A STATISTICAL TECHNIQUE FOR IDENTIFYING ARTICULATORY ROLES IN SPEECH PRODUCTION

Veena D. Singampalli and Philip J.B. Jackson [v.singampalli,p.jackson]@surrey.ac.uk

Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, U.K.

Abstract

A statistical approach for identifying critical articulators from annotated articulatory data and estimating the distributions of dependent articulators using articulatory correlations is presented.
Two versions of the algorithm are considered, 1D and 2D, which either treat x and y coordinates of the articulatory data independently or incorporate covariation within the motion of an articulator in the mid-saggital plane. The 2D case uses canonical correlation analysis.
The performance of the critical articulator identification approach is evaluated by comparing the results obtained with our algorithm to the distinctive phonetic features.

1 Background

- Feature based models of speech production where coarticulation is viewed as a spread of features [8, 15, 6]
  - Binary features [5, 7]
  - Formant based features [13, 16]
- Target or goal oriented models of speech production where coarticulation is viewed as coproduction of gestures [19, 4, 14, 12]
- Degree of articulatory constraint [17] and coarticulation resistance [3]
- Statistical models for speech dynamics [18, 20] and temporal aspects of coarticulation [2, 10]

2 Articulatory roles

2.1 Articulatory data

EMA recordings of one male (msak) and female (fsew) speaker uttering 460 TIMIT sentences each from MOCHA database [21]
Raw data is smoothed using a Hann window and sampled at 10ms frame rate.

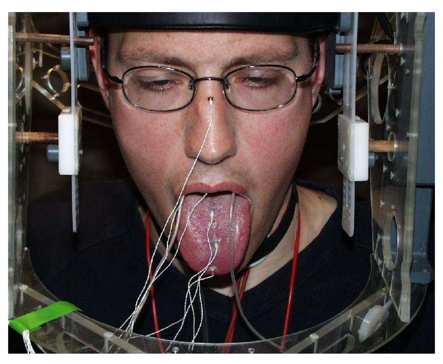

FIGURE 1: *Coils on upper lip* UL, *lower lip* LL, *lower incisor* LI, *tongue tip* TT, *tongue blade* TB, *tongue dorsum* TD, *and velum* V *with upper incisor* UI *and the bridge of the nose as reference points.*

REFERENCES

[1] T.W. Anderson. *An introduction to multivariate statistical analysis.* Wiley, New York, 2 edition, 1984.
[2] S.C. Blackburn and S. Young. A self-learning predictive model of articulator movements during speech production. *JASA*, 103(3):1659–70, March 2000.
[3] R. A. W. Bladon and A. Al-Bamerni. Coarticulation resistance in English /l/. *J. Phon.*, 4:135–50, 1976.
[4] C.P. Browman and L. Goldstein. Towards an articulatory phonology. *Phonology*, 3:219–52, 1986.
[5] N. Chomsky and M. Halle. *The sound pattern of English.* Harper & Row, New York, 1968.
[6] R. Daniloff and R. Hammarberg. On defining coarticulation. *J. Phon.*, 1:239–248, 1973.
[7] G. Fant. Distinctive features and phonetic dimensions. *Applications of Linguistics, Cambridge, UK)*, 1969.
[8] W. L. Henke. *Dynamic articulatory model of speech production using computer simulation.* PhD thesis, MIT, Cambridge, MA, 1965.
[9] R. A. Johnson and D. W. Wichern. *Applied multivariate statistical analysis.* Prentice Hall, New Jersey, 4 edition, 1998.
[10] P. A. Keating. The window model of coarticulation: articulatory evidence. *UCLA Working papers in Phonetics*, 69:3–29, 1988.
[11] S. Kullback. *Information theory and statistics.* Dover Pub., New York, 1 edition, 1968.
[12] A.M. Liberman. The grammars of speech and language. *Cog. Psych.*, 1:301–23, 1970.
[13] B. Lindblom. Spectrographic study of vowel reduction. *JASA*, 35:1773–81, 1963.
[14] P.F. MacNeilage. Motor control of serial ordering of speech. *Psychol. Rev.*, 77:182–196, 1970.
[15] K. Moll and R. Daniloff. Investigation of the timing of velar movements during speech. *JASA*, 50(2):678–84, 1971.
[16] S.E.G. Öhman. Coarticulation in VCV utterances: Spectrographic measurements. *JASA*, 39(1):151–68, 1966.
[17] D. Recasens, D.M. Pallarés, and J. Fontdevilla. A model of lingual coarticulation based on articulatory constraints. *JASA*, 102:544–561, 1997.
[18] M.J. Russell and P.J.B. Jackson. Models of speech dynamics in a segmental-HMM recogniser using intermediate linear representations. *Proc. ICSLP*, Denver, CO, 2002.
[19] E.L. Saltzman and K. Munhall. A dynamic approach to gestural patterning in speech production. *Ecology Psychology*, 1:333–82, 1989.
[20] K. Tokuda, H. Zen, and T. Kitamura. Reformulating the HMM as a trajectory model by imposing explicit relationships between static and dynamic vector feature sequences. *Comp. Speech & Lang.*, 21(1):153–73, 2007.
[21] A.A. Wrench. A new resource for production modelling in speech technology. *Proc. Inst. of Acoust.*, Stratford-upon-Avon, UK, 2001.

