



Εθνικό Μετσόβιο Πολυτεχνείο

Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Ομάδα Όρασης Υπολογιστών, Επικοινωνίας Λόγου & Επεξεργασίας Σήματος

<http://cvsp.cs.ntua.gr>

Αναγνώριση και Μοντελοποίηση Νοηματικής Γλώσσας με την Χρήση Οπτικής Επεξεργασίας και Στατιστικών Μεθόδων

Παρουσίαση Διδακτορικής Διατριβής

Θεοδωράκης Σταύρος

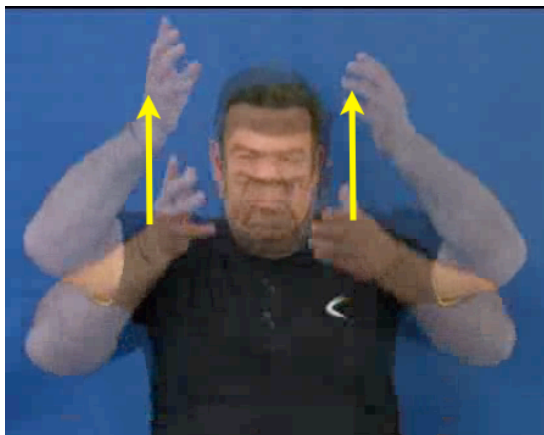
Ηλεκτρολόγος Μηχανικός & Μηχανικός Η/Υ,

Υποψήφιος Διδάκτορας ΕΜΠ

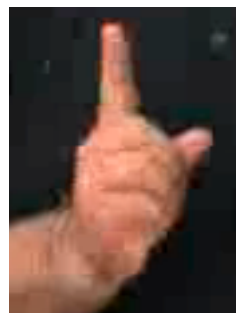
17 Ιουνίου 2014



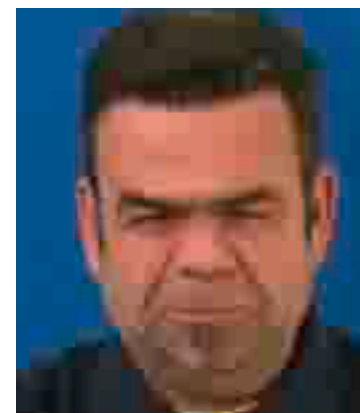
Νοηματική Γλώσσα



κίνηση χεριών



χειρομορφή



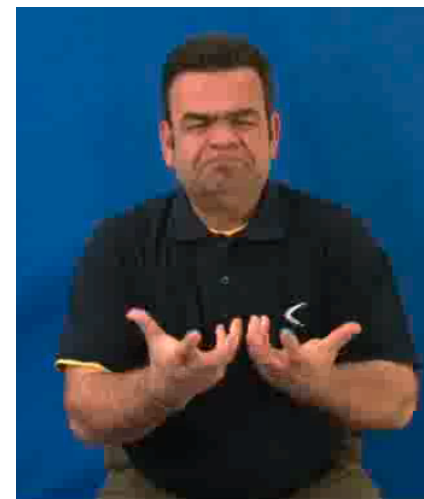
εκφράσεις προσώπου



θέση χεριών

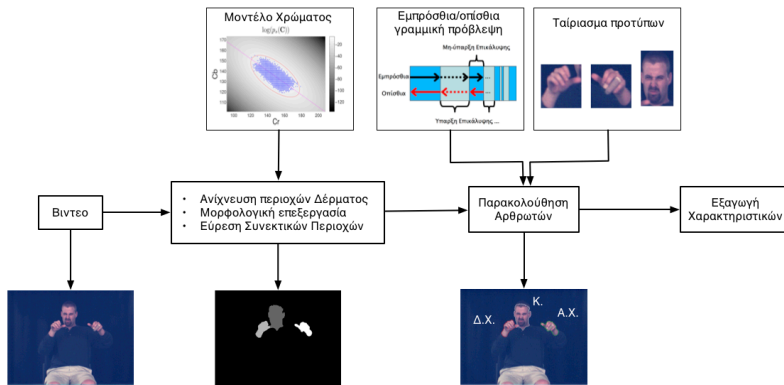


Προσανατολισμός
της παλάμης

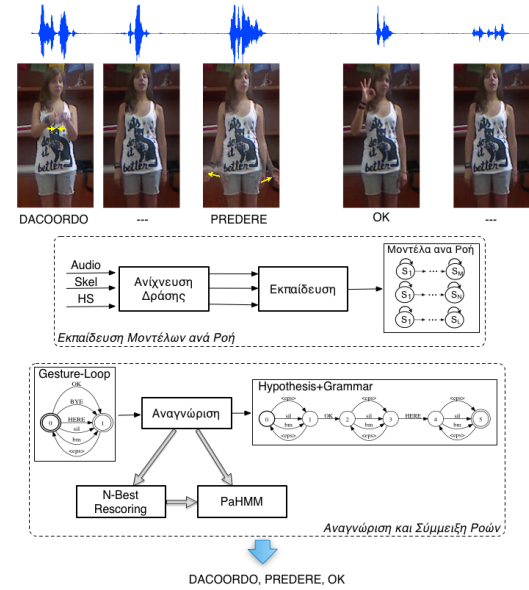


στάση σώματος

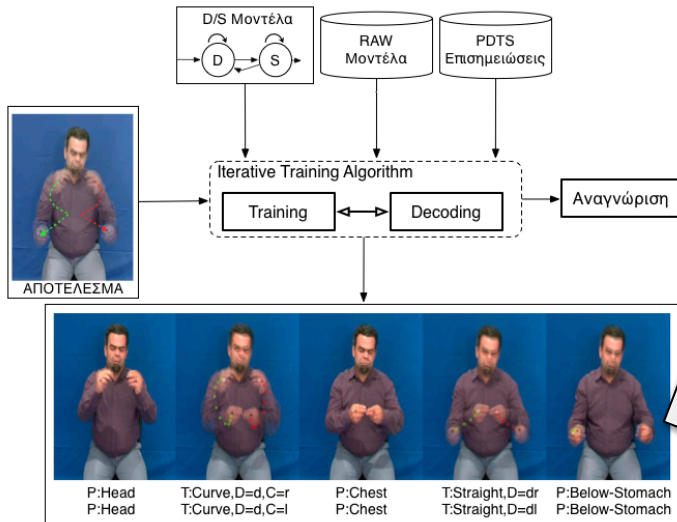
Εξαγωγή Χαρακτηριστικών



Πολυτροπική Αναγνώριση Χειρονομιών

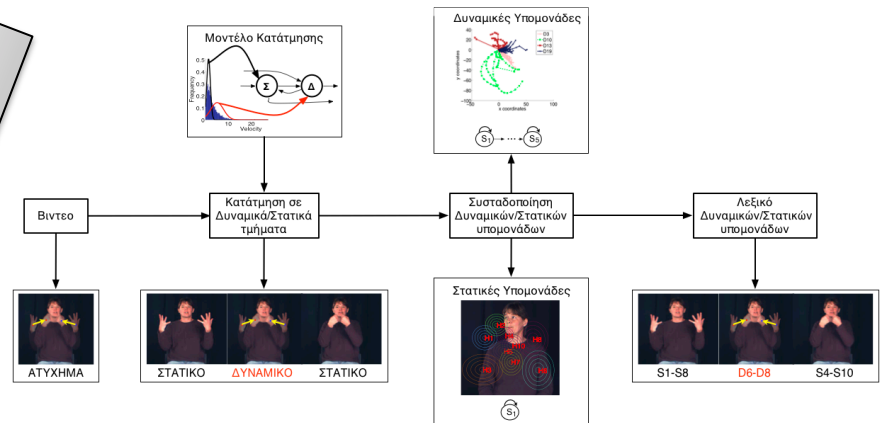


Φωνητικές Υπομονάδες



Αναγνώριση ΝΓ με Υπομονάδες

Δεδομενοκεντρικές Υπομονάδες



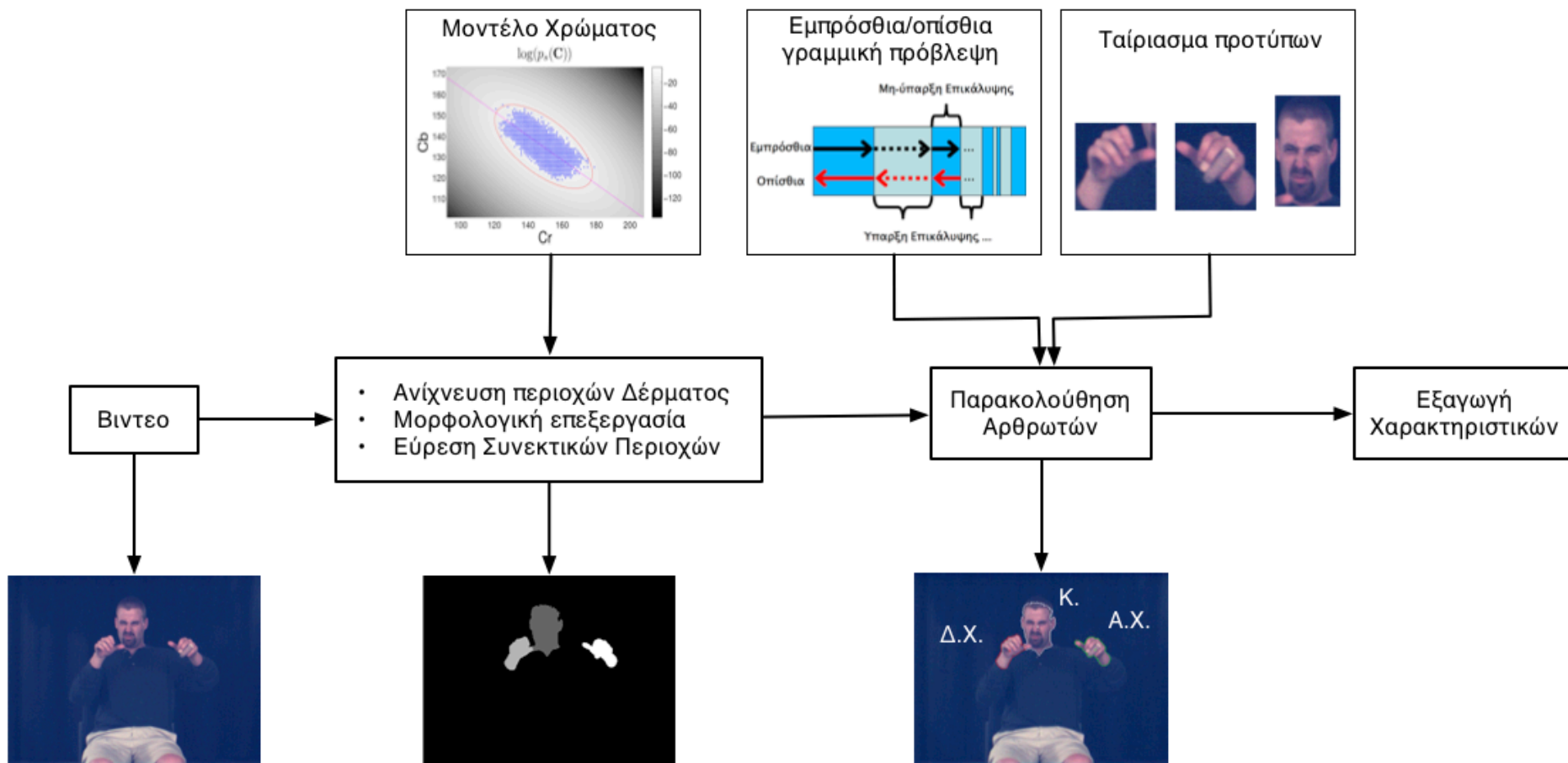


Εξαγωγή Χαρακτηριστικών

Ref: Rousos, Theodorakis, Pitsikalis and Maragos, ECCV SGA 2010

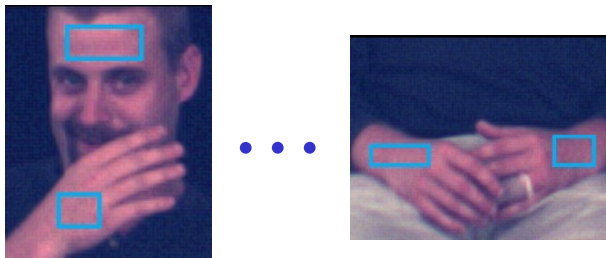
Ref: Rousos, Theodorakis, Pitsikalis and Maragos, JMLR 2013

Σύστημα Εξαγωγής Χαρακτηριστικών από βίντεο Νοηματισμού

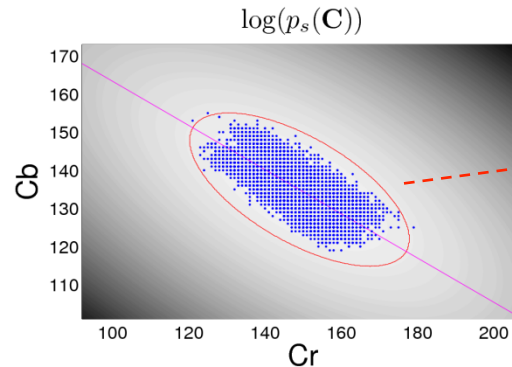


Ανίχνευση Χεριών και Κεφαλιού

■ Μοντελοποίηση Χρώμα δέρματος



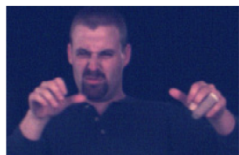
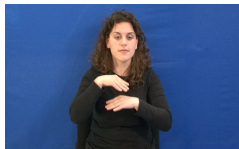
Δεδομένα Εκπαίδευσης



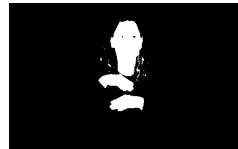
Γκαουσιανή Κατανομή

Κατωφλιοποίηση για την μάσκα δέρματος S_0

■ Μορφολογική Επεξεργασία μασκών δέρματος



Εικόνα



Μάσκα δέρματος S_0



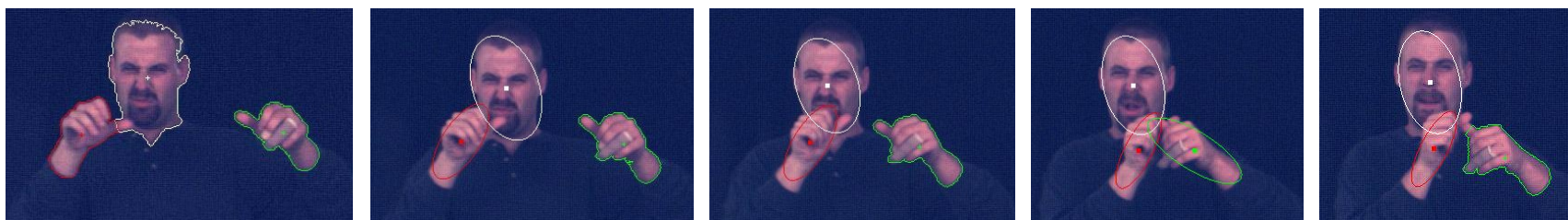
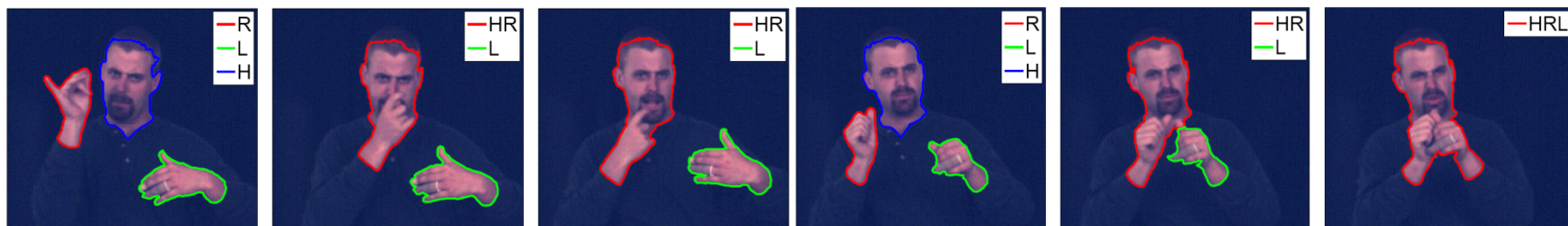
refinement of S_0
generalized hole filling
- area opening



Κατάτμηση
- Συνεκτικές περιοχές
- competitive rec. opening

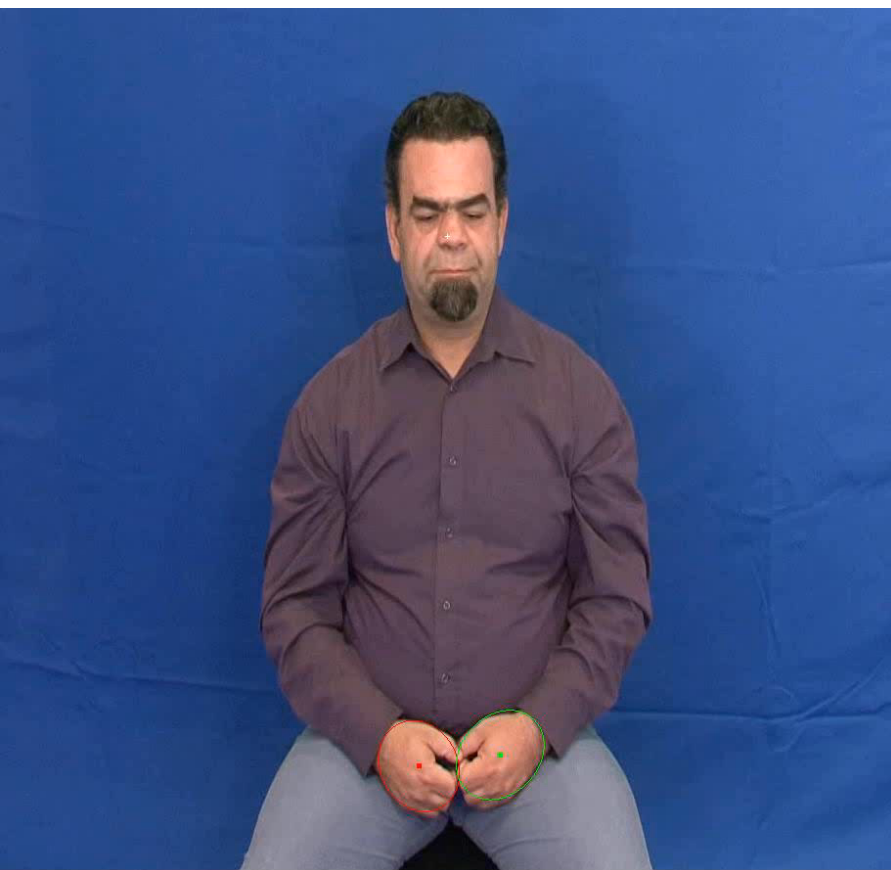
Παρακολούθηση Χεριών και Κεφαλιού

- Παρακολούθηση Αρθρωτών αποτελείται:
 - Εμπρόσθια και Οπίσθια Γραμμική πρόβλεψη,
 - Ταίριασμα προτύπου,
 - Ταίριασμα έλλειψης,
 - Πιθανοτικοί περιορισμοί
- Αποτέλεσμα: Ένα σύνολο από περιοχές δέρματος ακολουθούμενο από ετικέτες που υποδεικνύουν του αρθρωτές που περιλαμβάνουν



Παρακολούθηση Χεριών και Κεφαλιού Video

Dicta-Sign Lemmas



BU-400



Μοντελοποίηση Χειρομορφών

- Αναπαράσταση Σχήματος-Εμφάνισης



$$f(\mathbf{x}) = \begin{cases} g(I(\mathbf{x})), & \text{if } \mathbf{x} \in M \\ -c_b & \text{else} \end{cases}$$

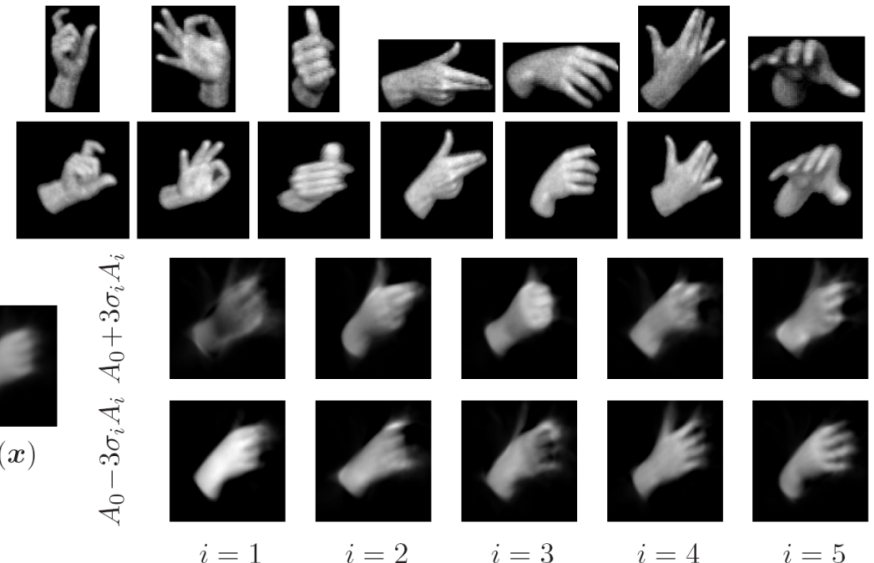
- Γεννητικό Μοντέλο

$$f(W_p(\mathbf{x})) \approx A_0(\mathbf{x}) + \sum_{i=1}^{N_c} \lambda_i A_i(\mathbf{x})$$

$W_p(\mathbf{x})$: 2D αφινικός μετασχηματισμός
Με παραμέτρους $\mathbf{p} \in \mathbb{R}^6$

