Μελέτη και χρήση ακουστικής πληροφορίας για τον εντοπισμό επιβλαβούς περιεχομένου και ενσωμάτωση με οπτική πληροφορία

Γιαννακόπουλος Θεόδωρος

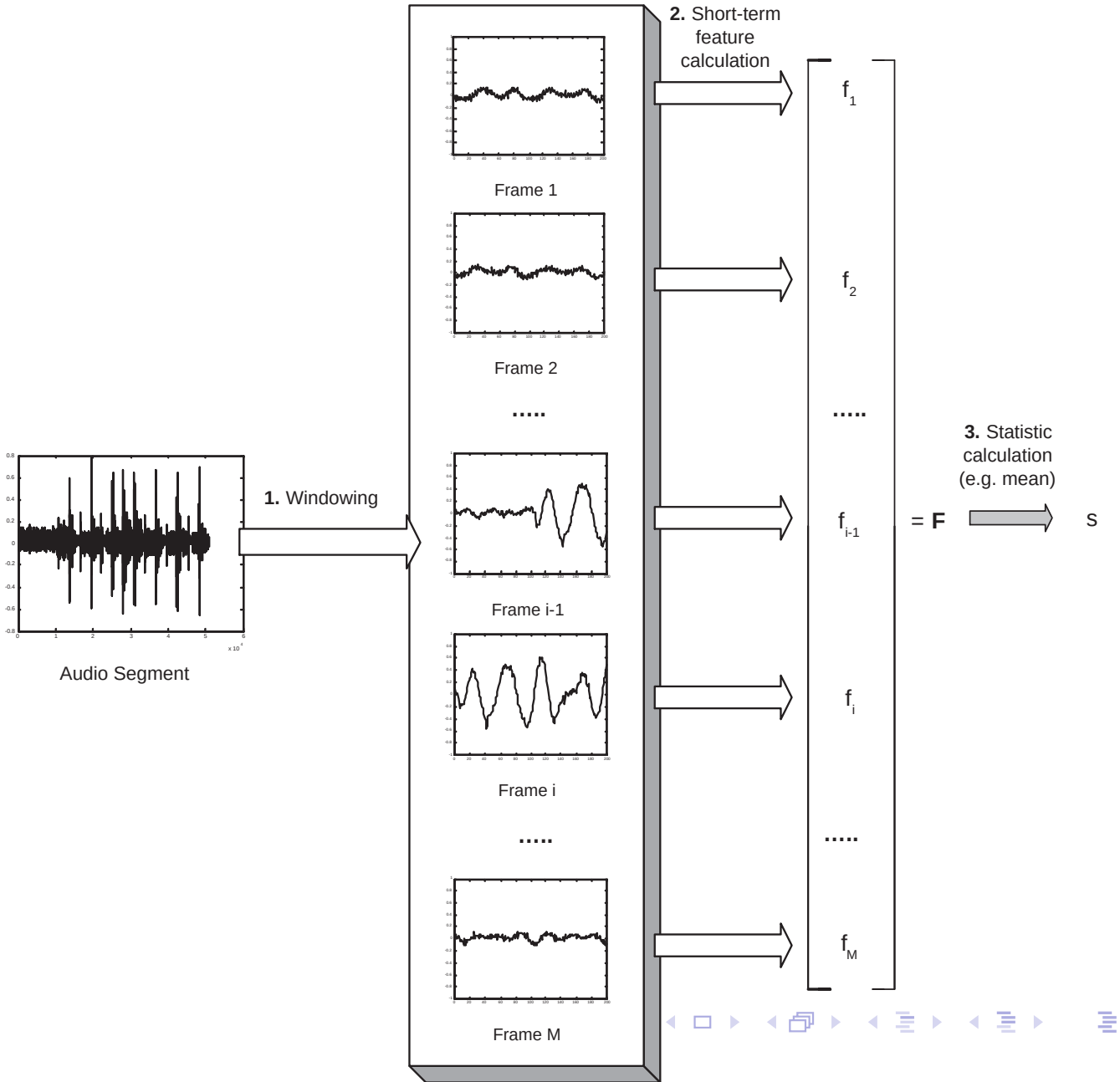
Παρασκευή 17 Ιουλίου 2009

Σκοπός

- **Κατάτμηση** και **ταξινόμηση** ηχητικών σημάτων, με βάση το **περιεχόμενο**.
- Έμφαση στην ανάλυση του περιεχομένου **ταινιών** (πολλαπλές κλάσεις) και **ραδιοφωνικών εκπομπών** (2 κλάσεις)
- Εντοπισμός ηχητικών κατηγοριών σχετικών με βίαιο περιεχόμενο (στις ταινίες).
- Ανάλυση συναισθηματικού περιεχομένου σε ταινίες

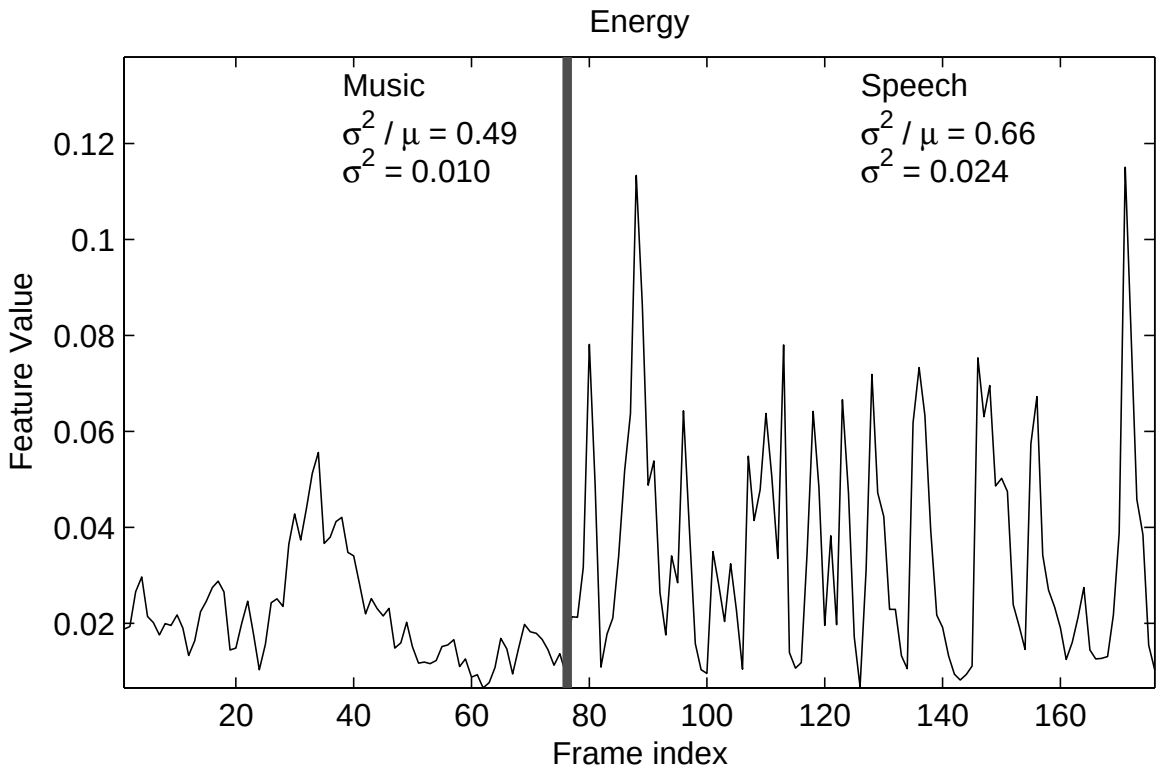
Περιεχόμενα

Long-term εξαγωγή ηχητικών χαρακτηριστικών



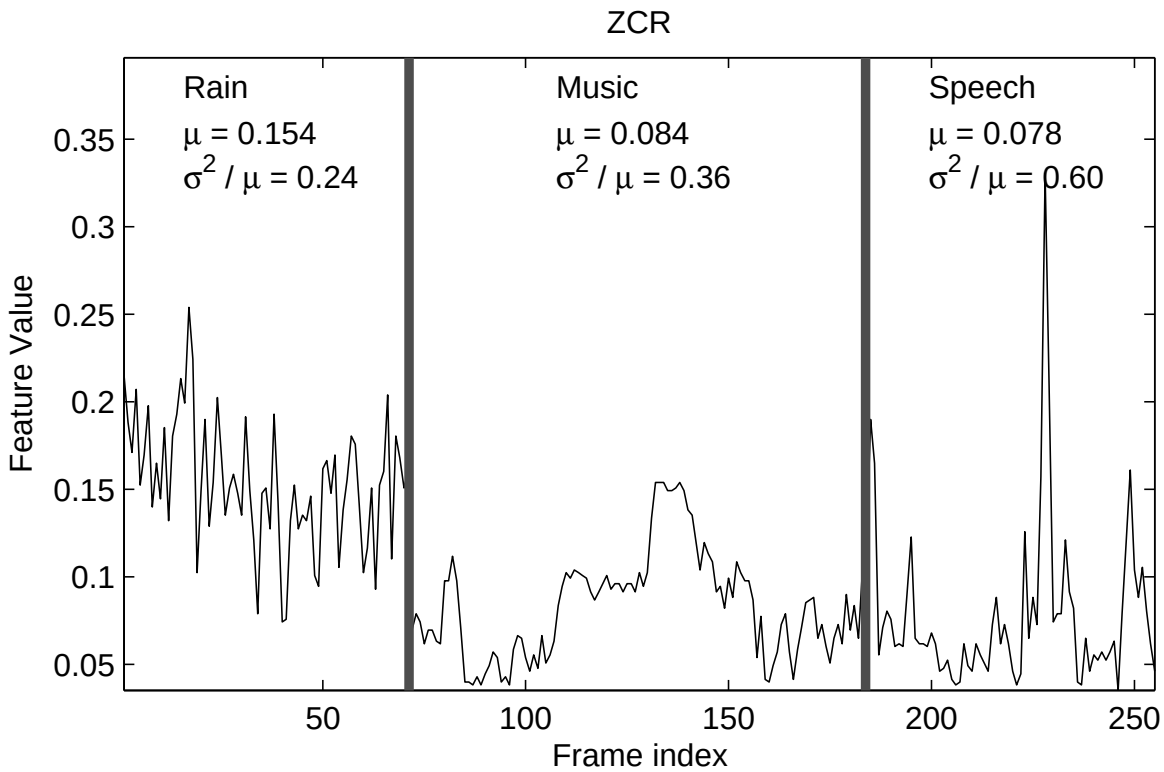
Ενέργεια σήματος

- $E(i) = \frac{1}{N} \sum_{n=1}^{N} |x_i(n)|^2$
- Παράδειγμα (σήμα μουσικής και ομιλίας):



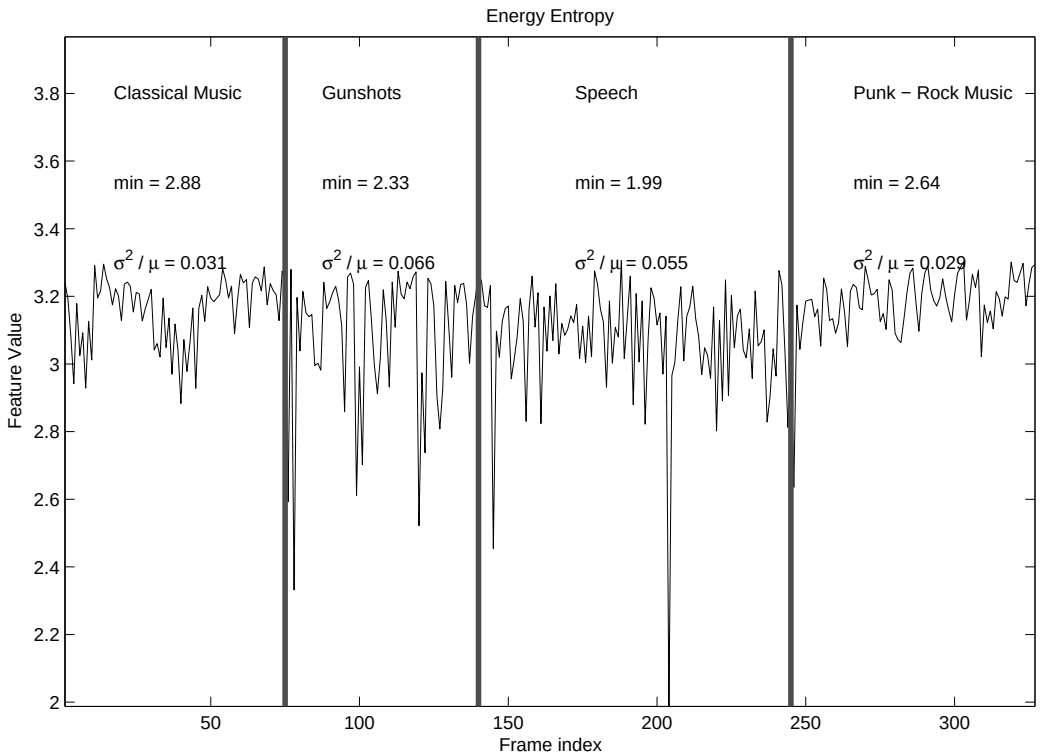
Ρυθμός διέλευσης από το μηδέν

- Zero Crossing Rate
- $Z(i) = \frac{1}{2N} \sum_{n=1}^{N} |sgn[x_i(n)] - sgn[x_i(n-1)]|$
- Παράδειγμα (σήμα: ήχος βροχής, μουσικής και ομιλίας):



