Delayed Decision CELP Speech Coding using Squared and Perceptual Error Criteria

Raquel Fandos Marín

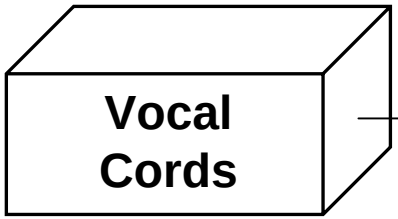
Examiner: Prof. W. Bastiaan Kleijn
Supervisors: Jan Plasberg
Harald Pobloth

Speech Processing Group
Department of Signals, Sensors and Systems (S3)
Royal Institute of Technology (KTH)

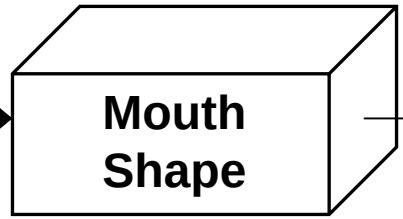
Outline

- The Speech Signal
- The Standard CELP Coder
- A Sophisticated Error Criterion: an Auditory Model
- Delayed Decision CELP Coder
- Evaluation.
- Conclusion and future work

The Speech Signal



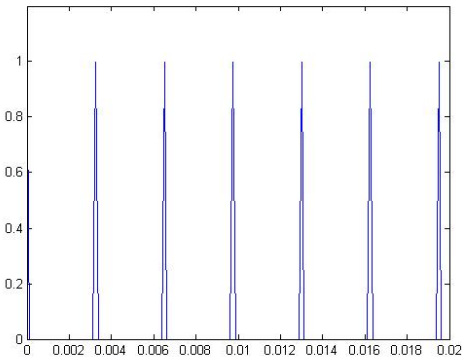
Excitation



H(z)=1/A(z)

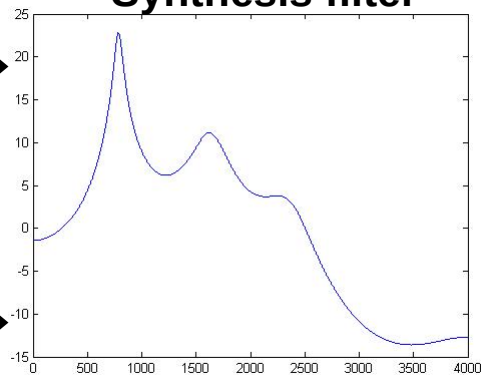
Speech

Voiced excitation



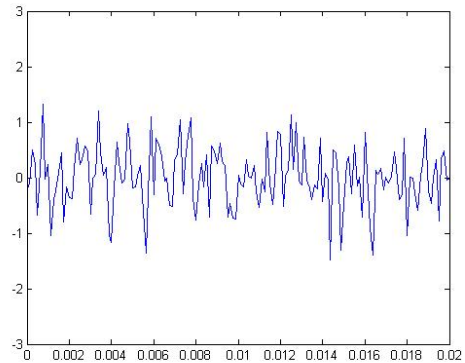
Typically
200Hz for male
400Hz for female

Synthesis filter

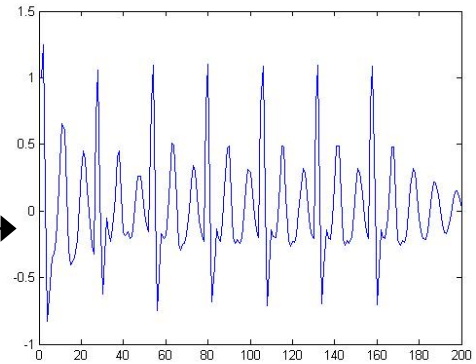


Significant BW= 4KHz
4 poles →
8 Linear Prediction Coefficients

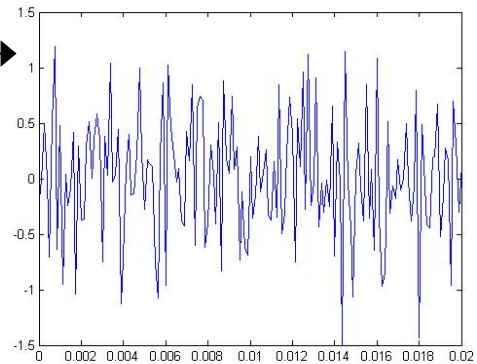
Unvoiced excitation



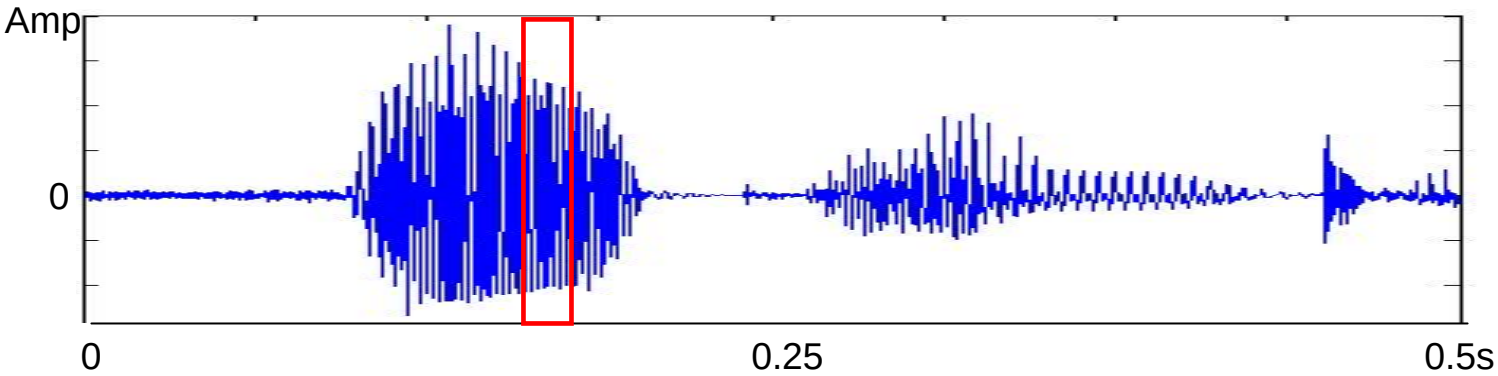
Voiced speech



Unvoiced speech



The Speech Signal



•Most of the information up to 4kHz → Sampling frequency=8kHz

•PCM: 16bits per sample → Transmission rate=128kbps

•Coding using the speech characteristics permits to reduce this rate. I.e. GSM rate=13kbps.

•The synthesis filter is stationary in blocks of 20 ms.

