Signal and Image Processing Lab

Very Low Bit Rate Coding using Temporal Decomposition

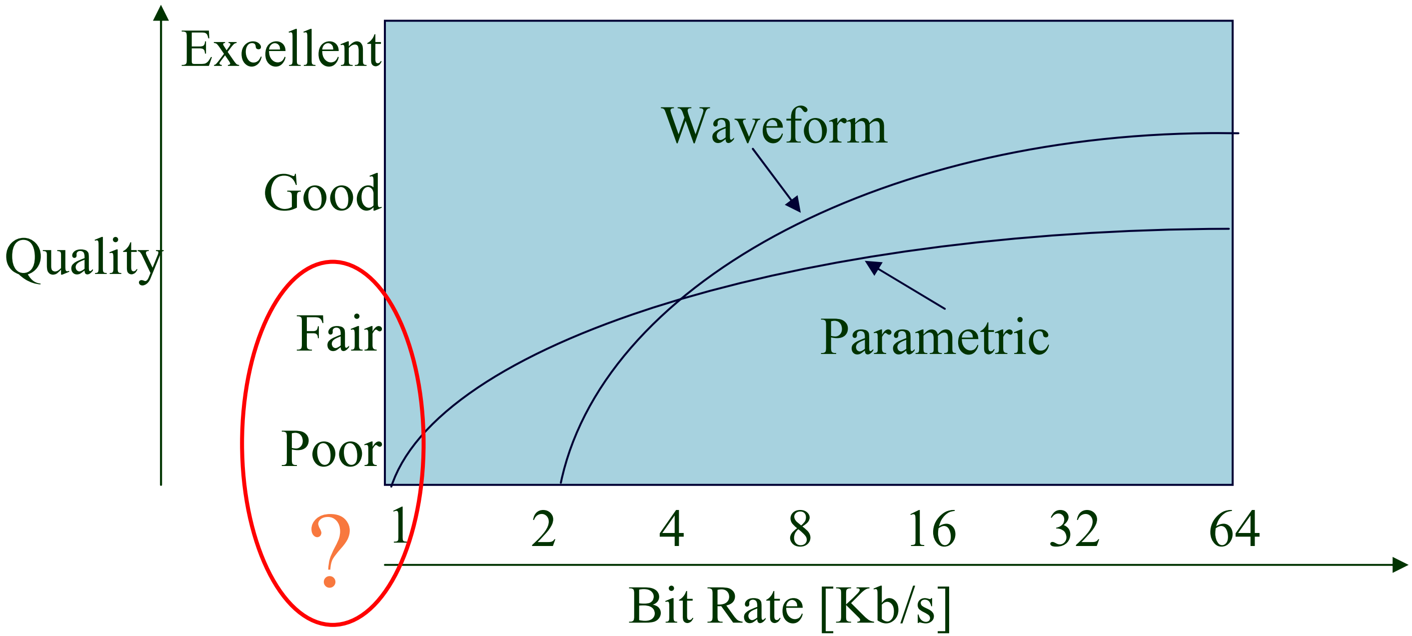
Slava Shechtman

Supervisor: Prof. David Malah

May,23, 2004

Objective

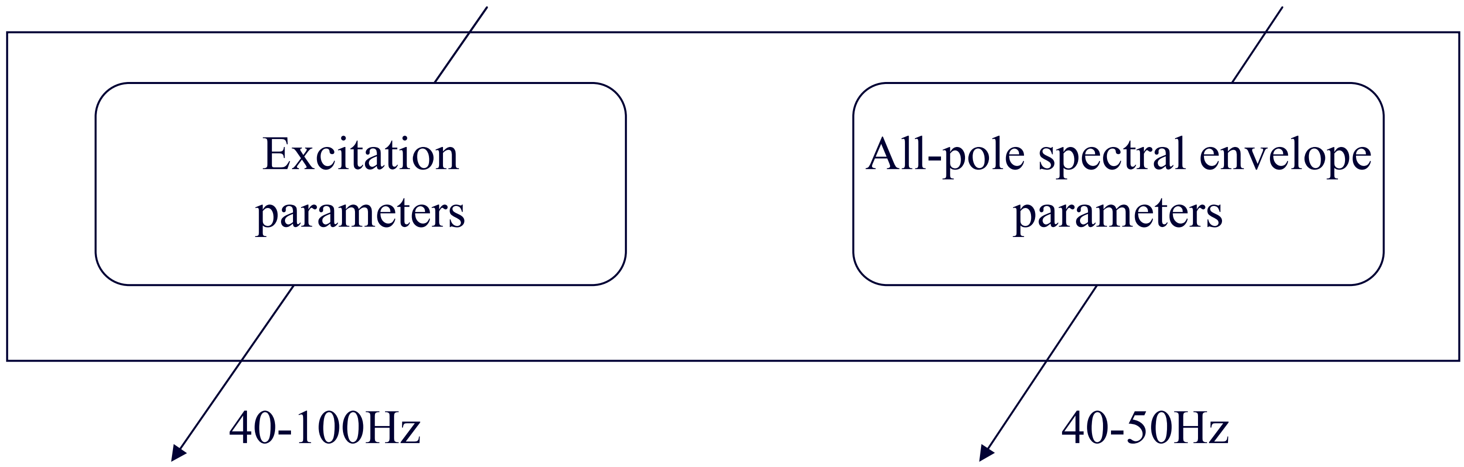
- Explore the possibility of speech coding at 600 bps with fair quality, based on common LPC parametric vocoder (300 bps for the spectral envelope)



Outline

- Introduction
  - Conventional bit-rate reduction schemes
  - Temporal Decomposition (TD) paradigm
- Dynamically Weighted Reduced TD (DW-RTD)
  - Optimized Reduced TD (ORTD)
  - ORTD with Dynamically Weighted MMSE
  - Computationally efficient sub-optimal algorithm (SORTeD)
- SORTeD-based speech coding
  - Spectral envelope coding
  - Excitation coding
- Performance evaluation

Low bit-rate (LBR) speech coding



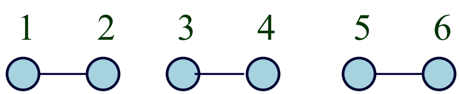
- Common LBR speech coders (LP based) require at least 1000 bps for spectral envelope representation.
- Usually 10 LSF coefficients are coded in each frame
- Excitation parameters depend on desired quality and excitation production model

Inter-frame redundancy removal

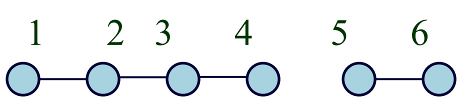
- Common rate reduction schemes exploit inter-frame redundancies and reach 500-600 bps for the envelope representation (speaker independent).
- Basically two approaches were explored:
  - **Joint frame representation**
    - Combine a number of parameter vectors to jointly represent them, using large codebooks.
  - **Frame skipping**
    - Skip frames. Skipped data is interpolated at the decoder.

Inter-frame redundancy removal-2

- Joint frame representation:
  - Matrix quantization- MQ [Tsao & Gray, 1985]
    - Joint quantization of **fixed-length** blocks of spectral parameter vectors.

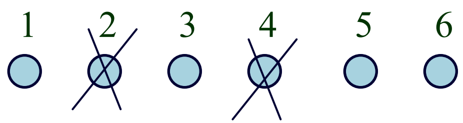


  - Segment quantization –SegQ [Honda & Shiraki, 1992]
    - Segmentation and joint quantization of variable-length blocks of spectral parameter vectors

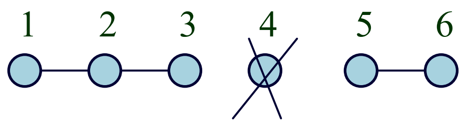


Inter-frame redundancy removal-3

- Frame skipping:
  - Optimal frame skipping [George, 1996]
    - Select M frames out of a block of N frames



  - Optimal combine & skip technique [Mayrench & Malah, 1999]
    - Select M representatives out of a block of N frames, allowing frame skipping.