- Εκπαίδευση Μοντέλου

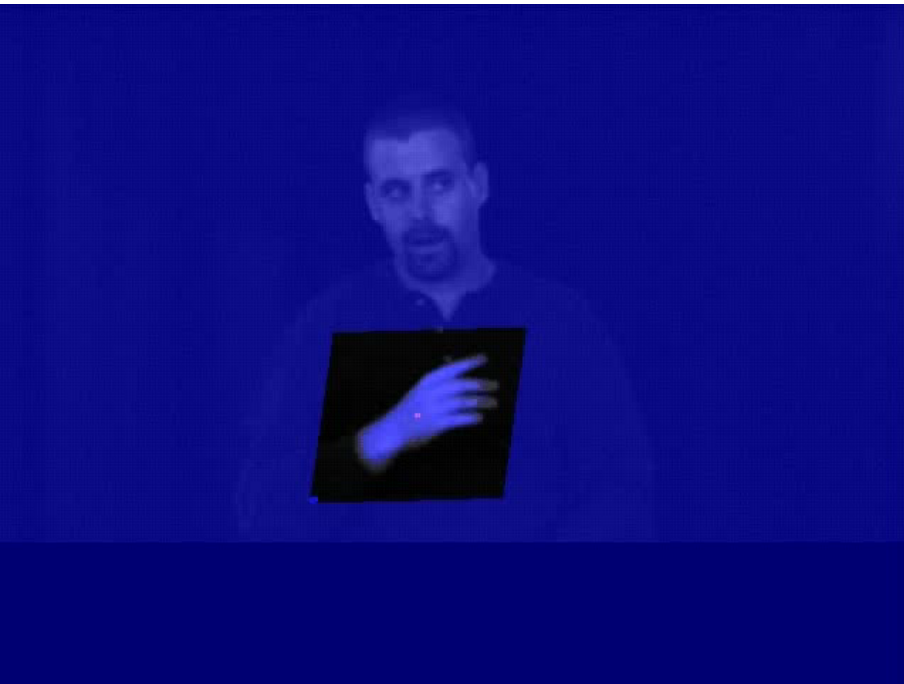
- Affine alignment of the training set
 - generalization of the procrustes analysis
 - iterative manual feedback
- PCA to learn $A_i(\mathbf{x})$
 - keep only $N_c=35$ components



- Ομαλοποιημένο Ταίριασμα

$$E(\boldsymbol{\lambda}, \mathbf{p}) = E_{rec}(\boldsymbol{\lambda}, \mathbf{p}) + w_S E_S(\boldsymbol{\lambda}, \mathbf{p}) + w_D E_D(\boldsymbol{\lambda}, \mathbf{p})$$

Video : Ταίριασμα Μοντέλου Σχήματος-Εμφάνιση



BU-400

Dicta-Sign Lemmas

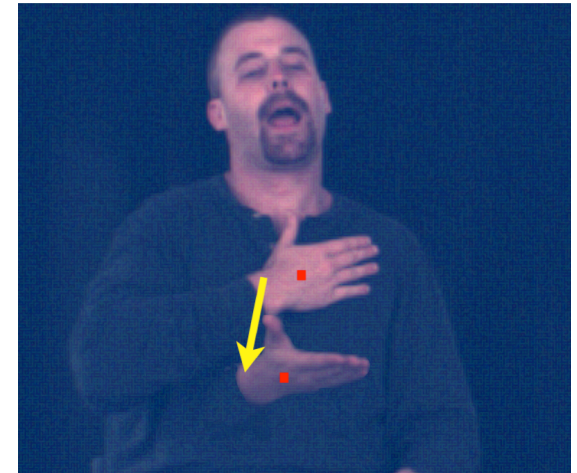




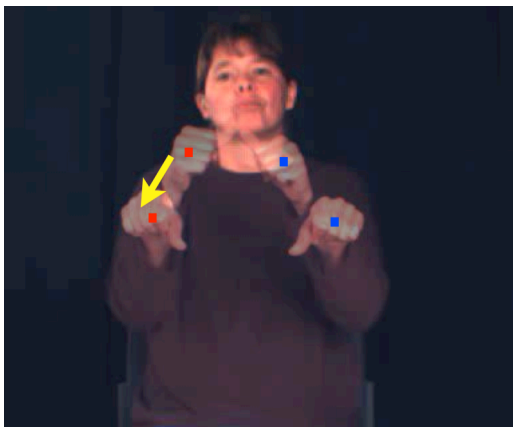
Εισαγωγή στις Υπομονάδες

Εισαγωγή στην έννοια των υπομονάδων (1)

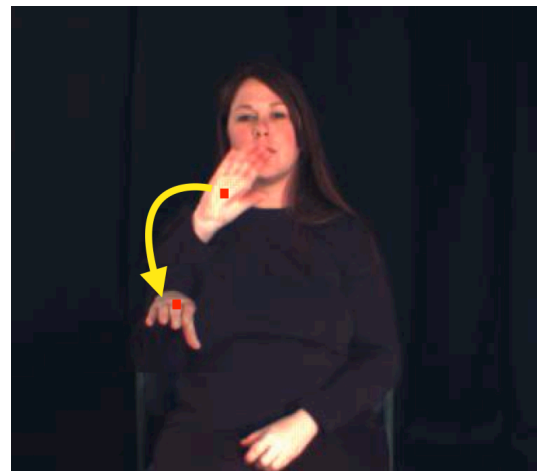
- Υπομονάδες στη Νοηματική
 - Η έννοια του φωνήματος/υπομονάδας στην νοηματική δεν είναι δεδομένη ούτε πλήρως καθορισμένη
 - Παράλληλα κατα την άρθρωση
- Stokoe παράμετροι
 - θέση των χεριών (TAB),
 - είδος της κίνησης τους (SIGS),
 - είδος της χειρομορφής (DEZ)



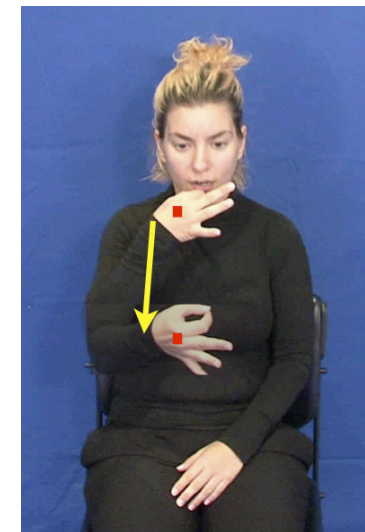
ΠΑΡΑΔΕΧΟΜΑΙ



ΚΑΤΑΘΕΤΩ



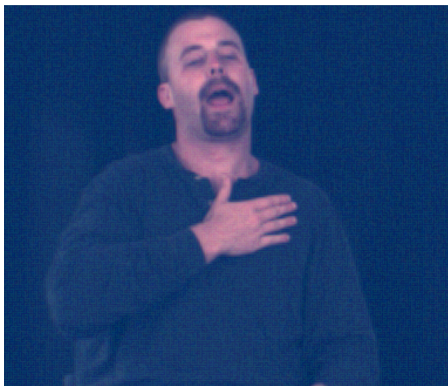
ΣΙΚΑΓΟ



ΛΕΩ

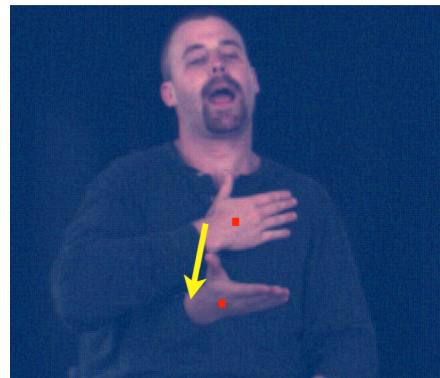
Εισαγωγή στην έννοια των υπομονάδων (2)

- Liddell και Johnson Movement-Hold
 - Κίνηση (Movement -M-), ύπαρξη μεταβολής
 - Στάση (Hold -H-), μη-ύπαρξη μεταβολής
- Liddell και Johnson PDTS
 - Στάσεις (Postures), δηλαδή η μη-ύπαρξη μεταβολής
 - Μικρές Στάσεις (Detentions) μικρής χρονικής διάρκειας
 - Κινήσεις (Transitions), δηλαδή η ύπαρξη μεταβολής
 - Σταθερής-Αλλαγής (Steady-Shift) αντίστοιχες με τις κινήσεις



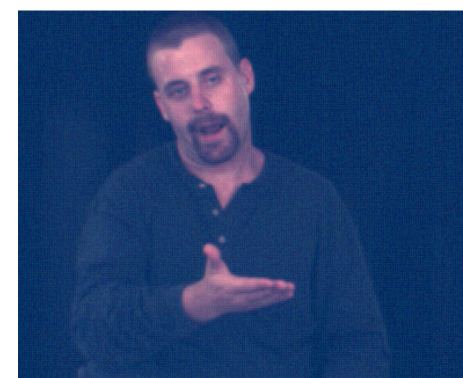
HOLD

POSTURE



MOVEMENT

TRANSITION



HOLD

POSTURE

Movement-Hold:

PDTS:

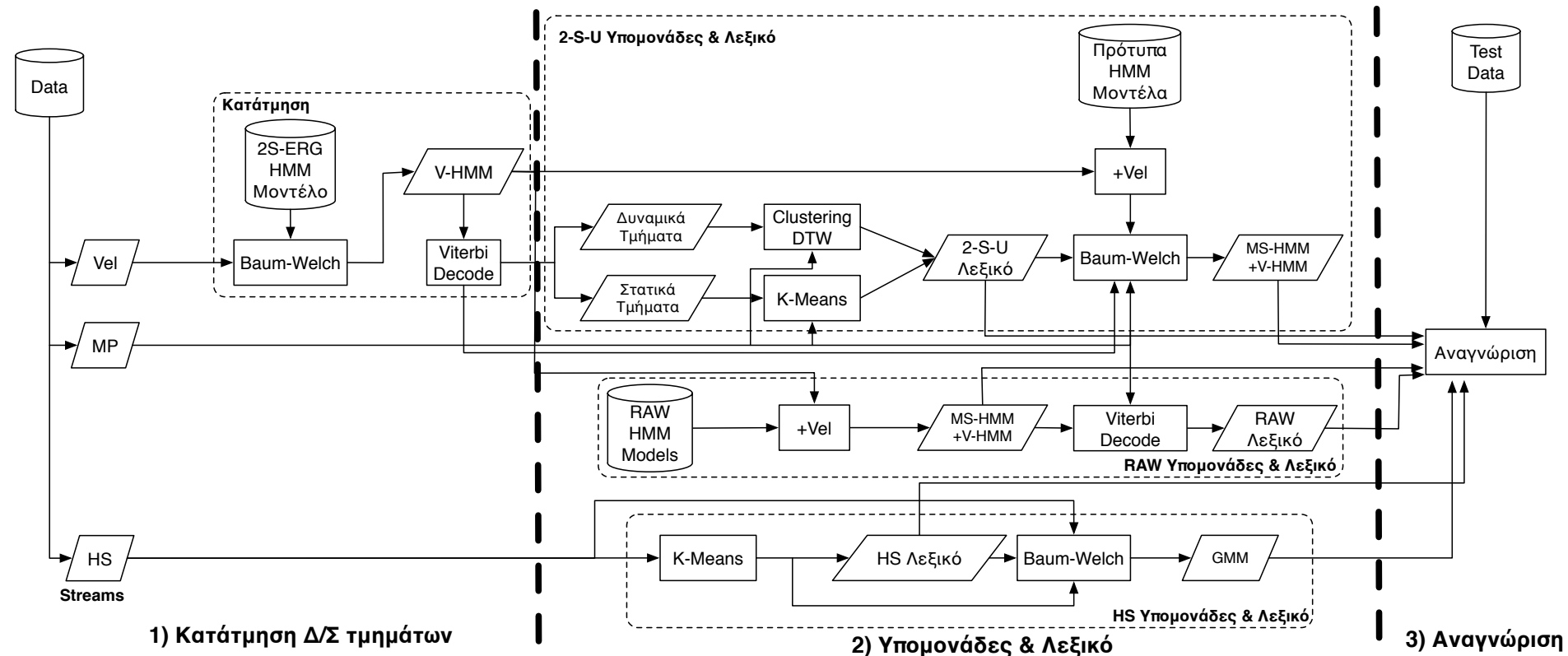


Δεδομενοκεντρικές Υπομονάδες

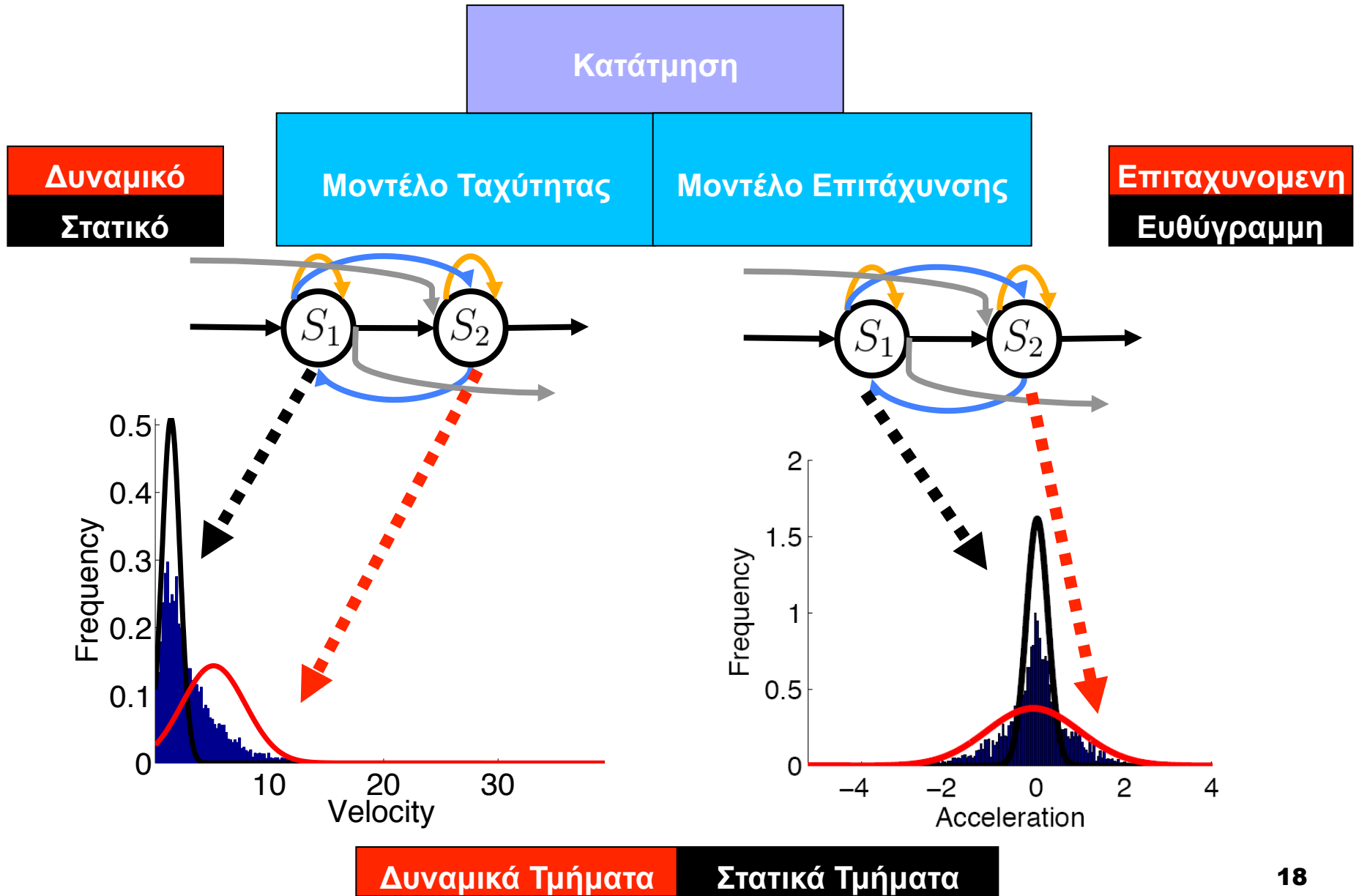
Ref: Theodorakis, Pitsikalis and Maragos, IVC 2014

Ref: Theodorakis, Pitsikalis and Maragos, ICIP 2012

Σύνοψη συστήματος Αναγνώρισης ΝΓ με δεδομοκεντρικές υπομονάδες



Αυτόματη Κατάτμηση (1)



Αυτόματη Κατάτμηση (2)

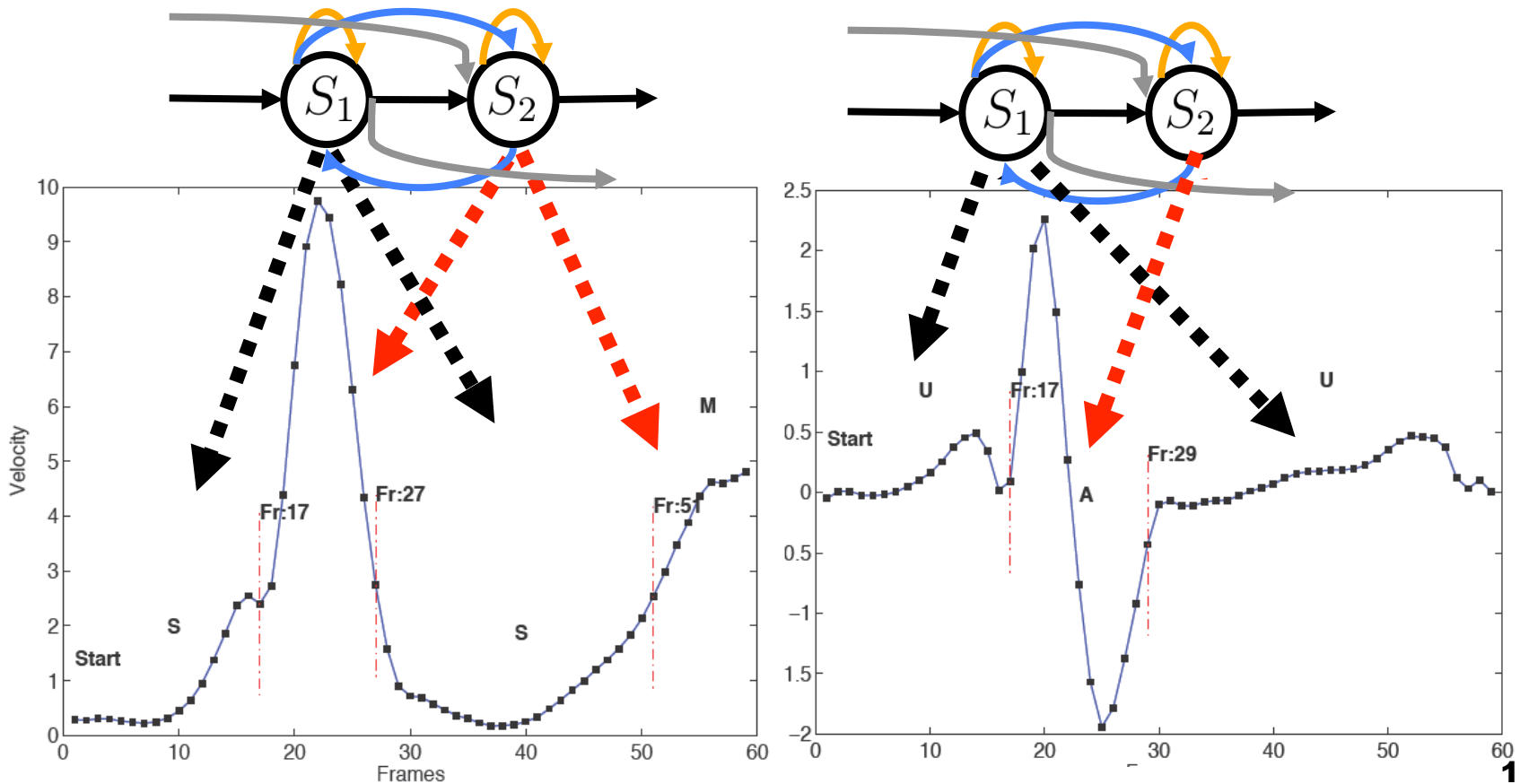
Κατάτμηση

Δυναμικό
Στατικό

Μοντέλο Ταχύτητας

Μοντέλο Επιτάχυνσης

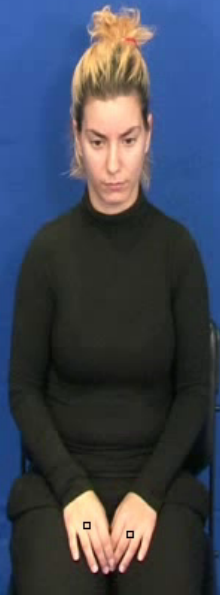
Επιταχυνόμενη
Ευθύγραμμη



Παράδειγματα Κατάτμησης

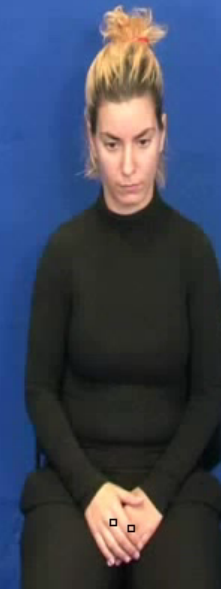
ΡΥΖΙ

Static



ΑΤΜΟΣ

Static



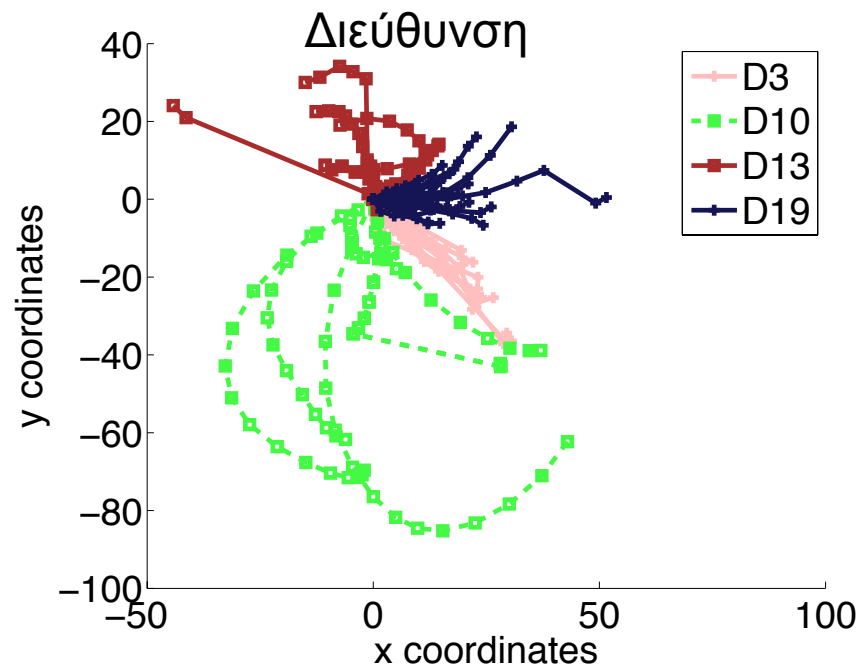
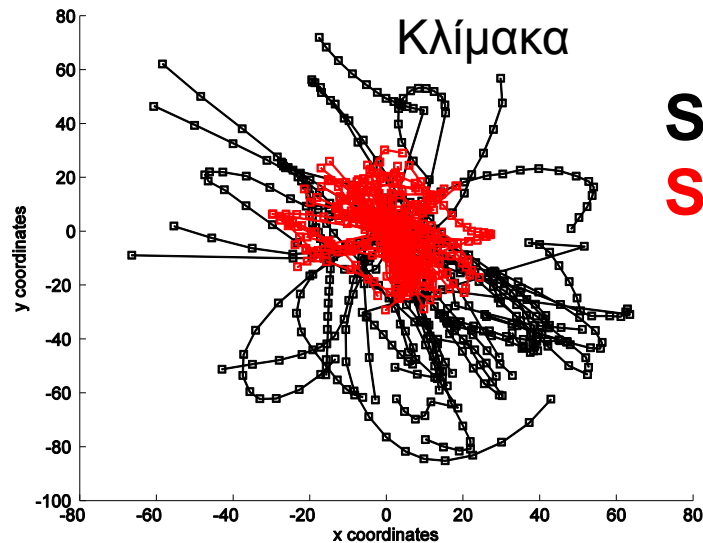
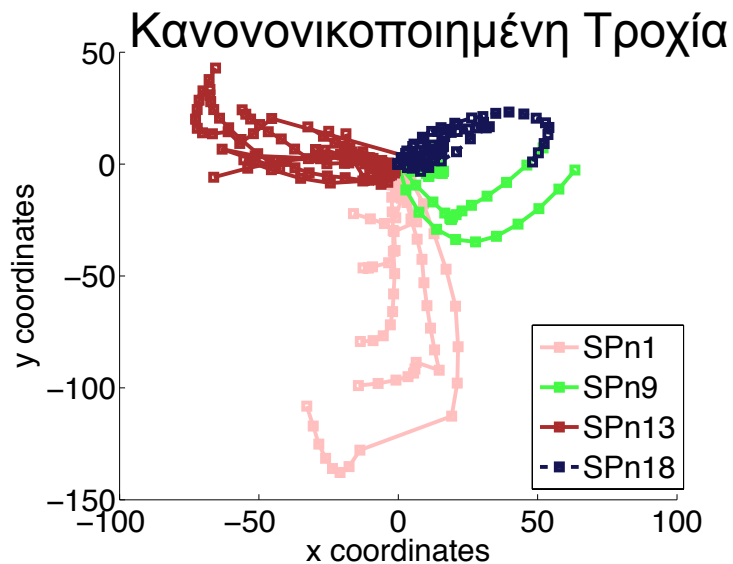
Δυναμικά (Dynamic) Τμήματα

VS.

