Coarticulatory constraints determined by automatic identification from articulograph data

Philip JB Jackson & Veena D Singampalli
CVSSP, University of Surrey, UK

p.jackson | v.singampalli@surrey.ac.uk
www.ee.surrey.ac.uk/personal/p.jackson/dansa

UNIVERSITY OF SURREY

Abstract: Identification of critical, dependent and redundant articulators from 1D and 2D statistics of EMA data. Compared with IPA. Interesting findings.

INTRODUCTION

Rôles in articulation

Critical: essential for a phone's production

Dependent: follows critical articulator due to connection with it

Redundant: free to move, unconstrained by current phone

Background on coarticulation

Coarticulation is a major challenge for speech technologies

Coarticulation affects the *planning stage* of speech production to simplify commands for sequence of speech gestures [1,7]

Coarticulation affects the *articulation stage* from stiffness, damping and inertia, and effort minimisation [2,5,6,8,9]

Present work proposes use of statistical models to uncover the likely representation of gestural commands for each phone [10]

METHOD

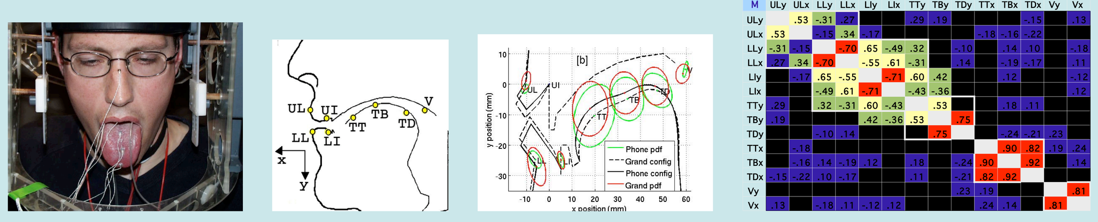
Electromagnetic articulograph (EMA) data [11]

Midsagittal trajectories of 7 glued coils

Male (msak) and female (fsew) British English subjects

460 sentences each with automatic phone transcription, which required some manual re-labelling

Calibrated data smoothed and re-sampled at 100 Hz



UL - upper lip
LL - lower lip
LI - lower incisor

TT - tongue tip
TB - tongue blade
TD - tongue dorsum

V - velum

Nose & upper incisor used as ref.

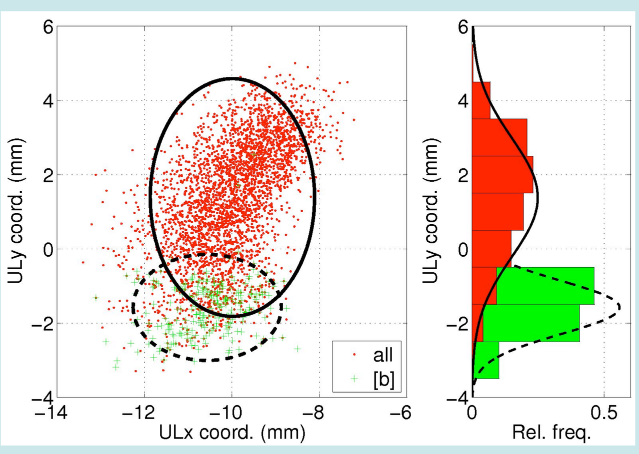
Symmetric Kullback-Leibler divergence (KLD)

Measures distance between two distributions in terms of average log-likelihood ratio [4]:

$$J = \int_{-\infty}^{\infty} f_1(x) \ln \frac{f_1(x)}{f_2(x)}\, dx + \int_{-\infty}^{\infty} f_2(x) \ln \frac{f_2(x)}{f_1(x)}\, dx$$
$$= \frac{1}{2}\left[\mathrm{tr}\left(\Sigma_1 - \Sigma_2\right)\left(\Sigma_2^{-1} - \Sigma_1^{-1}\right) + \mathrm{tr}\left(\Sigma_1^{-1} + \Sigma_2^{-1}\right)\left(\mu_1 - \mu_2\right)\left(\mu_1 - \mu_2\right)^{\mathrm{T}}\right]$$

where $f_1 = \mathcal{N}(\mu_1, \Sigma_1)$ and $f_2 = \mathcal{N}(\mu_2, \Sigma_2)$.

