Εύρωστα Ακουστικά Χαρακτηριστικά για Αυτόματη Αναγνώριση Φωνής από Απόσταση

Νικόλαος Φλεμοτόμος
Επιβλέπων: Καθ. Πέτρος Μαραγκός

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Σημάτων, Ελέγχου και Ρομποτικής
Εργαστήριο Όρασης Υπολογιστών, Επικοινωνίας Λόγου και Επεξεργασίας Σημάτων

29 Μαρτίου 2016

Περιεχόμενα

Περιεχόμενα

Αυτόματη Αναγνώριση Φωνής

Υπολογιστική διαδικασία, όπου:

- Είσοδος: Ακουστικό κύμα / Ηχητικό σήμα
- Έξοδος: Κείμενο

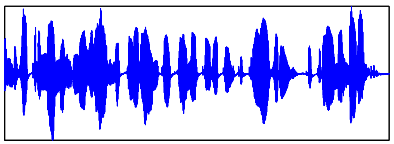
Στόχος: Ταύτιση μεταξύ παραγόμενου κειμένου και αρχικού λόγου.

Αυτόματη Αναγνώριση Φωνής

Υπολογιστική διαδικασία, όπου:

- Είσοδος: Ακουστικό κύμα / Ηχητικό σήμα
- Έξοδος: Κείμενο

Στόχος: Ταύτιση μεταξύ παραγόμενου κειμένου και αρχικού λόγου.

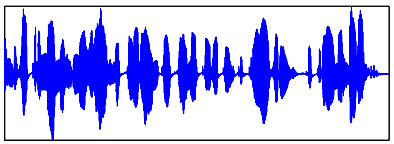


Αυτόματη Αναγνώριση Φωνής

Υπολογιστική διαδικασία, όπου:

- Είσοδος: Ακουστικό κύμα / Ηχητικό σήμα
- Έξοδος: Κείμενο

Στόχος: Ταύτιση μεταξύ παραγόμενου κειμένου και αρχικού λόγου.

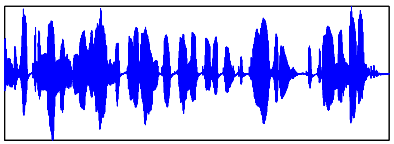


Αυτόματη Αναγνώριση Φωνής

Υπολογιστική διαδικασία, όπου:

- Είσοδος: Ακουστικό κύμα / Ηχητικό σήμα
- Έξοδος: Κείμενο

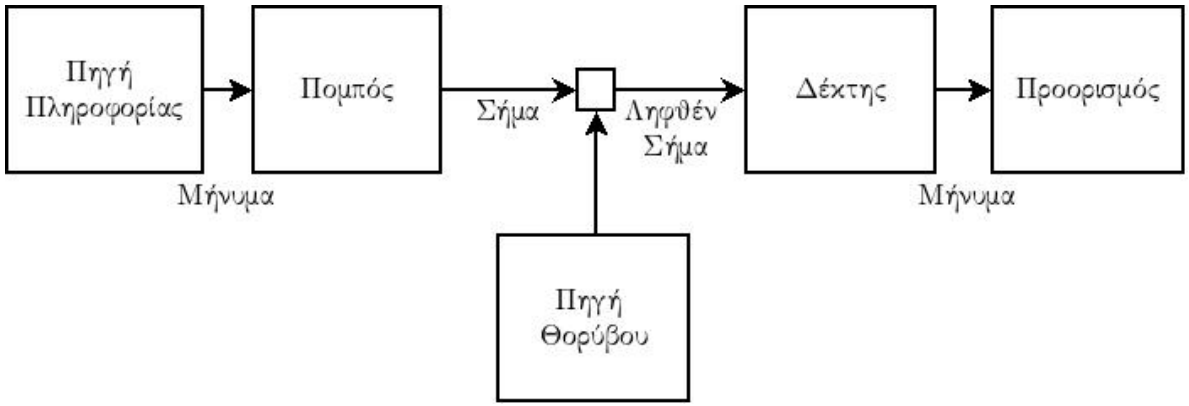
Στόχος: Ταύτιση μεταξύ παραγόμενου κειμένου και αρχικού λόγου.



Επιθυμητή έξοδος

Στη συγκεκριμένη συνεδρίαση κληρώνεται και ο αναπληρωτής του προέδρου που προαναφέραμε.

Αναγνώριση Φωνής από Απόσταση



Σχήμα : Ομιλία υπό τη σκοπιά της Θεωρίας Πληροφορίας.

Όταν το μικρόφωνο απομακρύνεται από το στόμα του ομιλητή, εισάγονται αλλοιώσεις που οφείλονται σε

- θόρυβο υποβάθρου,
- αντήχηση,
- διεύθυνση κεφαλιού, φαινόμενο Lombart, κ.λπ.

Βασικά Στάδια ενός Αναγνωριστή

0. Καταγραφή, Δειγματοληψία, Κβάντιση
1. Εξαγωγή χαρακτηριστικών
2. Ακουστικό μοντέλο

$$P(W|O) = \frac{\mathbf{p(O|W)}P(W)}{P(O)}$$

3. Γλωσσικό μοντέλο

$$P(W|O) = \frac{p(O|W)\mathbf{P(W)}}{P(O)}$$

4. Αποκωδικοποίηση

$$\hat{W} = \underset{W \in \mathcal{W}}{\operatorname{argmax}}\, p(O|W)P(W)^{LMSF}\, WIP^{N(W)}$$

Κίνητρα και Στόχος

- καλύτερο γλωσσικό μοντέλο $\Rightarrow \downarrow PP \Rightarrow \downarrow WER$
  $\Rightarrow$ Πολλές φορές η αξία των ακουστικών χαρακτηριστικών παραβλέπεται χάριν της καλύτερης γλωσσικής μοντελοποίησης.
- Η συσχέτιση $PP - WER$ είναι ίδια μεταξύ HSR και ASR.
  Όμως: απόδοση HSR $\gg$ απόδοση ASR
- Συνήθως η αξιολόγηση νέων συνόλων χαρακτηριστικών γίνεται σε καθαρές ή σε ελεγχόμενες συνθήκες θορύβου.
- Αδύνατη η εκπαίδευση σε πραγματικές συνθήκες, προβληματική με μεθόδους τεχνητής αλλοίωσης.

Στόχος: Συγκριτική μελέτη διαφορετικών συνόλων χαρακτηριστικών

- πραγματικές συνθήκες DSR
- μεγάλη αναντιστοιχία μεταξύ εκπαίδευσης και ελέγχου
- ελαχιστοποίηση επίδρασης λοιπών παραγόντων

Κεντρικοί Άξονες

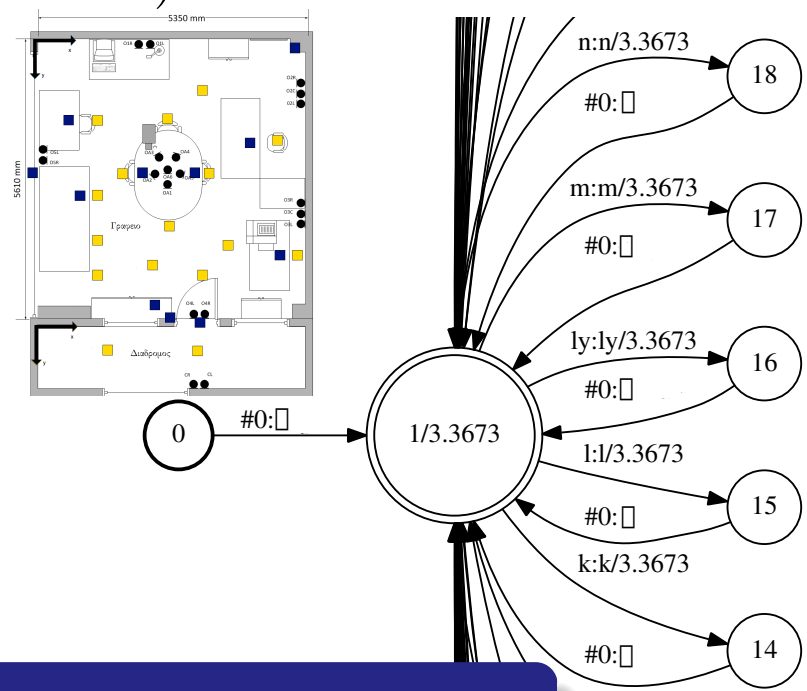
Αναγνώριση Φωνημάτων (Αξιολόγηση βάσει *PER*)

γλωσσικό μοντέλο

γραμματική ισοπίθανων μεταβάσεων ⇒ ελαχιστοποίηση επίδρασης του context

ακουστικό μοντέλο

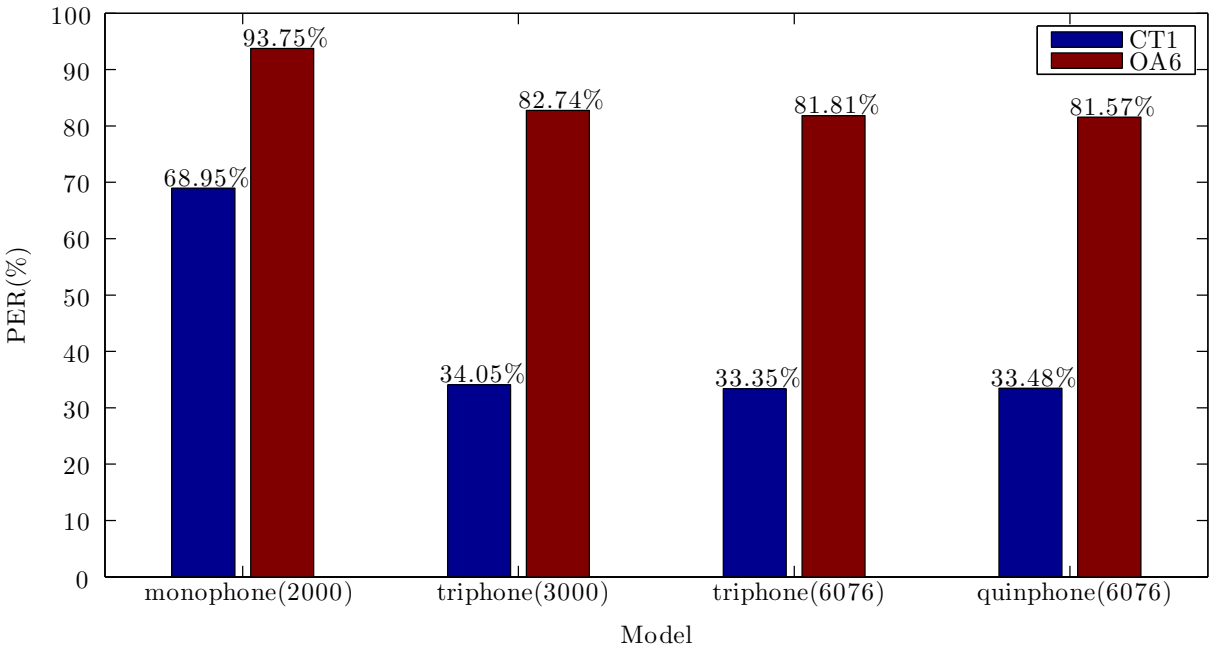
καθιερωμένο πρότυπο HMMs/GMMs, συνάρθρωση ⇒ τριφωνικό μοντέλο