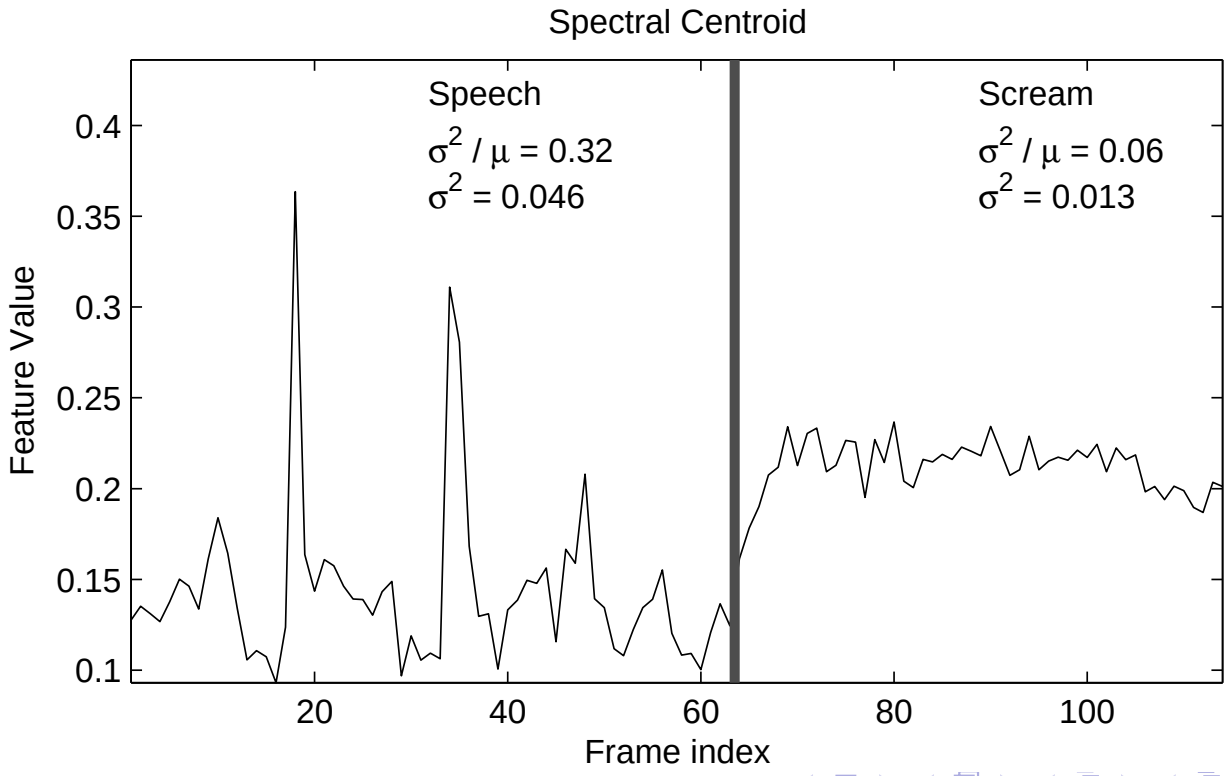
Εντροπία ενέργειας

- Μέτρο του βαθμού αλλαγών στην ενέργεια ενός σήματος
- Κάθε frame χωρίζεται σε subframes
- Για κάθε subframe : $e_j^2 = \frac{E_{subFrame_j}}{E_{shortFrame_i}}$
- Εντροπία: $H(i) = -\sum_{j=1}^{K} e_j^2 \cdot log_2(e_j^2)$
- Παράδειγμα (σήμα: κλασσικής μουσικής, πυροβολισμού, ομιλίας και punk-rock μουσικής):



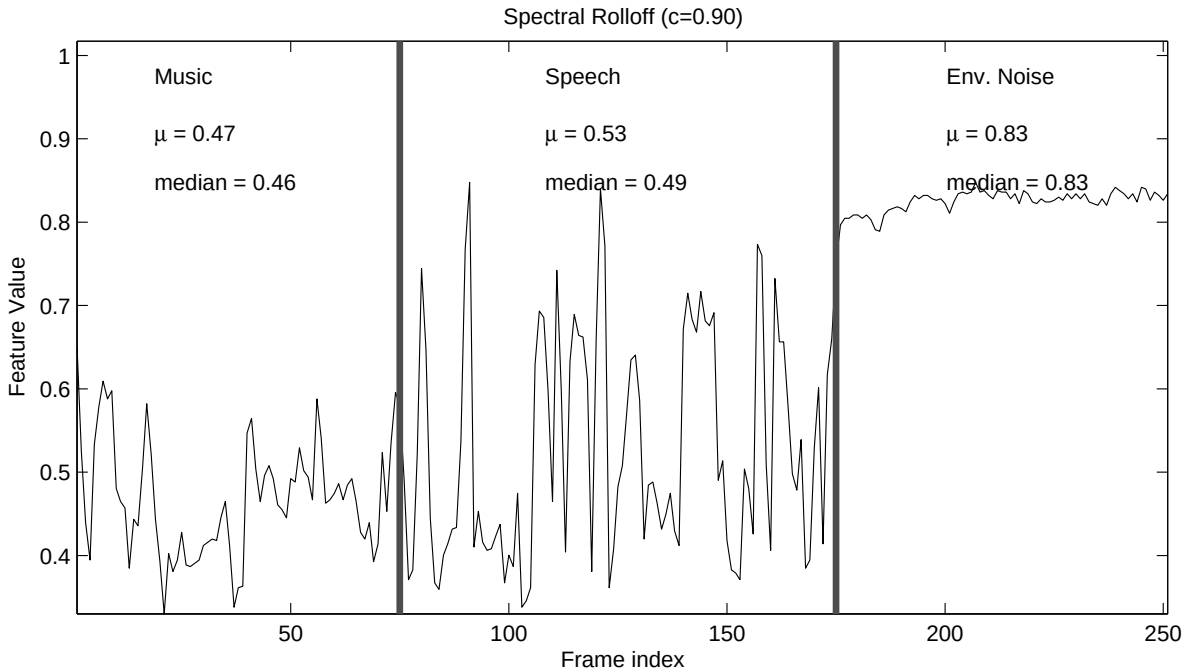
Φασματικό Κεντροϊδές

- Κέντρο βάρους του φάσματος
- $C_i = \frac{\sum_{k=1}^{N}(k+1)X_i(k)}{\sum_{k=1}^{N}X_i(k)}$
- Μέτρο της φασματικής θέσης
- Υψηλές διακυμάνσεις για σήματα ομιλίας
- Παράδειγμα ακολουθίας για ομιλία και κραυγή



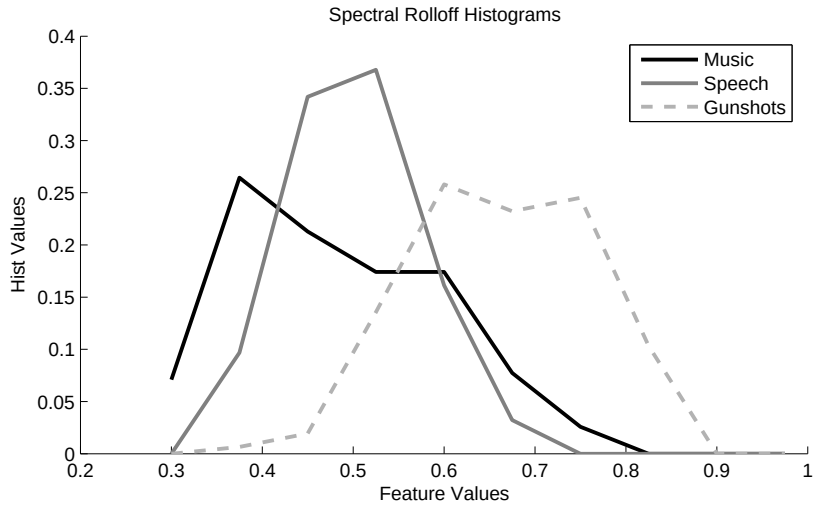
Φασματική Συγκέντρωση (1)

- Spectral Rolloff
- Η συχνότητα, κάτω από την οποία, συγκεντρώνεται το $C\%$ της φασματικής ενέργειας.
- Παράδειγμα ακολουθίας για μουσική, ομιλία και περιβαλλοντικό θόρυβο.

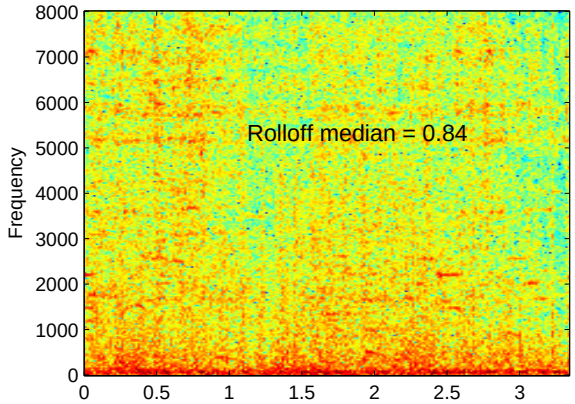


Φασματική Συγκέντρωση (2)

- Ιστογράμματα Median τιμών φασματικής συγκέντρωσης:

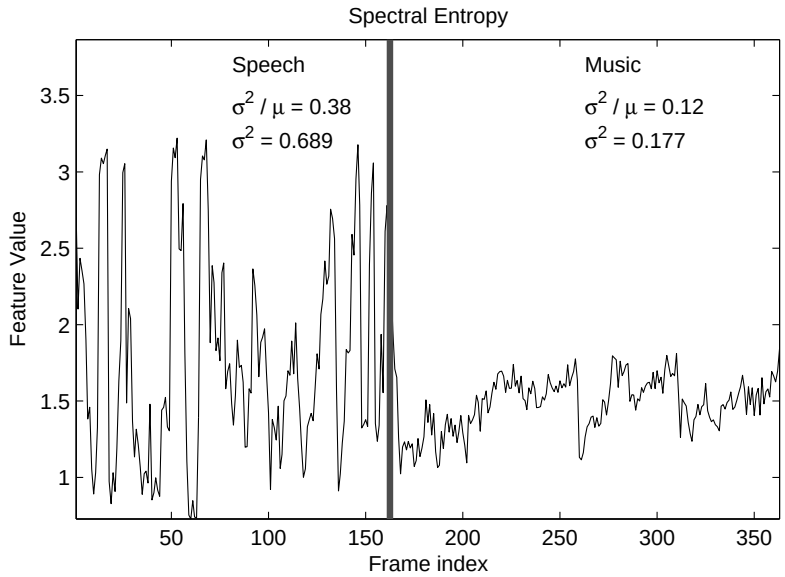


- 96% των πυροβολισμών παρουσιάζουν τιμές ≥ 0.5.
- Φασματογράφημα πυροβολισμών:



Φασματική Εντροπία

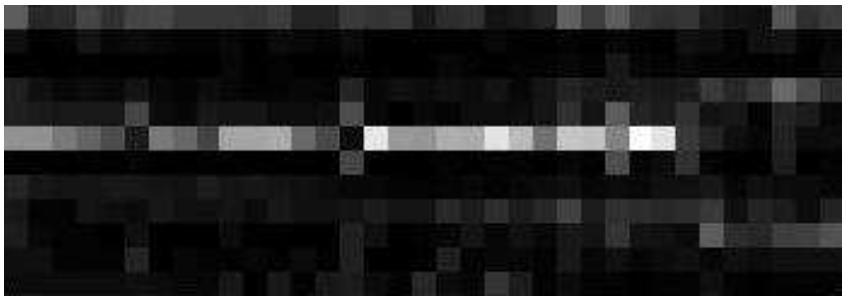
- Το φάσμα κάθε frame διαιρείται σε N υποσυχνότητες (bins)
- Για κάθε frame $n_f = \frac{E_f}{\sum_{i=0}^{L-1} E_f}$, $f = 0, \ldots, L-1$
- Εντροπία: $H = -\sum_{f=0}^{L-1} n_i \cdot log_2(n_i)$
- Φασματική εντροπία, για ομιλία και μουσική



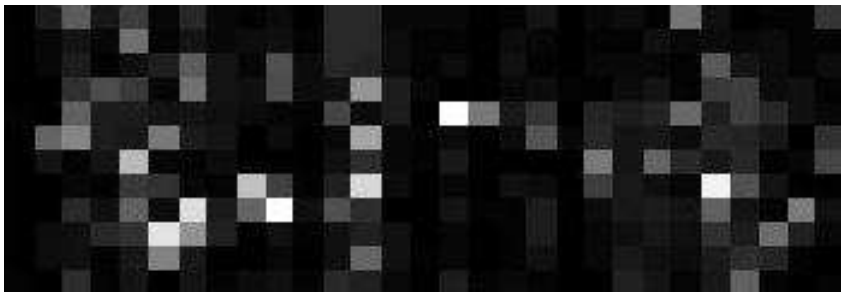
- Παραλλαγή: 'χρωματική εντροπία'

Χρωματικά χαρακτηριστικά (1)

- Χρωματικό διάνυσμα (**Chroma Vector - cv**): 12-διάστατη αναπαράσταση της φασματικής ενέργειας (κάθε στοιχείο αντιστοιχεί σε τονικές κλάσεις - 12 νότες δυτικής μουσικής).
- Αν το cv υπολογιστεί για κάθε frame : chromagram .