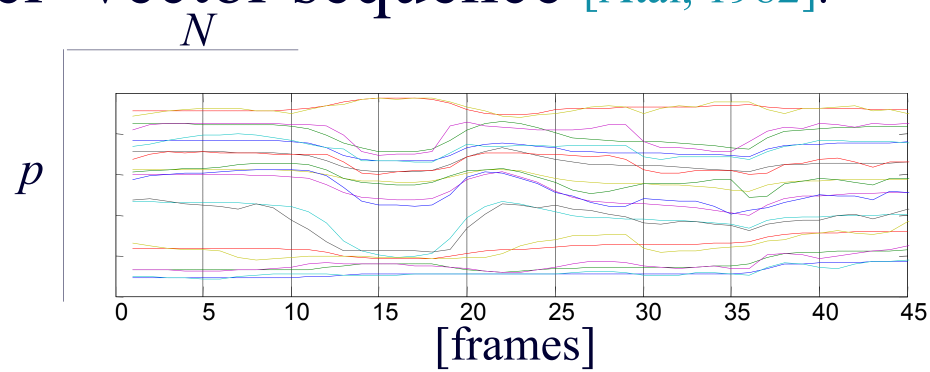
Inter-frame redundancy removal-4

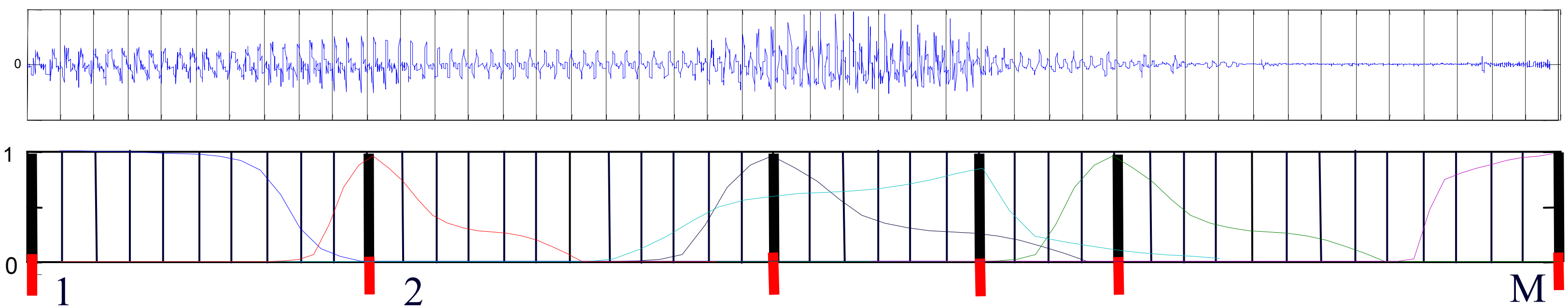
- Limitations
  - Huge codebooks
  - Complicated codebook training
  - Interpolation causes degradation

Temporal Decomposition (TD)

- Technique for temporal redundancies removal from spectral parameter vector sequence [Atal, 1982].

  - *N Input vectors*





  - *M<N Event instants & target vectors (targets)*
  - *M Event functions*

Temporal Decomposition-2

$$\begin{pmatrix} \vdots & \vdots & & \vdots \\ \mathbf{y}(1) & \mathbf{y}(2) & \cdots & \mathbf{y}(N) \\ \vdots & \vdots & & \vdots \end{pmatrix}_{p\times N} \approx \begin{pmatrix} \vdots & \vdots & & \vdots \\ \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_M \\ \vdots & \vdots & & \vdots \end{pmatrix}_{p\times M} \begin{pmatrix} \cdots & \boldsymbol{\varphi}_1 & \cdots \\ \cdots & \boldsymbol{\varphi}_2 & \cdots \\ & \vdots & \\ \cdots & \boldsymbol{\varphi}_M & \cdots \end{pmatrix}_{M\times N}$$

Parameter vectors

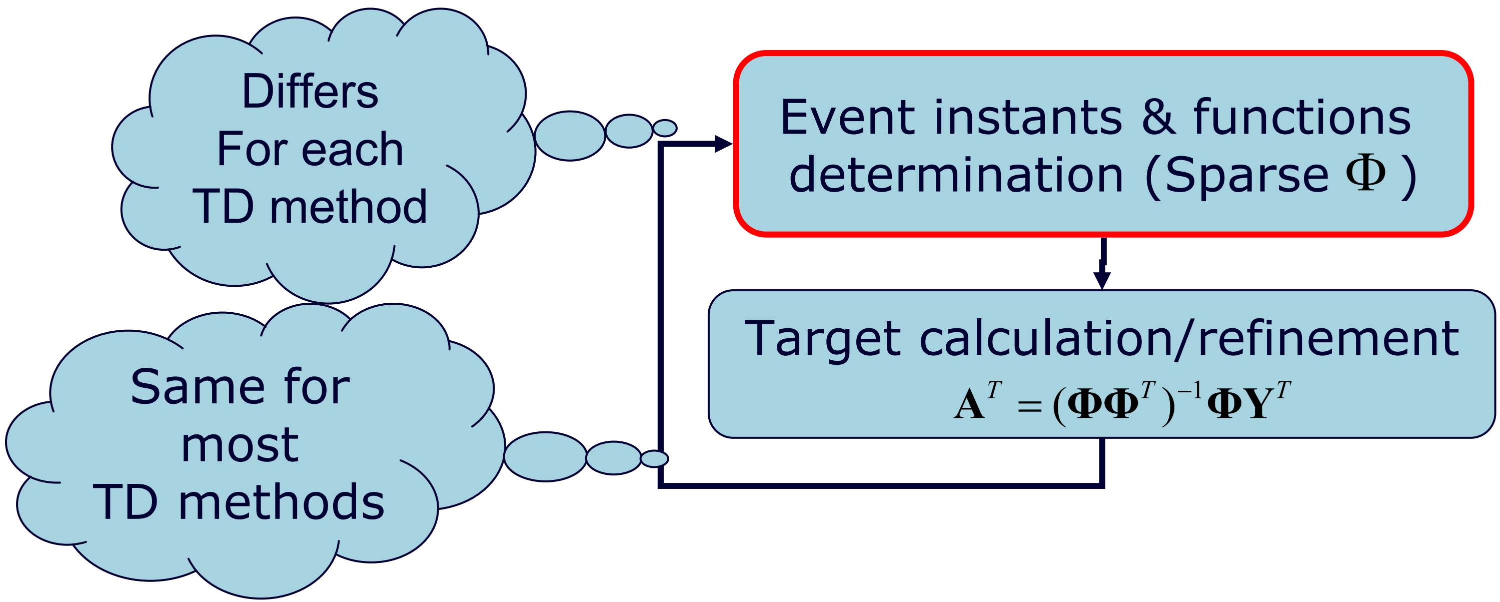
Target vectors

Event functions,
centered over event instants

$$\mathbf{Y} \overset{\text{MMSE}}{\approx} \mathbf{A} \times \mathbf{\Phi}$$

Temporal Decomposition-3

- Two major stages : *event functions* determination and *target* calculation/refinement

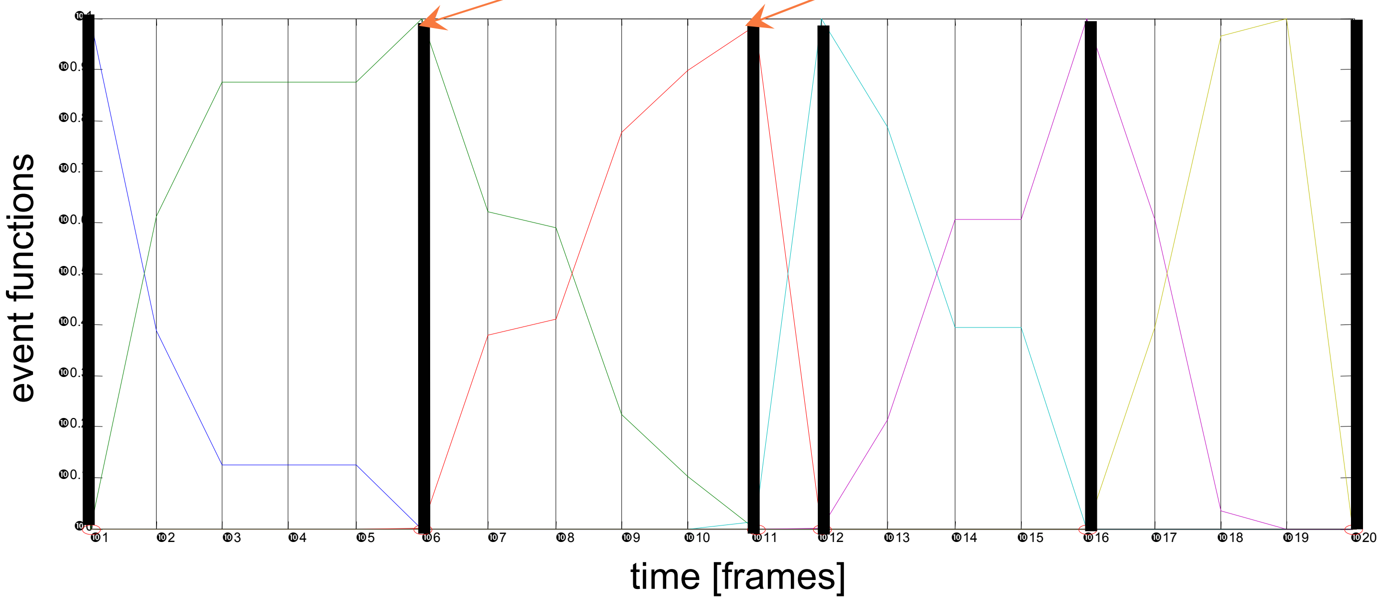


- $\Phi\Phi^T$ is sparse, i.e. target calculation is efficient

Reduced Temporal Decomposition (RTD)

- *Reduced TD* [Athaudage, 1999, Kim & Oh, 1999] - only adjacent event functions may overlap:

$$\hat{\mathbf{y}}(n) = \mathbf{a}_m \phi_m(n) + \mathbf{a}_{m+1}\phi_{m+1}(n), \quad n_m \le n < n_{m+1}$$