Στατικά (Static) Τμήματα

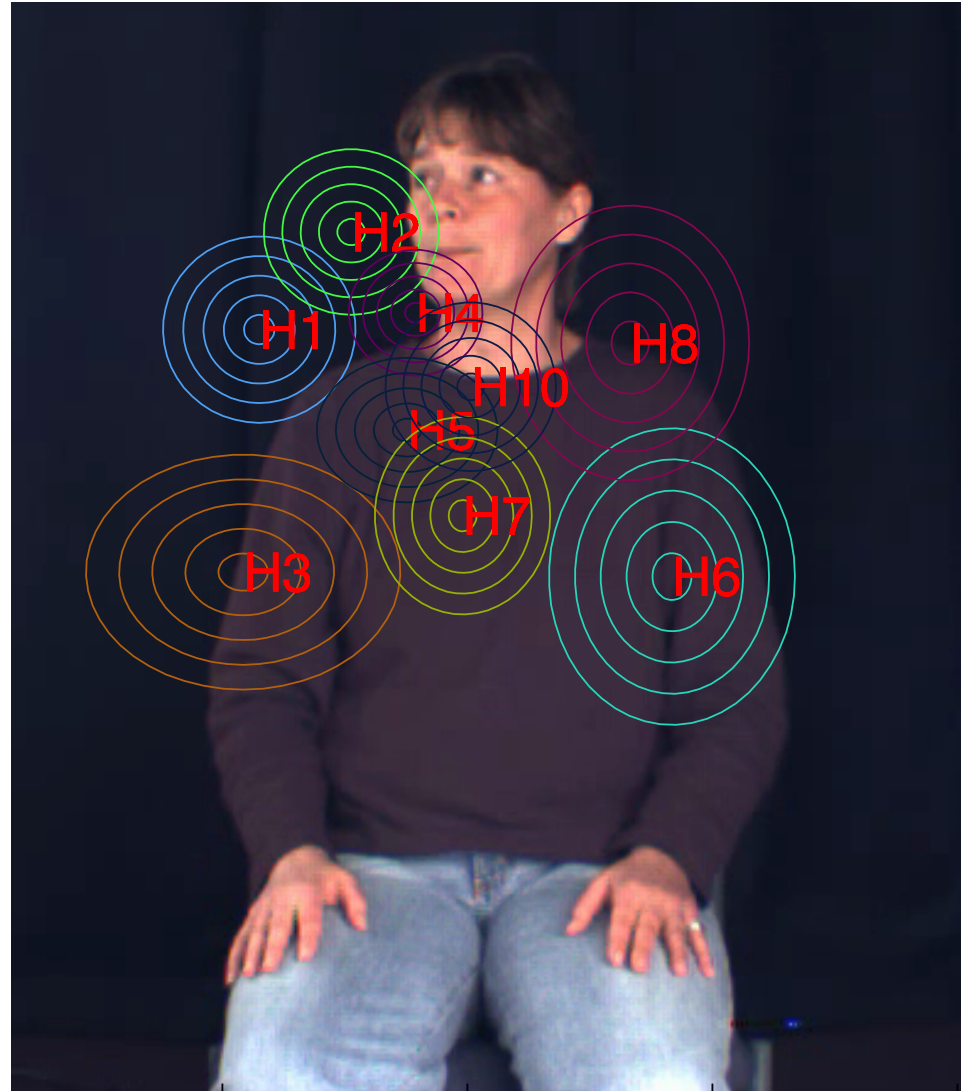
2-S-U Δυναμικές Υπομονάδες

- Κανονικοποιημένη Τροχιά, Διεύθυνση, κλίμακα
- Ιεραρχική συσταδοποίηση
 - DTW

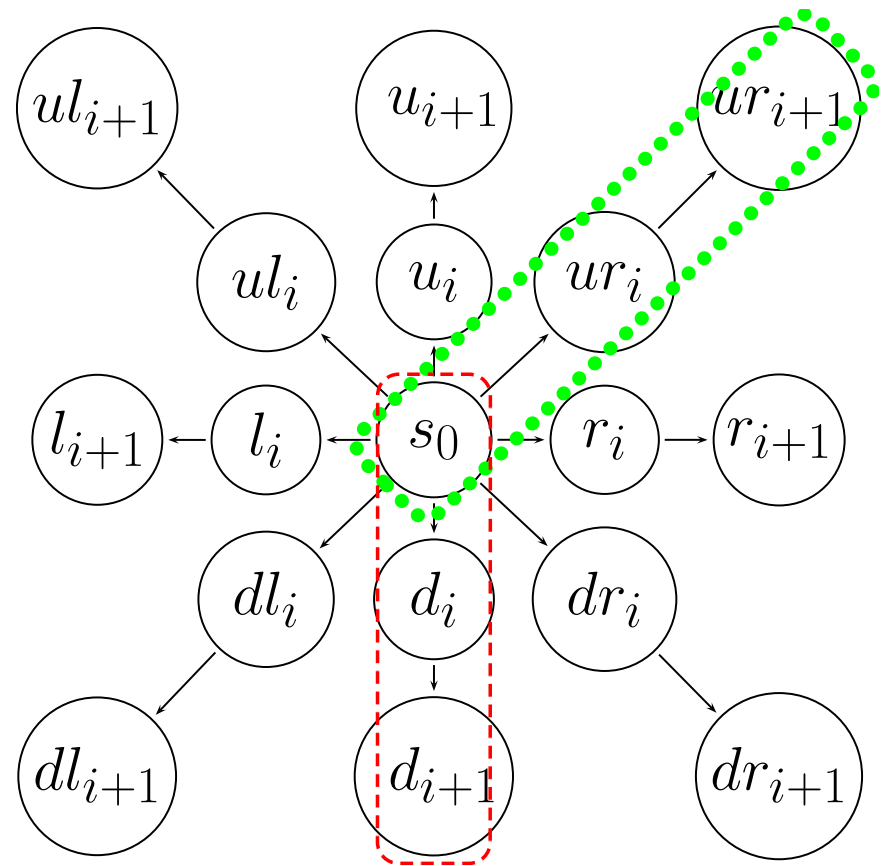
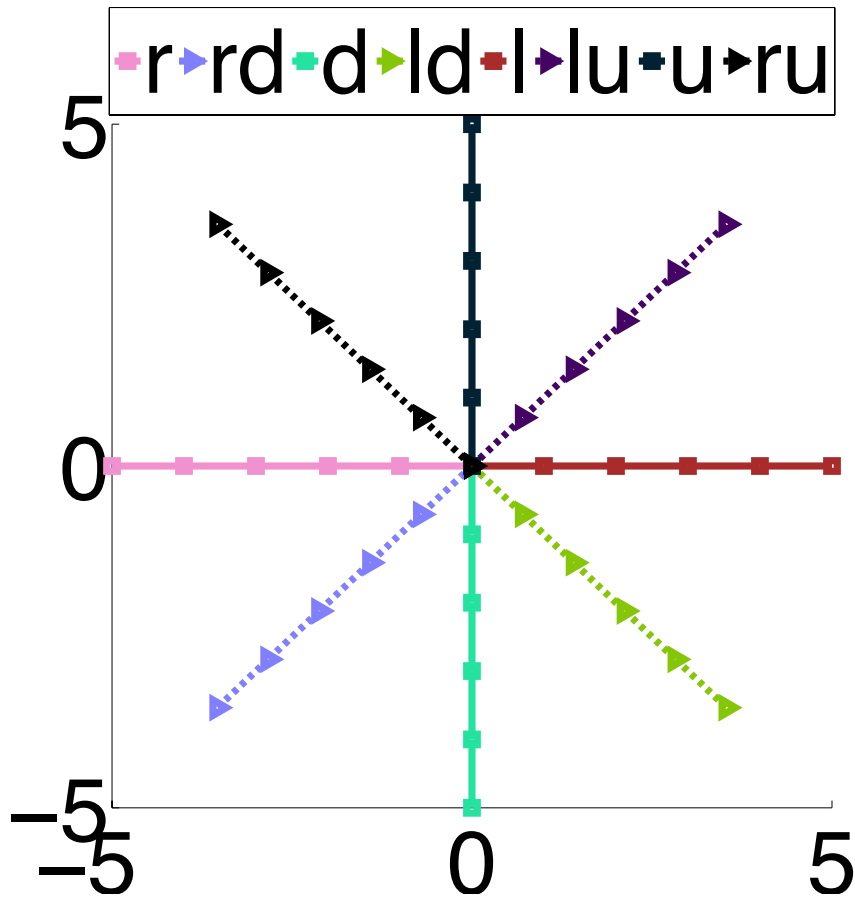


2-S-U Στατικές Υπομονάδες

- Δισδιάστατη Θέση
- Συσταδοποίηση με τον αλγόριθμο K-means

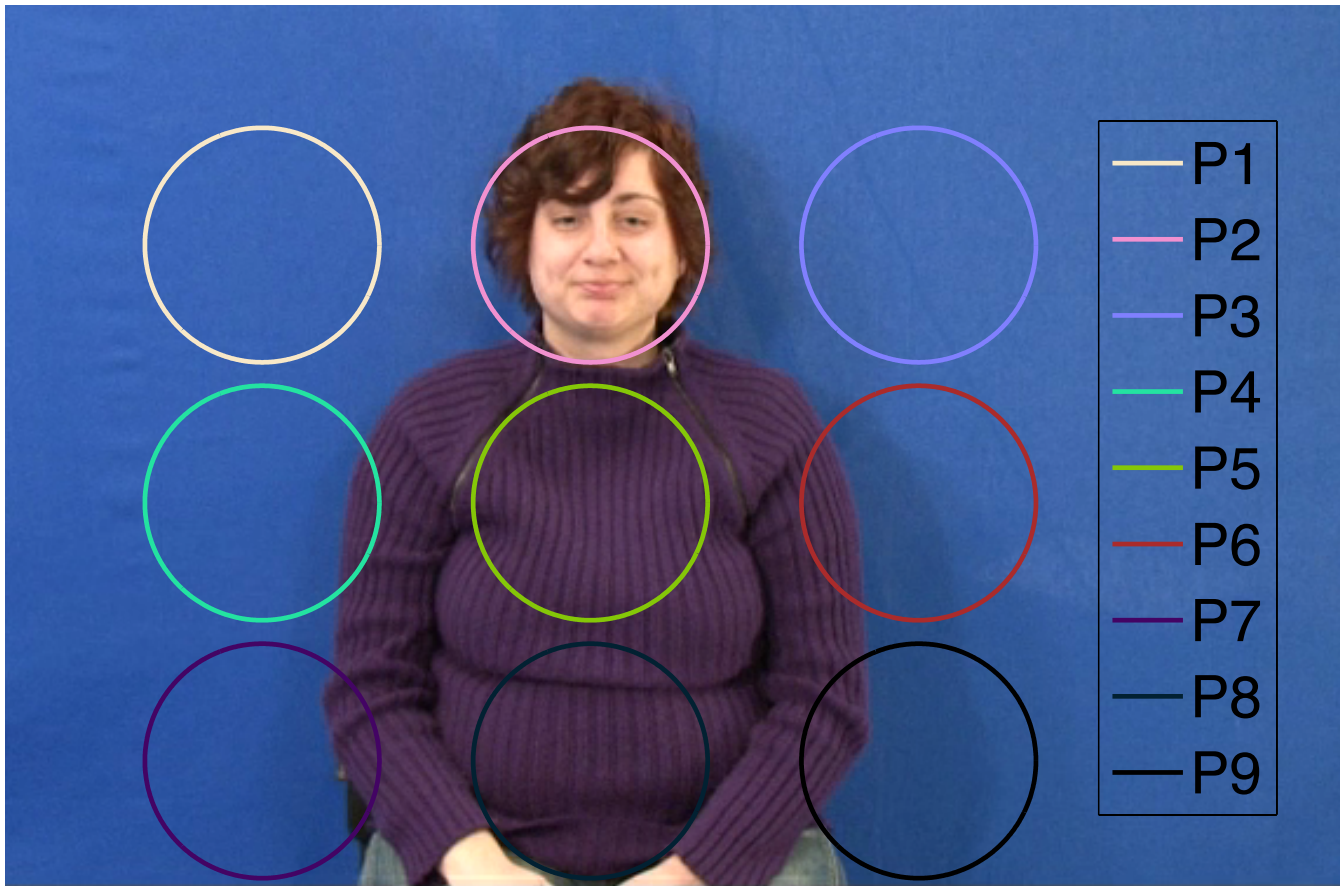


RAW Δυναμικές Υπομονάδες



- Ομοιόμορφη Διαμέριση του χώρου χαρακτηριστικών της Κατεύθυνσης
- Ευθείες, καμπύλες και κυκλικές κινήσεις

RAW Στατικές Υπομονάδες



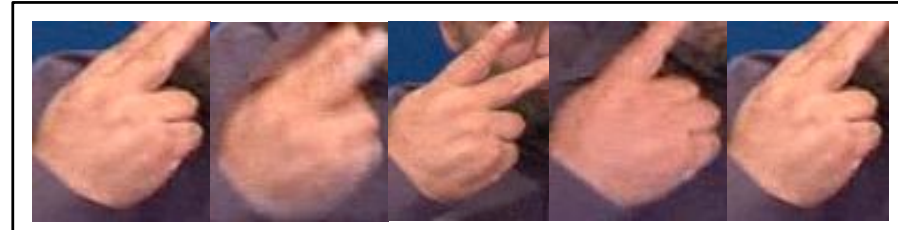
Ομοιόμορφη Διαμέριση του χώρου χαρακτηριστικών της δισδιάστατης Θέσης

Υπομονάδες Χειρομορφής

HS44



HS17



HS41



HS47



Spatial Pyramids Features

- Δεν λαμβάνεται υπόψη η κατάτμηση σε Δ/Σ τμήματα
- Συσταδοποίηση με τον αλγόριθμο K-means

Βίντεο Παράδειγματα για τα Λεξικά Υπομονάδας

ΡΥΖΙ

ΧΩΡΙΣ

2-S-U: S3
RAW: S13

HS: H5398



2-S-U: S3
RAW: S13

HS: H5495



Δυναμικές Υπομονάδες

Στατικές Υπομονάδες

Υπομονάδες Χειρομορφής

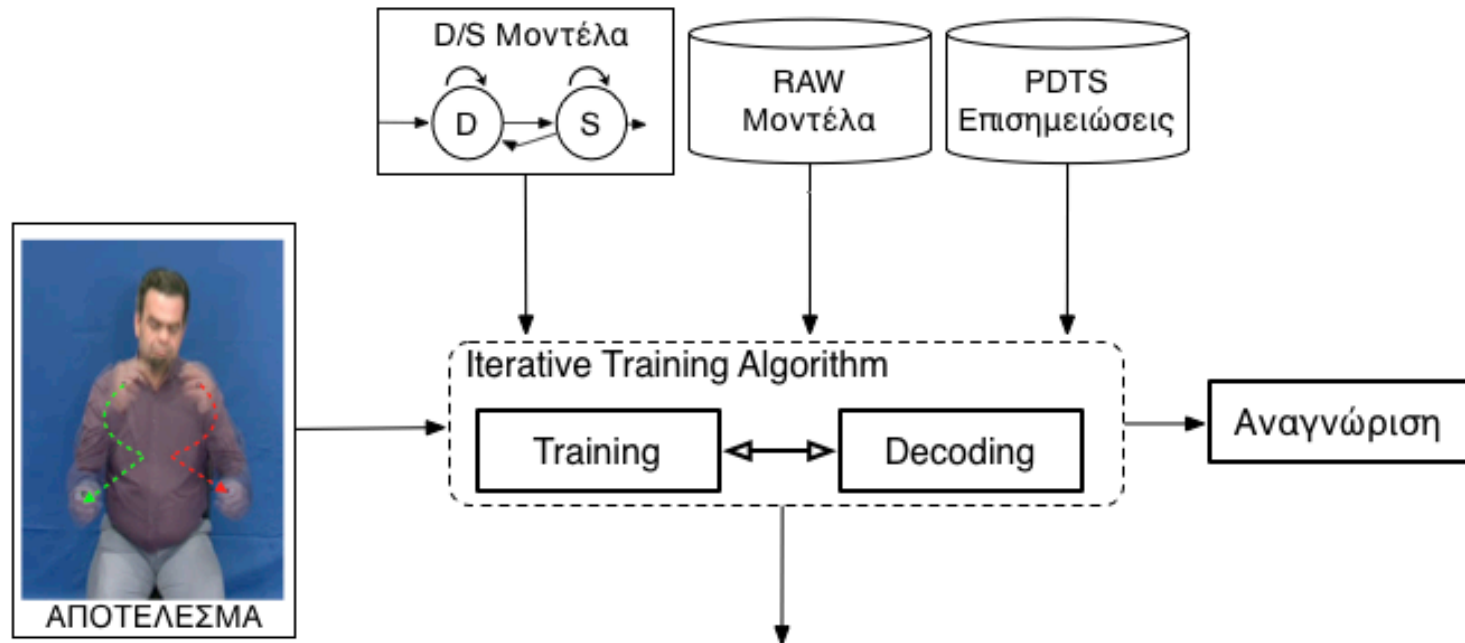


Φωνητικές Υπομονάδες

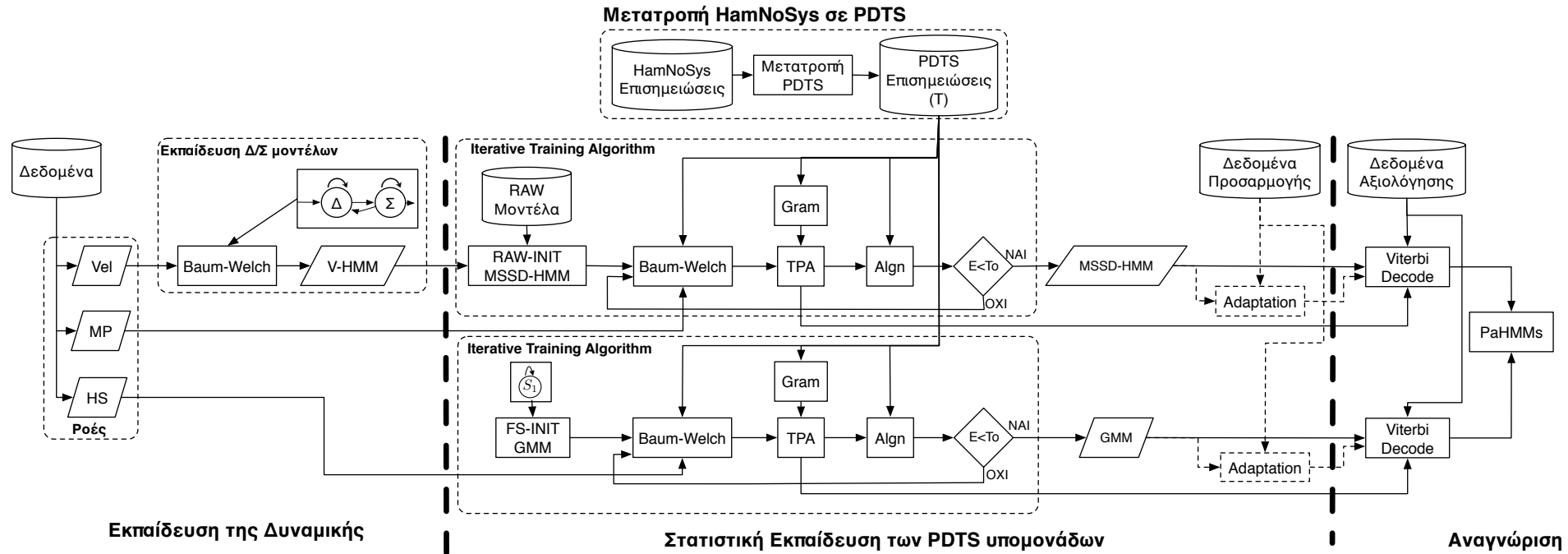
Ref: Theodorakis, Pitsikalis and Maragos, TASLP 2014