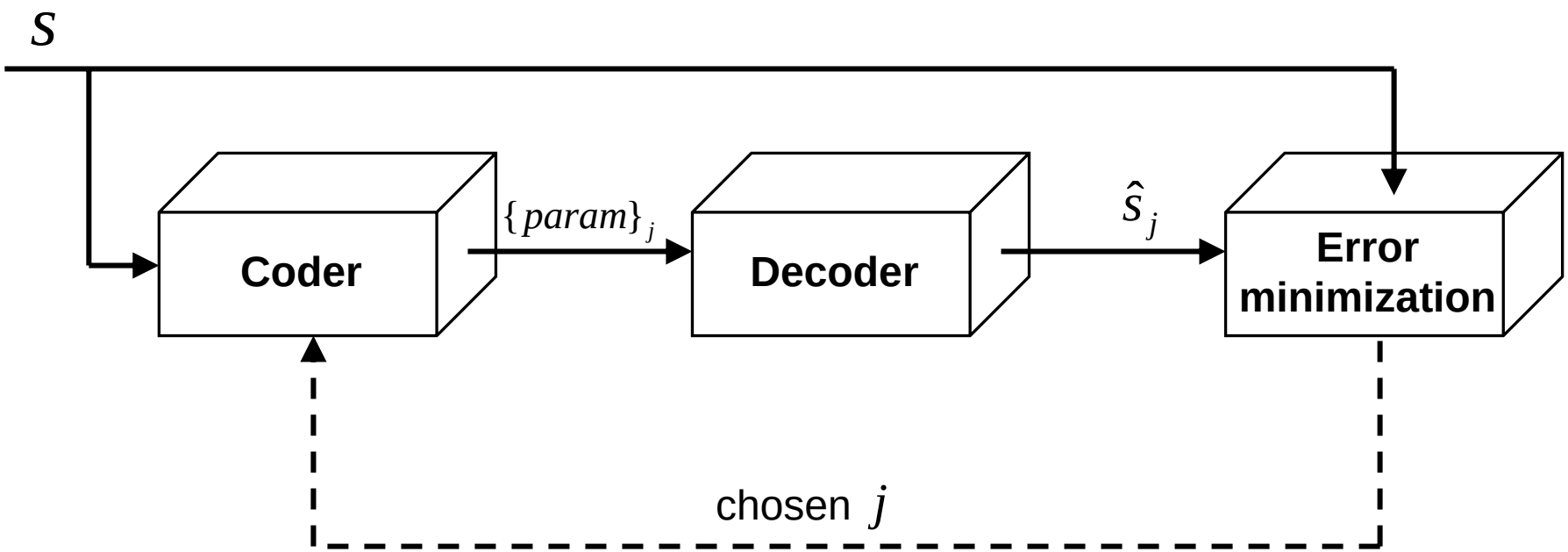
•The excitation is calculated in sub-blocks of 5ms.

Outline

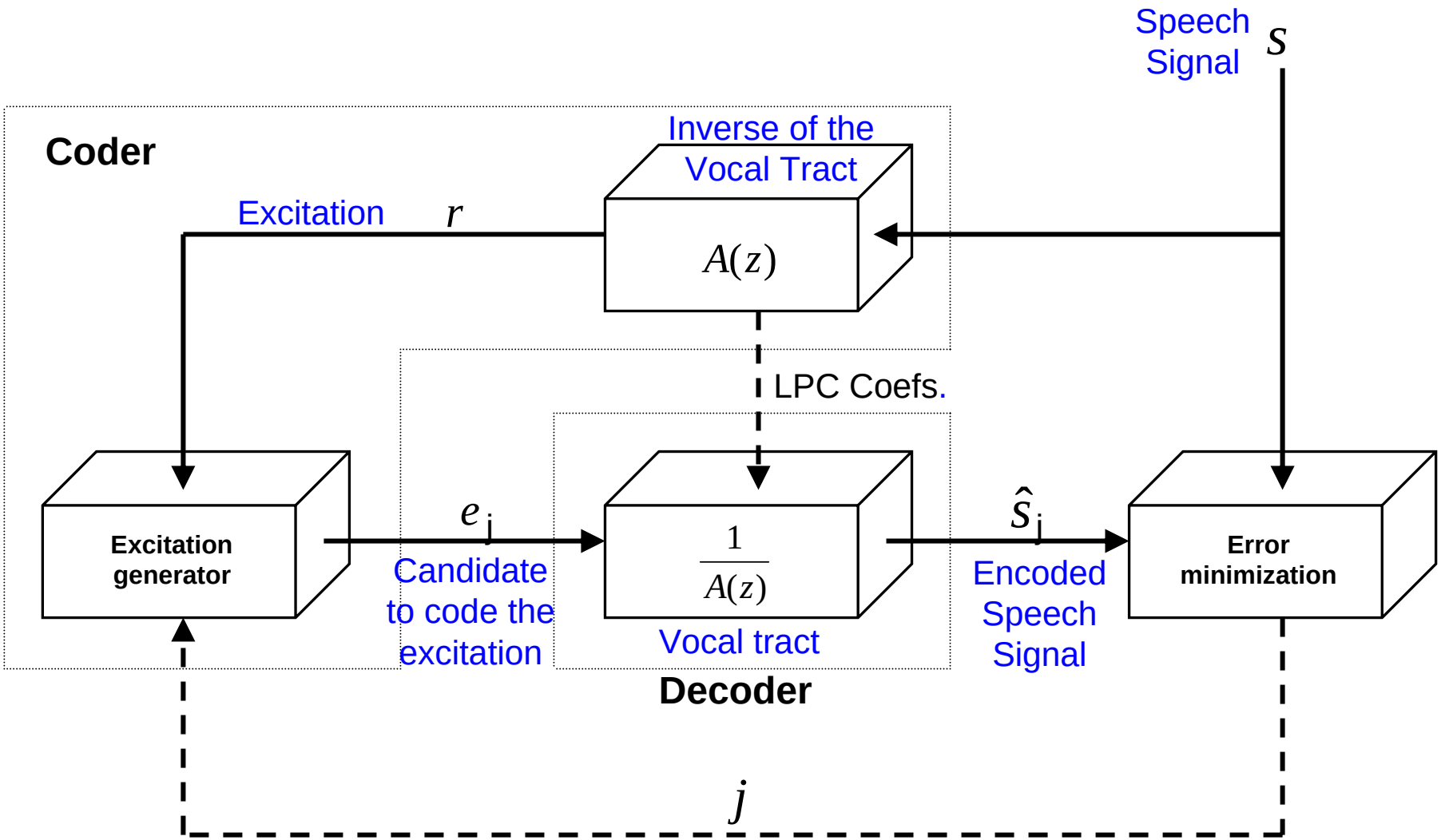
- The Speech Signal
- The Standard CELP Coder
- A Sophisticated Error Criterion: an Auditory Model
- Delayed Decision CELP Coder
- Evaluation.
- Conclusion and future work