βάσεις δεδομένων

εκπαίδευση: 6076 καθαρές εκφορές από Logotypografia ($\sim 8.6sec$)
ανάπτυξη: 190 εκφορές από ATHENA ($\sim 2.7sec$) [OA6 και CT1]
έλεγχος: 190 εκφορές από ATHENA ($\sim 2.9sec$) [OA6 και CT1]

Το Σύστημα Αναγνώρισης στην Πράξη



βασικές παράμετροι

- max #states=2000
- max #gaussians=10000
- $LMSF \in \{1, 2, \cdots, 20\}$
- $PIP \in \{0, 0.5, 1\}$

βασικές παράμετροι

- ευθυγράμμιση: beam Viterbi, beam width = 8 ($\rightarrow$ 40)
- αποκωδικοποίηση: beam Viterbi, beam width = 13

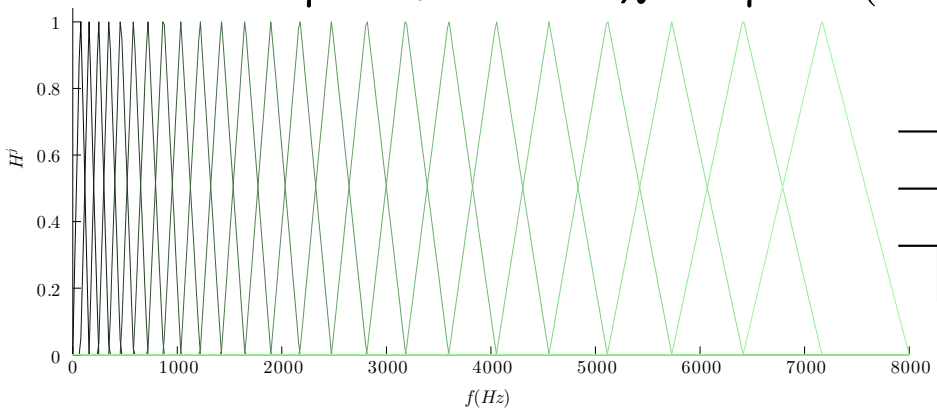
Περιεχόμενα

Εξαγωγή των MFCCs

- Προέμφαση (← ενίσχυση υψηλών συχνοτήτων)
  $H_{preemph}(z) = 1 - 0.97z^{-1}$
- Παραθύρωση Hamming (← quasistationarity)

	$15msec$	$20msec$	$25msec$	$32msec$
CT1	33.86	33.13	**32.95**	33.35
OA6	82.04	84.40	83.14	**81.81**

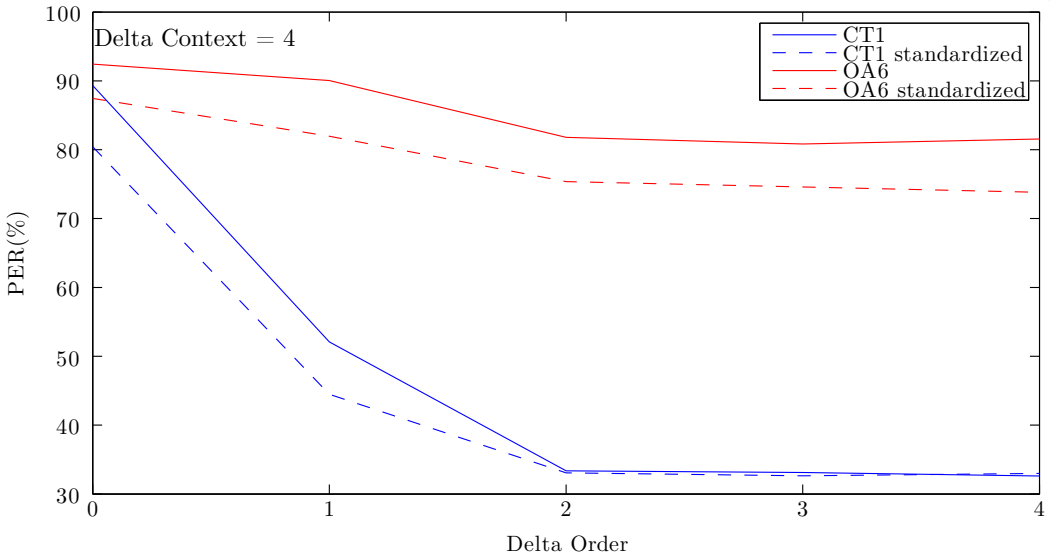
- Διάσπαση σε ζώνες συχνοτήτων (κλίμακα mel)

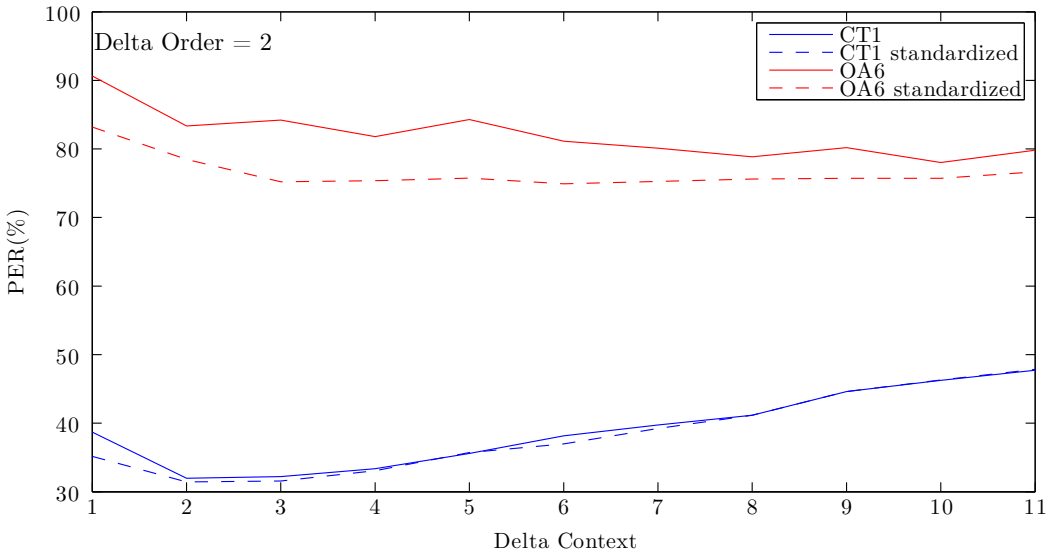


	CT1	OA6
$[0, 8000]Hz$	**33.35**	**81.81**
$[20, 7800]Hz$	33.60	84.03
$[130, 6800]Hz$	34.75	84.61

- Λογάριθμος του φάσματος ισχύος σε κάθε ζώνη
  $G_i(j) = \log\left\{\sum_{k=0}^{N/2} |S_i[k] \cdot H^j[k]|^2\right\}$
- DCT M/Σ (← συμπίεση + αποσυσχέτιση)
- dithering, liftering

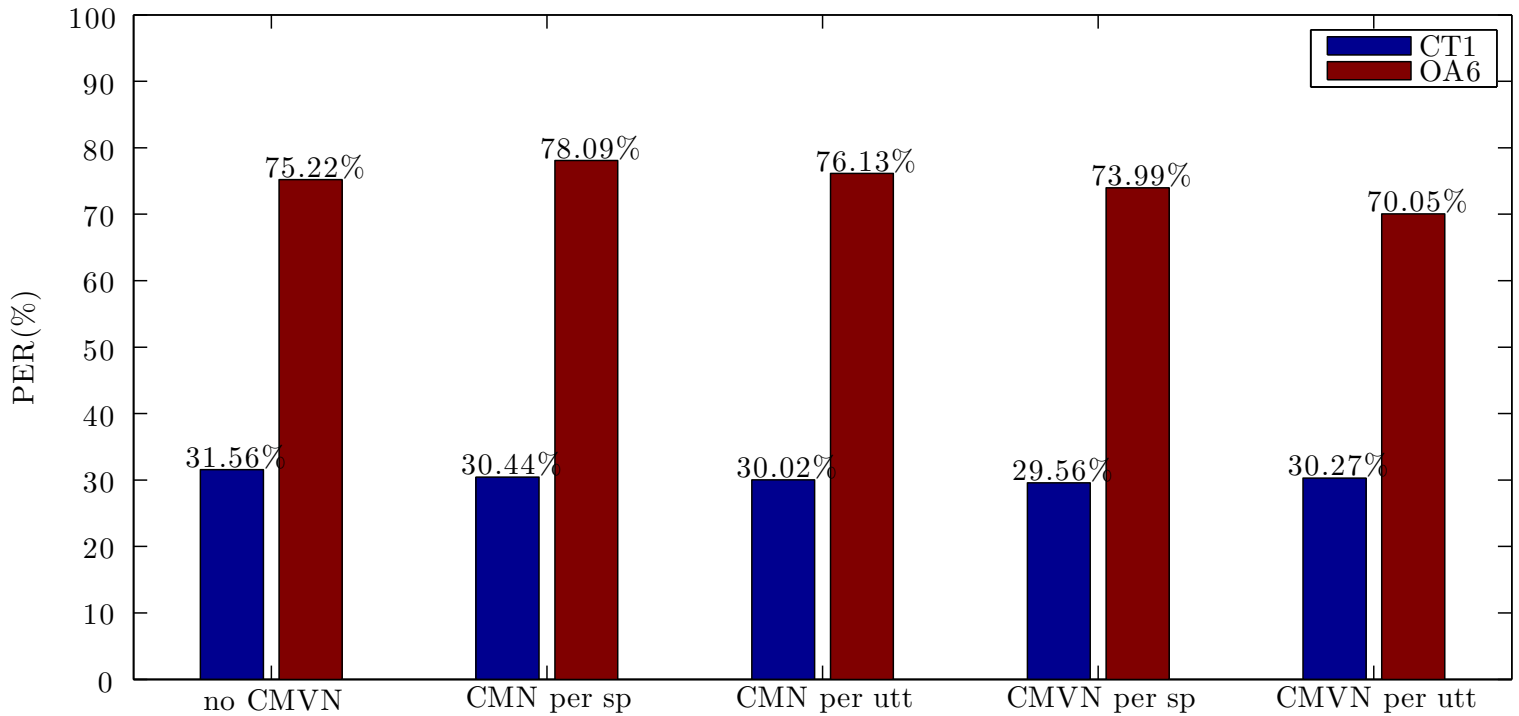
Παραγώγιση των MFCCs





$$\Delta x_i(k) = \frac{\sum_{m=-M}^{M} m \cdot x_{i+m}(k)}{\sum_{m=-M}^{M} m^2}$$