2.2 Articulatory roles

- *Critical articulator*: position or gesture is critical in production of speech sounds, e.g., TT for [s] and TD for [g].
- *Dependent articulator*: follows one or more critical articulators and its position is influenced by them, e.g., TT, TB for [g].
- *Redundant articulator*: moves in an unconstrained way and its position does not effect the production of sounds critically, e.g., LL, UL for [g].

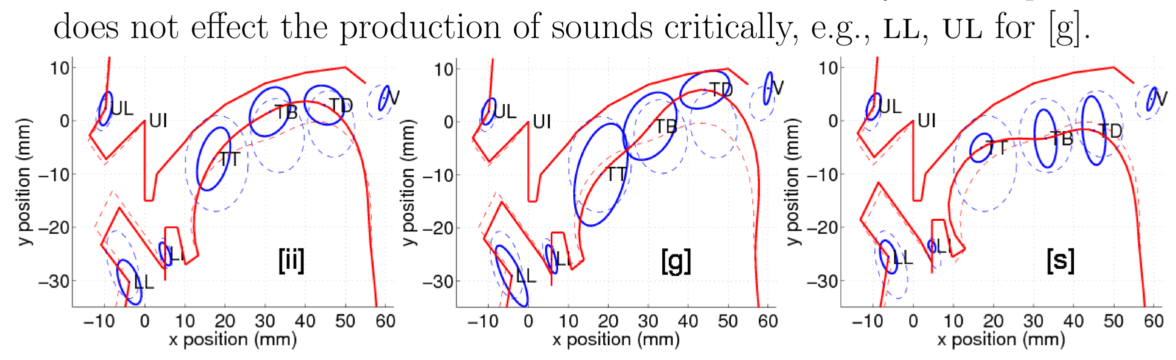

FIGURE 2: *Midsaggital display of different articulatory points and covariance ellipses corresponding to global (dotted) and phone specific distributions (solid).*

3 Modeling articulatory roles

Critical articulators are identified based on the strength of Kullback-Leibler divergence [11] between the model and phone pdfs. Model pdfs of dependent articulators are then conditioned using
- correlations of articulators in space
- correlations amongst articulators

3.1 Articulatory correlations

For any two articulators i and j, univariate correlation matrix is $R = \{r_{ij}\}\forall i,j \in 1..n$, $(n = 14)$; Statistically significant $(\alpha = 0.05)$ and strong $(r_{ij} = 0$ iff $|r_{ij}| < 0.1)$ correlations R^* and covariance $\Sigma_{ij} = \Sigma_{ii}^{1/2} r_{ij} \Sigma_{jj}^{1/2}$.