The Analysis-by-Synthesis Principle

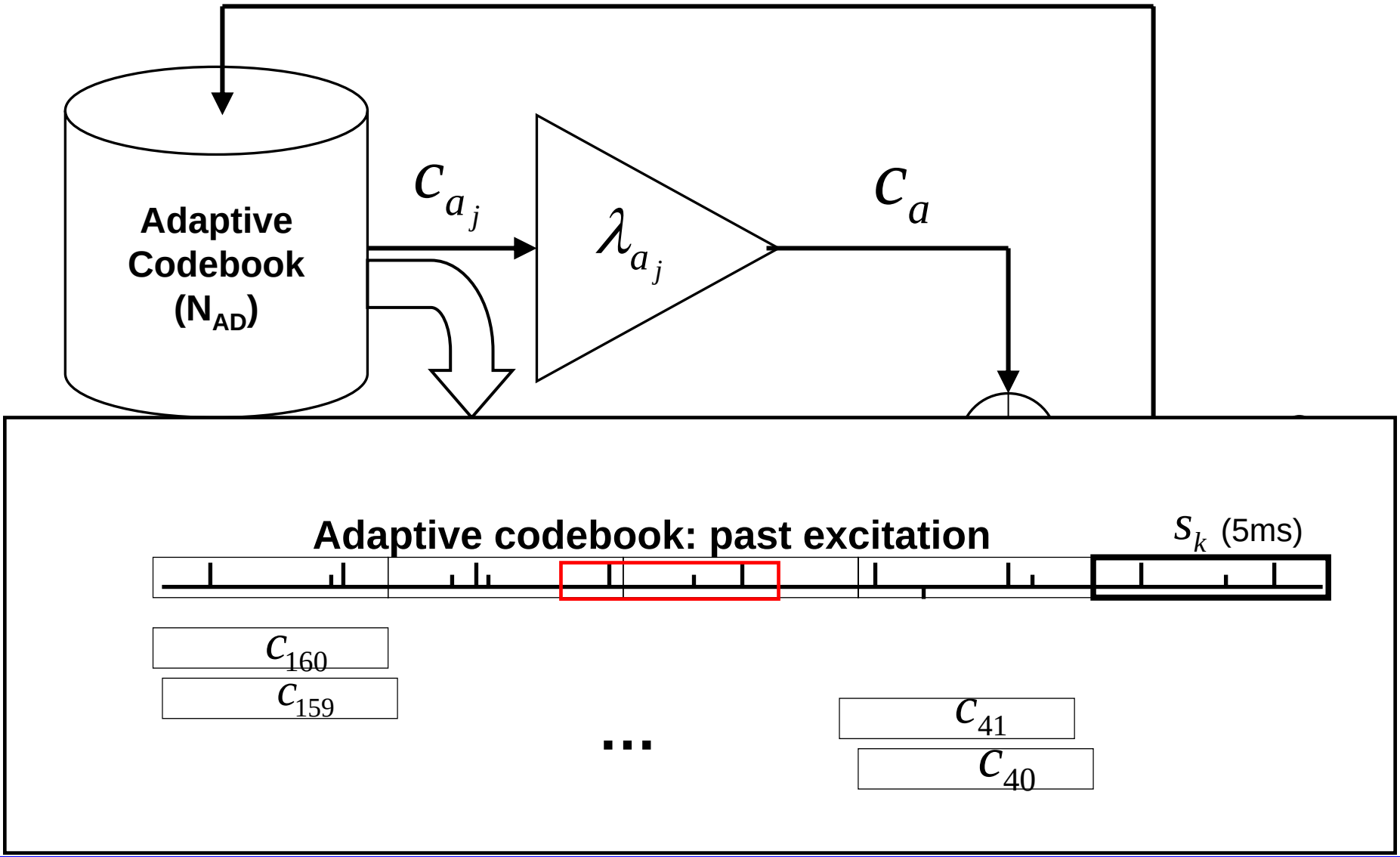
Encoder



An LPAS Coder



Excitation Generation for CELP Coding



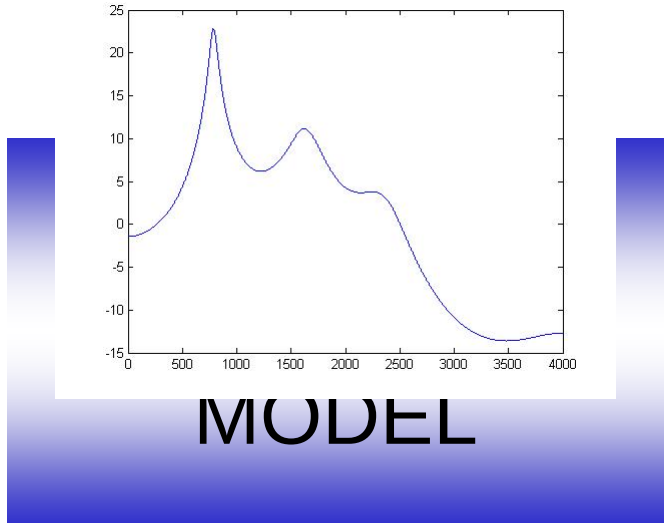
Excitation Generation Inefficiencies

1. c_a and c_f separate searches
2. c_a search priority
3. Separate sub-block processing
   - Adaptive Codebook
   - s_0 (Ringing of the synthesis filter)

DELAYED DECISIONS

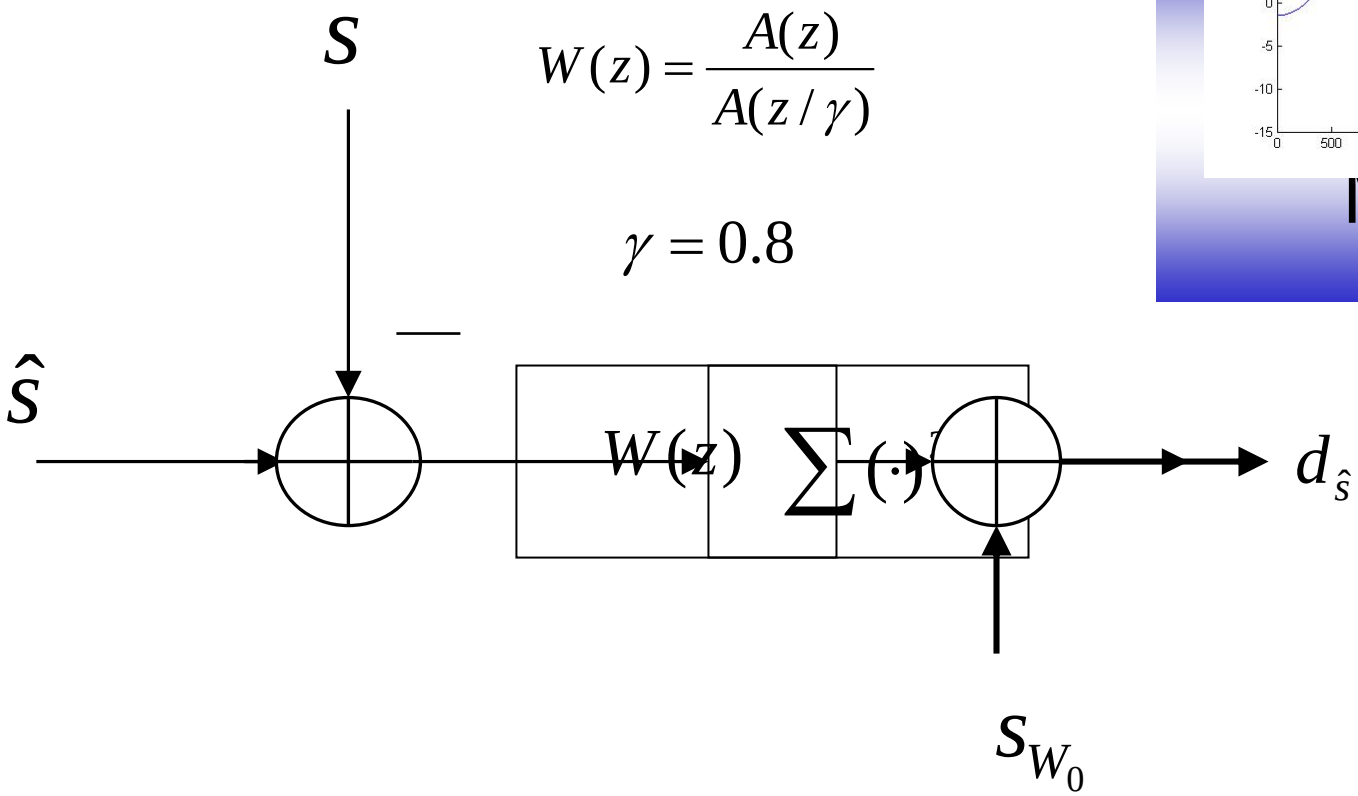
Standard Distortion Criteria

- MSE (Mean-Squared Error)
- WMSE (Weighted Mean-Squared Error)