Cepstral Mean (& Variance) Normalization



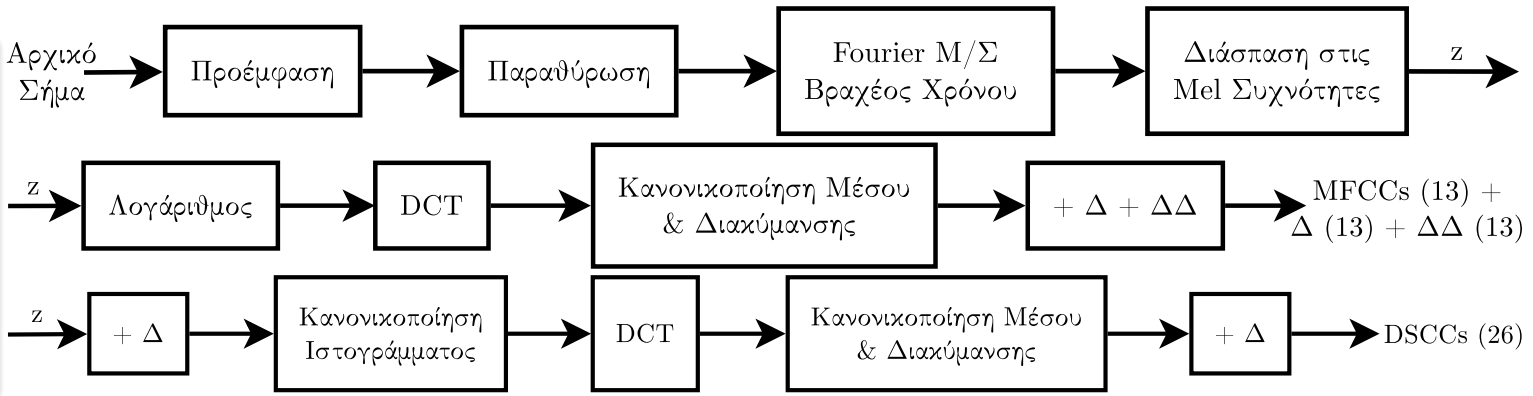
$$y_i(n) = x_i(n) * h(n) \Rightarrow Y_i(q) = X_i(q) + H(q)$$

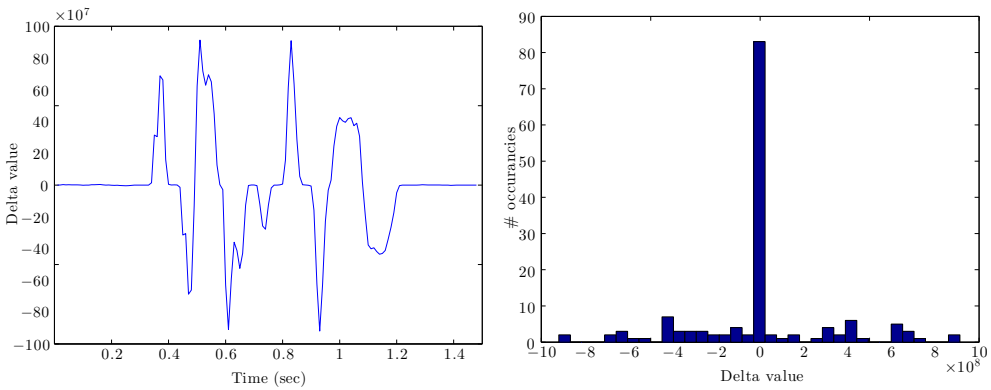
$$Y_i'(q) = Y_i(q) - \frac{1}{F}\sum_{i=1}^{F} Y_i(q) = X_i(q) - \frac{1}{F}\sum_{i=1}^{F} X_i(q)$$

Εξαγωγή των DSCCs

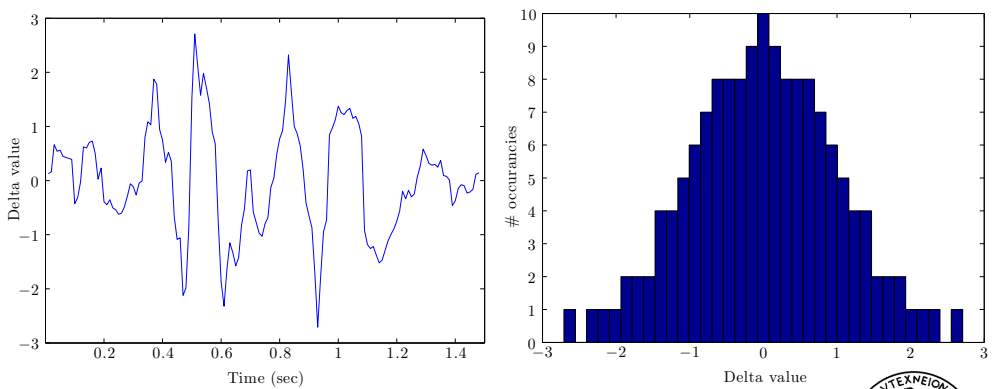
κινητήρια ιδέα

ρυθμός μεταβολής φασματικών χαρακτηριστικών φωνής vs θορυβώδους υποβάθρου



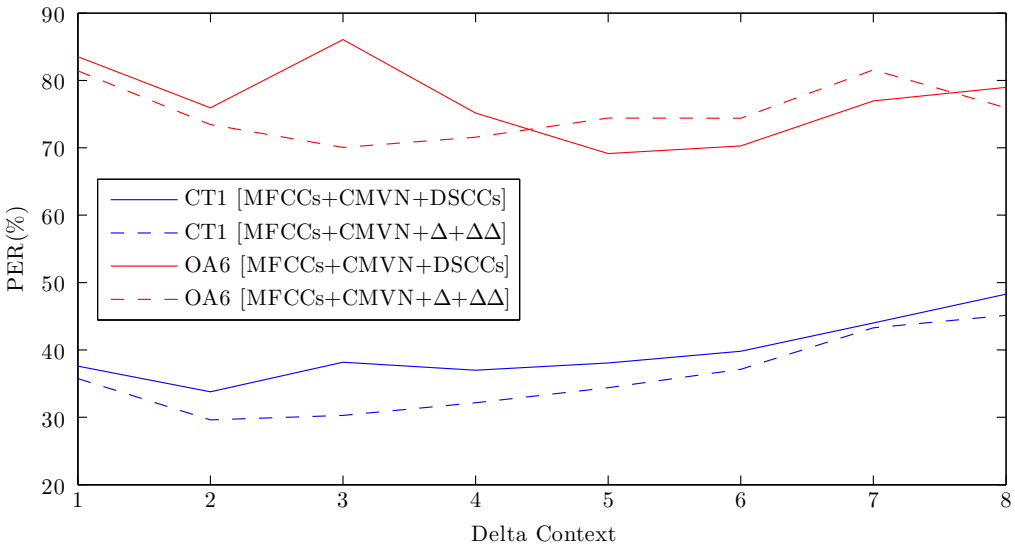


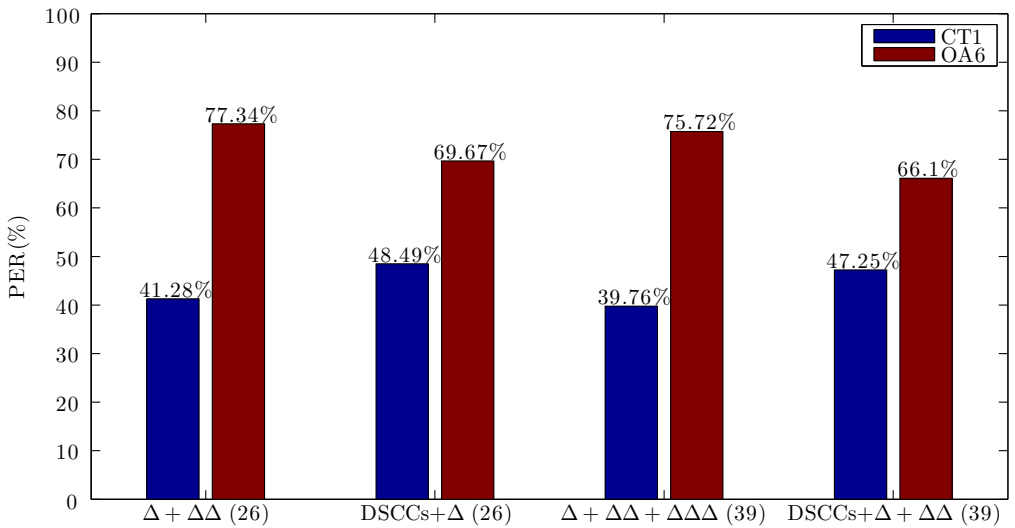
Πριν την Κανονικοποίηση Ι/Γ
(10ο φίλτρο)



Μετά την Κανονικοποίηση Ι/Γ
(10ο φίλτρο)

DSCCs - Πειραματικά Αποτελέσματα





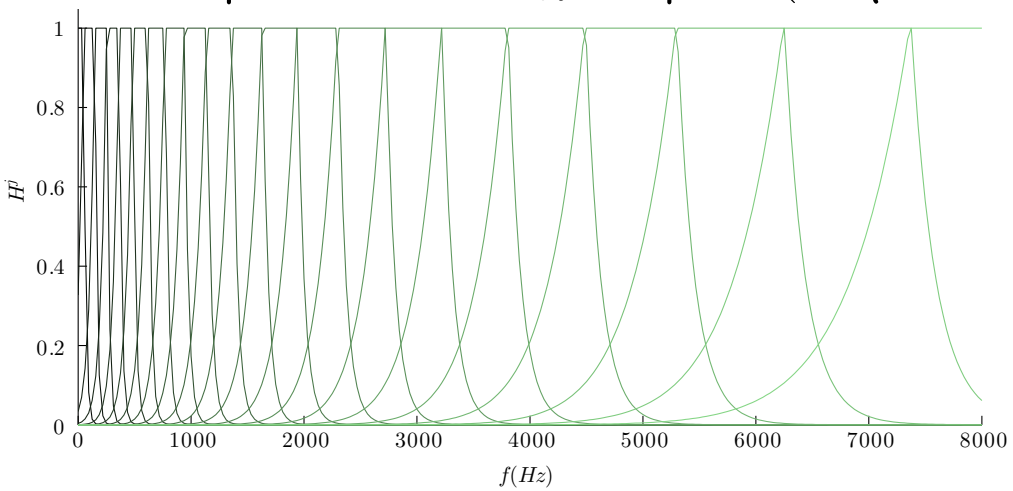
Εναλλακτικός υπολογισμός Δ

$$\Delta x_i(k) = x_{i+M} - x_{i-M}$$

Χρήση μόνο δυναμικών συντελεστών

Εξαγωγή των PLPs - I

- Παραθύρωση Hamming
- Διάσπαση σε ζώνες συχνοτήτων (κλίμακα Bark)



- Φάσμα ισχύος σε κάθε ζώνη $G_i(j) = \sum_{k=0}^{N/2} \left\{ |S_i[k]|^2 \cdot |H^j[k]| \right\}$
- Φίλτρο ίσων επιπέδων έντασης

$$E(\omega) = \frac{\omega^4 (\omega^2 + 56.8 \cdot 10^6)}{(\omega^2 + 6.3 \cdot 10^6)^2 (\omega^2 + 0.38 \cdot 10^9)(\omega^6 + 9.58 \cdot 10^{26})} \quad (\text{για } Fs > 10kHz)$$

$$\tilde{G}_i(j) = \begin{cases} \tilde{G}_i(2) & , j = 1 \\ \sum_{k=0}^{N/2} \left\{ |S_i[k]|^2 \cdot |E[k] H^j[k]| \right\} & , 2 \leq j \leq Q-1 \\ \tilde{G}_i(Q-1) & , j = Q \end{cases}$$

