Statistical identification of critical, dependent and redundant articulators

Veena D Singampalli and Philip JB Jackson

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

August 28, 2007

Interspeech 2007, August 27-31, Antwerp, Belgium

Overview

Introduction

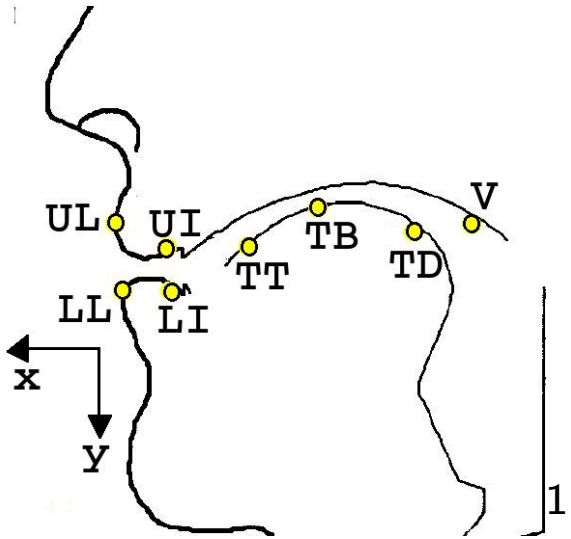
Articulatory representations in modelling coarticulation and in ASR:

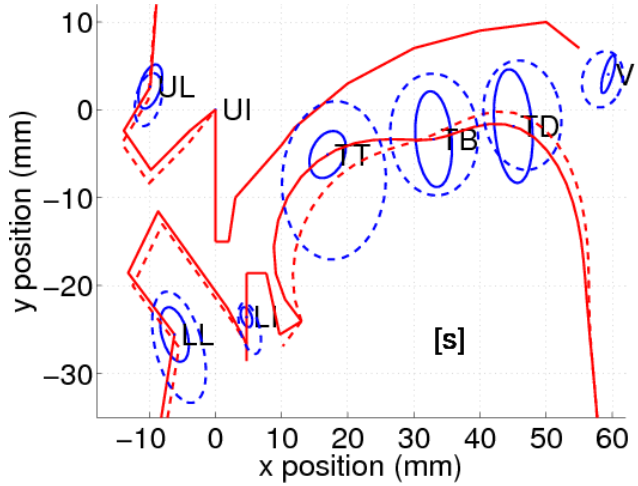
- distinctive phonetic features [Henke, 1965, Daniloff and Hammarberg, 1973, Eide, 2001, Kirchhoff, 1999]
- gestural constraints based on articulatory phonetic rules [Recasens et al., 1997, Erler and Freeman, 1996, Richardson et al., 2000]
- statistical descriptions [Blackburn and Young, 2000, Bakis, 1991].

Constraints on gestures determined by role:

- Critical
- Dependent
- Redundant

Midsaggital views of human speech production system:





- EMA data from MOCHA database [Wrench, 2001] for one male and one female speakers, Hann window smoothed, 10ms frame rate.

[1] [Saltzman and Munhall, 1989]

Our approach

Our algorithm

- is entirely statistical and data driven
- identifies articulatory roles (1D and 2D versions)
- captures the constraints in terms of:
  1. correlated movement of every articulator in space
  2. correlations amongst articulators

Articulatory correlations (1D)

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

	ULx	ULy	LLx	LLy	LIx	LIy	TTx	TTy	TBx	TBy	TDx	TDy	Vx	Vy
ULx	1.00	.53	.34	-.15	.03	-.17	-.18	.02	-.16	-.04	-.22	-.08	.05	-.08
ULy	.53	1.00	.27	-.31	-.04	.05	-.10	.29	-.08	.19	-.15	.08	.13	.02
LLx	.34	.27	1.00	-.70	.61	-.55	-.06	-.31	-.19	-.08	-.17	.14	.11	.02
LLy	-.15	-.31	-.70	1.00	-.49	.65	.01	.32	.14	.06	.10	-.10	-.18	-.09
LIx	.03	-.04	.61	-.49	1.00	-.71	.03	-.43	-.05	-.36	.03	-.02	.12	.05
LIy	-.17	.05	-.55	.65	-.71	1.00	-.01	.60	.12	.42	.04	.09	-.12	-.05
TTx	-.18	-.10	-.06	.01	.03	-.01	1.00	.06	.90	.06	.82	-.01	.24	.19
TTy	.02	.29	-.31	.32	-.43	.60	.06	1.00	.18	.53	.11	.08	-.05	-.06
TBx	-.16	-.08	-.19	.14	-.05	.12	.90	.18	1.00	-.03	.92	-.24	.14	.06
TBy	-.04	.19	-.08	.06	-.36	.42	.06	.53	-.03	1.00	-.02	.75	-.04	.08
TDx	-.22	-.15	-.17	.10	.03	.04	.82	.11	.92	-.02	1.00	-.21	.05	.01
TDy	-.08	.08	.14	-.10	-.02	.09	-.01	.08	-.24	.75	-.21	1.00	.06	.23
Vx	.05	.13	.11	-.18	.12	-.12	.24	-.05	.14	-.04	.05	.06	1.00	.81
Vy	-.08	.02	.02	-.09	.05	-.05	.19	-.06	.06	.08	.01	.23	.81	1.00

Articulatory correlations (1D)

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

	ULx	ULy	LLx	LLy	LIx	LIy	TTx	TTy	TBx	TBy	TDx	TDy	Vx	Vy
ULx	1.00	.53	.34	-.15	.03	-.17	-.18	.02	-.16	-.04	-.22	-.08	.05	-.08
ULy	.53	1.00	.27	-.31	-.04	.05	-.10	.29	-.08	.19	-.15	.08	.13	.02
LLx	.34	.27	1.00	-.70	.61	-.55	-.06	-.31	-.19	-.08	-.17	.14	.11	.00
LLy	-.15	-.31	-.70	1.00	-.49	.65	.00	.32	.14	.06	.10	-.10	-.18	-.09
LIx	.03	-.04	.61	-.49	1.00	-.71	.03	-.43	-.05	-.36	.03	-.02	.12	.05
LIy	-.17	.05	-.55	.65	-.71	1.00	.00	.60	.12	.42	.04	.09	-.12	-.05
TTx	-.18	-.10	-.06	.00	.03	.00	1.00	.06	.90	.06	.82	.00	.24	.19
TTy	.02	.29	-.31	.32	-.43	.60	.06	1.00	.18	.53	.11	.08	-.05	-.06
TBx	-.16	-.08	-.19	.14	-.05	.12	.90	.18	1.00	-.03	.92	-.24	.14	.06
TBy	-.04	.19	-.08	.06	-.36	.42	.06	.53	-.03	1.00	-.02	.75	-.04	.08
TDx	-.22	-.15	-.17	.10	.03	.04	.82	.11	.92	-.02	1.00	-.21	.05	.00
TDy	-.08	.08	.14	-.10	-.02	.09	.00	.08	-.24	.75	-.21	1.00	.06	.23
Vx	.05	.13	.11	-.18	.12	-.12	.24	-.05	.14	-.04	.05	.06	1.00	.81
Vy	-.08	.02	.00	-.09	.05	-.05	.19	-.06	.06	.08	.00	.23	.81	1.00

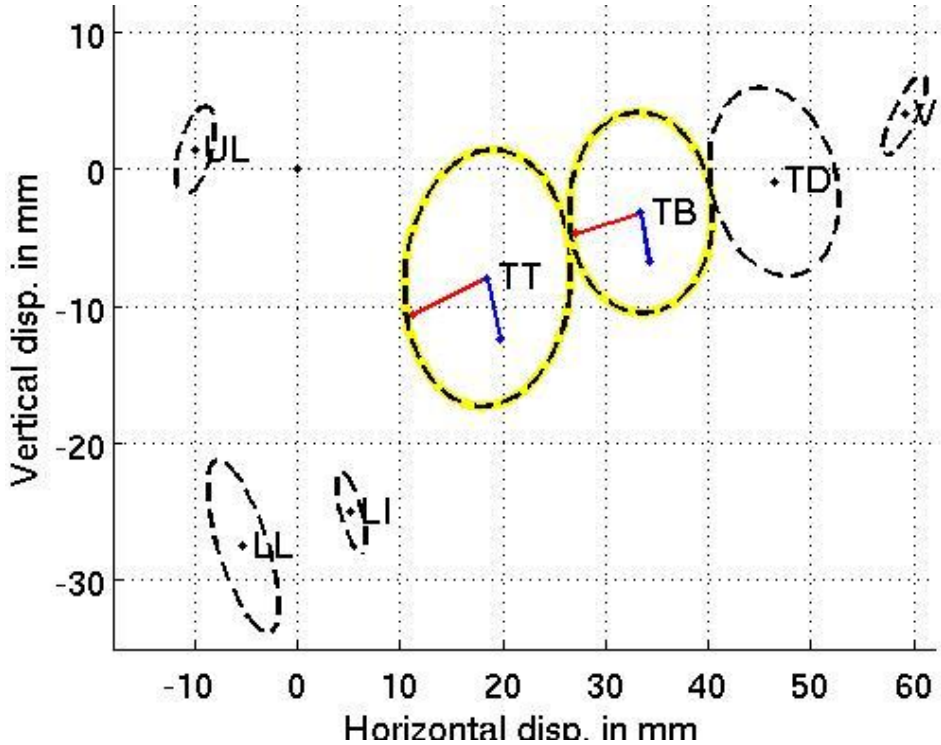
Articulatory correlations (1D)