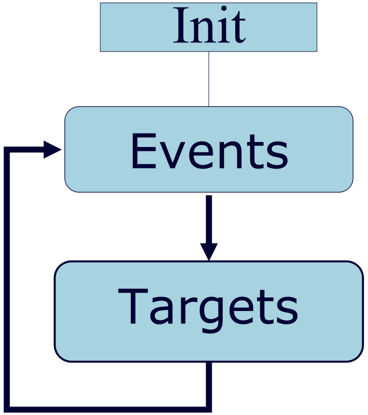


Reduced Temporal Decomposition-2

- Set Event instants, assume $\mathbf{a}_m = \mathbf{y}(n_m)$.
- Optimal event determination in MMSE sense
  - Closed form analytic solution for event functions, given targets and event instants.

$$\begin{pmatrix} \phi_k(n) \\ \phi_{k+1}(n) \end{pmatrix} = \begin{pmatrix} \mathbf{a}_k^T \mathbf{a}_k & \mathbf{a}_k^T \mathbf{a}_{k+1} \\ \mathbf{a}_k^T \mathbf{a}_{k+1} & \mathbf{a}_{k+1}^T \mathbf{a}_{k+1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{a}_k^T \mathbf{y}(n) \\ \mathbf{a}_{k+1}^T \mathbf{y}(n) \end{pmatrix},$$

$$n_{k-1} \le n < n_k$$



- Target refinement stage includes LS minimization: $(\mathbf{\Phi\Phi}^T)\mathbf{A}^T = \mathbf{\Phi Y}^T$
  - p sets of tri-diagonal linear equations – efficient solution

Constrained event function solutions

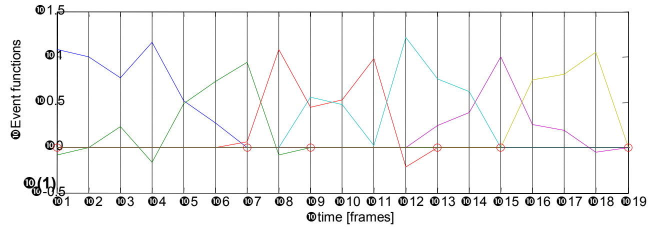
- Impose 1's complement constraint [Kim & Oh, 1999] (☞)

$$\hat{\mathbf{y}}(n) = \mathbf{a}_k \phi_k(n) + \mathbf{a}_{k+1}(1-\phi_k(n)),\ \ \phi_k(n) \geq 0$$

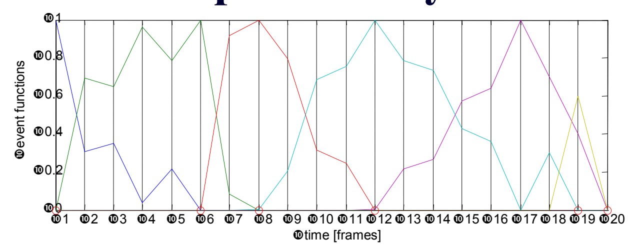
  - Code only right-hand branch of each event function

- Monotonicity of event function branches [Nguyen & Akagi, 1999] (☞)

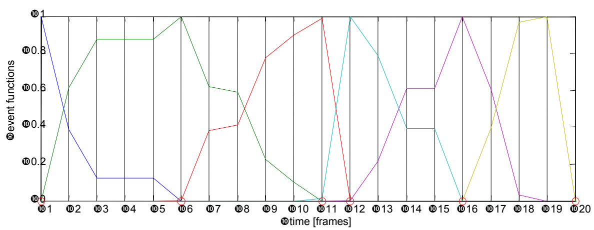
Unconstrained solution



1's complementary solution

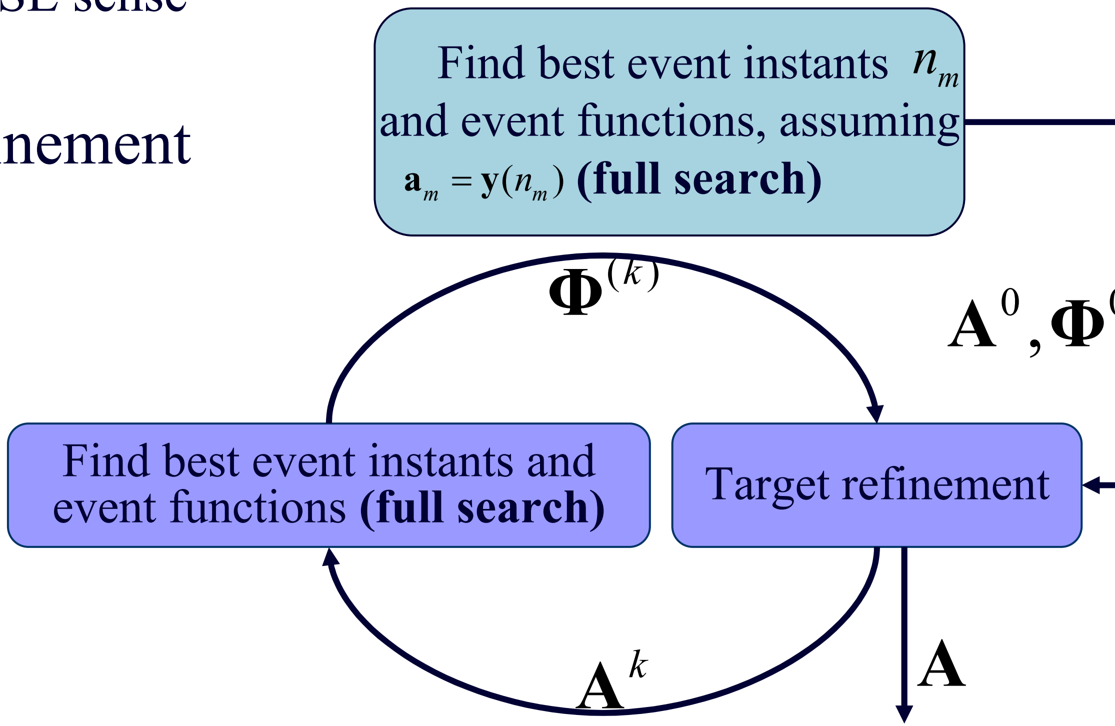


1's comp., monotonic solution



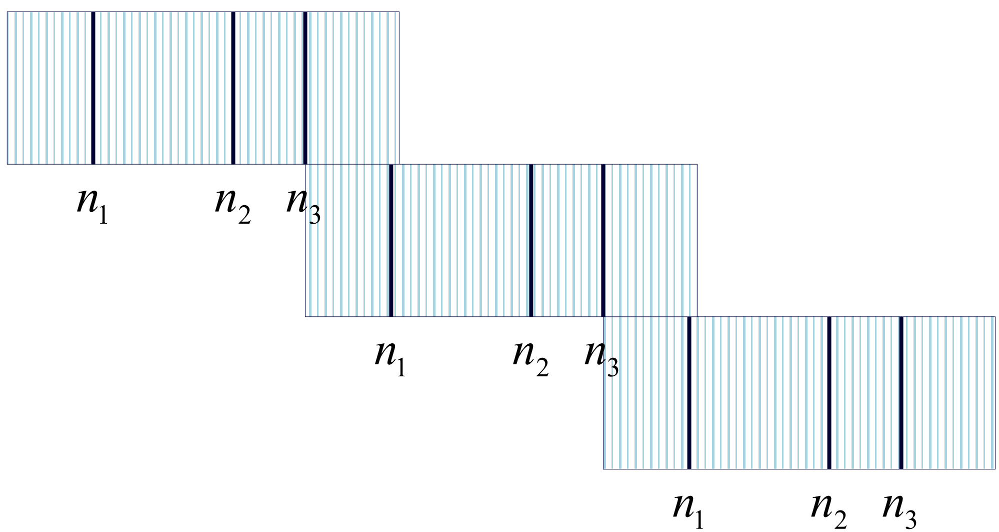
Optimized RTD (ORTD) [Athaudage, 1999]

- Perform RTD for all possible placements of M events in a block of N frames
  - Use Viterbi algorithm (trellis search)
  - Impose required event rate
  - Best solution in MMSE sense
  - High complexity
- Possible solution refinement iterations

