Speech Segmentation based on the Bayesian Information Criterion

Constantine Kotropoulos

Collaborators: Dr. George Almpanidis, Dr. Margarita Kotti, Emmanouil Benetos, Vassiliki Moschou, Nikoletta Bassiou

e-mail: *costas@aiia.csd.auth.gr*

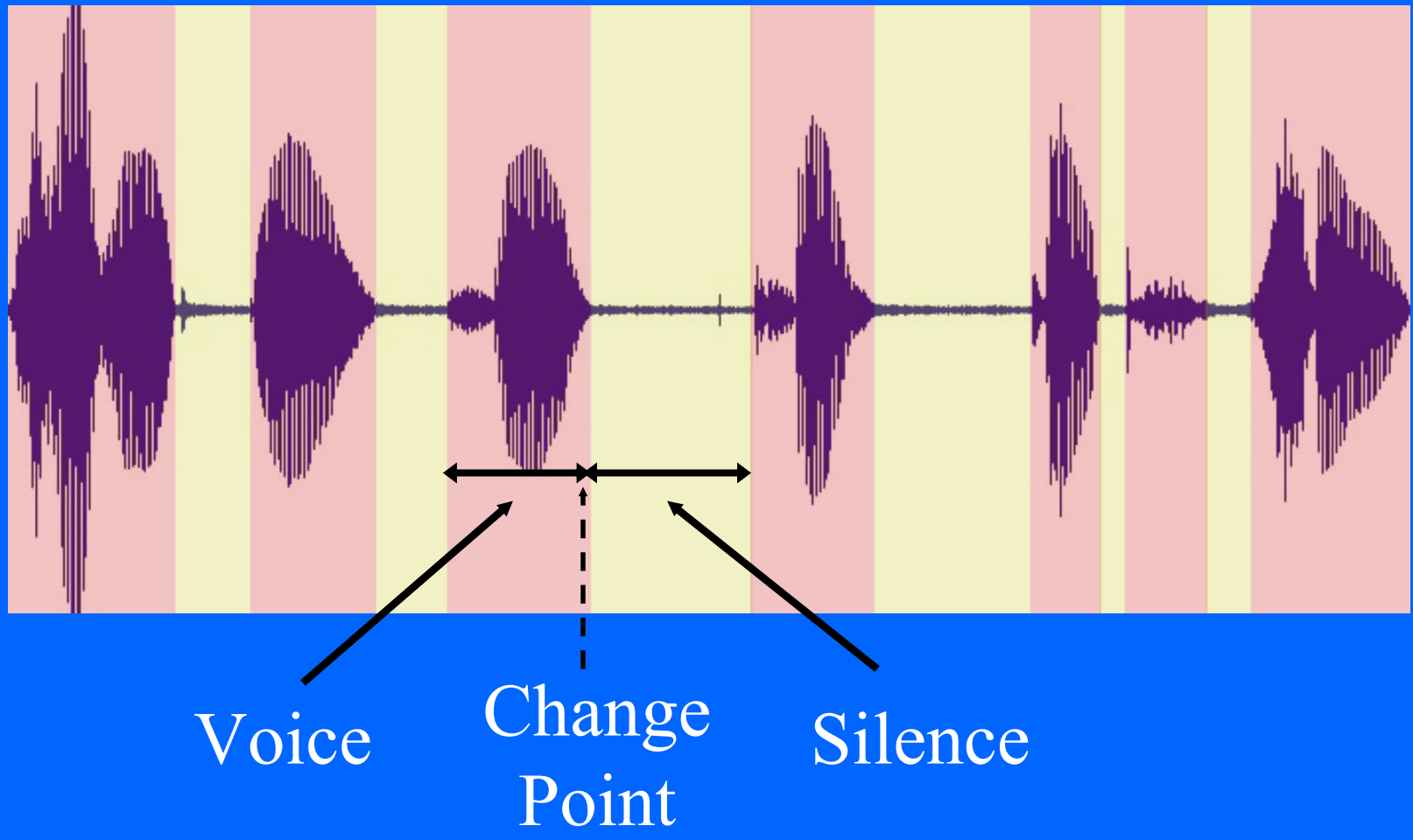
Informatics and Telematics Institute

Centre for Research & Technology Hellas

May 12th, 2010, Thessaloniki

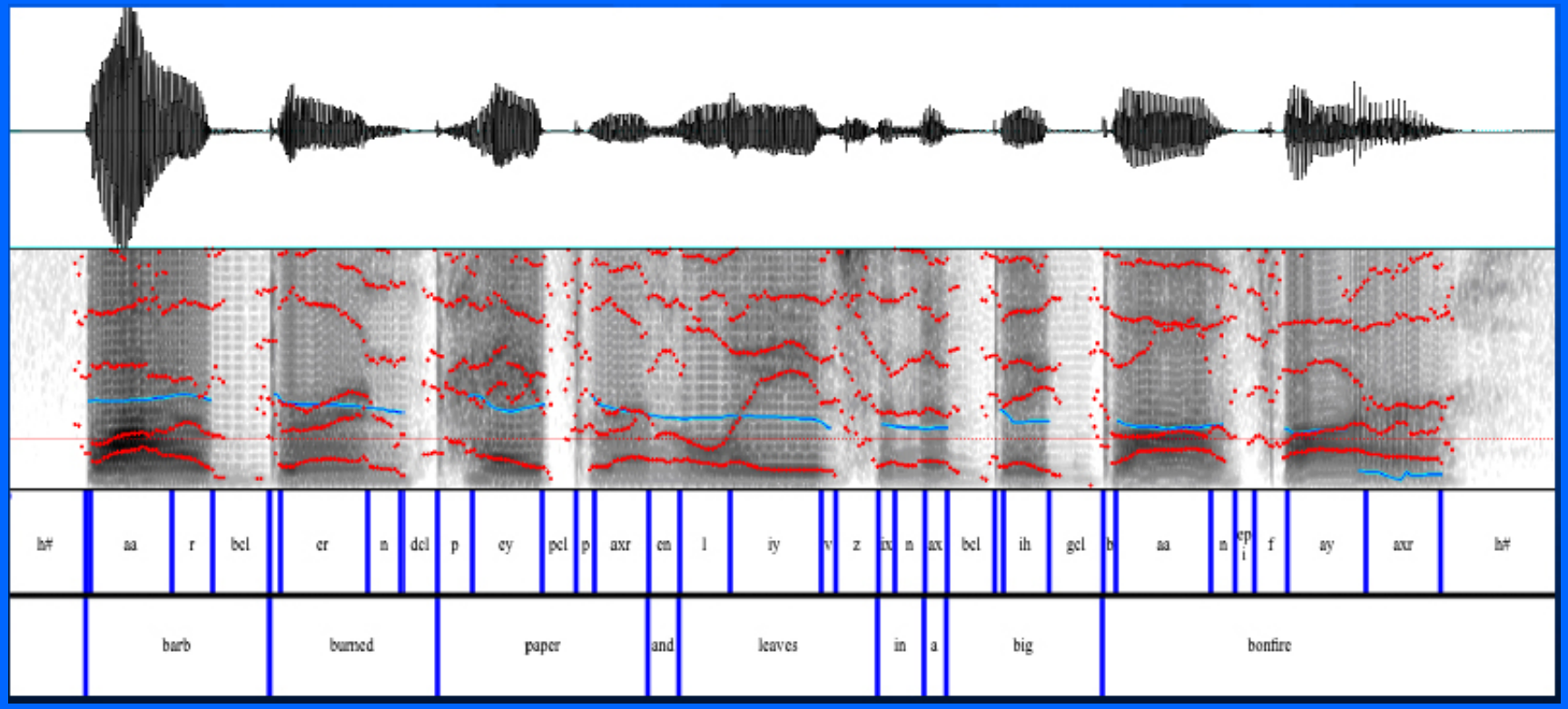
Dept. of Informatics, Aristotle Univ. of Thessaloniki, Greece

Voice Activity Detection (VAD)



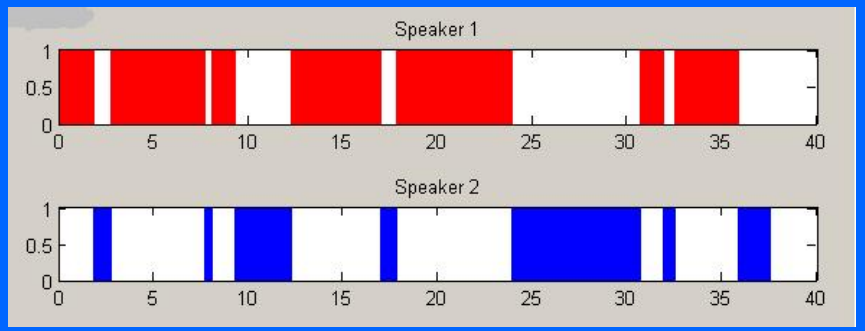
Phonemic Segmentation

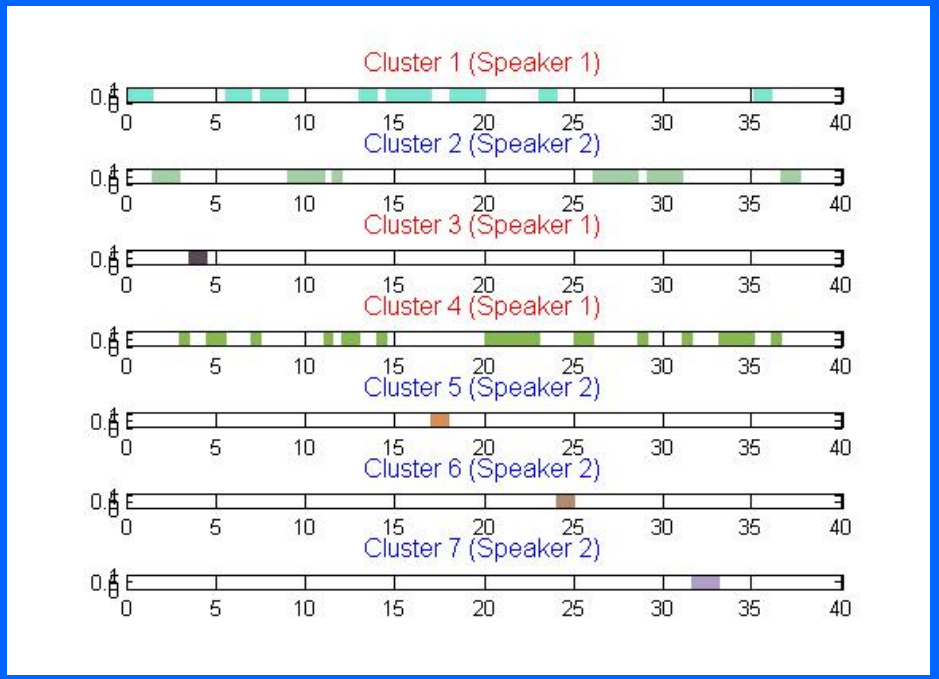
Spectrogram, formants,
Phonemic transcription (PRAAT, TIMIT)



Speaker Clustering

- Each cluster contains speech from only one speaker;
- Speech from the same speaker is gathered into the same cluster.





Assumptions

- Speakers do not speak simultaneously
- No real-time constraints (allow multi-pass algorithms)
- Mutually independent and identically distributed data
- Short- time processing
- Noise: additive, zero-mean and uncorrelated with the clean speech (in orthogonal transforms DFT, DCT, KLT)

Speech Segmentation Algorithms

Energy based

- use energy thresholds and heuristics to detect silence periods in the audio stream
- easy to implement, fast, computationally efficient
- online, real-time processing
- not robust under noise, misclassify fricatives, clicking, artifacts
- no direct connection between boundaries and acoustic changes

Model based

- assign statistical models to each acoustic class
- explore speech and noise statistics and use a decision rule usually derived from LRT
- slow, complex, computationally intensive, offline applications
- good precision even in low SNRs

Metric based

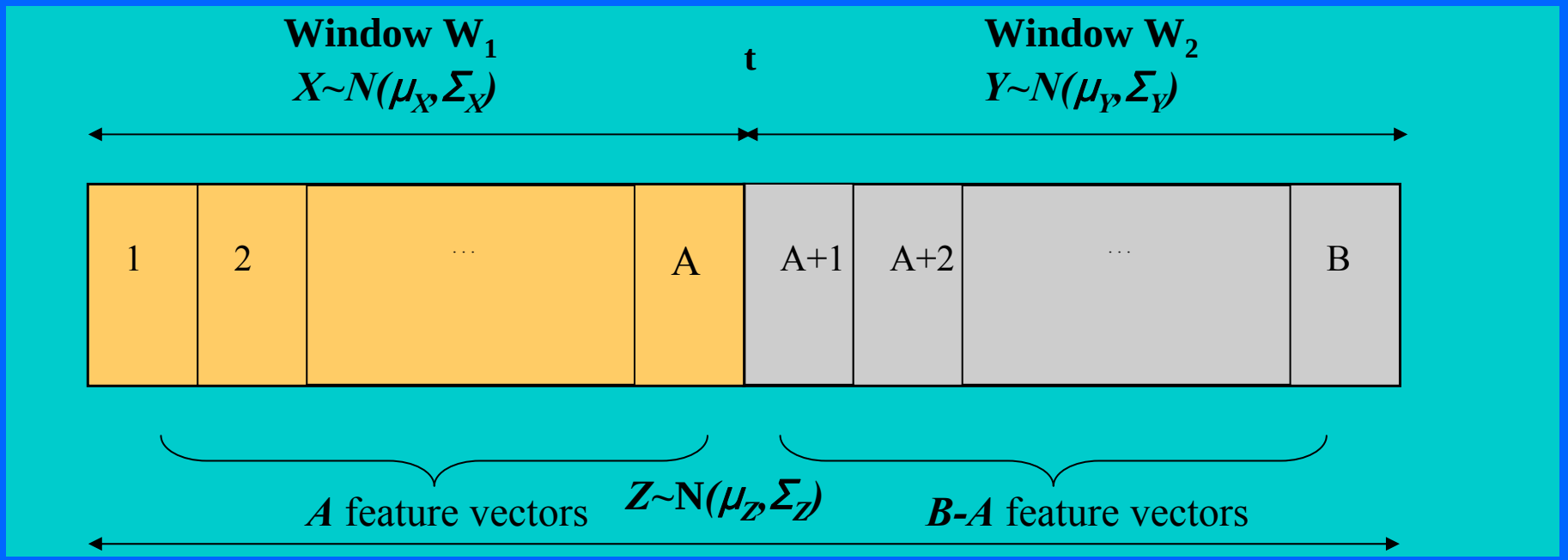
- measure the dissimilarity value between two consecutive parts of the parameterized signal
- slow, computationally intensive, offline applications
- good precision even in low SNRs

Hybrid

Hypothesis Testing Problem

H_0: $(x_1,x_2,\ldots,x_B) \sim N(\mu_Z, \Sigma_Z)$

H_1: $(x_1,x_2,\ldots,x_A) \sim N(\mu_X, \Sigma_X)$ & $(x_{A+1},x_{A+2},\ldots,x_B) \sim N(\mu_Y, \Sigma_Y)$



Log-likelihoods

- **Log-likelihood under the null hypothesis**

Let $\mathbf{z}_i$ be i.i.d. d-dimensional Gaussian random vectors.

$$\ell(Z; \underbrace{\boldsymbol{\mu}_Z, \boldsymbol{\Sigma}_Z}_{\boldsymbol{\theta}_Z}) = \sum_{i=1}^{A} \ln p(\mathbf{z}_i|\boldsymbol{\theta}_Z) + \sum_{i=A+1}^{B} \ln p(\mathbf{z}_i|\boldsymbol{\theta}_Z)$$

$$= -B\frac{d}{2}\ln(2\pi) - \frac{B}{2}\ln|\boldsymbol{\Sigma}_Z| - \frac{1}{2}\sum_{i=1}^{B}(\mathbf{z}_i - \boldsymbol{\mu}_Z)^T\boldsymbol{\Sigma}_Z^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_Z)$$

- **Log-likelihood under the alternative hypothesis**

$$\ell(X; \underbrace{\boldsymbol{\mu}_X, \boldsymbol{\Sigma}_X}_{\boldsymbol{\theta}_X}) + \ell(Y; \underbrace{\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y}_{\boldsymbol{\theta}_Y}) = \sum_{i=1}^{A} \ln p(\mathbf{z}_i|\boldsymbol{\theta}_X) + \sum_{i=A+1}^{B} \ln p(\mathbf{z}_i|\boldsymbol{\theta}_Y)$$

$$= -A\frac{d}{2}\ln(2\pi) - \frac{A}{2}\ln|\boldsymbol{\Sigma}_X| - \frac{1}{2}\sum_{i=1}^{A}(\mathbf{z}_i - \boldsymbol{\mu}_X)^T\boldsymbol{\Sigma}_X^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_X)$$

$$-(B-A)\frac{d}{2}\ln(2\pi) - \frac{B-A}{2}\ln|\boldsymbol{\Sigma}_Y| - \frac{1}{2}\sum_{i=A+1}^{B}(\mathbf{z}_i - \boldsymbol{\mu}_Y)^T\boldsymbol{\Sigma}_Y^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_Y)$$

Distance Measures (1)

- **Likelihood-ratio test**

$$
\begin{aligned}
R &= \frac{L(z;\boldsymbol{\theta}_Z)}{L(z;\boldsymbol{\theta}_X)L(z;\boldsymbol{\theta}_Y)} \\
d = -\ln R &= \ell(z;\boldsymbol{\theta}_X) + \ell(z;\boldsymbol{\theta}_Y) - \ell(z;\boldsymbol{\theta}_Z)
\end{aligned}
$$

Distance Measures (2)

- Kullback-Leibler (KL) divergence = relative entropy

$$KL(X,Y) = \int \cdots \int_{\mathcal{D}\{p(\mathbf{z};\boldsymbol{\theta}_X)\}} \ln \frac{p(\mathbf{z};\boldsymbol{\theta}_X)}{p(\mathbf{z};\boldsymbol{\theta}_Y)}\, p(\mathbf{z};\boldsymbol{\theta}_X)\, d\mathbf{z}$$

- Symmetric KL distance

$$KL2(X,Y) = KL(X,Y) + KL(Y,X)$$

- For Gaussian random vectors

$$\begin{aligned} KL2(X,Y) \;=\; & \frac{1}{2}(\boldsymbol{\mu}_Y - \boldsymbol{\mu}_X)^T(\boldsymbol{\Sigma}_X^{-1} + \boldsymbol{\Sigma}_Y^{-1})(\boldsymbol{\mu}_Y - \boldsymbol{\mu}_X) \\ & + \frac{1}{2}\mathrm{tr}\left((\boldsymbol{\Sigma}_X^{1/2}\boldsymbol{\Sigma}_Y^{-1/2})(\boldsymbol{\Sigma}_X^{1/2}\boldsymbol{\Sigma}_Y^{-1/2})^T\right) \\ & + \frac{1}{2}\mathrm{tr}\left((\boldsymbol{\Sigma}_X^{-1/2}\boldsymbol{\Sigma}_Y^{1/2})(\boldsymbol{\Sigma}_X^{-1/2}\boldsymbol{\Sigma}_Y^{1/2})^T\right) - d \end{aligned}$$

Information Criteria (1)

$$\text{IC} = \text{-GOF} + \text{complexity penalty}$$

- Estimates of the KL distance
  - Akaike Information Criterion [AIC]; Variants: Consistent AIC [CAIC], Quasi AIC [QAIC], Takeuchi information criterion [TIC]
  - Holistic: Competing models are assessed simultaneously and the best model is selected by applying a single rule; No constraint that one model is the "true".
- Dimension consistent criteria:
  - Bayesian IC [BIC], Minimum Description Length (MDL), Hannan-Quinn criterion.
  - Finds the true model, provided that such a model exists and it is in the set of candidate models

AIC

$$AIC(\mathcal{M}) = -2 \ln L(X; \mathcal{M}) + P$$

where

- X are the sample data,
- $L(X; \mathcal{M})$ is the maximized likelihood function under $\mathcal{M}$
- P is the number of the model parameters.

Bayes Factor

- Similar to an LRT, but instead of maximizing the likelihood, we average the likelihood over the parameters

$$BF = \frac{P(X|\mathcal{M}_0)}{P(X|\mathcal{M}_1)} = \frac{\int_{\Theta_0} p(X|\boldsymbol{\theta}_0, \mathcal{M}_0)\, p(\boldsymbol{\theta}_0|\mathcal{M}_0)\, d\boldsymbol{\theta}_0}{\int_{\Theta_1} p(X|\boldsymbol{\theta}_1, \mathcal{M}_1)\, p(\boldsymbol{\theta}_1|\mathcal{M}_1)\, d\boldsymbol{\theta}_1}$$

- Best model: That with the highest posterior probability

$$\frac{P(\mathcal{M}_0|X)}{P(\mathcal{M}_1|X)} = \underbrace{\frac{P(X|\mathcal{M}_0)}{P(X|\mathcal{M}_1)}}_{BF} \times \underbrace{\frac{P(\mathcal{M}_0)}{P(\mathcal{M}_1)}}_{\text{Prior odds}}$$

- Related to BIC

BIC

- Penalized ML technique (Schwarz 1978). Selects the true model asymptotically with probability 1. For *n* observations:

$$BIC(\mathcal{M}) = -2 \ln L(X; \mathcal{M}) + P \ln n$$

- BIC can be derived as an approximation of the BF

$$\text{Let } \Delta BIC = BIC(\mathcal{M}_1) - BIC(\mathcal{M}_0) \text{ then}$$

$$BIC(\mathcal{M}_1) = BIC(\mathcal{M}_0) - \Delta BIC \approx 2 \ln BF$$

- The approximation is close, when the prior over the parameters is the **unit information prior**, i.e. a multivariate normal prior with mean at the MLE and variance equal to the expected information matrix per observation.

BIC for Multivariate Gaussian Observations

$$\begin{aligned}\mathrm{BIC}(Z) &= \ell(z;\boldsymbol{\theta}_Z) - \frac{\lambda}{2}\left(d + \frac{d(d+1)}{2}\right)\ln B \\ &= -B\frac{d}{2}\ln(2\pi) - \frac{B}{2}\ln|\boldsymbol{\Sigma}_Z| - \frac{1}{2}\sum_{i=1}^{B}(\mathbf{z}_i - \boldsymbol{\mu}_Z)^T\boldsymbol{\Sigma}_Z^{-1}(\mathbf{z}_i - \boldsymbol{\mu}_Z) \\ &\quad - \frac{\lambda}{2}\underbrace{\left(d + \frac{d(d+1)}{2}\right)}_{P=\text{model parameters}}\ln B\end{aligned}$$

- ***When the covariance matrices are estimated by sample dispersion matrices***

$$\sum_{i=1}^{B}(\boldsymbol{z}_i - \boldsymbol{\mu}_Z)^T\boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{z}_i - \boldsymbol{\mu}_Z) = \mathrm{tr}\{\boldsymbol{\Sigma}_Z^{-1}\sum_{i=1}^{N_Z}(\boldsymbol{z}_i - \boldsymbol{\mu}_Z)(\boldsymbol{z}_i - \boldsymbol{\mu}_Z)^T\} = d\,B$$

- ***BIC is simplified to:***

$$\mathrm{BIC}(Z) = -B\frac{d}{2}\ln(2\pi) - \frac{B}{2}\ln|\boldsymbol{\Sigma}_Z| - \frac{B}{2}\,d - \frac{\lambda}{2}\,P\,\ln B$$

ΔBIC for Multivariate Gaussian Observations

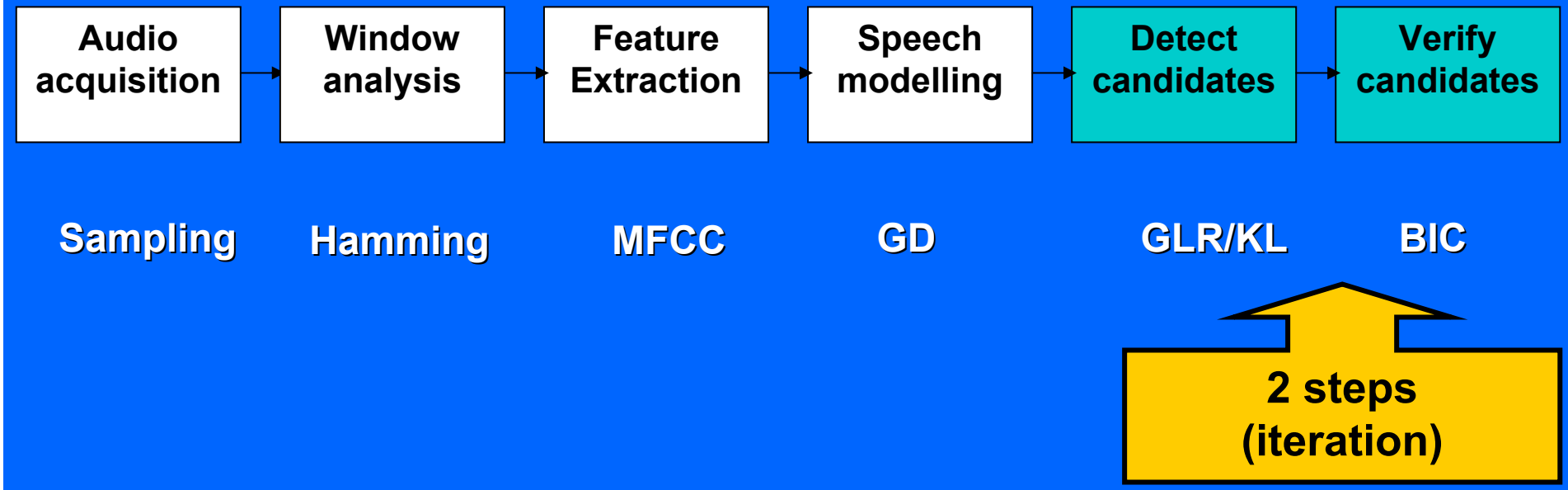
- *The BIC varies between the two models (i.e. one Gaussian vs. two different Gaussians) by*

$$\mathrm{BIC}(X,Y) \quad - \quad \mathrm{BIC}(Z) = \frac{B}{2}\ln|\mathbf{\Sigma}_Z| - \frac{A}{2}\ln|\mathbf{\Sigma}_X| - \frac{B-A}{2}\ln|\mathbf{\Sigma}_Y| - \frac{\lambda}{2}P\ \ln B$$

- **Positive ΔBIC values indicate that model transition from Z to (X,Y)**

DISTBIC for Multivariate Gaussian Observations

Combination of metric-based segmentation with the KL distance and the BIC model-based segmentation (Delacourt & Wellekens 2000).