Articulatory probability density functions (pdfs)



Low KLD (0.2) for [g] shows that ULy is redundant for velar

High KLD (9.1) for [b] marks ULy as critical for bilabial correlated LLy becomes dependent for [b]

RESULTS

Comparison with International Phonetic Alphabet (IPA)

Purpose of IPA:

- to capture linguistic distinctions
- from human speech
- observed by phoneticians

Speech technology needs:

- descriptions of typical phone characteristics in a voice
- to include effects of phoneme-to-phone mapping
- ways to incorporate knowledge from all sources

Critical articulators for consonants

fricatives and *affricates* were most constrained, setting jaw LI for [s,z,sh,zh,ch,jh], as well as expected IPA place

stops and *nasals* gave expected place, after labels corrected, and velum identified in nasals as secondary for one subject

liquids and *semivowels* similar to predictions, except lateral [l]

Φ	IPA 1D	IPA 2D	msak 1D	msak 2D	fsew 1D	fsew 2D
[f]	lly llx	ll	lly uly	ll	lly llx	ll
[v]	lly llx	ll	lly uly	ll	lly llx	ll
[th]	tty ttx	tt	ttx tty lly	tt ll	tty ttx tby	tt tb ll li
[dh]	tty ttx	tt	ttx tty	tt	ttx tty	tt
[s]	tty ttx	tt	liy ttx tty	tt li	liy tty tbx	li tt
[z]	tty ttx	tt	liy ttx tty	tt li	liy tty ttx	li tt
[sh]	tty ttx	tt	tty tbx liy tdy	tt li	tty liy ttx	li tt
[zh]	tty ttx	tt	liy tty tdy ttx lly	li tt td	tty liy tby tdx llx	tt li tb
[h]	-	-	-	-	tty	tt
[ch]	tty ttx	tt	liy tty tbx tby	tt li	tty liy ttx	li tt ul
[jh]	tty ttx	tt	tty tby ttx liy	tt tb	liy tty ttx	li tt
[p]	uly lly	ul ll	uly lly	ul	uly	ul
[b]	uly lly	ul ll	uly lly	ul ll	uly lly	ul ll
[t]	tty ttx	tt	tty	tt	tty liy	tt
[d]	tty ttx	tt	tty	tt	tty	tt
[k]	tdy	td	tdy	td	tdy	td
[g]	tdy	td	tdy	td	tdy	td
[m]	uly lly vx	ul ll v	uly lly vx	ul ll	lly uly	ll ul
[n]	tty ttx vx	tt v	tty vx	tt	tty	tt
[ng]	tdy vx	td v	tdy vx	td v	tdy	td
[w]	ulx llx tdy	ul ll td	uly	ul	uly tdy lly	ul td ll
[l]	tty ttx	tt	-	-	-	-
[r]	tty ttx	tt	tbx	tt	ttx	tt
[y]	tby tbx	tb	tby	tb	tby	tb tt

Critical articulator identification algorithm

Derive statistics → Initialise model → Compute KLD → Identify next critical articulator → Update any dependent articulators

C-step:
- Identifies most divergent articulator
- Sets its pdf equal to the phone pdf

D-step:
- Gets stats for correlated articulators
- Conditions their pdfs on critical pdf

Pseudocode description of the algorithm

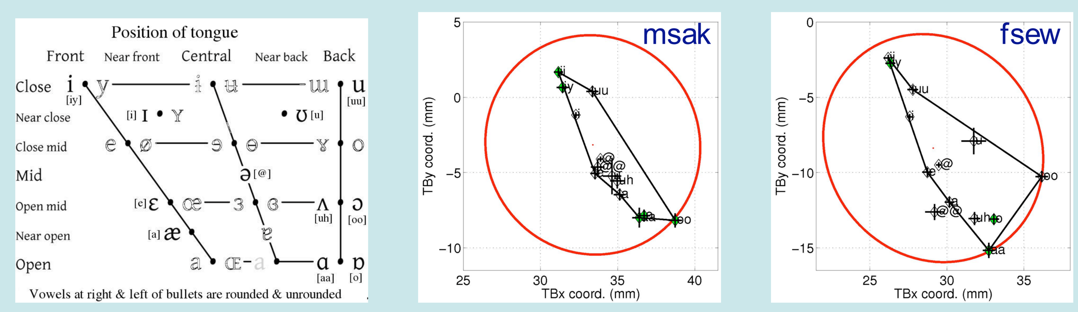
Global statistics: $\mathbf{\Gamma} = \{M, \Sigma\}$, grand correlation $\mathbf{R}^*$, sample size N
Phone statistics: $\mathbf{\Lambda}^\phi = \{\mu^\phi, \Sigma^\phi\}$, phone correlation R^ϕ, sample size N^ϕ
Model statistics $\mathbf{\Delta}_k^\phi = \{m_k^\phi, S_k^\phi\}$, sample size N_k
Threshold $\Theta = \{\theta_{\mathrm{C}}, \theta_{\mathrm{D}}\}$
Model initialisation:
level $k = 0$
$m_{k,i}^\phi = M_i,\ S_{k,i}^\phi = \Sigma_i,\ N_{k,i} = N,\ \forall i \in \{1..n\}$
Critical articulator list $C_k^\phi = \{\}$
while($k \leq n$)
    Identification divergence: $J_{k,i}^\phi = \mathrm{computeIdiv}(\mathbf{\Delta}_{k,i}^\phi, \mathbf{\Lambda}_i^\phi, N_{k,i}, N^\phi),\ \forall i \in \{1..n\}$
    Articulator with maximum divergence $j = \mathrm{argmax}\{J_{k,1}^\phi ..., J_{k,n}^\phi\}$
    C-step:
    if($J_{k,j}^\phi > \theta_{\mathrm{C}}$)
        Go to the next level $k = k + 1$
        $\mathbf{\Delta}_k^\phi = \mathbf{\Delta}_{k-1}^\phi$
        $N_k = N_{k-1}$
        Identify critical articulator: $C_k^\phi \leftarrow \{C_{k-1}^\phi\} \cup \{j\}$
        Update distribution: $m_{k,j}^\phi \leftarrow \mu_j^\phi,\ S_{k,j}^\phi \leftarrow \Sigma_j^\phi$
        $N_{k,j} \leftarrow N^\phi$
        D-step:
        $[N_k, \mathbf{\Delta}_k^\phi] = \mathrm{updateDep}(\mathbf{\Delta}_k^\phi, N_k, \mathbf{\Gamma}, \mathbf{\Lambda}^\phi, \Theta, C_k^\phi, R^*, R^\phi)$
    else
        Critical articulator list $\hat{C}^\phi = C_k^\phi$
        Model distribution $\hat{m}^\phi = m_k^\phi,\ \hat{S}^\phi = S_k^\phi$
        No: of critical articulators $\hat{N}^\phi = k$.
        break
    end if
end while