1D	Ulx	Uly	Llx	Lly	Lix	Liy	Ttx	Tty	Tbx	Tby	Tdx	Tdy	Vx	Vy
Ulx	1.00	.53	.34	-.15	.00	-.17	-.18	.00	-.16	.00	-.22	.00	.00	.00
Uly	.53	1.00	.27	-.31	.00	.00	.00	.29	.00	.19	-.15	.00	.13	.00
Llx	.34	.27	1.00	-.70	.61	-.55	.00	-.31	-.19	.00	-.17	.14	.11	.00
Lly	-.15	-.31	-.70	1.00	-.49	.65	.00	.32	.14	.00	.10	-.10	-.18	.00
Lix	.00	.00	.61	-.49	1.00	-.71	.00	-.43	.00	-.36	.00	.00	.12	.00
Liy	-.17	.00	-.55	.65	-.71	1.00	.00	.60	.12	.42	.00	.00	-.12	.00
Ttx	-.18	.00	.00	.00	.00	.00	1.00	.00	.90	.00	.82	.00	.24	.19
Tty	.00	.29	-.31	.32	-.43	.60	.00	1.00	.18	.53	.11	.00	.00	.00
Tbx	-.16	.00	-.19	.14	.00	.12	.90	.18	1.00	.00	.92	-.24	.14	.00
Tby	.00	.19	.00	.00	-.36	.42	.00	.53	.00	1.00	.00	.75	.00	.00
Tdx	-.22	-.15	-.17	.10	.00	.00	.82	.11	.92	.00	1.00	-.21	.00	.00
Tdy	.00	.00	.14	-.10	.00	.00	.00	.00	-.24	.75	-.21	1.00	.00	.23
Vx	.00	.13	.11	-.18	.12	-.12	.24	.00	.14	.00	.00	.00	1.00	.81
Vy	.00	.00	.00	.00	.00	.00	.19	.00	.00	.00	.00	.23	.81	1.00

2D	Ulx	Uly	Llx	Lly	Lix	Liy	Ttx	Tty	Tbx	Tby	Tdx	Tdy	Vx	Vy
Ulx	1.00	.53	.36	.06	-.07	-.25	-.16	-.09	-.15	-.13	-.22	-.14	.13	-.08
Uly	.53	1.00	.09	-.30	.02	.09	-.07	.32	-.05	.22	-.02	-.01	.14	-.09
Llx	.36	.09	1.00	-.70	.45	-.19	-.02	-.21	-.16	-.07	-.14	.12	.09	.00
Lly	.06	-.30	-.70	1.00	-.09	.59	.02	.27	.10	.04	.07	-.05	-.18	.00
Lix	-.07	.02	.45	-.09	1.00	-.71	.01	-.17	-.05	-.18	.00	.00	.10	.00
Liy	-.25	.09	-.19	.59	-.71	1.00	-.03	.58	.11	.39	.00	.00	-.12	.00
Ttx	-.16	-.07	-.02	.02	.01	-.03	1.00	.06	.90	.05	.83	.08	.21	.13
Tty	-.09	.32	-.21	.27	-.17	.58	.06	1.00	.17	.53	.10	.01	-.05	-.03
Tbx	-.15	-.05	-.16	.10	-.05	.11	.90	.17	1.00	-.03	.91	-.15	.12	-.07
Tby	-.13	.22	-.07	.04	-.18	.39	.05	.53	-.03	1.00	.07	.75	-.17	.09
Tdx	-.22	-.02	-.14	.07	.00	.00	.83	.10	.91	.07	1.00	-.21	.00	.00
Tdy	-.14	-.01	.12	-.05	.00	.00	.08	.01	-.15	.75	-.21	1.00	-.11	.31
Vx	.13	.14	.09	-.18	.10	-.12	.21	-.05	.12	-.17	.00	-.11	1.00	.81
Vy	-.08	-.09	.00	.00	.00	.00	.13	-.03	-.07	.09	.00	.31	.81	1.00

TABLE 1: *Univariate R^* (left) and bivariate $\mathbf{R}^*$ (right) correlation matrices of strong and statistically-significant correlations.*