ANIMATED EXAMPLE

DISTBIC

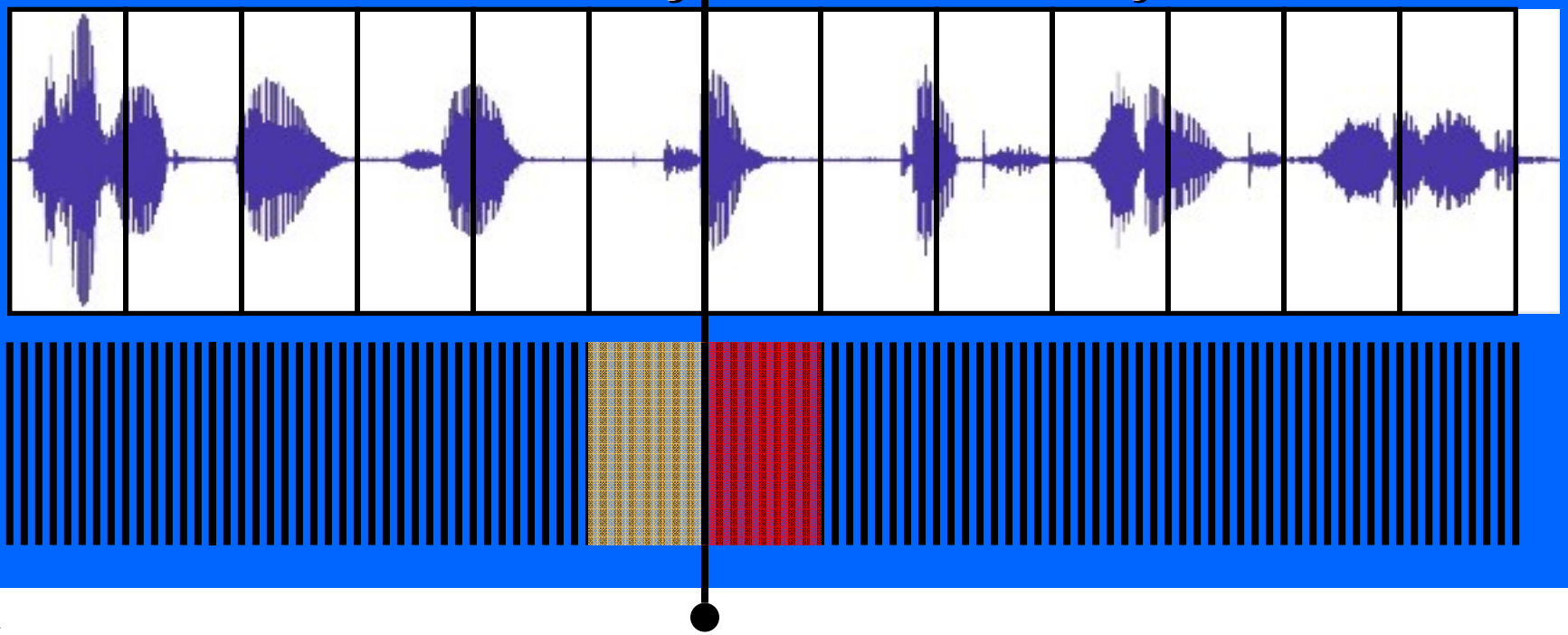
- Feature extraction (MFCC, DCT, DFT, …)

sequence of acoustic vectors

d-dimensional vectors (multivariate Gaussians)