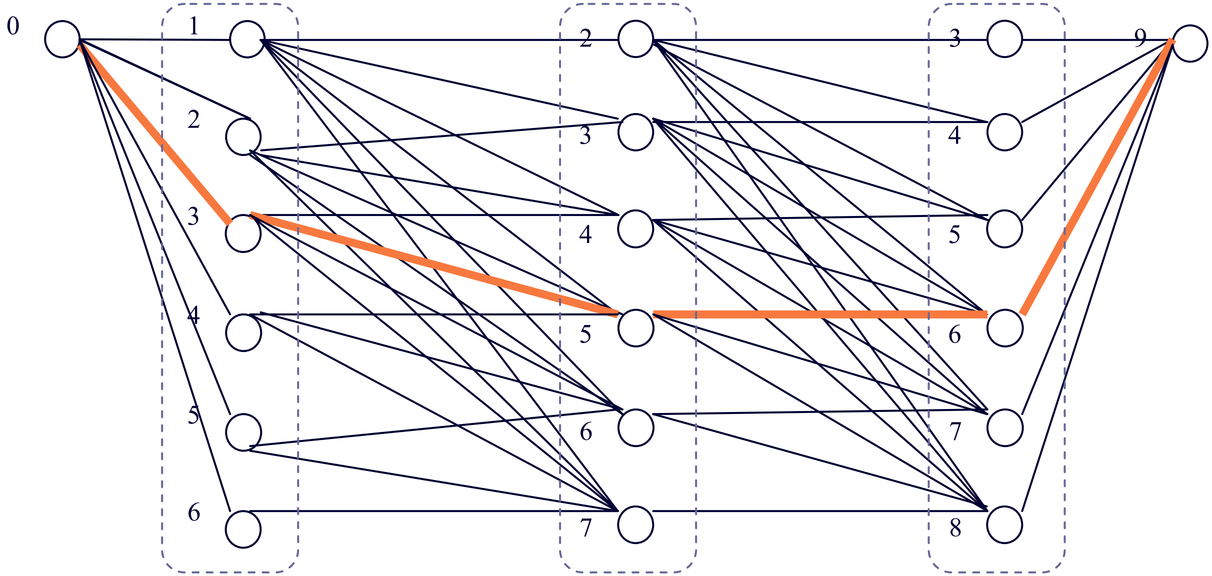


Optimized RTD (ORTD)-2

- Boundary conditions:
  - Block overlap: Last event of previous block = beginning of current block (zero event)
    - Slightly increases event rate
    - Improves overall quality
  - Dummy event at the block end (*M*+1 event)



Optimized RTD (ORTD)-3



- **Trellis stages**: events (M+2)
- **Nodes:** possible event instants (~NM)
- **Branch cost:** sum of best instant errors

Dynamically Weighted ORTD - motivation

- MMSE criterion for spectral envelope parameters (i.e. LSF) may not correlate well with human perception.
- Log Spectral Distance (LSD) is highly correlated with human perception, but is complicated for practical design.

$$d_{LSD}(A,\hat{A}) = \sqrt{\frac{1}{2\pi}\int_{-\theta_1}^{\theta_2}\left(10\log_{10}\left|\frac{1}{A(\omega)}\right|^2 - 10\log_{10}\left|\frac{1}{\hat{A}(\omega)}\right|^2\right)^2 d\omega}$$

- Usually, WMSE is used in practical designs, where the weights depend on the input vector.

$$d_{WMSE}(\mathbf{a},\hat{\mathbf{a}}) = (\mathbf{a}-\hat{\mathbf{a}})W_{\mathbf{a}}(\mathbf{a}-\hat{\mathbf{a}})^T$$

WMSE for LSF vectors

- Atal & Paliwal's Weighting [1993]
  - W is a diagonal matrix with elements proportional to the synthesis filter spectrum.
- Gardner's Weighting [1994]
  - Approximate LSD using WMSE (for low distortions)
- Modified Gardner's Weighting
  - Modify Gardner's weights by a fixed attenuation of their high frequency components
- Ranking of weighting performance :

1.Modified Gardner weights

2. Paliwal-Atal weights

3.Gardner weights

4. No weights

Reduce LSD

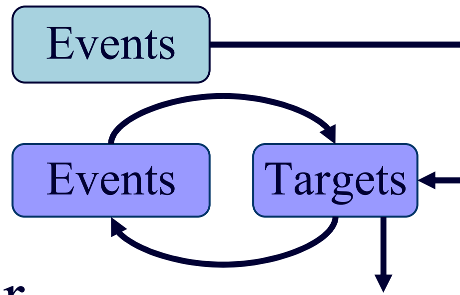
(→)

Dynamically Weighted ORTD (DW-ORTD)

- Event determination
  - Simple modification of event function calculation (→)
- Target Refinement
  - Revise target refinement stage by minimization of

$$E_{block}^{(i)} = \sum_{k=0}^{M} \sum_{n=n_k}^{n_{k+1}-1} w_i(n)(y_i(n) - a_{i,k}\phi_k(n) - a_{i,k+1}\phi_{k+1}(n))^2,$$

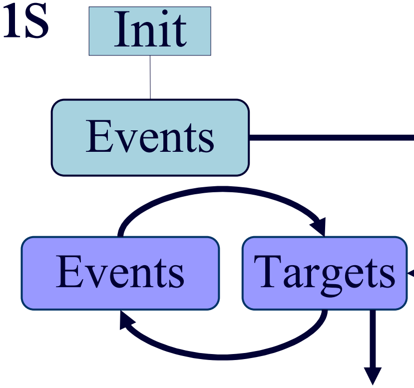
$$1 \le i \le p$$

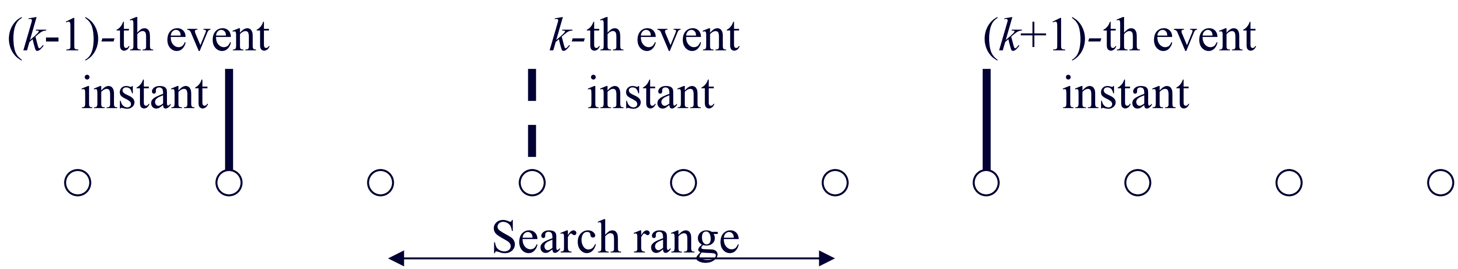


  - Solve p sets of tri-diagonal, symmetric linear equations (→)
  - Similar complexity as in MMSE criterion

Sub-optimal RTD algorithm (SORTeD)

- **ORTD:** Full Search event-determination is not suited for real-time implementation.
- **SORTeD**: Apply <u>partial search</u> of event instants with initialization



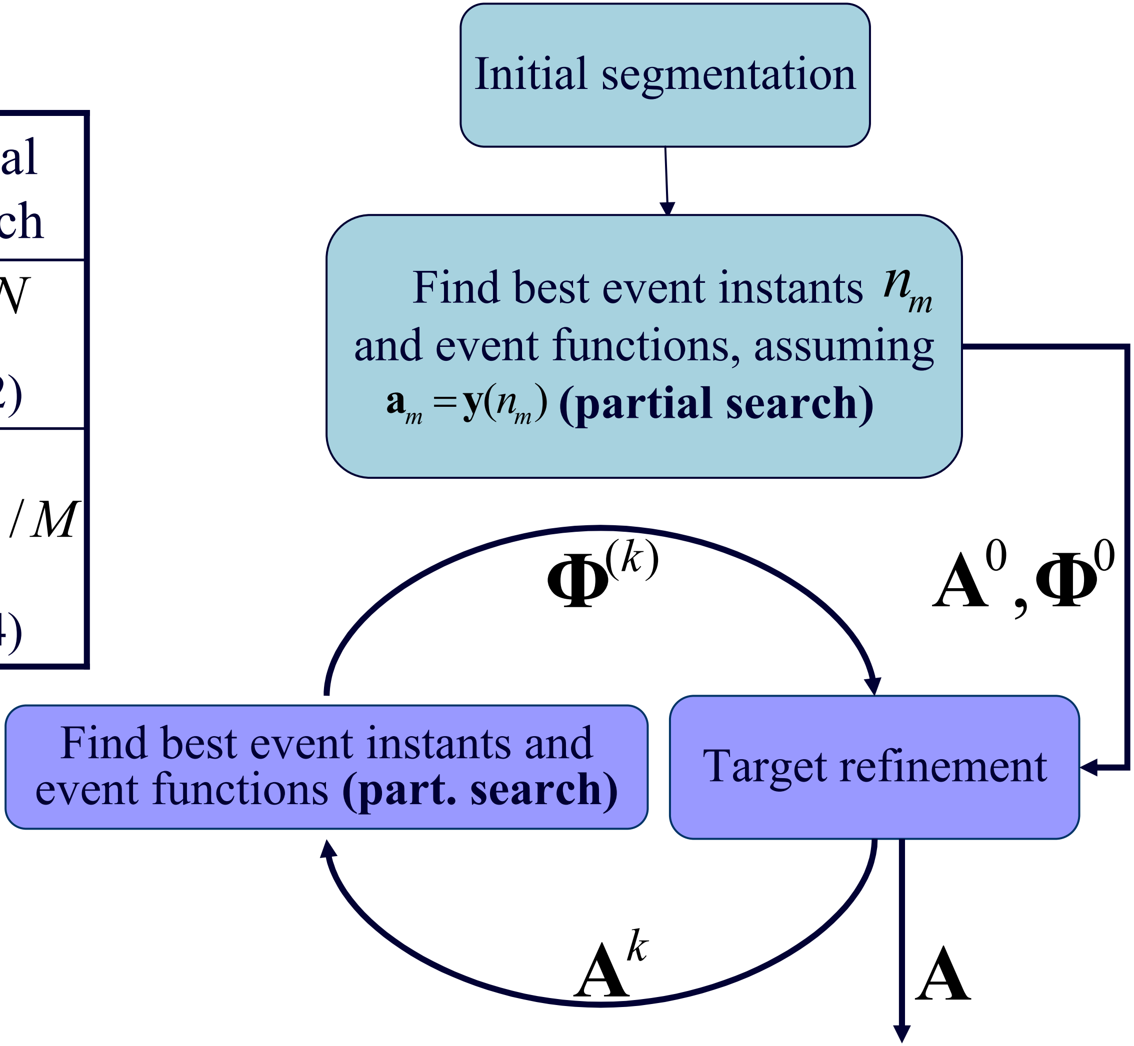


- Initial event instants are uniformly spaced or based on any input vector stability criteria.