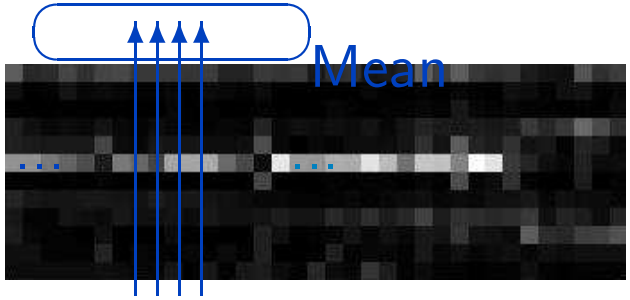
Σχήμα: Μουσική

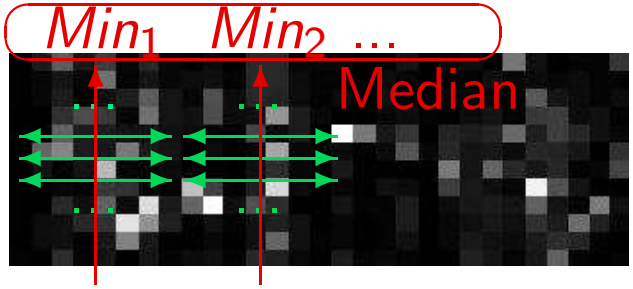


Σχήμα: Ομιλία

Χρωματικά χαρακτηριστικά (2)

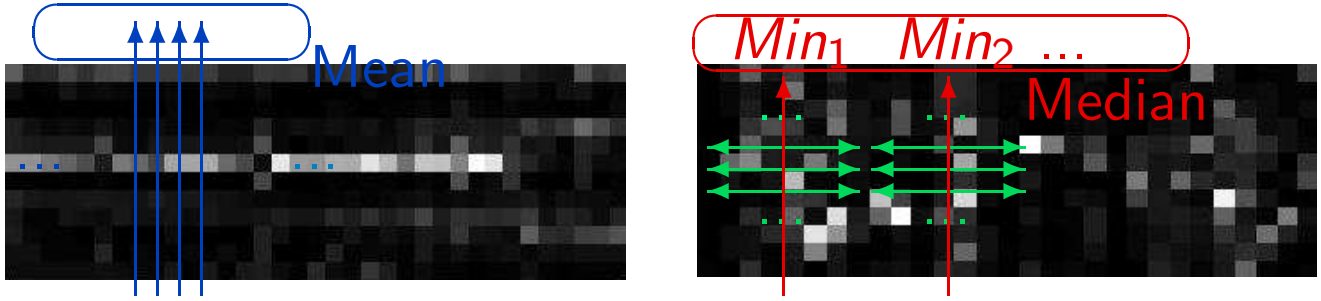
- 2 χαρακτηριστικά:





Χρωματικά χαρακτηριστικά (2)

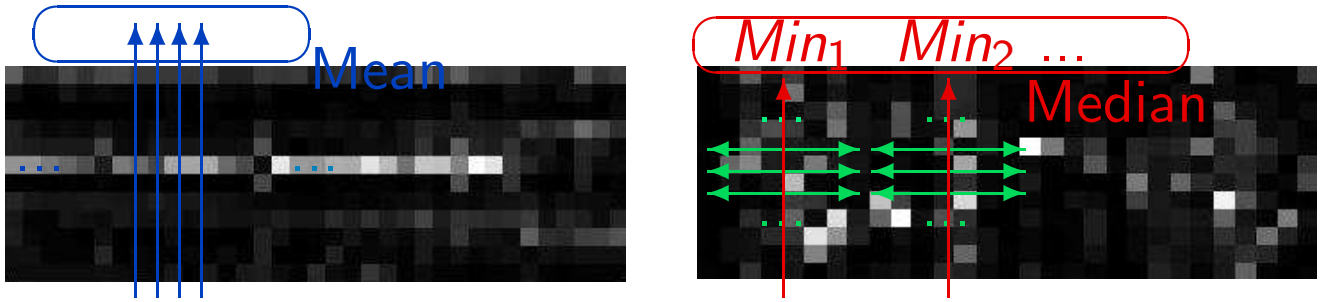
- 2 χαρακτηριστικά:



- 1ο **χαρακτηριστικό**: Υπολογισμός του STD των συντελεστών και έπειτα της **μέσης τιμής** αυτής της ακολουθίας (100ms frames).

Χρωματικά χαρακτηριστικά (2)

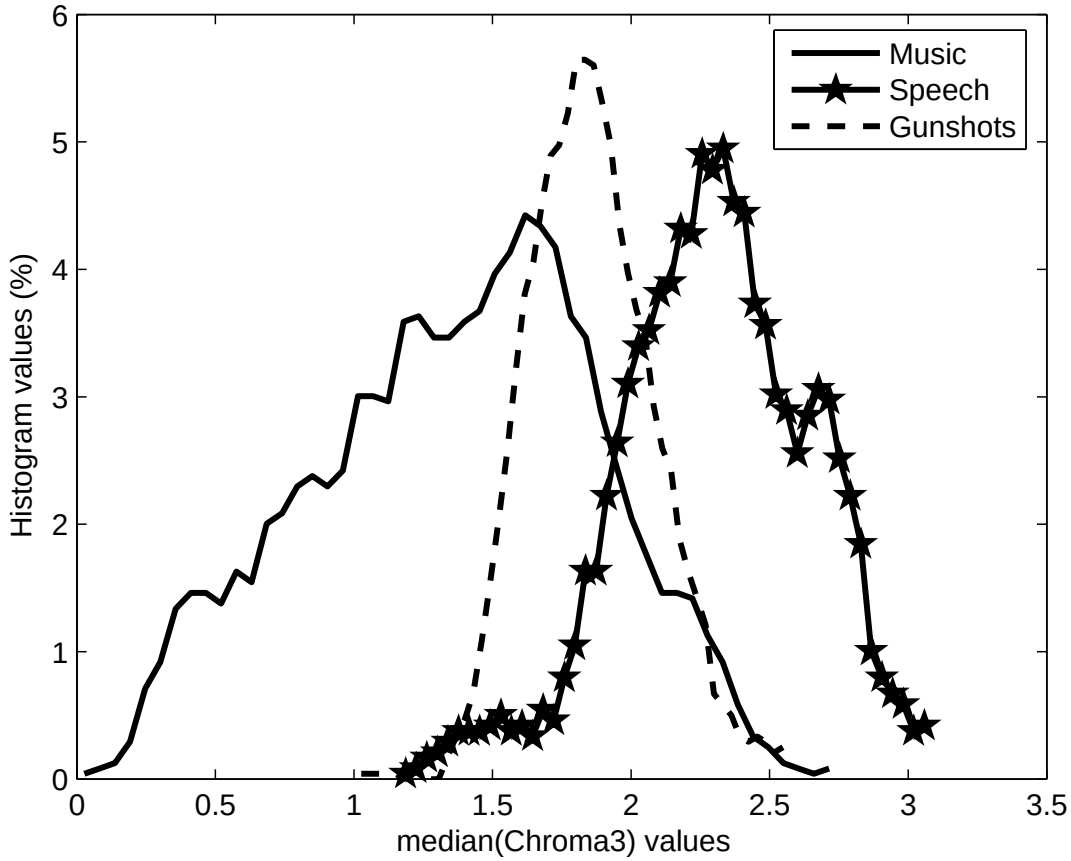
- 2 χαρακτηριστικά:



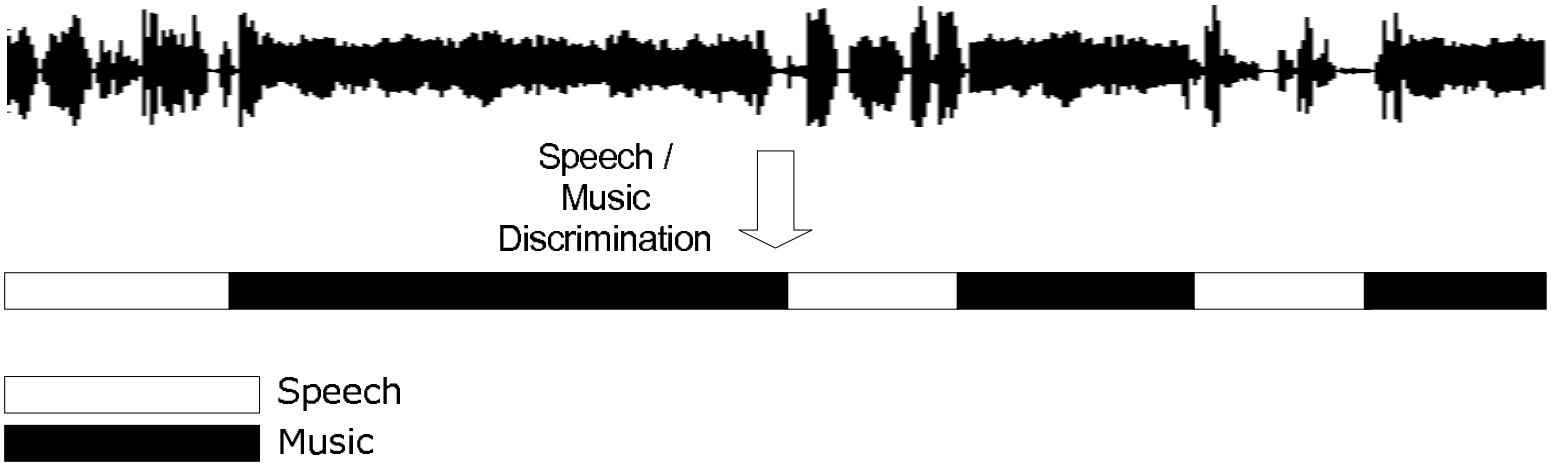
- 1ο **χαρακτηριστικό**: Υπολογισμός του STD των συντελεστών και έπειτα της **μέσης τιμής** αυτής της ακολουθίας (100ms frames).
- 2ο **χαρακτηριστικό**: Για κάθε 200ms midterm frame , υπολογισμός του STD (για κάθε συντελεστή - 20ms short-term frame). Η ελάχιστες τιμές αποθηκεύονται και η **median** αυτής τις ακολουθίας είναι το τελικό χαρακτηριστικό.

Χρωματικά χαρακτηριστικά (3)

- Ιστογράμματα του 2ου χαρακτηριστικού:



- Κατάτμηση ενός ηχητικού σήματος και ταξινόμηση των επί μέρους τμημάτων σαν μουσική ή ομιλία.



Γενικά: Μεθοδολογίες

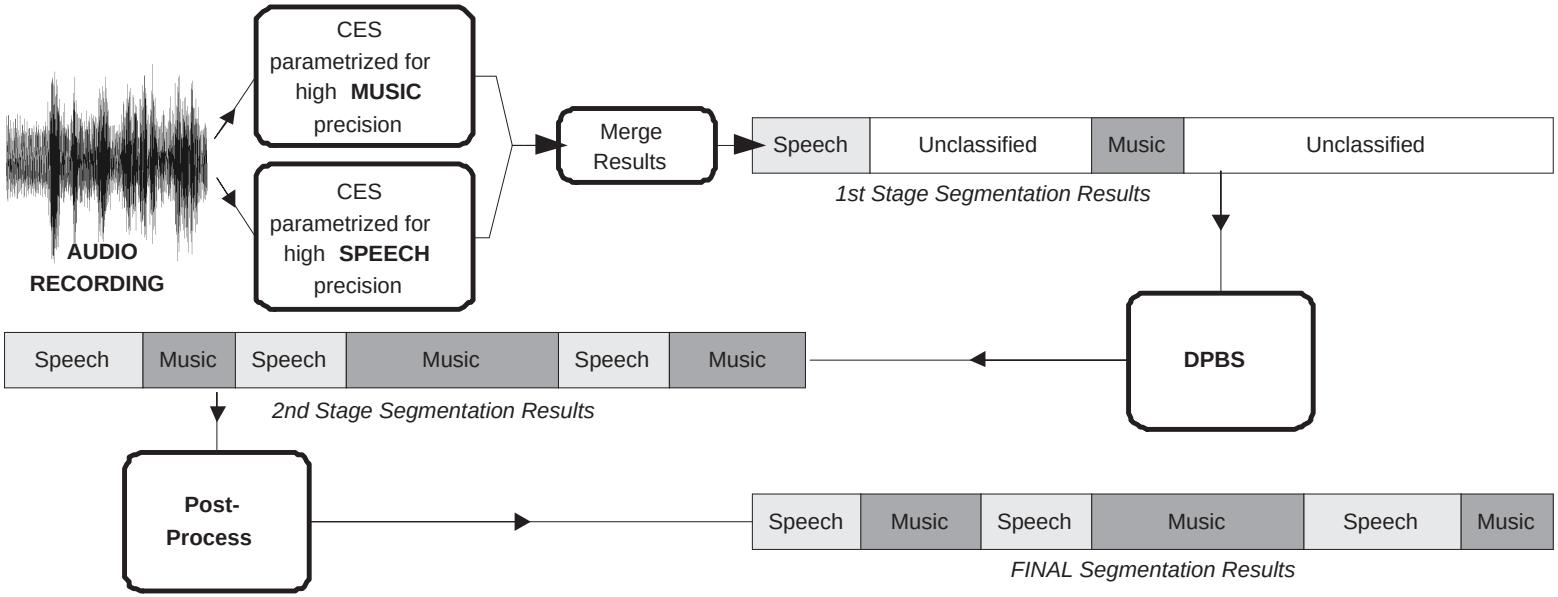
Σειριακές μέθοδοι:

- 1ο Βήμα: Εφαρμογή μεθόδων **κατάτμησης** ⇒ εντοπισμός ορίων ομοιογενών ηχητικών τμημάτων (με βάση το περιεχόμενο).
- 2ο Βήμα: Εφαρμογή μεθόδων **ταξινόμησης** (music vs speech) στα ομοιογενή ηχητικά τμήματα.

Προτεινόμενες μέθοδοι:

- Βασική μέθοδος: αναγωγή σε πρόβλημα μεγιστοποίησης πιθανοτήτων.
- Δευτερεύουσα μέθοδος: βασισμένη σε τεχνικές region growing .

Προτεινόμενη μέθοδος: Γενικά



Σχήμα: Γενική περιγραφή: μία μέθοδος κατάτμησης - ταξινόμησης χαμηλής υπολογιστικής πολ/τας (CES) εντοπίζει τμήματα μουσικής και ομιλίας με υψηλά ποσοστά ακρίβειας, αφήνοντας μη-ταξινομημένα ορισμένα τμήματα, τα οποία δίνονται σαν είσοδο στην βασική μέθοδο (DPBS), βασισμένη σε δίκτυα Bayes και δυναμικό προγραμματισμό.

CES

Γενικά:

- CES: Chromatic Entropy Segmenter
- Παρόμοιες τεχνικές σε επεξεργασία εικόνας (regions grow).

Βήματα:

- Υπολογισμός χρωματικής εντροπίας C (short-term)
- Αρχικοποίηση "seeds" ανά $T_{seed} secs$.
- Όσο $std(C) \leq T_h$: εξάπλωση seeds
- Διαγραφή μικρών ($\leq T_{min}$) segments
- Όσα segments επιβιώνουν: μουσική.

Μεγιστοποίηση Precision(για μουσική και ομιλία ξεχωριστά).