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

	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

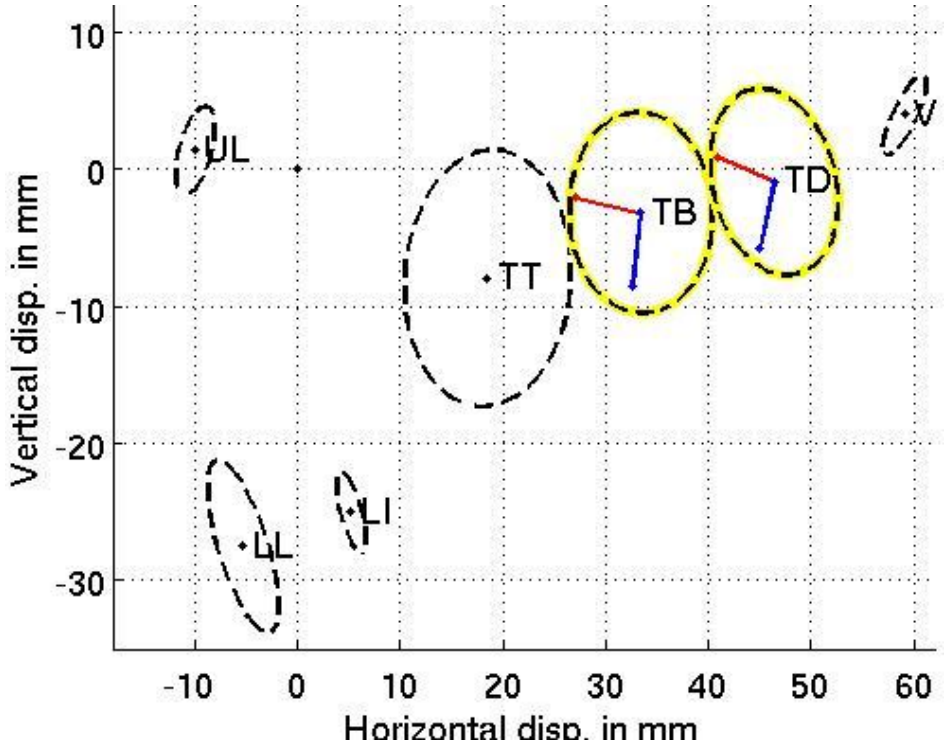
Articulatory correlations (2D)

- Canonical correlation analysis: $(\rho_{ij}^1, \rho_{ij}^2)$; $\mathbf{U}_i$, $\mathbf{U}_j$
- 2D correlations $\mathbf{R} = \{\mathbf{U}_i \mathrm{diag}(\rho_{ij}^1, \rho_{ij}^2)\mathbf{U}_j'\} \forall i, j \in 1..n$.



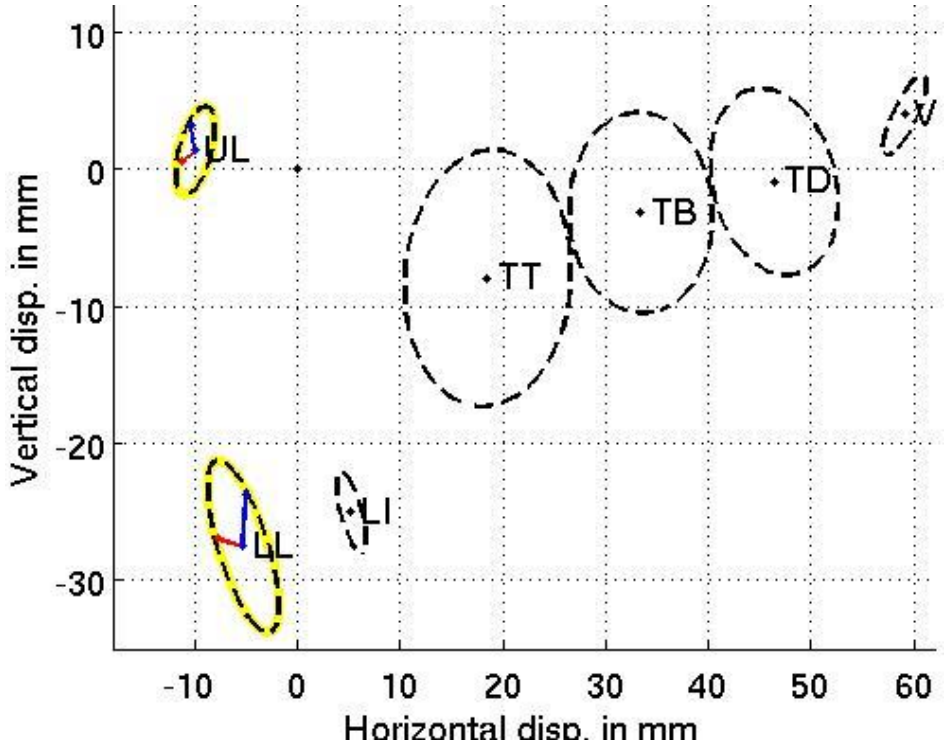
Articulatory correlations (2D)

- Canonical correlation analysis: $(\rho_{ij}^1, \rho_{ij}^2)$; $\mathbf{U}_i$, $\mathbf{U}_j$
- 2D correlations $\mathbf{R} = \{\mathbf{U}_i \mathrm{diag}(\rho_{ij}^1, \rho_{ij}^2)\mathbf{U}_j'\} \forall i, j \in 1..n$.

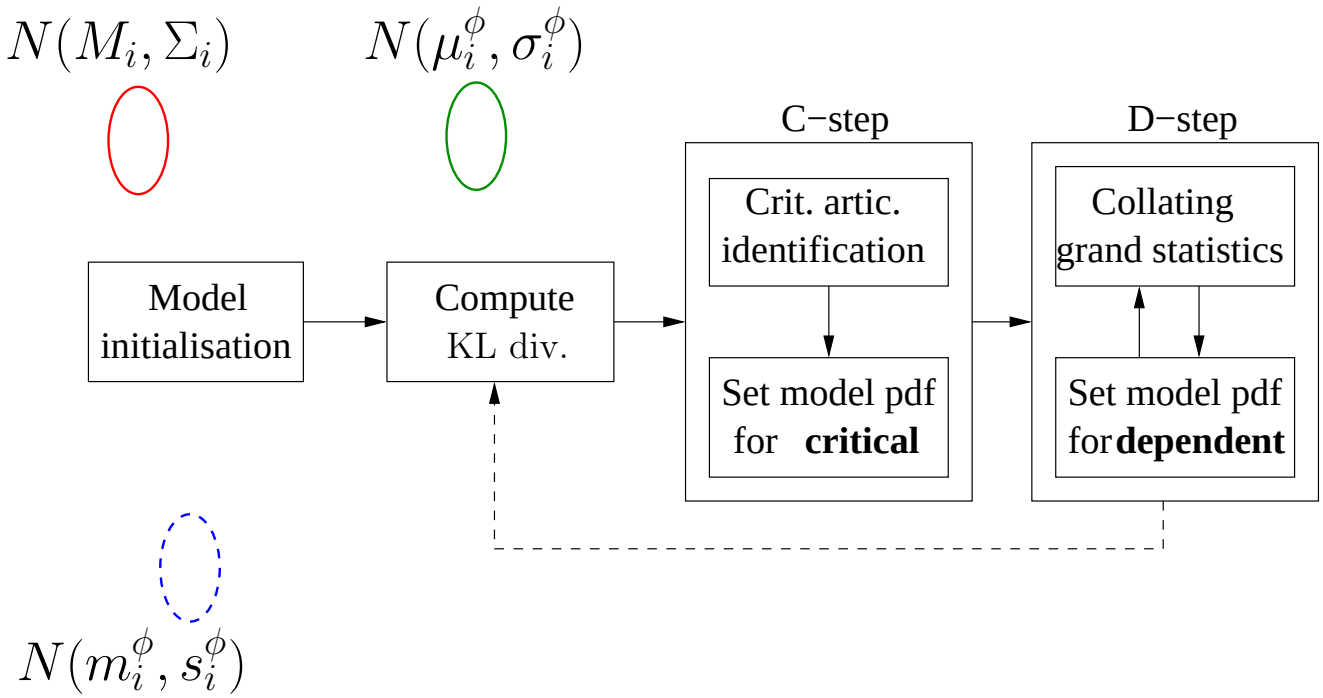


Articulatory correlations (2D)

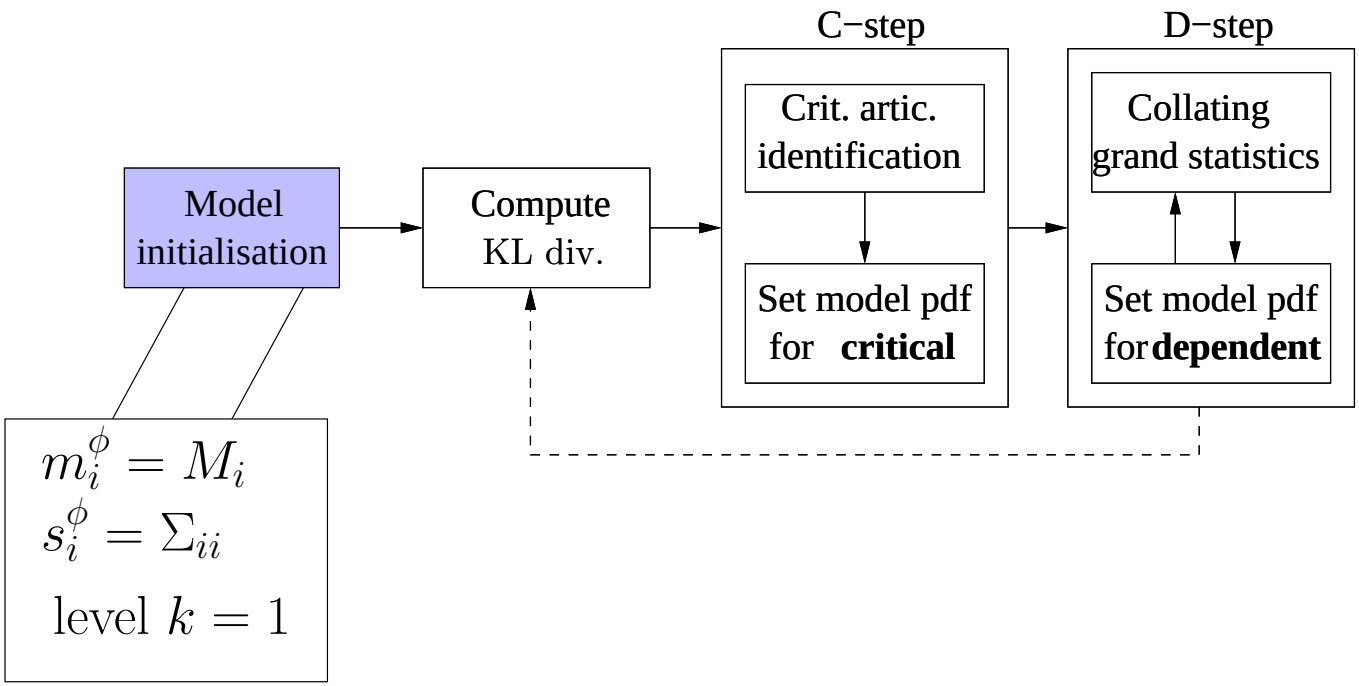
- Canonical correlation analysis: $(\rho_{ij}^1, \rho_{ij}^2)$; $\mathbf{U}_i$, $\mathbf{U}_j$
- 2D correlations $\mathbf{R} = \{\mathbf{U}_i \mathrm{diag}(\rho_{ij}^1, \rho_{ij}^2)\mathbf{U}_j'\} \forall i, j \in 1..n$.