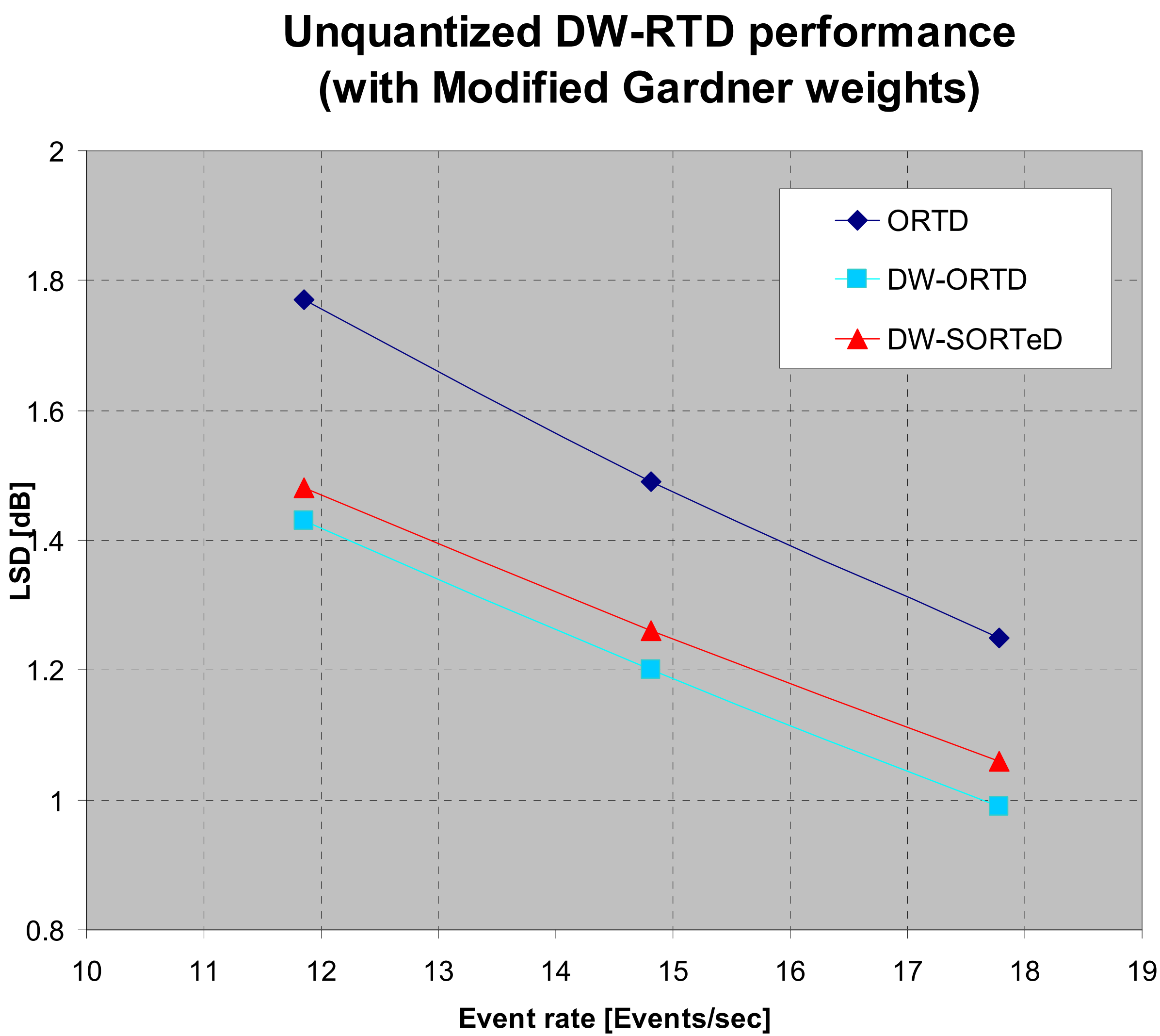
Sub-optimal RTD algorithm (SORTeD)-2

Number of operations

	Full Search	Partial Search
Comparisons (3/11)	$\sim N^2M/2$ (135)	$\sim 2N$ (12)
Error calc. (1st run) (3/11)	$\sim N^3/2$ (212)	$\sim 4N^2/M$ (54)