Ref: Pitsikalis, Theodorakis, Vogler and Maragos, CVPR Workshop GR 2011

Αναγνώριση ΝΓμε φωνητικές υπομονάδες

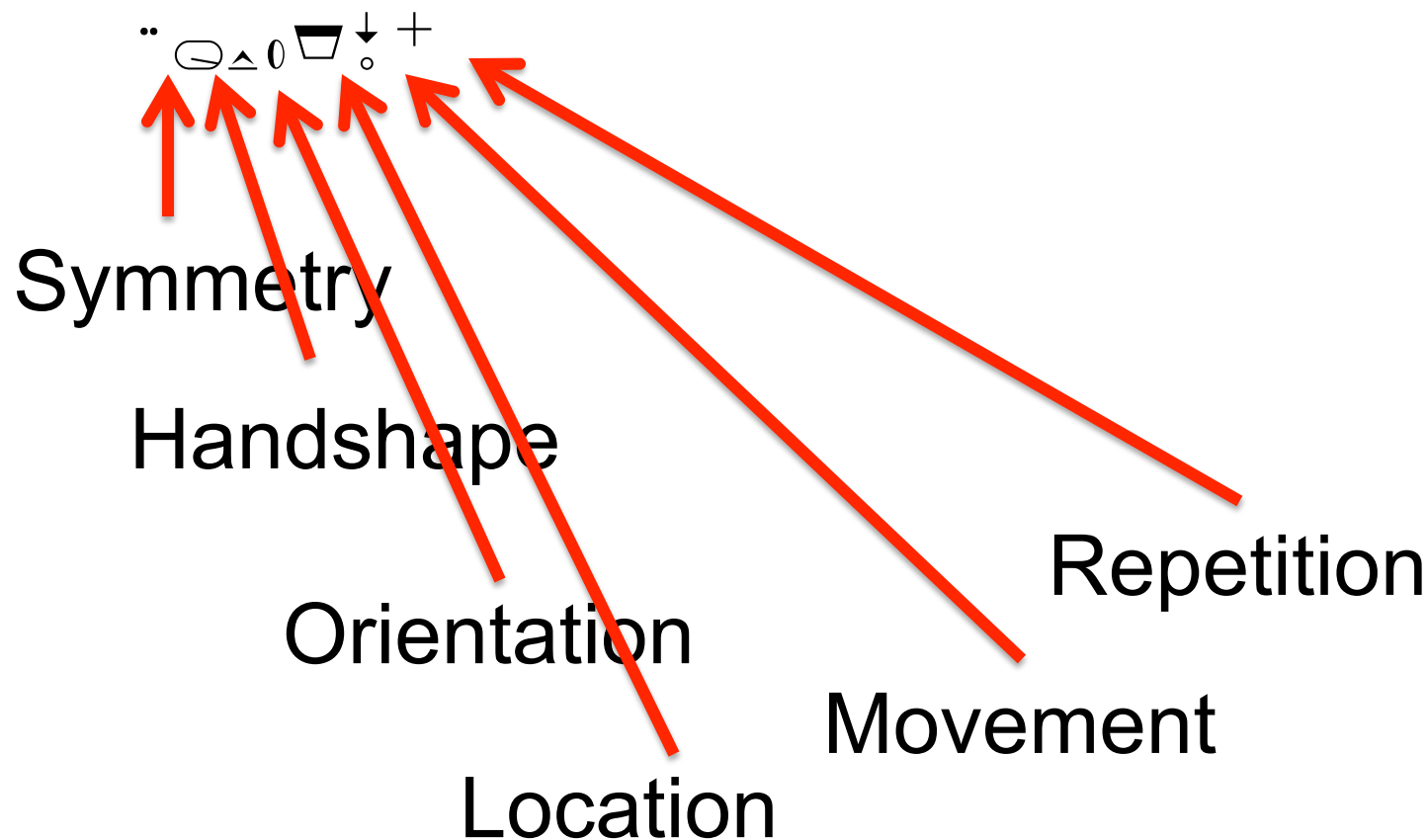


Σύνοψη συστήματος

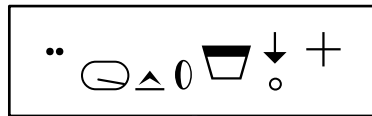


Επισημείωση HamNoSys

■ HamNoSys Επισημείωση ΚΑΡΕΚΛΑΣ:



Μετατροπή Hamnosys σε PDTS



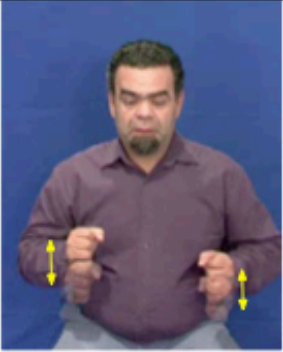


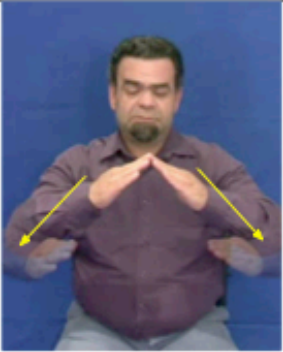
Hamnosys ΚΑΡΕΚΛΑ



PDTS ΚΑΡΕΚΛΑ

P--location_bodyarm,location=shoulders
T--directedmotion,direction=d,size=small
P--location_bodyarm,location=shoulders--directedmotion,direction=d,size=small
T--location_bodyarm,location=shoulders--location_bodyarm,location=shoulders
P--location_bodyarm,location=shoulders
T--directedmotion,direction=d,size=small
P--location_bodyarm,location=shoulders--directedmotion,direction=d,size=small

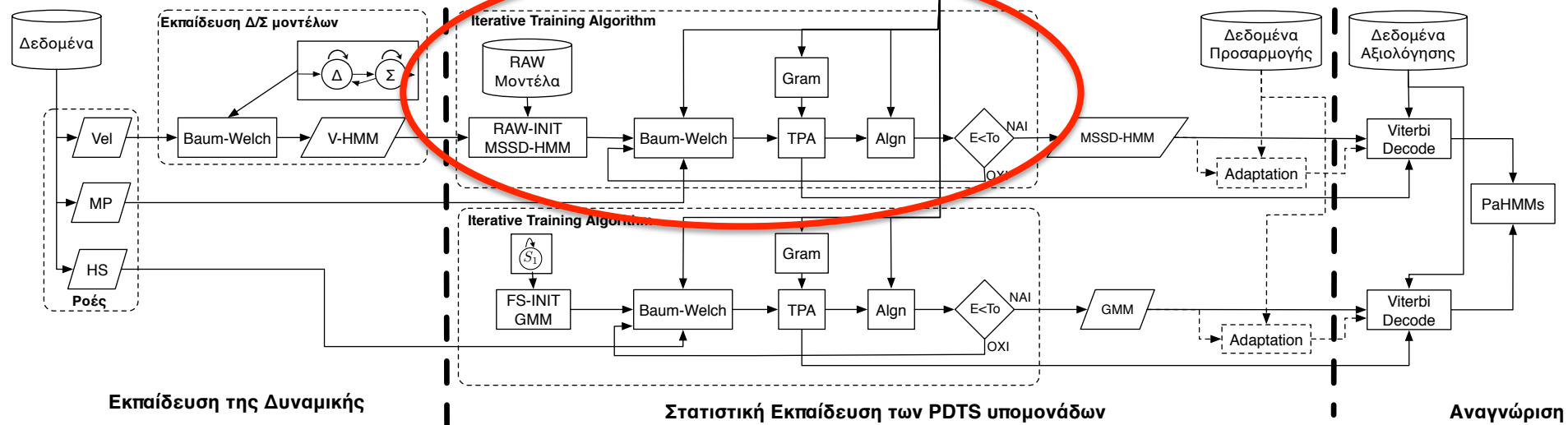
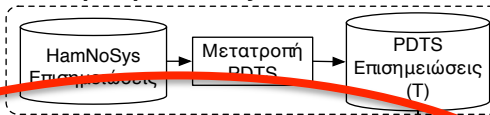
Παραδείγματα Νοημάτων με PDS επισημειώσεις

Νόημα	HamNoSys	PDS			Παράδειγμα
		D/ND	M-P/HS	Ακολουθία υπομονάδων ^{††}	
ΚΑΡΕΚΛΑ	" ⓪ △ 0 ▽ ↓ +	D	M-P	P-RShoulder T-d P-N/A T-N/A P-RShoulder T-d P-N/A	
		ND	M-P	P-LShoulder T-d P-N/A T-N/A P-LShoulder T-d P-N/A	
		D	HS	P-Fist T-N/A P-Fist T-N/A P-Fist T-d P-Fist	
		ND	HS	P-Fist T-N/A P-Fist T-N/A P-Fist T-d P-Fist	
ΒΛΕΠΩ		D	M-P	P-Eye T-do P-N/A	
		ND	M-P	-	
		D	HS	P-Finger23 T-N/A P-Finger23	
		ND	HS	-	
ΣΤΕΓΗ	" ⓪ r ⓪ [3 ⓪ 3 ⓪] X ⓪ > x	D	M-P	P-HeadHandTouch T-dr P-N/A	
		ND	M-P	P-HeadHandTouch T-dl P-N/A	
		D	HS	P-Flat T-N/A P-Flat	
		ND	HS	P-Flat T-N/A P-Flat	

Iterative Training Algorithm (ITA)

zoom

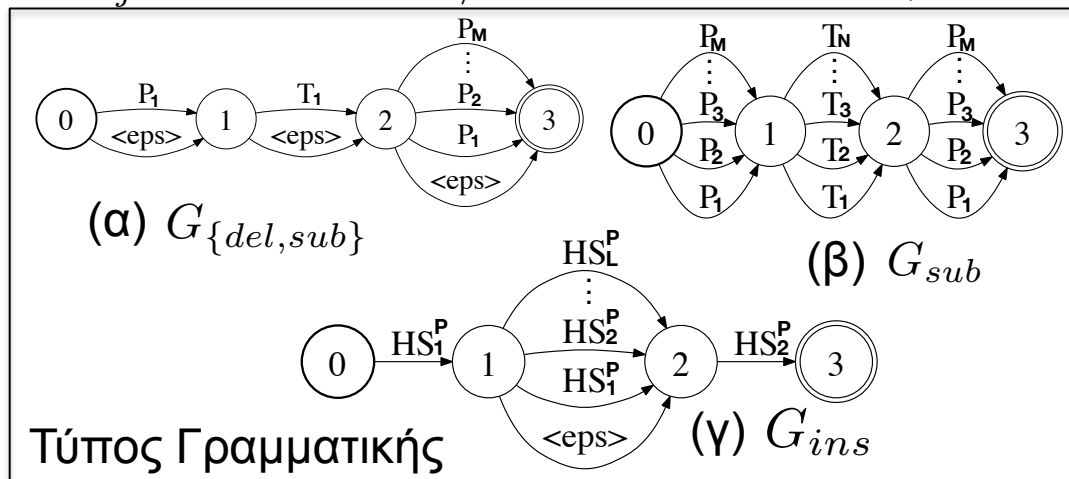
Μετατροπή HamNoSys σε PDTS



Iterative Training Algorithm (ITA)

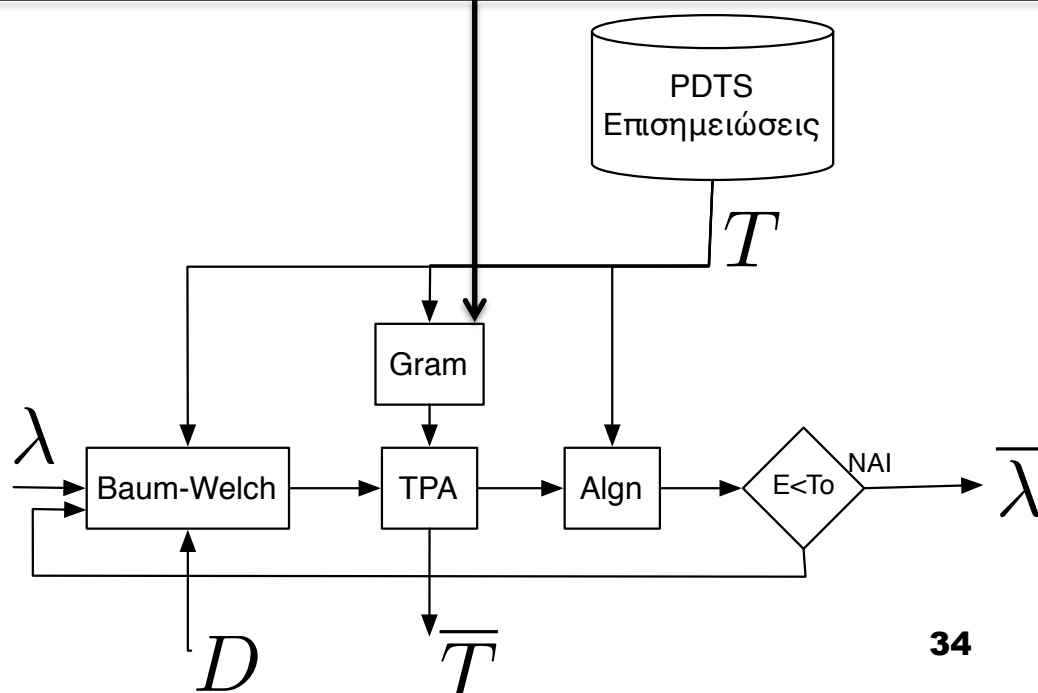
$$T_j^{MP} = P_1 T_1 P_{N/A} \quad T_j^{HS} = HS_1^P HS_{N/A}^T HS_2^P$$

- Αρχικοποίηση των HMM λ
- Τύπος Γραμματικής
 $G_{sub} \quad G_{\{del,sub\}} \quad G_{ins}$

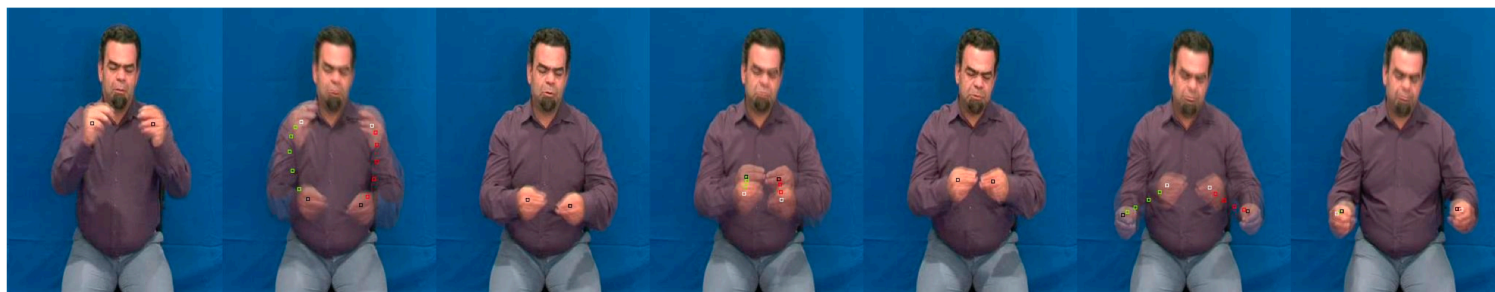


Algorithm 1 $[\bar{\lambda}, \bar{T}] = ITA(\lambda, T, D)$

- 1: $E = 1, T_0 = 0.05, \bar{T} = T$
- 2: $\lambda = Init(D, T)$
- 3: **while** $E > T_0$ **do**
- 4: $\bar{\lambda} = BW(\lambda, \bar{T}, D)$
- 5: **for** $i = 1$ **to** N **do**
- 6: $G = Gram(T_i)$
- 7: $\bar{T}_i = Dec(\bar{\lambda}, G, D_i)$
- 8: **end for**
- 9: $E = Algn(T, \bar{T})$
- 10: $T = \bar{T}, \lambda = \bar{\lambda}$
- 11: **end while**



Παράδειγμα εφαρμογής του ΙΤΑ (1)



Right-Hand
Left-Hand

P:Head
P:Head

T:Curve,D=d,C=r
T:Curve,D=d,C=l

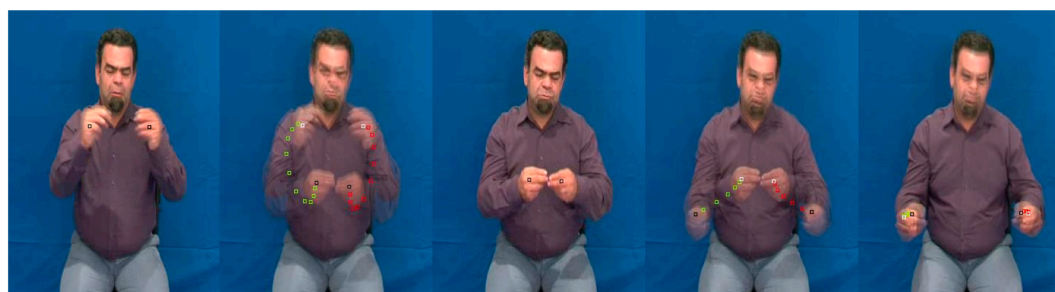
P:Head
P:Head

T:Straight,D=r
T:Straight,D=l

P:Head
P:Head

T:Curve,D=l,C=r
T:Curve,D=l,C=l

P:NaN
P:NaN



Right-Hand
Left-Hand

P:Head
P:Head

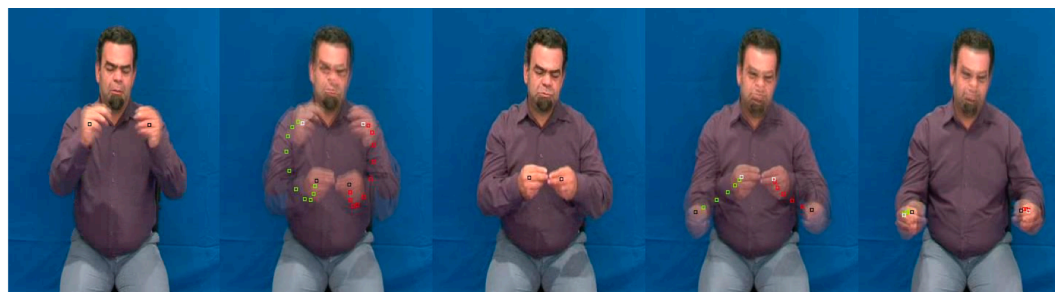
T:Curve,D=d,C=r
T:Curve,D=d,C=l

P:Head
P:Head

T:Straight,D=r
T:Straight,D=l

P:Below-Stomach
P:Below-Stomach

$G_{\{del,sub\}}$
PDTS Grammar
2 deletions and
1 substitution



Right-Hand
Left-Hand

P:Head
P:Head

T:Curve,D=d,C=r
T:Curve,D=d,C=l

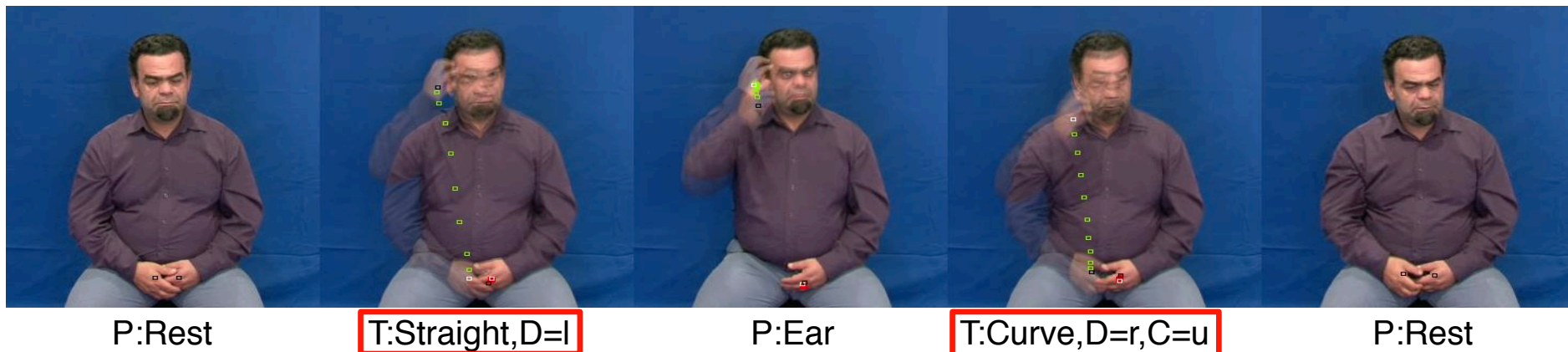
P:Chest
P:Chest

T:Straight,D=dr
T:Straight,D=dl

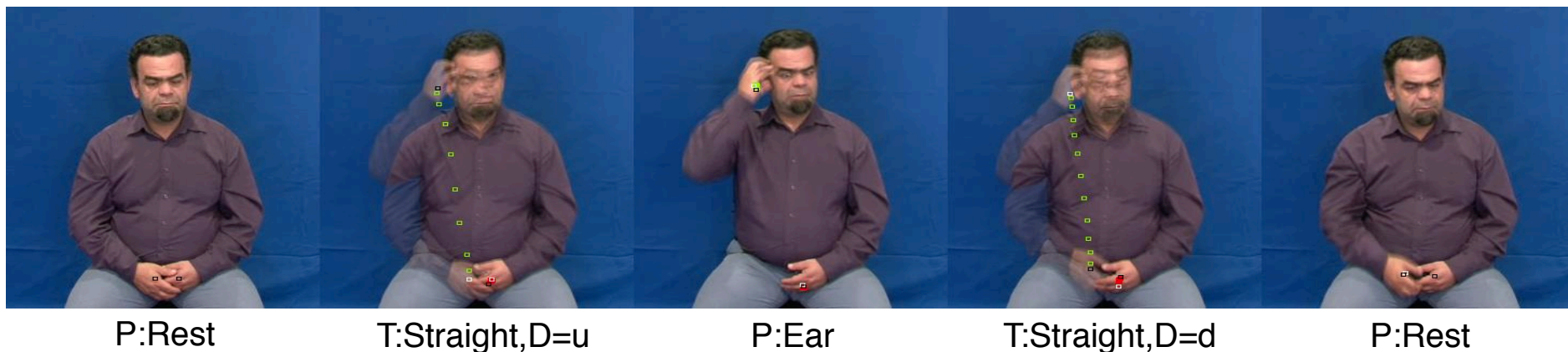
P:Below-Stomach
P:Below-Stomach

G_{sub}
PDTS Grammar
2 substitution

Παράδειγμα εφαρμογής του ΙΤΑ (2)



