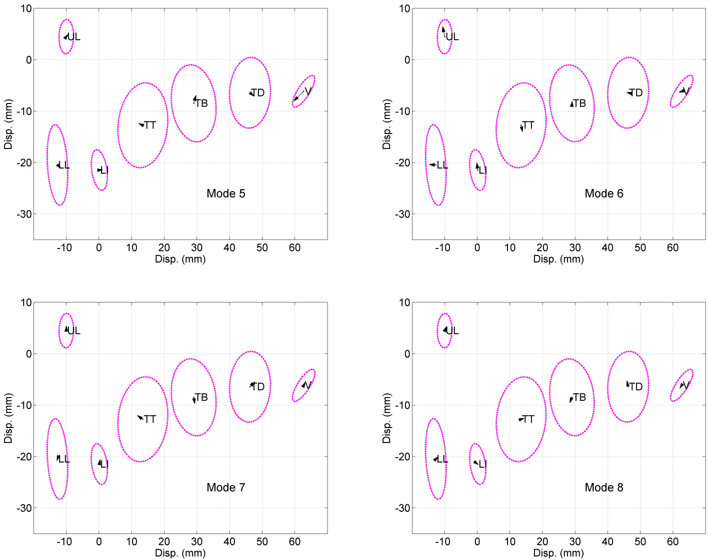
A.A. Wrench. A new resource for production modelling in speech technology. *Proc. Inst. of Acoust.,* Stratford-upon-Avon, UK, 2001.

http://www.ee.surrey.ac.uk/Personal/P.Jackson/Dansa/

PCA modes -male



PCA modes - female



Alternative principal component analyses