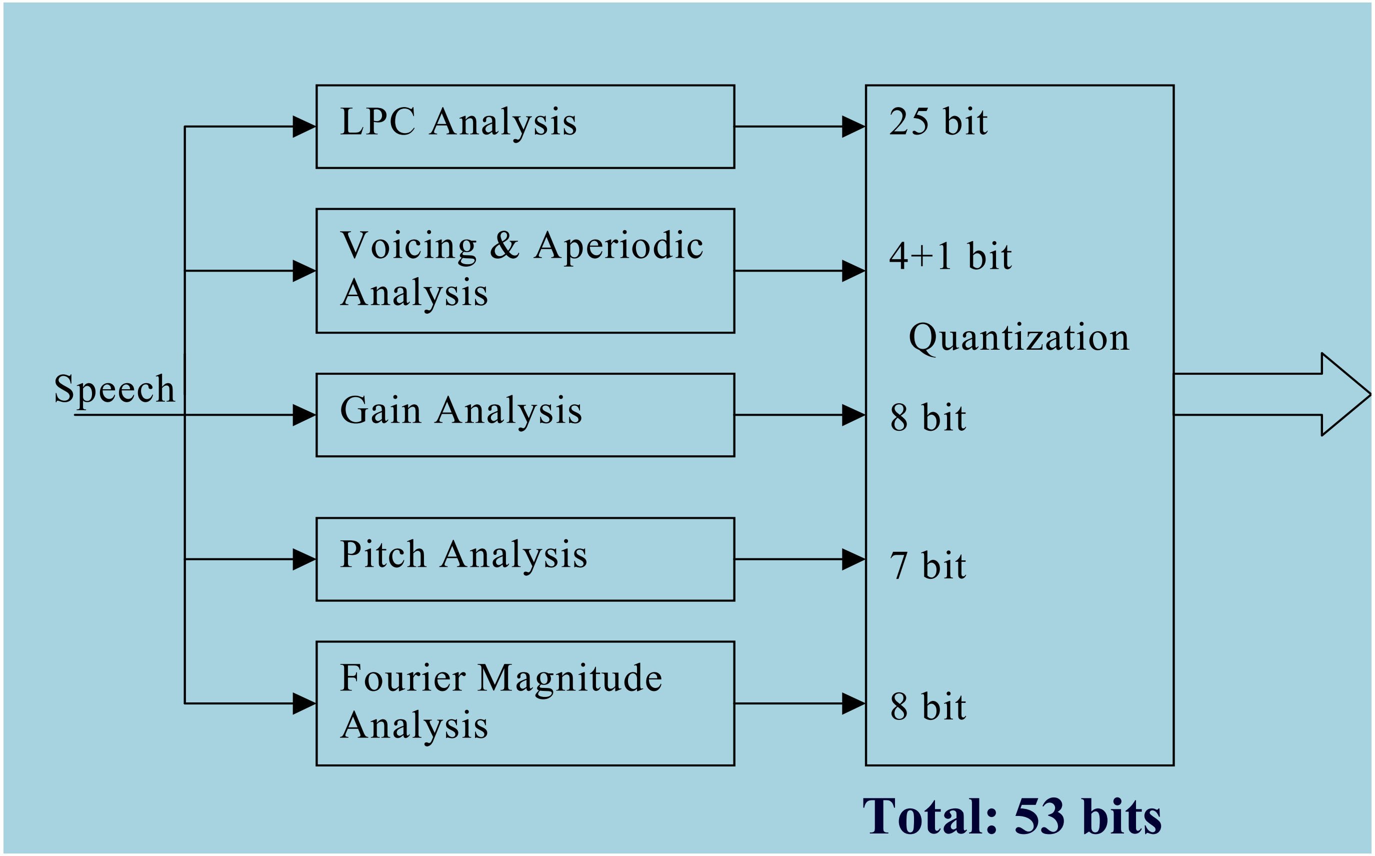


Different RTD models of LSF parameters



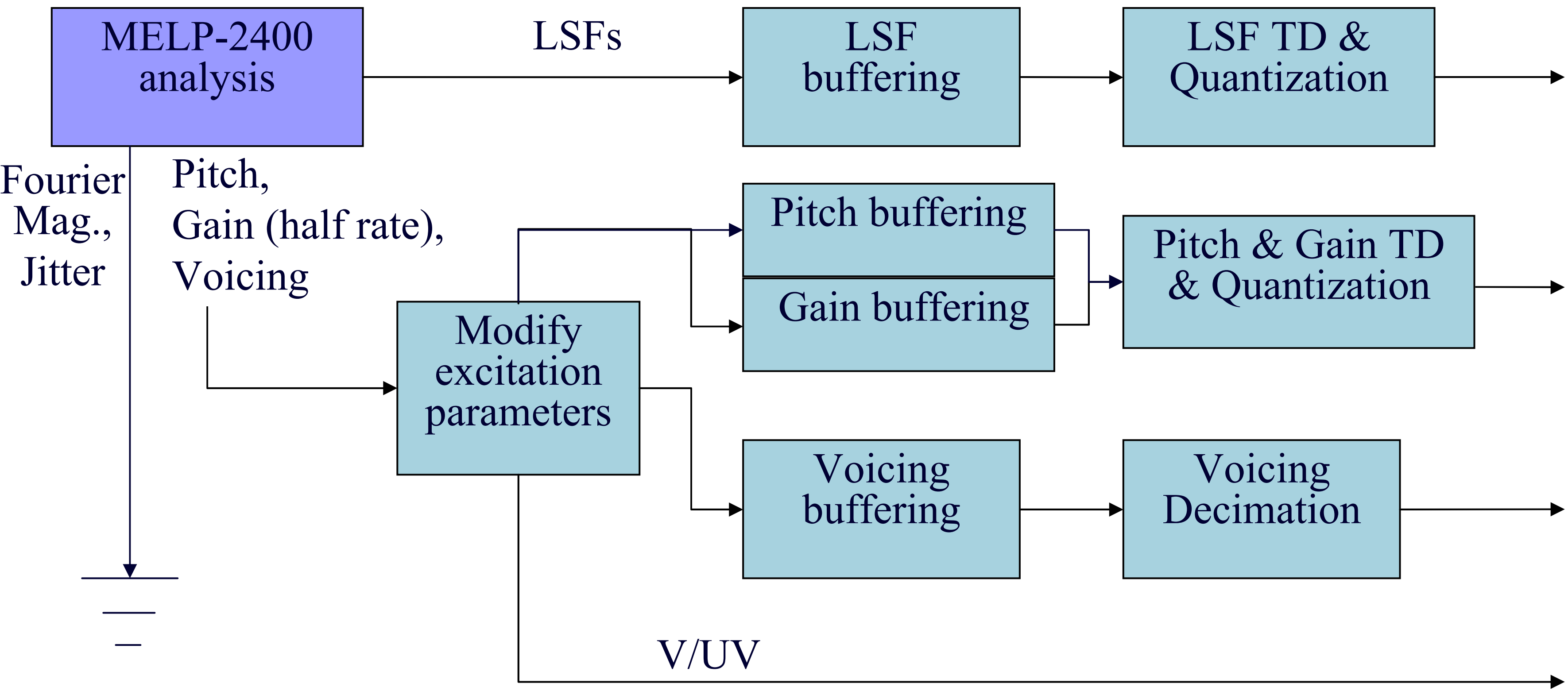
Speech Coding with DW-SORTeD

- Based on MELP-2400 standard



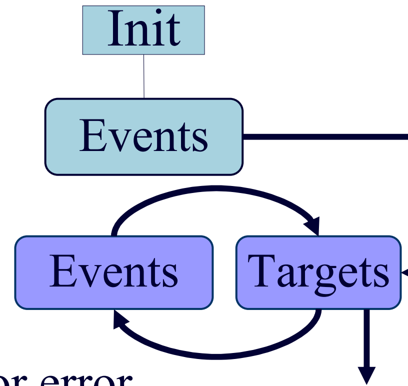
- Frame length is 22.5 ms (44.44 frames/sec)

Speech Coding with DW-RTD-2



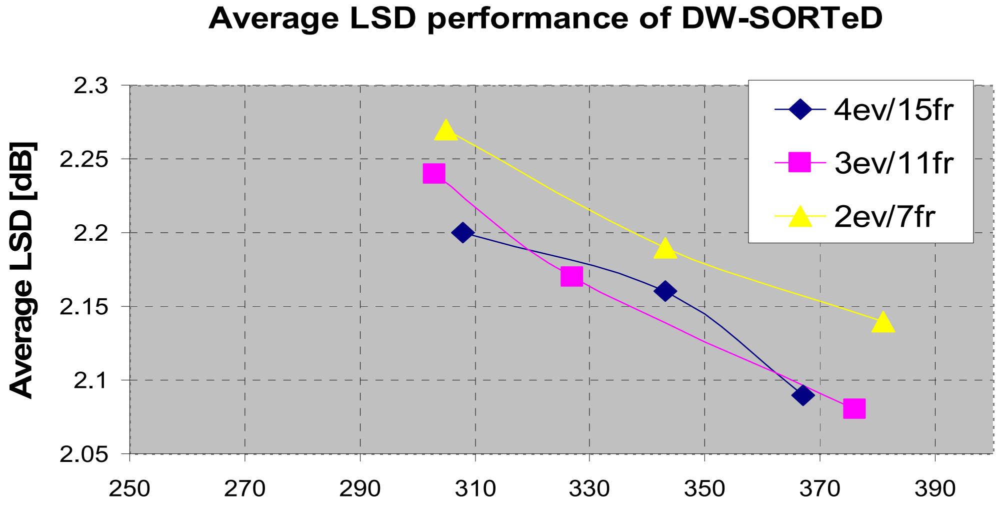
Speech Coding: Spectral Envelope

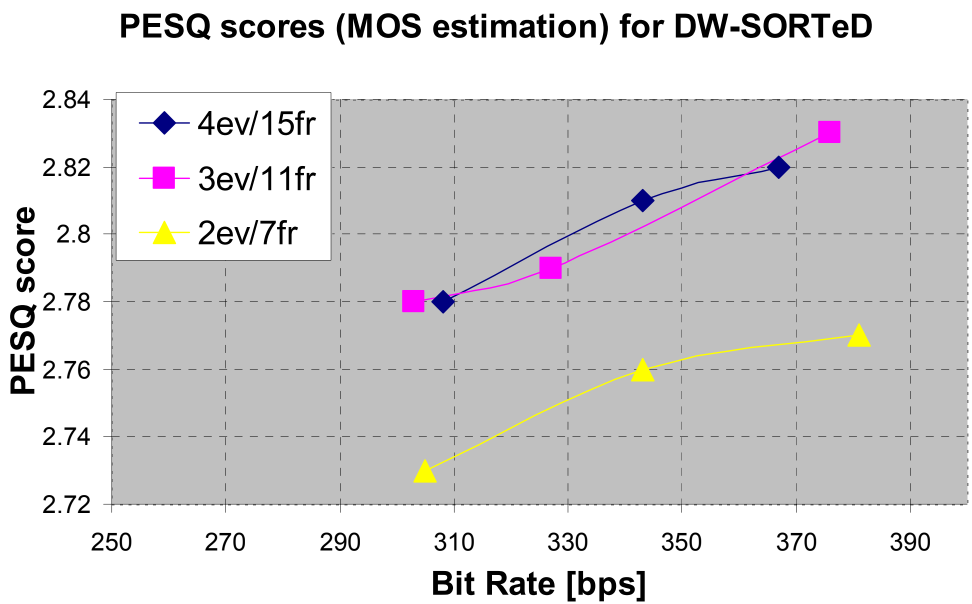
- DW-SORTeD scheme with quantization
  - **Targets**: Split-VQ
  - **Event functions**: multi-codebook VQ
    - The codebooks are trained on <u>constrained</u> DW-SORTeD.
  - **Embedded quantization:**
    - Use quantized target candidates and unquantized inputs for error calculations.
    - Substitute analytic solution for event functions by codebook search
    - Quantize refined targets
    - Allow "early escape"



M/N	Target Codebook-1	Target Codebook-2	Event functions	Event length	Rate [Bps]
3/11	11	9	4	3	327
	10	8	4	3	303
2/7	11	9	4	3	343
	10	8	3	3	305