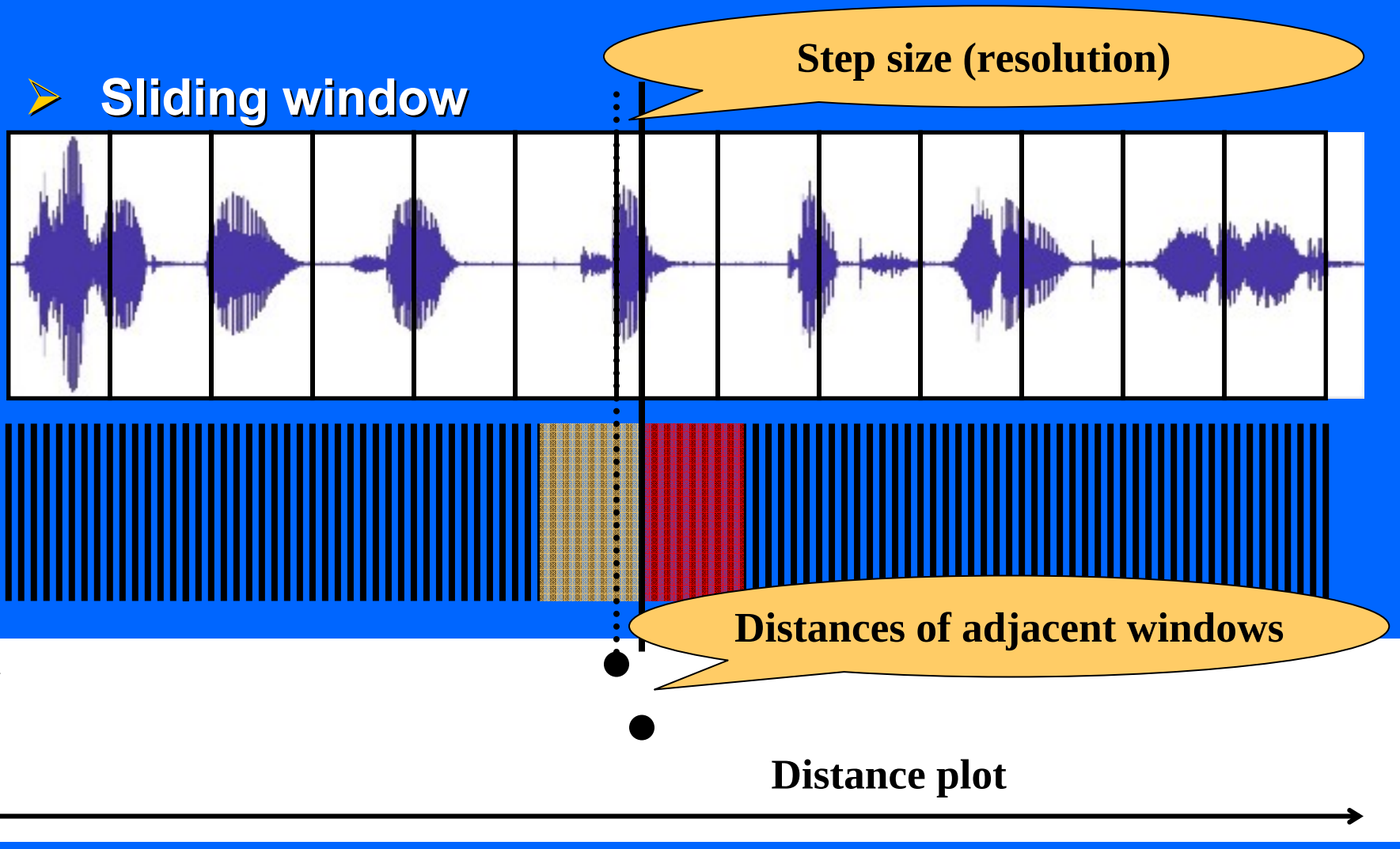
DISTBIC

- Calculate dissimilarity distances of adjacent windows



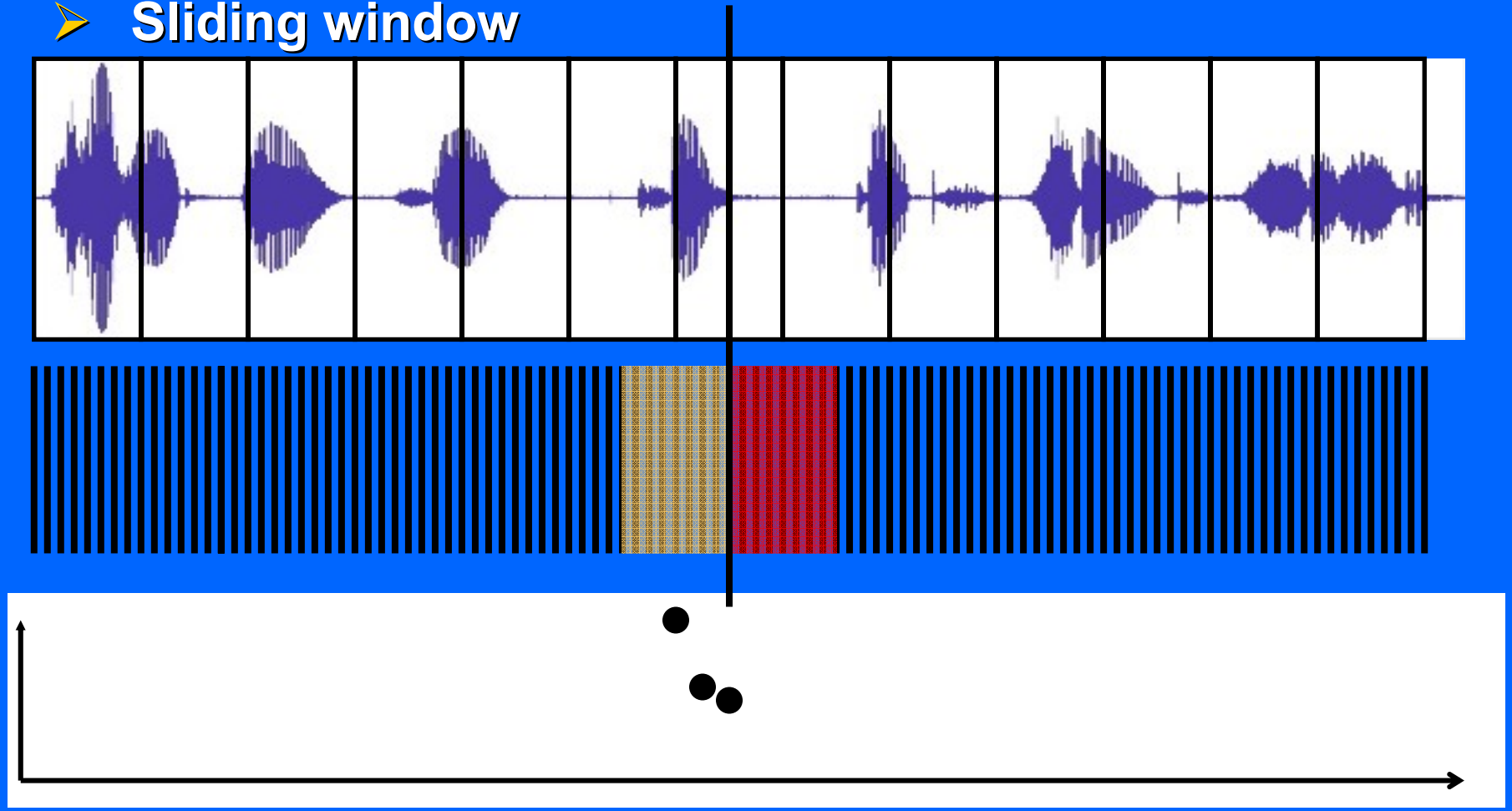
DISTBIC

- Sliding window



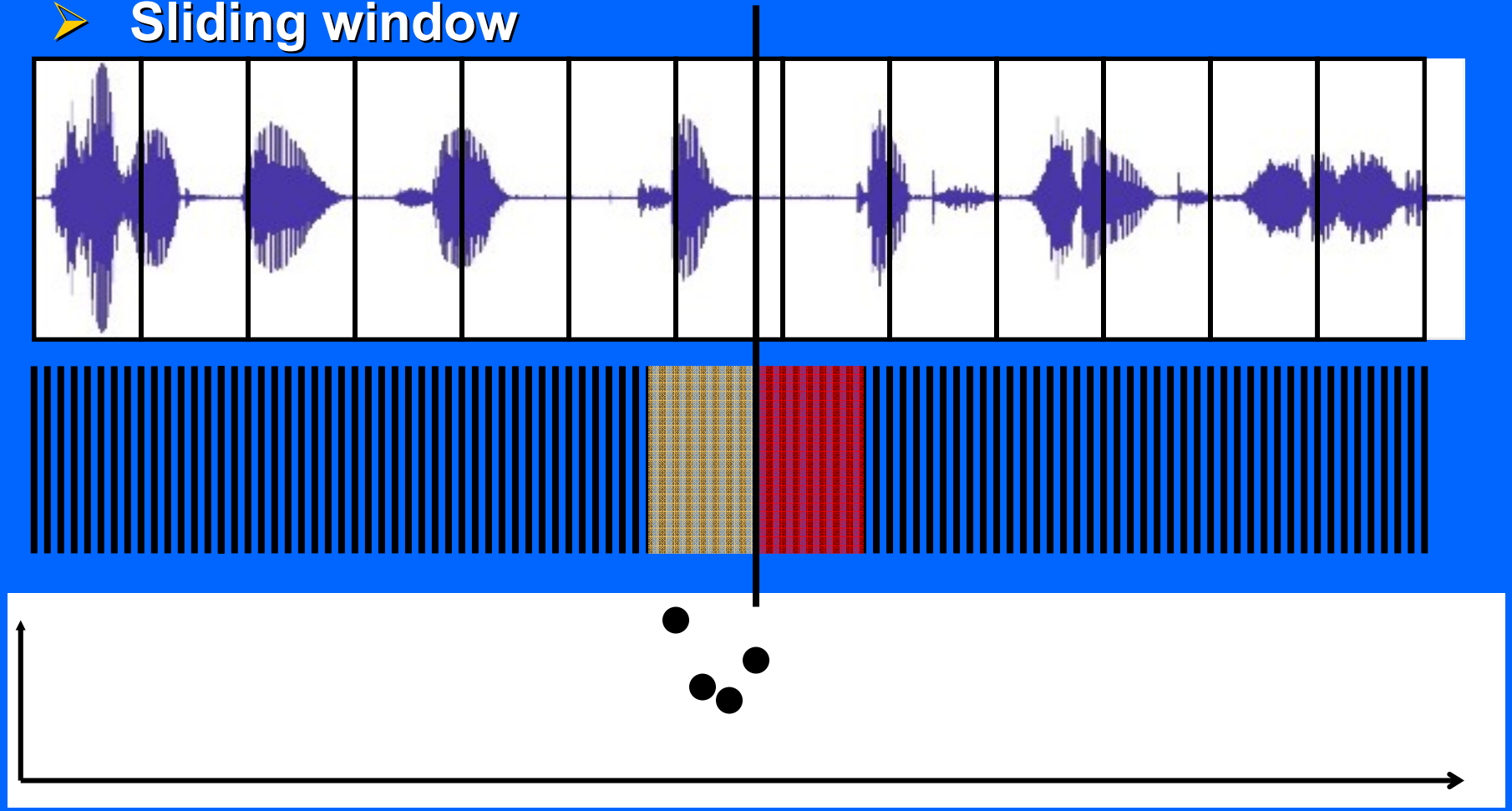
DISTBIC

- Sliding window



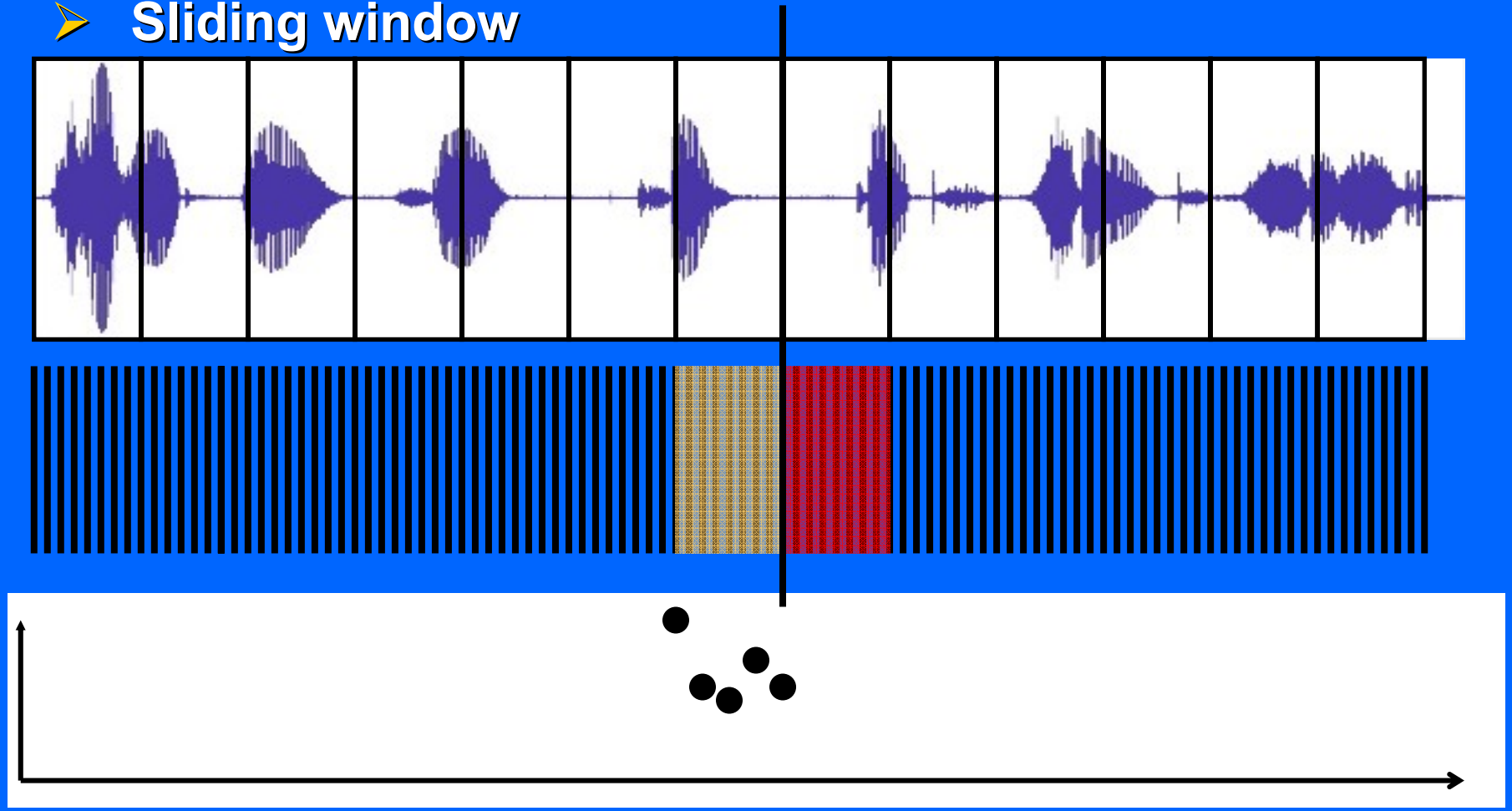
DISTBIC

- Sliding window



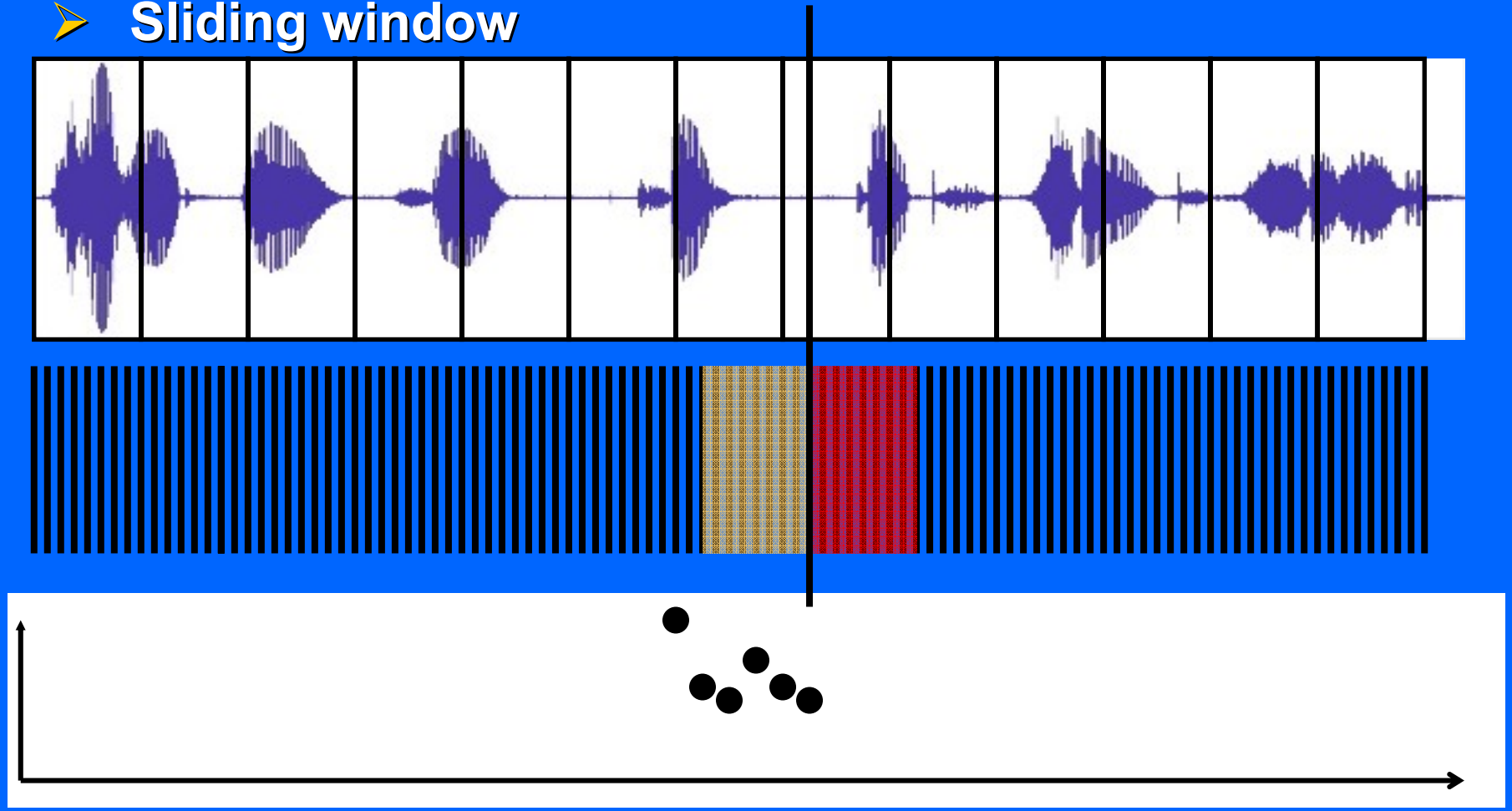
DISTBIC

- Sliding window



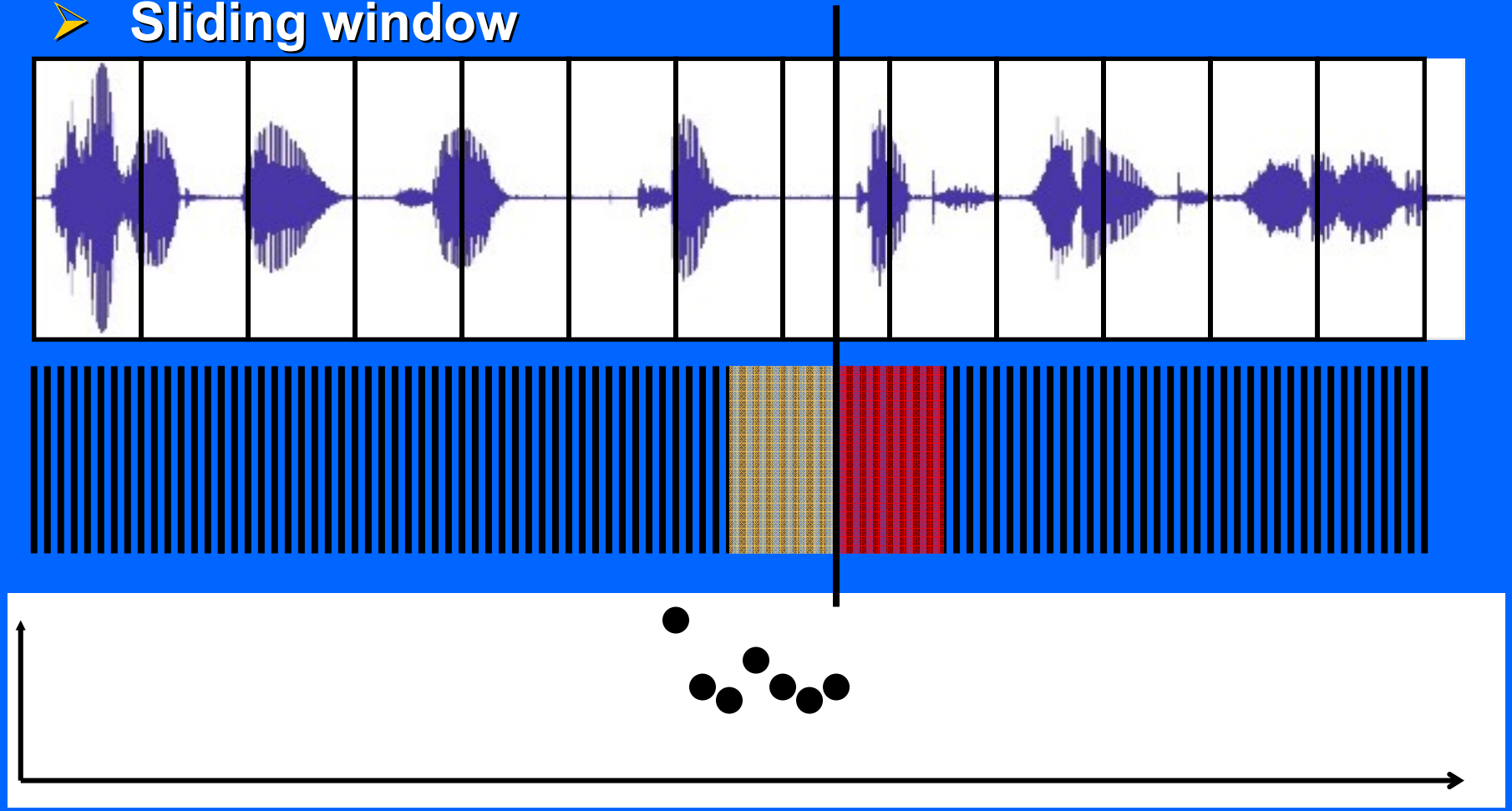
DISTBIC

- Sliding window



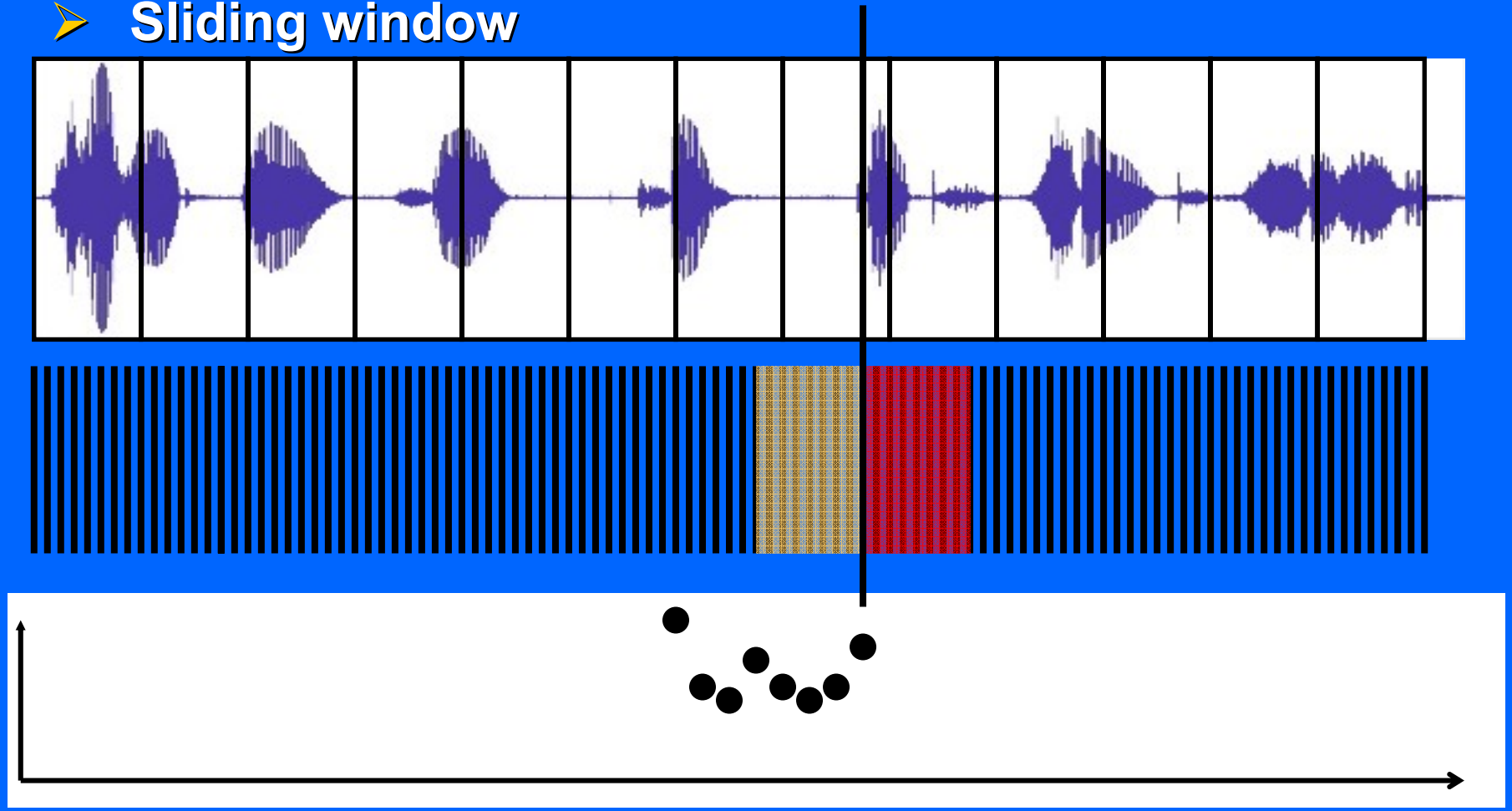
DISTBIC

- Sliding window



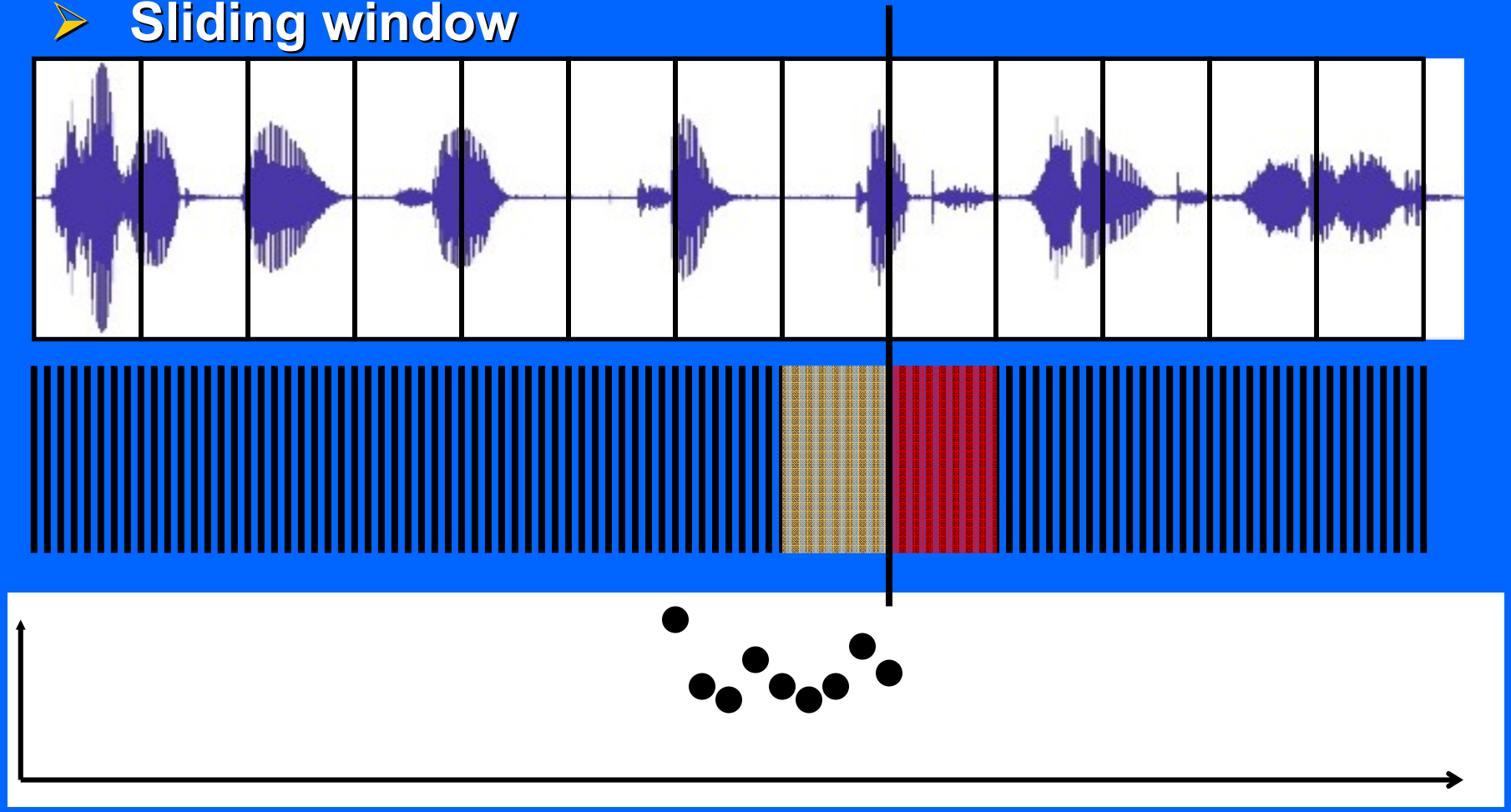
DISTBIC

- Sliding window



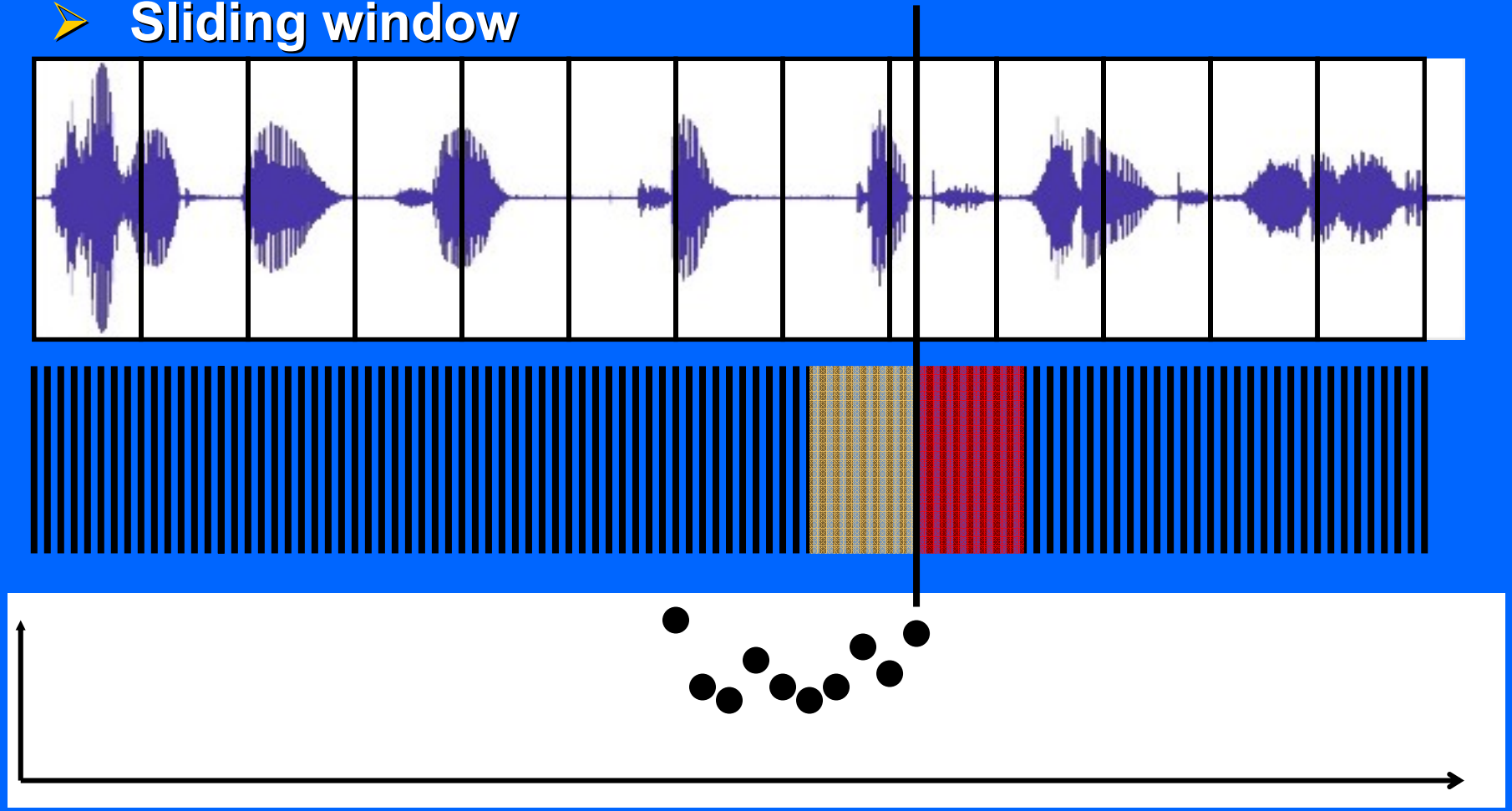
DISTBIC

- Sliding window



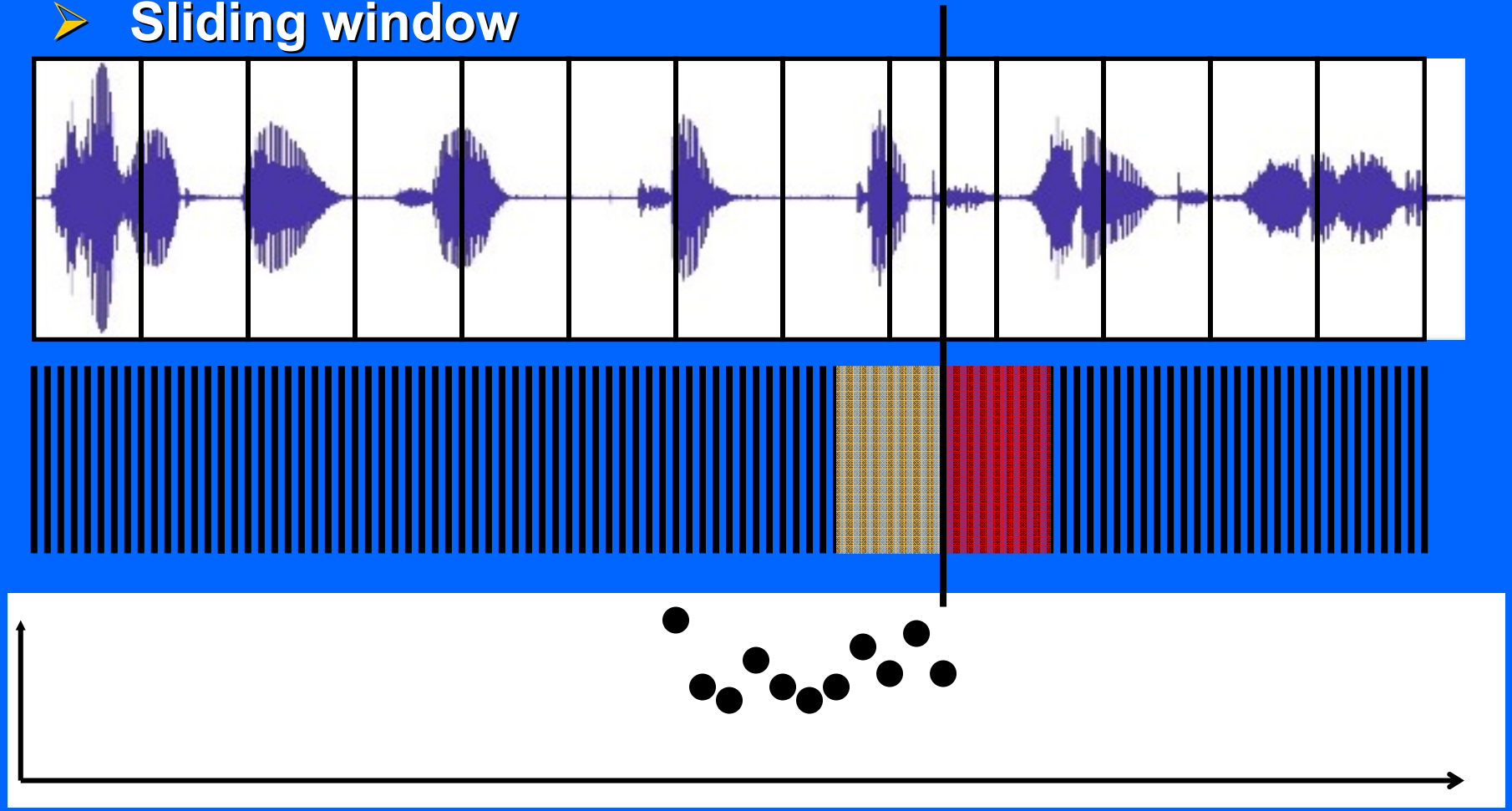
DISTBIC

- Sliding window



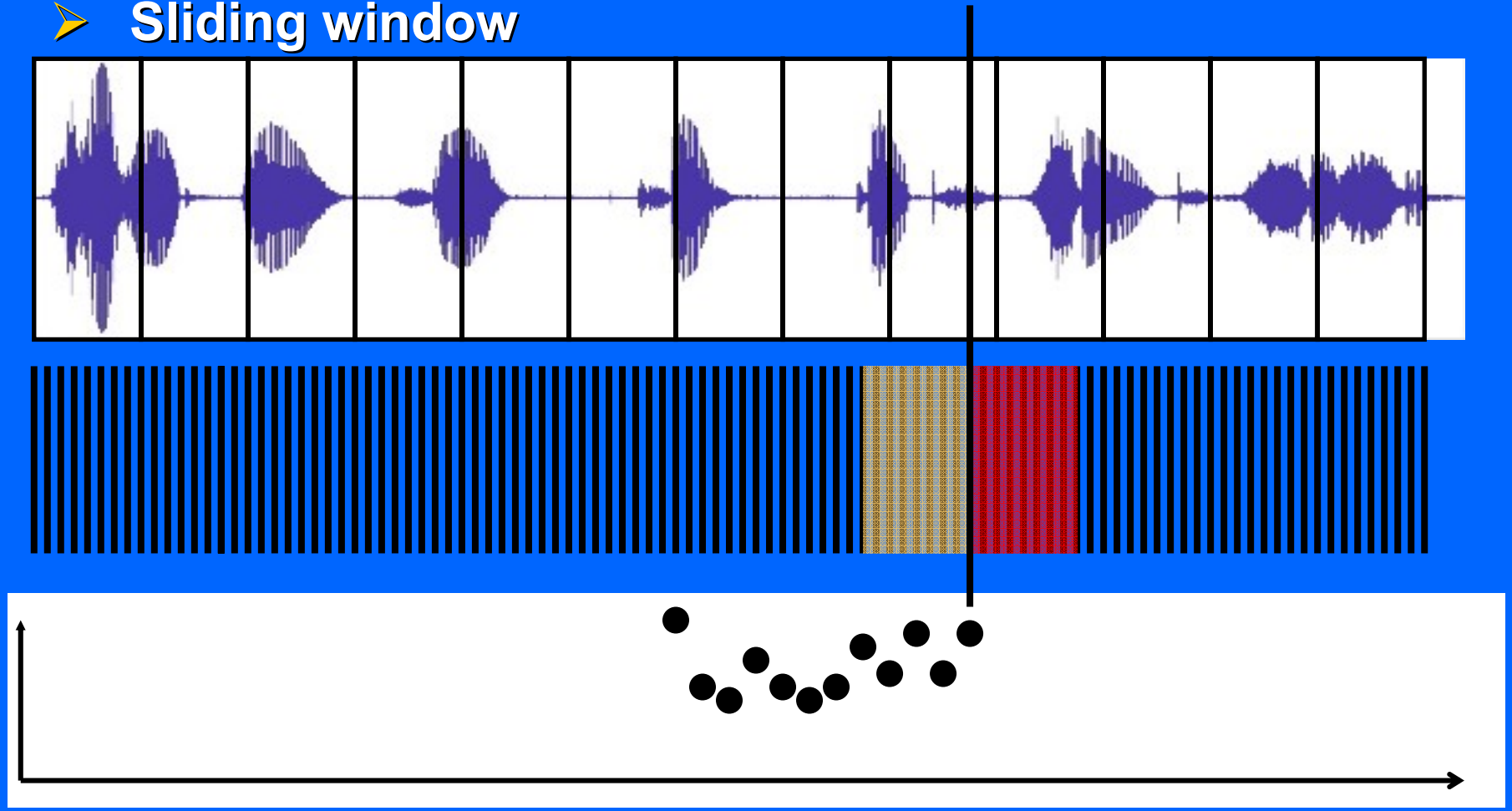
DISTBIC

- Sliding window



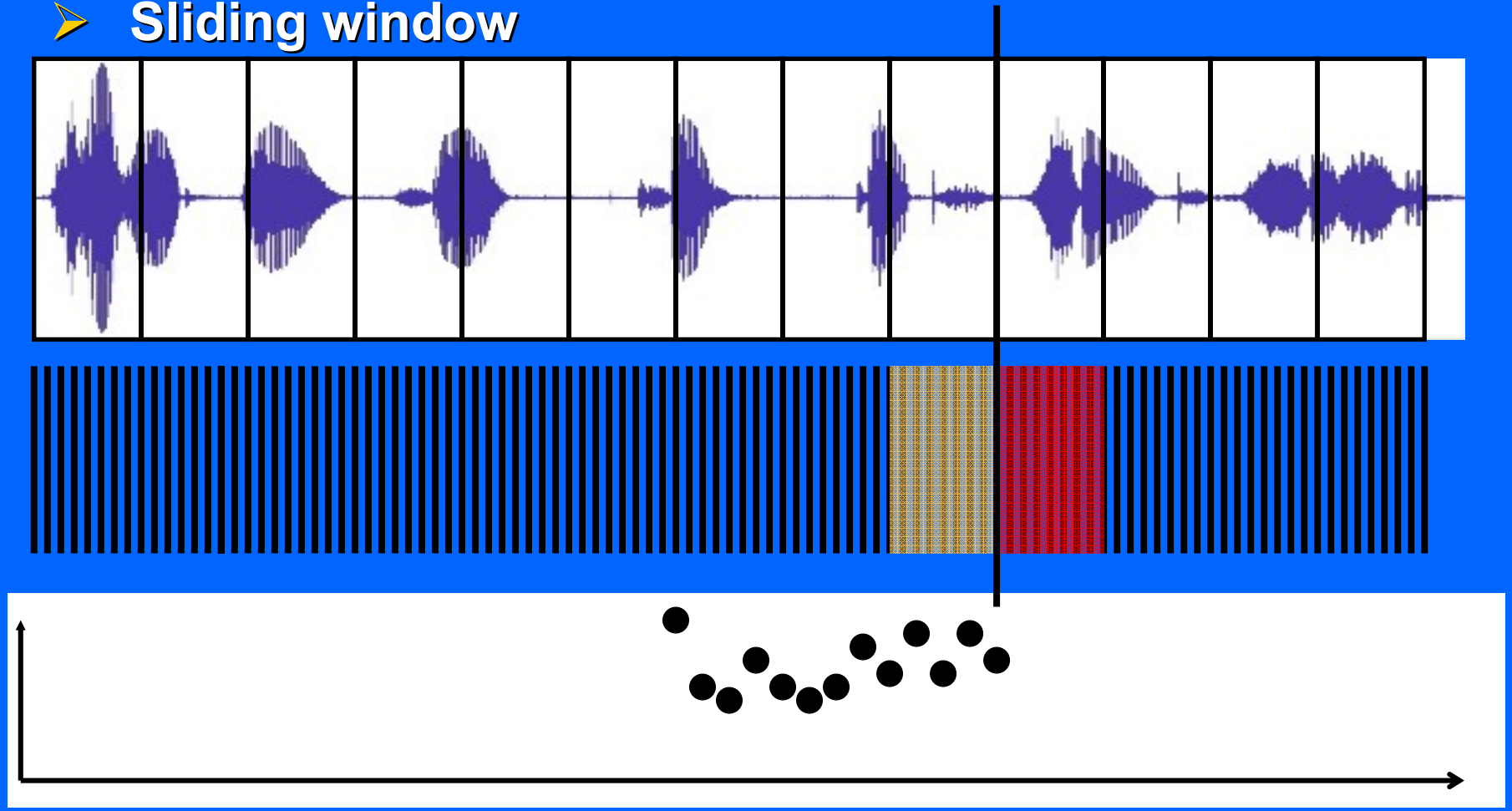
DISTBIC

- Sliding window



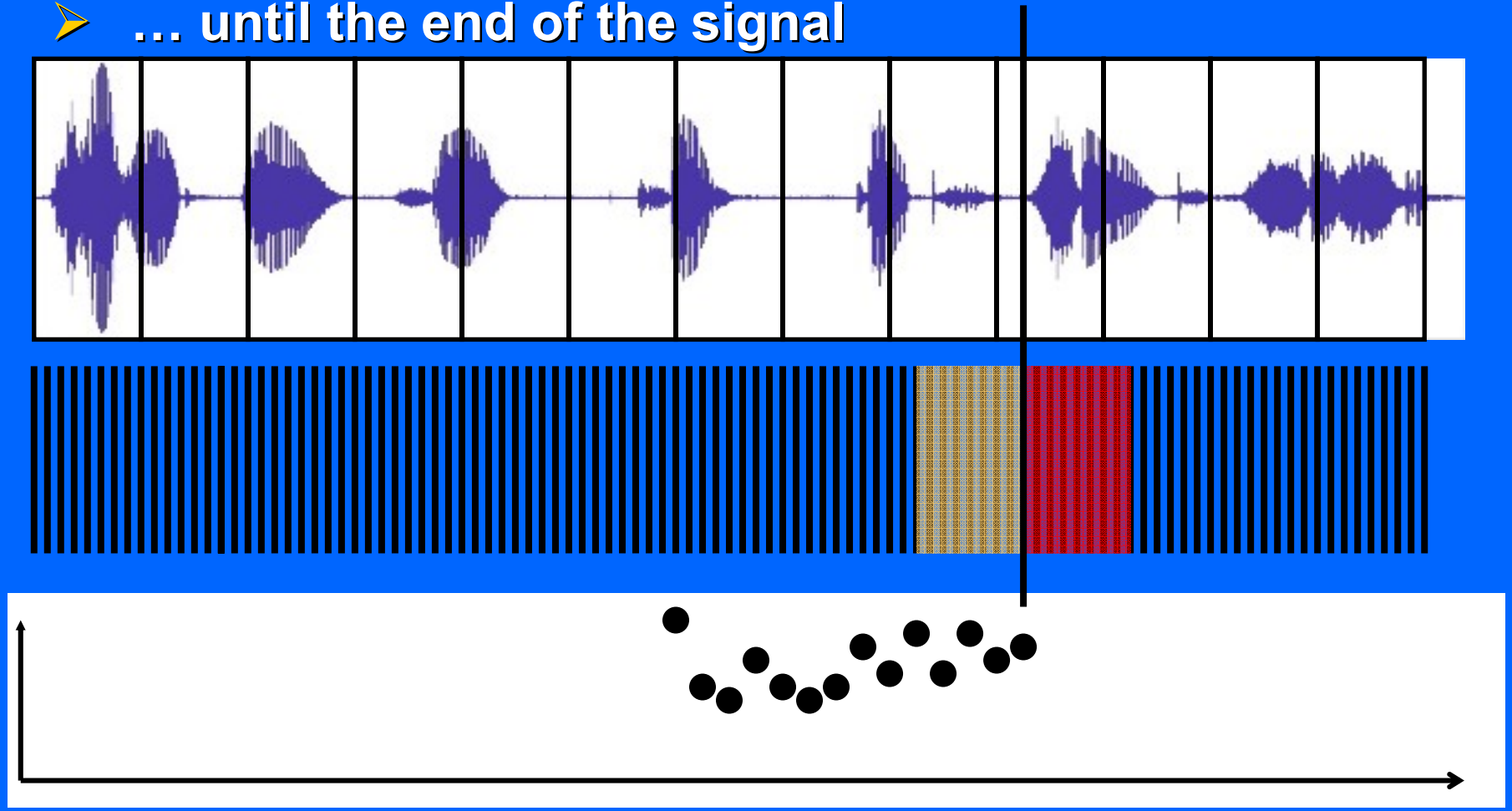
DISTBIC

- Sliding window



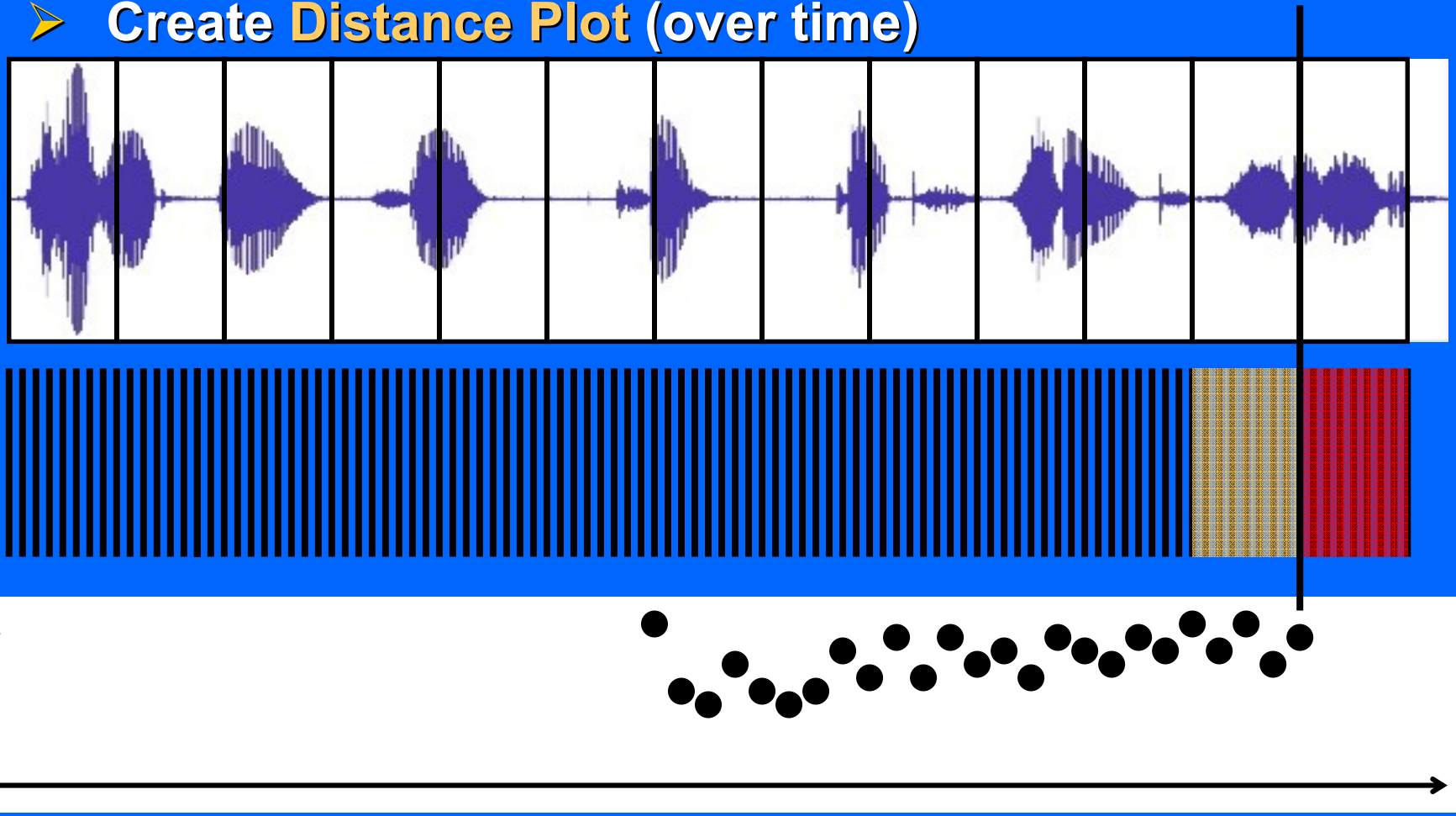
DISTBIC

- … until the end of the signal



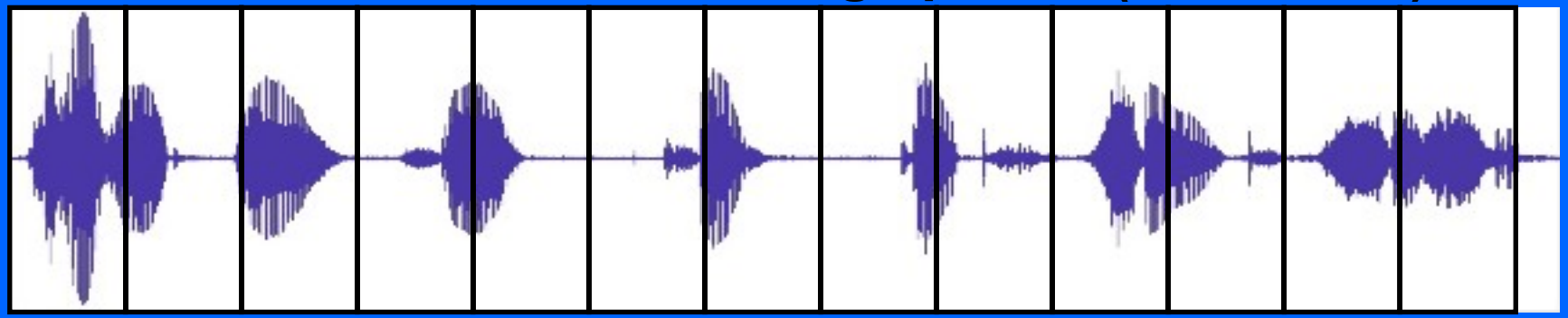
DISTBIC (step1: distance plot)

- Create Distance Plot (over time)



DISTBIC (smoothing)

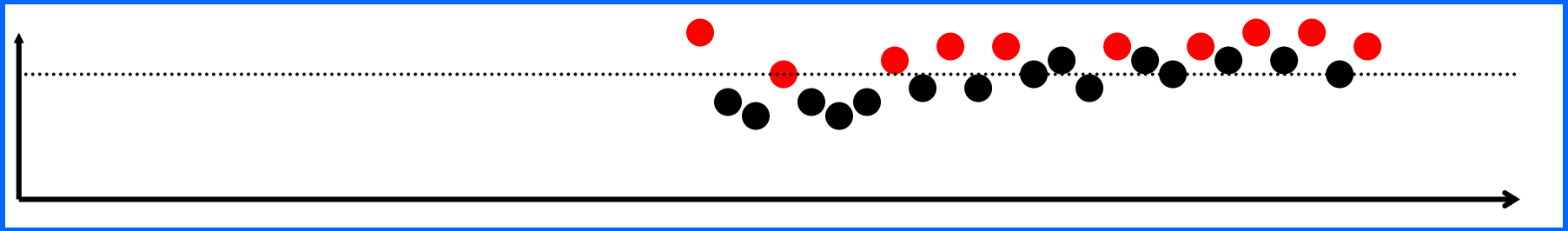
- Determine candidate change points (Heuristics)



Heuristics

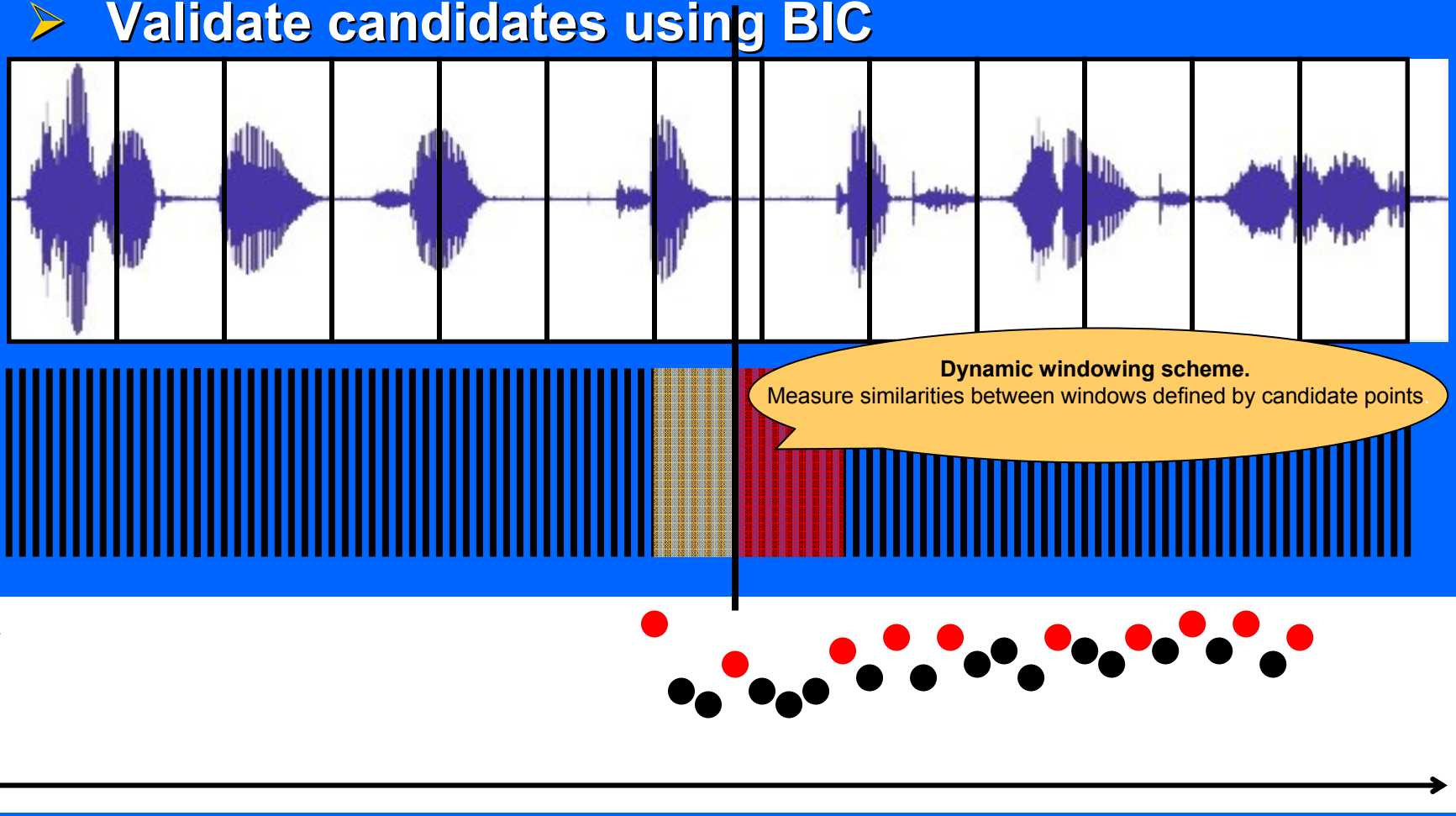
- select max values (vertical threshold)