$$W(z) = \frac{A(z)}{A(z/\gamma)}$$

$$\gamma = 0.8$$

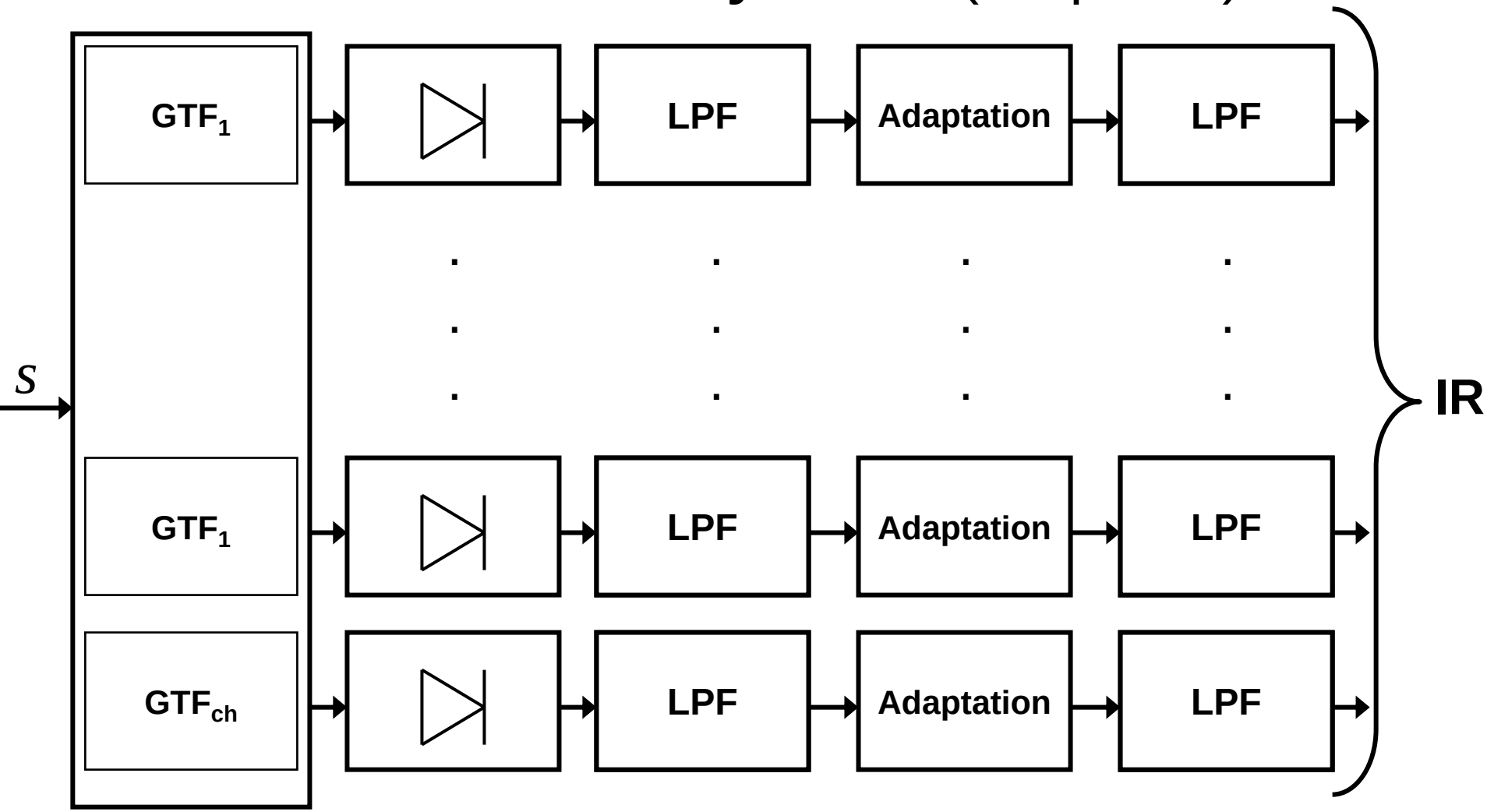


Outline

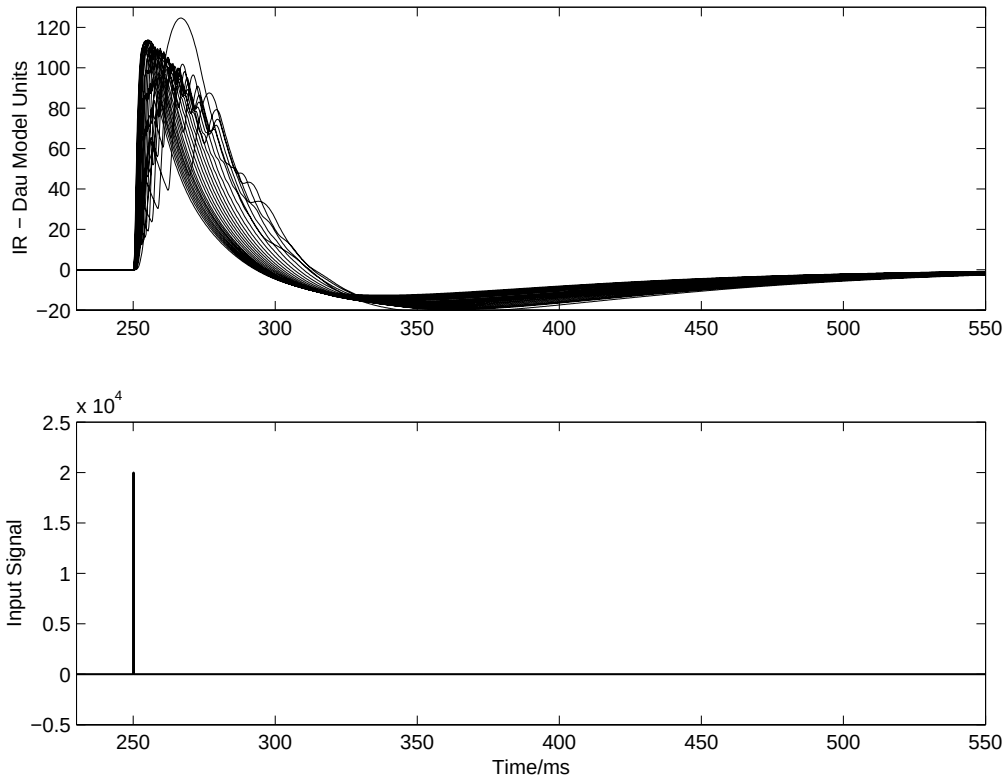
- The Speech Signal
- The Standard CELP Coder
- A Sophisticated Error Criterion: an Auditory Model
- Delayed Decision CELP Coder
- Evaluation.
- Conclusion and future work

A Sophisticated Distortion Measure

the DAU Auditory Model (simplified)



The Support Problem



- Long support
- No single-letter distortion measure

DELAYED DECISIONS

Outline

- The Speech Signal
- The Standard CELP Coder
- A Sophisticated Error Criterion: an Auditory Model
- Delayed Decision CELP Coder
- Evaluation
- Conclusions and future work

Why DDCELP?