Speech Coding: Spectral Envelope-2

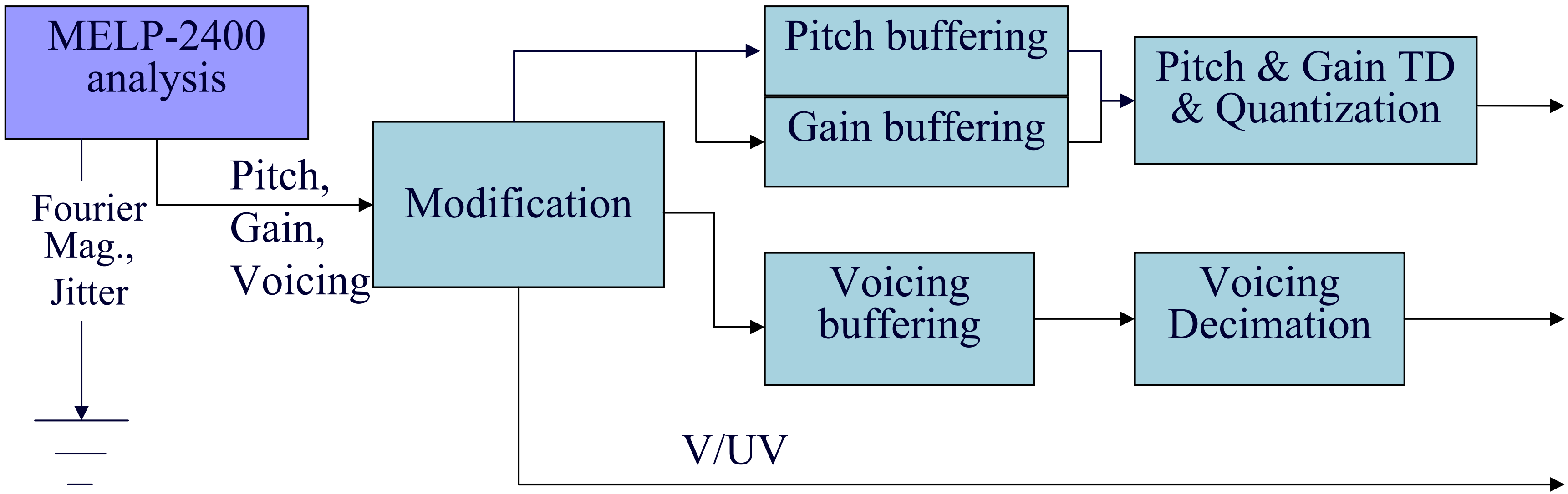




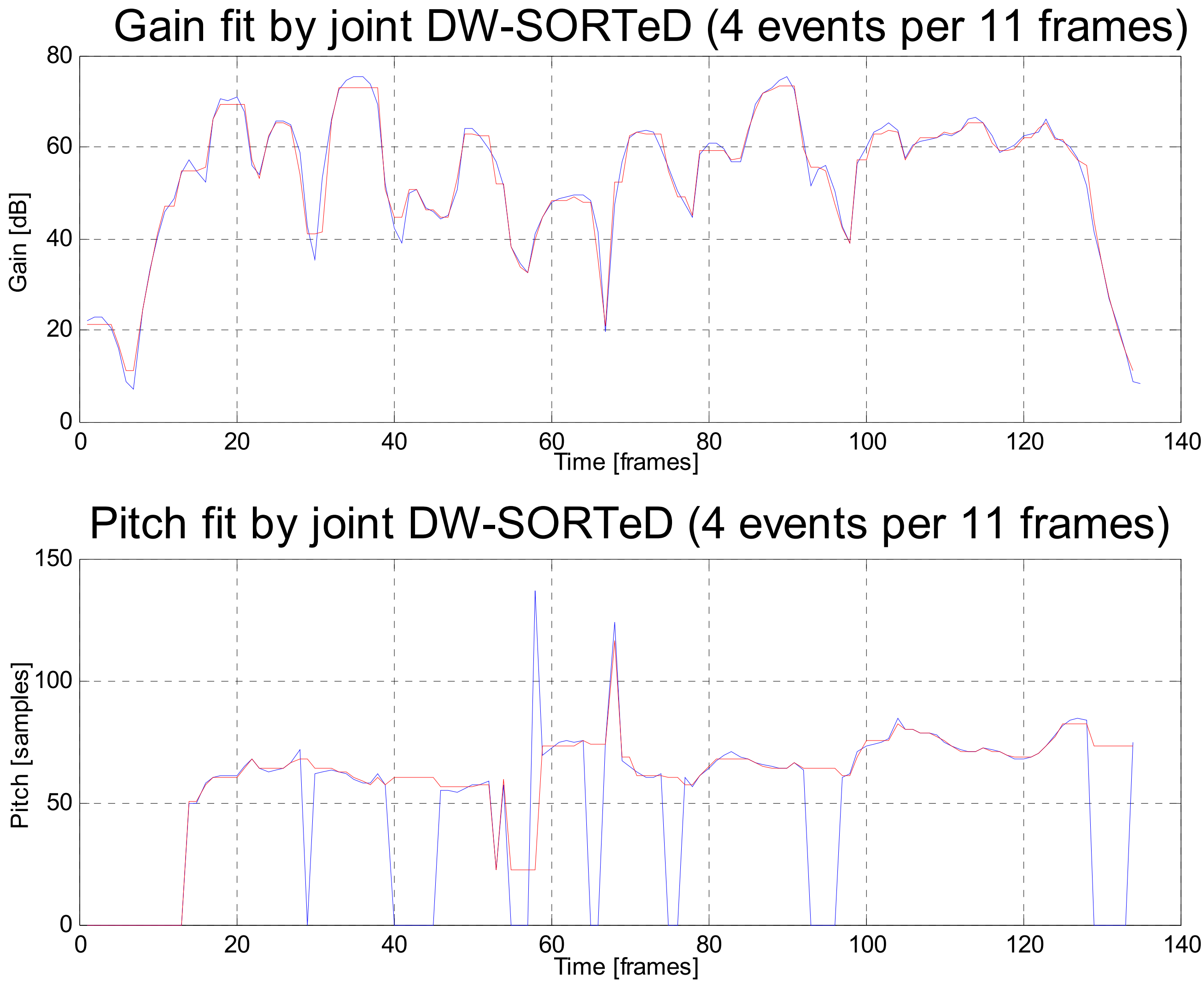
- Averaged on 20 sentences

Speech Coding: Excitation-1

- Code pitch and gain with a DW-SORTeD (jointly or separately)



Speech Coding: Excitation-2



Speech Coders: Bit Assignment examples

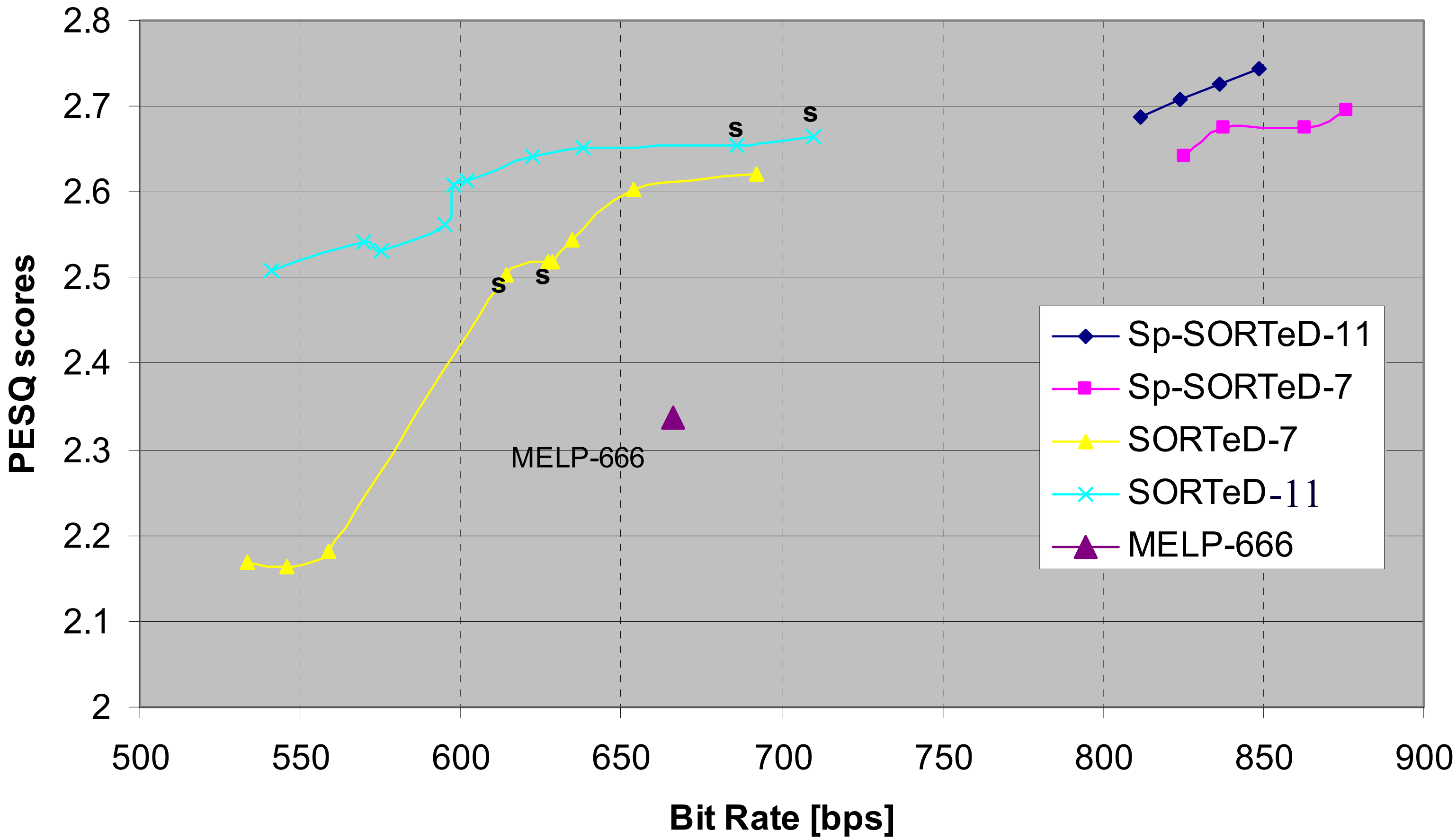
- Codec 1 (250 ms buffer)