function computeIdiv$(\mathbf{\Delta}_{k,i}^\phi, \mathbf{\Lambda}_i^\phi, N_{k,i}, N^\phi)$
Incorporating standard error:
$S_1 = S_{k,i}^\phi + (S_{k,i}^\phi / N_{k,i}),\ S_2 = \Sigma_i^\phi + (\Sigma_i^\phi / N^\phi)$
$J = \frac{1}{2}\left(tr(S_1 - S_2)(S_2^{-1} - S_1^{-1}) + tr(S_1^{-1} + S_2^{-1})(m_{k,i}^\phi - \mu_i^\phi)(m_{k,i}^\phi - \mu_i^\phi)'\right)$
return J

function updateDep$(\mathbf{\Delta}_k^\phi, N_k, \mathbf{\Gamma}, \mathbf{\Lambda}^\phi, \Theta, C, R^*, R^\phi)$
Collate critical grand statistics: $M_{\{C\}}, \Sigma_{\{C\}\{C\}}$ (from Σ and R^*)
Collate critical phone statistics: $\mu_{\{C\}}^\phi$ and covariance matrix $\Sigma_{\{C\}\{C\}}^\phi$ (from Σ^ϕ and R^ϕ)
for $i \in \{1..., n\} \cap \{C\}$
    $J_{k,i}^\phi = \mathrm{computeIdiv}(\mathbf{\Delta}_{k,i}^\phi, \mathbf{\Lambda}_i^\phi, N_{k,i}, N^\phi)$
    if($J_{k,i}^\phi > \theta_D$)
        Dependent covariance : $\Sigma_{i\{C\}}$
        update mean: $m_{k,i}^\phi \leftarrow M_i + \Sigma_{i\{C\}} \Sigma_{\{C\}\{C\}}^{-1} (\mu_{\{C\}}^\phi - M_{\{C\}})$
        update variance: $S_{k,i}^\phi \leftarrow \Sigma_i + \Sigma_{i\{C\}} \Sigma_{\{C\}\{C\}}^{-1} (\Sigma_{\{C\}\{C\}}^\phi - \Sigma_{\{C\}\{C\}}) \Sigma_{\{C\}\{C\}}^{-1} \Sigma_{i\{C\}}'$
        $N_{k,i} \leftarrow N^\phi$
    end if
end for
return $\mathbf{\Delta}_k^\phi, N_k$

Critical articulators for vowels

Targets are in both acoustic and articulatory domains



high front and back vowels selected TBy orTDy/TB

central and reduced vowels picked none

open and low vowels tended to choose LLy/LL

low back vowels also picked TBx or TBy/TB

Φ	IPA 1D	IPA 2D	msak 1D	msak 2D	fsew 1D	fsew 2D
[ii]	tty	tt	tdy lly ttx	tb	tby	tb td
[iy]	tty	tt	tdy	tb	tdy ttx	tb td
[uu]	tdy ulx llx	td ul ll	tdy	-	tby	td
[i]	tty	tt	-	-	-	-
[u]	tdy ulx llx	td ul ll	-	-	-	-
[@]	tby	tb	-	-	-	-
[@@]	tby	tb	lly	-	tty llx	tb ll
[uh]	tby	tb	-	-	tby	tb
[oo]	tdy ulx llx	td ul ll	lly tbx tty	tb ll	tdx lly tdy tby	td ll
[o]	tdy ulx llx	td ul ll	tby	tb	tby tdx	tb
[aa]	tdy	td	lly tby tbx	ll tb	tby tdx	tb
[a]	tty	tt	lly	ll	lly tdy	ll td
[e]	tty	tt	lly	-	lly	ll