- **Excitation Generation Inefficiencies.**
  - c_a and c_f separate searches
  - c_a search priority
  - Separate sub-block processing
    - Adaptive Codebook
    - Ringing of the synthesis filter
    - WMSE: Ringing of the weighting filter
- **Long support of DAU Model Distortion Criterion**

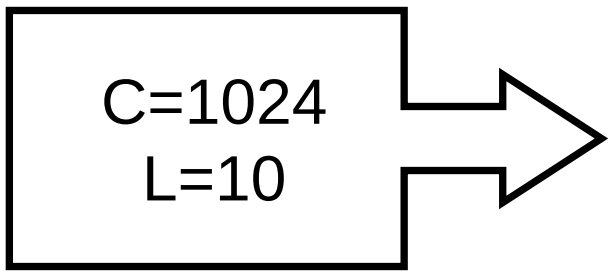
How does it do it?

- Multiple preserved candidates
- Cumulative distortion estimation

DDCELP Limitations

Theoretically optimal but…

- Computational complexity and memory size



$1024^{10} = 1.3*10^{30}$ alternatives !!!!!!

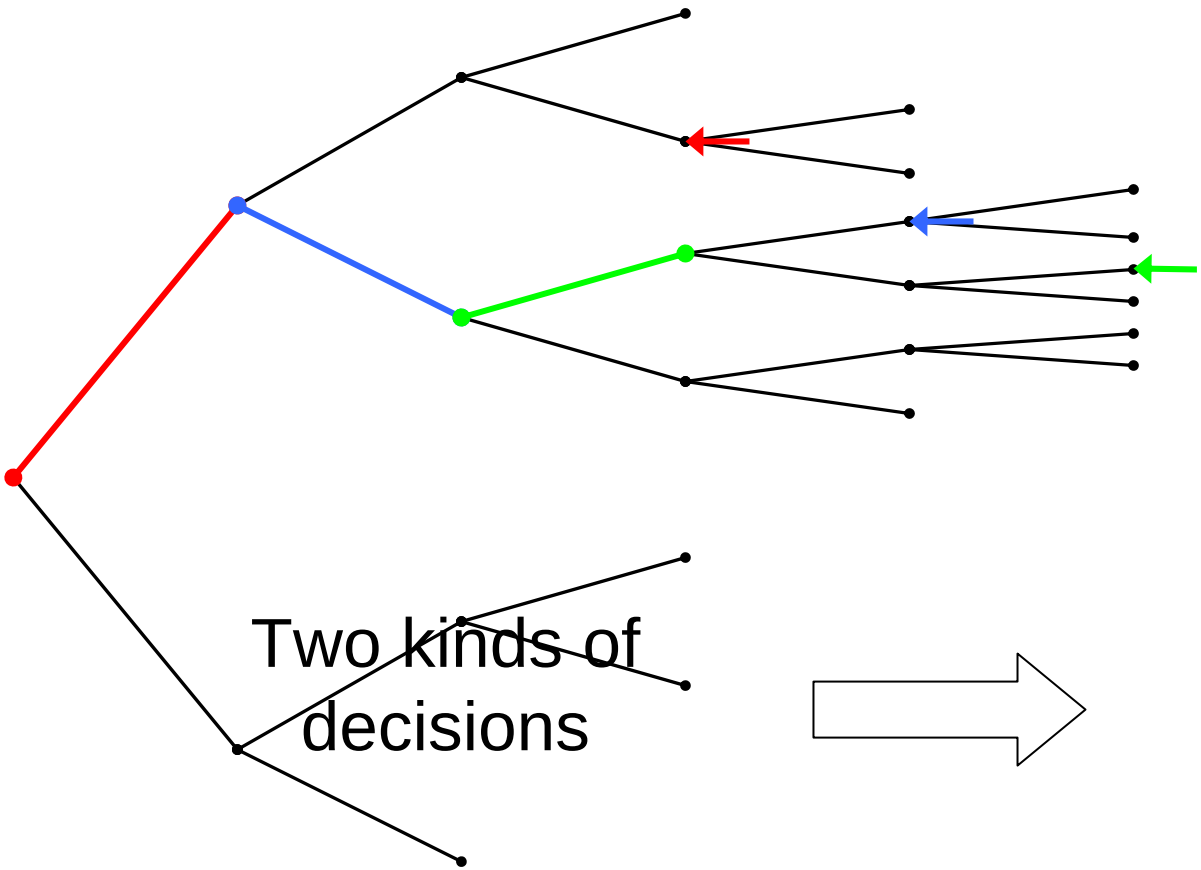
Choose only the M best candidates per sub-block