Εξαγωγή των PLPs - II

- Σχέση διέγερσης - ψυχολογικής έντασης $\Phi_i(j) = \left(\tilde{G}_i(j)\right)^{0.33}$
- Εκτίμηση συντελεστών LP $s_i(n) = \sum_{k=1}^{p} a_{i,k} s_i(n-k) + G_i u_i(n)$

  (θεώρημα Wiener-Khinchin και αλγόριθμος Levinson-Durbin)

  $R_{\phi\phi} = IDFT\{\tilde{\Phi}\}$

  $\tilde{\Phi}_i = [\Phi_i(1), \Phi_i(2), \cdots, \Phi_i(Q-1), \Phi_i(Q), \Phi_i(Q-1), \cdots, \Phi_i(2)]$
- Μετασχηματισμός στο πεδίο cepstrum

$$\hat{\phi}_i(j) = \begin{cases} \log G_i & , j = 0 \\ a_{i,j} + \sum_{k=1}^{j-1} \left(\frac{k}{j}\right) \hat{\phi}_i(k) a_{i,j-k} & , 1 \le j \le p \end{cases}$$

Ιδέα της RASTA Ανάλυσης

γιατί;

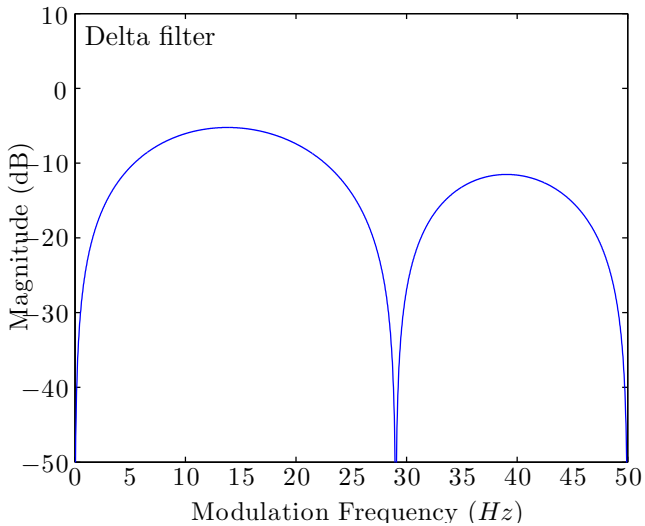
μέγιστο ευαισθησίας ακοής στα $\sim 4Hz$

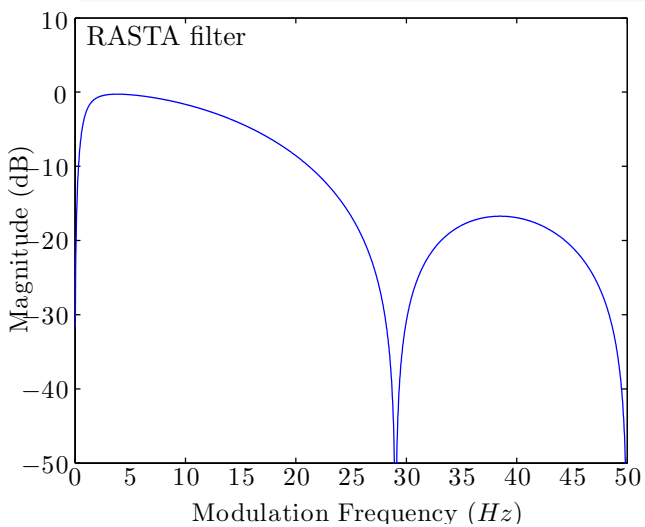
πότε;

μετά τη διάσπαση σε ζώνες συχνοτήτων *πριν* το φίλτρο ίσων επιπέδων έντασης

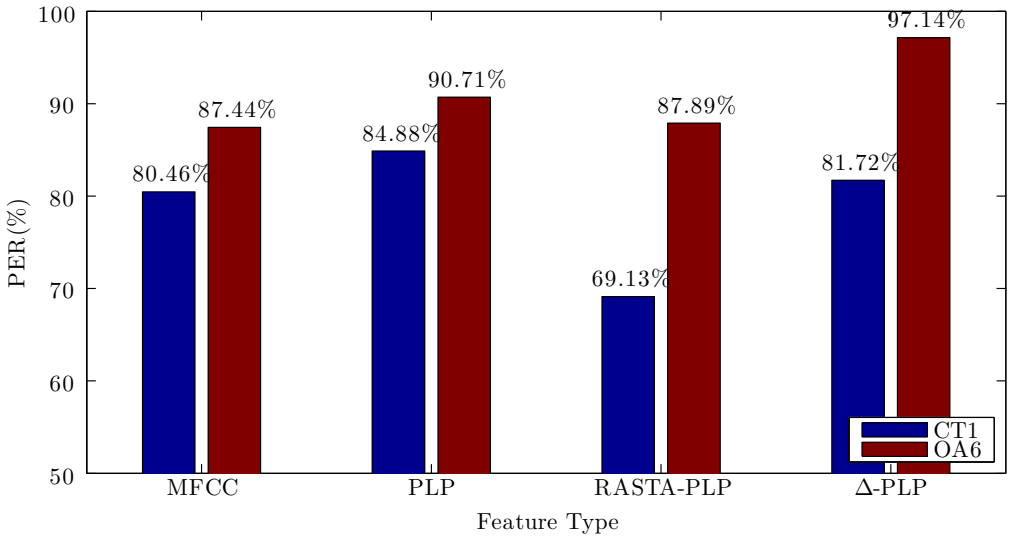
πώς;

- $T[x] = y = \log(x)$
  $T[x] = \log(1 + Jx)$ [J-RASTA]
- $H(z) = 0.1z^4 \frac{2+z^{-1}-z^{-3}-2z^{-4}}{1-0.94z^{-1}}$
- $T^{-1}[y] = x = e^y$
  $T^{-1}[y] \approx \frac{e^y}{J}$ [J-RASTA]



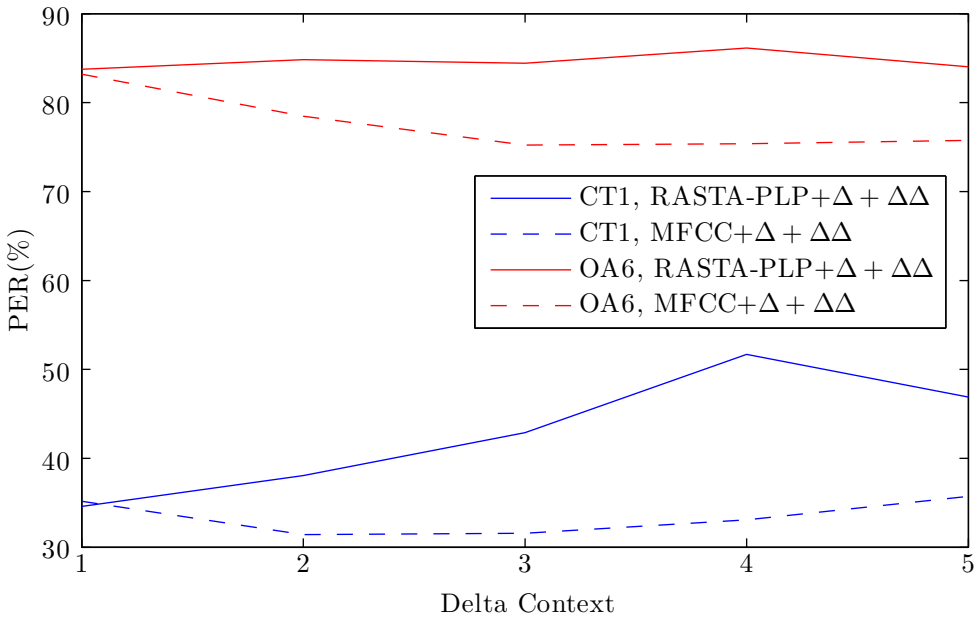


PLPs και RASTA-PLPs: Αποτελέσματα



	10^{-8}	10^{-7}	10^{-6}	10^{-5}
CT1	84.94	80.63	69.22	68.87
OA6	89.51	87.87	88.46	87.82
	10^{-4}	10^{-3}	10^{-2}	10^{-1}
CT1	75.72	68.51	78.12	69.10
OA6	88.79	88.40	89.09	87.45
	10^{1}	10^{2}	10^{3}	10^{4}
CT1	70.19	**68.45**	69.64	84.67
OA6	**87.39**	87.86	87.56	90.04

J-RASTA για διάφορα J



Περιεχόμενα

Λίγη Θεωρία - I

Στόχος

Εύρεση μήτρας $\mathbf{T}$ ώστε

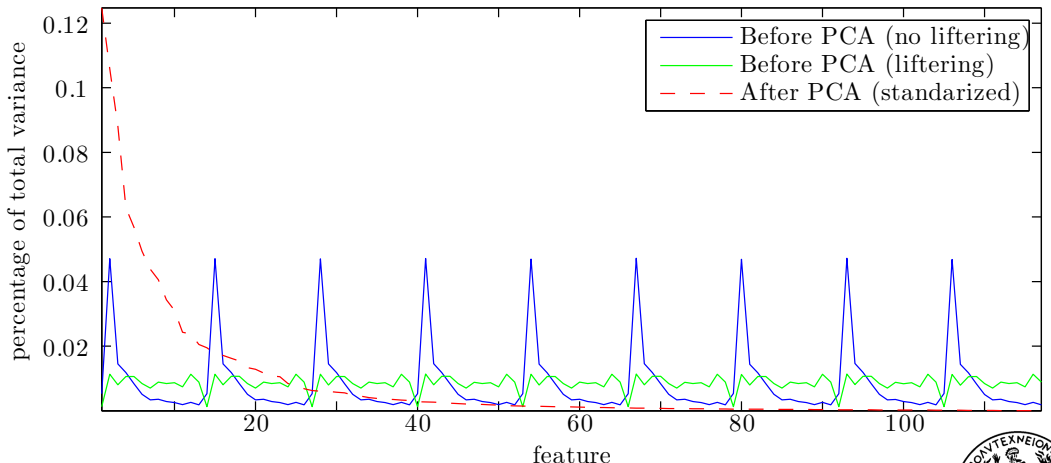
$\tilde{\mathbf{x}}_i = \mathbf{T}\left[\mathbf{x}_{i-M}^T, \mathbf{x}_{i-M+1}^T, \cdots, \mathbf{x}_i^T, \cdots, \mathbf{x}_{i+M-1}^T, \mathbf{x}_{i+M}^T\right]^T$

$\Rightarrow$ καλύτερη περιγραφή της δυναμικής του σήματος

Δ συντελεστές $\rightarrow$ μια υποπερίπτωση

PCA

- μη-επιβλεπόμενη μάθηση
- μεγιστοποίηση του ρυθμού μείωσης της διακύμανσης
- προβολή στον υποχώρο όπου εκτείνονται τα πρώτα ιδιοδιανύσματα του πίνακα συσχέτισης των δεδομένων