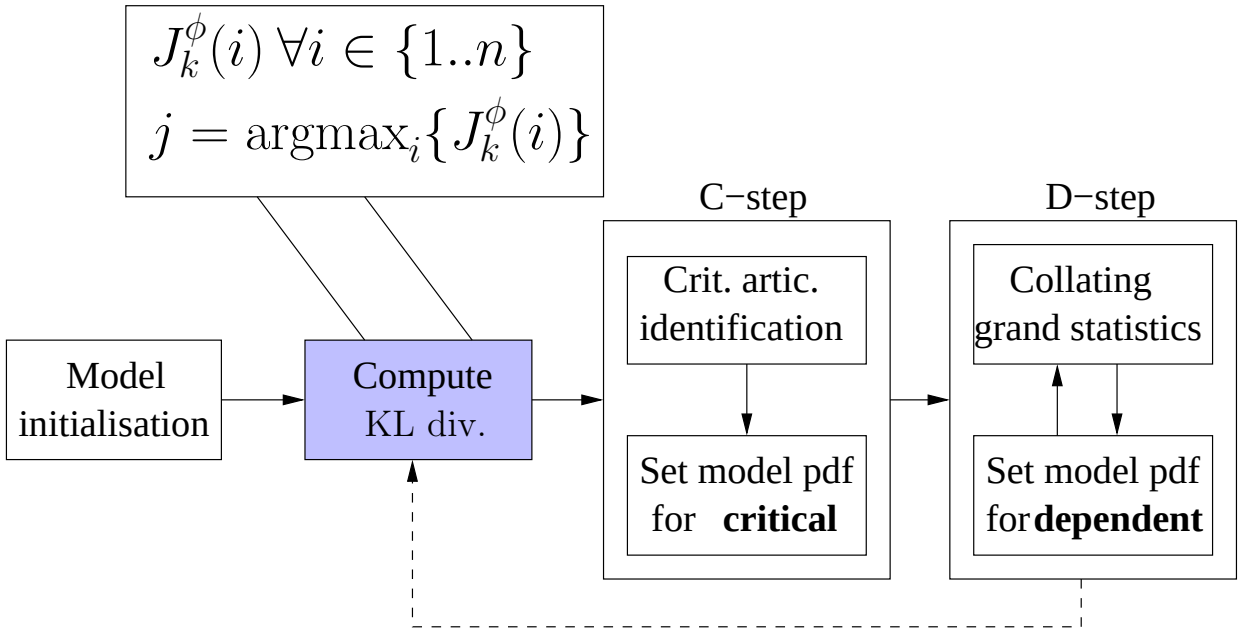
Algorithm



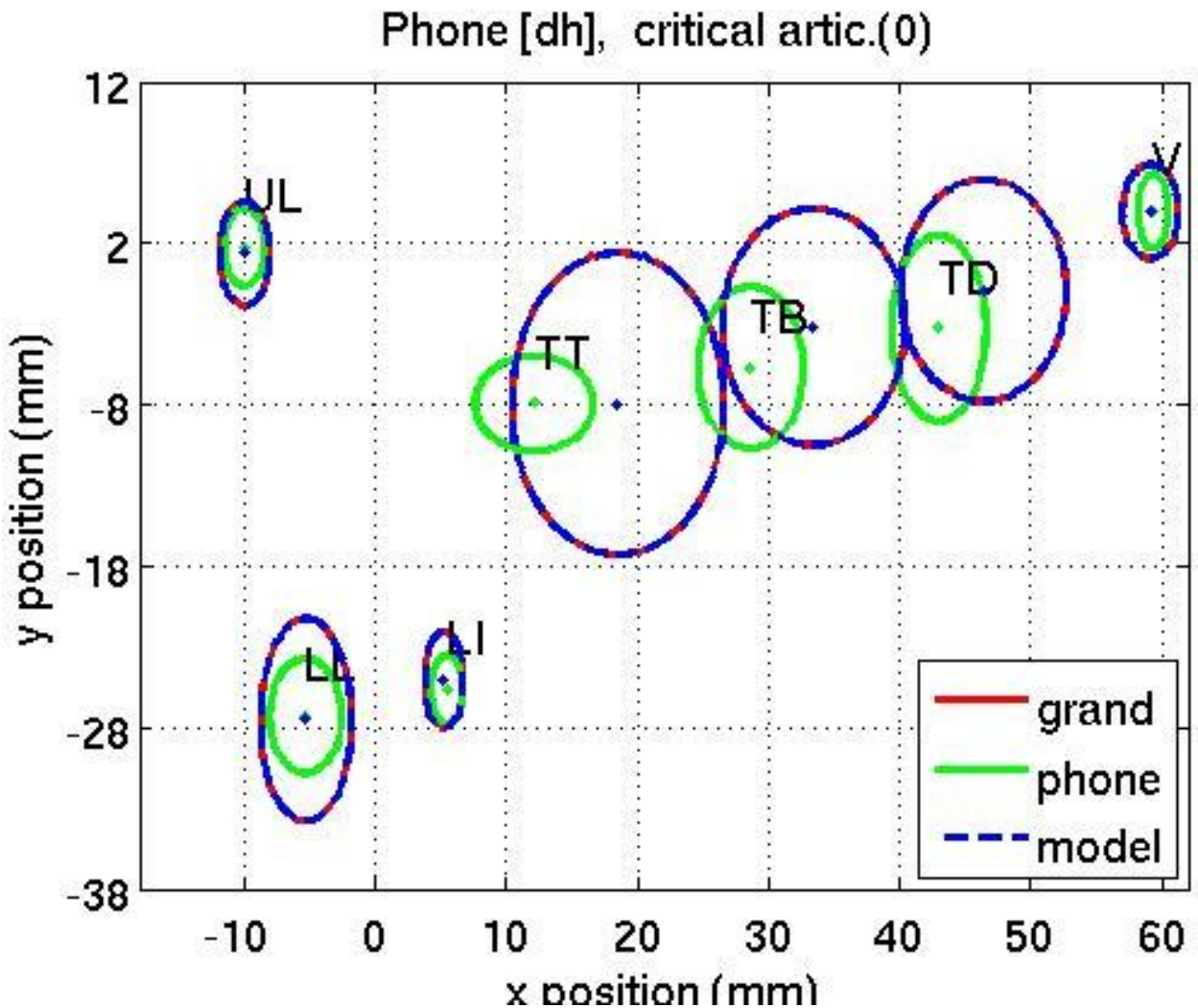
Algorithm



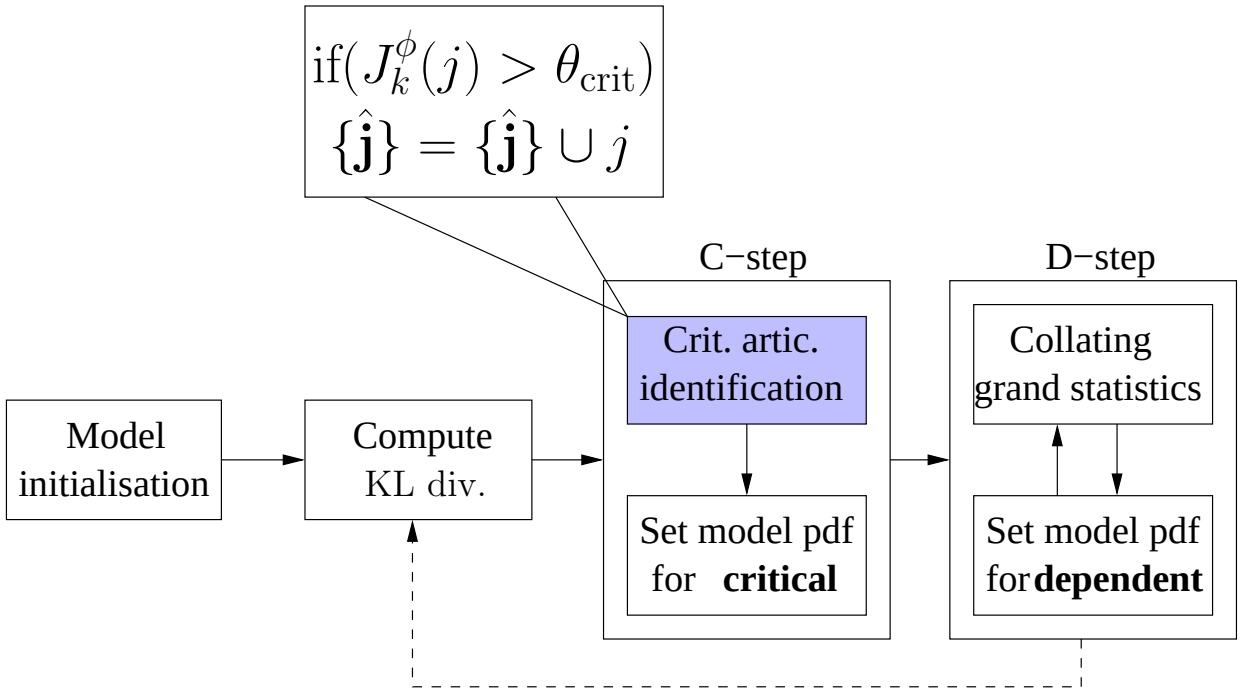
Algorithm



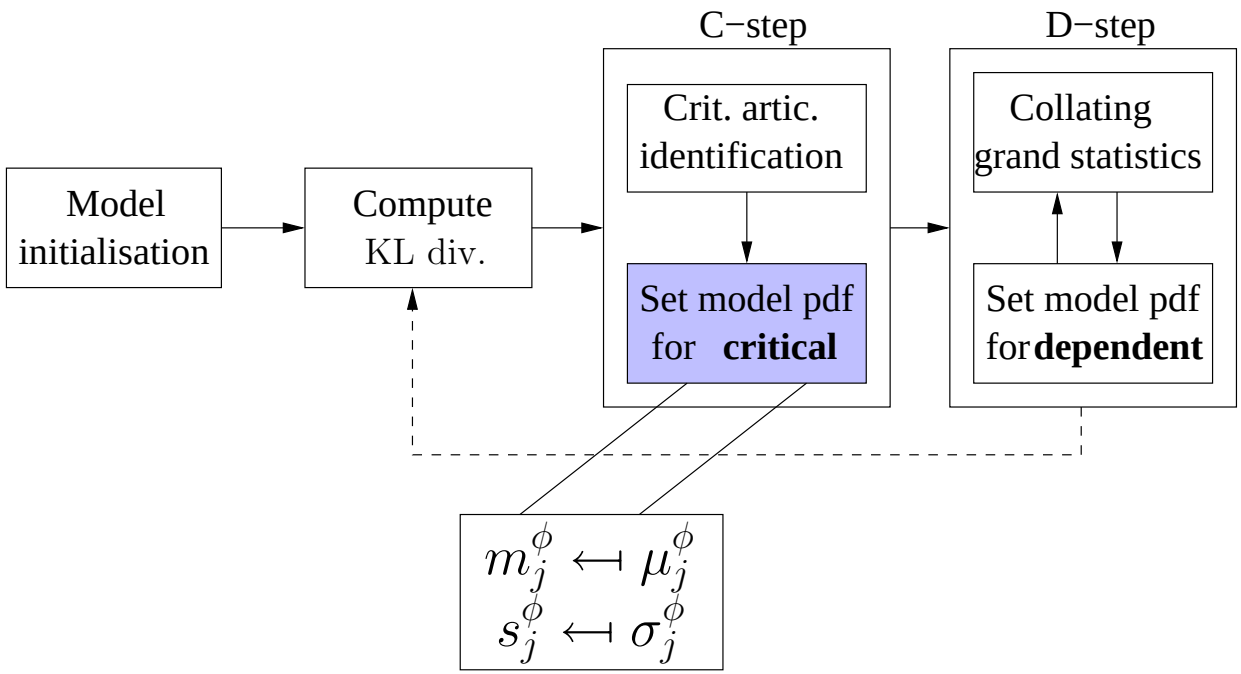
1D model convergence