	T_h	T_{min}	T_{seed}	Precision	Recall
Music	0.3	9.0	2.0	99.5%	45.1%
Speech	0.6	4.0	2.0	98.5%	75.5%

CES - Ρύθμιση παραμέτρων

Μεγιστοποίηση Precision(για μουσική και ομιλία ξεχωριστά).

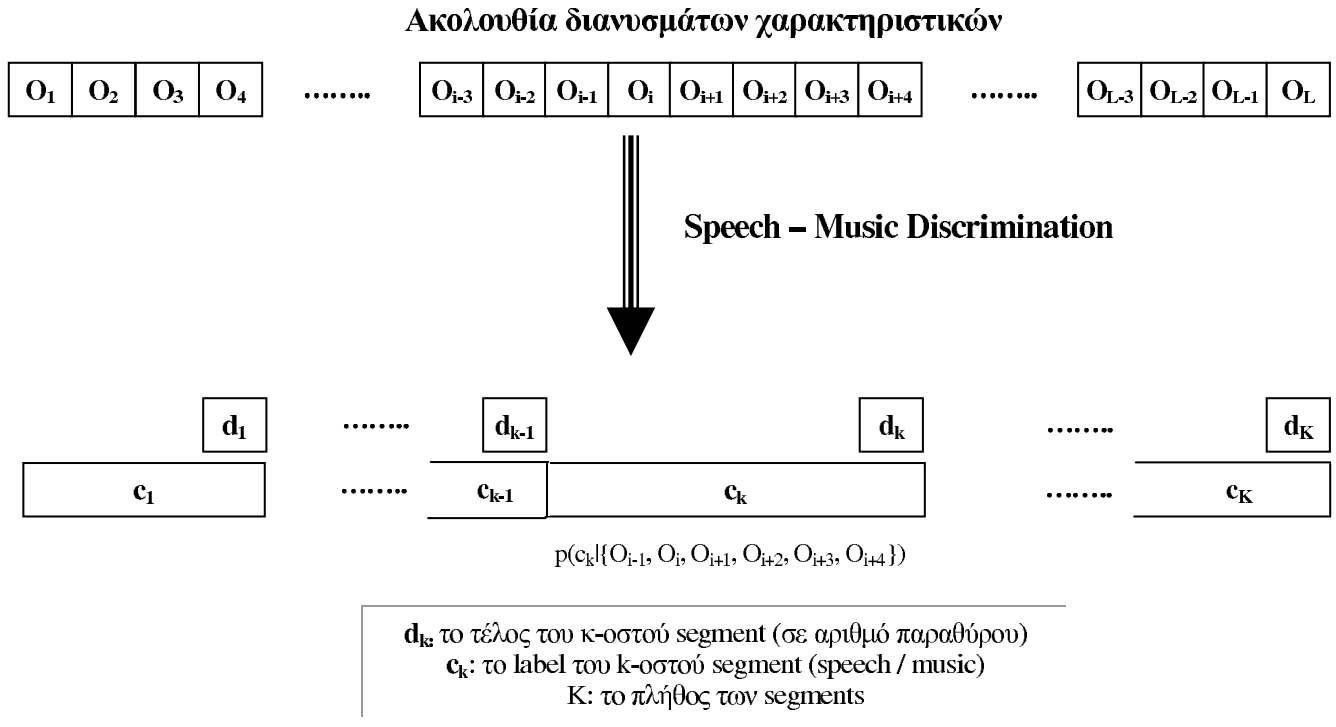
	T_h	T_{min}	T_{seed}	Precision	Recall
Music	0.3	9.0	2.0	99.5%	45.1%
Speech	0.6	4.0	2.0	98.5%	75.5%

Μεγιστοποίηση Overall Accuracy .

T_h	T_{min}	T_{seed}
0.50	3.0 sec	2.0 sec

Dynamic Programming Based Segmenter (1)

Εξαγωγή χαρακτηριστικών: Ενέργεια, 2 βασισμένα στο cv, τα πρώτα 2 MFCCs.



$p(c_k \mid \{O_{d_{k-1}+1}, \ldots, O_{d_k}\})$: posterior πιθανότητα του c_k δεδομένων παρατηρήσεων (χαρακτηριστικών).

Dynamic Programming Based Segmenter (2)

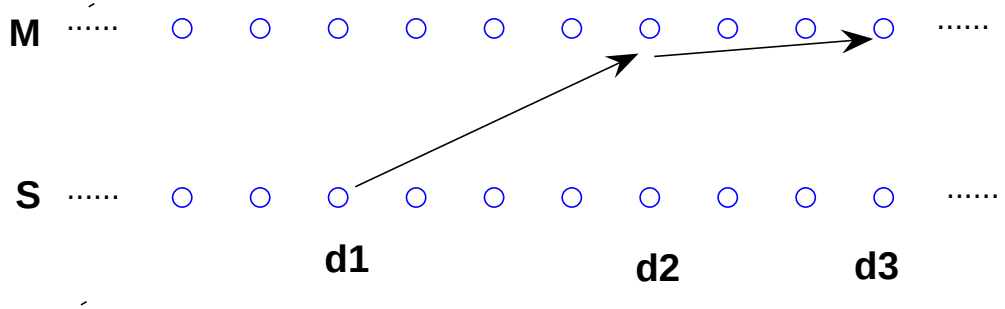
- Για κάθε ακολουθία από segments ορίζεται:

$$J(\{d_1, \ldots, d_K\}, \{c_1, \ldots, c_K\}, K) \equiv p(c_1 \mid \{O_1, \ldots, O_{d_1}\}) \cdot \prod_{k=2}^{K} p(c_k \mid \{O_{d_{k-1}+1}, \ldots, O_{d_k}\})$$

- Μεγιστοποίηση του J για όλες τις δυνατές τιμές των $\{d_1, d_2, \ldots, d_{K-1}, d_K\}$, $\{c_1, c_2, \ldots, c_{K-1}, c_K\}$ και K

Dynamic Programming Based Segmenter (3)

- Χρήση grid :



- Κόμβος (O_{d_k}, S): τμήμα ομιλίας που τελειώνει στο d_k.
- Μονοπάτι: $\{(O_{d_1}, c_1), (O_{d_2}, c_2), \ldots, (O_{d_K}, c_K)\}$
- Μετάβαση: $(O_{d_{k-1}}, c_{k-1}) \rightarrow (O_{d_k}, c_k)$ με κόστος: $T(\cdot) = p(c_k \mid \{O_{d_{k-1}+1}, \ldots, O_{d_k}\})$

Dynamic Programming Based Segmenter (4)

- Η συνάρτηση γίνεται:

$$p(c_1 \mid \{O_1, \ldots, O_{d_1}\}) \cdot \prod_{k=2}^{K} T((O_{d_{k-1}}, c_{k-1}) \rightarrow (O_{d_k}, c_k))$$

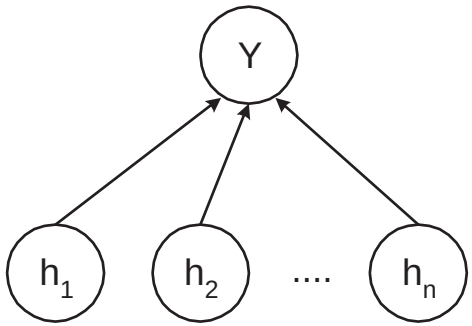
- Δυναμικός προγραμματισμός
- Για κάθε παράθυρο:
  - Επιλογή του βέλτιστου προγόνου με βάση την αντίστοιχη πιθανότητα
  - Περιορισμός: $T_{dmax} \geq d_k - d_{k-1} \geq T_{dmin}$
- Backtracking .
- Εκτίμηση του: $p(c_k \mid \{O_{d_{k-1}+1}, \ldots, O_{d_k}\})$: χρήση δικτύων Bayes .

BN probability estimator (1)

- Δίκτυα Bayes για την εκτίμηση του: $p(c_k \mid \{O_{d_{k-1}+1}, \ldots, O_{d_k}\})$.
- Το BN αποφασίζει για κάθε υποψήφιο **segment** .
- Δύο βήματα: (1) Individual Classifiers (2) BN combiner

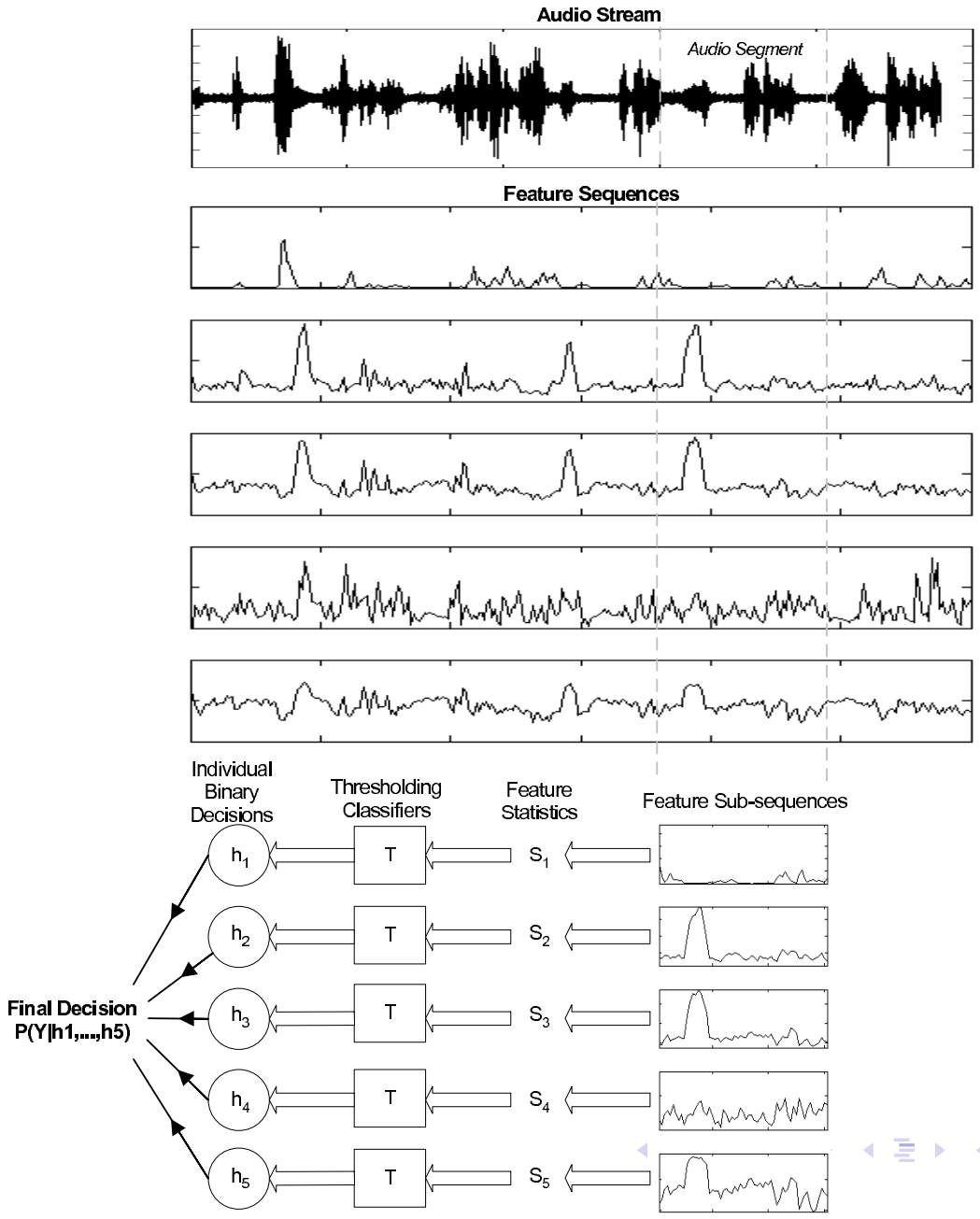
BN probability estimator (2)

- Για κάθε υποψήφιο segment : 5 ακολουθίες χαρακτηριστικών $\Rightarrow$ 5 στατιστικά
- Κάθε στατιστικό ταξινομείται (μουσική - ομιλία) από έναν απλό ταξινομητή κατωφλίου $\Longrightarrow$ 5 δυαδικές αποφάσεις $(h_1, ..., h_5)$.
- Οι 5 αποφάσεις συνδυάζονται από ένα BN : εκτίμηση του $P_{dec} = P(Y|h_1, ..., h_5)$ (Y: ετικέτα της πραγματικής κλάσης).