- discard peaks too close to each other (horizontal threshold)



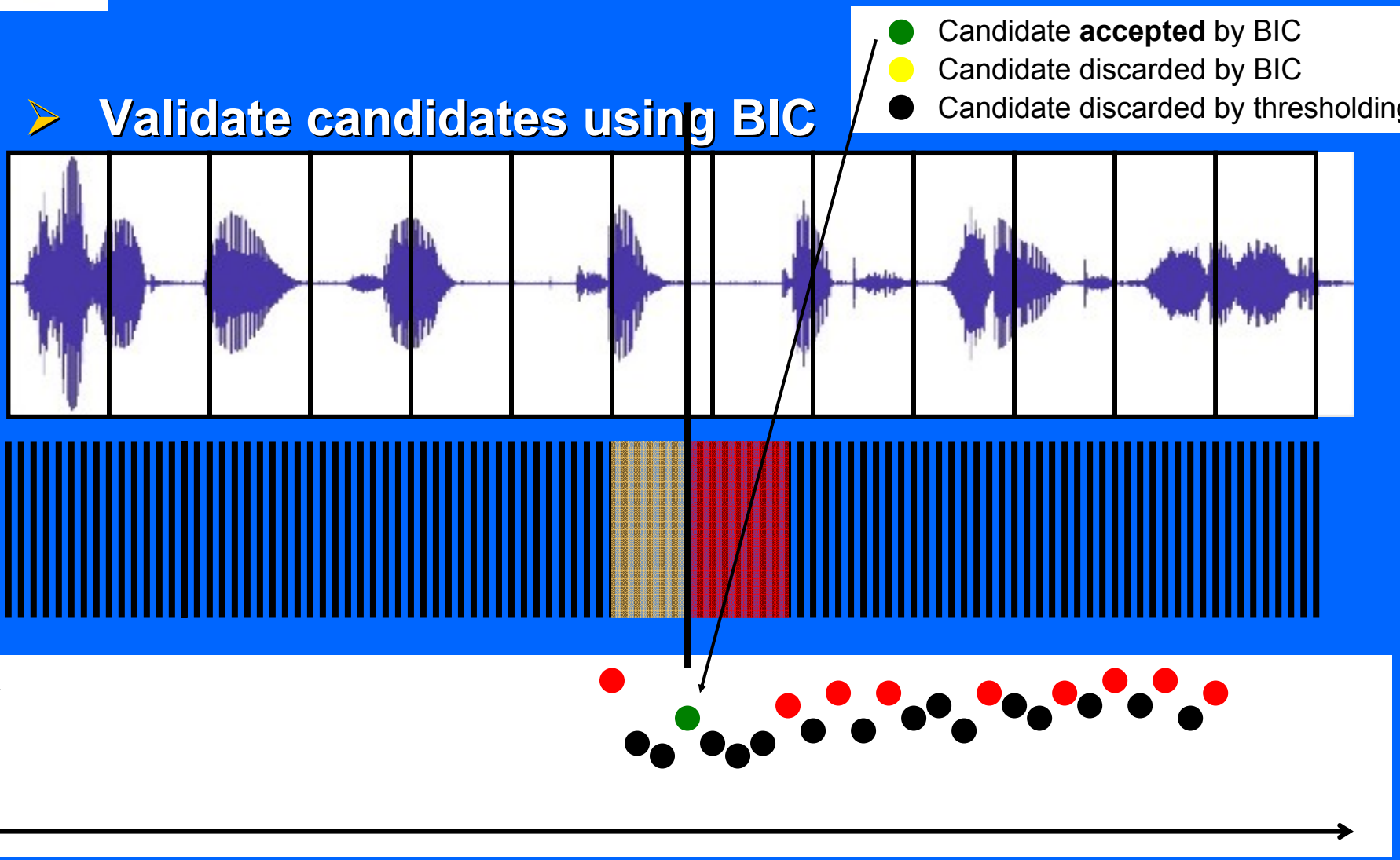
DISTBIC (step2: validation)

- Validate candidates using BIC



DISTBIC (validation)

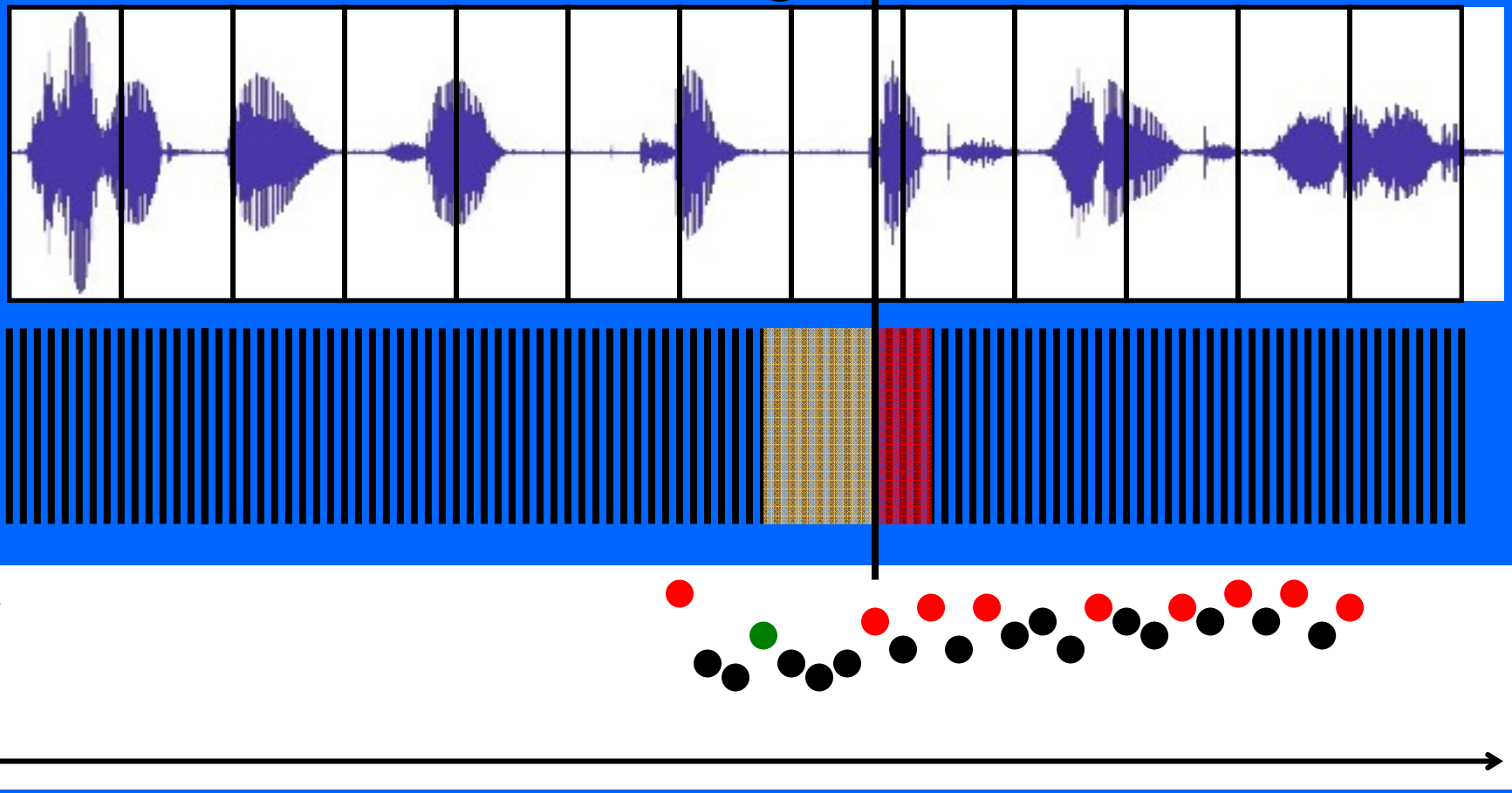
- Validate candidates using BIC



DISTBIC (validation)

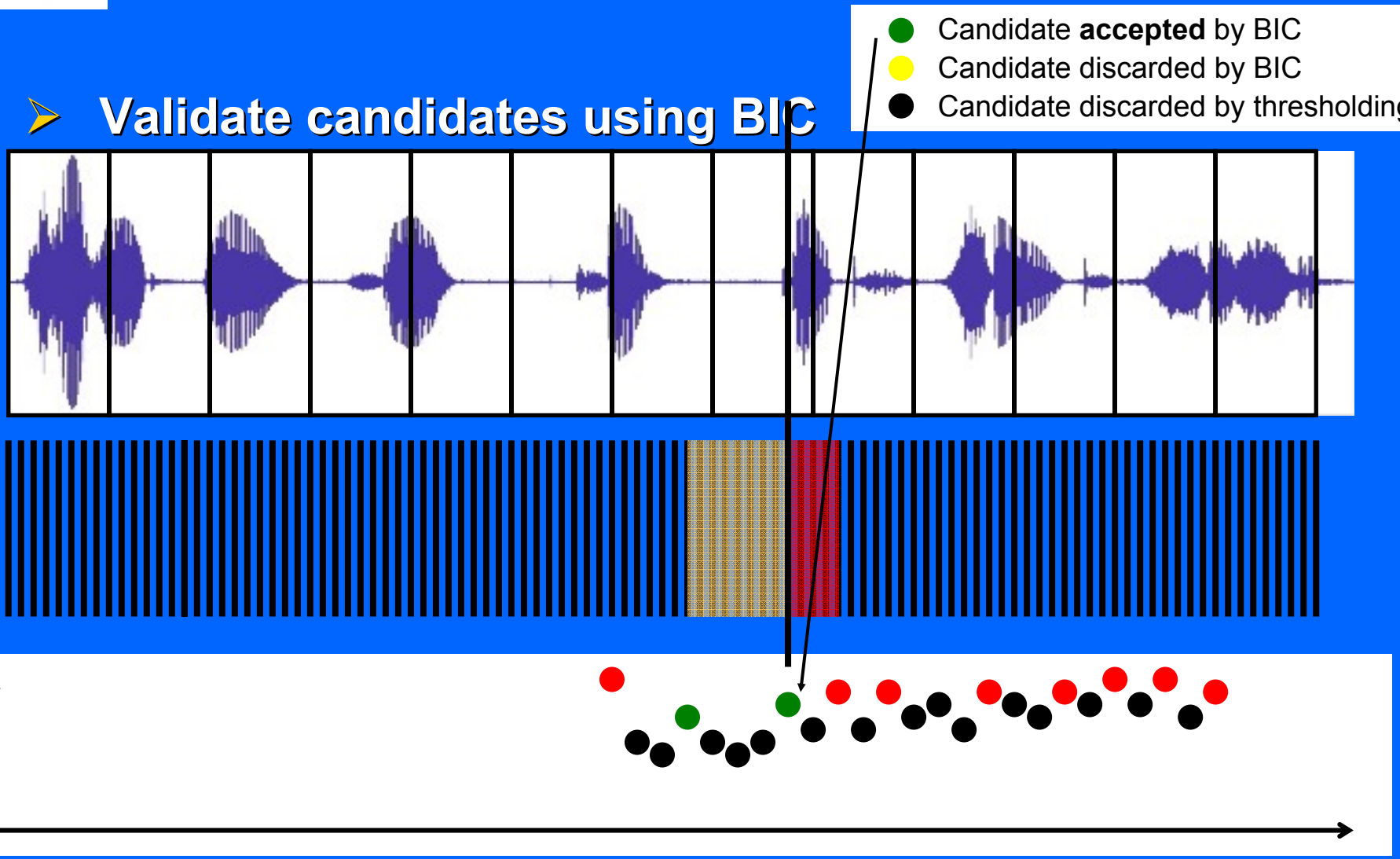
- Candidate accepted by BIC
- Candidate discarded by BIC
- Candidate discarded by thresholding

- Validate candidates using BIC



DISTBIC (validation)

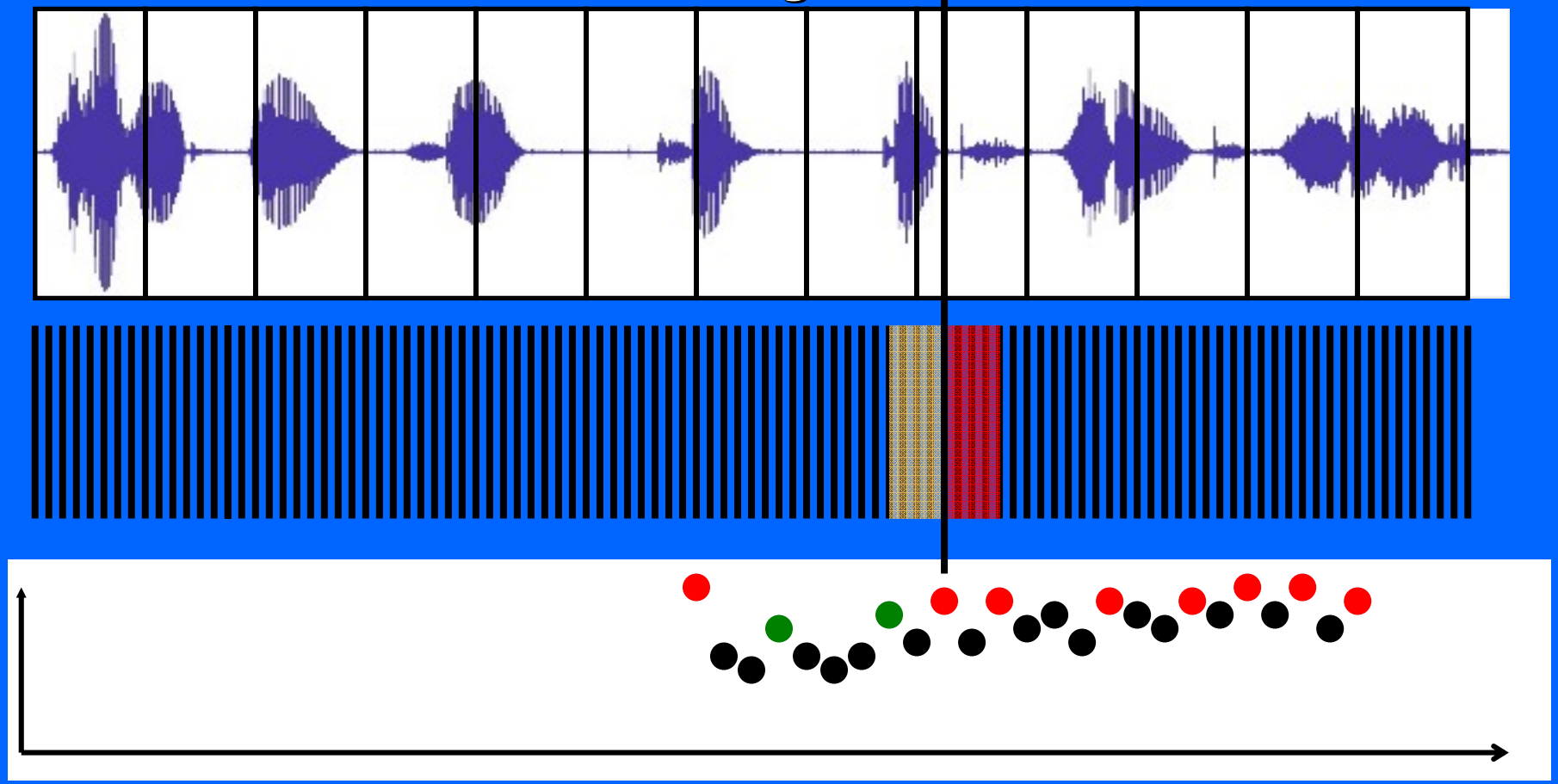
- Validate candidates using BIC



DISTBIC (validation)

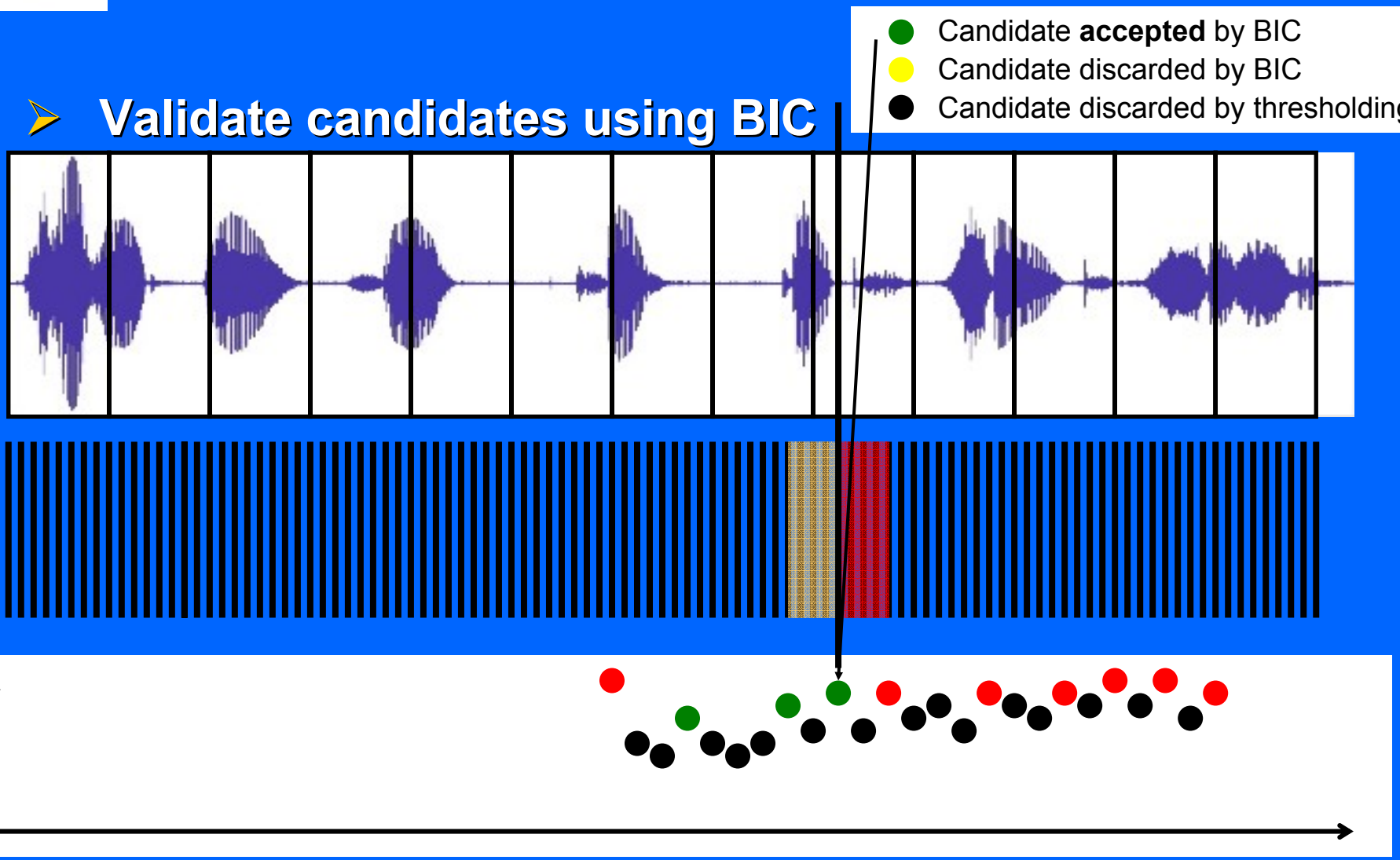
- Candidate accepted by BIC
- Candidate discarded by BIC
- Candidate discarded by thresholding

- Validate candidates using BIC



DISTBIC (validation)

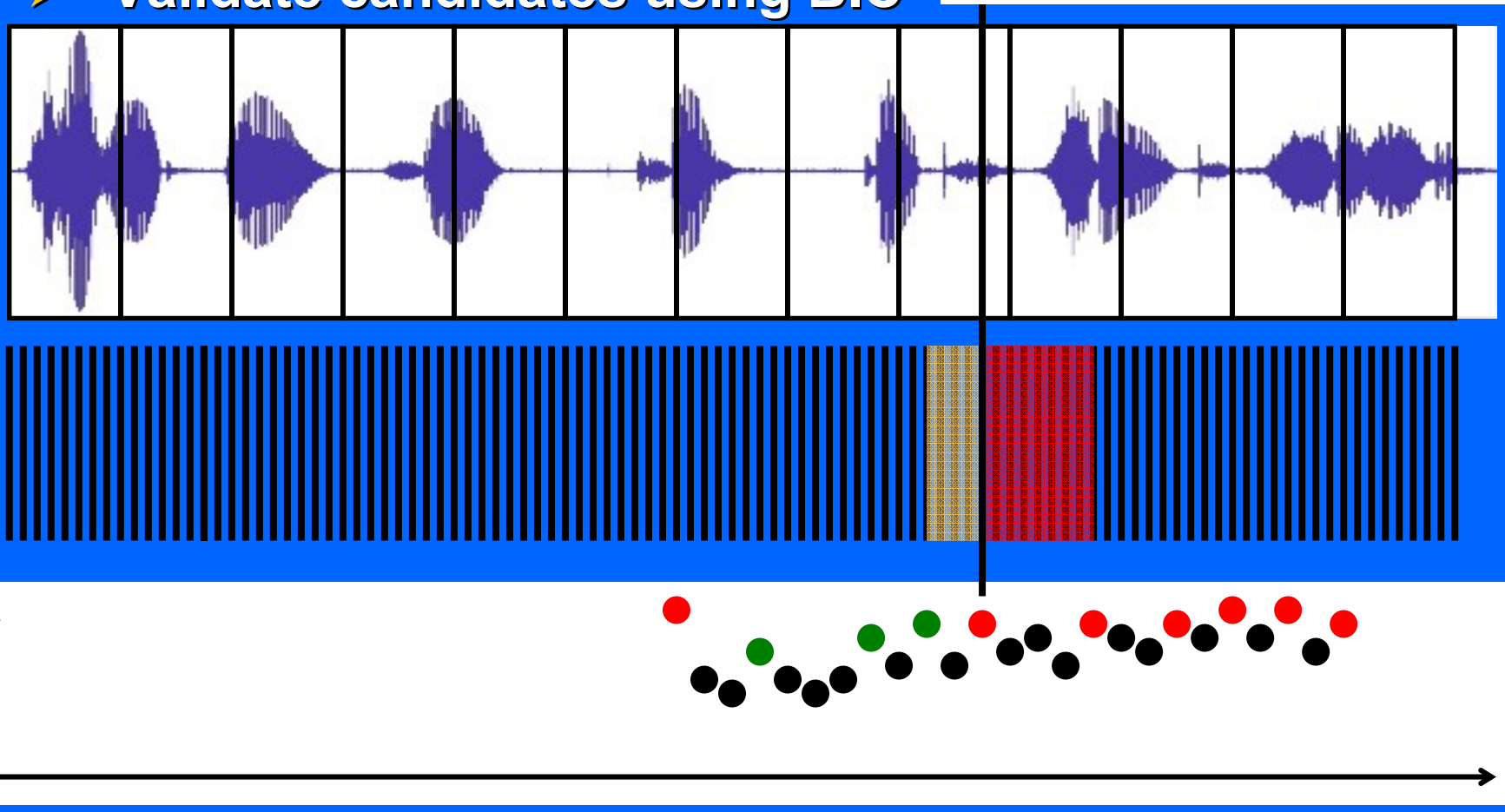
- Validate candidates using BIC



DISTBIC (validation)

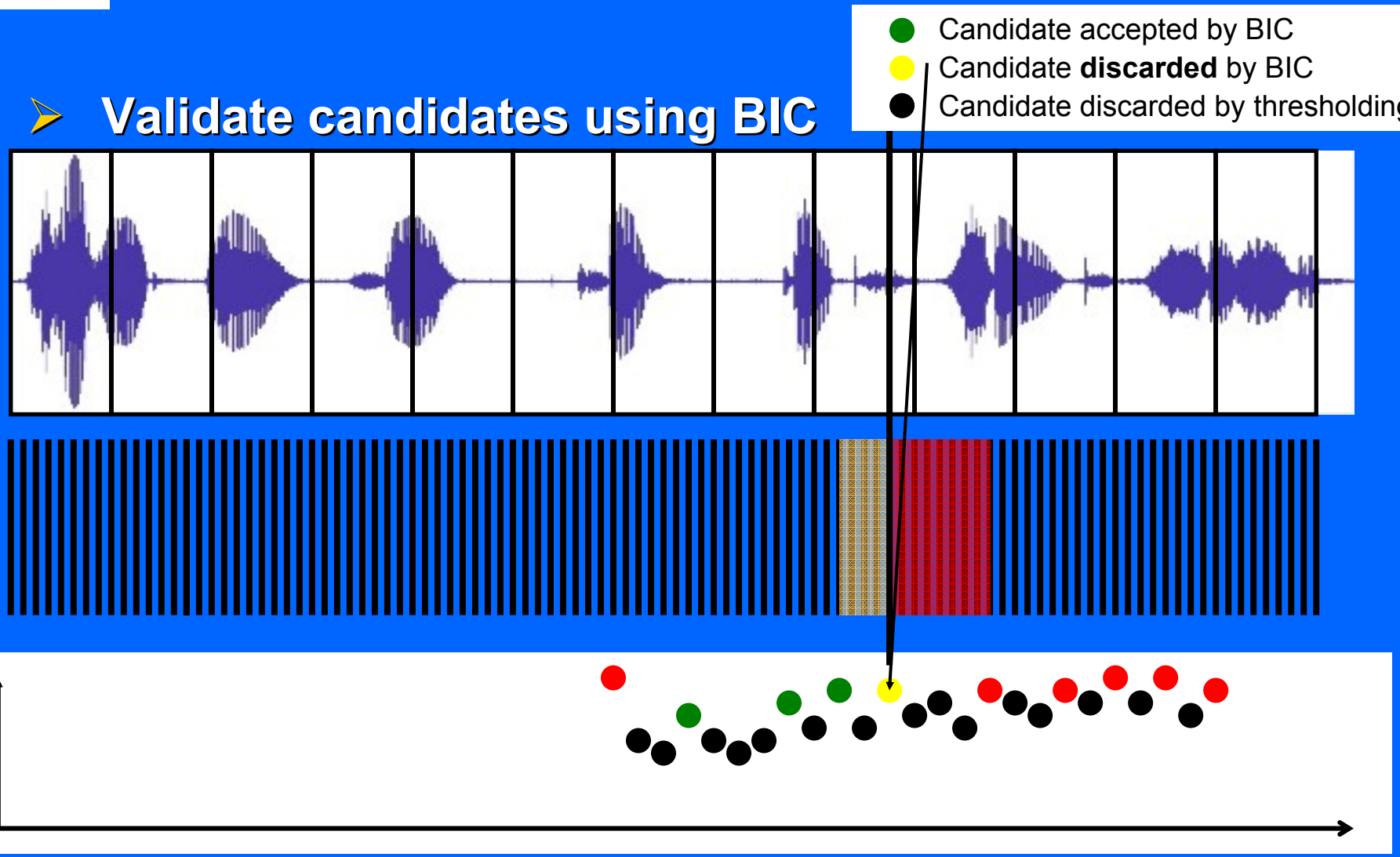
- Candidate accepted by BIC
- Candidate discarded by BIC
- Candidate discarded by thresholding

- Validate candidates using BIC



DISTBIC (validation)

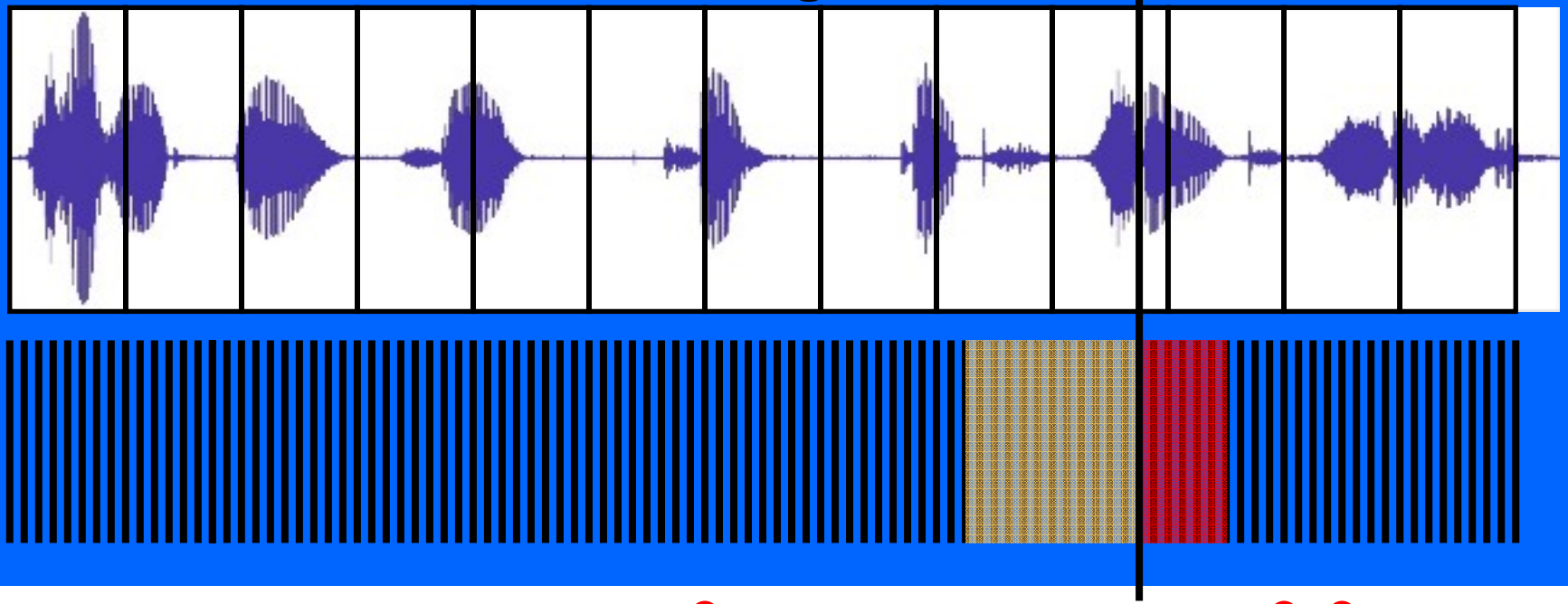
- Validate candidates using BIC



DISTBIC (validation)