- Delay constraints for practical applications

Limit the delay L

M-L tree algorithm

The M-L Algorithm



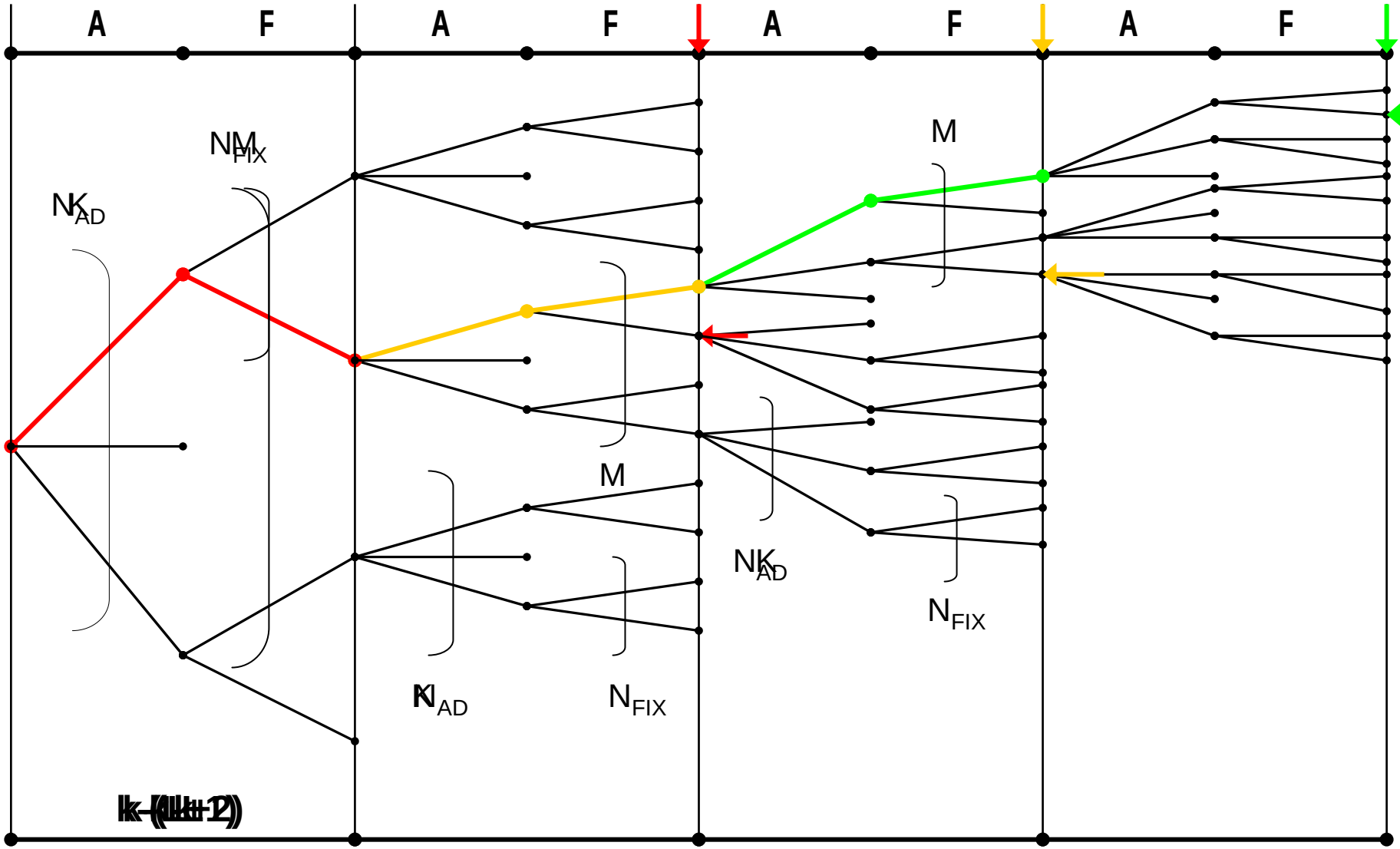
■■■

Two kinds of decisions

M-K-L Algorithm

M: 3, L: 2

The M-K-L Algorithm (N_{AD}=3,N_{FIX}=2,M=3,K=2,L=1)



MSE and WMSE DDCELP Evaluation

- Convergence

$L \uparrow \Rightarrow$ **Weaker dependency**

$M, K \uparrow \Rightarrow$ **Useless candidates**

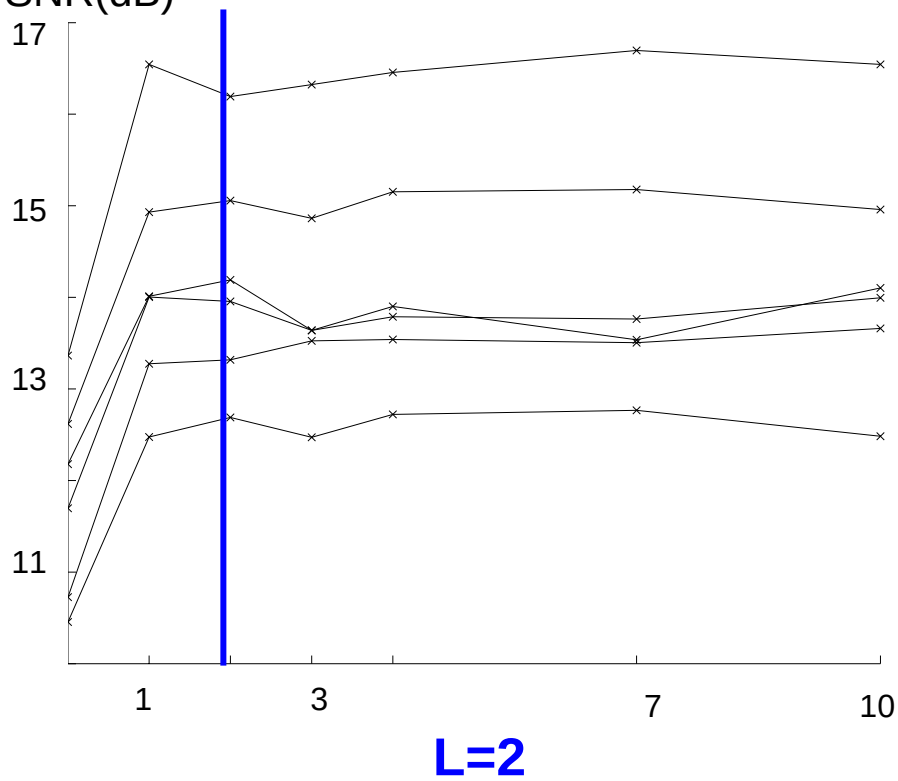
Outline

- The Speech Signal
- The Standard CELP Coder
- A Sophisticated Error Criterion: an Auditory Model
- Delayed Decision CELP Coder
- Evaluation
- Conclusion and future work

MSE and WMSE DDCELP Evaluation

MSE. Dependence with L.

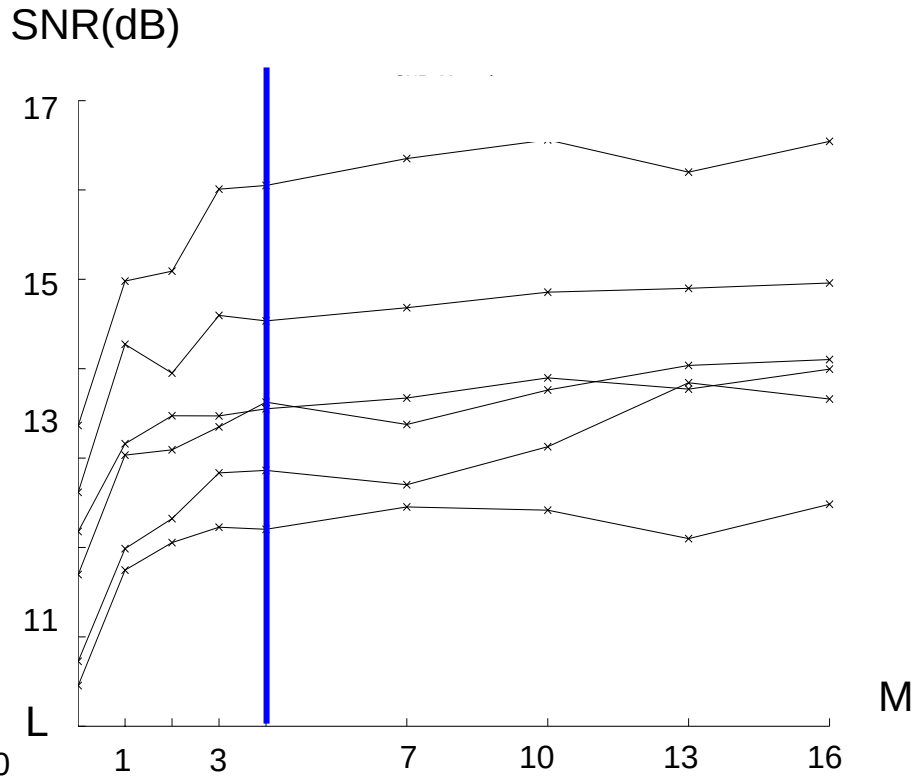
M=K=16



L=2

MSE. Dependence with M.