In bivariate case, ρ_{ij}^1 and ρ_{ij}^1 are the canonical correlations [9] between them and $\mathbf{U}_i$ and $\mathbf{U}_j$ are the corresponding eigenvectors. Bivariate correlation matrix is $\mathbf{R} = \{\mathbf{r}_{ij}\}\forall i,j \in 1..n$, $(n = 7)$ where $\mathbf{r}_{ij} = \mathbf{U}_i \mathrm{diag}(\rho_{ij}^1, \rho_{ij}^2)\mathbf{U}_j'$; Statistically significant $(\alpha = 0.05)$ and strong $(\rho_{ij} = 0$ iff $\rho_{ij} < 0.15)$ correlations $\mathbf{R}^*$ and covariance $\mathbf{\Sigma}_{ij} = \mathbf{\Sigma}_{ii}^{1/2} \mathbf{r}_{ij} \mathbf{\Sigma}_{jj}^{1/2}$.

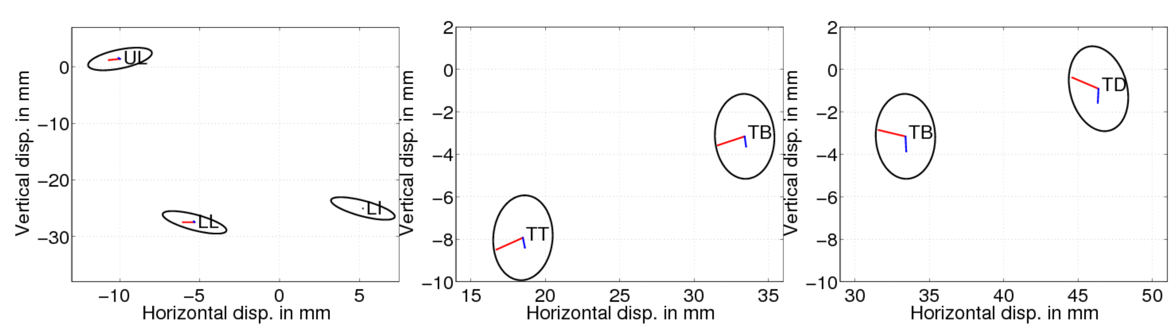

FIGURE 3: *Directions of first and second significant canonical correlations between* UL, LL *(left),* TT, TB *(centre) and* TB, TD *(right), grand covariance ellipses are plotted in black.*

3.2 Identifying articulatory roles and updating models

1. **Initialisation** Data samples at the phone midpoints provide reference statistics of normal articulatory distributions: mean $\boldsymbol{\mu}_i^\phi$ and covariance $\boldsymbol{\sigma}_i^\phi$ (2D).
   The model pdfs are initialised with global statistics: $m_i^\phi = M_i$ and $s_i^\phi = \Sigma_{ii}$ (1D), $\boldsymbol{m}_i^\phi = \boldsymbol{M}_i$ and $\boldsymbol{s}_i^\phi = \boldsymbol{\Sigma}_i$ (2D).
2. **Critical identification (C-step)**
   Identify critical articulator:
   At each level, $k = 1..n$, calculate bivariate KL divergence between model and phone distributions $J_k^\phi(i)$ for each articulator i, incorporating the standard error by multiplying the variance of N data points by $(N+1)/N$. The next critical articulator $j = \mathrm{argmax}_i\{J_k^\phi(i)\}$ is selected iff $J_k^\phi(j) > \theta_{\mathrm{crit}}$.
   Set critical articulator distribution: Model statistics are set to phone statistics :
$$\boldsymbol{m}_j^\phi \leftarrow \boldsymbol{\mu}_j^\phi, \boldsymbol{s}_j^\phi \leftarrow \boldsymbol{\sigma}_j^\phi. \qquad (1)$$
3. **Dependent update (D-step)**
   Gather dependency statistics: The covariations are collated from the grand statistics, Σ_{ii}, Σ_{ij} and Σ_{ii}, for all dependent (iff $J_k^\phi(i) > \theta_{\mathrm{dep}}$.) and critical articulators i and j; $\mathbf{\Sigma}_{ii}$, $\mathbf{\Sigma}_{ij}$ and $\mathbf{\Sigma}_{jj}$.
   Update dependent articulator pdfs: Combining these statistics with the phone ones for any critical articulator j, the conditional pdfs get re-estimated, according to [1], which for 2D gives:
$$\mathbf{m}_i^\phi \leftarrow \boldsymbol{M}_i + \mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1}\left(\boldsymbol{\mu}_j^\phi - \boldsymbol{M}_j\right)$$
$$\mathbf{s}_i^\phi \leftarrow \mathbf{\Sigma}_{ii} + \mathbf{\Sigma}_{ij}\mathbf{\Sigma}_{jj}^{-1}\left(\boldsymbol{\sigma}_j^\phi - \mathbf{\Sigma}_{jj}\right)\mathbf{\Sigma}_{jj}^{-1}\mathbf{\Sigma}_{ij}'. \qquad (2)$$