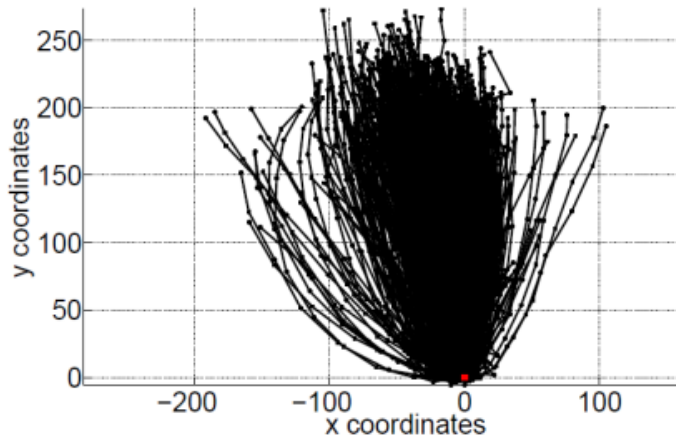
(α) Flat-Start Αρχικοποίηση των HMMs



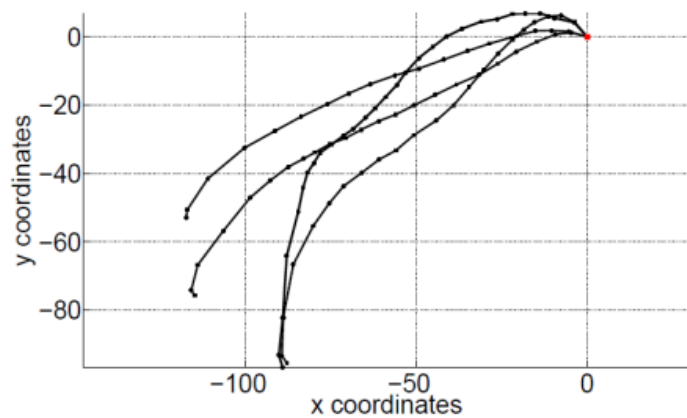
(β) Αρχικοποίηση των HMMs χρησιμοποιώντας τα RAW HMMs

Φωνητικές Υπομονάδες

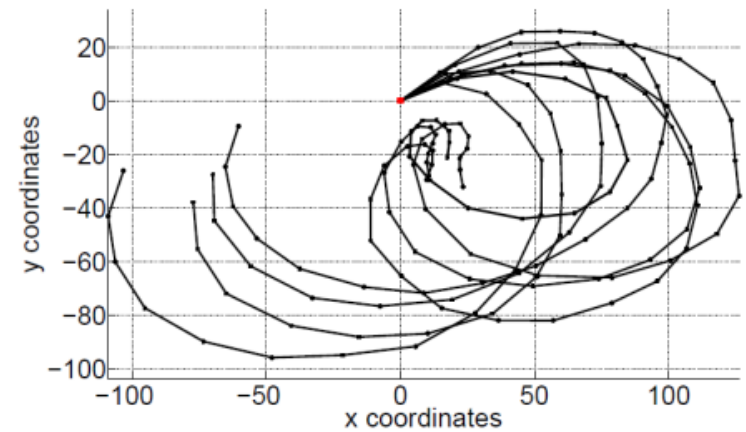
Transition/Epenthesis



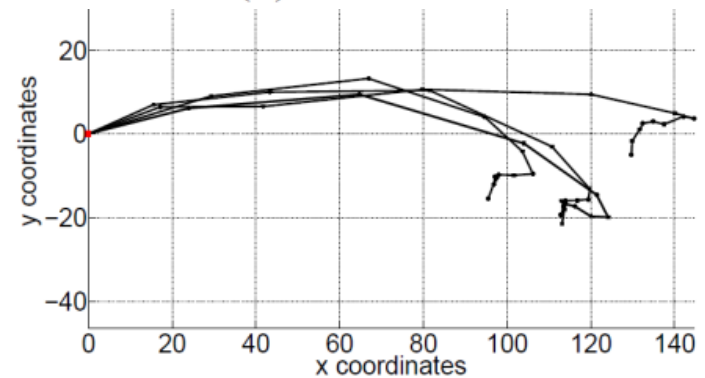
(a) E-to-head



(c) T-down-right



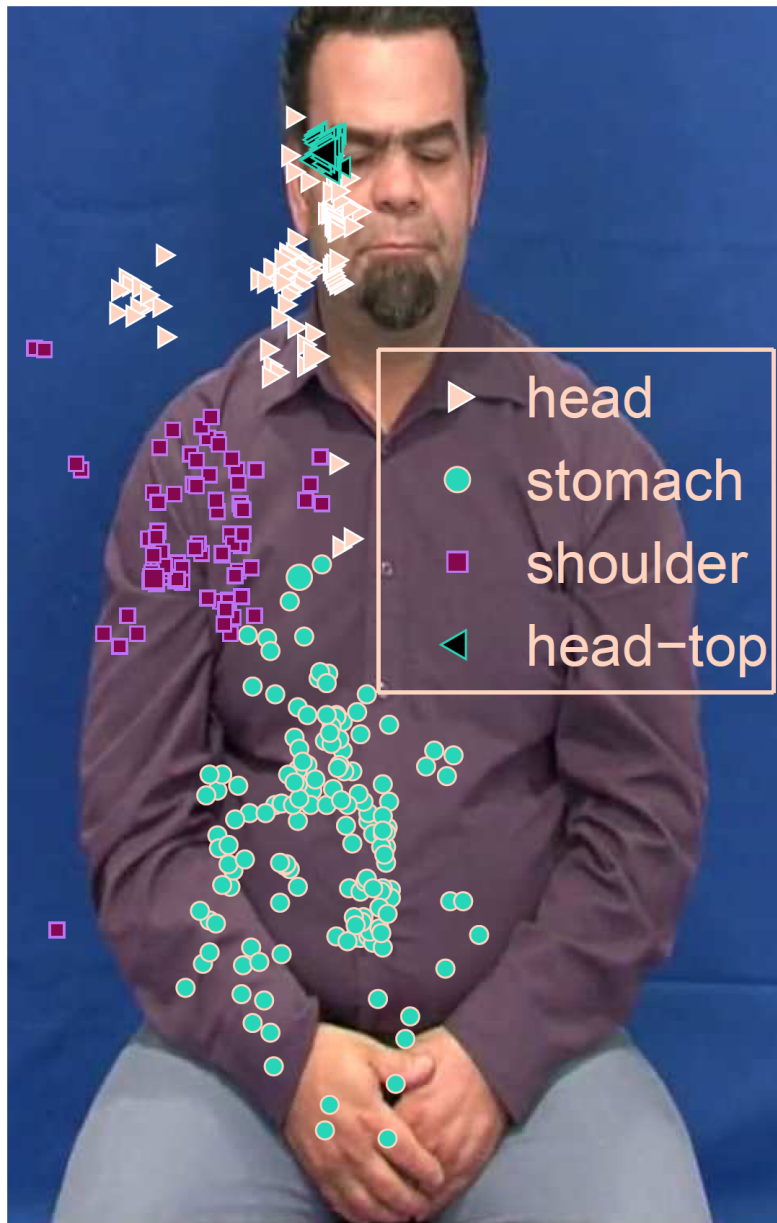
(b) T-circular



(d) T-in-left

PSU	Sign	Type	PDTS Label
E-to-head	*	E	rest-position — location-head
T-circular	EUROPE	T	circularmotion, axis=i
T-down-right	PILE	T	directedmotion, direction=dr, small
T-in-left	IMMEDIATELY	T	directedmotion, direction=il, fast=true, halt=true

Φωνητικές Υπομονάδες



Postures

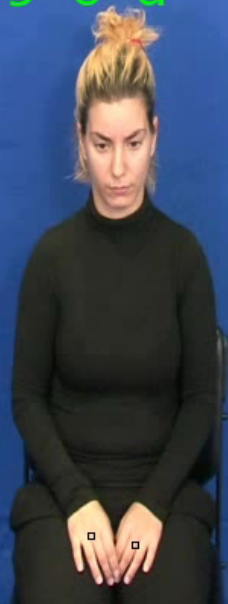
PSU	Type	PDTS Label
P-forehead	P	location=forehead
P-stomach	P	location=stomach
P-shoulder	P	location=shouldertop, side=right_beside
P-head-top	P	location=head-top

Βίντεο Παράδειγματα για τα PDTS Λεξικά

ΡΥΖΙ

MP: P:none

HS: finger2345-o-d



ΧΩΡΙΣ

MP: P:none

HS: flat-u-ul



Transition Υπομονάδες

Posture Υπομονάδες

Υπομονάδες Χειρομορφής



Μοντέλα, Σύμμεριξη, Προσαρμογή σε Νοηματιστή

Ref: Theodorakis, Pitsikalis and Maragos, IVC 2014

Ref: Theodorakis, Pitsikalis and Maragos, TASLP 2014

Multi-Stream Switching Probability Distribution

Επέκταση της multi-space probability distribution (MSD)
[Tokuda 2002]

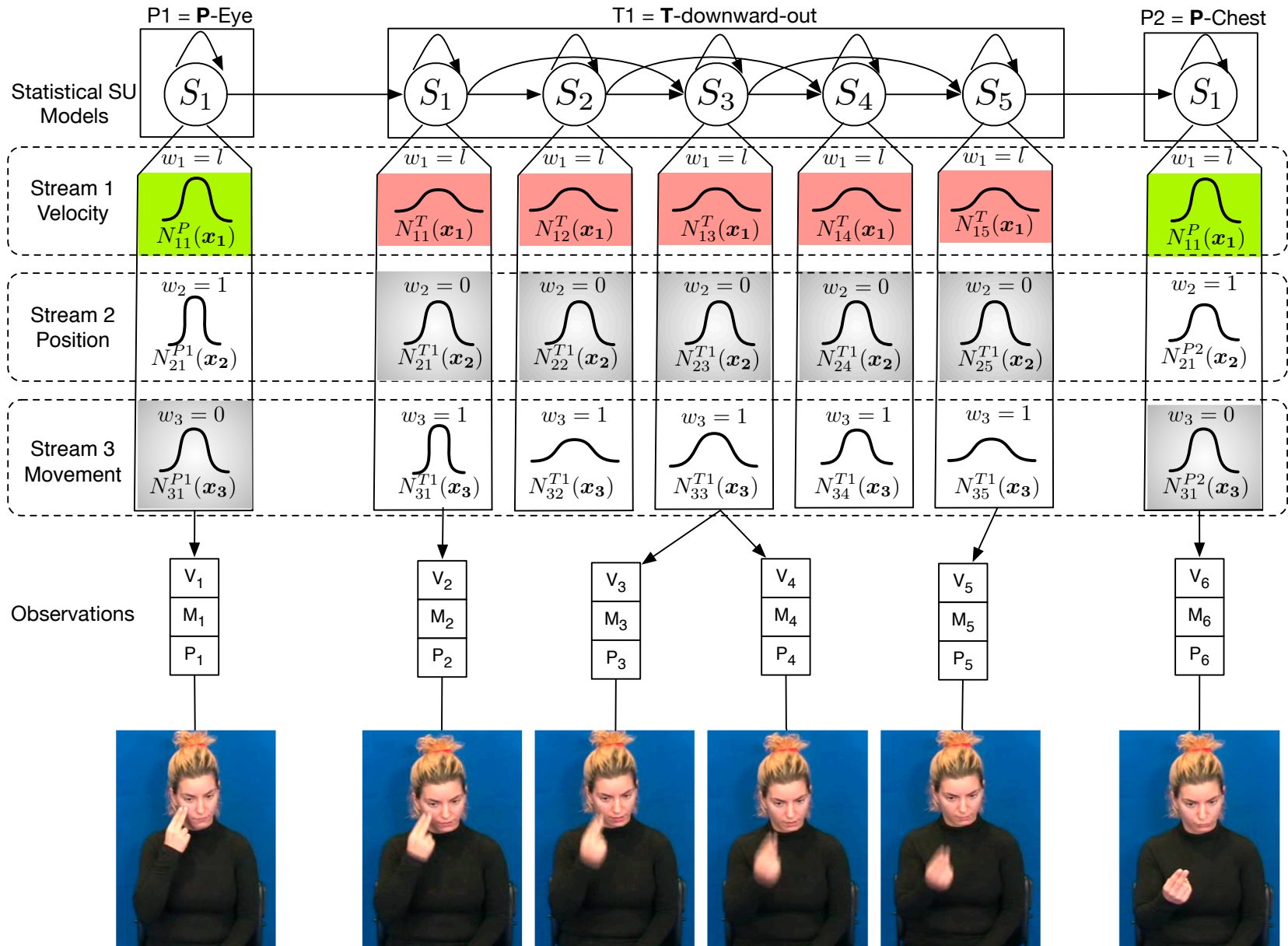
Έστω $\mathbf{o} = (\mathbf{x}, \mathbf{y})$ όπου $\mathbf{x} \equiv (x_1; \dots; x_G)$, $\mathbf{y} \in \{l_1, l_2, \dots, l_M\}$

G διαφορετικά streams, M διαφορετικοί τύποι κλάσεων

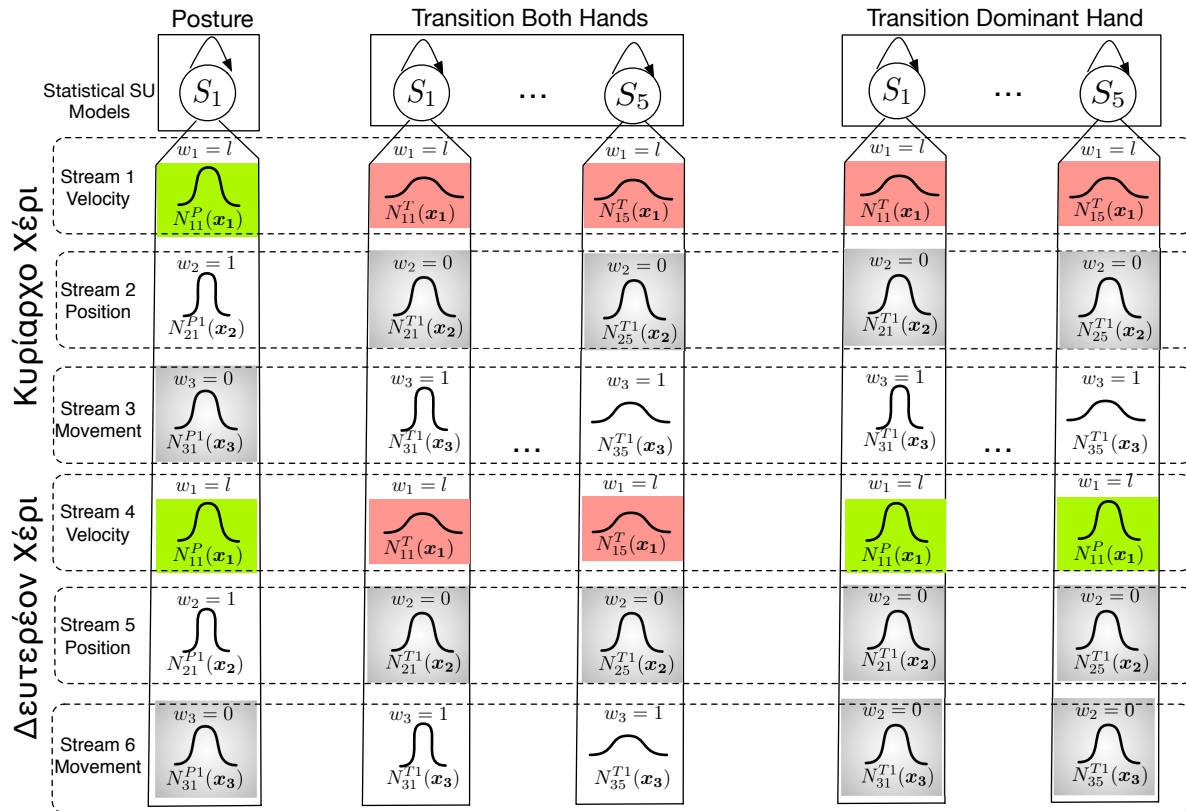
Η πιθανότητα της παρατήρησης \mathbf{o} είναι:

$$b(\mathbf{o}) = \prod_{g \in I(\mathbf{y})} \mathcal{N}_g(\mathbf{x}_g)^{w_g}$$

Παράδειγμα Υπομονάδων με MSSD-HMMs

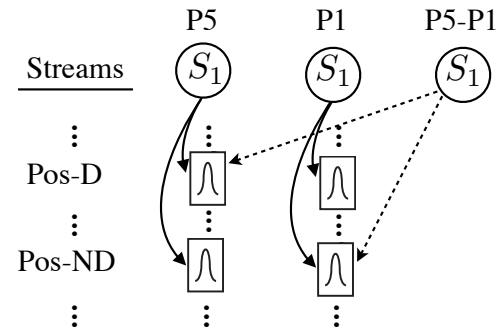
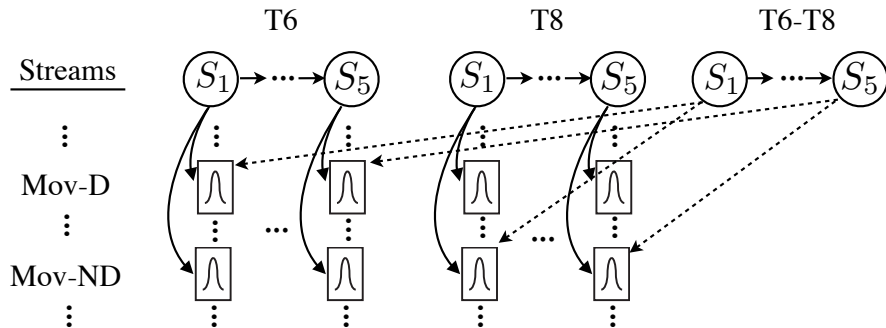