L=10, K=16

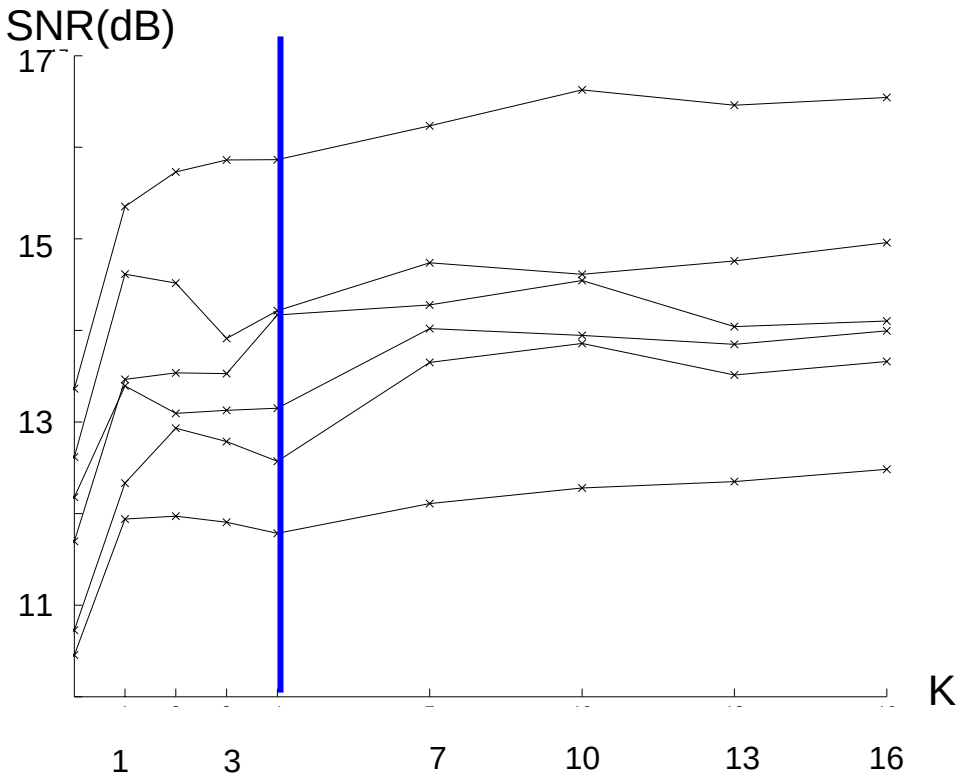


M=4

MSE and WMSE DDCELP Evaluation

MSE. Dependence with K.