- Συμπερασμός: $P_{dec} = P(Y|h_1, ..., h_5)$

BN probability estimator (3)

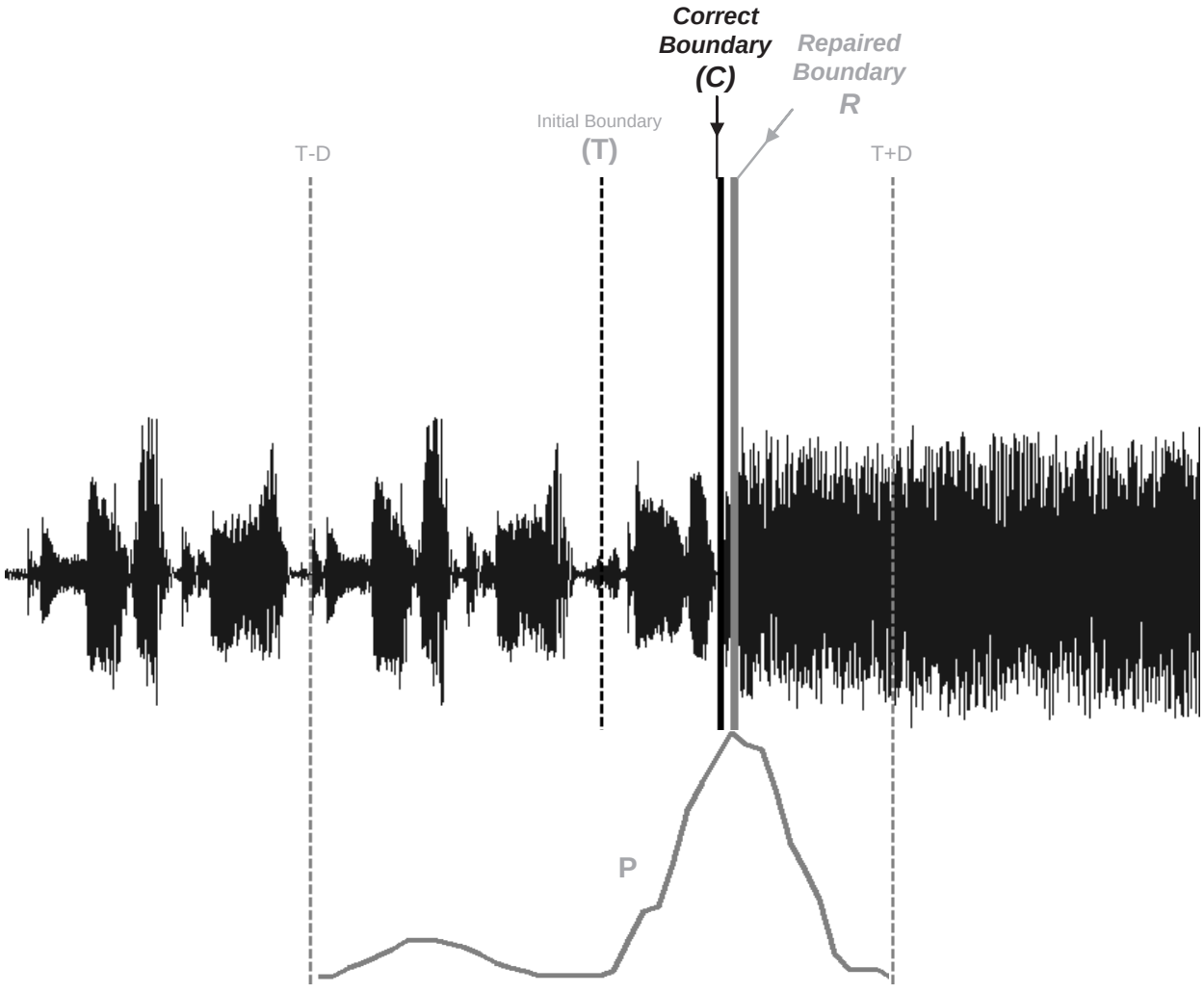


Post-processing (1)

Διόρθωση ορίων. Για κάθε όριο:

- Έστω T η θέση του ορίου, c_{left} και c_{right} οι ετικέτες των τμημάτων αριστερά και δεξιά του T.
- $t = T - D$, όπου D παράμετρος και $i = 0$.
- Όσο $t \leq T + D$:
  - x_{left}: τα ηχητικά δείγματα στο $[t - D, t]$.
  - x_{right}: τα ηχητικά δείγματα στο $[t, t + D]$.
  - Υπολογισμός των: $P_{left} = P(Y = c_{left}|x_{left})$ και $P_{right} = P(Y = c_{right}|x_{right})$
  - $P_i = P_{left} \cdot P_{right}$.
  - $i = i + 1$, $t = t + 0.050$.
- $maxPos = \arg\max(P)$.
- Το νέο όριο: $R = T + (maxPos \cdot 0.050 - D)$

Post-processing (2)



Αποτελέσματα (1)

Περιγραφή ηχητικών δεδομένων που χρησιμοποιήθηκαν για δοκιμή:

Genre	Διάρκεια (λ)	Μουσική	Ομιλία
POP - ROCK	125	83.02%	16.98%
JAZZ-BLUES	90	67.19%	32.81%
DANCE	85	76.81%	23.19%
NEWS	80	16.17%	83.83%
H. METAL - H. ROCK	80	94.11%	5.89%
CLASSICAL	75	78.64%	21.36%

Αποτελέσματα (2)

	CES		DPBS	
	M	**S**	**M**	**S**
M	69.09	1.59	69.24	1.44
S	6.74	22.58	4.18	25.14
	A: 91.67		**A: 94.38**	

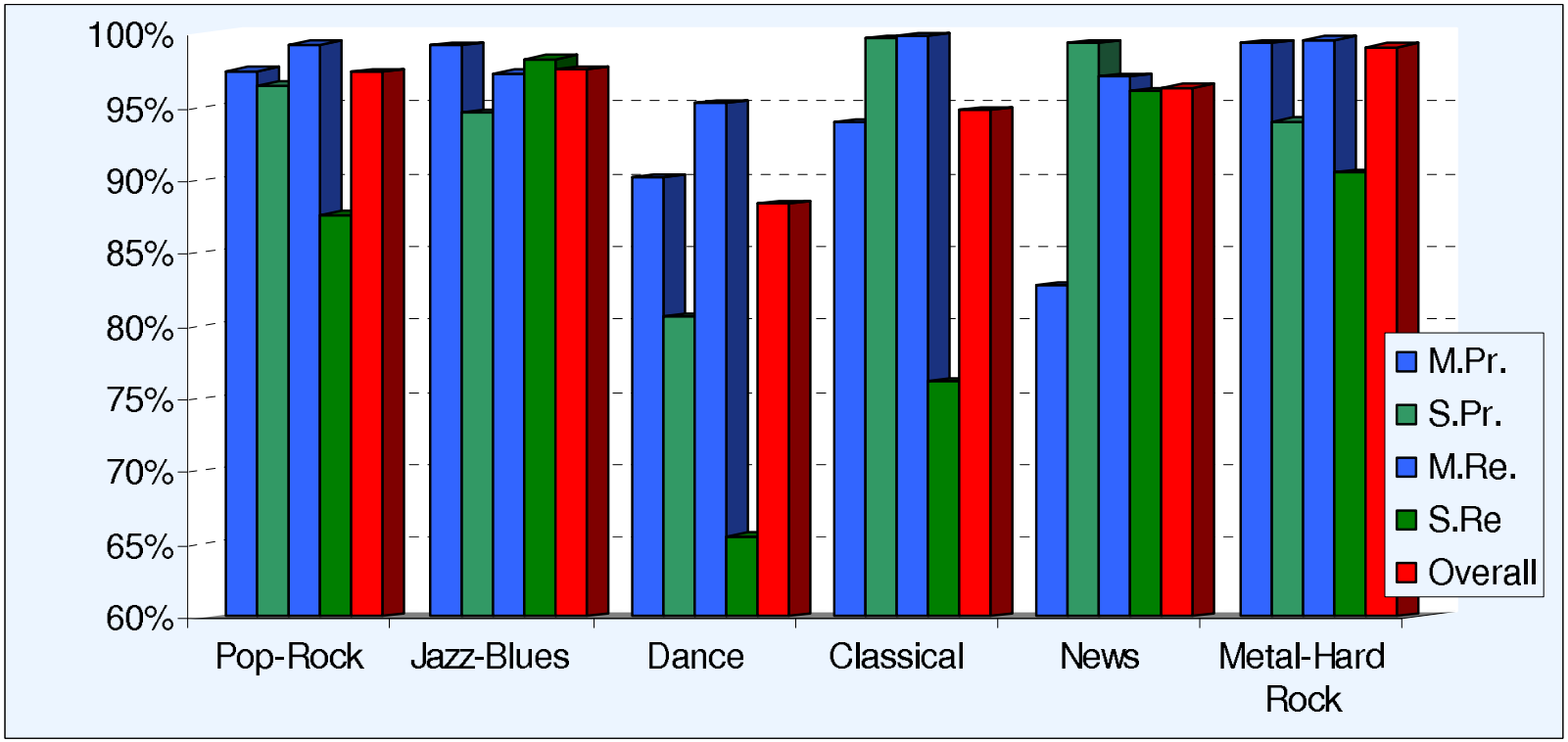
Αποτελέσματα (2)

	CES		DPBS	
	M	**S**	**M**	**S**
M	69.09	1.59	69.24	1.44
S	6.74	22.58	4.18	25.14
	A: 91.67		**A: 94.38**	

	Overall		Overall2	
	M	**S**	**M**	**S**
M	69.34	1.34	69.53	1.15
S	3.51	25.80	3.17	26.15
	A: 95.15		**A: 95.68**	

Αποτελέσματα (3)

Αποτελέσματα ανά ραδιοφωνικό είδος:



Εισαγωγή

Στόχοι:

- Αναγνώριση περιεχομένου ταινιών με βάση το ηχητικό μέσο.
- Έμφαση σε εντοπισμό περιεχομένου βίας.

Εισαγωγή

Στόχοι:

- Αναγνώριση περιεχομένου ταινιών με βάση το ηχητικό μέσο.
- Έμφαση σε εντοπισμό περιεχομένου βίας.

Βία:

- Συμπεριφορά ατόμων απέναντι σε άτομα με σκοπό την απειλή ή την φυσική (ή και λεκτική) βλάβη.
- Βίαιες σκηνές ταινιών: συνήθως δύσκολο να εντοπιστούν μέσω της οπτικής πληροφορίας.
- Προηγούμενες εργασίες: έμφαση σε οπτική πληροφορία, κυρίως βασισμένη σε μοντελοποίηση ανθρώπινης κίνησης.

Ορισμός ηχητικών κλάσεων (1)

Πολλαπλές ηχητικές κλάσεις:

- Λεπτομερής περιγραφή του περιεχομένου
- Βελτίωση σε σχέση με δυαδική προσέγγιση (βία-όχι βία)

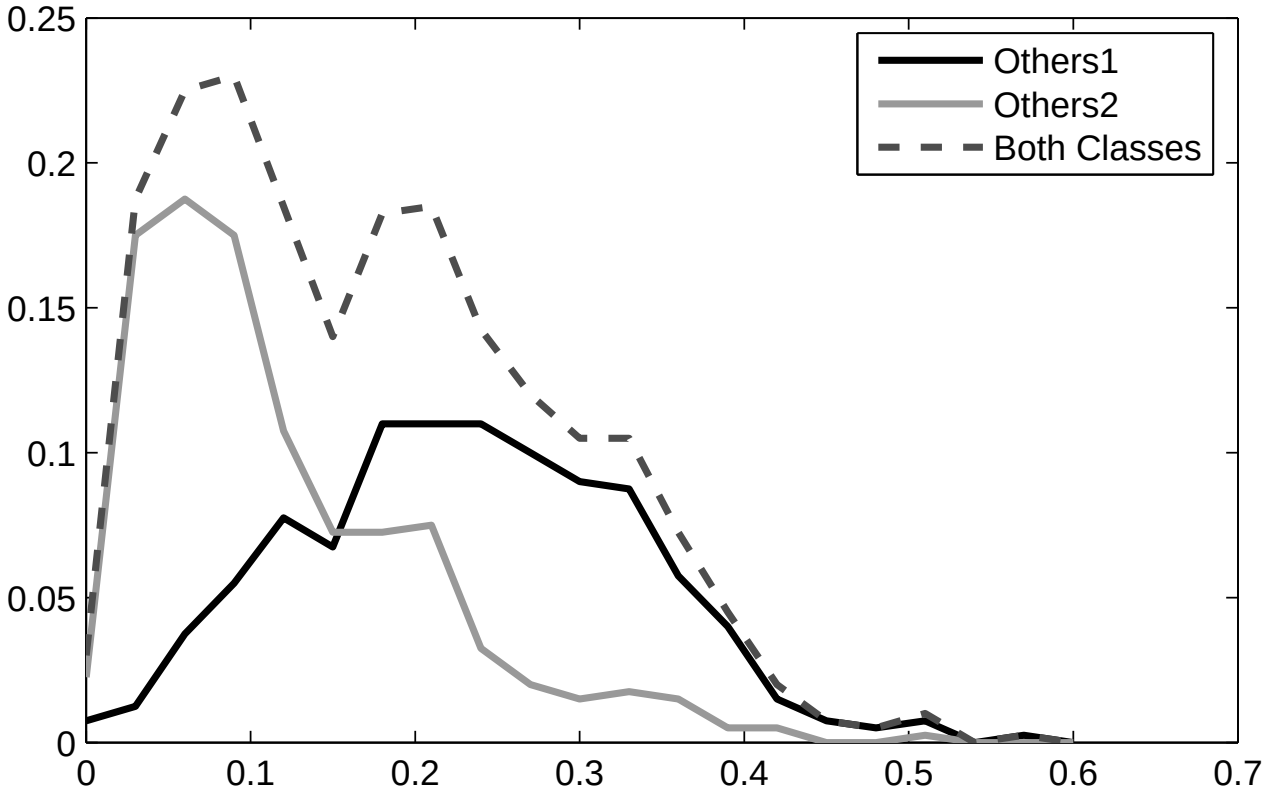
	Class Name	Class Description
1	Music	Μουσική και μουσικά εφέ
2	Speech	Ομιλία από διαφορετικούς ομιλητές, συναισθηματικές καταστάσεις και επίπεδα θορύβου.
3	Others 1	Περιβαλλοντικοί ήχοι χαμηλών διακυμάνσεων
4	Others 2	Περιβαλλοντικοί ήχοι με απότομες αλλαγές στο σήμα
5	Gunshots	Συνεχόμενοι ή μεμονωμένοι πυροβολισμοί
6	Fights	Ήχοι ξυλοδαρμών, συμπλοκών
7	Screams	Ανθρώπινες κραυγές

12-διάστατο διάνυσμα χαρακτηριστικών για κάθε ηχητικό τμήμα:

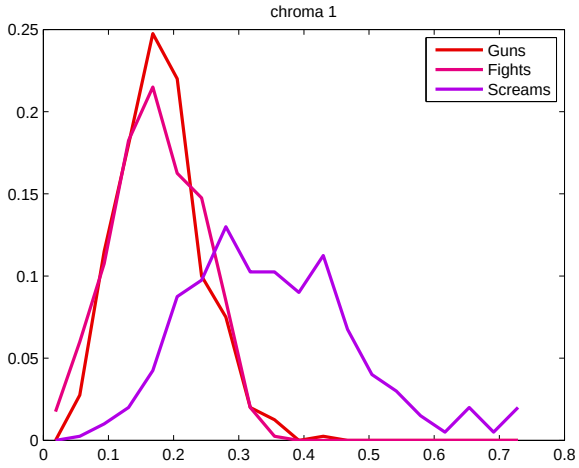
	Feature	Statistic	Window (msecs)
1	Spectrogram	σ^2	20
2	Chroma 1	μ	100
3	Chroma 2	*median*	20 (mid term:200)
4	Energy Entropy	*min*	20
5	MFCC 2	σ^2	20
6	MFCC 1	*max*	20
7	ZCR	μ	20
8	Sp. RollOff	*median*	20
9	Non Zero Pitch Ratio	–	20
10	MFCC 1	max/μ	20
11	Spectrogram	*max*	20
12	MFCC 3	*median*	20