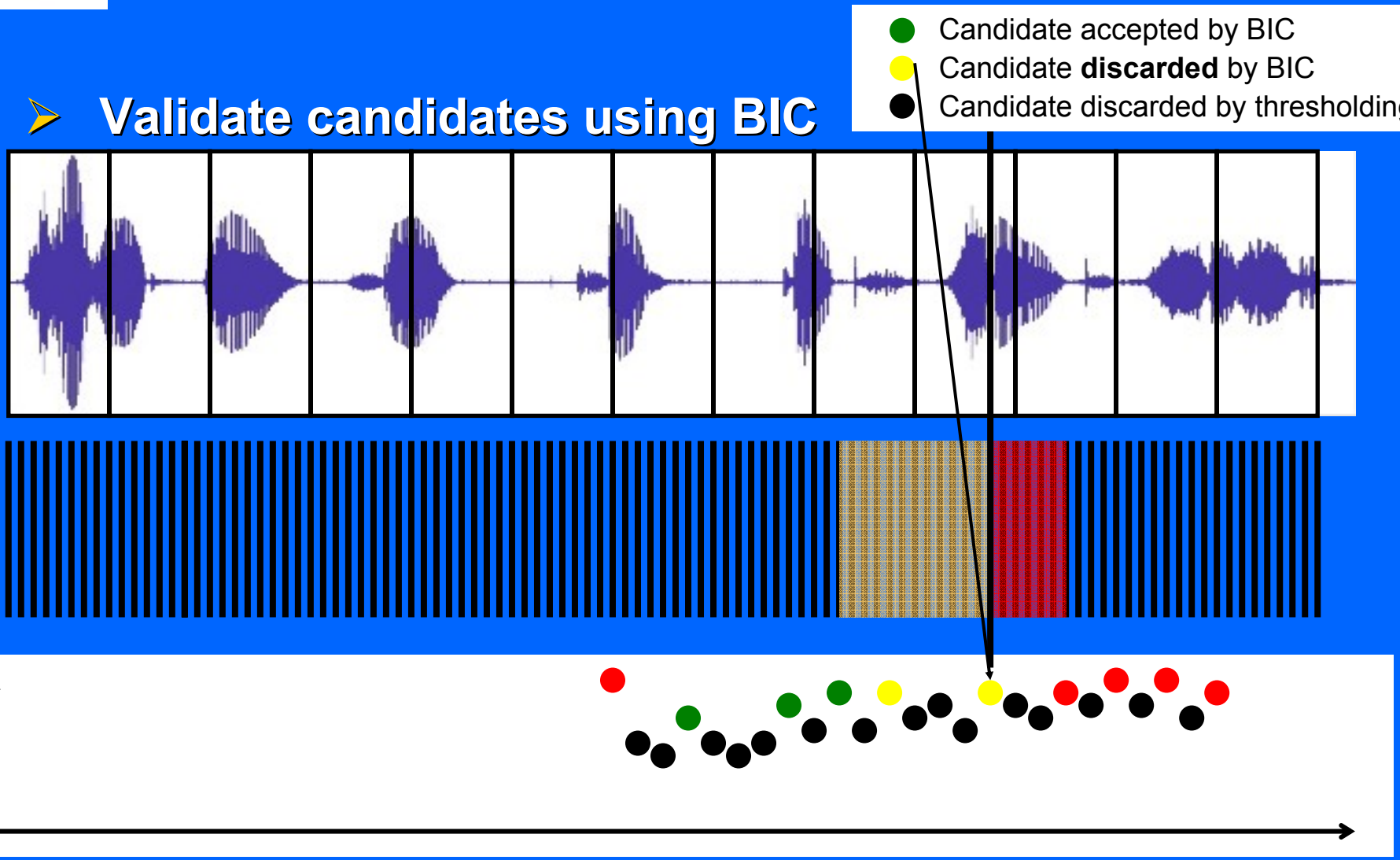
- Candidate accepted by BIC
- Candidate discarded by BIC
- Candidate discarded by thresholding

➢ **Validate candidates using BIC**



DISTBIC (validation)

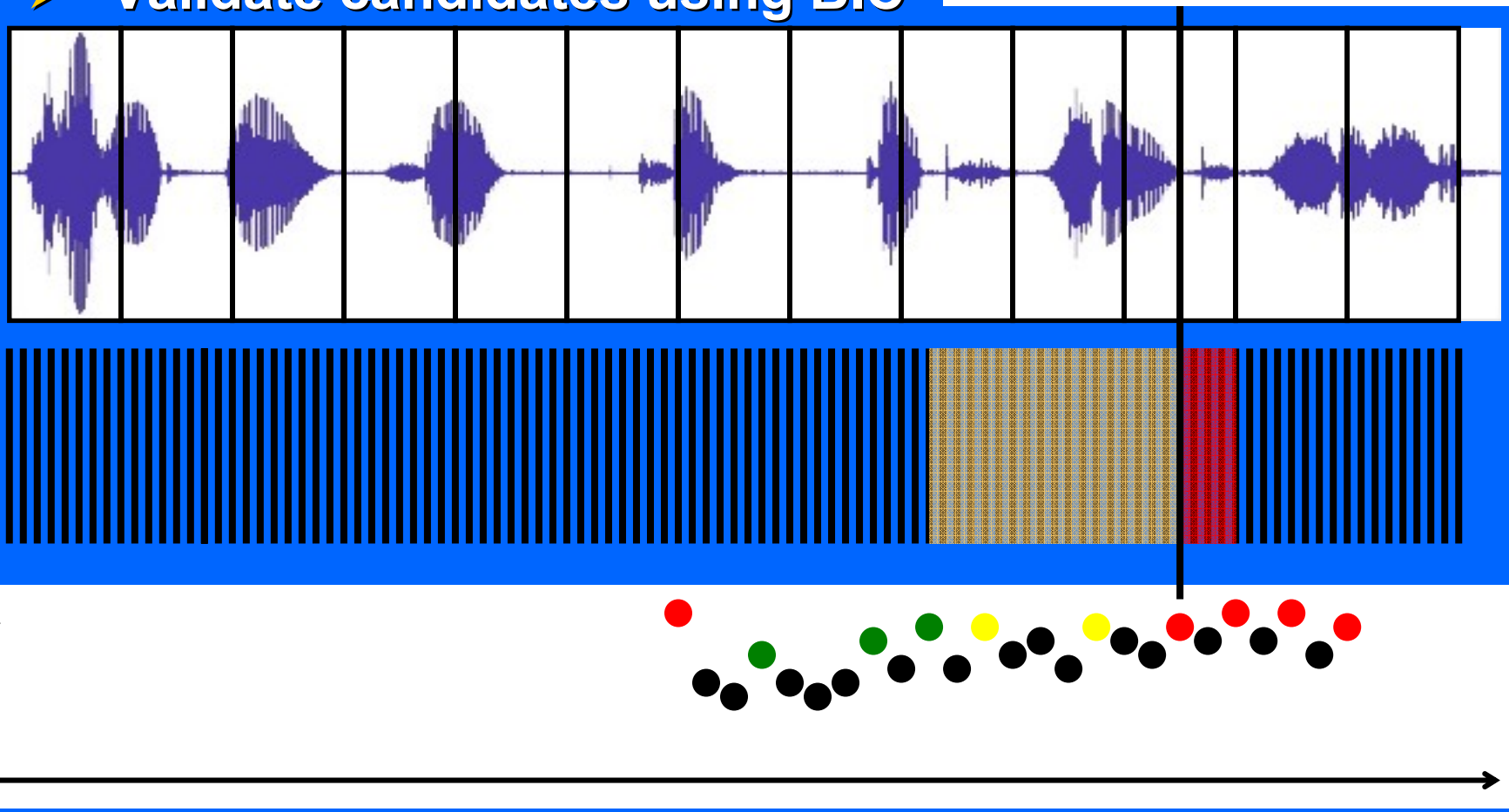
- Validate candidates using BIC



DISTBIC (validation)

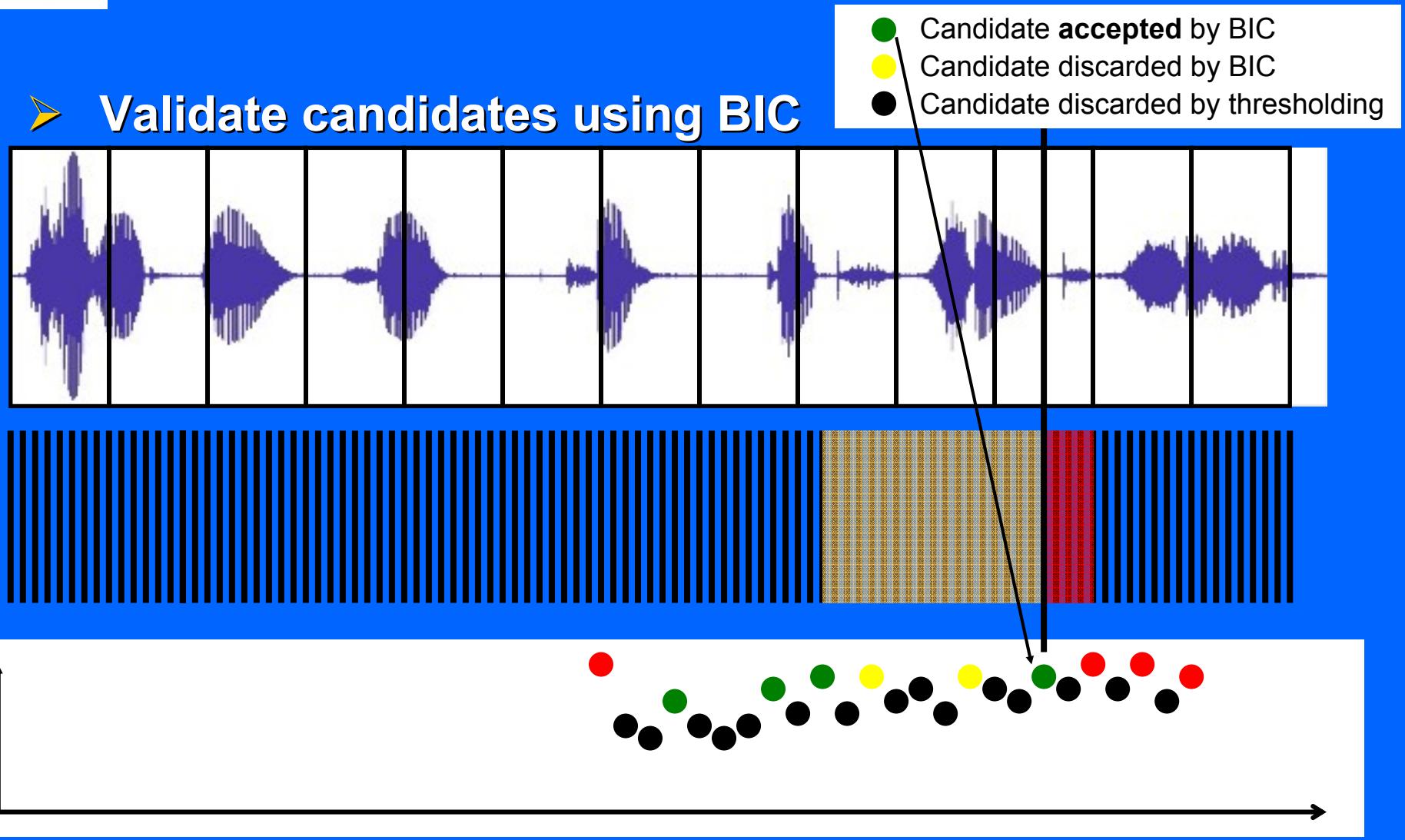
- Candidate accepted by BIC
- Candidate discarded by BIC
- Candidate discarded by thresholding

- Validate candidates using BIC



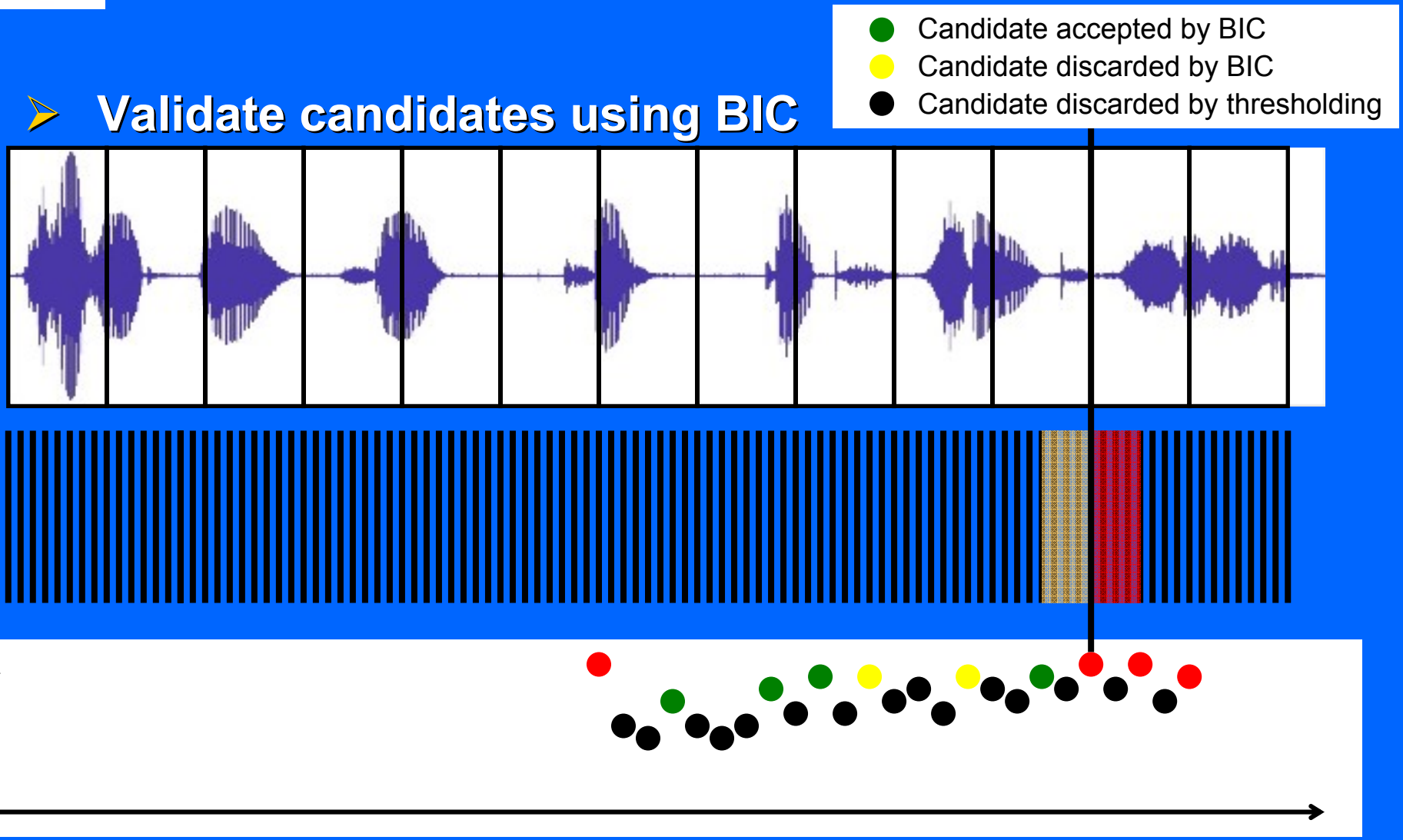
DISTBIC (validation)

- Validate candidates using BIC



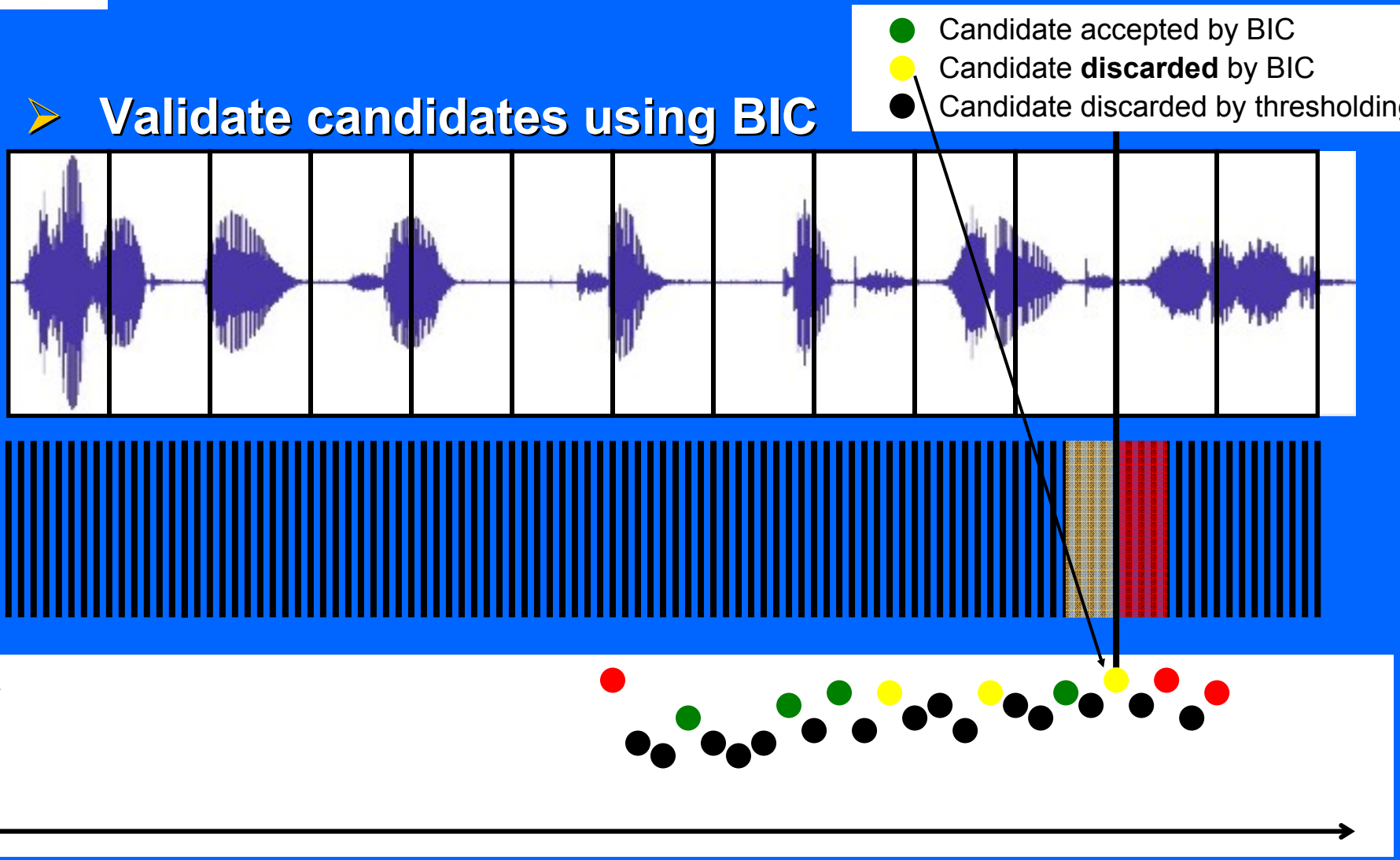
DISTBIC (validation)

- Validate candidates using BIC



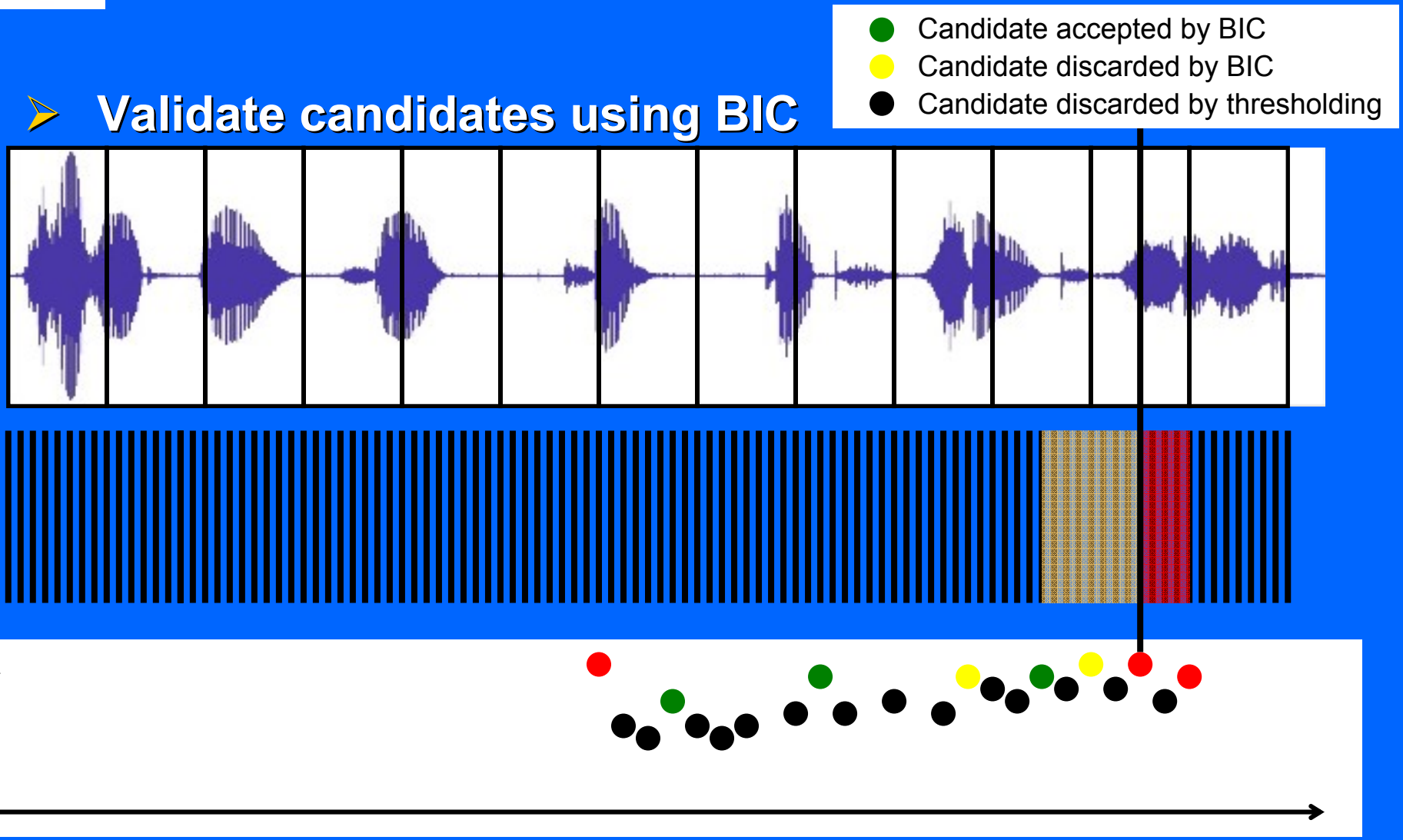
DISTBIC (validation)

- Validate candidates using BIC



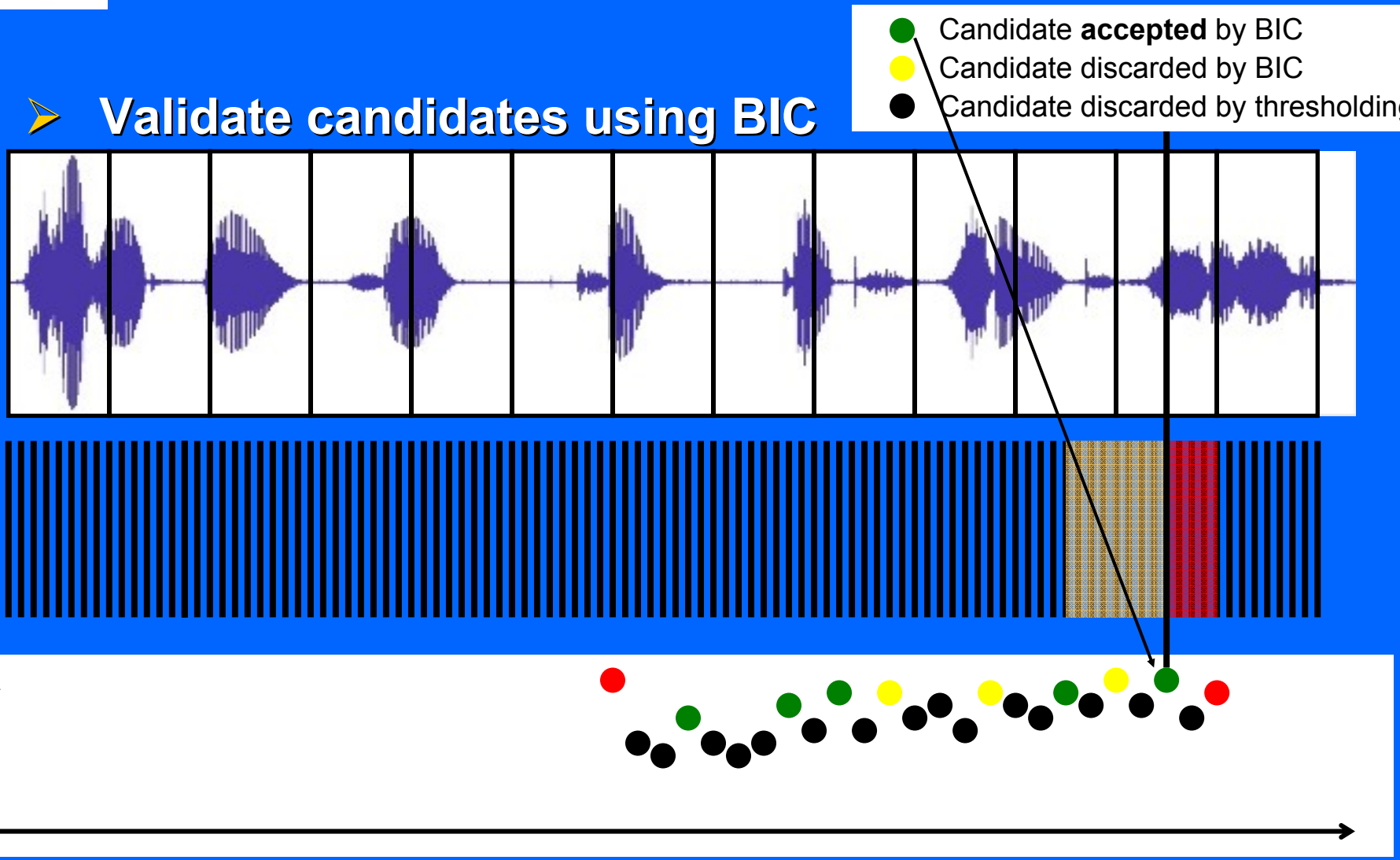
DISTBIC (validation)

- Validate candidates using BIC



DISTBIC (validation)

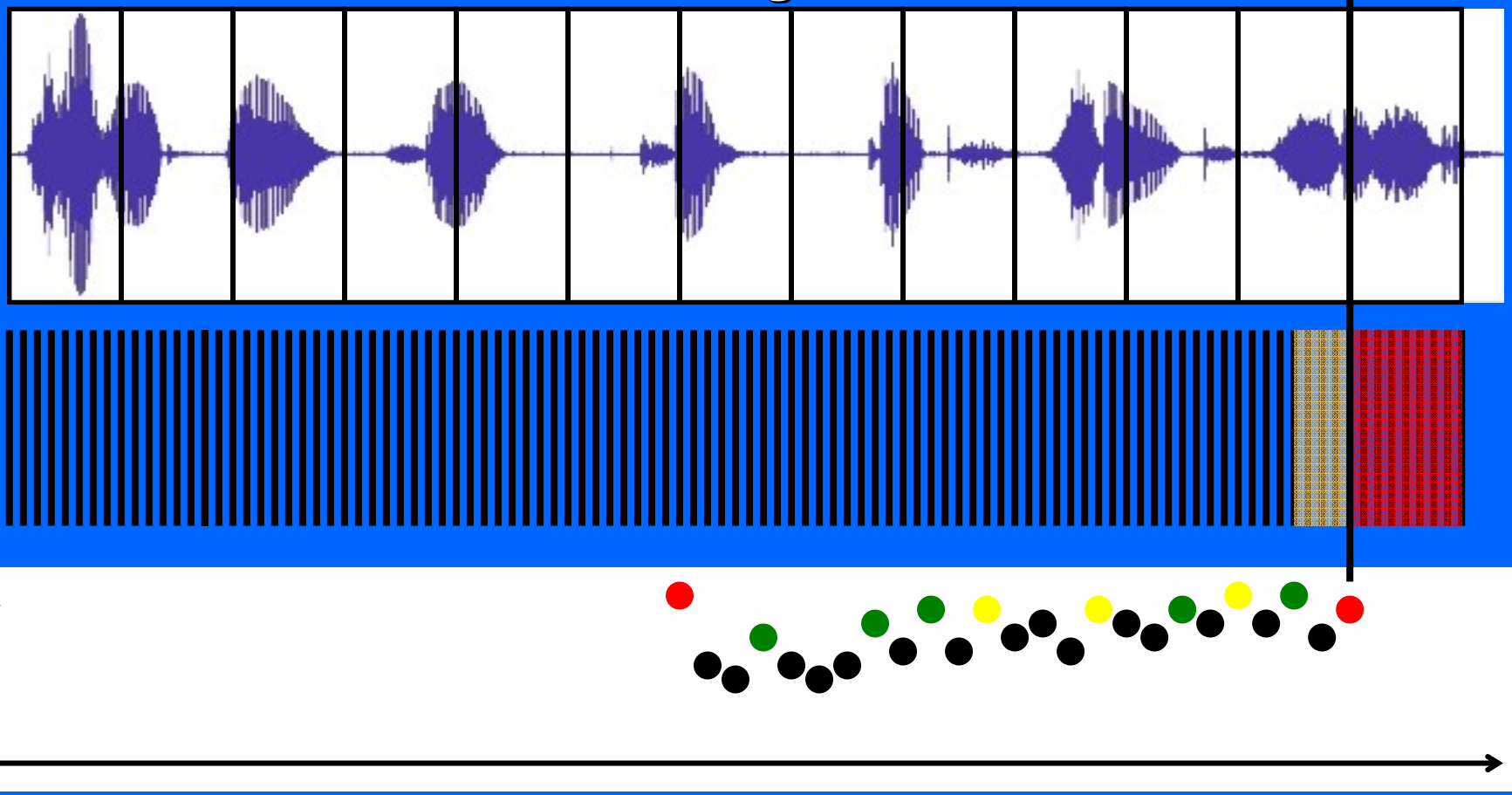
- Validate candidates using BIC



DISTBIC (validation)