Σύμμειξη κυρίαρχου και δευτερεύοντος χεριού

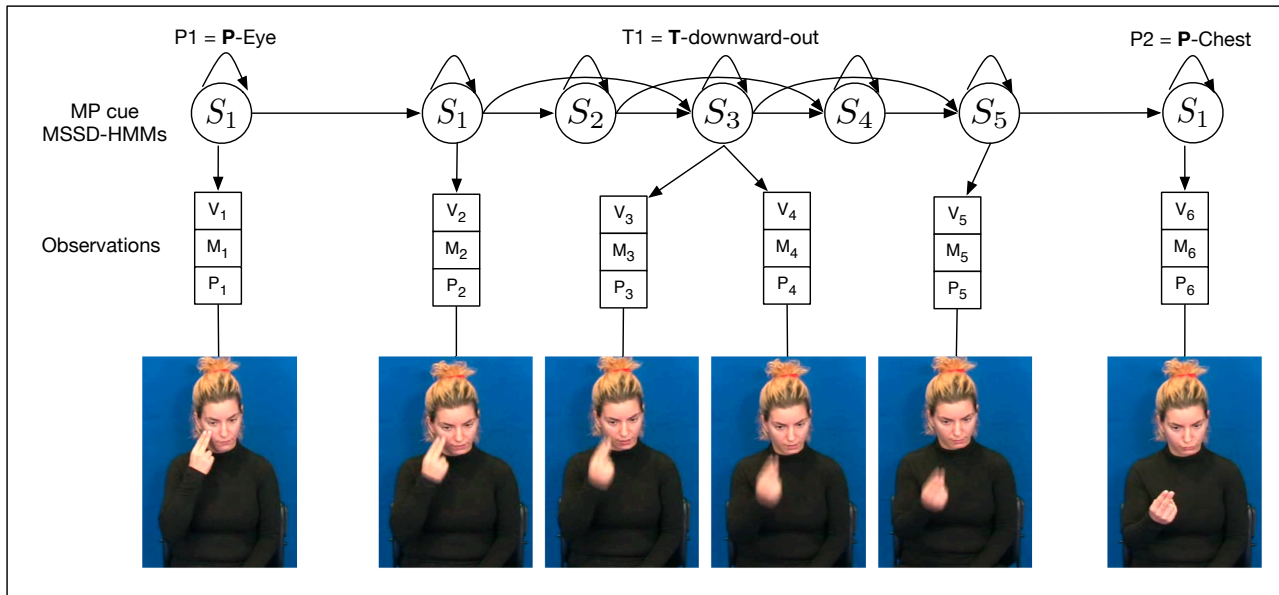


Transition SUs examples

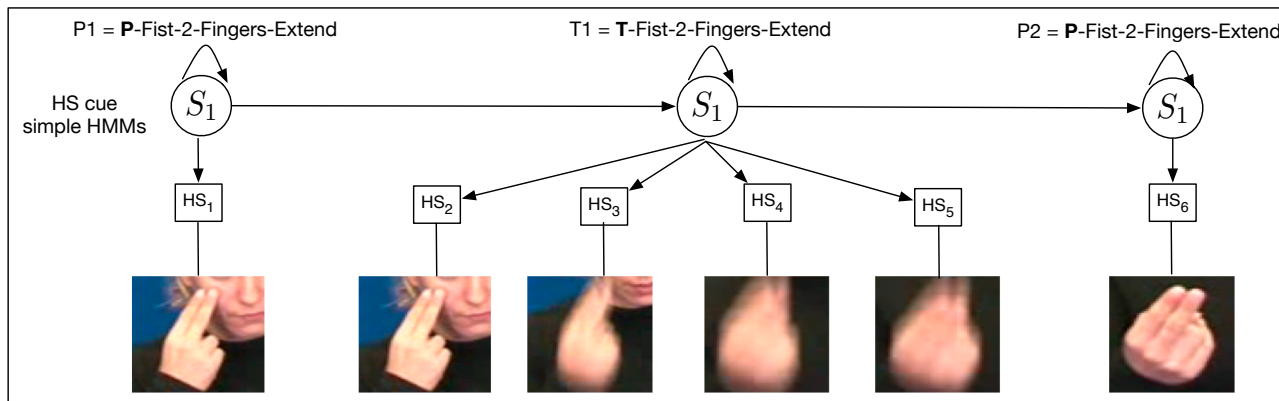
Posture SUs examples



Σύμμειξη ροών κίνησης-θέσης και χειρομορφής



$$\sum_{s \in \{MP, HS\}} w^s * \log(\max_{Q^s} P(Q^s, O^s | \lambda^s))$$



Προσαρμογή σε Νοηματιστή



S5

D14

S1

D21

S1

D16

S4

D21

S5



S5

D14

S1

D29

S5

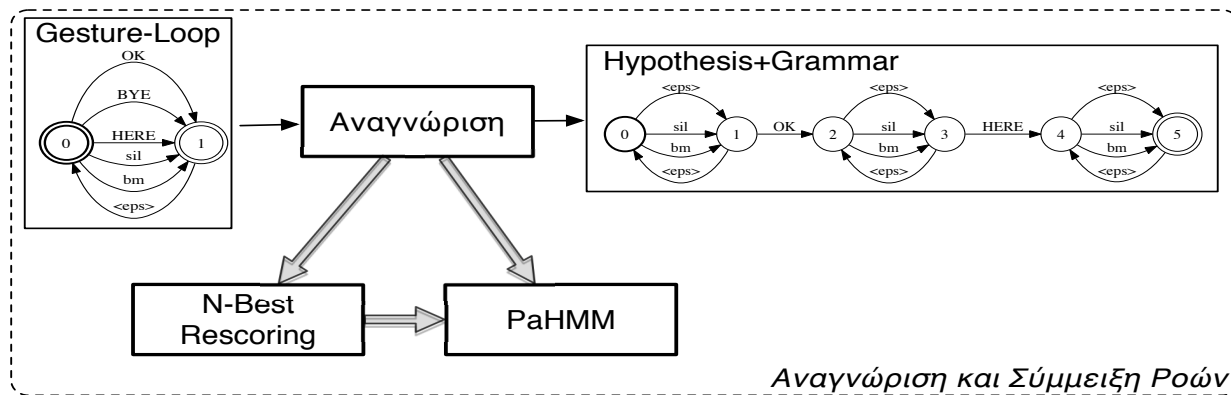
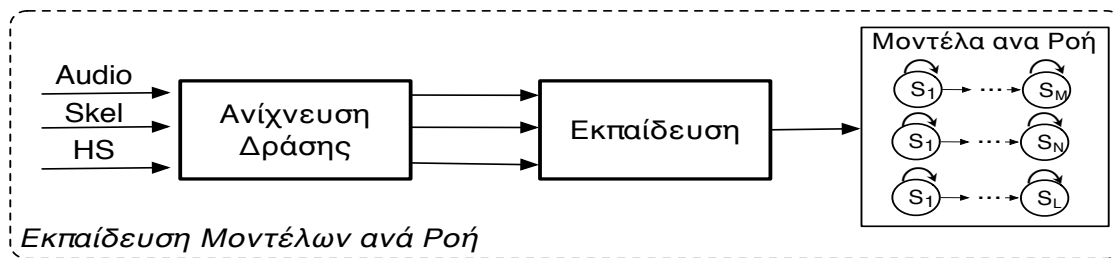
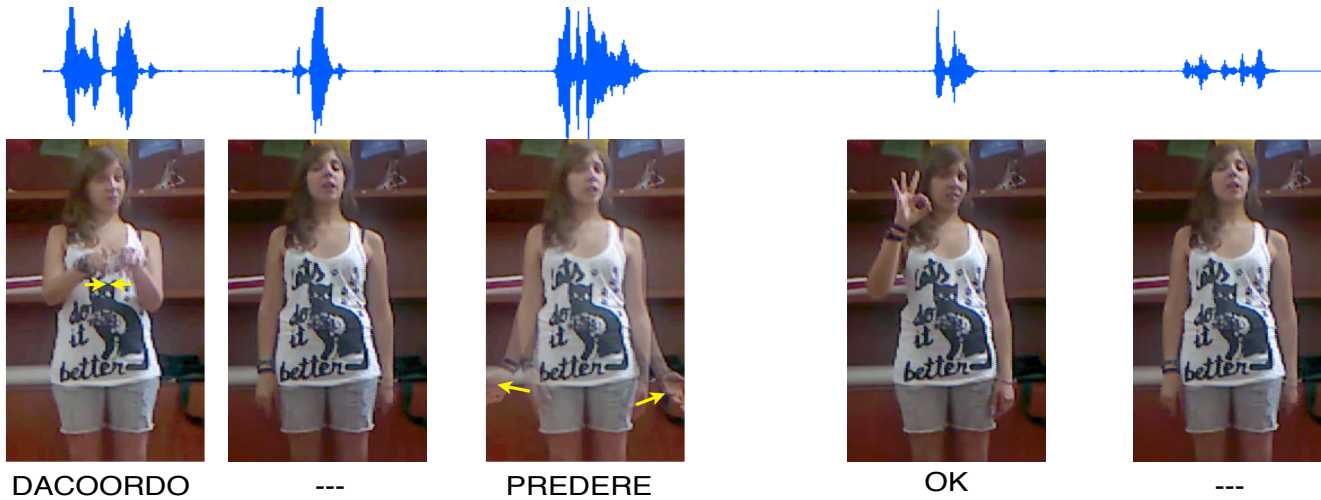
- Επιλογή σύνολου δεδομένων προσαρμογής από νέο νοηματιστή
- Προσαρμογή των μοντέλων υπομονάδας με MLLR
- Παραγωγή νέων προφορών
- Εισαγωγή νέων προφορών ως νέες εγγραφές στο λεξικό



Πολυτροπική Αναγνώριση Χειρονομιών

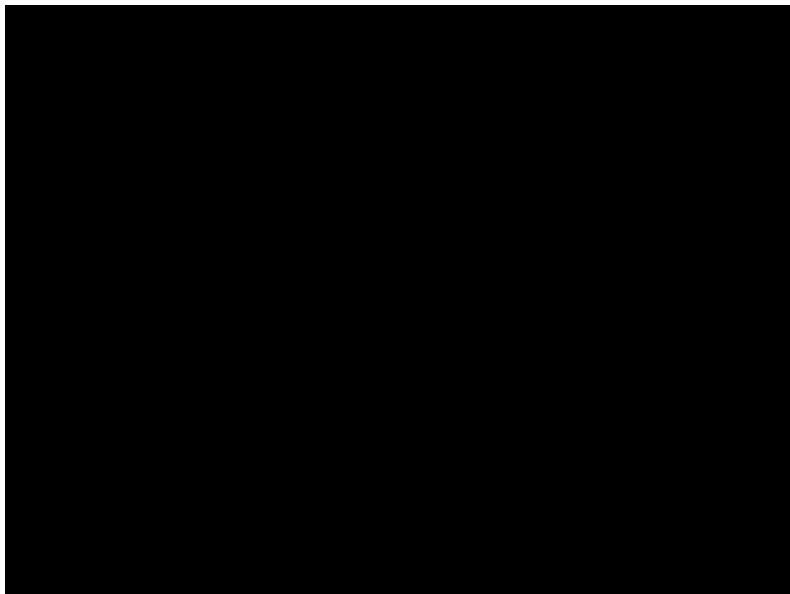
Ref: Pitsikalis, Katsamanis, Theodorakis and Maragos, JMLR 2014

Ref: Pavlakos, Theodorakis, Pitsikalis, Katsamanis and Maragos, ICIP 2014

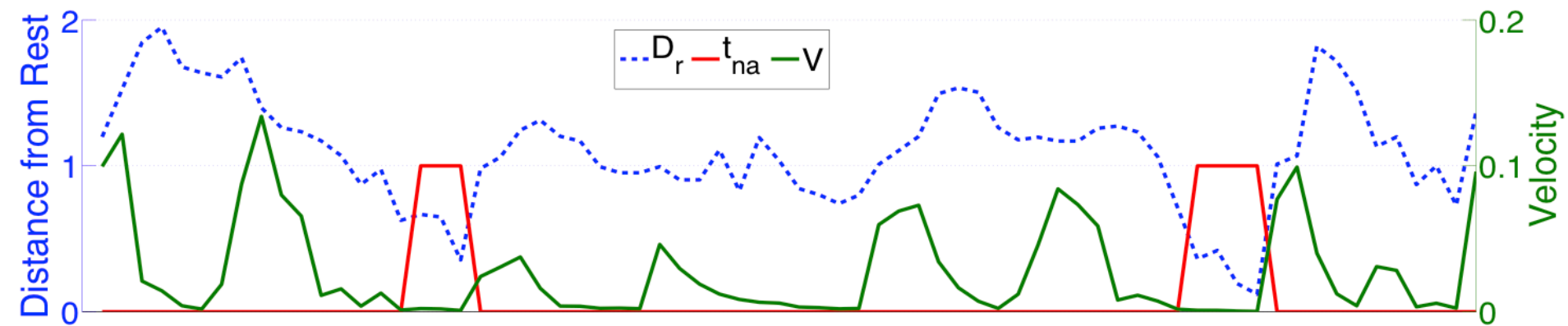


DACOORDO, PREDERE, OK

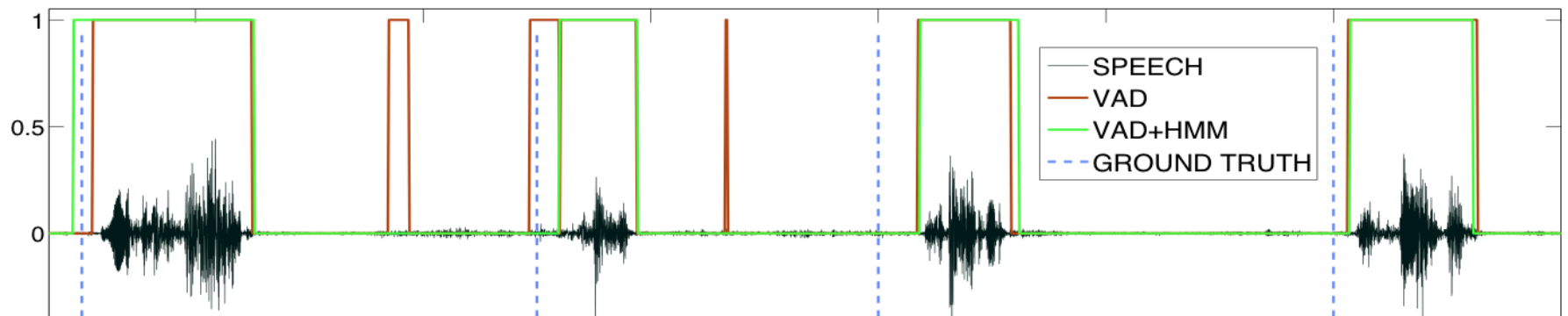
Παραδείγματα από την ChaLearn βάση



Πολυτροπική Ανίχνευση Δράσης



Activity	Non-Activity	Activity	Non-Activity	Activity	Non-Activity	Activity
----------	--------------	----------	--------------	----------	--------------	----------

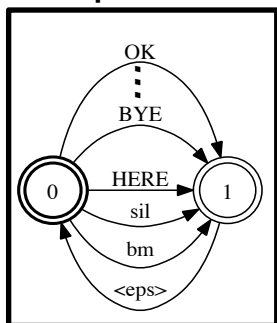


Παραγωγή Υποθέσεων Αναγνώρισης

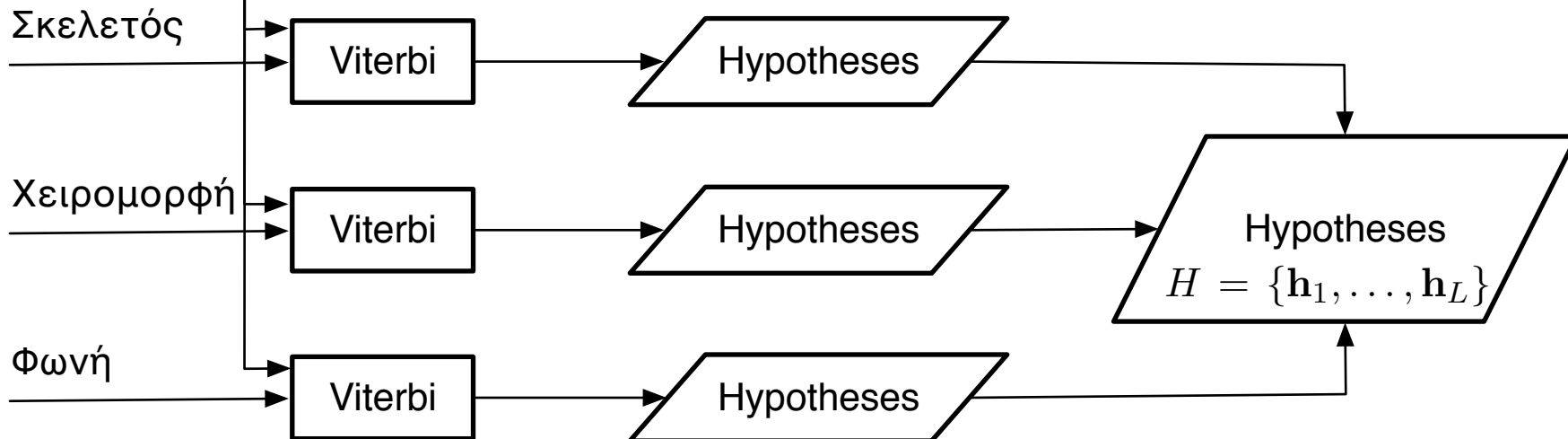
Έστω $G = \{g_i\}, i = 1, \dots, |G|$ χειρονομίες και $S = \{O_i\}, i = 1, \dots, |S|$ ροές

Παραγωγή Υποθέσεων

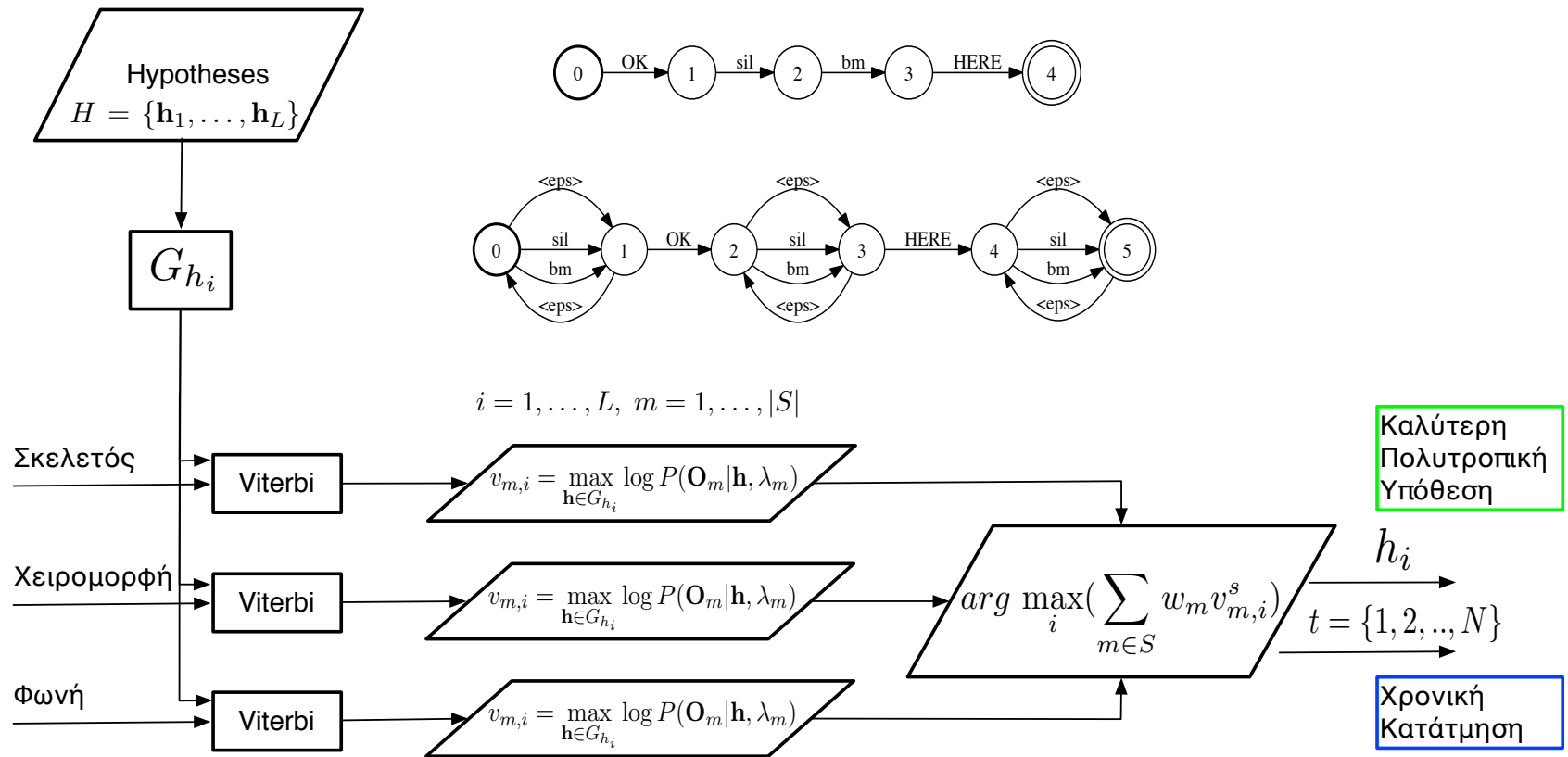
Loop-Gram



$$\mathbf{h} = [bm, g_1, sil, g_5, \dots, bm, sil, g_3]$$



Πολυτροπικό Σκοράρισμα και Τμηματική Παράλληλη Σύμμειξη



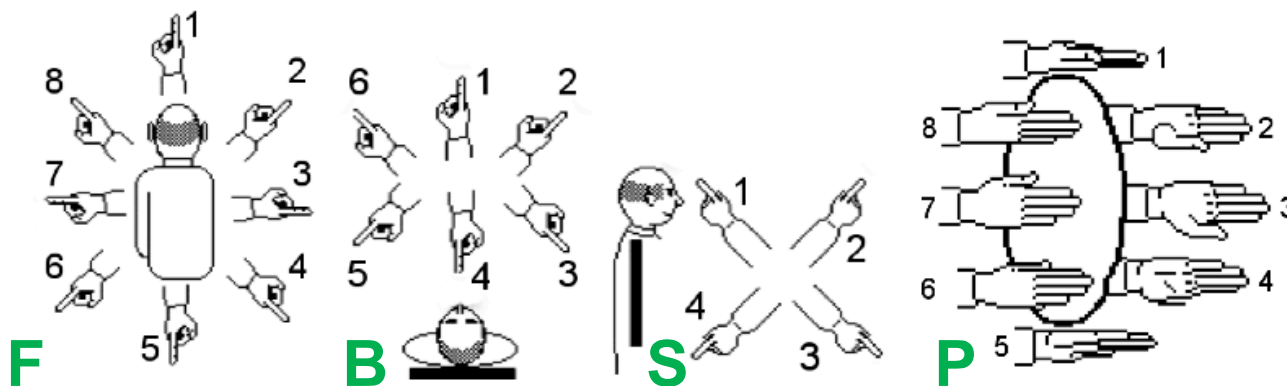
$$LL_{m,j}^t = \max_{\mathbf{q} \in Q} \log P(\mathbf{O}_m^t, \mathbf{q} | \lambda_{m,j}), \quad j = 1, \dots, |G| + 2,$$

$$LL_j^t = \sum_{m \in S} w'_m LL_{m,j}^t$$

Πειράματα

Επισημείωση Χειρομορφών

- **Handshape Id** based on **SignStream** Conventions: e.g. ('A', 'B', '1', 'C' etc)
- Handshape **Orientation Parameters** based on **Hamnosys**



■ # of Instances

HS: 5

HS: BL

	HS: 5				HS: BL								
	5C	5	5	5	5C	BL	BL	BL	BL	BL	BL	BL	
H	1	1	4	4	5C	5	5	5	BL	BL	BL	BL	BL
F	8	1	7	6	1	7	8	1	8	7	8	8	1
S	0	0	0	3	1	0	2	2	0	0	0	0	0
B	0	0	0	6	4	0	1	1	0	0	0	0	0
P	1	8	3	1	3	3	1	5	2	3	3	4	5
N	14	24	10	12	27	38	14	19	10	15	23	30	22



Ταξινόμηση Χειρομορφών

BU400

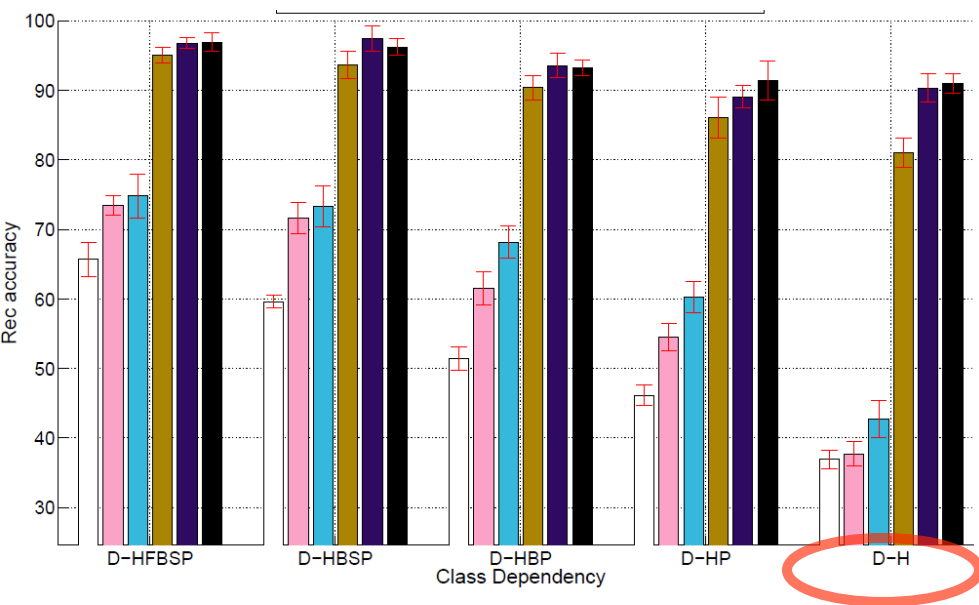
Proposed: Affine SAM
 Direct Similarity SAM
 Direct Translation Scale SAM

Aff-SAM
 DTS-SAM
 DS-SAM

Fourier Descriptors (FD)
 Moments (M)
 Region Based (RB)

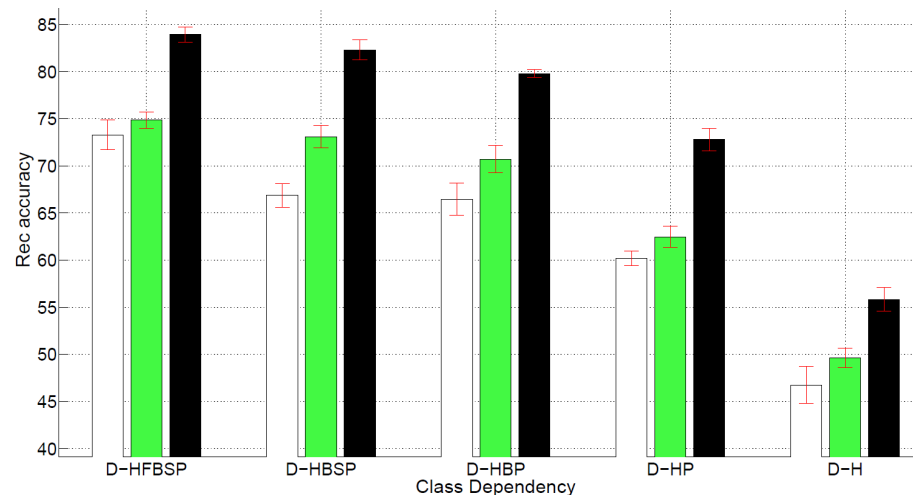
FD
 RB
 M

Aff-SAM
 DTS-SAM
 DS-SAM



ΟΧΙ Επικαλύψεις

+ Επικαλύψεις



Class Dependency

Class Dependency	Annotation Parameters				
	F	S	B	P	H
D-HFSBP	D	D	D	D	D
D-HSBP	*	D	D	D	D
D-HBP	*	*	D	D	D
D-HP	*	*	*	D	D
D-H	*	*	*	*	D

Δεδομενοκεντρικές Υπομονάδες (1)

DS-Lem
300 Νοήματα

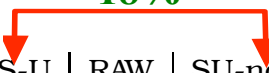
Γνωστός Νοηματιστής

2-S-U	RAW	SU-Frame [1]	MC [2]	SPs [2]	SU-Segm [3]	Sign-DTW [4]
96.98	96.9	96.2	71.4	74.1	96.2	99

- [1] B. Bauer and K. F. Kraiss, GW, 2001.
 [2] H. Cooper et. al. JMLR, 2012.
 [3] G. Fang, et. al, PR 2004.
 [4] H. Wang, et. al. ECCV Workshops 2010.