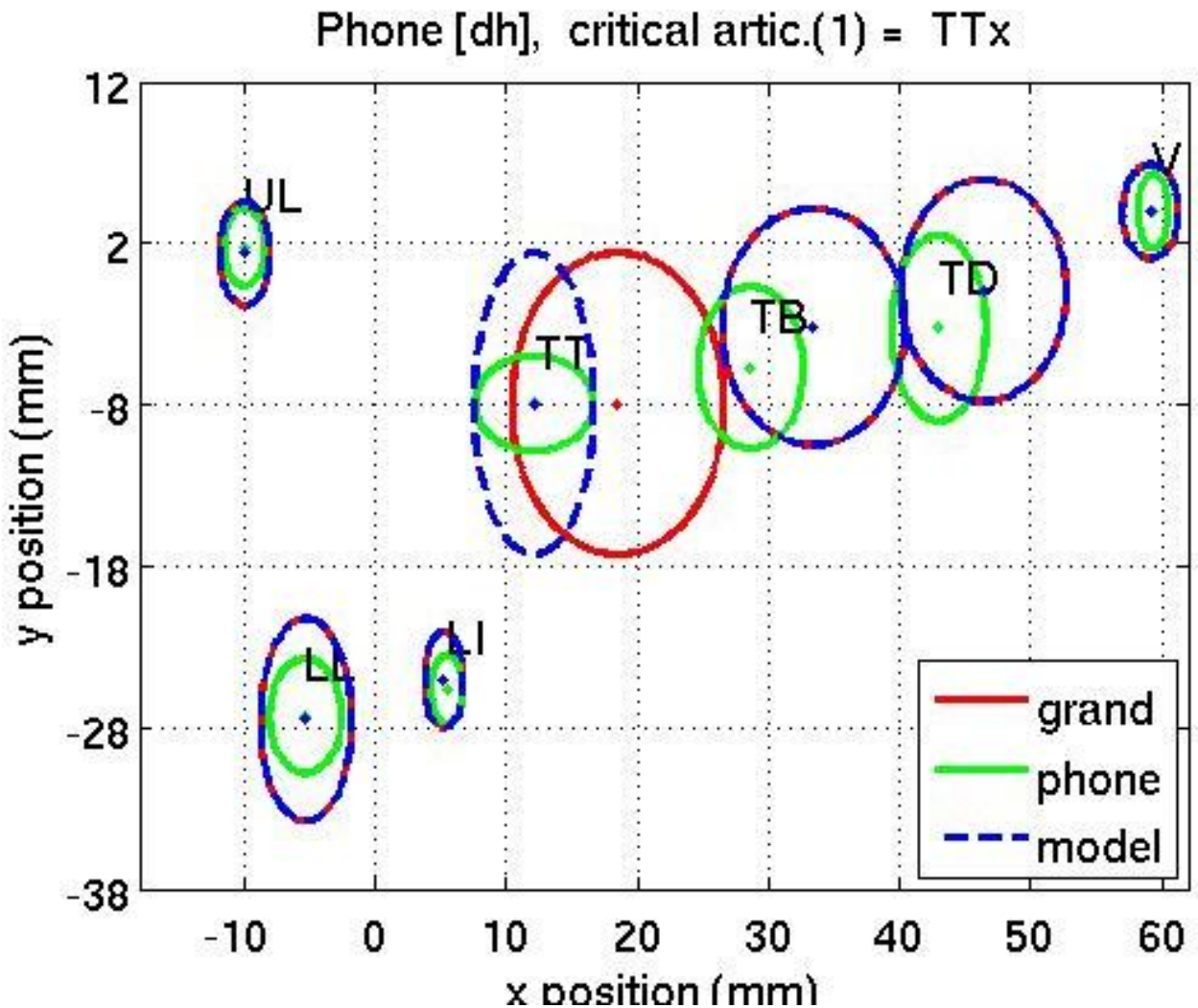
Algorithm



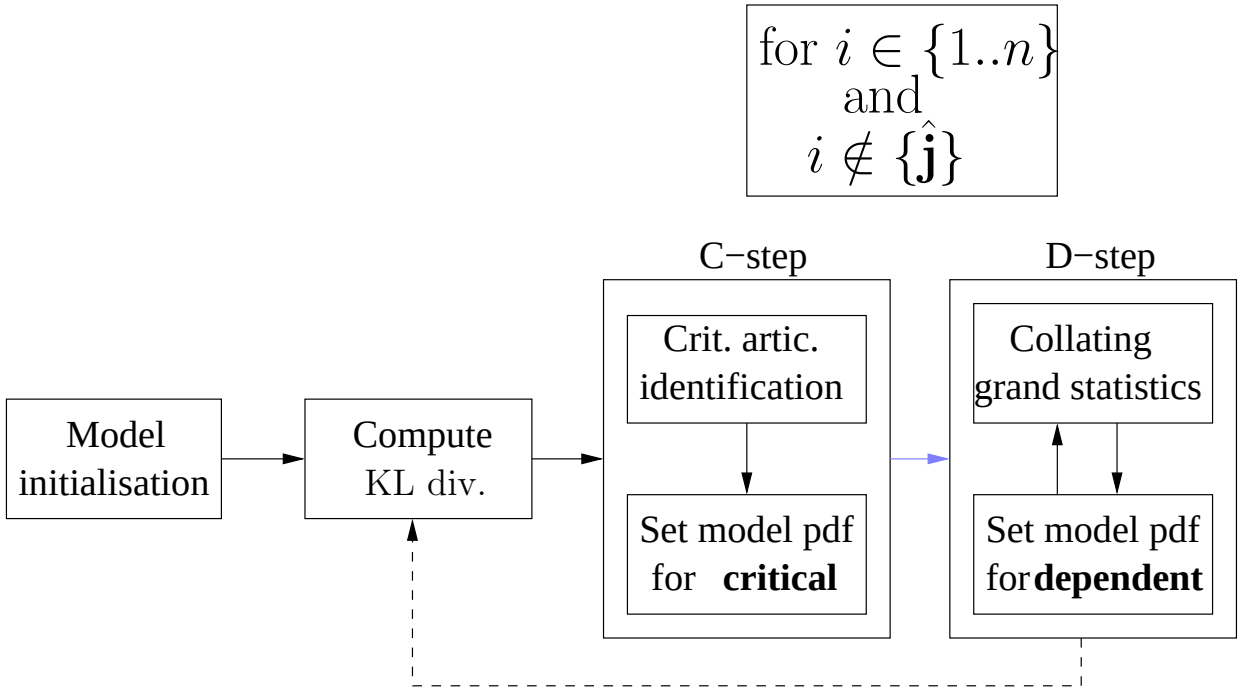
Algorithm



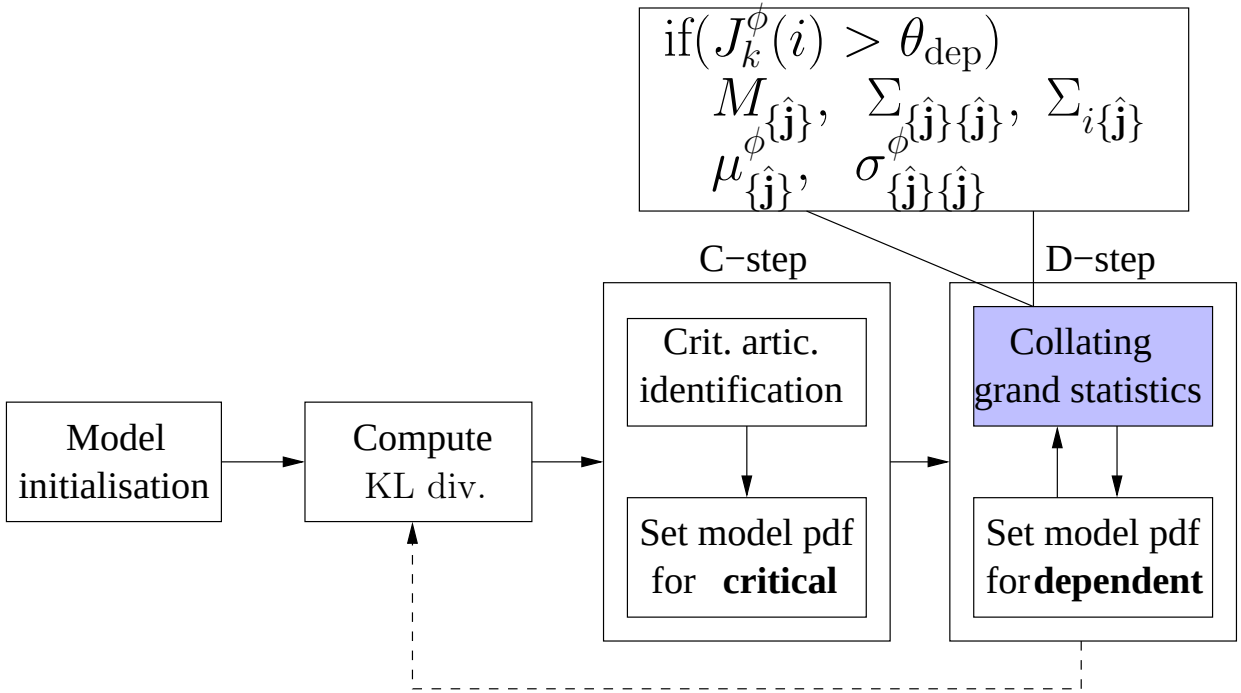
1D model convergence



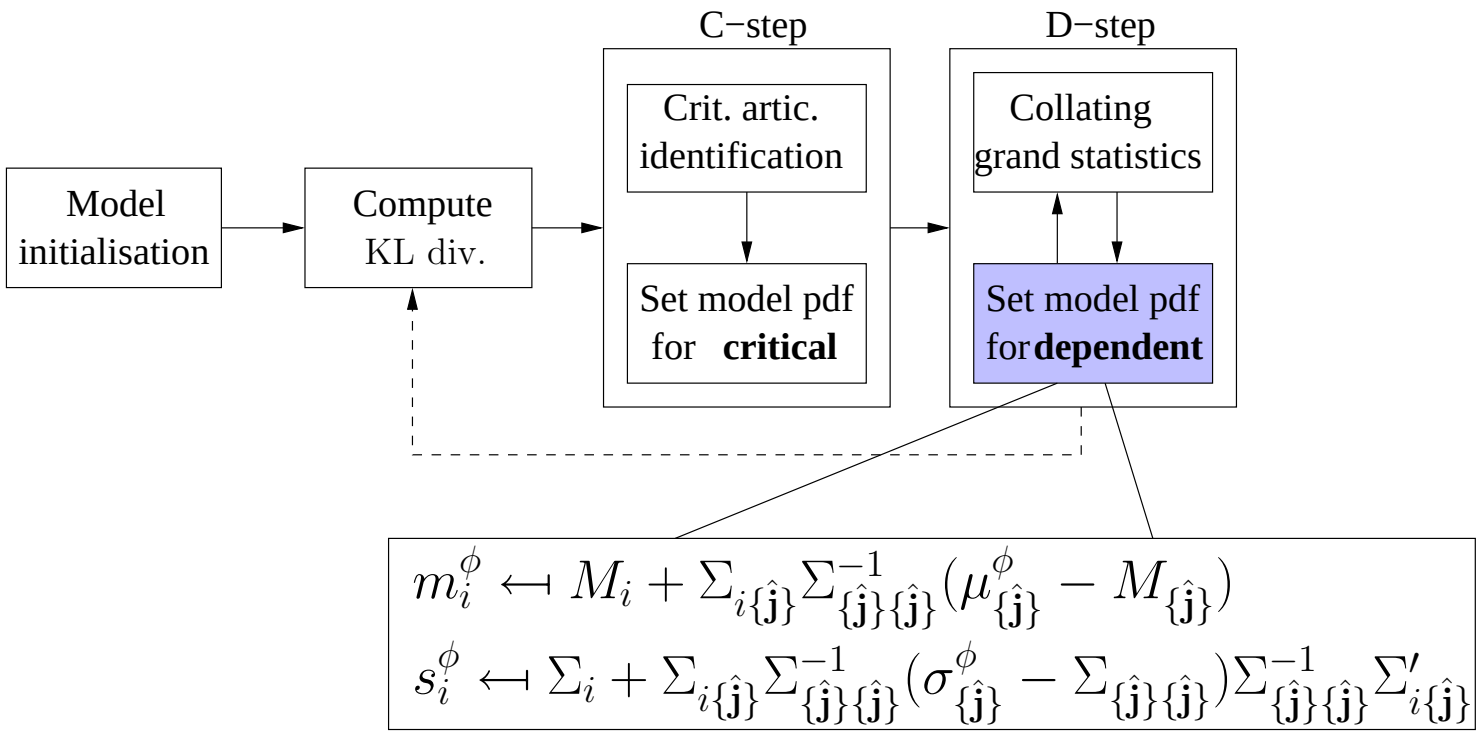