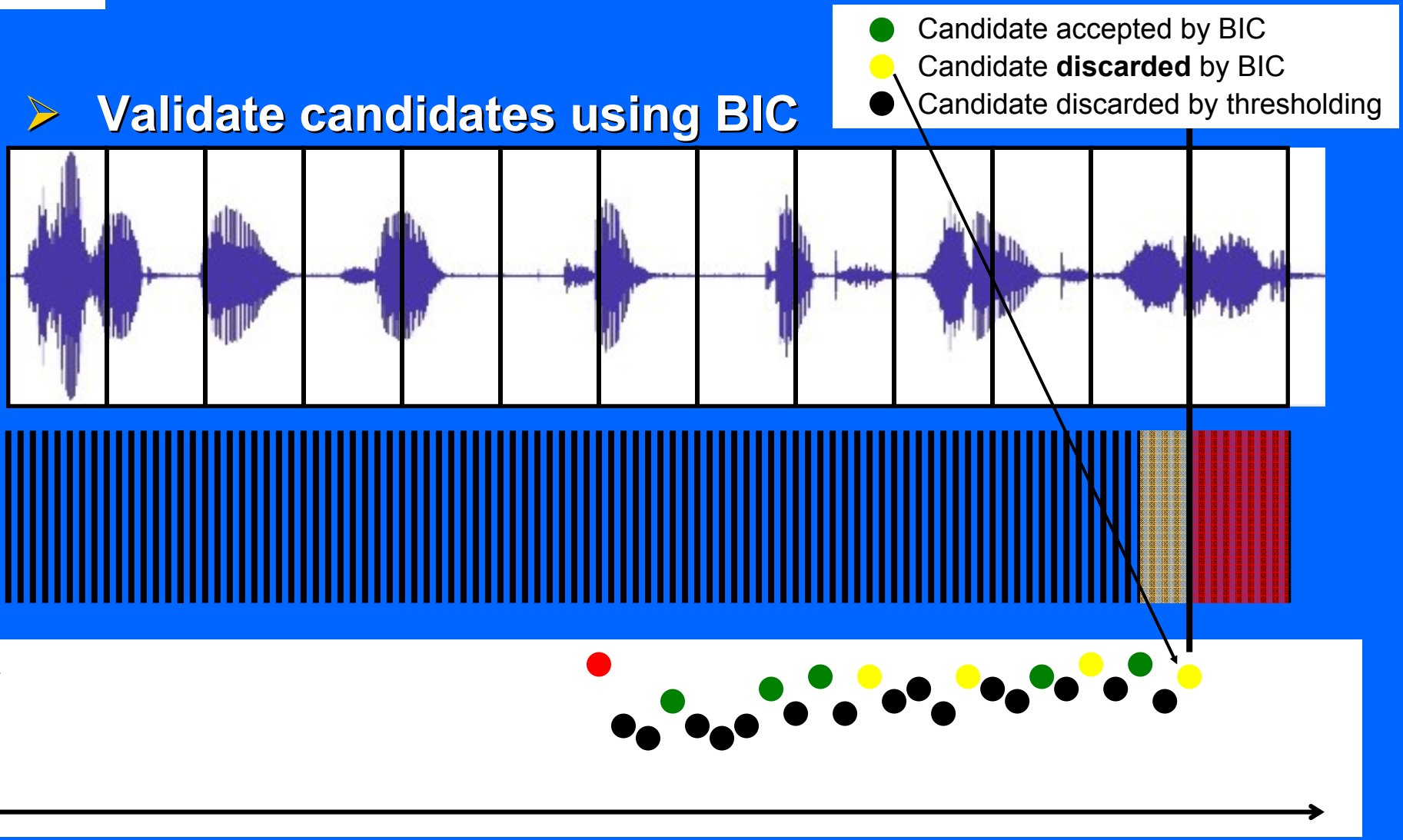
- Candidate accepted by BIC
- Candidate discarded by BIC
- Candidate discarded by thresholding

- Validate candidates using BIC



DISTBIC (validation)

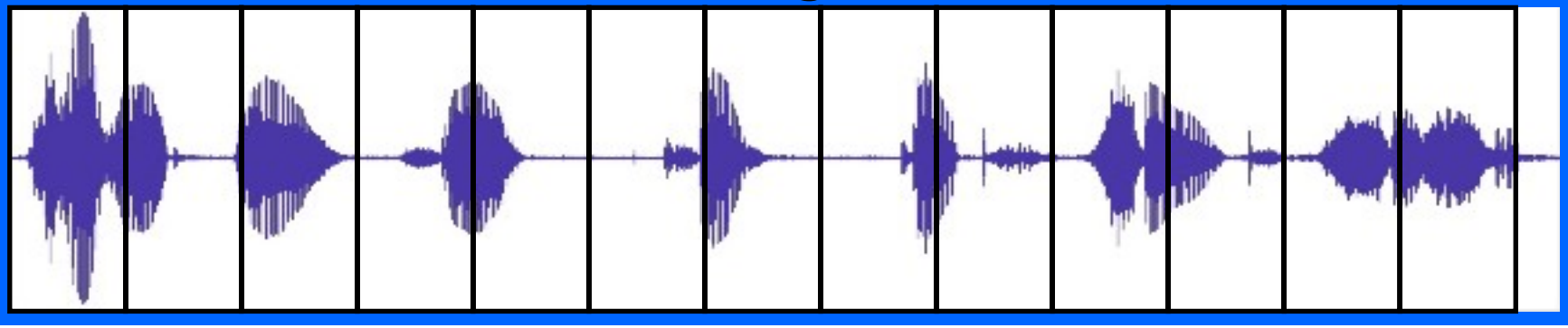
- Validate candidates using BIC



DISTBIC (validation)

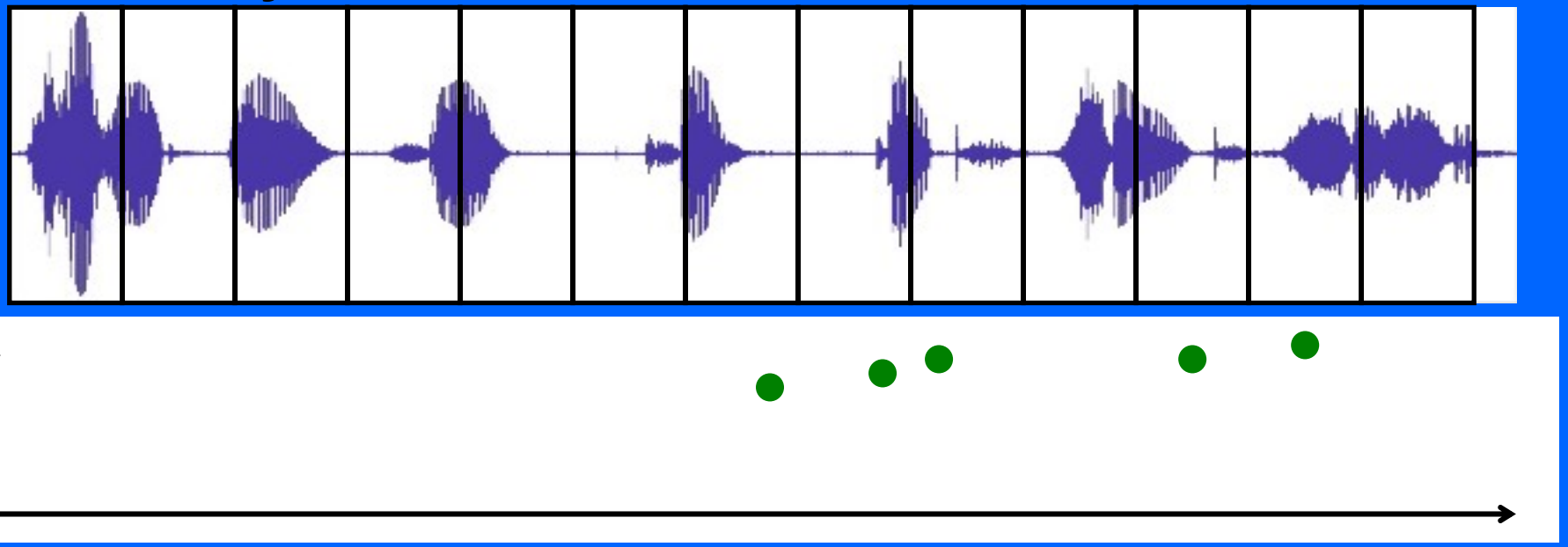
- Candidate accepted by BIC
- Candidate discarded by BIC
- Candidate discarded by thresholding

- Validate candidates using BIC



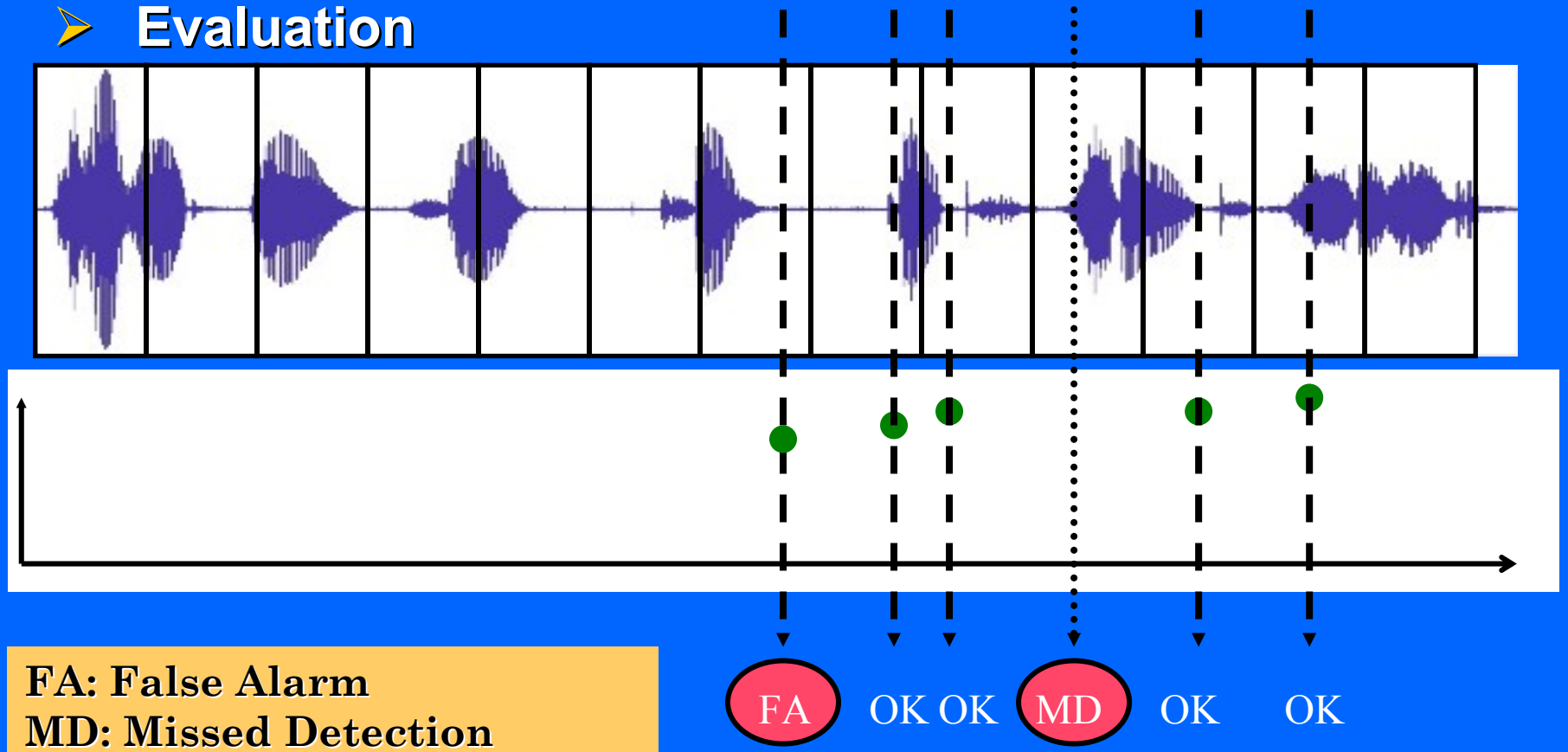
DISTBIC (validation)

- Finally:



DISTBIC (Evaluation)

- Evaluation



FA: False Alarm
MD: Missed Detection
OK: Correctly Found change

PROBLEM (1)

- Covariance matrices have been estimated by sample dispersion matrices.
- Alternative estimators: Robust estimates (e.g. M-estimates), regularized MLEs.

- Can we formulate efficiently BIC?
- SOLUTION: Simultaneous diagonalization of covariance matrices Σ_X and Σ_Z as well as Σ_Y and Σ_Z)

Transformed BIC (1)

- Standard BIC (for sample dispersion matrices)

$$\Delta BIC = \frac{B}{2}\ln|\mathbf{\Sigma}_Z| - \frac{A}{2}\ln|\mathbf{\Sigma}_X| - \frac{B-A}{2}\ln|\mathbf{\Sigma}_Y| - \frac{\lambda}{2}P\ln B > 0$$

- Simultaneous diagonalization

$$\begin{aligned}
\mathbf{\Sigma}_Z &= \mathbf{\Phi}\mathbf{\Lambda}_Z\mathbf{\Phi}^T \\
\mathbf{K} &\equiv \mathbf{\Lambda}_Z{}^{-\frac{1}{2}}\mathbf{\Phi}^T\mathbf{\Sigma}_X\mathbf{\Phi}\mathbf{\Lambda}_Z{}^{-\frac{1}{2}} \\
\mathbf{\Lambda}_K &= \mathbf{\Psi}^T\mathbf{K}\mathbf{\Psi} \\
\mathbf{W} &\equiv \mathbf{\Phi}\mathbf{\Lambda}_Z{}^{-\frac{1}{2}}\mathbf{\Psi}.
\end{aligned}$$

$$\begin{aligned}
\mathbf{W}^T\mathbf{\Sigma}_Z\mathbf{W} &= \mathbf{I} \\
\mathbf{W}^T\mathbf{\Sigma}_X\mathbf{W} &= \mathbf{\Lambda}_K
\end{aligned}$$

$$\begin{aligned}
\mathbf{H} &\equiv \mathbf{\Lambda}_Z{}^{-\frac{1}{2}}\mathbf{\Phi}^T\mathbf{\Sigma}_Y\mathbf{\Phi}\mathbf{\Lambda}_Z{}^{-\frac{1}{2}} \\
\mathbf{\Lambda}_H &= \mathbf{\Xi}^T\mathbf{H}\mathbf{\Xi} \\
\mathbf{\Omega} &\equiv \mathbf{\Phi}\mathbf{\Lambda}_Z{}^{-\frac{1}{2}}\mathbf{\Xi}.
\end{aligned}$$

$$\begin{aligned}
\mathbf{\Omega}^T\mathbf{\Sigma}_Z\mathbf{\Omega} &= \mathbf{I} \\
\mathbf{\Omega}^T\mathbf{\Sigma}_Y\mathbf{\Omega} &= \mathbf{\Lambda}_H
\end{aligned}$$

- Transformed BIC

$$\Delta BIC = \frac{A}{2}\sum_{j=1}^{d}\ln\Lambda_{K_{jj}} + \frac{B-A}{2}\sum_{j=1}^{d}\ln\Lambda_{H_{jj}} + \frac{\lambda}{2}P\ln B < 0$$

Transformed BIC (2)

- Let us introduce the centered feature vectors:

$$\begin{aligned}\tilde{\mathbf{z}}_i &= \mathbf{z}_i - \boldsymbol{\mu}_Z, \quad i = 1, 2, \ldots, A \\ \tilde{\mathbf{z}}_i &= \mathbf{z}_i - \boldsymbol{\mu}_Z, \quad i = A+1, A+2, \ldots, B\end{aligned}$$

$$\begin{aligned}\boldsymbol{\mu}'_X &= \boldsymbol{\mu}_X - \boldsymbol{\mu}_Z = \frac{1}{A}\sum_{i=1}^{A} \tilde{\mathbf{z}}_i \\ \boldsymbol{\mu}'_Y &= \boldsymbol{\mu}_Y - \boldsymbol{\mu}_Z = \frac{1}{(B-A)}\sum_{i=A+1}^{B} \tilde{\mathbf{z}}_i\end{aligned}$$

- then

$$\begin{aligned}\Delta BIC' = {} & \frac{A}{2}\ln\frac{|\boldsymbol{\Sigma}_Z|}{|\boldsymbol{\Sigma}_X|} + \frac{1}{2}\sum_{i=1}^{A}\tilde{\mathbf{z}}_i^T\boldsymbol{\Sigma}_Z^{-1}\tilde{\mathbf{z}}_i - \frac{1}{2}\sum_{i=1}^{A}\tilde{\mathbf{z}}_i^T\boldsymbol{\Sigma}_X^{-1}\tilde{\mathbf{z}}_i + \frac{A}{2}\boldsymbol{\mu}'_X{}^T\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\mu}'_X \\ & \frac{(B-A)}{2}\ln\frac{|\boldsymbol{\Sigma}_Z|}{|\boldsymbol{\Sigma}_Y|} + \frac{1}{2}\sum_{i=A+1}^{B}\tilde{\mathbf{z}}_i^T\boldsymbol{\Sigma}_Z^{-1}\tilde{\mathbf{z}}_i - \frac{1}{2}\sum_{i=A+1}^{B}\tilde{\mathbf{z}}_i^T\boldsymbol{\Sigma}_Y^{-1}\tilde{\mathbf{z}}_i \\ & + \frac{(B-A)}{2}\boldsymbol{\mu}'_Y{}^T\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\mu}'_Y - \frac{\lambda}{2}P\ln B > 0\end{aligned}$$

Transformed BIC (3)

$$\Delta BIC' = \underbrace{\frac{A}{2}\ln\frac{|\mathbf{\Sigma}_Z|}{|\mathbf{\Sigma}_X|} + \frac{(B-A)}{2}\ln\frac{|\mathbf{\Sigma}_Z|}{|\mathbf{\Sigma}_Y|} - \frac{\lambda}{2}P\ln B}_{\Delta BIC}$$
$$+\frac{1}{2}\sum_{i=1}^{B}\tilde{\mathbf{z}}_i^T\mathbf{\Sigma}_Z^{-1}\tilde{\mathbf{z}}_i - \frac{1}{2}\sum_{i=1}^{A}\tilde{\mathbf{z}}_i^T\mathbf{\Sigma}_X^{-1}\tilde{\mathbf{z}}_i - \frac{1}{2}\sum_{i=A+1}^{B}\tilde{\mathbf{z}}_i^T\mathbf{\Sigma}_Y^{-1}\tilde{\mathbf{z}}_i$$
$$+\frac{A}{2}\boldsymbol{\mu}_X'^T\mathbf{\Sigma}_X^{-1}\boldsymbol{\mu}_X' + \frac{(B-A)}{2}\boldsymbol{\mu}_Y'^T\mathbf{\Sigma}_Y^{-1}\boldsymbol{\mu}_Y' > 0$$

$$-\underbrace{\frac{A}{2}\sum_{j-1}^{d}\ln\Lambda_{K_{jj}} - \frac{B-A}{2}\sum_{j=1}^{d}\ln\Lambda_{H_{jj}} - \frac{\lambda}{2}P\ln B}_{\Delta BIC} - \frac{1}{2}\sum_{i=1}^{A}\tilde{\mathbf{w}}_i^T(\mathbf{\Lambda}_K^{-1}-\mathbf{I})\tilde{\mathbf{w}}_i$$
$$-\frac{1}{2}\sum_{i=A+1}^{B}\tilde{\mathbf{v}}_i^T(\mathbf{\Lambda}_H^{-1}-\mathbf{I})\tilde{\mathbf{v}}_i + \frac{A}{2}\boldsymbol{\mu}_X'^T\mathbf{\Sigma}_X^{-1}\boldsymbol{\mu}_X' + \frac{(B-A)}{2}\boldsymbol{\mu}_Y'^T\mathbf{\Sigma}_Y^{-1}\boldsymbol{\mu}_Y' > 0$$

where

$$\tilde{\mathbf{w}}_i = \mathbf{W}^T\tilde{\mathbf{z}}_i, \quad i = 1,2,\ldots,A$$
$$\tilde{\mathbf{v}}_i = \mathbf{\Omega}^T\tilde{\mathbf{z}}_i, \quad i = A+1, A+2,\ldots,B$$

Computational cost (1)

- Standard BIC

$$-\sum_{i=1}^{B}(\boldsymbol{z}_i-\boldsymbol{\mu}_Z)^T\boldsymbol{\Sigma}_Z^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_Z)+\sum_{i=1}^{A}(\boldsymbol{z}_i-\boldsymbol{\mu}_X)^T\boldsymbol{\Sigma}_X^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_X)$$

$$+\sum_{i=A+1}^{B}(\boldsymbol{z}_i-\boldsymbol{\mu}_Y)^T\boldsymbol{\Sigma}_Y^{-1}(\boldsymbol{z}_i-\boldsymbol{\mu}_Y)<B\ln|\boldsymbol{\Sigma}_Z|-A\ln|\boldsymbol{\Sigma}_X|-(B-A)\ln|\boldsymbol{\Sigma}_Y|-\lambda P\ln B=\gamma$$

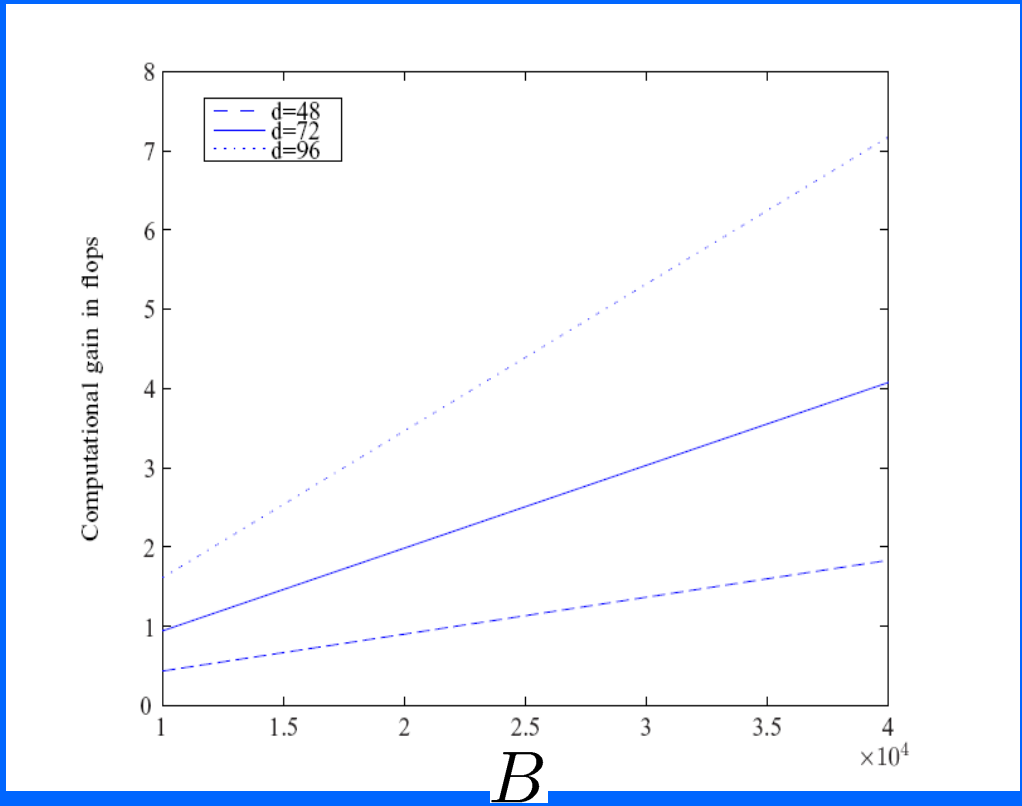
- Transformed BIC

$$\sum_{i=1}^{A}\tilde{\mathbf{w}}_i^T(\boldsymbol{\Lambda}_K^{-1}-\mathbf{I})\tilde{\mathbf{w}}_i\sum_{i=A+1}^{B}\tilde{\mathbf{v}}_i^T(\boldsymbol{\Lambda}_H^{-1}-\mathbf{I})\tilde{\mathbf{v}}_i<\gamma+A\boldsymbol{\mu}'_X{}^T\boldsymbol{\Sigma}_X^{-1}\boldsymbol{\mu}'_X+(B-A)\boldsymbol{\mu}'_Y{}^T\boldsymbol{\Sigma}_Y^{-1}\boldsymbol{\mu}'_Y$$

Computational cost (2)