Χρήση 13 MFCCs από τα δεδομένα του συνόλου ελέγχου

Λίγη Θεωρία - II

LDA

- επιβλεπόμενη μάθηση
- μεγιστοποίηση διαταξικών αποστάσεων ελαχιστοποίηση ενδοταξικών
- πρόβλημα γενικευμένων ιδιοτιμών

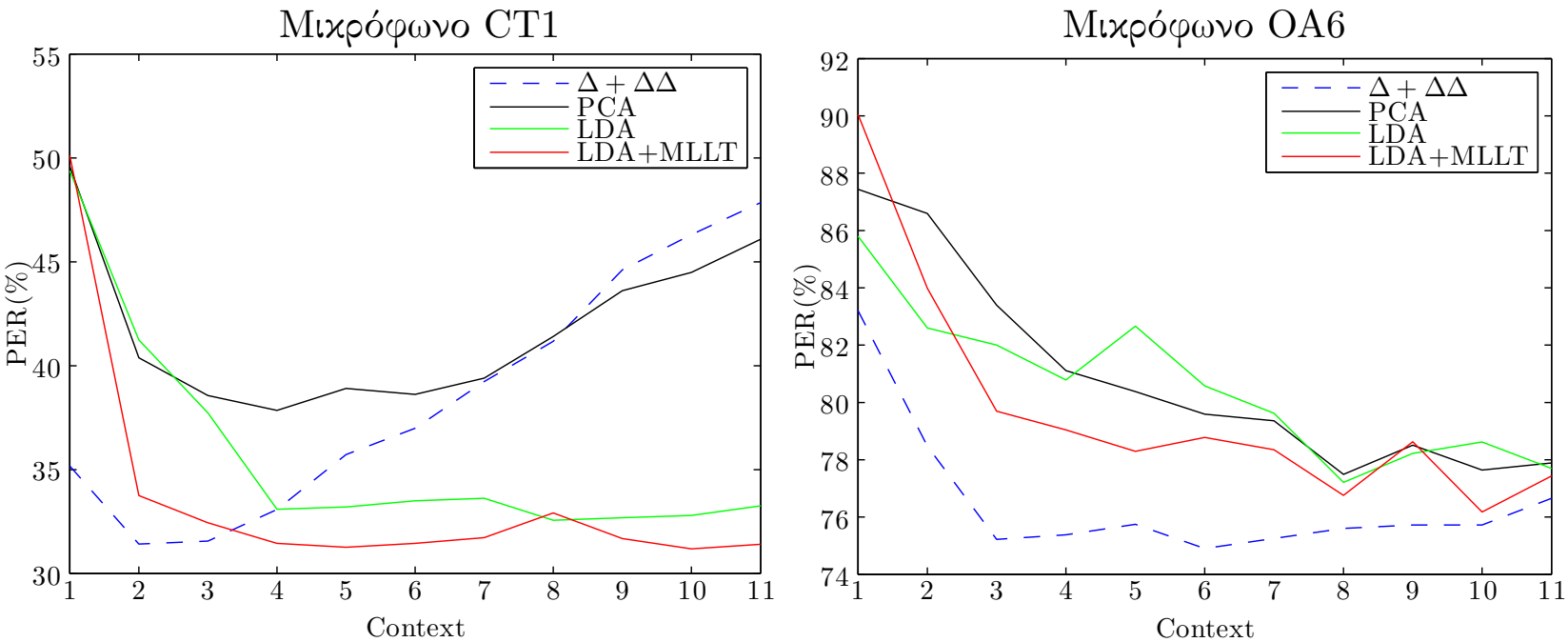
HLDA

- γενίκευση LDA σε περίπτωση ετεροσκεδαστικών δεδομένων
- δεν υπάρχει αναλυτική λύση

STC / MLLT

- STC συμβιβαστική λύση μεταξύ διαγωνίων και πλήρων πινάκων συμμεταβλητότητας
- 1 πλήρης πίνακας $\Rightarrow$ global STC ή MLLT
- ελαχιστοποίηση διαφοράς στην πιθανοφάνεια μεταξύ μοντελοποίησης με πλήρεις και με διαγώνιους πίνακες
- τελική συνάρτηση προς βελτιστοποίηση παρόμοια σε HLDA και MLLT
- στην πράξη: LDA+MLLT (HLDA χειρότερα αποτελέσματα)

PCA vs LDA vs LDA+MLLT και Context



- εξαγωγή των 13 MFCCs σε κάθε πλαίσιο
- μονοφωνικό + πρώτο πέρασμα τριφωνικού μοντέλου: MFCCs+Δ+ΔΔ (με Context=3) ⇒ εξαγωγή κλάσεων
- MLLT 5 επαναλήψεων

Περιεχόμενα

TEO vs SEO

Ενεργειακό περιεχόμενο σήματος → σημαντικό στοιχείο κατά την εξαγωγή χαρακτηριστικών για ASR
⇒ Πώς θα αποδοθεί καλύτερα;

SEO

$$S_c[x(t)] \triangleq x^2(t)$$
$$S_d[s[n]] \triangleq s^2[n]$$

TEO

$$\Psi_c[x(t)] \triangleq \dot{x}^2(t) - x(t)\ddot{x}(t)$$
$$\Psi_d[s[n]] \triangleq s^2[n] - s[n-1]s[n+1]$$

ελεύθερος ταλαντωτής

$$x(t) = A\cos(\omega_0 t + \theta), \omega_0 = \sqrt{\frac{k}{m}}$$
$$E_0 = \frac{m}{2}A^2\omega_0^2$$
$$\Psi_c[x(t)] = A^2\omega_0^2 = \frac{E_0}{(m/2)}$$

AM-FM σήματα

$$\Psi_c\left[a(t)\cos\left(\int_0^t \omega(\tau)d\tau + \theta\right)\right] \approx a^2(t)\omega^2(t)$$

σήματα φωνής:

$$s(t) = \sum_{r=1}^{R} a_r(t)\cos(\phi_r(t))$$

TEO στη Συχνότητα και TPS - I

TEO στο χρόνο $\Rightarrow$ συνελίξεις με συστοιχία φίλτρων $\Rightarrow$ $\uparrow$υπολογιστική πολυπλοκότητα

Υπάρχουσες προσεγγίσεις για χρήση TEO στη συχνότητα

- $|\Psi[S[k]]| = \left|S^2[k] - S[k-1]S[k+1]\right|$
- $\Phi[S[k]] \triangleq \Psi[\mathrm{Re}\{S[k]\}] + \Psi[\mathrm{Im}\{S[k]\}]$

Χρήση Θ. Parseval και Plancherel

$$\mathrm{SEO} \rightarrow \sum_{n=0}^{N-1} s^2[n] = \frac{1}{N}\sum_{k=0}^{N-1} |S[k]|^2 \text{ (Φάσμα Ισχύος)}$$

$$\mathrm{TEO} \rightarrow \sum_{n=0}^{N-1} \Psi[s[n]] = \sum_{n=0}^{N-1} s^2[n] - \sum_{n=0}^{N-1} s[n-1]s[n+1]$$

$$= \frac{1}{N}\sum_{k=0}^{N-1} \left\{ |S[k]|^2 - S_-[k]S_+^*[k] \right\} \text{ (Φάσμα Teager Ισχύος)}$$

TEO στη Συχνότητα και TPS - II

- Έστω σήμα s N' δειγμάτων
$$s = \{s[1], s[2], s[3], \cdots, s[N'-2], s[N'-1], s[N']\}$$
- Ορίζουμε τα σήματα
$$\tilde{s} \triangleq \{s[2], s[3], \cdots, s[N'-2], s[N'-1]\},$$
$$\tilde{s}_- \triangleq \{s[1], s[2], s[3], \cdots, s[N'-2]\},$$
$$\tilde{s}_+ \triangleq \{s[3], \cdots, s[N'-2], s[N'-1], s[N']\}$$
- Παραθυρώνονται όμοια και τα 3 και σε κάθε δείγμα k του i-πλαισίου ορίζεται το TPS:
$$S_{(t)i}[k] \triangleq |\tilde{S}_i[k]|^2 - \tilde{S}_{i-}[k]\tilde{S}^*_{i+}[k]$$
- Μπορεί να χρησιμοποιηθεί στη ροή εργασίας γνωστών μεθόδων αντί του Φάσματος Ισχύος:

$$G_i(j) = \log \sum_{k=0}^{N/2} \left\{ \left| S_i^2[k] \cdot H^j[k] \right| \right\} \rightsquigarrow G_i(j) = \log \sum_{k=0}^{N/2} \left\{ \left| S_{(t)i}[k] \cdot H^j[k] \right| \right\}$$

Συνδυασμός TEO και SEO

κίνητρο & ιδέα

- σφάλματα διακριτοποίησης TEO μεγαλύτερα στις υψηλές συχνότητες
- προσθετικός θόρυβος επηρεάζει τα αποτελέσματα του TEO περισσότερο στις υψηλές συχνότητες

$\Rightarrow$ χρήση TEO για πρώτα φίλτρα της συστοιχίας, SEO για τα υπόλοιπα:

$$G_i(j) = \begin{cases} \log \sum_{k=0}^{N/2} \left\{ \left| S_{(t)i}[k] \cdot H^j[k] \right| \right\}, j \leq M \\ \log \sum_{k=0}^{N/2} \left\{ \left| S_i^2[k] \cdot H^j[k] \right| \right\}, j > M \end{cases}$$

	$M=0$	Best M	$M=Q$
		CT1	
MFCC	31.56	**30.17** ($M=27$)	31.09
PLP	45.64	**35.43** ($M=13$)	43.70
SPNCC	30.76	**29.46** ($M=11$)	30.03
		OA6	
MFCC	77.67	**75.71** ($M=23$)	77.25
PLP	90.03	**80.97** ($M=16$)	87.28
SPNCC	78.60	**75.75** ($M=23$)	78.02

Αποτελέσματα της μεθόδου - PER(%)

Αποδιαμόρφωση AM-FM Σημάτων

Hilbert M/Σ

$$|a(t)| \approx \sqrt{s^2(t) + \hat{s}^2(t)}$$
$$\omega(t) \approx \frac{d}{dt}\left(\arctan\frac{\hat{s}(t)}{s(t)}\right)$$

ESA

$$|a(t)| \approx \frac{\Psi[s(t)]}{\sqrt{\Psi[\dot{s}(t)]}}$$
$$\omega(t) \approx \sqrt{\frac{\Psi[\dot{s}(t)]}{\Psi[s(t)]}}$$

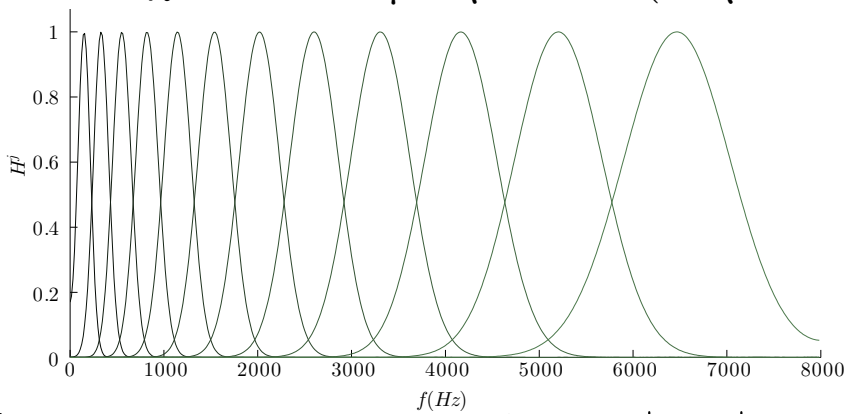
- παρόμοια αποτελέσματα
- ESA $\downarrow$ υπολογιστική πολυπλοκότητα