Algorithm



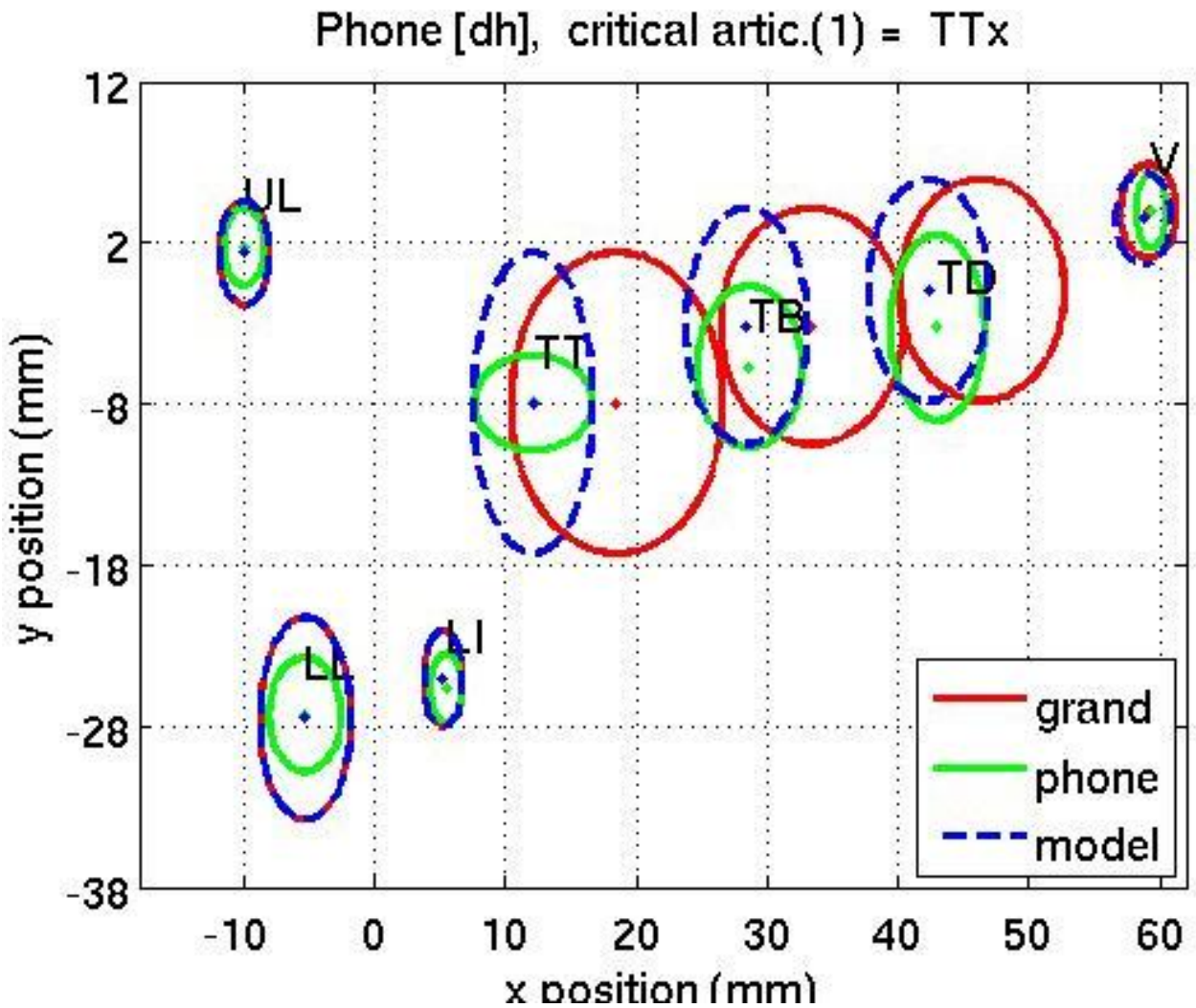
Algorithm



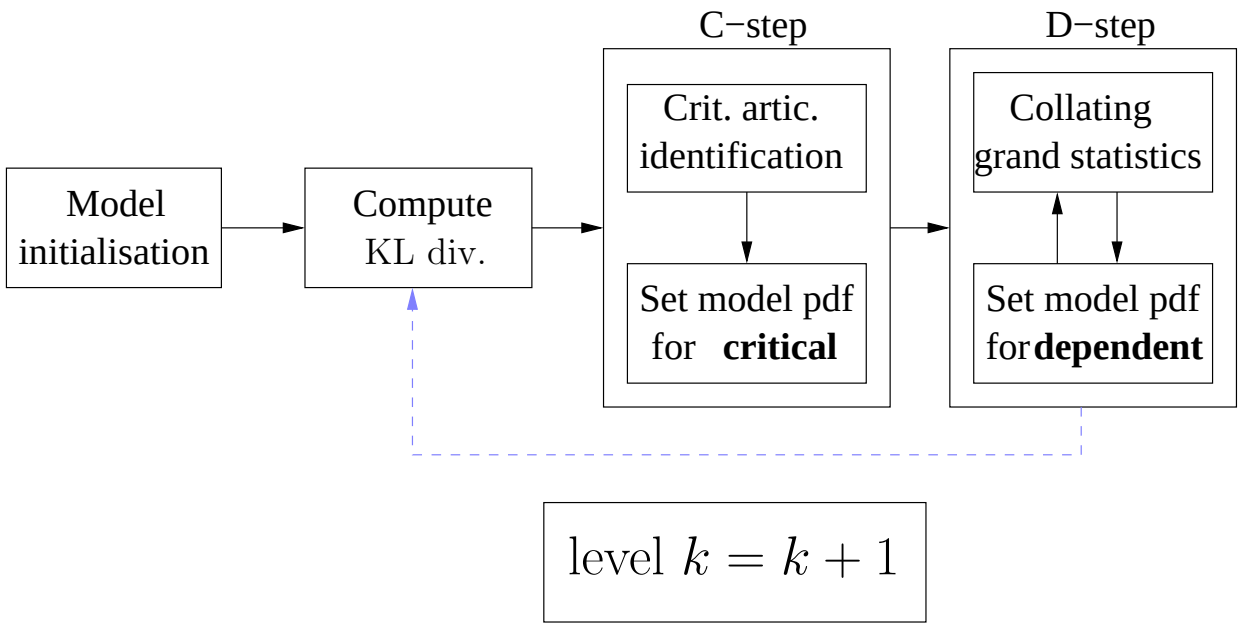
Algorithm



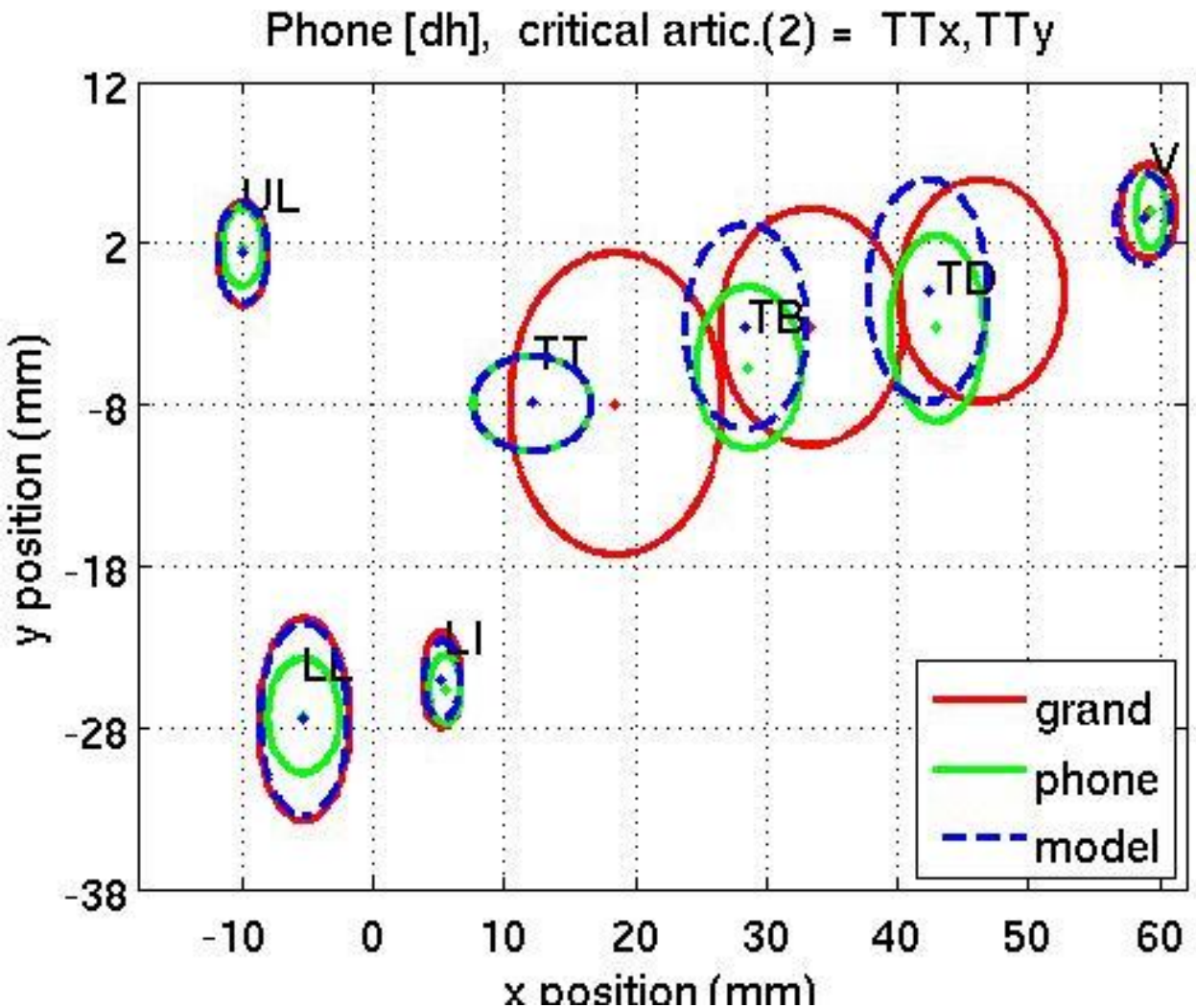
1D model convergence



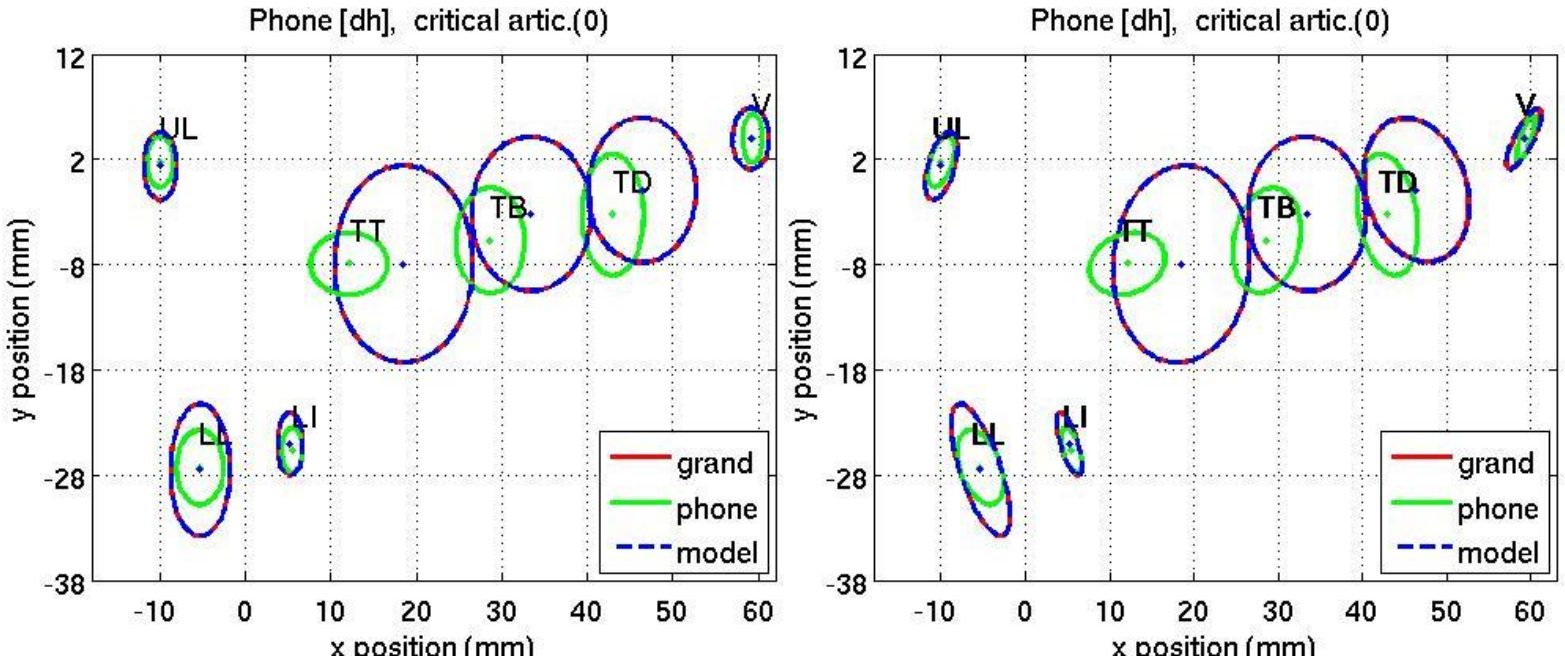
Algorithm


C−step
D−step
Model initialisation
Compute KL div.
Crit. artic. identification
Set model pdf for **critical**
Collating grand statistics
Set model pdf for **dependent**
level $k = k + 1$