- Standard BIC $3d^3 + 6Bd^2 + (8B + 3)d + 2$

Transformed BIC $30d^3 + (4B + 4)d^2 + (7B + 9)d + 5$

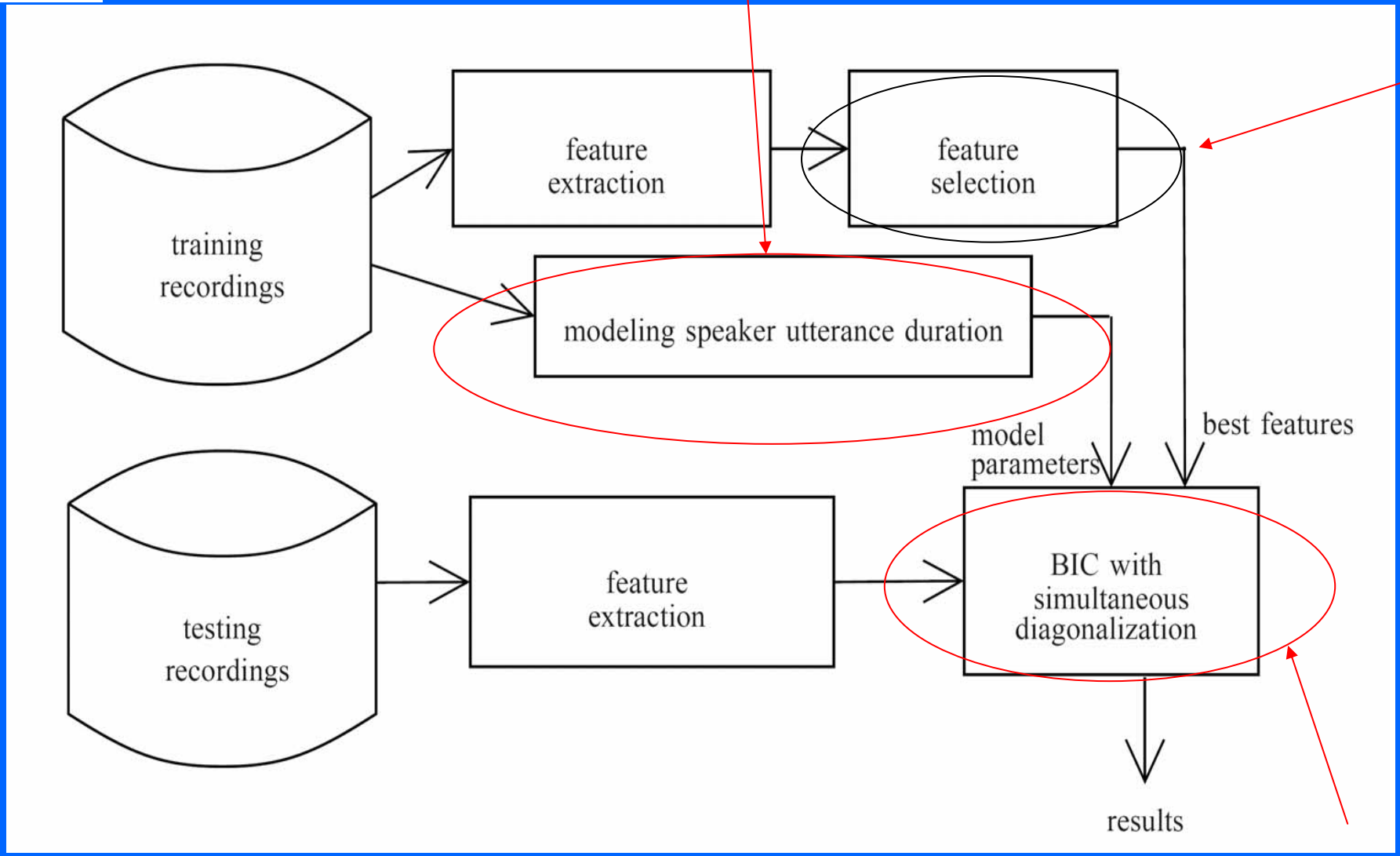


$B \gg d$

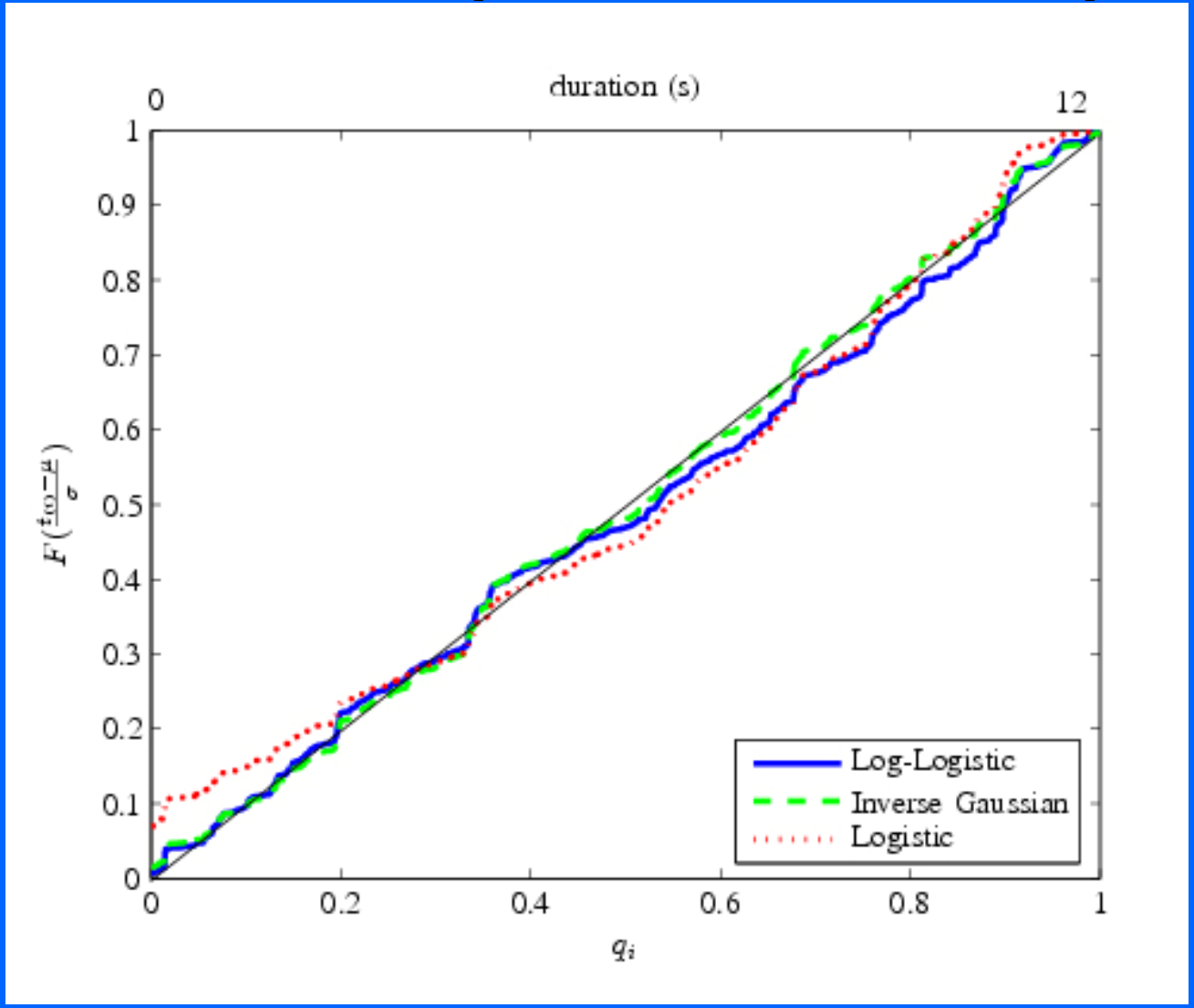
BIC-based Speaker Segmentation

- To improve performance
  - estimate speaker utterance duration
  - select the most efficient features
  - derive a new BIC formulation

Speaker Segmentation System



P-P plot (TIMIT subset)



MFCC feature selection (1)

- Instead of trying to reveal the MFCC *order* that yields the most accurate results, an MFCC *subset* that is more suitable for detecting a speaker change is computed
- From an initial set of 36 MFCCs, the 24 more suitable MFCCs are derived

MFCC feature selection (2)

- Branch and bound, depth-first search, and backtracking
- Selection criterion $J = \mathrm{tr}(\mathbf{S}_w^{-1}\mathbf{S}_b)$
- Selected subset: 1, 3-13, 16, 22-29, 31, 33, 35, 36
- Used in conjunction with their delta- and delta-delta coefficients

Evaluation metrics

- **False Alarm Rate (FAR)**

$$FAR = \frac{\text{number of FA}}{\text{number of ACP} + \text{number of FA}} 100\%$$

- **Missed Detection Rate (MDR)**

$$MDR = \frac{\text{number of MD}}{\text{number of ACP}} 100\%$$

FA= False Alarms, MD=Missed Detections, ACP=Actual Change Points

high value of FAR -> over-segmentation of the speech signal
high value of MDR -> algorithm does not segment the signal properly

Evaluation metrics (2)

- Precision (PRC) and Recall (RCL)

$$PRC = \frac{CFC}{DET}100\% \qquad RCL = \frac{CFC}{ACP}100\%$$

CFC=Correctly Found Changes
DET= Changes detected by the system
ACP=Actual Change Points

- F_1-measure (F_1)

$$F_1 = \frac{2 \cdot PRC \cdot RCL}{PRC + RCL}$$

F_1: overall objective effectiveness of the system

Experimental results

Table 1: conTIMIT

	PRC	RCL	F_1	FAR	MDR
mean	0.670	0.949	0.777	0.289	0.051
standard deviation	0.106	0.056	0.069	0.139	0.056

Table 2: HUB-4 1997 English Broadcast News Speech dataset

	PRC	RCL	F_1	FAR	MDR
mean	0.634	0.922	0.738	0.309	0.078
standard deviation	0.131	0.148	0.112	0.174	0.148

Performance comparison

Table 3: Performance comparison on broadcasts.

System	Database used	PRC	RCL	F_1	FAR	MDR
Proposed system	HUB-4 1997 English Broadcast News Speech	0.634	0.922	0.738	0.309	**0.078**
Lu and Zhang	HUB-4 1997 English Broadcast News Speech		0.89		0.15	
Ajmera et al.	HUB-4 English Evaluation Speech and Transcripts	0.68	0.65	0.67		
Cheng and Wang	MATBN-2002				0.289	0.100
Kim et al.	audio track from television talk show program	0.754	0.864	0.805		

PROBLEM (2)

- **The binary hypothesis test for phonemic segmentation requires small windows.**

- **Phone durations: are short as 10-20ms -> need for efficient signal modelling at this level**
  - GD is not a good fit for speech in small frame sizes.
  - What about noisy speech ?
  SOLUTION: model noisy speech with Generalized Gamma (GΓD)

- Use GΓD for GLRT tests in the first pass of DISTBIC (DISTBIC-Γ)

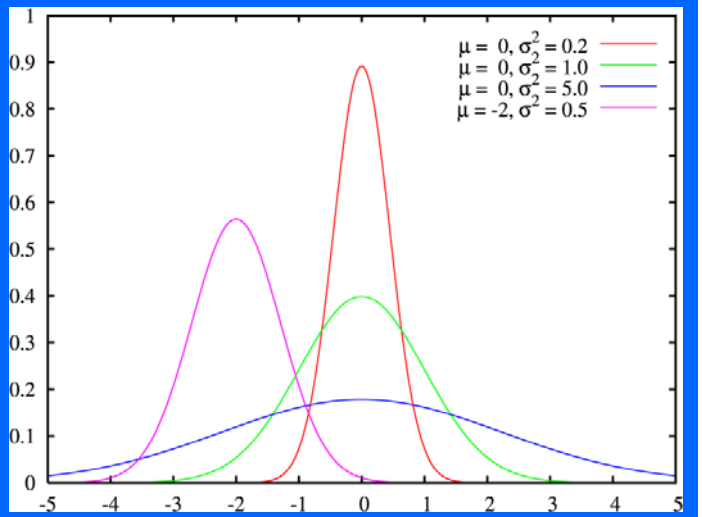
Speech Distributions (1)

- **Gaussian Distribution (GD)**

$$f(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

μ = mean

σ = standard deviation

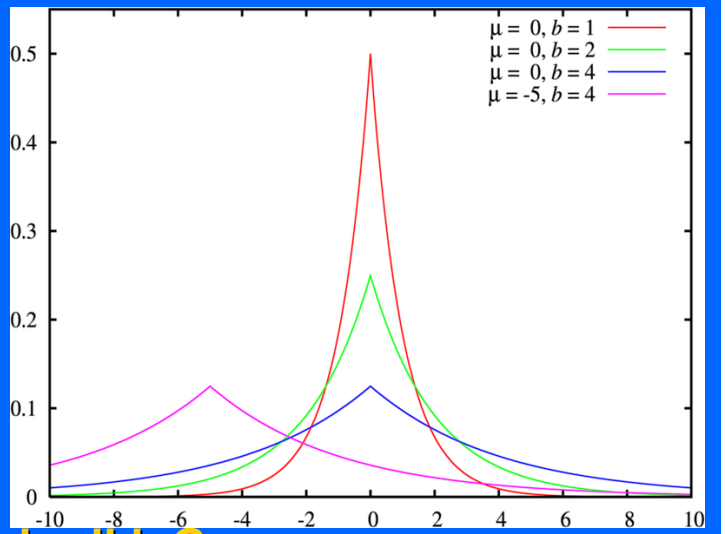


- **Laplace Distribution (LD)**

$$f(x;m,b) = \frac{1}{2b} e^{\frac{|x-\mu|}{b}}$$

μ = location

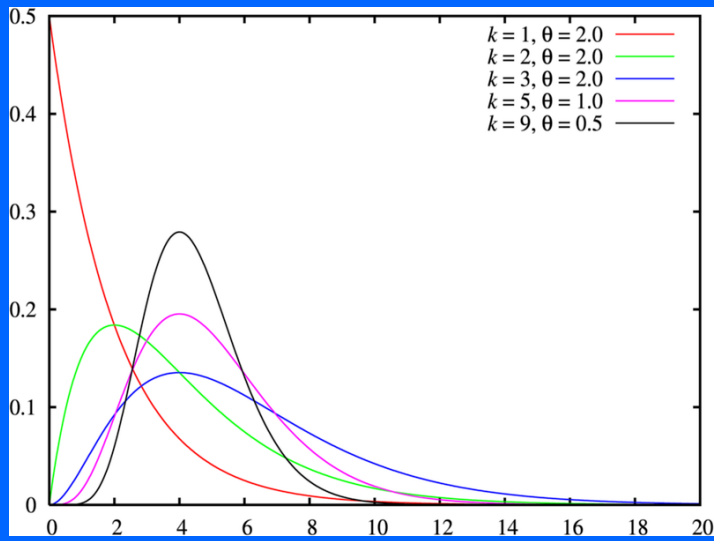
b = scale



Speech Distributions (2)

- **Gamma Distribution (ΓD)**

$$f(x;k,\theta) = x^{k-1}\frac{e^{-x/\theta}}{\theta^{k}\Gamma(k)}, x > 0$$

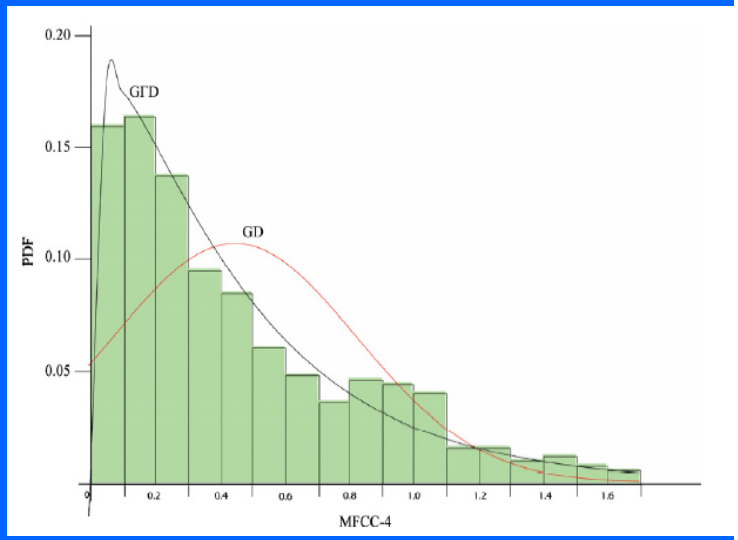


k>0 : shape
θ>0 : scale

- **Gazor (2003), Shin (2005), Martin (2005), Nakamura (2002), and others: ΓD > LD > GD (in speech signals)**

Generalized Gamma Distribution (GΓD)

- special cases: GD (*c*=2, *a*=0.5), LD (*c*=1, *a*=1), ΓD (*c*=1, *a*=0.5), Weibull, etc



$$f_x(x) = \frac{cb^a}{2\Gamma(a)} |x|^{ac-1} e^{-b|x|^c}$$

- (*a*: scale, *b*,*c*: shape)

Maximum Likelihood Estimation (MLE) of GΓD parameters

Given n observations (univariate i.i.d. RVs), the loglikelihood is given by:

$$\ell(a,b,c) = \ln f(\underline{x};a,b,c) = n \ln \frac{cb^a}{2\Gamma(a)} + (ac-1)\sum_{i=1}^{n} \ln|x_i| - b\sum_{i=1}^{n} |x_i|^c$$

$$\frac{\partial L(a,b,c)}{\partial a} = \frac{\partial L(a,b,c)}{\partial b} = \frac{\partial L(a,b,c)}{\partial c} = 0$$

- **System of nonlinear equations**

$$\begin{aligned} \psi_0(a) &= \ln b + \frac{1}{n}\sum_{i=1}^{n} \ln|x_i|^c \\ b &= \frac{a\,n}{\sum_{i=1}^{n}|x_i|^c} \\ 0 &= \frac{1}{a} + \psi_0(a) - \ln b - \frac{b}{a\,n}\sum_{i=1}^{n} |x_i|^c \ln|x_i|^c \end{aligned}$$

$$\psi_0(x) \equiv \frac{d\ln\Gamma(x)}{d\,x} = \frac{\Gamma'(x)}{\Gamma(x)}$$

Digamma function

Gradient Ascend Algorithm

Let ***ξ*** be **a forgetting factor** and ***μ*** *be the* ***l*****earning rate**

Start with an **initial** set of parameters and **iterate** for i

$$
\begin{aligned}
S_1(i) &= (1-\xi)S_1(i-1) + \xi\,|x_i|^{\hat{c}(i)} \\
S_2(i) &= (1-\xi)S_2(i-1) + \xi\,\ln|x_i|^{\hat{c}(i)} \\
S_3(i) &= (1-\xi)S_3(i-1) + \xi\,|x_i|^{\hat{c}(i)}\ln|x_i|^{\hat{c}(i)} \\
\psi_0(\hat{a}(i)) - \ln\hat{a}(i) &= S_2(i) - \ln S_1(i) \\
\hat{b}(i) &= \frac{\hat{a}(i)}{S_1(i)} \\
\hat{c}(i+1) &= \hat{c}(i) + \mu\left(\frac{1}{\hat{a}(i)} - S_2(i)\frac{S_3(i)}{S_1(i)}\right)
\end{aligned}
$$

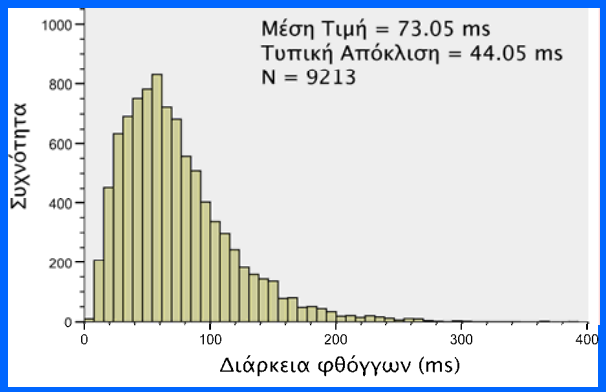
Monotonically increasing function of â*(i)* (use an inverse table)

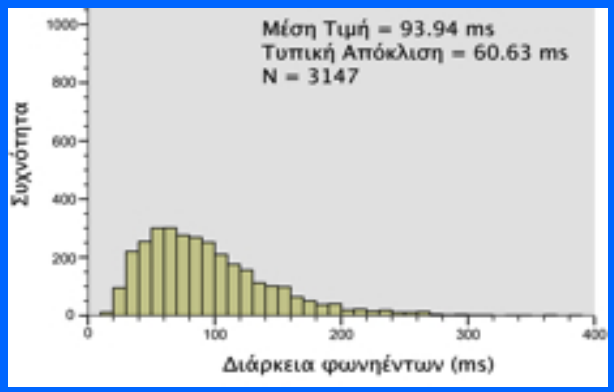
PROBLEM (3)

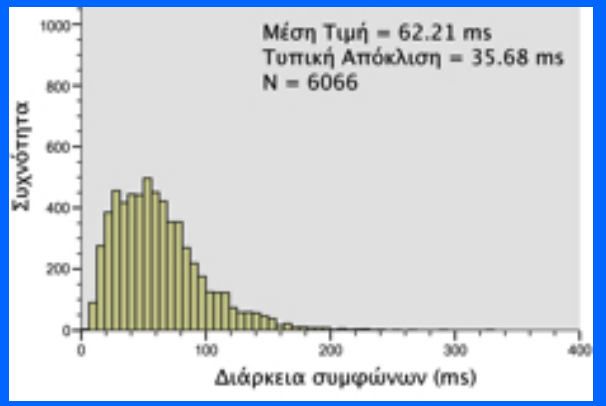
- **The binary hypothesis test for phonemic segmentation requires small windows.**

- **Phone durations: are short as 10-20ms**
  BIC underperforms for few observations
  SOLUTION: use criteria corrected for small samples.

Histograms of phone (vowel, consonant) durations







BIC Criticism

- BIC+C also approximates 2 ln BF, for any constant C
- While the BIC target model does not depend on the sample size *n*, the parameters, which can be reliably estimated do depend on *n*
- *n* can be replaced by:
  - the number of observed statistics per parameter;
  - the rate at which the Hessian matrix of the log-likelihood grows

Log-likelihood in BIC (1)

$$BIC(\mathcal{M}_k) = -2 \ln L(X|\mathcal{M}_k) + P \ln n$$

$$\begin{aligned} P(X|\mathcal{M}_k) &= \int_{\Theta_k} p(X|\boldsymbol{\theta}_k, \mathcal{M}_k)\, p(\boldsymbol{\theta}_k|\mathcal{M}_k) d\,\boldsymbol{\theta}_k \\ &\approx (2\pi)^{P/2} |\tilde{\mathbf{I}}_0(\tilde{\boldsymbol{\theta}}_k)|^{-P/2} P(X|\tilde{\boldsymbol{\theta}}_k, \mathcal{M}_k) P(\tilde{\boldsymbol{\theta}}_k|\mathcal{M}_k)|_{\tilde{\boldsymbol{\theta}}_k = \hat{\boldsymbol{\theta}}_k} \end{aligned}$$

where

- $p(\boldsymbol{\theta}_k|\mathcal{M}_k)$: prior distribution of model parameters $\boldsymbol{\theta}_k$ given $\mathcal{M}_k$
- $P = \dim(\boldsymbol{\theta}_k)$ number of model parameters
- $\tilde{\boldsymbol{\theta}}_k$ mode of $p(\boldsymbol{\theta}_k|\mathcal{M}_k)$
- $\tilde{\mathbf{I}}_0$ observed Fisher information matrix of the posterior distribution
- $\hat{\boldsymbol{\theta}}_k$ MLE
- n number of observations

Log-likelihood in BIC (2)

$$BIC(\mathcal{M}_k) = -2 \ln L(X|\mathcal{M}_k) + P \ln n$$

$$\begin{aligned} -2 \ln P(X|\mathcal{M}_k) &= -2 \ln P(X|\hat{\boldsymbol{\theta}}_k, \mathcal{M}_k) - 2 \ln P(\hat{\boldsymbol{\theta}}|\mathcal{M}_k) \\ &\quad -P \ln 2\pi + P \ln n + \ln |\bar{\mathbf{I}}_E(\hat{\boldsymbol{\theta}}_k)| + \mathcal{O}(n^{-1/2}) \end{aligned}$$

where

- $\bar{\mathbf{I}}_E$ expected information matrix per observation with ij element

$$\left(\bar{\mathbf{I}}_E(\hat{\boldsymbol{\theta}}_k)\right)_{ij} = -\frac{1}{n} E\left\{ \frac{\partial^2 \ell(X|\boldsymbol{\theta}_k, \mathcal{M}_k)}{\partial(\boldsymbol{\theta}_k)_i \partial(\boldsymbol{\theta}_k)_j} \right\} |_{\boldsymbol{\theta}_k = \hat{\boldsymbol{\theta}}_k}$$

BIC corrected for small-samples (BICC)

$$BIC(\mathcal{M}_k) = -2 \ln L(X|\mathcal{M}_k) + P \ln n = -2\ell(\hat{\boldsymbol{\theta}}_k) + P \ln n$$

$$BICC(\mathcal{M}_k) = -2\ell(\hat{\boldsymbol{\theta}}_k) + P\frac{n \ \ln n}{n - P - 1}$$

- Tremblay and Wallach (2004)
- DISTBICC-Γ

Bollen's approximation of BF (ABF-2)

$$BIC(\mathcal{M}_k) = -2\ln L(X|\mathcal{M}_k) + P\ln n = -2\ell(\hat{\boldsymbol{\theta}}_k) + P\ln n$$

$$ABF2 = \begin{cases} -2\ell(\hat{\boldsymbol{\theta}}_k) + P(1 + \ln \frac{P}{\hat{\boldsymbol{\theta}}_k^T \bar{\mathbf{I}}_E(\hat{\boldsymbol{\theta}}_k)\hat{\boldsymbol{\theta}}_k}) & \text{if } P > \hat{\boldsymbol{\theta}}_k^T \bar{\mathbf{I}}_E(\hat{\boldsymbol{\theta}}_k)\hat{\boldsymbol{\theta}}_k \\ -2\ell(\hat{\boldsymbol{\theta}}_k) - \hat{\boldsymbol{\theta}}_k^T \bar{\mathbf{I}}_E(\hat{\boldsymbol{\theta}}_k)\hat{\boldsymbol{\theta}}_k & \text{otherwise} \end{cases}$$

- DISTABF2-Γ

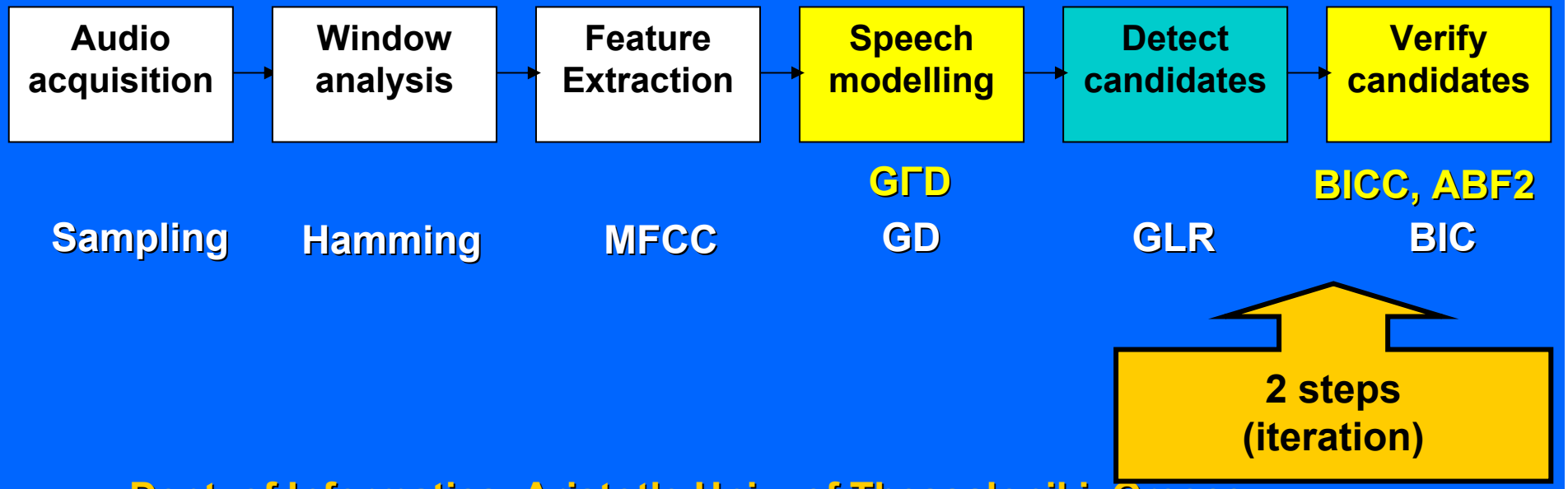
Bollen's approximation of BF (ABF-2) for GΓD

$$\bar{\mathbf{I}}_E(\hat{\boldsymbol{\theta}}_k) = n \begin{bmatrix} \psi_1(a) & -\frac{1}{b} & -\frac{\psi_0(a)+\ln b}{c} \\ -\frac{1}{b} & \frac{a}{b^2} & \frac{a\psi_0(a)-a\ln b+1}{bc} \\ -\frac{\psi_0(a)+\ln b}{c} & \frac{a\psi_0(a)-a\ln b+1}{bc} & \frac{\eta}{c^2} \end{bmatrix}$$

$$\begin{aligned} \eta &= a\psi_0^2(a) - 2(a\ln b - 1)\psi_0(a) + a\psi_1(a) \\ &\quad -2\ln b + a\ln^2 b + 1 \\ \psi_0(x) &= \frac{d}{dx}\ln\Gamma(x) \\ \psi_1(x) &= \frac{d^2}{dx^2}\ln\Gamma(x) \\ \hat{\boldsymbol{\theta}}_k^T\bar{\mathbf{I}}_E(\hat{\boldsymbol{\theta}}_k)\hat{\boldsymbol{\theta}}_k &= \hat{a}\psi_0(\hat{a})^2 - 2(\hat{a}\ln\hat{b} - 1)\psi_0(\hat{a}) + (\hat{a}^2+\hat{a})\psi_1(\hat{a}) \\ &\quad +\hat{a}\ln^2\hat{b} - \hat{a} - 2\ln\hat{b} + 3 \end{aligned}$$

Revised DISTBIC

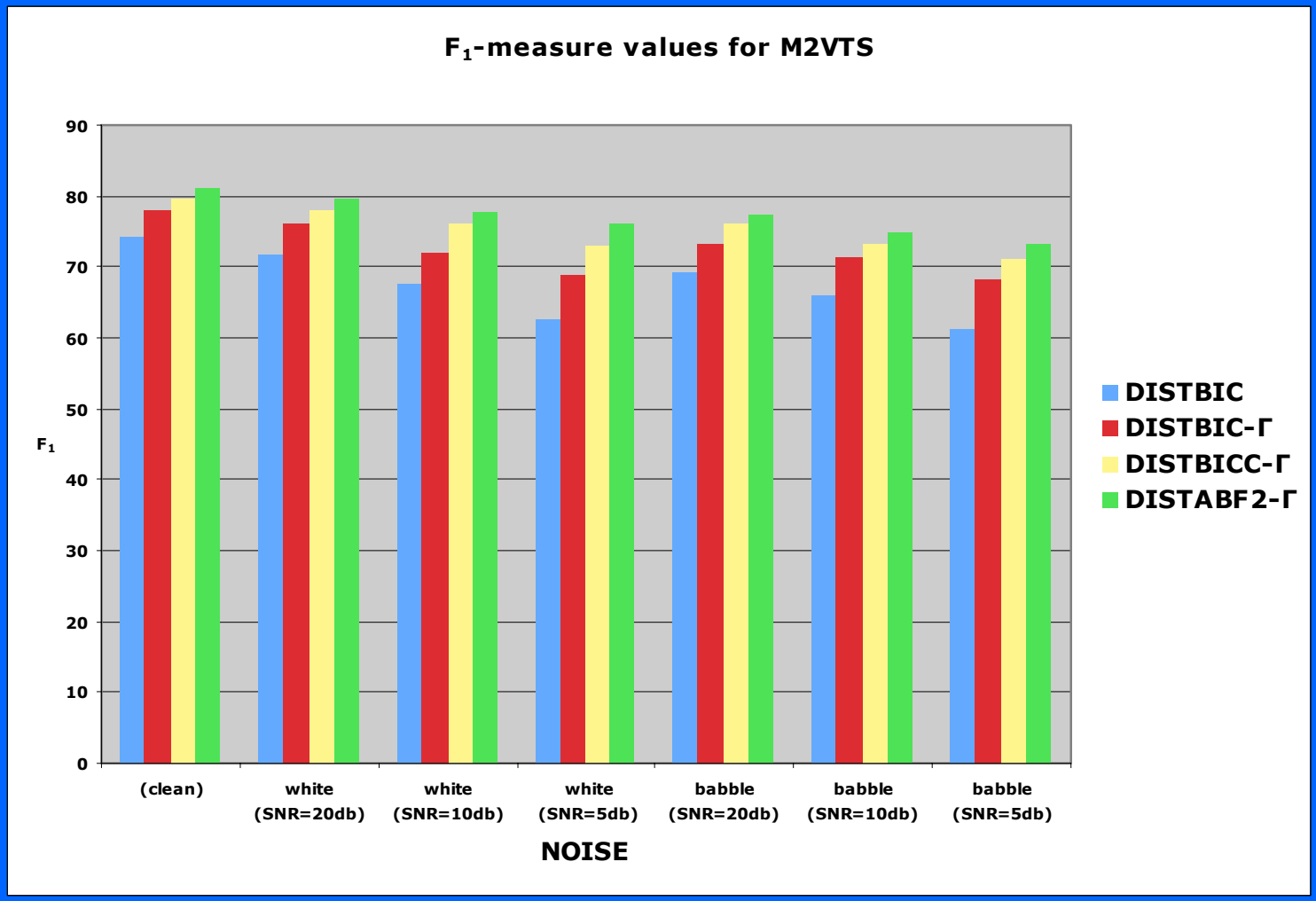
- DISTBIC-Γ = DISTBIC with GΓD priors
- DISTBICC-Γ= BICC instead of BIC, GΓD instead of GD
- DISTABF2-Γ= ABF2 instead of BIC, GΓD instead of GD