Gabor ESA

- ζωνοπερατό φιλτράρισμα με Gabor φίλτρο
  $g(t) = e^{-a^2t^2}\cos(2\pi f_c t + \phi)$
- εφαρμογή TEO στο ζωνοπεριορισμένο σήμα $s(t) * g(t)$
  $$\Psi[s(t)*g(t)] = \left(s(t) * \frac{dg(t)}{dt}\right)^2 - (s(t)*g(t))\left(s(t) * \frac{d^2g(t)}{dt^2}\right)$$
  στην πράξη διακριτή συνέλιξη με $g[n] = g(t)|_{t=nT}$
- εφαρμογή ESA

Εξαγωγή AM-FM Χαρακτηριστικών

- Κανονικοποίηση σήματος στο πεδίο του χρόνου
- Εφαρμογή Gabor ESA σε κάθε πλαίσιο i (διάρκειας $32msec$)
  - βαθυπερατό φιλτράρισμα αποτελεσμάτων TEO με διωνυμικό φίλτρο $\frac{1}{16}[1\ 4\ 6\ 4\ 1]$
  - median φίλτρο μήκους 7 στα τελικά αποτελέσματα
  - συστοιχία Gabor φίλτρων H^j (κλίμακα mel)



  $\Rightarrow$ εξαγωγή στιγμιαίου πλάτους $|a_{i,j}|$ και συχνότητας $f_{i,j}$
- Υπολογισμός AM-FM χαρακτηριστικών
- Χαρακτηριστικά που αφορούν στο πλάτος $\rightsquigarrow$ λογαριθμικό πεδίο
- Κανονικοποίηση χαρακτηριστικών
- Επαύξηση με Δ και $\Delta\Delta$ συντελεστές (Context $= 3$)

Πρώτα Αποτελέσματα

AM-FM Χαρακτηριστικά

$$B_i^2(j) = \frac{\sum(\dot{a}_{i,j}^2 + (f_{i,j} - F_i(j))^2 |a_{i,j}|^2}{\sum |a_{i,j}|^2}$$

$$F_i(j) = \frac{\sum f_{i,j} |a_{i,j}|^2}{\sum |a_{i,j}|^2}$$

$$FMP_i(j) = \frac{B_i(j)}{F_i(j)}$$

$$AMP_i(j) = |a_{i,j}|^T |a_{i,j}|$$

$$MAXA_i(j) = \max |a_{i,j}|$$

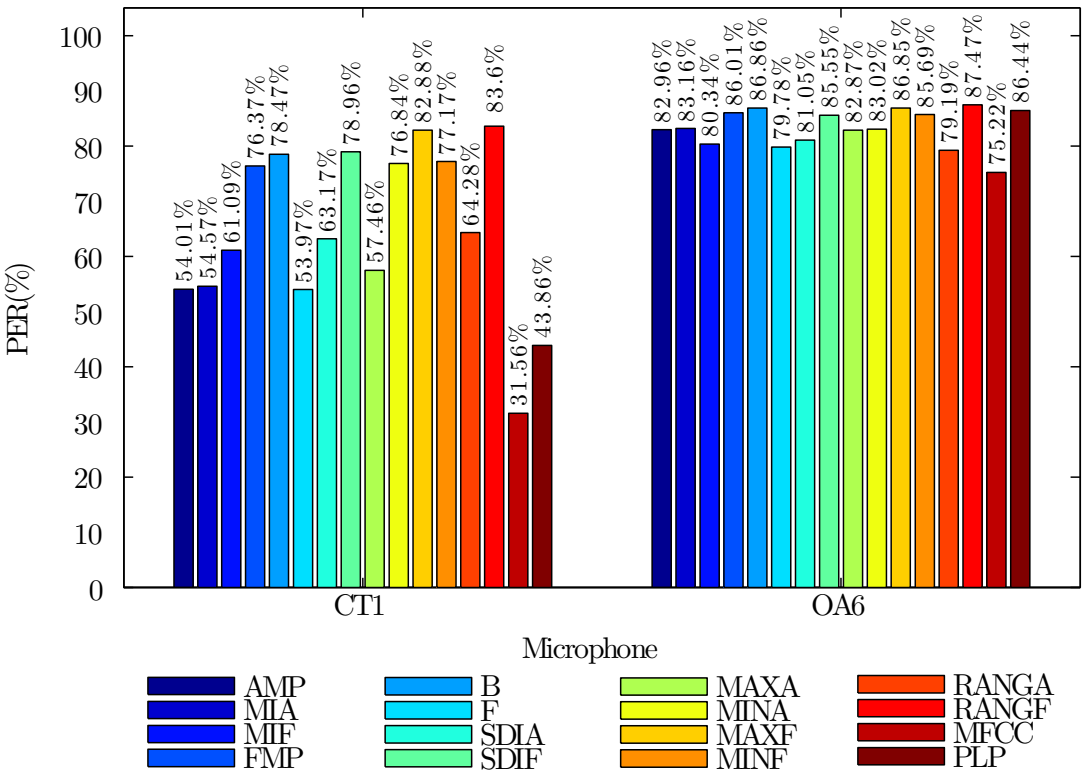
$$MINA_i(j) = \min |a_{i,j}|$$

$$RANGA_i(j) = MAXA_i(j) - MINA_i(j)$$

$$MIA_i(j) = \frac{1}{K} \sum |a_{i,j}|$$

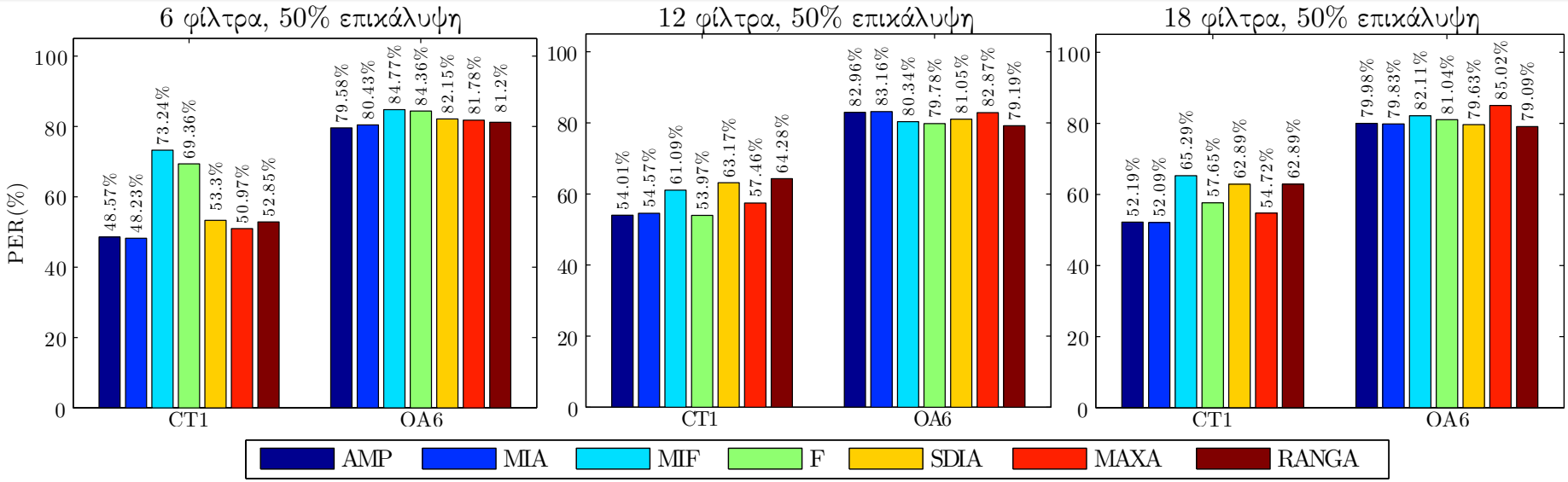
$$SDIA_i(j) = \sqrt{\frac{\sum(|a_{i,j}| - MIA_i(j))^2}{K-1}}$$

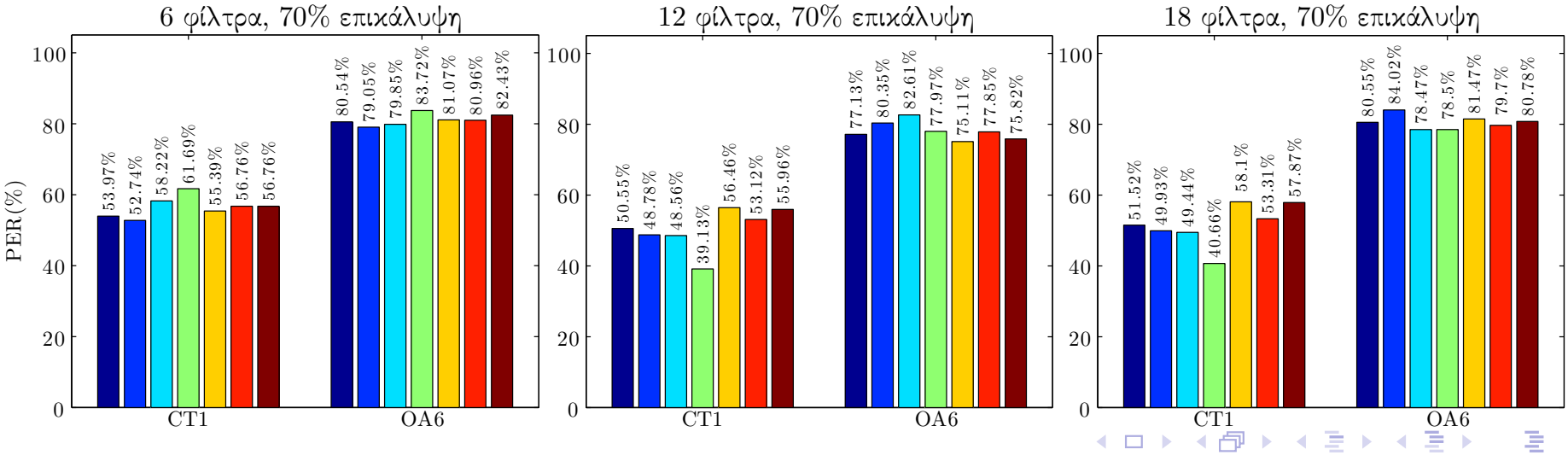
ομοίως $MAXF$, $MINF$, $RANGF$
MIF, $SDIF$