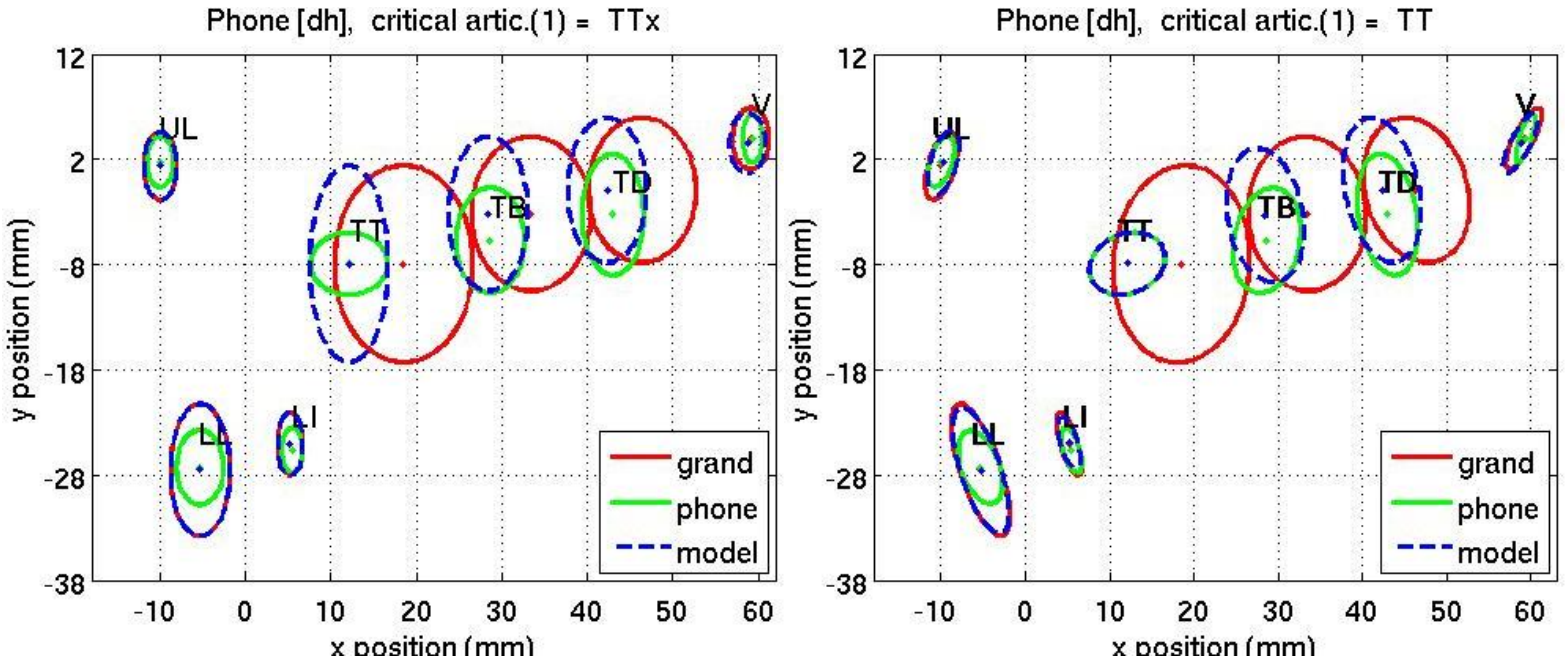
1D model convergence



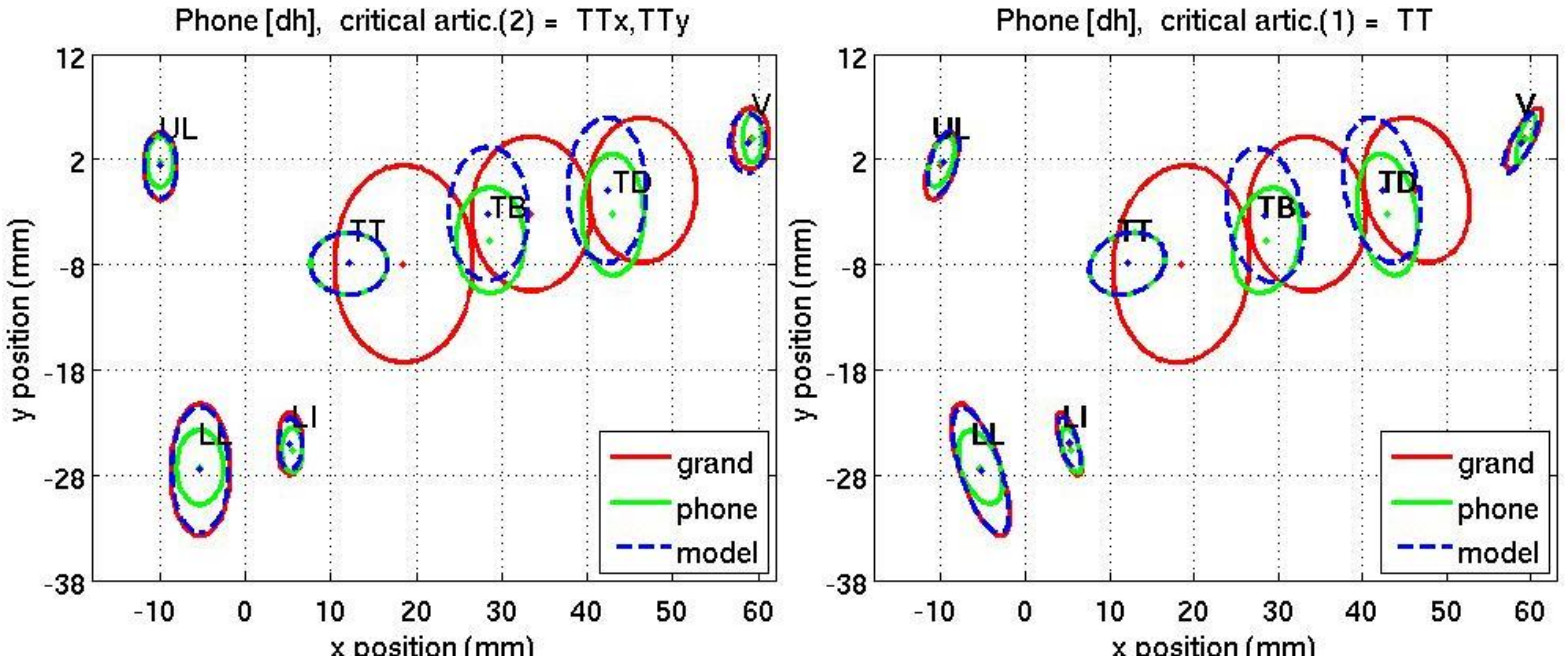
2D model convergence



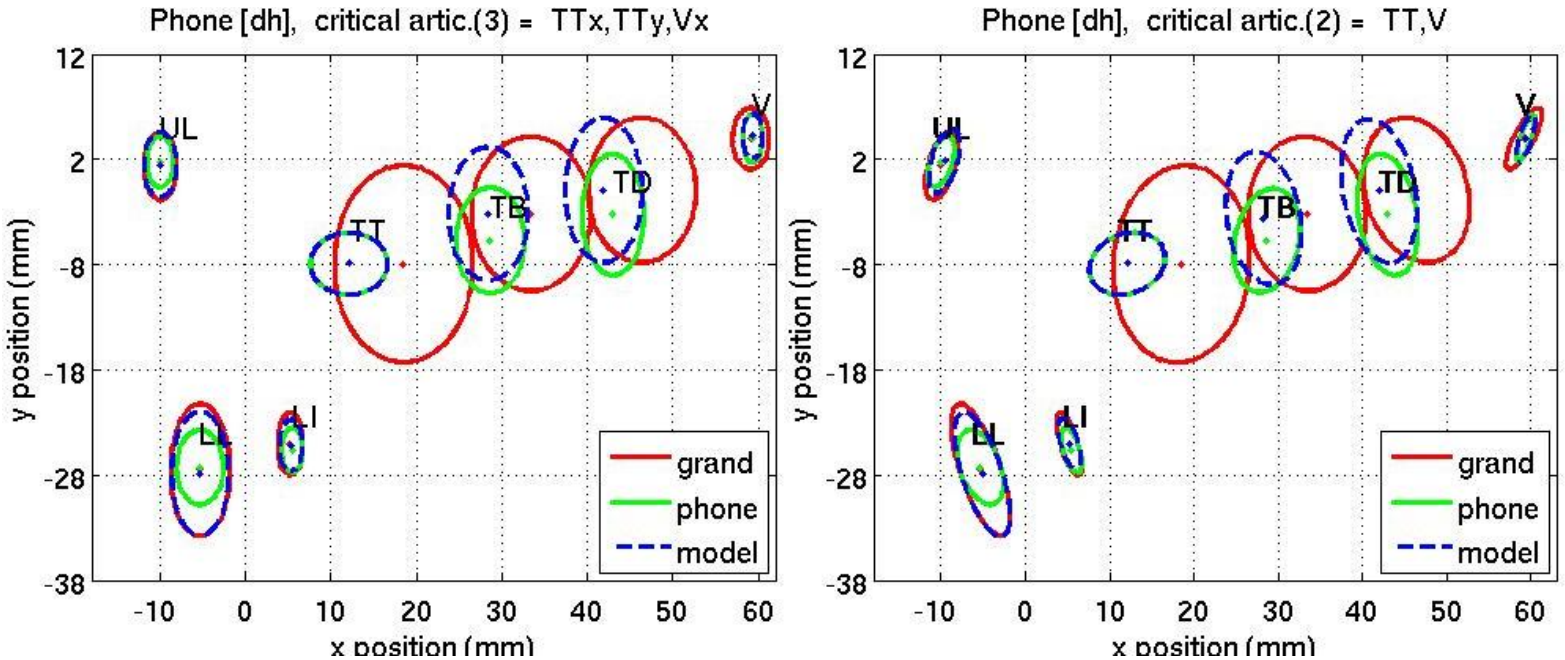
2D model convergence



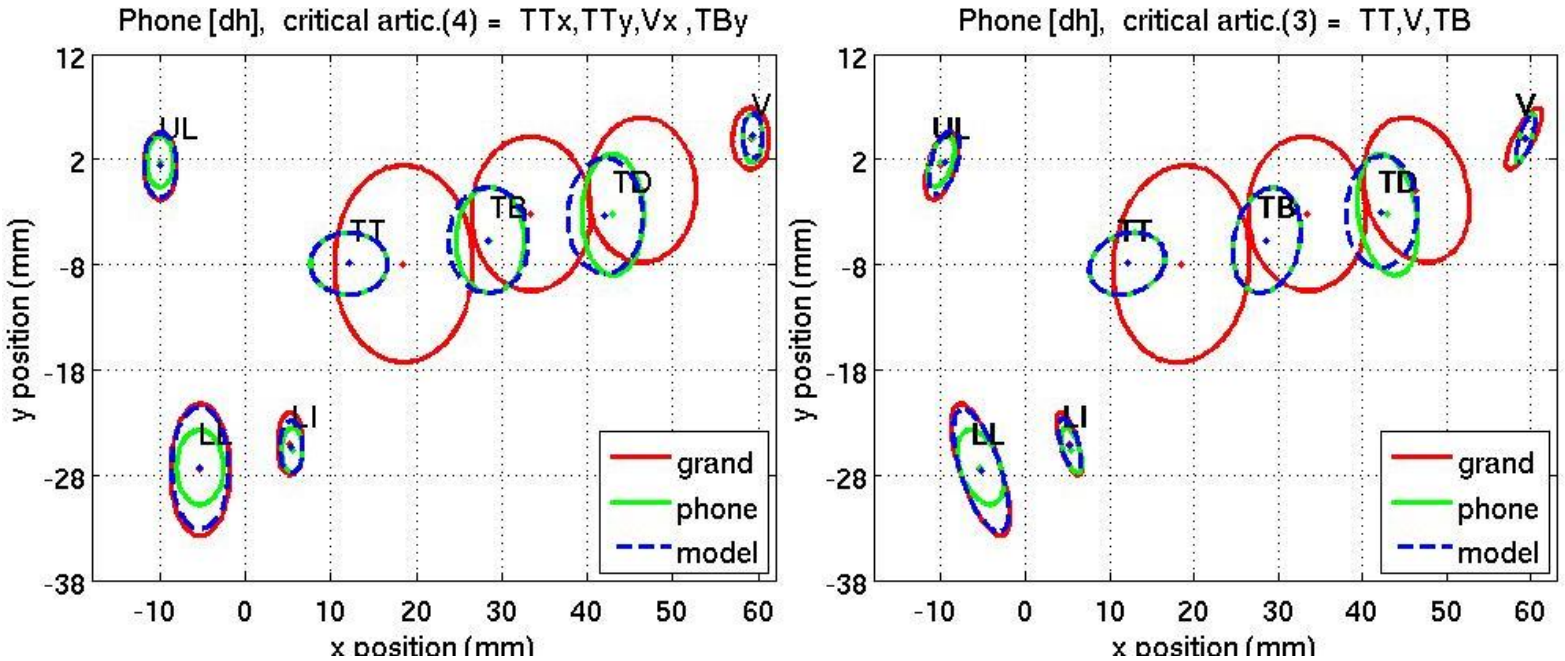
2D model convergence



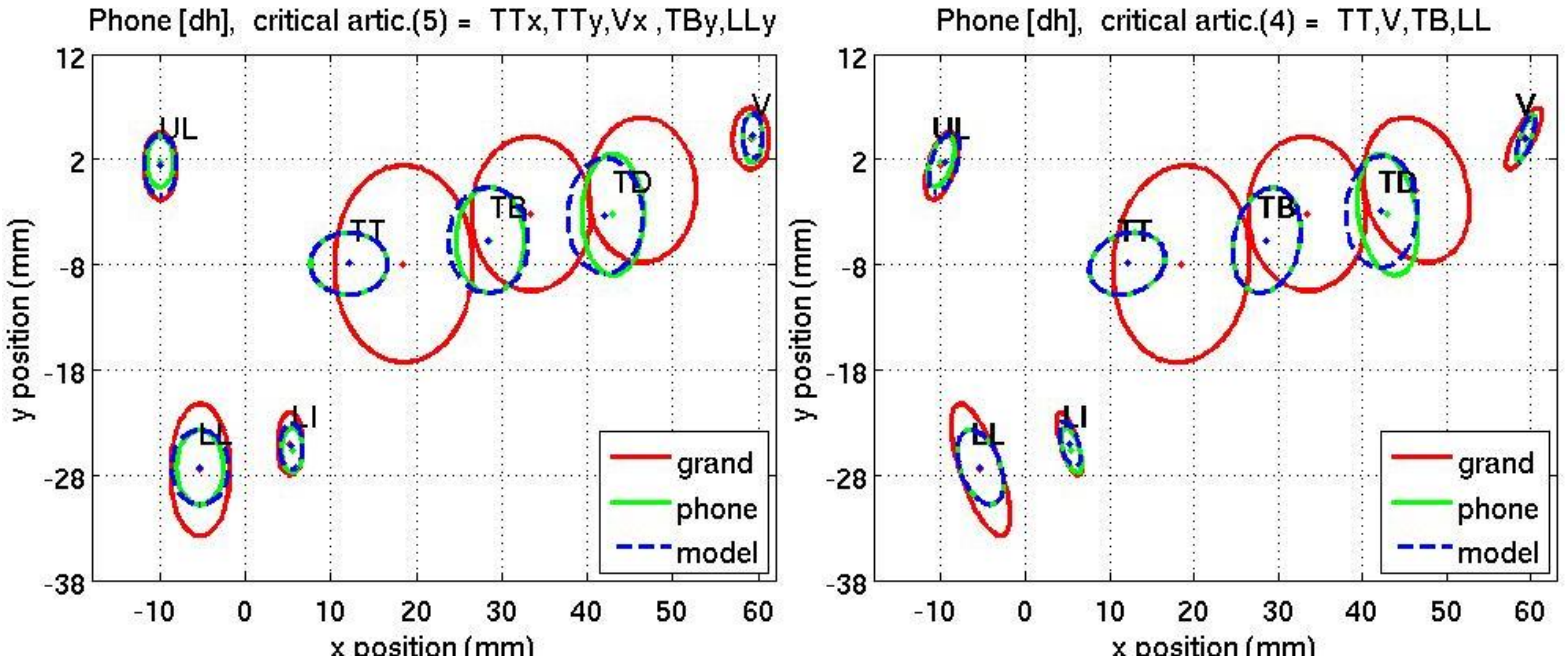
2D model convergence



2D model convergence

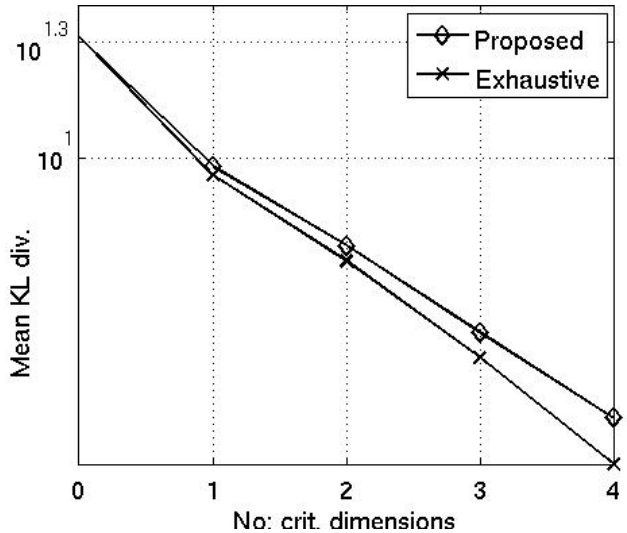


2D model convergence



Exhaustive search

- 1D articulatory data of the male speaker(msak) was used.



- An Exhaustive Search (ES) for the best critical articulator combinations was performed at every level k, $0 < k < 4$.

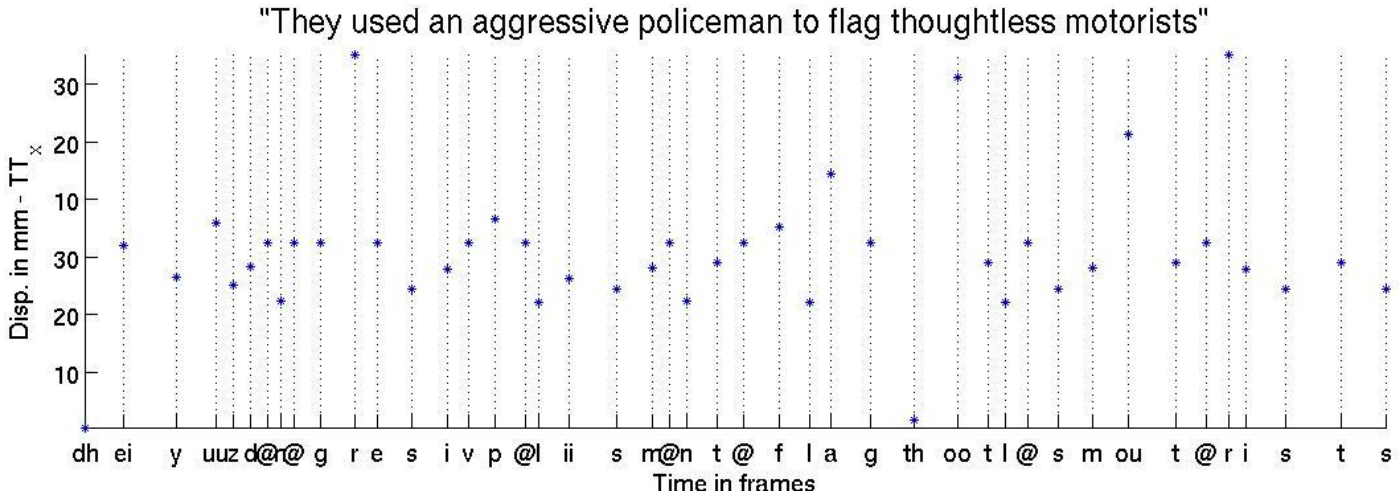
	Proposed	Exhaustive
Time(s)	1×10^2	3×10^5

Generation of Trajectories

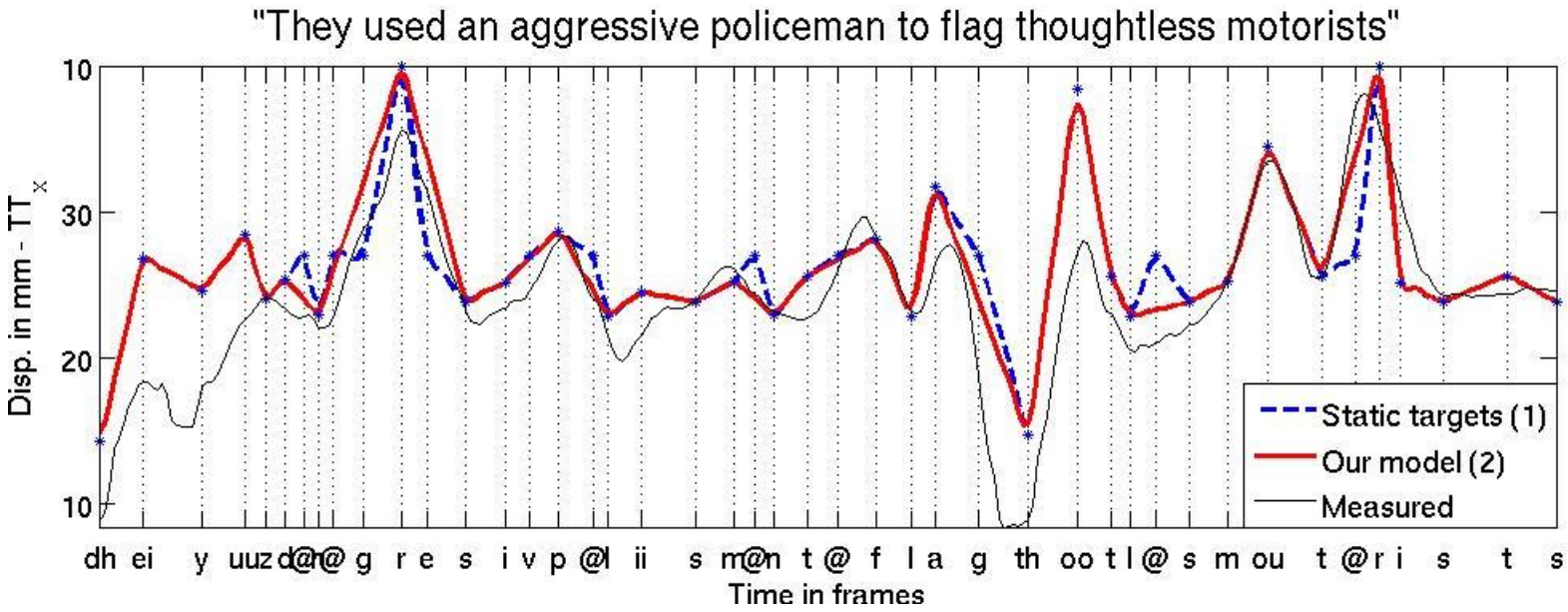
Linear interpolation between

1. static target positions of all articulators.
2. successive target positions of critical and dependent articulators; **no target positions were specified for redundant articulators**.

• Evaluation of errors: Normalised RMSE and correlation between filtered synthetic and original trajectories.



Results



Evaluation of errors

- Improvement obtained using our model over static targets:

Method	Male	Female
Norm. RMSE	2.3%	2.3%
Correlation	2.7%	3.2%

Summary

Conclusions:

- Proposed a compact, data driven algorithm that
  - captures the statistical dependencies
  - identifies the changing articulatory roles
- Preliminary evaluation of the model showed improvement

Future work: Using the knowledge of articulatory roles to

- better model the constraints on articulatory movements
- integrate into our trajectory based recognizer.

http://personal.ee.surrey.ac.uk/Personal/P.Jackson/Dansa/ [2]

[2] Acknowledgements: Funded in the UK by EPSRC (GR/S85511/01).

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.

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

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