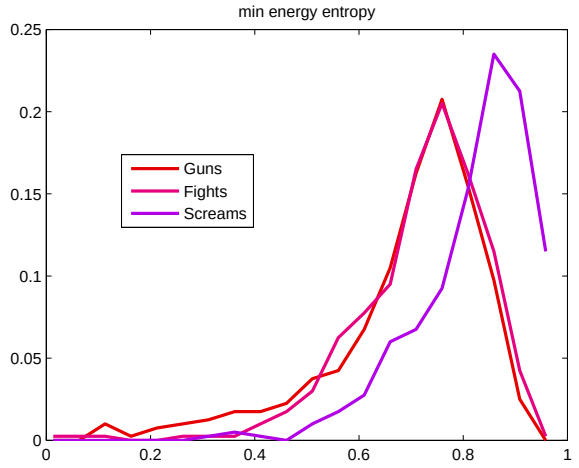
Ορισμός ηχητικών κλάσεων (2)

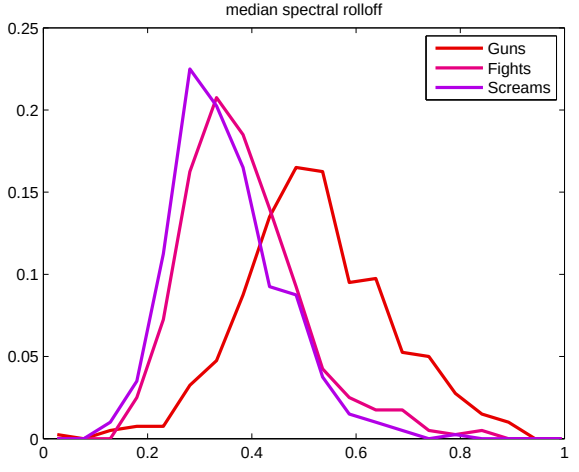
Παράδειγμα τιμών του 2ου χαρακτηριστικού για τις 2 περιβαλλοντικές κλάσεις:

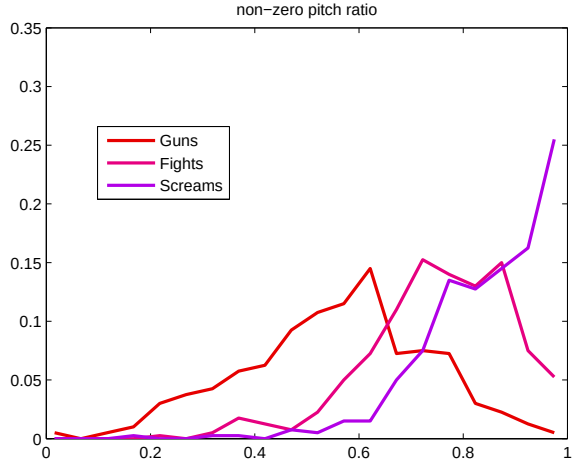


Μερικά ιστογράμματα για τις βίαιες κλάσεις









Ταξινόμηση (1)

- "One-vs-All" (OVA) .
- Ανάλυση σε 7 δυαδικά υπο-προβλήματα.
- Κάθε δυαδικό πρόβλημα: BN .

Ταξινόμηση (1)

- "One-vs-All" (OVA) .
- Ανάλυση σε 7 δυαδικά υπο-προβλήματα.
- Κάθε δυαδικό πρόβλημα: BN .
- Τα χαρακτηριστικά: $v_i, i = 1 \ldots 12$ ομαδοποιούνται σε 3 $4D$ διανύσματα:

$$V^{(1)} = [v_1, v_4, v_7, v_{10}]$$
$$V^{(2)} = [v_2, v_5, v_8, v_{11}]$$
$$V^{(3)} = [v_3, v_6, v_9, v_{12}]$$

- Για κάθε δυαδικό υπο-πρόβλημα: 3 kNN Classifiers (ένα για κάθε υπο-διάνυσμα)

Ταξινόμηση (1)

- "One-vs-All" (OVA) .
- Ανάλυση σε 7 δυαδικά υπο-προβλήματα.
- Κάθε δυαδικό πρόβλημα: BN .
- Τα χαρακτηριστικά: $v_i, i = 1 \ldots 12$ ομαδοποιούνται σε 3 $4D$ διανύσματα:

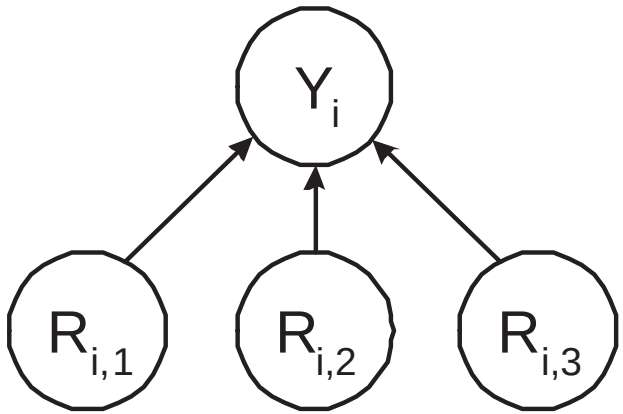
$$V^{(1)} = [v_1, v_4, v_7, v_{10}]$$
$$V^{(2)} = [v_2, v_5, v_8, v_{11}]$$
$$V^{(3)} = [v_3, v_6, v_9, v_{12}]$$

- Για κάθε δυαδικό υπο-πρόβλημα: 3 kNN Classifiers (ένα για κάθε υπο-διάνυσμα)
- 7×3 πίνακας δυαδικών αποφάσεων:

$$R_{i,j} = \begin{cases} 1, & \text{ταξινόμηση στην κλάση } i, \text{ δεδομένου του } V^{(j)} \\ 0, & \text{ταξινόμηση στην κλάση } i', \text{ δεδομένου του } V^{(j)} \end{cases}$$

Ταξινόμηση (2)

- Κάθε σειρά ($\mapsto$κάθε δυαδικό υποπρόβλημα) του πίνακα R συνδυάζεται από ένα BN .



- Μία πιθανότητα για κάθε δυαδικό υποπρόβλημα:

$$P_i(k) = P(Y_i(k) = 1 | R_{i,1}^{(k)}, R_{i,2}^{(k)}, R_{i,3}^{(k)})$$

- Νικήτρια κλάση:

$$WinnerClass(k) = \arg\max_i P_i(k)$$

Αποτελέσματα (1)

Confusion matrix:

	Classified						
True ↓	Mu	Sp	Ot1	Ot2	Sh	Fi	Sc
Mu	68.22	2.36	13.60	1.76	3.27	3.83	6.95
Sp	1.66	81.96	6.38	4.75	0.23	2.08	2.95
Ot1	4.59	1.90	70.24	11.20	5.44	2.52	4.11
Ot2	2.00	3.15	15.21	59.83	10.30	8.57	0.94
Sh	1.26	0.19	3.00	6.66	79.10	9.68	0.11
Fi	1.70	2.23	0.89	11.81	26.38	52.29	4.71
Sc	9.18	3.44	4.00	1.29	2.20	7.86	72.04

Recall και Precision ανά κλάση :

	Mu	Sp	Ot1	Ot2	Sh	Fi	Sc
Recall:	68.2	82.0	70.2	59.8	79.1	52.3	72.0
Precision:	77.0	86.1	62.0	61.5	62.3	60.2	78.5

Αποτελέσματα (2)

	Ov. Accuracy	V. Recall*	V. Precision**
BN	69.1%	83.2%	84.8%

- * $Re_{violence} = \frac{\sum_{i=5}^{7}\sum_{j=5}^{7} C_{ij}}{\sum_{i=5}^{7}\sum_{j=1}^{7} C_{ij}}$
- ** $Pr_{violence} = \frac{\sum_{i=5}^{7}\sum_{j=5}^{7} C_{ij}}{\sum_{i=1}^{7}\sum_{j=5}^{7} C_{ij}}$

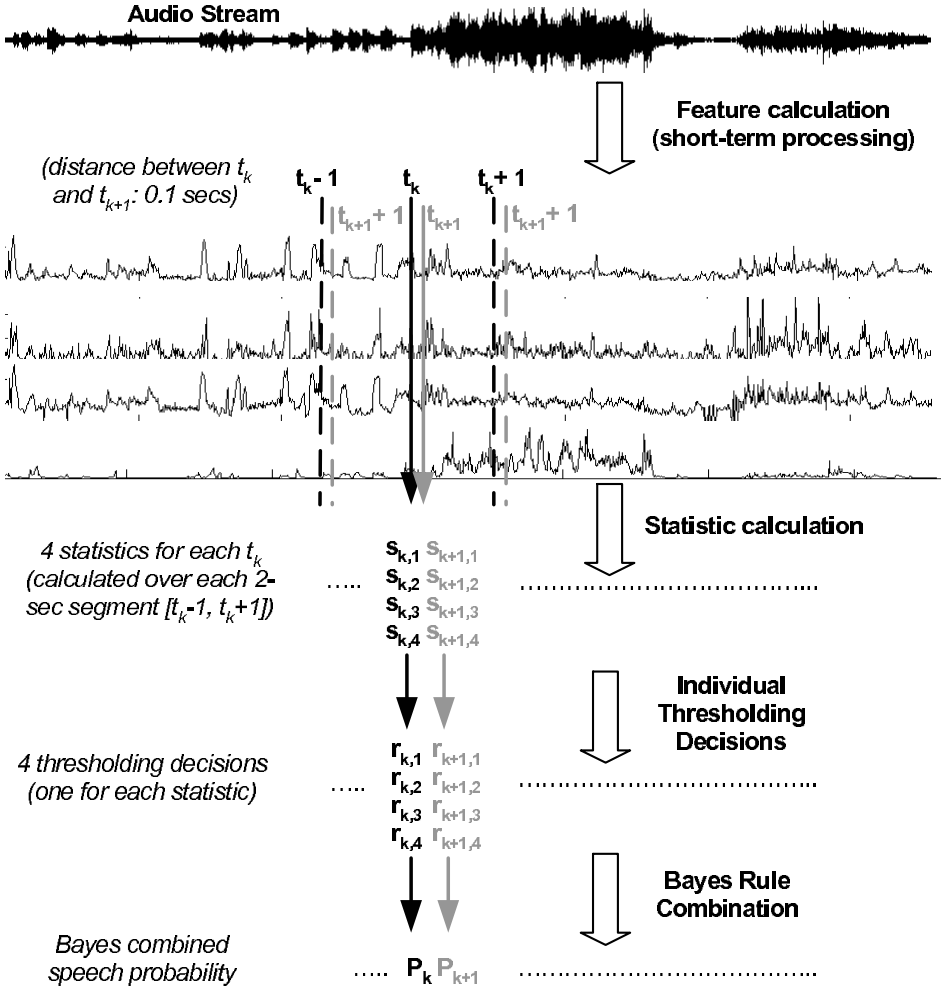
Εισαγωγή

1. Αναγνώριση πολυμεσικού περιεχομένου: γεγονότα, δομές, είδη.
2. Επίσης: αναγνώριση συναισθήματος
3. Εφαρμογές: μουσική, διαδραστικά περιβάλλοντα, ταινίες, σπορ.

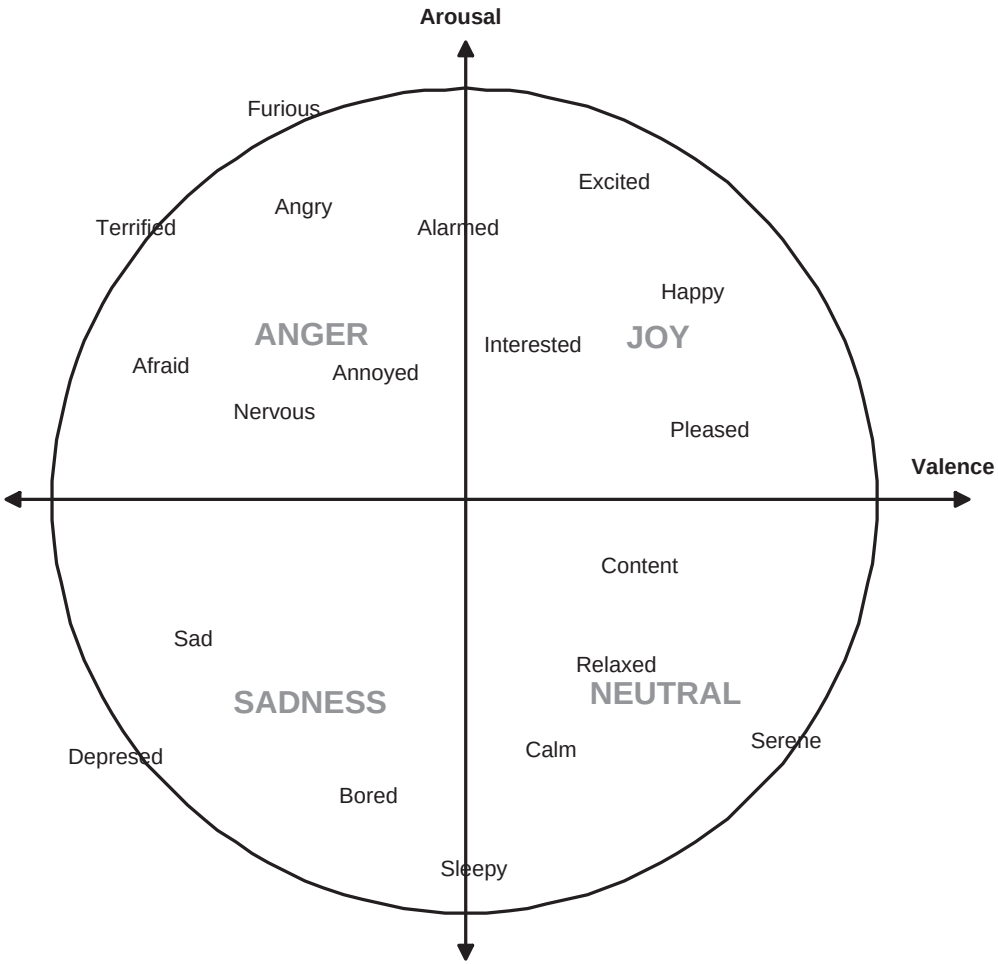
Σκοπός:

1. Εντοπισμός ομιλίας σε ταινίες (ακρίβεια 95%)
2. Έλεγχος 2-διάστατης αναπαράστασης (Arousal-Valence)
3. Μελέτη ηχητικών χαρακτηριστικών
4. Δοκιμή 3 μεθόδων regression
5. Ταξινόμηση / ανάκτηση ταινιών με βάση το συναισθηματικό περιεχόμενο.