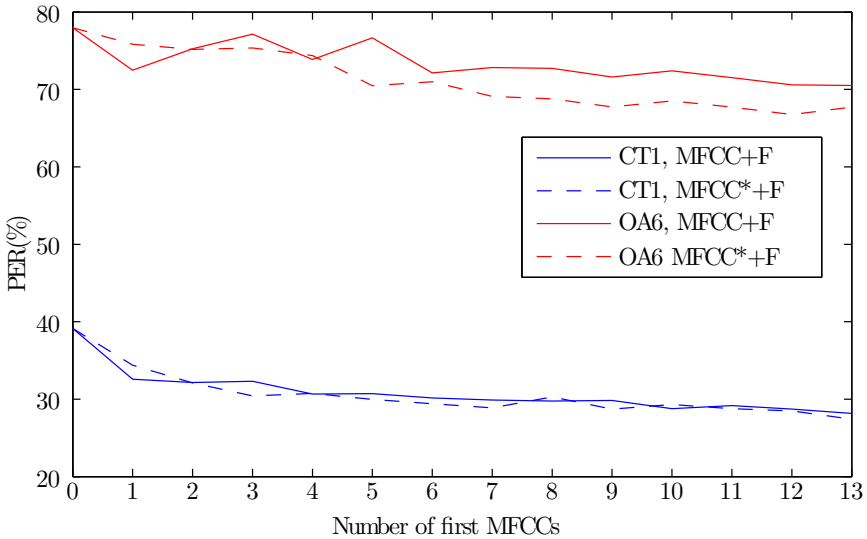
συστοιχία 12 Gabor φίλτρων, 50% επικάλυψη

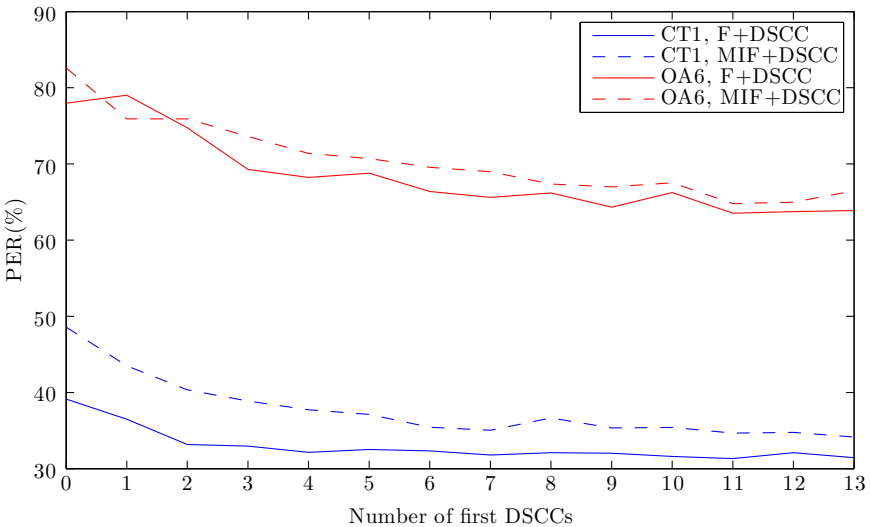
Αριθμός Φίλτρων και Ποσοστό Επικάλυψης

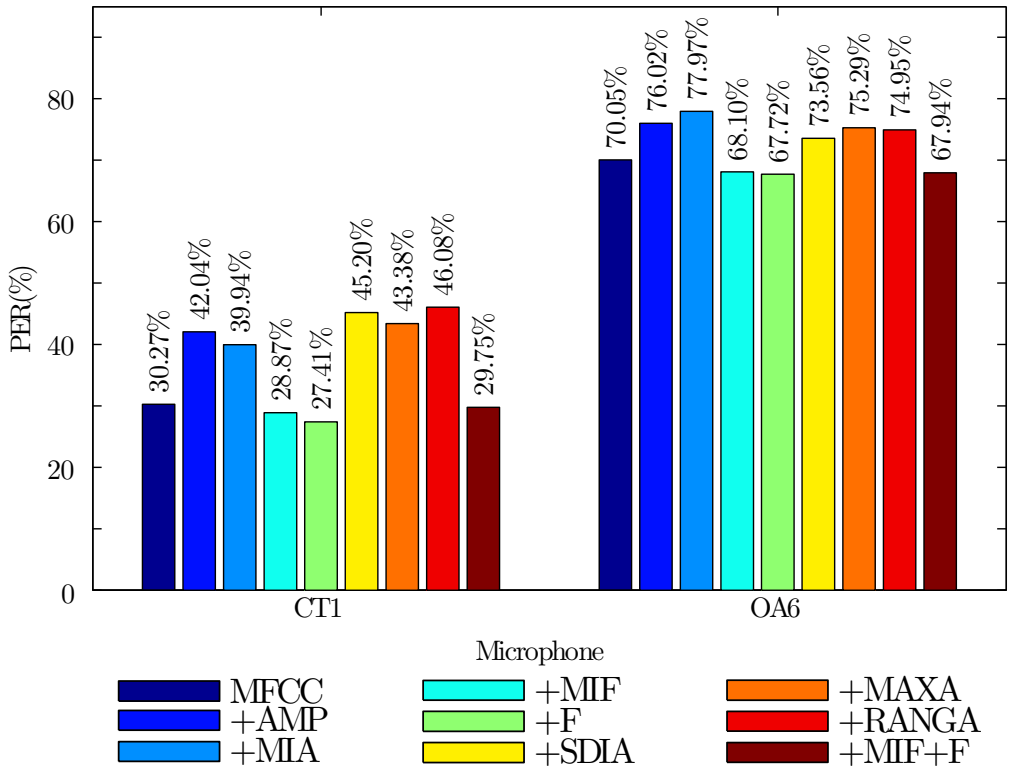




Συνδυασμός με MFCCs και DSCCs





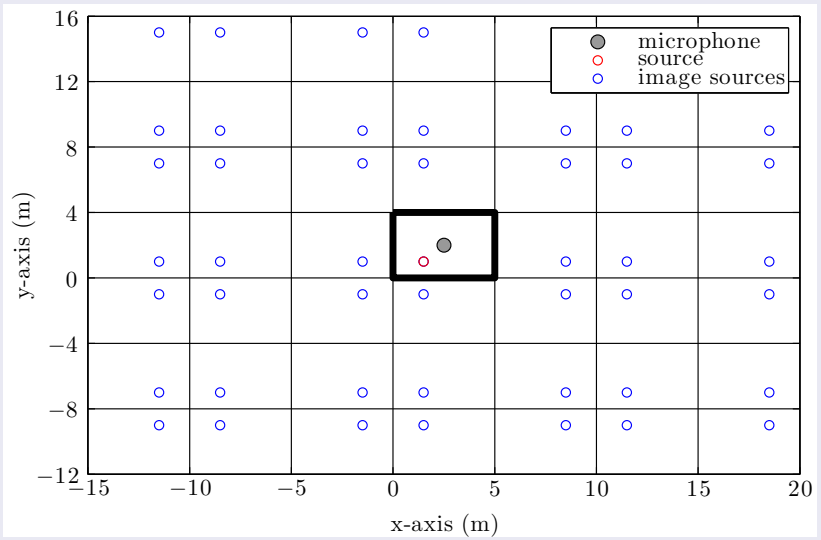


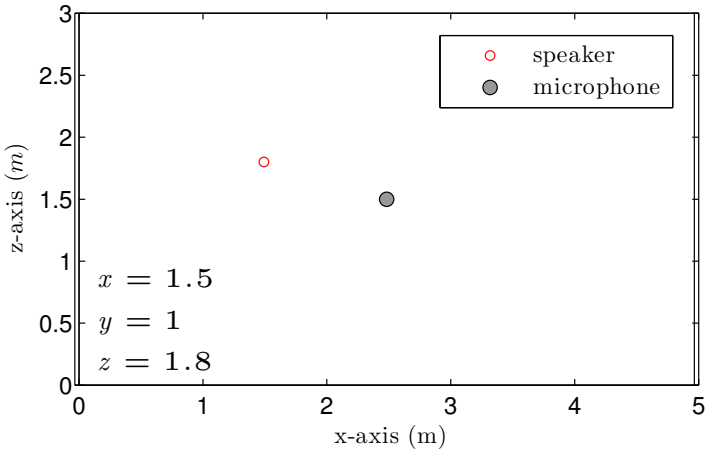
Σε κάθε περίπτωση, επαύξηση του τελικού διανύσματος με τους Δ και ΔΔ συντελεστές.

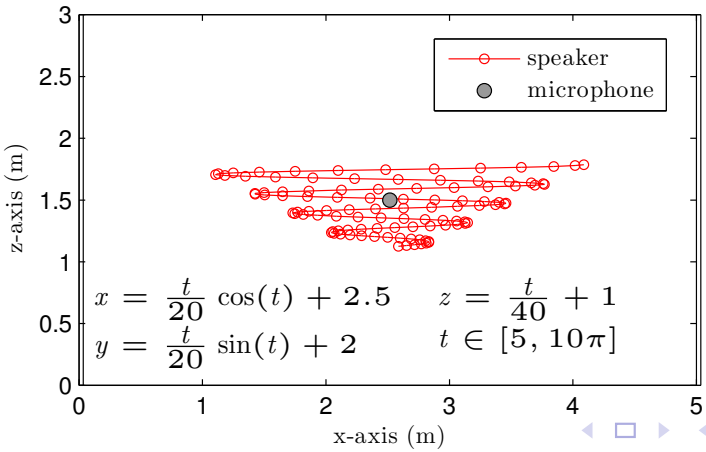
Παραγωγή Συνθετικών Δεδομένων

Image-Source Method

- Έστω s άμεσο κύμα $\Rightarrow y = s * RIR$
  $\rightsquigarrow$ εκτίμηση RIR;
- 1 τοίχος: συνεισφορά ανακλώμενου κύματος = συνεισφορά 2ης, κατοπτρικής πηγής (χρονική καθυστέρηση λόγω απόστασης + εξασθένηση λόγω απορρόφησης)
- 6 τοίχοι: πολλαπλές ανακλάσεις $\rightarrow$ κατοπτρισμός κάθε φανταστικής πηγής σε νέες $\Rightarrow$ 3D άπειρο πλέγμα