Discussion

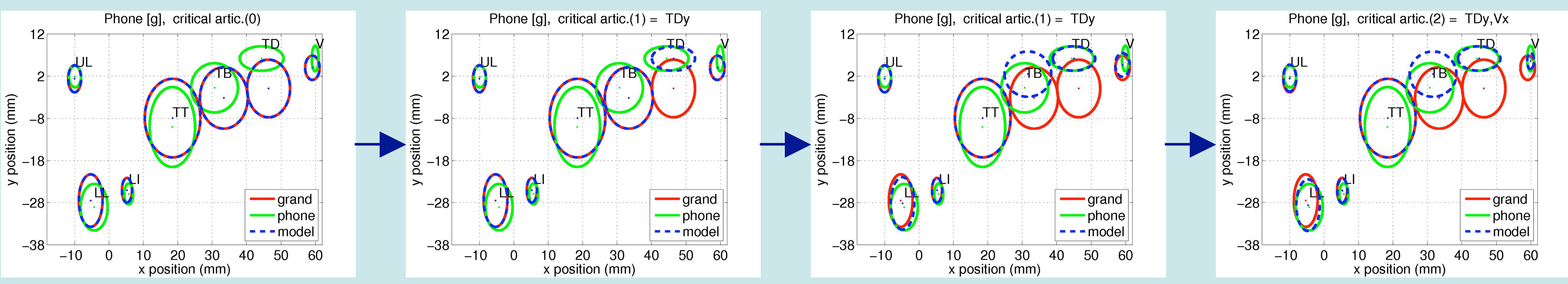
Limitations of the MOCHA-TIMIT data

- labelling alignment and transcription errors
- sensor calibration and attachment problems
- pronunciation of read sentences with EPG
- number of sentences, subjects, speaking styles

Extensions of the technique

- more accurate pdfs, e.g., GMM or numerical methods [3]
- from 1D and 2D to 3D data

Convergence of 1D pdfs as algorithm identifies critical articulators for phone [g]



SUMMARY

Conclusions

Presented algorithm for identifying critical, dependent and redundant roles of articulators during speech production

Results with EMA data compared well to IPA for consonants; fricatives claimed additional critical articulators

Central and reduced vowels had no critical articulator

Future perspectives

Potential to exploit new knowledge of articulatory constraints as conditional dependencies in probabilstic speech models

Need to investigate dynamic behaviour of articulatory constraints

Opportunity to expand critical articulator analysis to other data

REFERENCES

[1] N. Chomsky and M. Halle. The sound pattern of English. Harper & Row, New York, 1968.
[2] J. Dang, M. Honda, and K. Honda. Investigation of coarticulation in continuous speech of Japanese. Acoust. Sci. & Tech., 25(5):318 – 329, 2004.
[3] J. R. Hershey and P. A. Olsen. Approximating the Kullback Leibler divergence between Gaussian mixture models. Proc. ICASSP, 4:317–320, 2007.
[4] S. Kullback. Information theory and statistics. Dover Pub., New York, 1 edition, 1968.
[5] B. Lindblom. Spectrographic study of vowel reduction. J. Acoust. Soc. Am., 35:1773–81, 1963.
[6] A. Löfqvist. Speech as audible gestures. In W.J. Hardcastle and A. Marchal (Eds.), Speech production and Speech Modeling. Dordrecht: Kluwer Academic Publishers, 1990.
[7] K. Moll and R. Daniloff. Investigation of the timing of velar movements during speech. J. Acoust. Soc. Am., 50(2):678–84, 1971.
[8] S. Öhman. Numerical model of coarticulation. J. Acoust. Soc. Am., 41(2):310–20, 1967.
[9] E. Saltzman and K. Munhall. A dynamic approach to gestural patterning in speech production. Ecology Psychology, 1(4):333–82, 1989.
[10] V. D. Singampalli and P. J. B. Jackson. Statistical identification of critical, dependent and redundant articulators. Proc. Interspeech, Antwerp, pages 70–73, 2007.
[11] A. Wrench. A new resource for production modeling in speech technology. Proc. Inst. of Acoust., Stratford-upon-Avon, UK, 2001.

ISSP'08
Poster 2.12 presented at 8th International Seminar on Speech Production (ISSP'08) in Strasbourg, France on Wednesday 10 December 2008