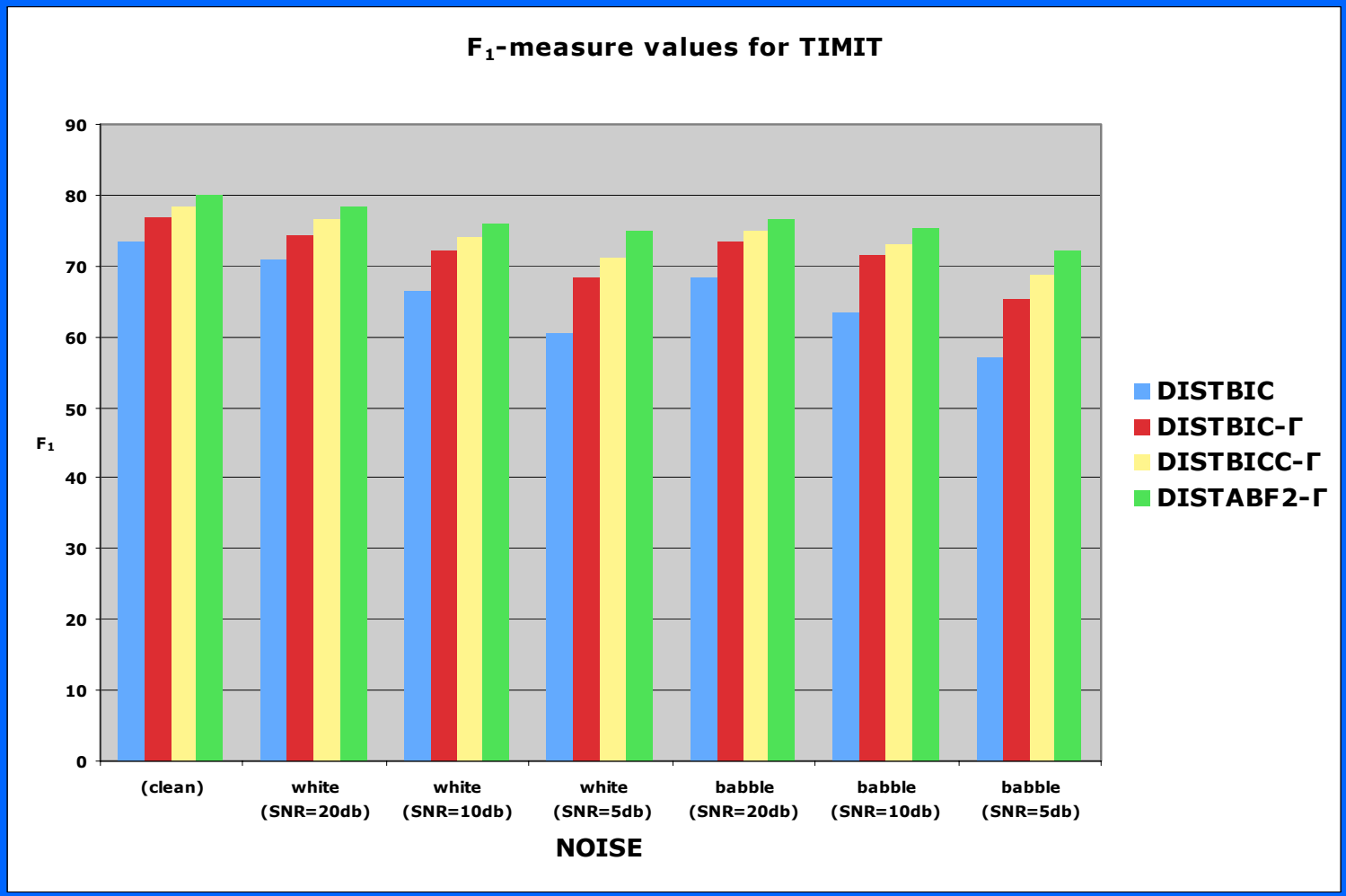
Experiments

- Datasets: M2VTS and TIMIT. Pre-existing hand-labeling.
- Compare DISTBIC-Γ, DISTBICC-Γ, DISTABF2-Γ against DISTBIC.
- Phone boundaries detection. 20ms tolerance (human error).
- Additive noise (white, babble) at various SNRs (20dB, 10dB, 5dB). NOISEX-92 database.

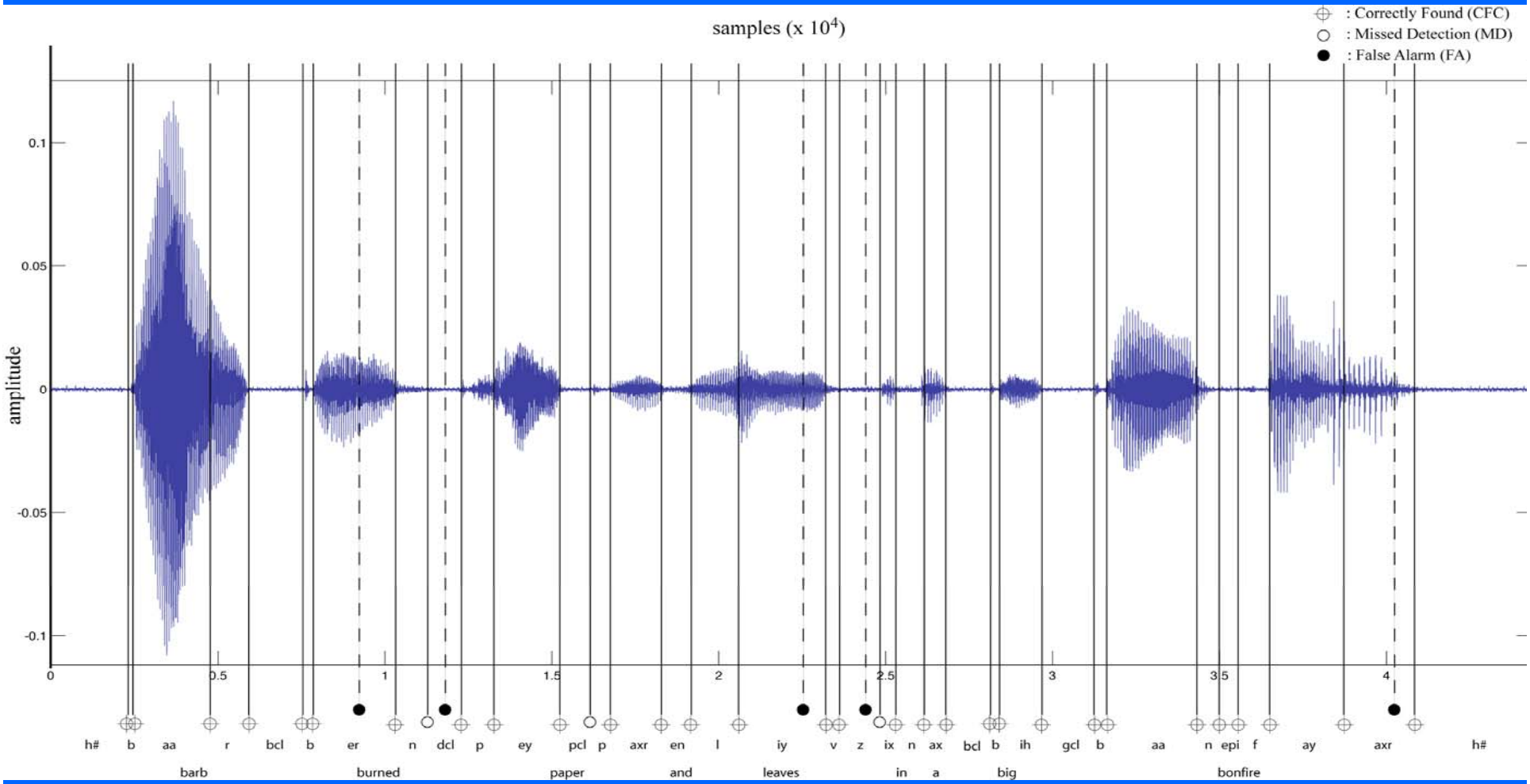
Results (M2VTS)



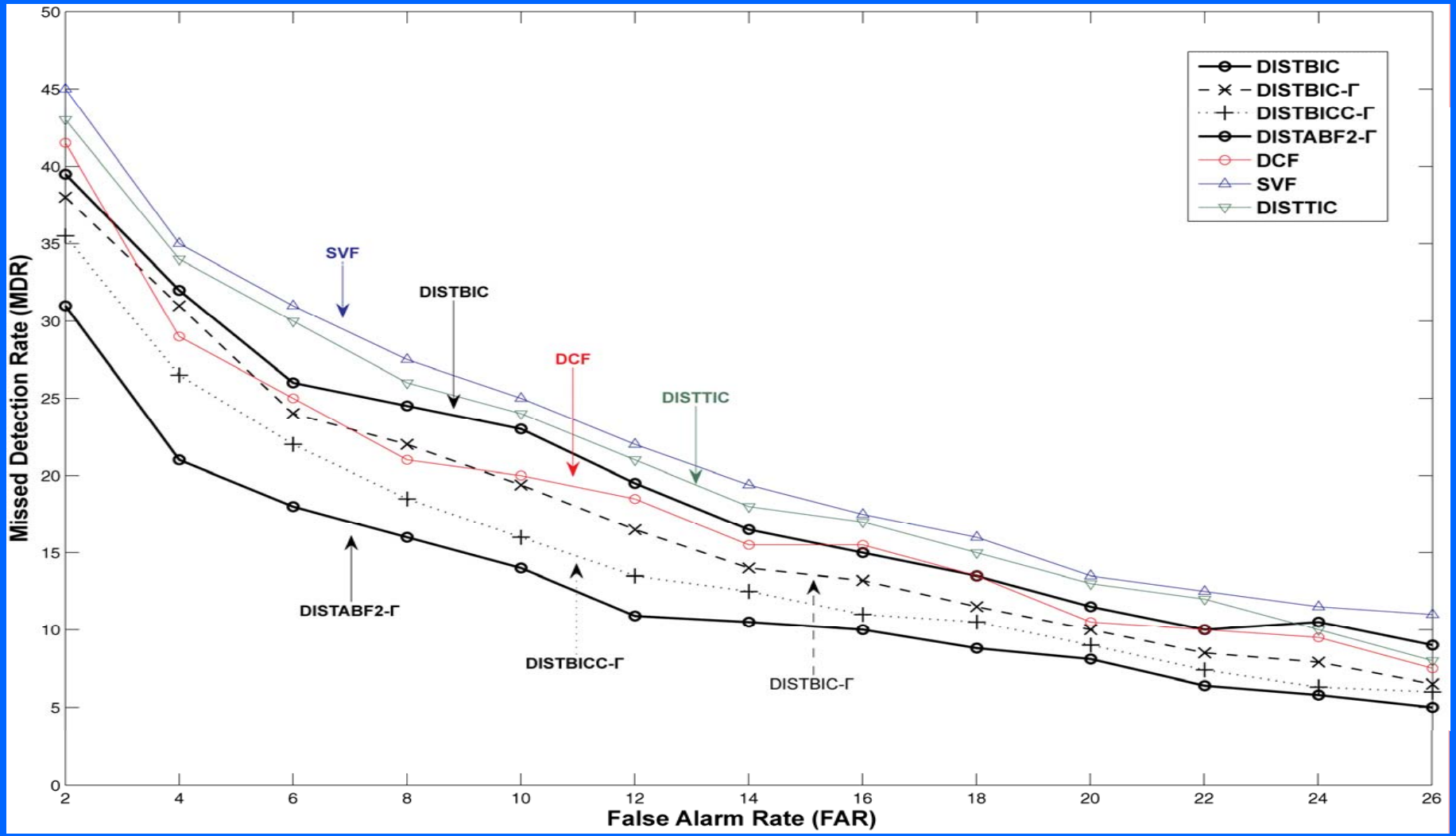
Results (TIMIT)



Results (NTIMIT)



DET curve (NTIMIT)



AICC

$$AIC(\mathcal{M}) = -2\ln L(X;\mathcal{M}) + 2P$$

small-sample corrected AIC (AICC)

$$AICC(\mathcal{M}) = -2\ln L(X;\mathcal{M}) + 2\frac{Pn}{n-P-1}$$

Hurvich and Tsai (1989); Cavanaugh (1997);

Bozdogan's Information Complexity Criterion (ICOMP)

$$ICOMP(IFIM) = \underbrace{-2\ell(X;\mathcal{M})}_{-2\ell(\hat{\boldsymbol{\theta}})} + P\ln\left(\frac{\mathrm{tr}(\mathbf{F}^{-1}(\hat{\boldsymbol{\theta}}))}{P}\right) - \ln|\mathbf{F}^{-1}(\hat{\boldsymbol{\theta}})|$$

Penalizes ellipsoidal dispersion

where $\mathbf{F}$: Fisher Information matrix (FIM) and IFIM: inverse FIM

- Penalizes the interdependence between the parameters by estimating the covariance complexity of the model
- Offers a judicious balance between GoF, model complexity, and accuracy of the parameter estimates
- IFIM can be approximated by a Monte Carlo resampling method (Spall) 2004.

PROBLEM (4)

- **Robust estimates of the covariance matrix?**
  SOLUTION: Use an M-estimator; The minimum covariance determinant estimator (MCD)

$$\hat{\boldsymbol{\mu}}_{\mathrm{MCD}} = \sum_{j \in J^*} t_j \, \mathbf{x}_j$$

$$\hat{\boldsymbol{\Sigma}}_{\mathrm{MCD}} = \sum_{j \in J^*} t_j \, (\mathbf{x}_j - \hat{\boldsymbol{\mu}}_{\mathrm{MCD}})(\mathbf{x}_j - \hat{\boldsymbol{\mu}}_{\mathrm{MCD}}))^T$$

where J^* is the r-element set, where the determinant of $\hat{\boldsymbol{\Sigma}}_{\mathrm{MCD}}$ is minimized; $t_j = 1/r$ and $r \approx n/2$.

Fast MCD: Rouseeuw and Van Driessen (1999)

PROBLEM (5)

- **Robust versions of AIC and BIC**

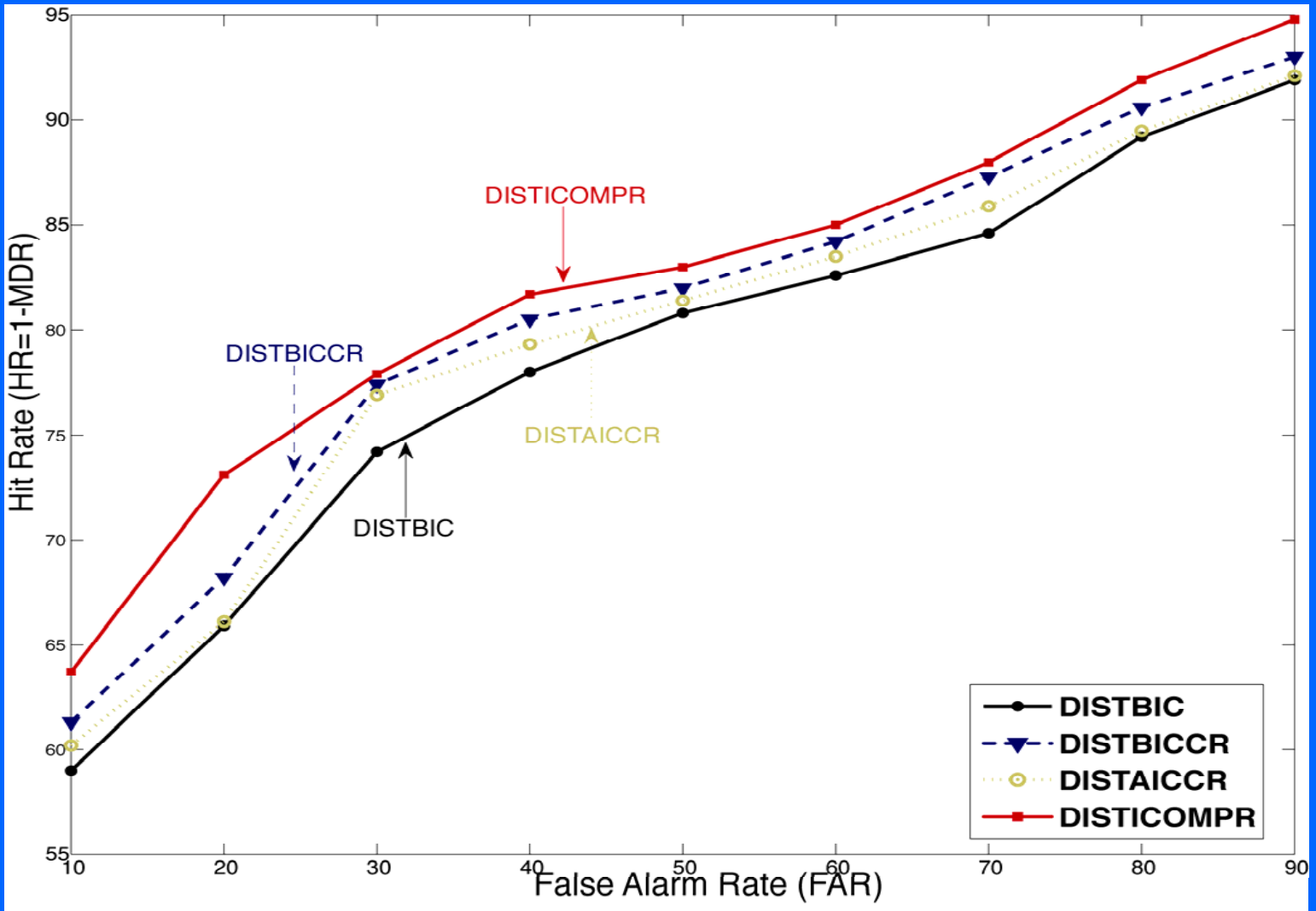
See discussion Almpanidis et al. (2009)

Experimental results (NTIMIT)

Table 4: Average phonemic segmentation in NTIMIT

	PRC	RCL	F_1	FAR	MDR
DISTBIC	0.659	0.718	0.687	0.271	0.282
DISTBICC	0.681	0.726	0.702	0.254	0.274
DISTBICR	0.710	0.743	0.726	0.233	0.257
DISTAIC	0.658	0.706	0.618	0.268	0.294
DISTAICCR	0.685	0.709	0.696	0.246	0.291
DISTICOMP	0.727	0.745	0.736	0.218	0.255
DISTICOMPR	0.739	0.754	0.747	0.210	0.246
SVF	0.647	0.704	0.677	0.277	0.296
DCF	0.672	0.733	0.701	0.264	0.267

ROC curves (NTIMIT)



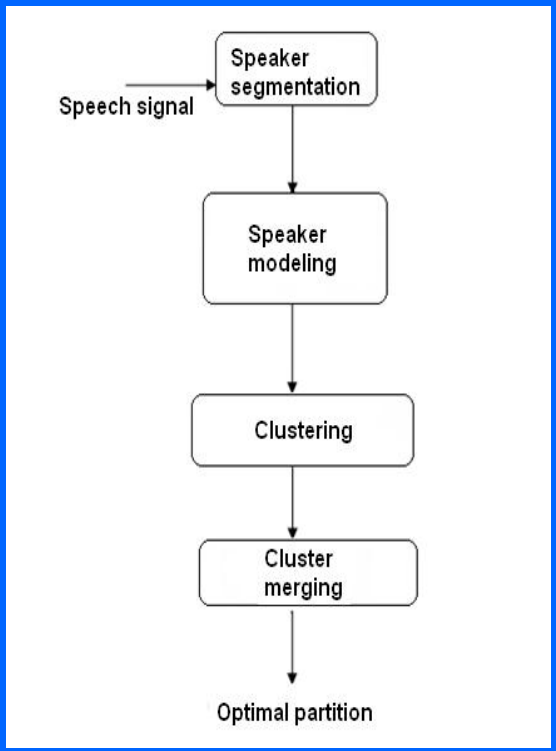
Average F_1 rates for phoneme transitions

TABLE III
AVERAGE F_1 RATES FOR THE TOP MOST FREQUENT PHONEME CLASS TRANSITIONS

		Percent of occurrences	F_1 (DISTBIC)	F_1 (DISTICOMPR)	F_1 (DIST)
1	silence -> stop	11.8	64.04	70.10	62.67
2	vowel -> silence	9.1	71.53	78.16	67.45
3	semivowel&glide -> vowel	8.9	65.48	71.11	65.08
4	stop -> vowel	8.4	69.44	77.75	67.72
5	vowel -> fricative	8.1	71.01	75.65	69.72
6	fricative -> vowel	7.5	70.76	77.02	67.78
7	vowel -> nasal	7.1	70.57	74.90	67.38
8	vowel -> semivowel&glide	4.8	63.37	68.77	63.22
9	fricative -> silence	4.0	67.16	72.74	64.44
10	nasal -> vowel	3.4	67.15	74.39	65.49

Speaker Diarization

- Speaker diarization: Speaker segmentation followed by speaker clustering.



Let $\mathcal{N}_i(\mu_i, \Sigma_i)$, $i = 1, 2$ be a multivariate Gaussian distribution with mean vector μ_i and covariance matrix Σ_i modeling each class. Let also N_i, $i = 1, 2$ be the number of feature vectors assigned to each class.

$$d_{COVMEAN} = d_{COV} \cdot d_{MEAN}$$

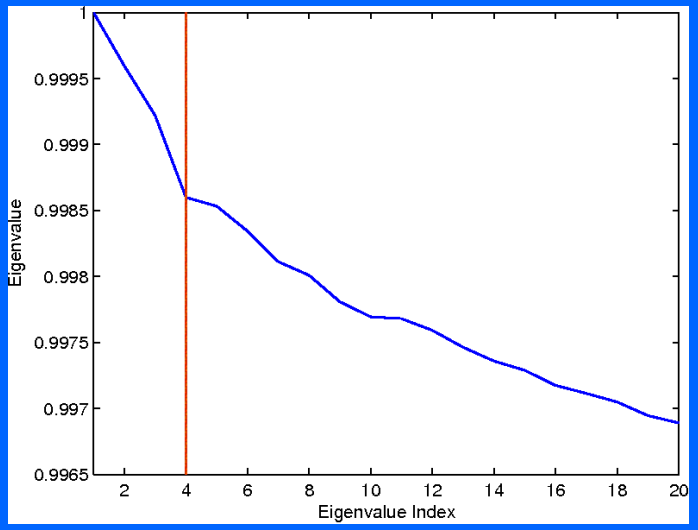
where

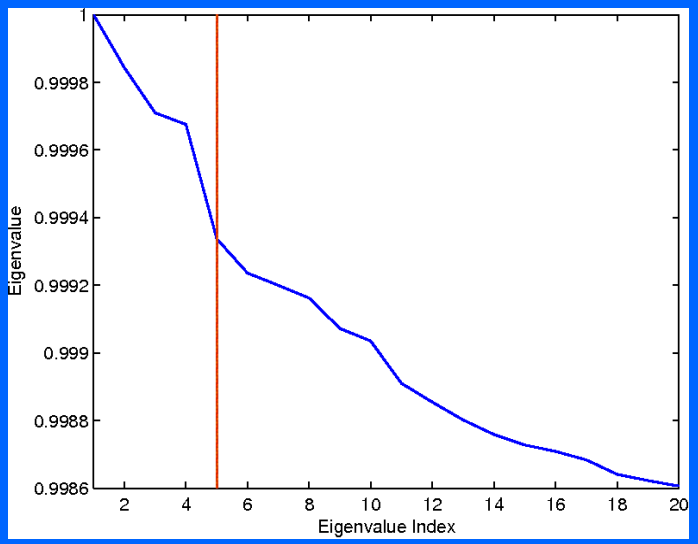
$$d_{COV} = \left(\frac{|\Sigma_1|^a \, |\Sigma_2|^{1-a}}{|\mathbf{W}|} \right)^{\frac{N_T}{2}}$$

$$d_{MEAN} = \left(1 + \frac{N_1 N_2}{N_T^2} (\mu_1 - \mu_2)^T \mathbf{W}^{-1} (\mu_1 - \mu_2) \right)^{-\frac{N_T}{2}}$$

with $N_T = N_1 + N_2$, $a = \frac{N_1}{N_T}$, and $\mathbf{W} = a\Sigma_1 + (1-a)\Sigma_2$.

Spectral Graph Theory





Let $\mathbf{A}$ be the $N \times N$ adjacency matrix with elements equal to the distance between two speech segments $d_{COVMEAN}$. The un-normalized Laplacian of the graph is given by $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the $N \times N$ diagonal matrix with $D(i,i) = \sum_j A(i,j)$. That is, $D(i,i)$ is the sum of the weights of the edges that are incident to vertex i. The normalized Laplacian of the graph can be derived as

$$\mathbf{L}_{norm} = \mathbf{D}^{-\frac{1}{2}} \mathbf{L} \mathbf{D}^{-\frac{1}{2}}.$$

In the ideal case of N_c completely disconnected clusters, the eigenvalue 0 has multiplicity N_c, and then there is a gap to the (N_c+1)th eigenvalue. Accordingly, the most stable clustering is generally obtained for k that maximizes the difference between two consecutive eigenvalues $(\beta_k - \beta_{k-1})$.

The top 20 eigenvalues of $\mathbf{I}-\mathbf{L}norm$ for two sound files. The vertical line indicates the actual number of speakers, which coincides with the drastic drop in the magnitude of the eigenvalues.

Performance evaluation (1)

Let n_{ij} be the total number of audio segments in cluster i uttered by actor j; N_a be the total number of actors; N_c be the total number of clusters; N be the total number of audio segments; $n_{.j}$ be the total number of audio segments uttered by actor j; $n_{i.}$ be the total number of audio segments in cluster i.

The error for cluster i, CE_i, is defined as the percentage of the total time spoken by actor whose speech segments appear in majority in cluster i, that has not been clustered to this cluster. The average classification error, ace, is defined as

$$ace = \frac{1}{N_c} \sum_{i=1}^{N_c} CE_i.$$

The average cluster purity provides a measure of how well a cluster is limited to only one speaker:

$$acp = \frac{1}{N} \sum_{i=1}^{N_c} \pi_{i.}\ n_{i.} \quad \text{where } \pi_{i.} = \textstyle\sum_{j=1}^{N_a} n_{ij}^2 / n_{i.}^2.$$

The average speaker purity provides a measure of how well a speaker is limited to only one cluster:

$$asp = \frac{1}{N} \sum_{j=1}^{N_a} \pi_{.j}\ n_{.j} \quad \text{where } \pi_{.j} = \textstyle\sum_{i=1}^{N_c} n_{ij}^2 / n_{.j}^2.$$

Diarization error rate:

$$derr = \frac{T_{FA} + T_{MS} + T_{wrong}}{T_{total}}$$

where T_{FA} is the total duration of the non-speech segments that were classified as speech, T_{MS} is the total duration of the speech segments that were classified as either non-speech or silence, T_{wrong} is the total duration of speech segments that were correctly classified as speech, but that were clustered into wrong speaker groups, and T_{total} is the total duration of all the speech segments.

Performance evaluation (2)

Movie Dialogue Database:

Figure of merit	**Proposed system**		**Baseline system**	
	Outliers excluded	Outliers included	Outliers excluded	Outliers included
N_c	2	2	3	3
$ace(\%)$	(**19.69, 12.82**)	(19.71, 11.50)	(25.30, 12.59)	(28.57, 12.33)
acp	(**0.74, 0.14**)	(0.73, 0.13)	(0.67, 0.11)	(0.64, 0.08)
asp	(0.70, 0.17)	(0.69, 0.10)	(**0.77, 0.22**)	(0.66, 0.27)

MDE RT-03 Training Data Speech Corpus subset:

Figure of merit	Proposed system	Baseline system
$d_{err}(\%)$	**20.2**	28.3
acp	**0.81**	0.72
asp	**0.59**	0.52

Conclusions - Discussion

- Modelling the speaker utterance and selecting features within BIC improves speech segmentation.
- The Generalized Gamma model is a more adequate model for the noisy speech than the Gaussian one. Offline DISTBIC-Γ = DISTBIC + GΓD offers better discriminative ability for phone segmentation.
- ABF2 and BICC yield better results than BIC for phone segmentation. ICOMP and its robust variant ICOMP-R provide the best results for phone segmentation.
- Spectral graph theory improves speaker diarization.

References

- G. Almpanidis and C. Kotropoulos, "Phonemic segmentation using the generalised Gamma distribution and small-sample Bayesian information criterion," *Speech Communication*, vol. 50, no. 1, pp. 38-55, January 2008.
- M. Kotti, V. Moschou, and C. Kotropoulos, "Speaker segmentation and clustering," *Signal Processing*, vol. 88, no. 5, pp. 1091-1124, May 2008.
- M. Kotti, E. Benetos, and C. Kotropoulos, "Computationally efficient and robust BIC-based speaker segmentation," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 16, no. 5, pp. 920-933, July 2008.
- G. Almpanidis, M. Kotti, and C. Kotropoulos, "Robust detection of phone segments in continuous speech using model selection criteria with few observations," *IEEE Trans. Audio, Speech, and Language Processing*, vol. 17, no. 2, pp. 287-298, February 2009.
- N. Bassiou, V. Moschou, and C. Kotropoulos, "Speaker diarization exploiting the eigengap criterion and cluster ensembles," *IEEE Trans. Audio, Speech, and Language Processing*, to appear 2010.