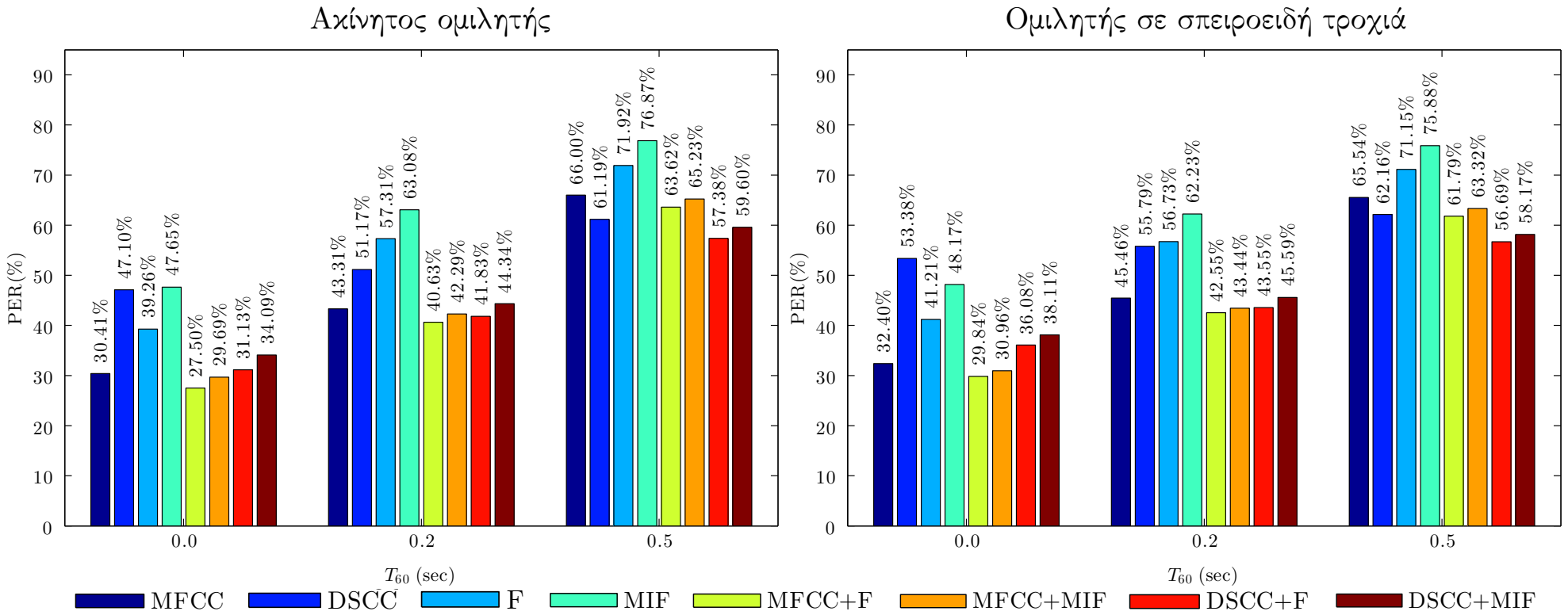






δωμάτιο: $5m \times 4m \times 3m$
θέση μικροφώνου: $(2.5, 2, 1.5)$

Αποτελέσματα σε Συνθετικά Δεδομένα - I

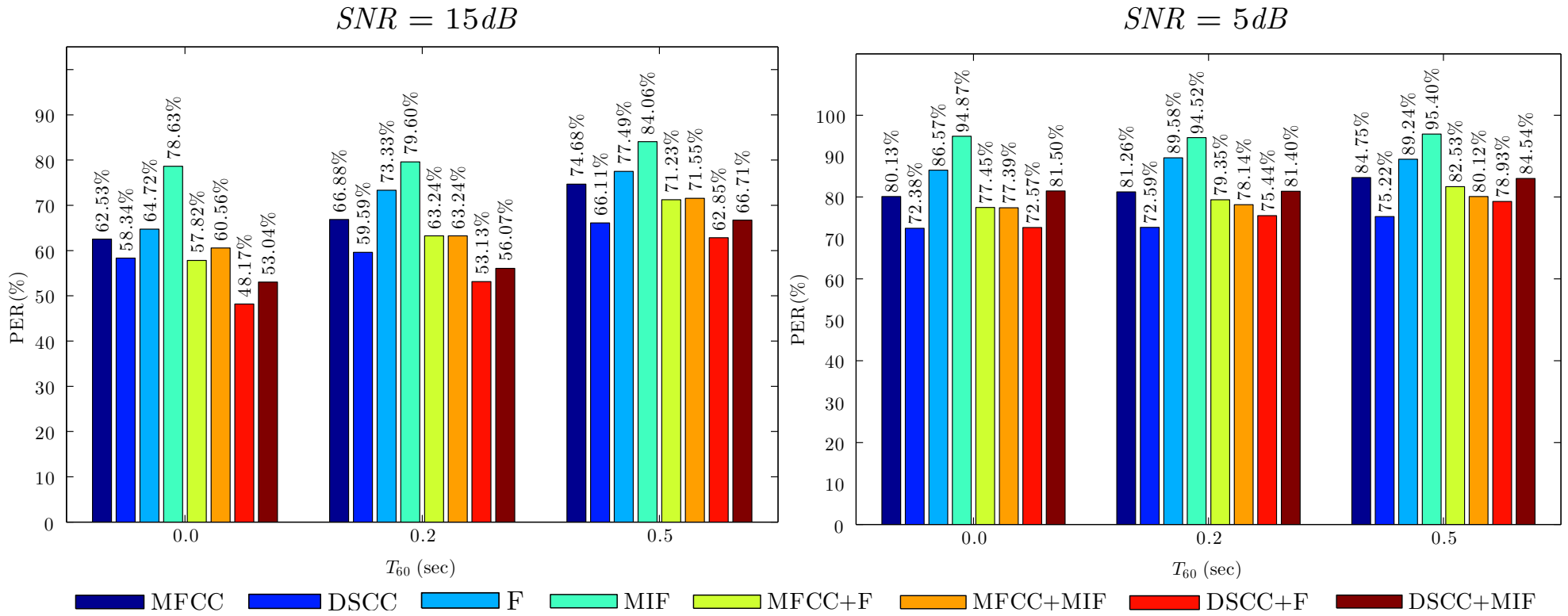


Χρόνος αντήχησης (T_{60})

Χρόνος που απαιτείται για την εξασθένηση ενός σήματος $60dB$ χαμηλότερα από το αρχικό του επίπεδο ακουστικής πίεσης

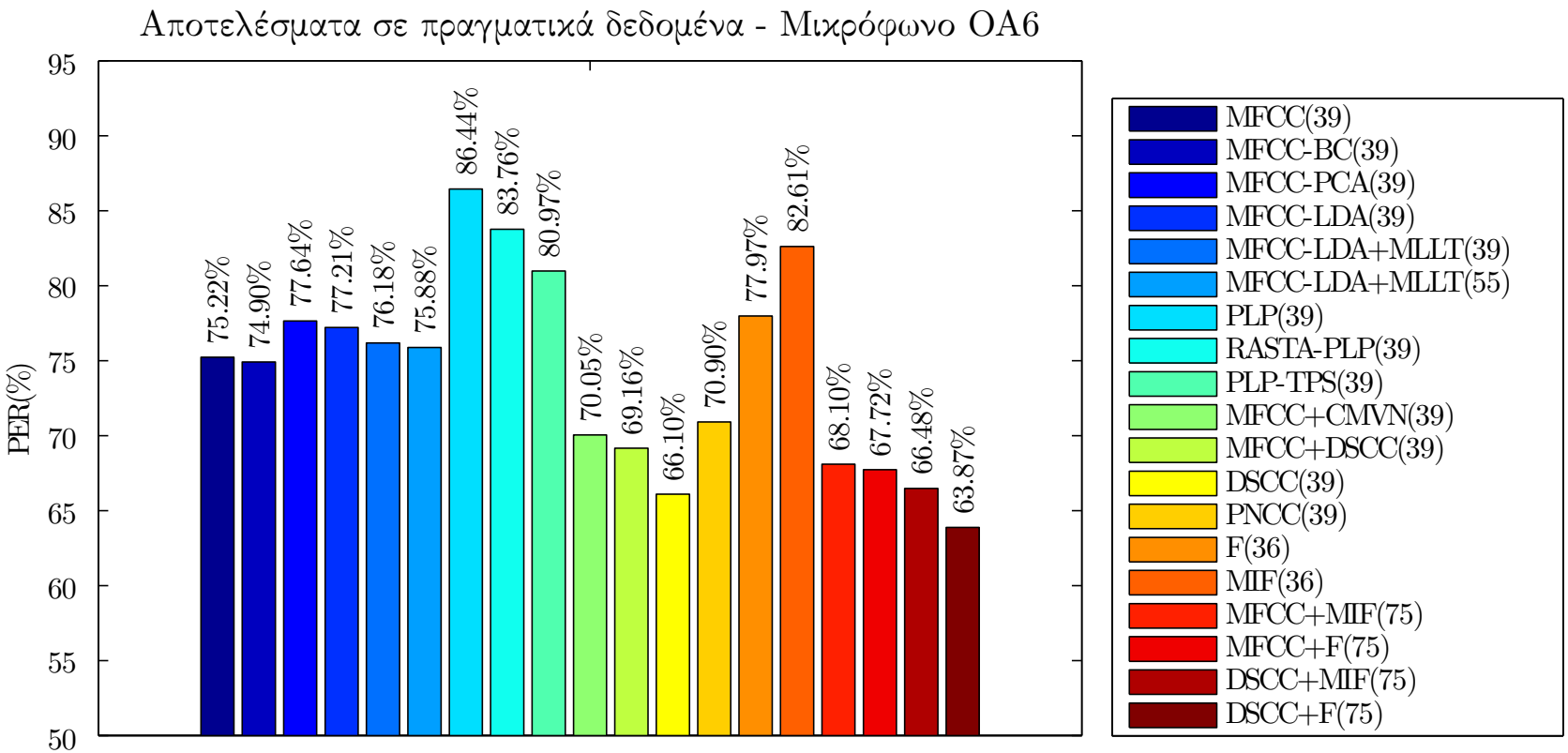
Αποτελέσματα σε Συνθετικά Δεδομένα - II

Ακίνητος ομιλητής



Περιεχόμενα

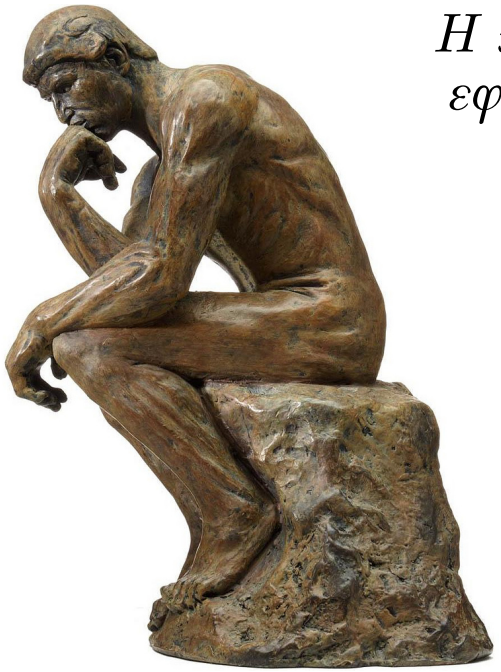
Συγκεντρωτικά Αποτελέσματα



Κατευθύνσεις για Μελλοντική Έρευνα

- Νέες μέθοδοι εξαγωγής δυναμικών χαρακτηριστικών. Στην αναγνώριση από απόσταση, ένα σύνολο που αντικατοπτρίζει αξιόπιστα τη δυναμική ίσως είναι πιο χρήσιμο από την όποια στατική πληροφορία.
- Επιπλέον συνδυασμοί συμπληρωματικών συνόλων χαρακτηριστικών.
- Χρήση συστοιχιών μικροφώνων με εξαγωγή διαφορετικών χαρακτηριστικών από κάθε μικρόφωνο, αναλόγως της αναμενόμενης επίδρασης του θορύβου στην εκάστοτε θέση του δωματίου.
- Εναλλακτικά κριτήρια συνδυασμού των SEO και TEO.

Ευχαριστώ



Η πιο ταπεινή και επικερδής εφεύρεση από όλες τις άλλες ήταν αυτή της ομιλίας.

Thomas Hobbes, 1651

Ερωτήσεις και Συζήτηση