The algorithm stops in C-step when $J_k^\phi(i) < \theta_{\mathrm{crit}} \forall i \in 1..n$.

4 Results

4.1 Model convergence

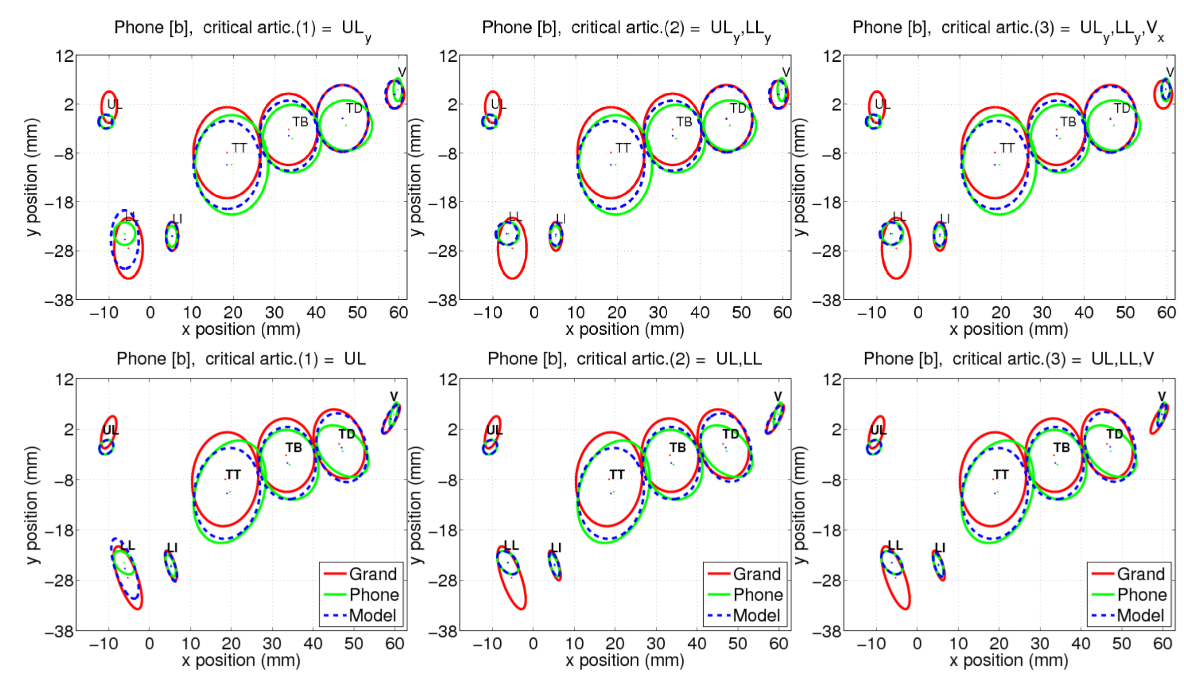

FIGURE 4: *Convergence of 1D (top) and 2D (bottom) models.*

4.2 Comparison with phonetic features

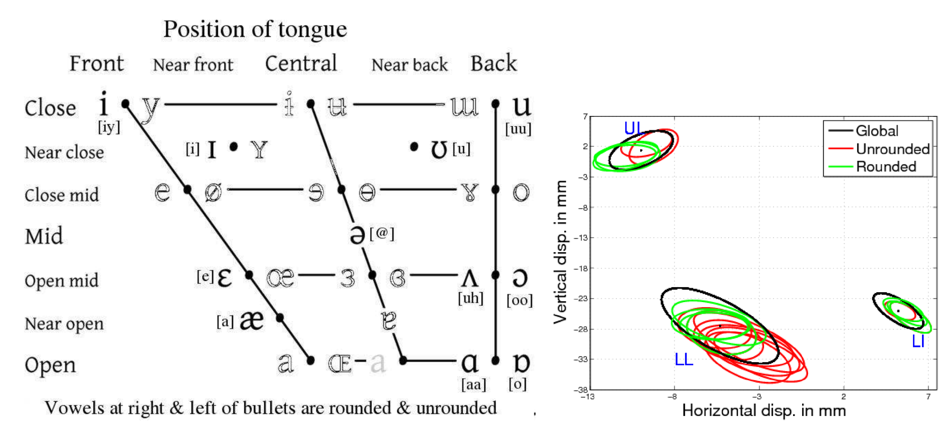

FIGURE 5: *IPA vowel chart with vowels in the database highlighted and MOCHA notation in [ ] (left); Critical lip and jaw pdfs for rounded and unrounded vowels (right).*

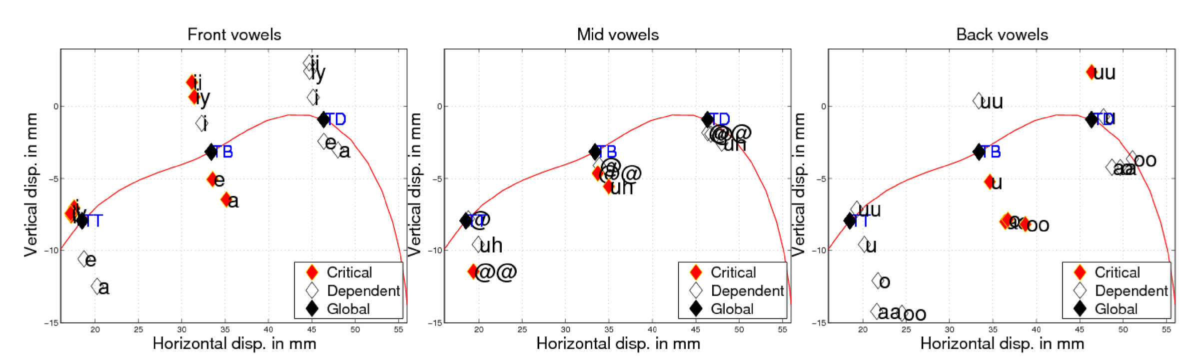

FIGURE 6: *Critical and non-critical tongue positions for front, mid and back vowels*

	1D							2D						
	UL	LL	LI	TT	TB	TD	V	UL	LL	LI	TT	TB	TD	V
[p]	1y	3y					2x	1	3					2
[b]	1y	2y					3x	1	2					3
[m]	1y	2y						1	2					3
[t]				2y			1x				1			2
[d]				1y			2x				1			2
[n]				1y			2x				1			2
[k]					3y	1y	2x						1	2
[g]					3y	1y	2x						1	2
[ng]						1y	2x	3					1	2
[f]	2y	1y	4y				3x	3	1		4			2
[v]	3y	1y	4y				2x	3	1					2
[th]		3y		1x,4y			2x	5	3	4	1			2
[dh]				1x,2y			3x				1			2
[s]		5y	2y	1y,3x			4x		4	2	1			3
[z]		5y	1y	2x,3y			4x		4	2	1			3
[sh]			3y	1y,2x		4y	5x			2	1		3	4
[zh]		5y	1y	2y,4x		3y			4	1	2		3	5
[ch]			1y	2y,3x	5y		4x	5		2	1	4		3
[jh]			5y	1y,3x	2y		4x		5	4	1	2		3

FIGURE 7: *Critical articulators identified for consonants using 1D and 2D techniques.*

5 Conclusion and future work

Proposed 1D and 2D statistical techniques for identifying articulatory roles in a data-driven way
Fit of the models to the respective phone pdfs is presented and the results obtained are compared with IPA phonetic features.
Future work will focus on incorporating temporal aspects of coarticulation, finding ways to exploit redundant degrees of freedom within dependent articulators.