Εντοπισμός ομιλίας

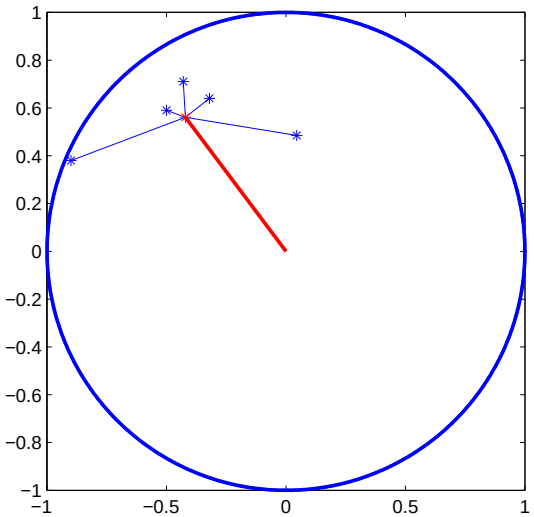


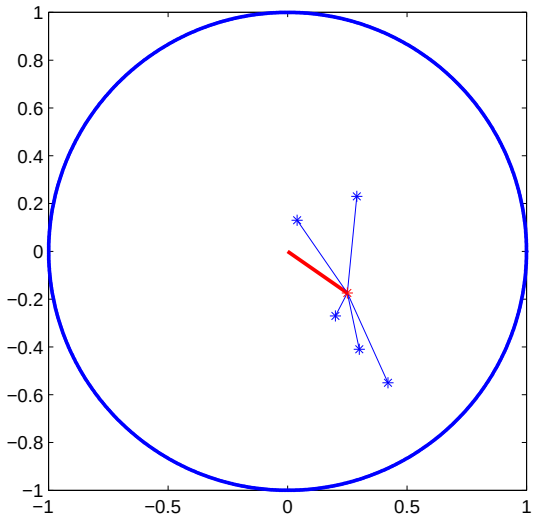
Αναπαράσταση συναισθήματος



Συλλογή δεδομένων

- 2000 δείγματα ομιλίας
- Χειροκίνητος χαρακτηρισμός από 50 ανθρώπους
- Σκοπός:
  1. Μέσες τιμές για εκπαίδευση - δοκιμή
  2. Εκτίμηση βαθμού διαφωνίας ανάμεσα στους χρήστες (D)
  3. Εκτίμηση βαθμού διαφωνίας ανάμεσα σε πολλαπλές αποφάσεις του ίδιου χρήστη (DS)



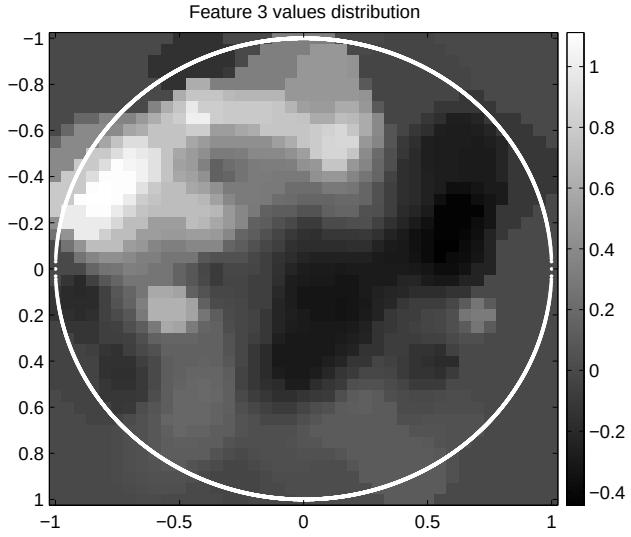


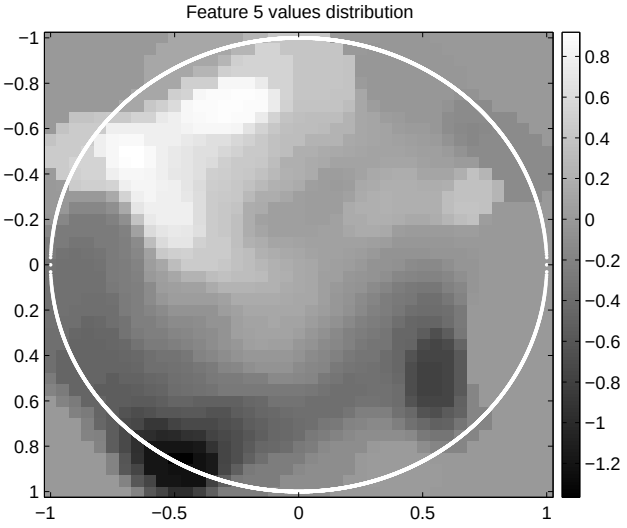
Ηχητικά χαρακτηριστικά (1)

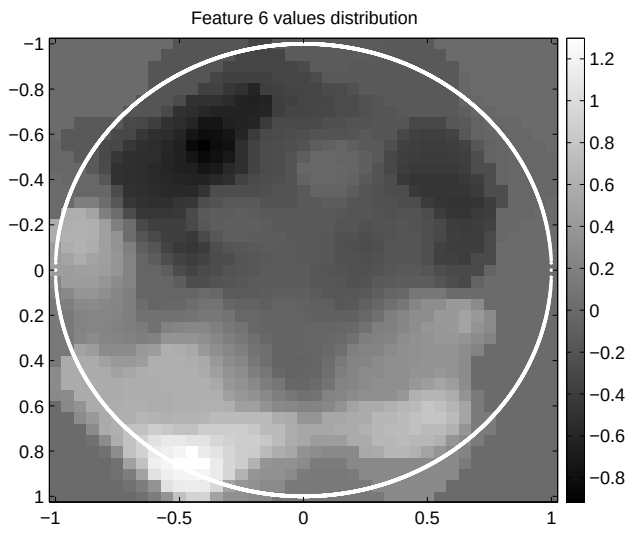
10-διάστατο διάνυσμα χαρακτηριστικών για κάθε τμήμα ομιλίας:

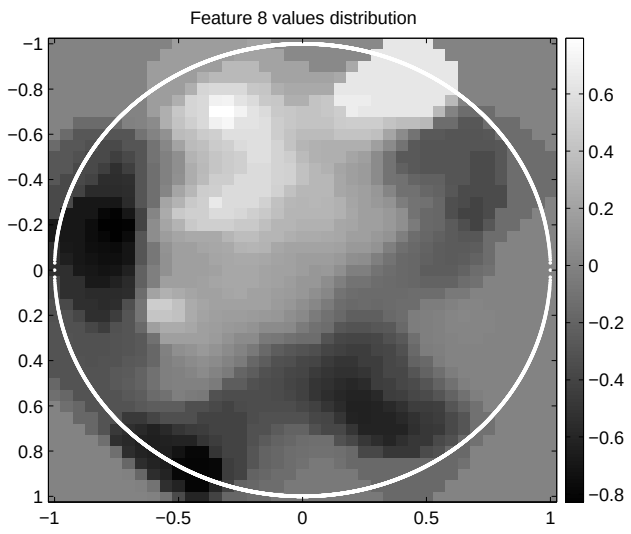
	Feature	Statistic
1	3rd MFCC	μ
2	2nd MFCC	*max*
3	maxFFTPos	*max*
4	maxFFTPos	*std*
5	ZCR	$\frac{\sigma^2}{\mu}$
6	ZCR	*median*
7	Sp. Centroid	$\frac{\sigma^2}{\mu}$
8	Pitch	$\frac{max}{\mu}$
9	Pitch	$\frac{\sigma^2}{\mu}$
10	2nd Chroma-based	—

Ηχητικά χαρακτηριστικά (2)