L=10, M=16



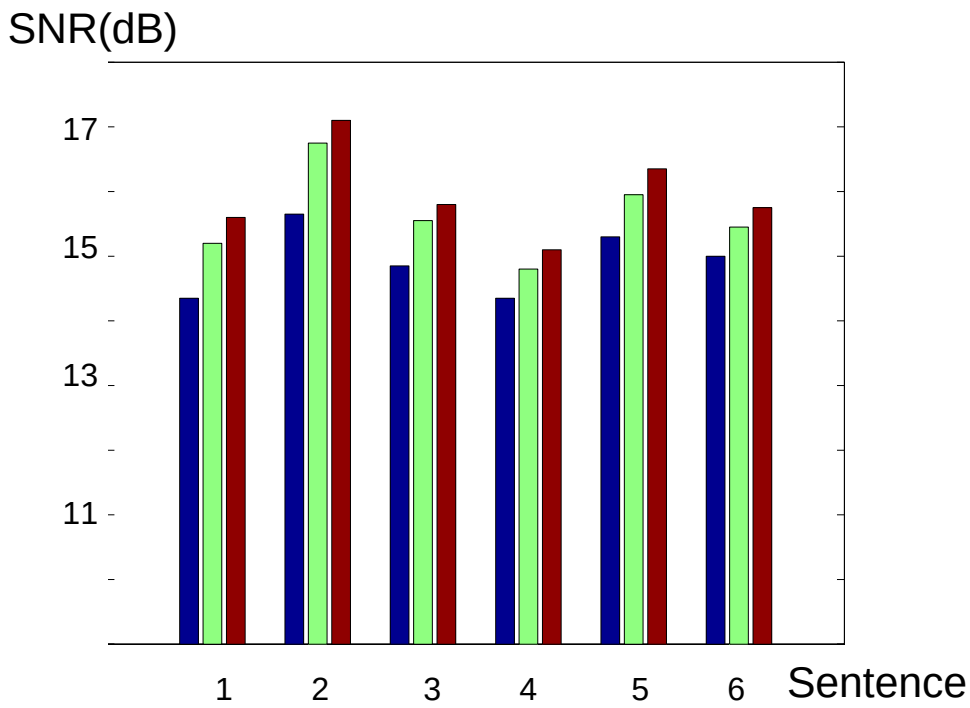
K=4

MSE and WMSE DDCELP Evaluation

Standard CELP

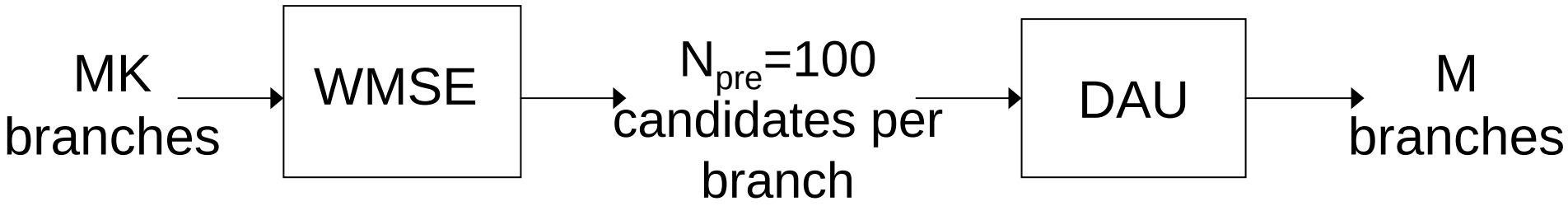
DDCELP (M=K=4,L=2) → **+1.5dB**

DDCELP (M=K=16,L=10) → **+2dB**



DAU-DDCELP Evaluation

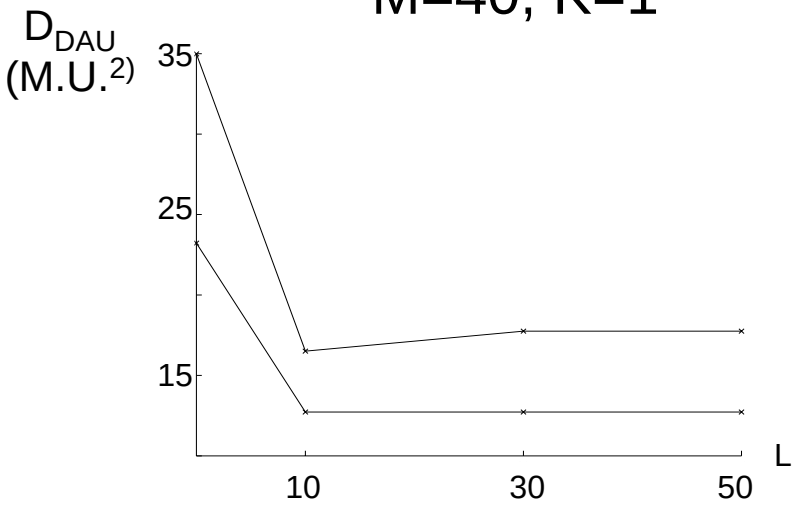
- High Complexity ⇨ WMSE-DAU-DDCELP



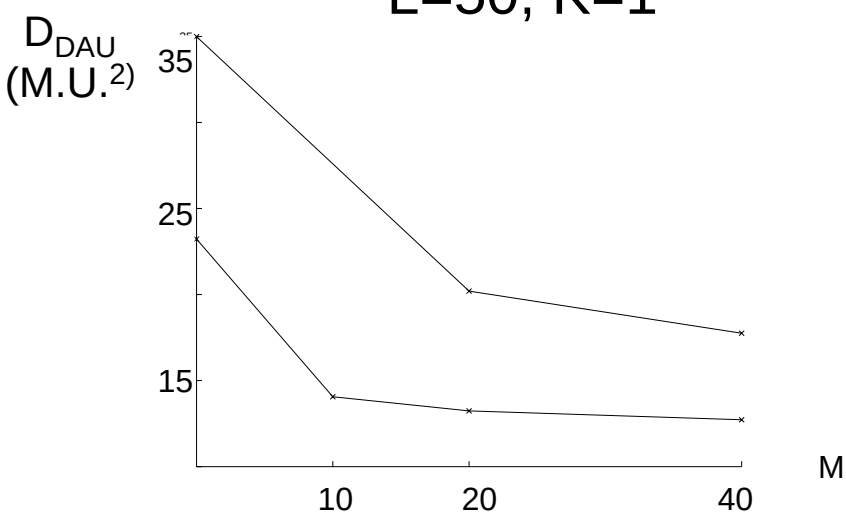
- Bad pitch estimation ⇨ K=1
- Long support ⇨ L ↑
- Effective delay ⇨ M ↑

WMSE-DAU-DDCELP Evaluation

Dependence with L.
M=40, K=1



Dependence with M.
L=50, K=1



- Not expected performance
  - Too low M, K, L
  - WMSE-DAU-DDCELP spoils DAU criterion
  - Dau not suitable for speech coding?
  - Bad distortion measure?

Outline

- The Speech Signal
- The Standard CELP Coder
- A Sophisticated Error Criterion: an Auditory Model
- Delayed Decision CELP Coder
- Evaluation
- Conclusion and future work

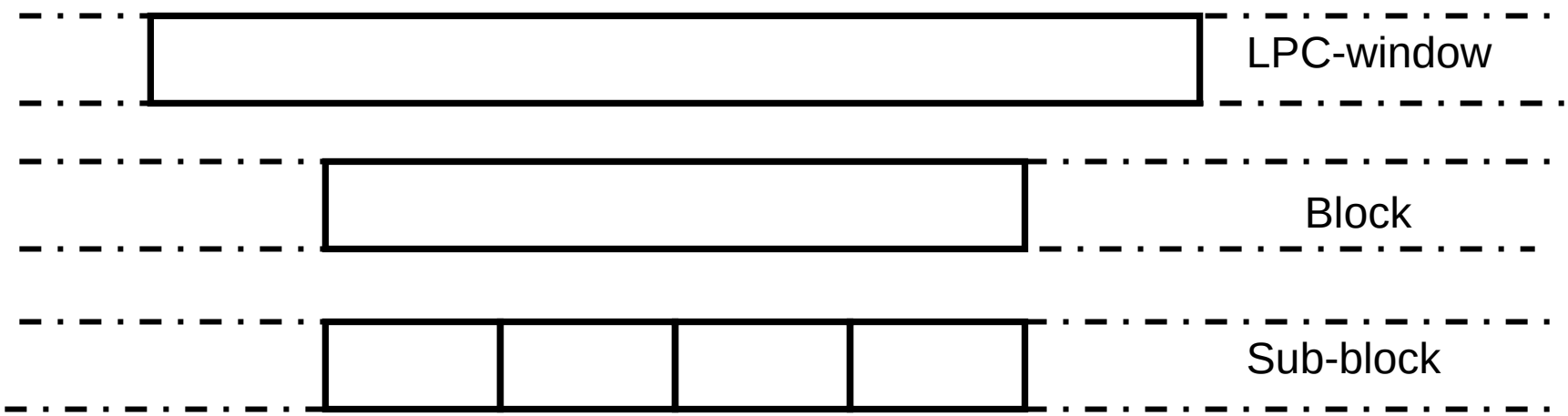
Conclusion and Future Work

- Conclusion
  - MSE-DDCELP and WMSE-DDCELP
    - Up to +2 dB of improvement
    - +1.5dB with 10ms of delay and reasonable complexity
  - DAU-DDCELP
    - Very high complexity
    - DDCELP decreases DAU distortion
    - So far bad results
- Future Work
  - DAU-DDCELP
    - Study influence of N_{pre}
    - Run simulations with higher M, L, K
    - Try different distortion measures

Questions

?

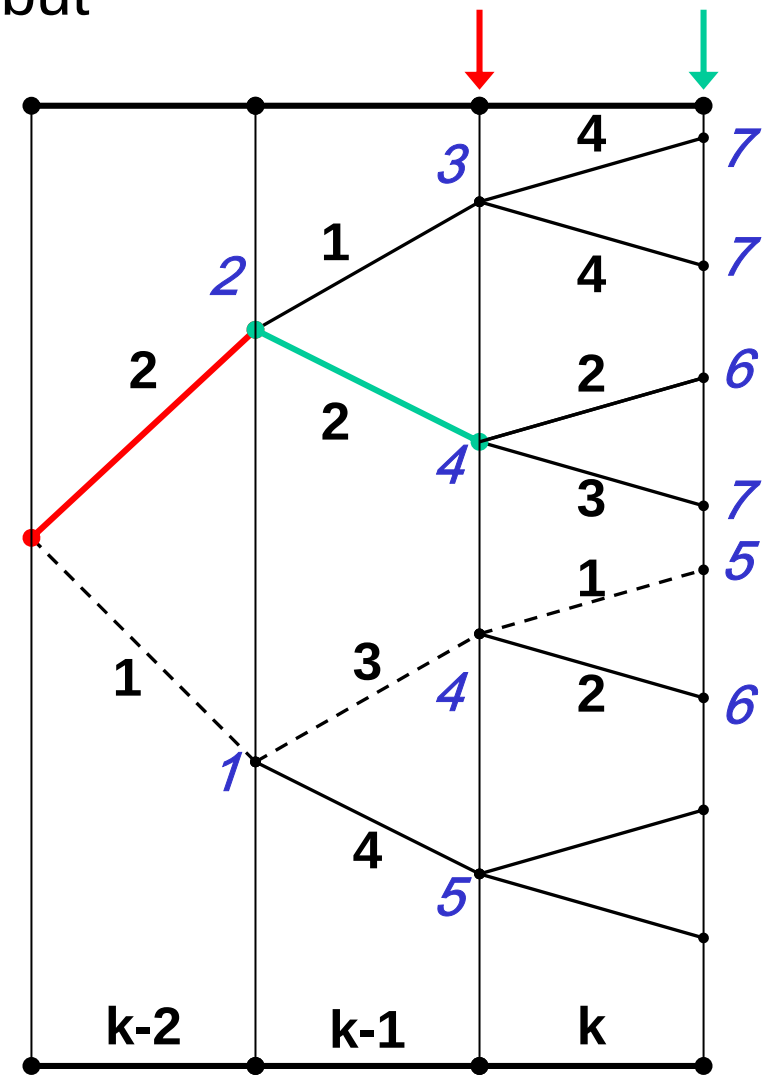
Appendix 1: Block Basis



- Symmetric window for LPC calculation
- Block of 20 ms for LPC actualization
- Sub-block of 5 ms for excitation calculation

Appendix 2: MSE and WMSE DDCELP Evaluation Considerations

- Growing tendency but fluctuations



Appendix 3: Previous DDCELP vs. M-K-L