Άγνωστος Νοηματιστής

Νοηματιστής	Ποή	2-S-U	RAW	SU-noDSC	SU-Segm	SU-Frame	Sign-DTW
Όλγα	M-P	30.1	22.2	11.3	14.23	11.4	25.8
	HS	38.8	38.8	38.8	38.8	38.8	42.2
	M-P+HS	61.2	52.4	46.6	54.4	40.53	57.9
Κώστας	M-P	29	23.3	11.8	9.1	11.9	24.4
	HS	28.8	28.8	28.8	28.8	28.8	32.7
	M-P+HS	50.1	42.3	33.2	32.6	35.53	46.3

18%


7.8% 15.8% 12.1% 17.6% 3.5%

Διανύσματα
χαρακτηριστικών

MP: **P, D, V, L**
 HS: **SP**

Προσαρμογή σε Νοηματιστή

	Νοηματιστής	Όλγα			Κώστας			
	Ποή	M-P	HS	M-P+HS	M-P	HS	M-P+HS	
Μέθοδοι	2-S-U	30.9	40.6	58.6	26.1	34	49.6	
	2-S-U+MLLR	33	54	67.9	28.8	59.2	71.7	+15.7%
	2-S-U+MLLR+IP	67.3	89.1	92.7	63.1	88.2	92.3	+22.7%

Δεδομενοκεντρικές Υπομονάδες (2)

Άγνωστος Νοηματιστής

ASLLVD
97 νοήματα

	Νοηματιστής	M-P	HS	M-P+HS
2-S-U	Dana	40.31	44.21	63.15
2-S-U	Lana	38.2	40.1	61.3
Sign-DTW	Dana	26.3	41	55.78
Sign-DTW	Lana	33.6	35.7	53.6
		+9.3%	+4.8%	+7.5%

Μια εκτέλεση ανα νόημα
για την εκπαίδευση

Γνωστός Νοηματιστής

2-S-U	SU-noDSC	SU-Frame	SU-Segm
82.5	77.8	80.9	72
	+4.7%	+1.6%	+10.5%

BU400
94 νοήματα

Γλωσσικές-Φωνητικές Υπομονάδες (1)

DS-Lem

300 Νοήματα

Άγνωστος Νοηματιστής

Μέθοδος	Παράμετροι του ITA				Νοηματιστής						
	G-MP	I-MP	G-HS	I-HS	Όλγα			Κώστας			
SU-P	-	-	-	-	9	35.5	34.8	10.6	31.6	37.9	
SU-P+ITA	$G_{\{del,sub\}}$	FS	G_{ins}	FS	21.7	38.7	43.6	23.8	41	44.7	+12.9 +6.3 +7.8
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	FS	G_{ins}	FS	22	38.7	60	23.2	41	57	
SU-P+ITA	$G_{\{del,sub\}}$	FS	$G_{ins} + G_{sub}$	FS	21.7	42.4	46.6	23.8	42.8	46.2	
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	FS	$G_{ins} + G_{sub}$	FS	22	42.4	60.7	23.2	42.8	57.4	0 +2.7 +14.9
SU-P+ITA	$G_{\{del,sub\}}$	RAW	G_{ins}	FS	21.3	38.7	47.5	20	41	47	
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	RAW	G_{ins}	FS	26.2	38.7	61.6	27.8	41	58	
SU-P+ITA	$G_{\{del,sub\}}$	RAW	$G_{ins} + G_{sub}$	FS	21.3	42.4	49.7	20	42.8	48.8	
SU-P+ITA	$G_{\{del,sub\}} + G_{sub}$	RAW	$G_{ins} + G_{sub}$	FS	26.2	42.4	62.9	27.8	42.8	58.6	+4.4 0 +1.7

Σχετικές αυξήσεις

M-P HS M-P+HS

Γλωσσικές-Φωνητικές Υπομονάδες (2)

DS-Lem

300 Νοήματα

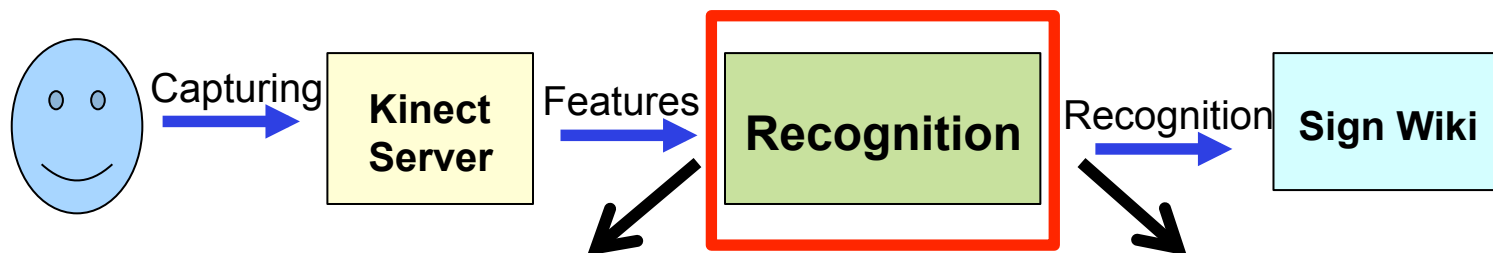
Άγνωστος Νοηματιστής

Νοηματιστής	Μέθοδοι				
	SU-P+ITA	2-S-U	SU-Segm	SU-Frame	Sign-DTW
Όλγα	62.9	61.2	54.4	40.5	57.9
Κώστας	58.6	50.1	32.6	35.5	46.3
		+5.1	+17.2	+22.7	+8.6

Προσαρμογή σε Νοηματιστή

	Νοηματιστής	Όλγα			Κώστας			
	Ροή	M-P	HS	M-P+HS	M-P	HS	M-P+HS	
Μέθοδοι	SU-P+ITA	28.5	42	62.6	27.1	44.3	57	
	SU-P+ITA+MLLR	37.8	60.9	76.3	34.8	66	78.4	+17.5
	SU-P+ITA+MLLR+IP	73	88	94.2	66.3	89.5	94	+16.7

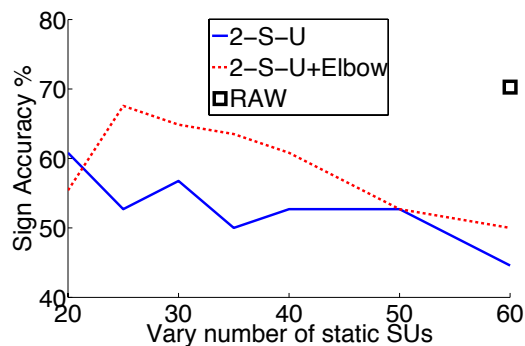
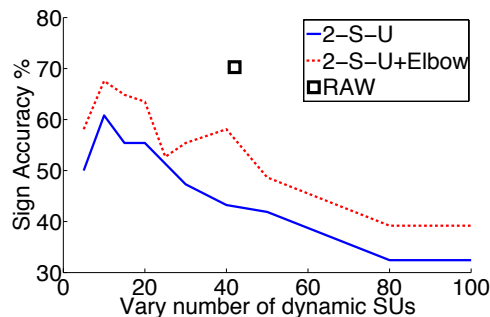
Αναγνώριση Συνεχούς ΕΝΓ (SignWiki-Demo)



GSL-Phrases
33 νοήματα
30 προτάσεις
1 νοηματιστής

Αποτελέσματα Αναγνώρισης DEMO

- RAW και 2-S-U υπομονάδες
- Κίνηση-Θέση



Βίντεο αναγνώρισης σε πραγματικό χρόνο



Πολυτροπική Αναγνώριση Χειρονομιών

ChaLearn
20 χειρονομίες
39 χρήστες

ΑΔ	Μεμονωμένες ροές			Σύμμιξη ροών		
	Φωνή	Σκελετικός	Χειρομορφή	Γραμματική	MHS	MHS+SPF
✗	78.4	47.6	13.3	✗	85.8	87.1
✓	87.2	49.1	20.2	✓	91.92	92.28
	+8.8	+1.5	+6.9		93.06	93.33

SS vs. Fusion **+7.4%**

Gram vs No-Gram **+1.14**

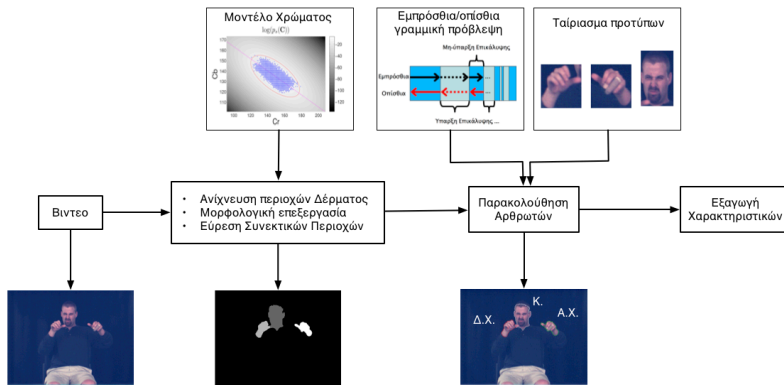
Θέση	Μέθοδος	Lev. Dist.	Acc.%	RER
-	MHS+SPF	0.0667	93.33	-
1	iva.mm [141]	0.12756	87.244	+47.6
2	wweight	0.15387	84.613	+56.6
3	E.T. [12]	0.17105	82.895	+60.9
4	MmM	0.17215	82.785	+61.2
5	pptk	0.17325	82.675	+61.4

- Ανίχνευση Πολυτροπικής Δράσης
- Εφαρμογή της SPF μεθόδου
- Εισαγωγή της Γραμματικής
- Αξιοποίηση του συνόλου δεδομένων επικύρωσης

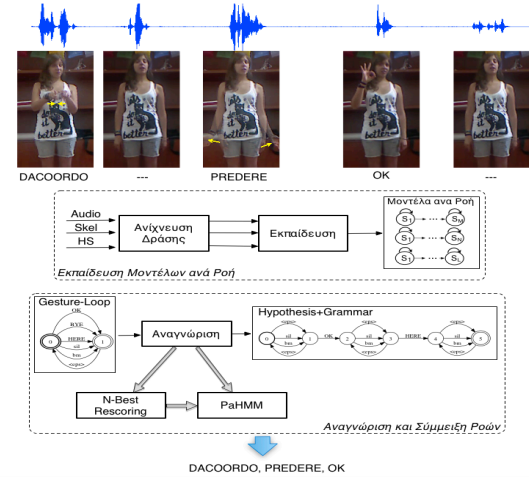
Παραδειγμα Πολυτροπικής Αναγνώρισης Χειρονομιών



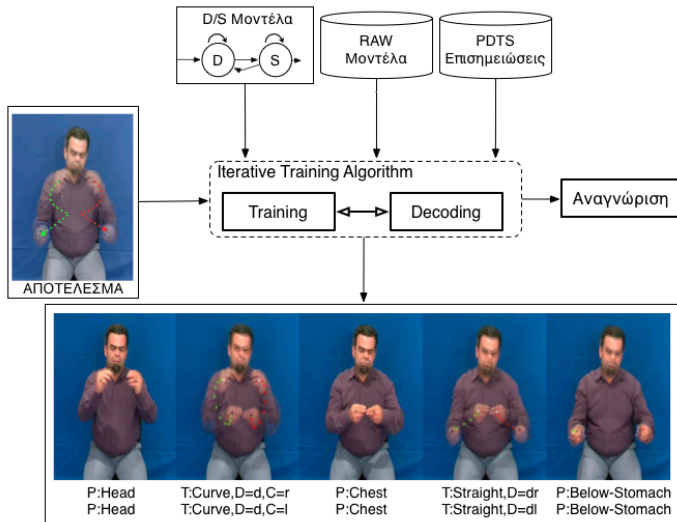
Εξαγωγή Χαρακτηριστικών



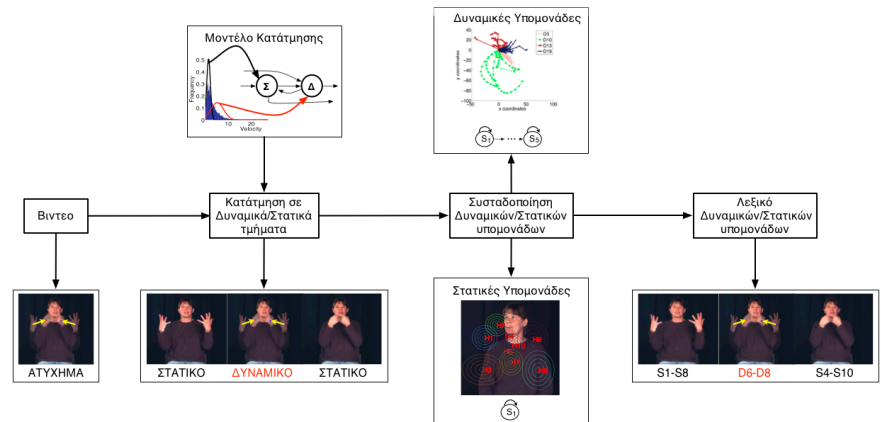
Πολυτροπική Αναγνώριση Χειρονομιών



Φωνητικές Υπομονάδες



Δεδομενοκεντρικές Υπομονάδες



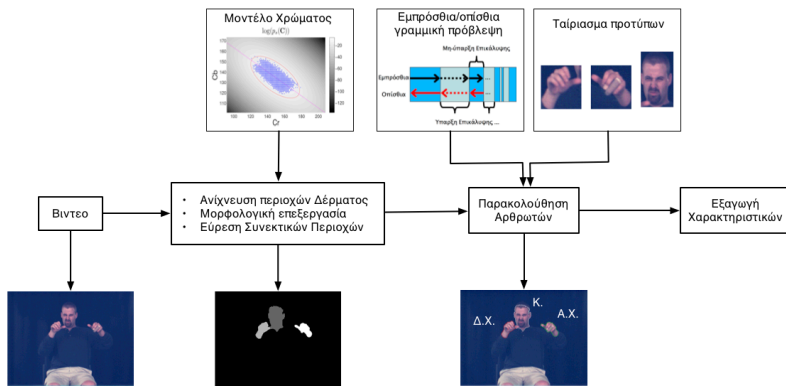
Μελλοντική Έρευνα

- Περαιτέρο συστηματική εξερεύνηση του προβλήματος αναγνώρισης συνεχούς νοηματικής γλώσσας.
- Σύμμιξη των πολλαπλών ροών πληροφορίας εκμεταλλευόμενοι πρότερη γλωσσολογική γνώση
- Εκμετάλευση επιπλέον ροών πληροφορίας όπως π.χ. εκφράσεις του προσώπου, στάση σώματος
- Αυτόματη φωνητική επισημείωση
- Εκμετάλευση στατιστικών εξαρτήσεων μεταξύ των πολλαπλών ροών πληροφορίας όπως π.χ. η φωνή σε σχέση με τις χειρονομίες

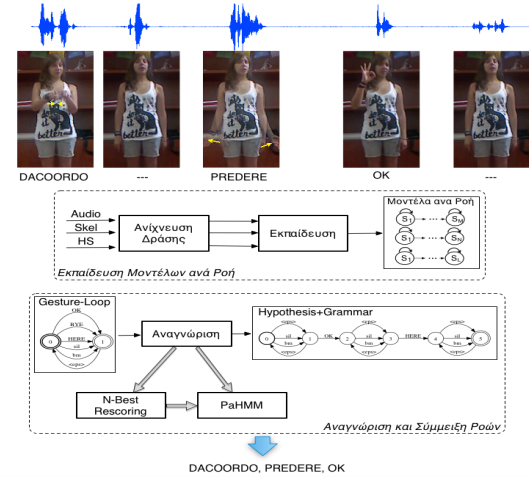
Δημοσιεύσεις Διατριβής

- Δημοσιεύσεις σε Διεθνή Περιοδικά με Κριτές
 - A. Roussos, S. Theodorakis, V. Pitsikalis, and P. Maragos. **Dynamic affine-invariant shape-appearance handshape features and classification in sign language videos**. The Journal of Machine Learning Research 14 (1), 1627-1663. 2013.
 - S. Theodorakis and V. Pitsikalis, and P. Maragos. **Dynamic–Static Unsupervised Sequentiality, Statistical Subunits and Lexicon for Sign Language Recognition**. Image and Vision Computing 32 (8), 533-549. 2014.
 - V. Pitsikalis, A. Katsamanis, S. Theodorakis, and P. Maragos. **Multimodal Gesture Recognition via Multiple Hypotheses Rescoring**. The Journal of Machine Learning Research (under review).
 - S. Theodorakis, V. Pitsikalis, and P. Maragos. **Linguistic Phonetic Subunits and Lexicon for Sign Language Recognition**. IEEE Transactions on Audio Speech and Language Processing (under submission).
- Επιλογή Δημοσιεύσεων σε Διεθνή Συνέδρια με Κριτές (11 δημοσιεύσεις συνολικά)
 - G. Pavlakos, S. Theodorakis, V. Pitsikalis, A. Katsamanis, and P. Maragos. **Kinect-Based Multimodal Gesture Recognition Using a Two-Pass Fusion Scheme**. In Proc. Int'l Conf. on Image Processing, 2014.
 - S. Theodorakis, V. Pitsikalis, I. Rodomagoulakis, and P. Maragos. **Recognition with raw canonical phonetic movement and handshape subunits on videos of continuous sign language**. In Proc. Int'l Conf. on Image Processing, 2012.
 - V. Pitsikalis, S. Theodorakis, C. Vogler, and P. Maragos. **Advances in phonetics-based sub-unit modeling for transcription alignment and sign language recognition**. In Proc. CVPR Workshop Gesture Recognition, 2011. (**Best Paper Award**)