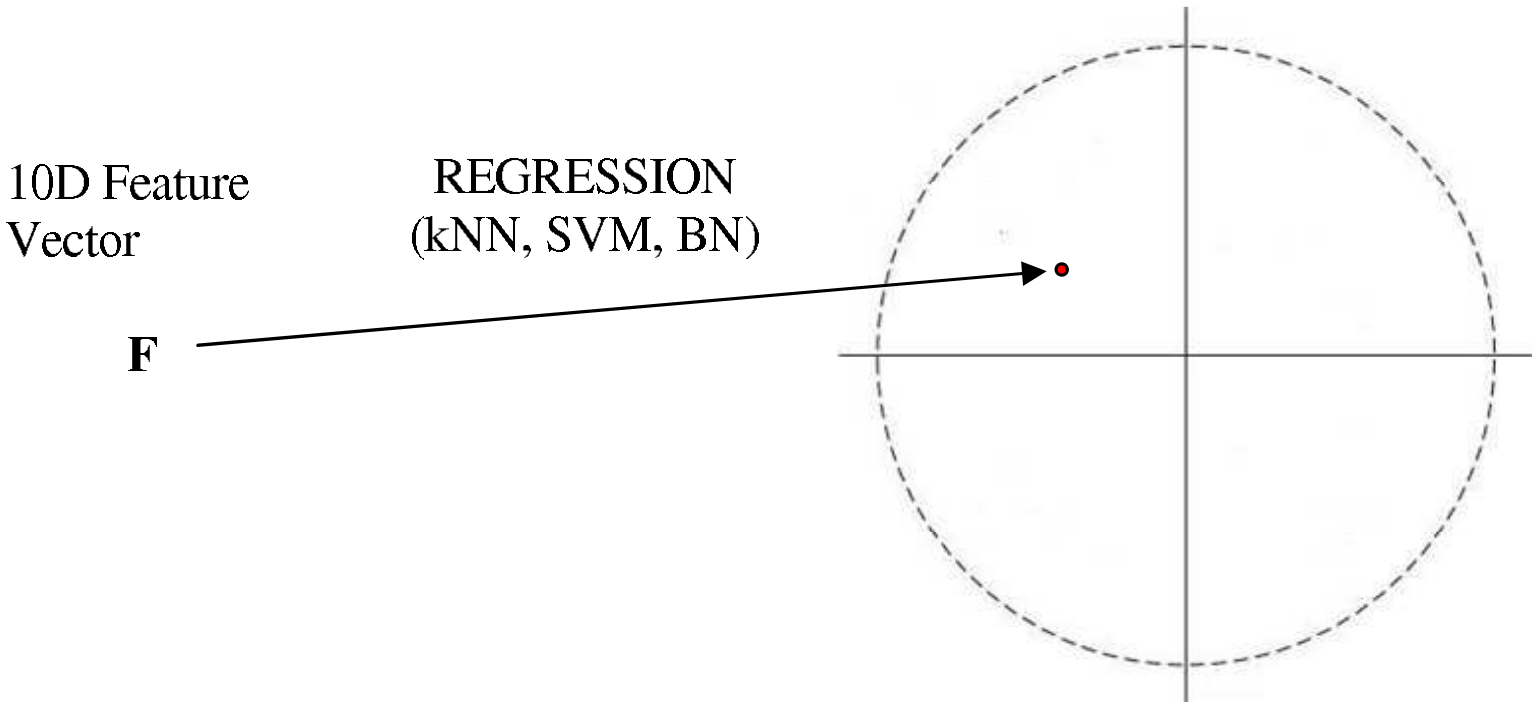




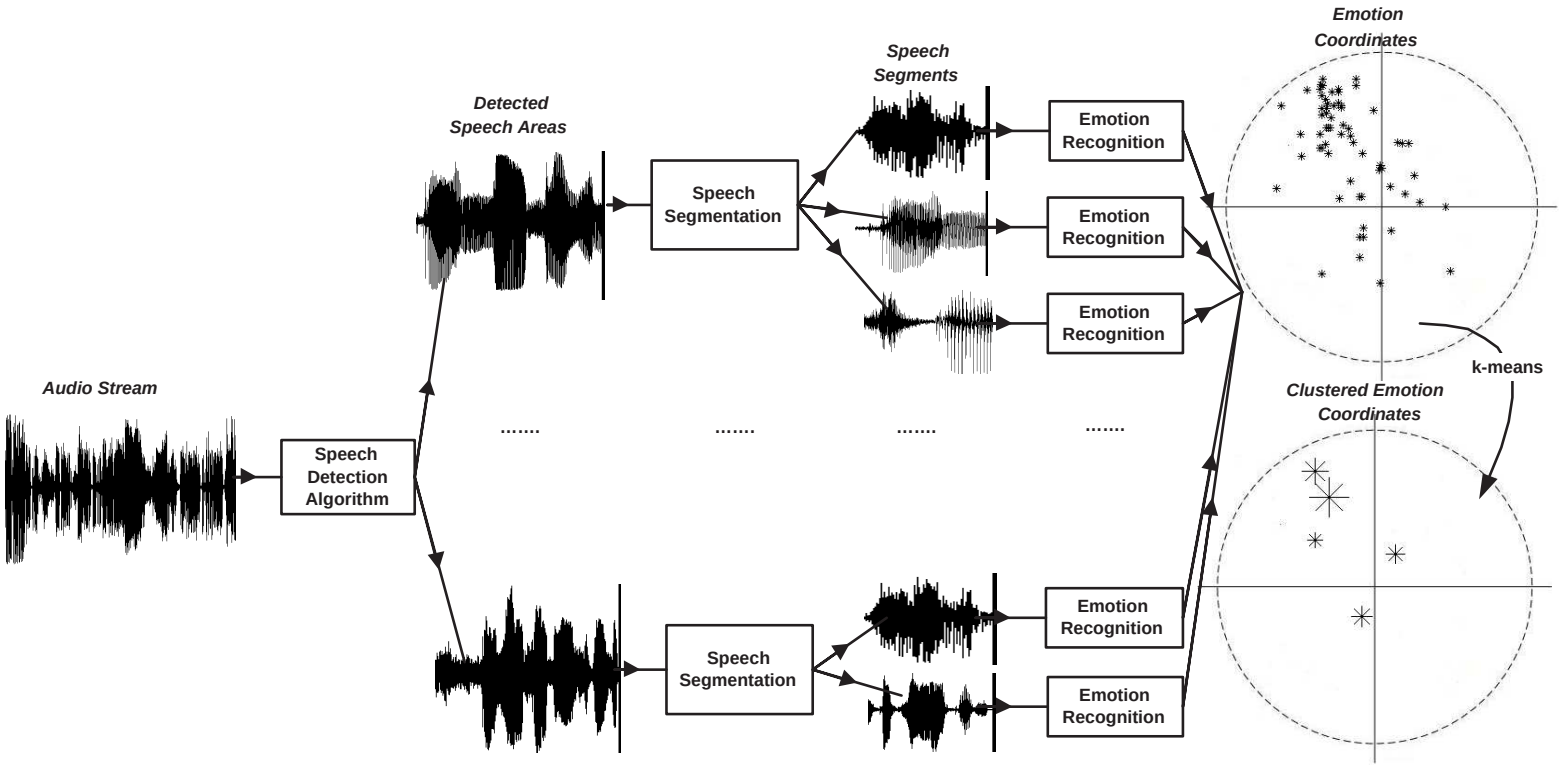




Regression

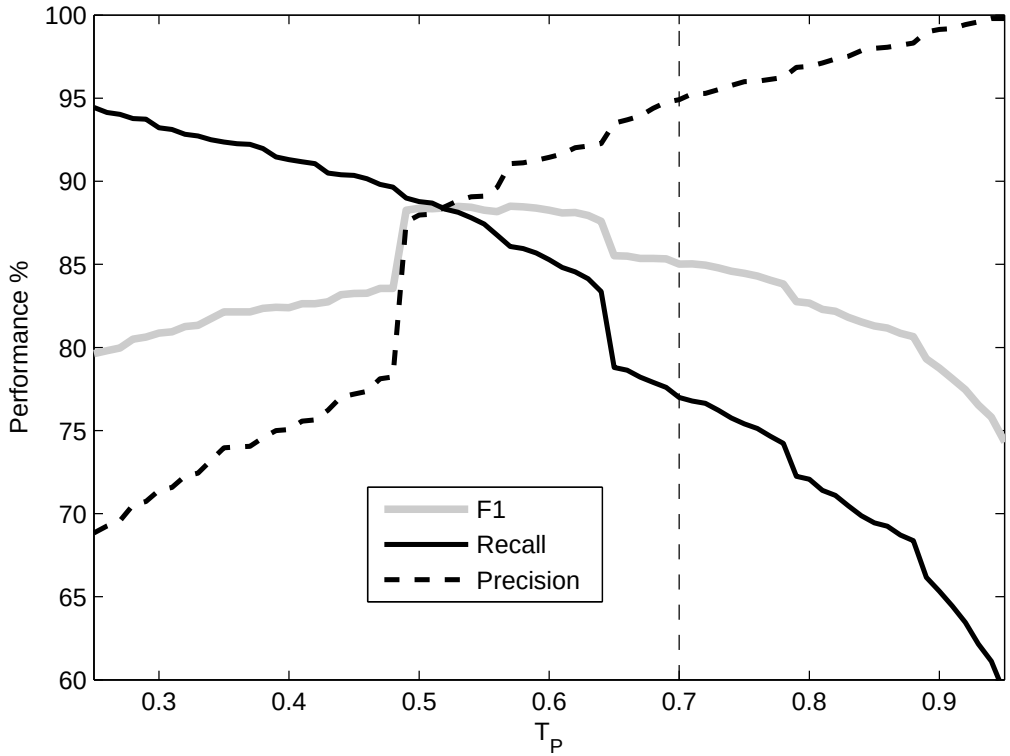


Συνολικό σύστημα αναγνώρισης συναισθήματος



Αποτελέσματα (1)

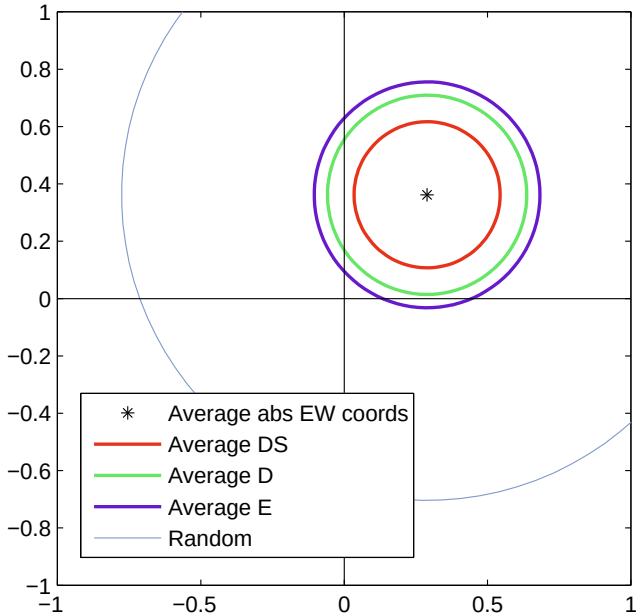
- Εντοπισμός ομιλίας (βέλτιστο κατώφλι: 0.55, επιλεγμένο (μεγαλύτερη ακρίβεια: 0.55):



Αποτελέσματα (2)

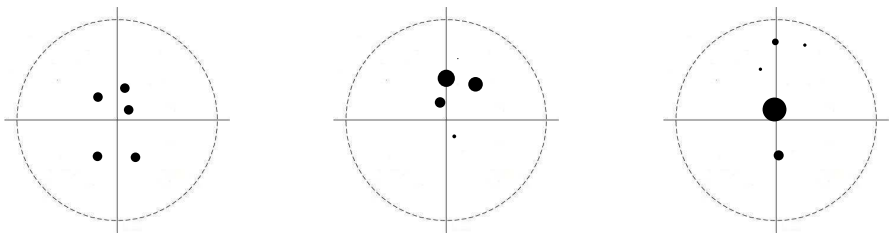
- Αναγνώριση συναισθήματος:

	E	R_X^2	R_Y^2
User	0.75	-	-
kNN	0.92	0.21	0.34
SVM	0.87	0.23	0.36
BN	0.88	0.23	0.35
Random	2.3	-3	-2.2

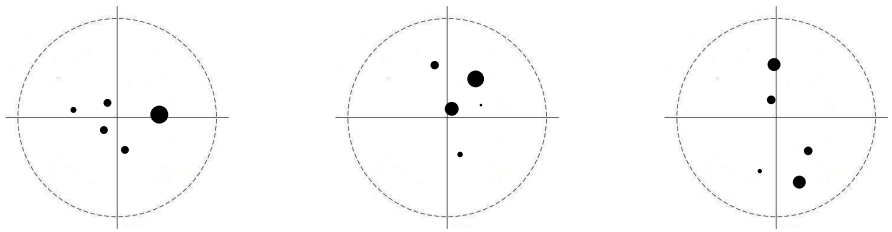


Αποτελέσματα (3)

- Παραδείγματα σε εφαρμογή του συνολικού συστήματος σε ηχητικές ακολουθίες:

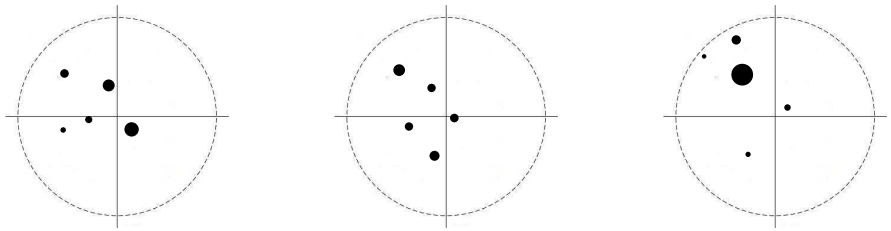


Ειδήσεις

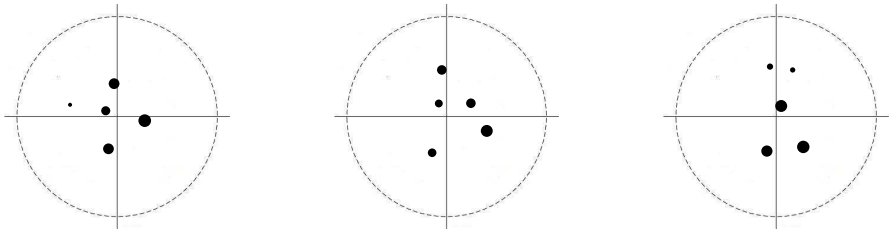


Διαφημίσεις

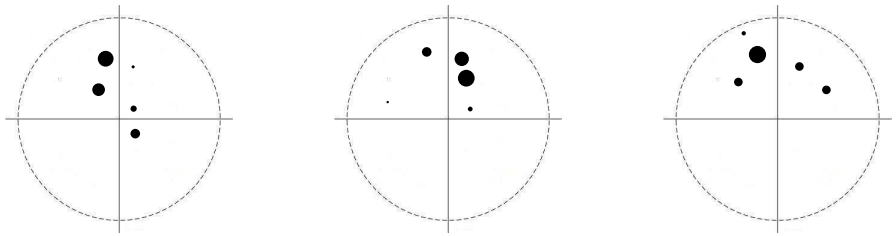
Αποτελέσματα (4)



Ταινίες με λεκτική βία



Ντοκιμαντέρ



Αθλητικές μεταδόσεις

Δημοσιεύσεις (1)

- A. Pikrakis, T. Giannakopoulos, S. Theodoridis, "A Speech/Music Discriminator of Radio Recordings based on Dynamic Programming and Bayesian Networks", IEEE Transactions on Multimedia, Volume: 10 Issue: 5, Aug. 2008, Page(s): 846-857.
- T. Giannakopoulos , A. Pikrakis, S. Theodoridis, "Speech emotion recognition in audio streams from movies", IEEE trans on Audio, Speech and Language Processing, (under review, IEEE Transactions on Speech, Audio and language Processing)
- A. Pikrakis, T. Giannakopoulos, S. Theodoridis, "An Overview of Speech - Music Discrimination Techniques in the Context of Audio Recordings" in Multimedia Services in Intelligent Environments (Advanced Tools and Methodologies, Studies in Computational Intelligence), Publisher: Springer Berlin / Heidelberg, Volume 120 / 2008, ISBN: 978-3-540-78491-3
- T. Giannakopoulos, A. Pikrakis and S. Theodoridis "A Novel Efficient Approach for Audio Segmentation", 19th International Conference on Pattern Recognition, 2008 (ICPR2008).

Δημοσιεύσεις (2)

- T. Giannakopoulos, A. Pikrakis and S. Theodoridis "A dimensional approach to emotion recognition of speech from movies" , 34th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2009).
- T. Giannakopoulos, A. Pikrakis and S. Theodoridis "Music Tracking in Audio Streams from Movies" , 2008 International Workshop on Multimedia Signal Processing, IEEE Signal Processing Society (MMSP2008).
- A. Pikrakis, T. Giannakopoulos and S. Theodoridis "Gunshot detection in audio streams from movies by means of dynamic programming and Bayesian networks", 33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP2008)

Δημοσιεύσεις (3)

- A. Pikrakis, T. Giannakopoulos and S. Theodoridis, "A computationally efficient speech/music discriminator for radio recordings", International Conference on Music Information Retrieval and Related Activities (ISMIR2006), 8-12 October 2006, Victoria, BC, Canada.
- T. Giannakopoulos; A. Pikrakis, S. Theodoridis, "A multi-class audio classification method with respect to violent content in movies using bayesian networks" 2007 IEEE International Workshop on Multimedia Signal Processing, Chania, Crete, Greece, October 1-3, 2007 (MMSP2007), [1]
- A. Pikrakis, T. Giannakopoulos and S. Theodoridis, "A Dynamic Programming Approach to Speech/Music Discrimination of Radio Recordings, 15th European Signal Processing Conference (EUSIPCO2007), Poznan, Poland from Sept 3 - 7, 2007

[1] Student paper award in IEEE 2007 International Workshop on Multimedia Signal Processing

Δημοσιεύσεις (4)

- A. Pikrakis, T. Giannakopoulos and S. Theodoridis: "Speech/Music Discrimination for radio broadcasts using a hybrid HMM-Bayesian Network architecture", 14th European Signal Processing Conference (EUSIPCO06), September 4-8, 2006, Florence, Italy.
- T. Giannakopoulos, A. Pikrakis, S. Theodoridis: "A Speech/music Discriminator for Radio Recordings Using Bayesian Networks", 2006 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2006), May 14-19, Toulouse, France.
- Giannakopoulos T, Kosmopoulos D., Aristidou A., Theodoridis S.: "Violence Content Classification Using Audio Features", 4th Hellenic Conference on Artificial Intelligence (SETN2006), Heraklion, Crete, Greece, May 18-20, 2006.

Μελλοντικοί Στόχοι

- Εντοπισμός βίας: χρήση άλλων μέσων: κείμενο, εικόνα
- Χρήση μεθόδου δυναμικού προγραμματισμού στο πρόβλημα πολλαπλών κλάσεων
- Αναζήτηση - ταξινόμηση ταινιών με κριτήρια σχετικά με βία + συναισθηματικό περιεχόμενο
- Χρήση μουσικών θεμάτων από ταινίες για συναισθηματική αναγνώριση