Εξαγωγή Χαρακτηριστικών

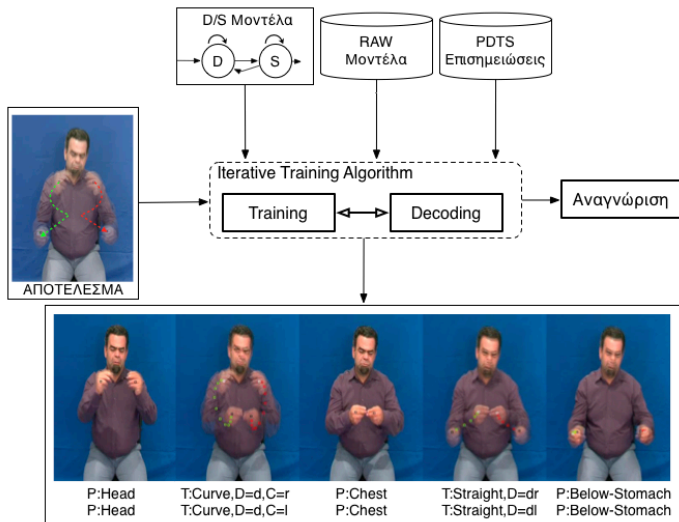


Πολυτροπική Αναγνώριση Χειρονομιών

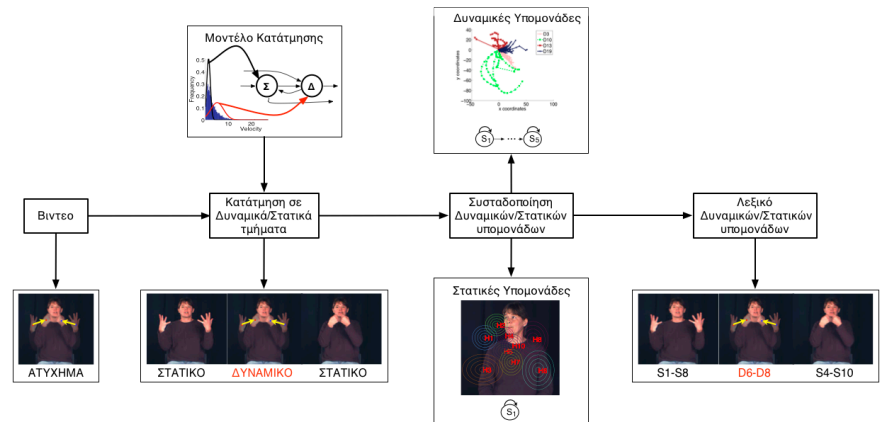


Ευχαριστώ πολύ

Φωνητικές Υπομονάδες



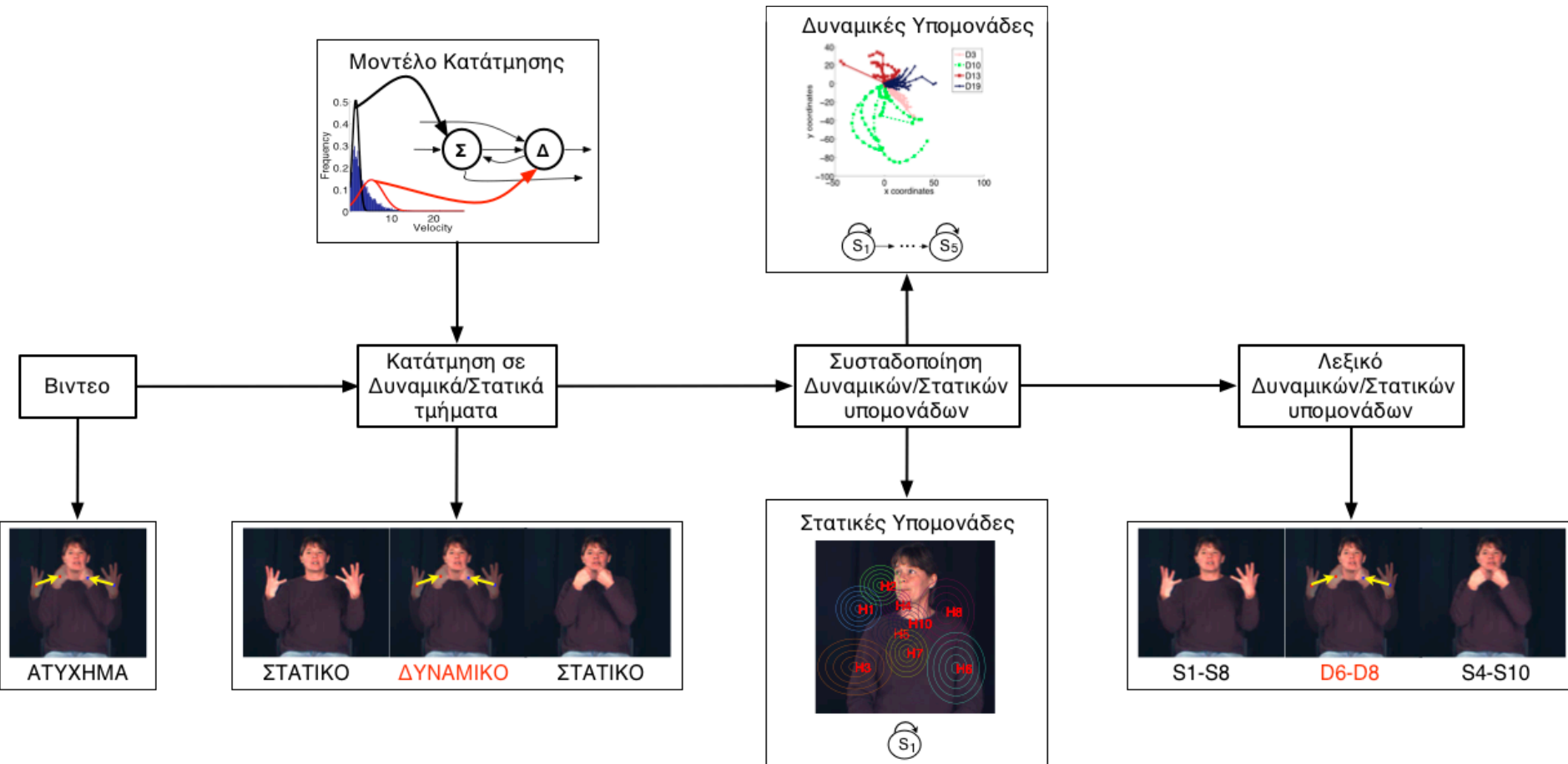
Δεδομενοκεντρικές Υπομονάδες



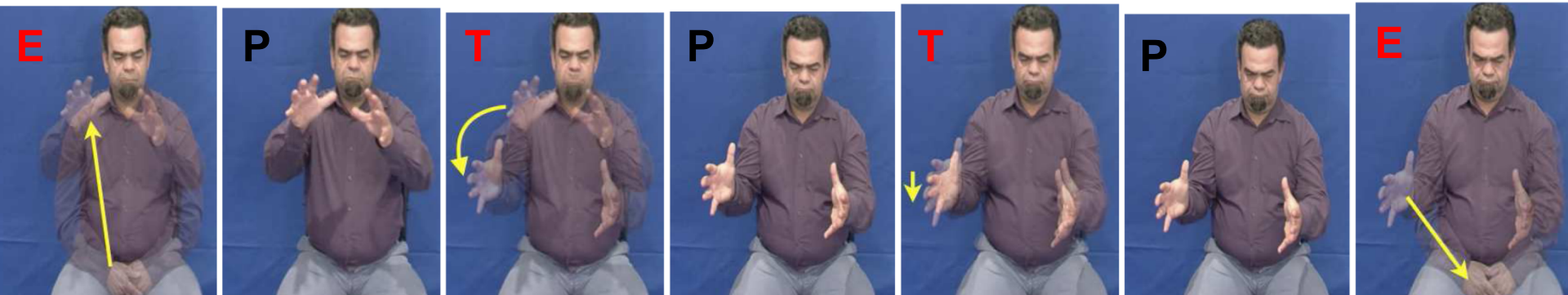


Apendix

Σύνοψη συστήματος Αναγνώρισης ΝΓ με δεδομοκεντρικές υπομονάδες



Φωνητικές Υπομονάδες



Transition/Epenthesis Segments

Superimposed Initial-End Frames + Arrow

Posture/Detention Segments

Single Frame

Frames	Type	PDTS label
1:12	E	rest-position — location-head
13:13	P	location-head
14:25	T	directedmotion, curve-r, direction-o, second-direction-do, tense-true
26:27	P	location-torso, side=right_beside
28:50	T	directedmotion, direction-dr, small
51:51	P	location-torso, side=right_beside_down
52:66	E	location-torso, side=right_beside_down — rest-position

Συνοψη Πειραμάτων

- Ταξινόμηση Χειρομορφών
- Αναγνώριση ΝΓ με δεδομενοκεντρικές υπομονάδες
- Αναγνώριση ΝΓ με γλωσσικές-φωνητικές υπομονάδες
- Πολυτροπική Αναγνώριση Χειρονομιών

Απεικόνιση των PDTS υπομονάδων

Posture

Trans

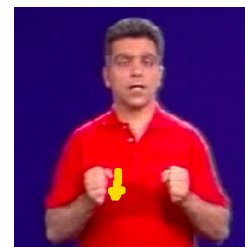
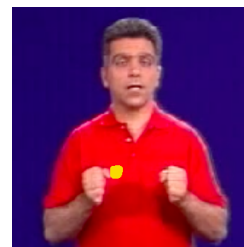
Posture

Trans

Posture

Trans

Det



shoulder

Straight
down

chest

Back up

shoulder

Straight
down

chest

Σύνοψη Σχετικής Έρευνας

Works	Sensor/FE	SU Segm.	SU Constr.	Modeling	Linguistic Transcriptions	SU Seq.	Parallel Fusion	Unseen Signer
[136]	Vis.	X	X	exemplar based (DTW)	X	X	X	✓
[150]	Vis.	X	X	HMMs	X	X	X	✓
[135]	Vis.	X	X	HMMs/RNN	X	X	X	n.a.
[114]	Vis.	X	X	HMMs	X	X	X	n.a.
[45]	d-gloves	X	X	SRN/HMMs	X	X	X	✓
[6]	Vis.	X	X	Multi-Class Fisher Score	X	X	X	✓
[146]	d-gloves	SBHMMs	DIST	HMMs	X	X	X	n.a.
[54]	Vis.	X	X	MLP/HMM	X	X	X	X
[98]	MoCap	X	X	DBN/MH-HMM	X	X	✓	X
[55]	Vis.	motion disc.	DTW	WC/Adaboost	X	X	X	X
[39]	Vis.	X	X	Tree-based	X	X	X	✓
[28]	Vis.	rule-based	WC	SP,MC	X	X	X	✓
[66]	Vis.	rule-based	WC	MC	X	X	X	X
[11]	Vis.+c-gloves	K-means	K-means	HMM	X	X	X	n.a.
[47]	d-gloves	LR-HMM	MKM-DTW	HMM	X	X	X	n.a.
[46]	d-gloves	LR-HMM	MKM-DTW	HMM + Epenthesis	X	X	X	n.a.
[72]	MoCap	rule-based	K-means	HMM	X	X	X	X
[9]	Vis.	motion disc.	DTW, hier. clust.	Adaboost	X	X	X	X
[124]	Vis.	2S-ERG HMM	K-means, DTW	MSSD-HMM, PaHMM	X	✓	✓	✓
[134]	MoCap	X	X	PaHMM + Epenthesis	M-H Model	✓	✓	n.a.
[71]	Vis.	X	X	HMMs	SignWriting	X	X	n.a.
[100]	Vis.	X	X	HMMs	PDTS	✓	X	X
[4]	Vis.	X	X	MSSD-HMM, PaHMM	PDTS	✓	✓	✓

Σύνοψη Σχετικής Έρευνας

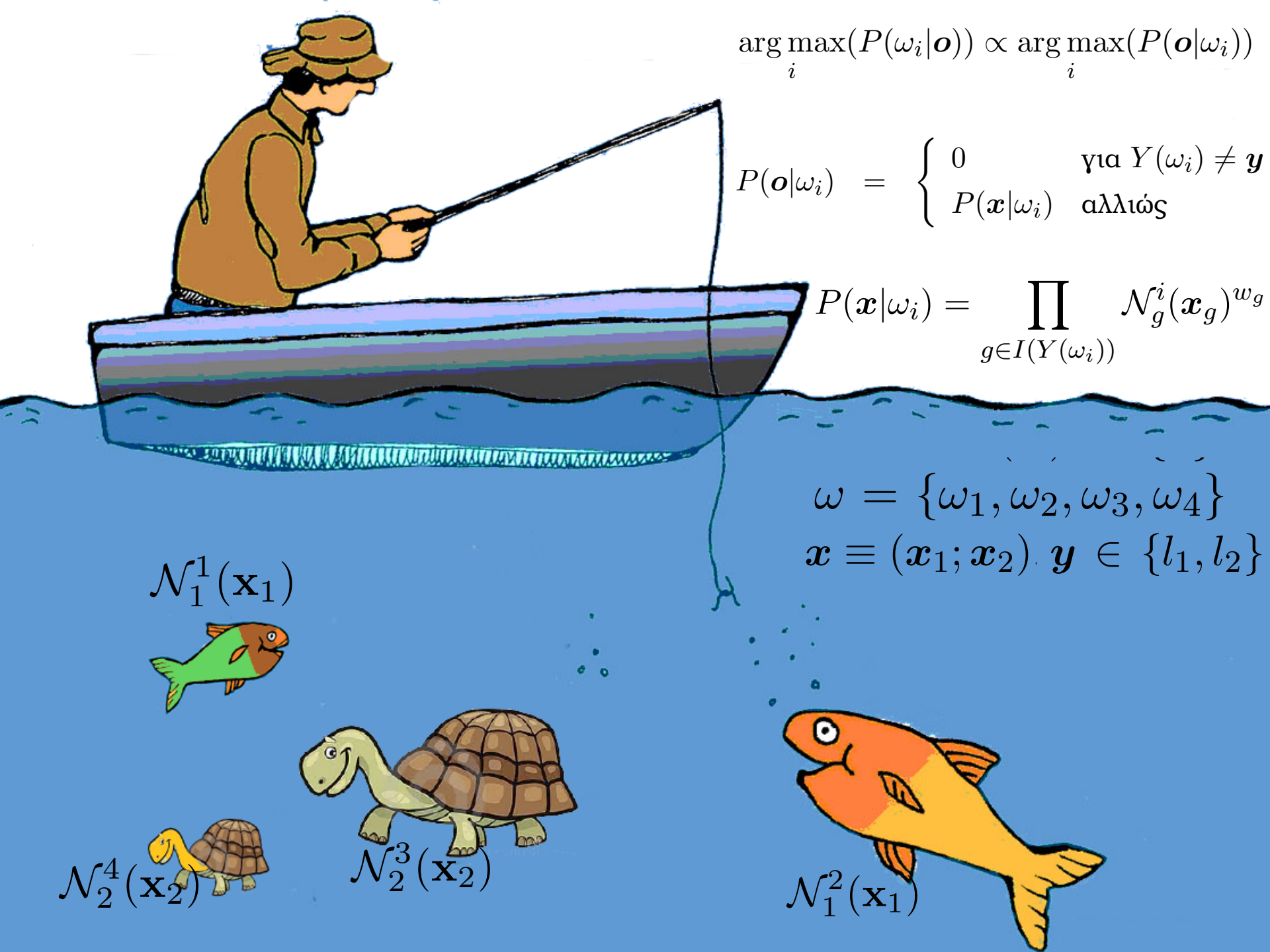
Works	Sensor/FE	SU Segm.	SU Constr.	Modeling	Linguistic Transcriptions	SU Seq.	Parallel Fusion	Unseen Signer
[136]	Vis.	X	X	exemplar based (DTW)	X	X	X	✓
[150]	Vis.	X	X	HMMs	X	X	X	✓
[135]	Vis.	X	X	HMMs/RNN	X	X	X	n.a.
[114]	Vis.	X	X	HMMs	X	X	X	n.a.
[45]	d-gloves	X	X	SRN/HMMs	X	X	X	✓
[6]	Vis.	X	X	Multi-Class Fisher Score	X	X	X	✓
[146]	d-gloves	SBHMMs	DIST	HMMs	X	X	X	n.a.
[54]	Vis.	X	X	MLP/HMM	X	X	X	X
[98]	MoCap	X	X	DBN/MH-HMM	X	X	✓	X
[55]	Vis.	motion disc.	DTW	WC/Adaboost	X	X	X	X
[39]	Vis.	X	X	Tree-based	X	X	X	✓
[28]	Vis.	rule-based	WC	SP,MC	X	X	X	✓
[66]	Vis.	rule-based	WC	MC	X	X	X	X
[11]	Vis.+c-gloves	K-means	K-means	HMM	X	X	X	n.a.
[47]	d-gloves	LR-HMM	MKM-DTW	HMM	X	X	X	n.a.
[46]	d-gloves	LR-HMM	MKM-DTW	HMM + Epenthesis	X	X	X	n.a.
[72]	MoCap	rule-based	K-means	HMM	X	X	X	X
[9]	Vis.	motion disc.	DTW, hier. clust.	Adaboost	X	X	X	X
[124]	Vis.	2S-ERG HMM	K-means, DTW	MSSD-HMM, PaHMM	X	✓	✓	✓
[134]	MoCap	X	X	PaHMM + Epenthesis	M-H Model	✓	✓	n.a.
[71]	Vis.	X	X	HMMs	SignWriting	X	X	n.a.
[100]	Vis.	X	X	HMMs	PDTS	✓	X	X
[4]	Vis.	X	X	MSSD-HMM, PaHMM	PDTS	✓	✓	✓

Σύνοψη Σχετικής Έρευνας

Works	Sensor/FE	SU Segm.	SU Constr.	Modeling	Linguistic Transcriptions	SU Seq.	Parallel Fusion	Unseen Signer
[136]	Vis.	X	X	exemplar based (DTW)	X	X	X	✓
[150]	Vis.	X	X	HMMs	X	X	X	✓
[135]	Vis.	X	X	HMMs/RNN	X	X	X	n.a.
[114]	Vis.	X	X	HMMs	X	X	X	n.a.
[45]	d-gloves	X	X	SRN/HMMs	X	X	X	✓
[6]	Vis.	X	X	Multi-Class Fisher Score	X	X	X	✓
[146]	d-gloves	SBHMMs	DIST	HMMs	X	X	X	n.a.
[54]	Vis.	X	X	MLP/HMM	X	X	X	X
[98]	MoCap	X	X	DBN/MH-HMM	X	X	✓	X
[55]	Vis.	motion disc.	DTW	WC/Adaboost	X	X	X	X
[39]	Vis.	X	X	Tree-based	X	X	X	✓
[28]	Vis.	rule-based	WC	SP,MC	X	X	X	✓
[66]	Vis.	rule-based	WC	MC	X	X	X	X
[11]	Vis.+c-gloves	K-means	K-means	HMM	X	X	X	n.a.
[47]	d-gloves	LR-HMM	MKM-DTW	HMM	X	X	X	n.a.
[46]	d-gloves	LR-HMM	MKM-DTW	HMM + Epenthesis	X	X	X	n.a.
[72]	MoCap	rule-based	K-means	HMM	X	X	X	X
[9]	Vis.	motion disc.	DTW, hier. clust.	Adaboost	X	X	X	X
[124]	Vis.	2S-ERG HMM	K-means, DTW	MSSD-HMM, PaHMM	X	✓	✓	✓
[134]	MoCap	X	X	PaHMM + Epenthesis	M-H Model	✓	✓	n.a.
[71]	Vis.	X	X	HMMs	SignWriting	X	X	n.a.
[100]	Vis.	X	X	HMMs	PDTS	✓	X	X
[4]	Vis.	X	X	MSSD-HMM, PaHMM	PDTS	✓	✓	✓

Σύνοψη Σχετικής Έρευνας

Works	Sensor/FE	SU Segm.	SU Constr.	Modeling	Linguistic Transcriptions	SU Seq.	Parallel Fusion	Unseen Signer
[136]	Vis.	X	X	exemplar based (DTW)	X	X	X	✓
[150]	Vis.	X	X	HMMs	X	X	X	✓
[135]	Vis.	X	X	HMMs/RNN	X	X	X	n.a.
[114]	Vis.	X	X	HMMs	X	X	X	n.a.
[45]	d-gloves	X	X	SRN/HMMs	X	X	X	✓
[6]	Vis.	X	X	Multi-Class Fisher Score	X	X	X	✓
[146]	d-gloves	SBHMMs	DIST	HMMs	X	X	X	n.a.
[54]	Vis.	X	X	MLP/HMM	X	X	X	X
[98]	MoCap	X	X	DBN/MH-HMM	X	X	✓	X
[55]	Vis.	motion disc.	DTW	WC/Adaboost	X	X	X	X
[39]	Vis.	X	X	Tree-based	X	X	X	✓
[28]	Vis.	rule-based	WC	SP,MC	X	X	X	✓
[66]	Vis.	rule-based	WC	MC	X	X	X	X
[11]	Vis.+c-gloves	K-means	K-means	HMM	X	X	X	n.a.
[47]	d-gloves	LR-HMM	MKM-DTW	HMM	X	X	X	n.a.
[46]	d-gloves	LR-HMM	MKM-DTW	HMM + Epenthesis	X	X	X	n.a.
[72]	MoCap	rule-based	K-means	HMM	X	X	X	X
[9]	Vis.	motion disc.	DTW, hier. clust.	Adaboost	X	X	X	X
[124]	Vis.	2S-ERG HMM	K-means, DTW	MSSD-HMM, PaHMM	X	✓	✓	✓
[134]	MoCap	X	X	PaHMM + Epenthesis	M-H Model	✓	✓	n.a.
[71]	Vis.	X	X	HMMs	SignWriting	X	X	n.a.
[100]	Vis.	X	X	HMMs	PDTS	✓	X	X
[4]	Vis.	X	X	MSSD-HMM, PaHMM	PDTS	✓	✓	✓



$$\arg \max_i (P(\omega_i | \mathbf{o})) \propto \arg \max_i (P(\mathbf{o} | \omega_i))$$

$$P(\mathbf{o} | \omega_i) = \begin{cases} 0 & \text{για } Y(\omega_i) \neq \mathbf{y} \\ P(\mathbf{x} | \omega_i) & \text{αλλιώς} \end{cases}$$

$$P(\mathbf{x} | \omega_i) = \prod_{g \in I(Y(\omega_i))} \mathcal{N}_g^i(\mathbf{x}_g)^{w_g}$$

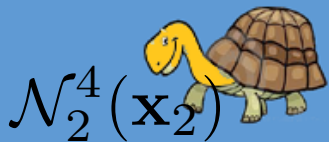
$$\omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$$

$$\mathbf{x} \equiv (\mathbf{x}_1; \mathbf{x}_2), \mathbf{y} \in \{l_1, l_2\}$$

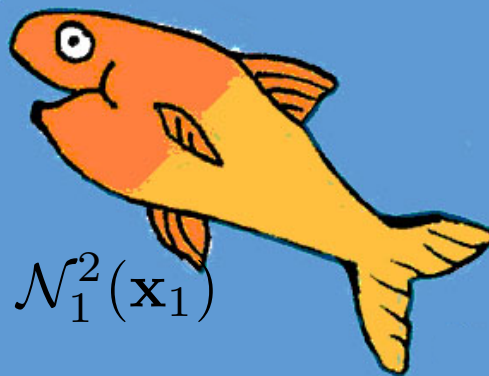
$$\mathcal{N}_1^1(\mathbf{x}_1)$$



$$\mathcal{N}_2^3(\mathbf{x}_2)$$

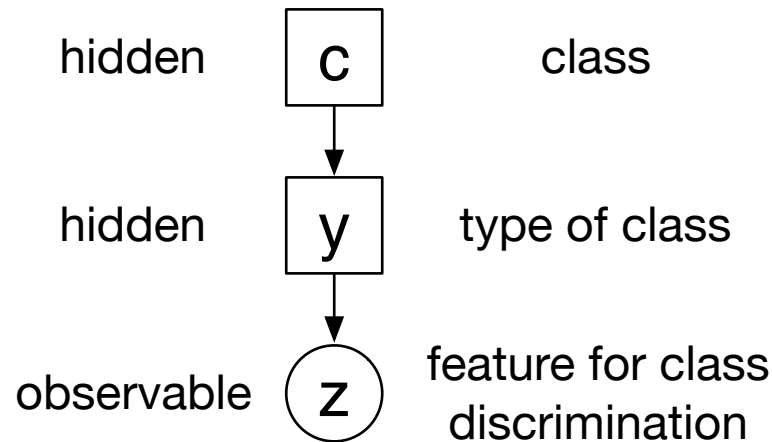


$$\mathcal{N}_2^4(\mathbf{x}_2)$$



$$\mathcal{N}_1^2(\mathbf{x}_1)$$

Multi-Stream Switching Probability Distribution



$$P(\mathbf{z}|\omega_i) = \sum_y (P(\mathbf{z}|\mathbf{y})P(\mathbf{y}|\omega_i)) = P(\mathbf{z}|Y(\omega_i)) = \mathcal{N}^{Y(\omega_i)}(\mathbf{z})$$