Param.	Bits/Block (11 frames)	Bit-Rate [bps]
LSF (3 events)	(10+8+2+3)*3= 69	278.8
Gain & Pitch (4 events)	(5+5+4)*4+7= 63	254.6
UV/V	11	44.4
Voicing	6	24.2
Total	149	602

- Codec 2 (160 ms buffer)

Param.	Bits/Block (7 frames)	Bit Rate [bps]
LSF(2 events)	(10+8+2+3)*2= 46	292
Gain & Pitch (3 events)	(5+5+3)*3+4= 43	273
UV/V	7	44.4
Voicing	4	25.4
Total	100	634.8

Speech coding: performance-1



- S – separate pitch & energy TD
- Sp – spectral envelope coding, with reduced MELP mexcitation

Hearing Examples

Coders	Rate	PESQ	Samples
Original			
MELP-2400	2400	3.22	
MELP Exc + SORTeD spectrum	1550	2.92	
MELP-1600 (reduced excitation)	1600	2.86	
11-frames delayed DW-SORTeD	602	2.58	
7-framed delayed DW-SORTeD	635	2.56	
MELP-666 (Harris, 4 frames MQ)	667	2.34	

Summary

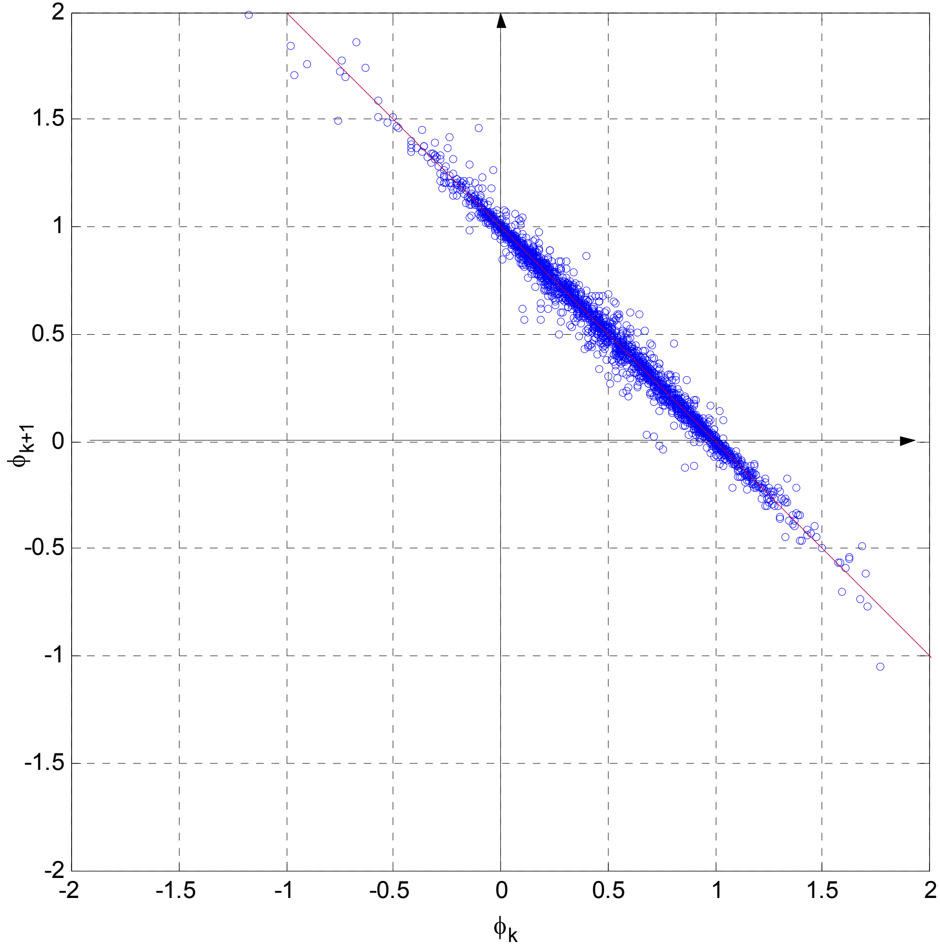
- A 600 bps coding scheme, based on Temporal Decomposition (TD) concept was developed.
  - TD with dynamic weighting
    - Uses Mod. Gardner weights
    - Improves the LSF fit by 0.3 dB (LSD)
  - Suboptimal scheme for Optimized Reduced TD
    - Only slightly deteriorates the model fit
    - Meaningful reduction in complexity
  - Incorporated into MELP vocoder to obtain a 600 bps coder
    - LSF quantization at 280-300 bps
    - Gain & pitch quantization at 250-300 bps.
    - Additional excitation parameters - 70 bps.
    - PESQ of 2.6

Suggestions for further research

- Explore DW-SORTeD power for high quality/high-rate coders
- Improve excitation coding; explore other excitation models (e.g. sinusoidal model, etc.)
- Extend the system by allowing variable rate coding
- Develop low-delay schemes, based on SORTeD concept

The End

Optimal instant event function scatter for RTD model



(⟵)

Instant event functions for RTD - optimal

- Optimal event functions :

$$\begin{pmatrix} \phi_k(n) \\ \phi_{k+1}(n) \end{pmatrix} = \begin{pmatrix} \mathbf{a}_k^T \mathbf{W}(n)\mathbf{a}_k & \mathbf{a}_k^T \mathbf{W}(n)\mathbf{a}_{k+1} \\ \mathbf{a}_k^T \mathbf{W}(n)\mathbf{a}_{k+1} & \mathbf{a}_{k+1}^T \mathbf{W}(n)\mathbf{a}_{k+1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{a}_k^T \mathbf{W}(n)\mathbf{y}(n) \\ \mathbf{a}_{k+1}^T \mathbf{W}(n)\mathbf{y}(n) \end{pmatrix},$$

$$n_{k-1} \le n < n_k$$

(←)

Instant event functions for RTD - constrained

- Constrained event functions :

$$\tilde{\phi}_k(n) = \begin{pmatrix} 1-\tilde{\phi}_{k-1}(n), & n_{k-1} \le n < n_k \\ 1, & n = n_k \\ \min(1, \max(0, \bar{\phi}_k(n))) & n_k \le n < n_{k+1} \\ 0, & else \end{pmatrix},$$

$$\bar{\phi}_k(n) = \frac{\left(\mathbf{y}(n) - \mathbf{a}_{k+1}\right)^T \left(\mathbf{a}_k - \mathbf{a}_{k+1}\right)}{\left(\mathbf{a}_k - \mathbf{a}_{k+1}\right)^T \left(\mathbf{a}_k - \mathbf{a}_{k+1}\right)}$$

(⟵)

DW-RTD Target Calculation

- Target refinement:

$$\begin{pmatrix} d_1 & x_1 & 0 & \mathbf{0} \\ x_1 & \ddots & \ddots & 0 \\ 0 & \ddots & d_{M-1} & x_{M-1} \\ \mathbf{0} & 0 & x_{M-1} & d_M \end{pmatrix} \begin{pmatrix} a_{i,1} \\ \vdots \\ a_{i,M-1} \\ a_{i,M} \end{pmatrix} = \begin{pmatrix} b_1 - x_0 a_{i,0} \\ \vdots \\ b_{M-1} \\ b_M \end{pmatrix},$$

$$d_k = \sum_n \phi_k^2(n) w_i(n), \quad x_k = \sum_n \phi_k(n)\phi_{k+1}(n) w_i(n), \quad b_k = \sum_n \phi_k(n) y_i(n) w_i(n)$$

(⟵)

WMSE for LSF vectors - formulae

- Atal & Paliwal's Weighting [1993]

$$w_i = [P(f_i)]^r, \quad P(f) = \frac{1}{\left|A(e^{j2\pi f/F_s})\right|^2}$$

$$r = 0.15$$

- Gardner's Weighting [1994]

$$d(a,\hat{a}) \cong \frac{1}{2}(a-\hat{a})W(a-\hat{a})^T, \qquad W = \left.\frac{\partial^2 d_{LSD}(a,\overline{a})}{\partial \hat{a}_k \partial \hat{a}_l}\right|_{a=\hat{a}} = 4\beta R_A(k-l)$$

$$R_A(k) = \sum_{n=0}^{\infty} h(n)h(n+k), \qquad h(n) = F^{-1}\left\{\frac{1}{A(z)}\right\}$$

β=constant

- Modified Gardner's Weighting

$$\tilde{w}_i(n) = (c_i)^2 w_i(n),$$

$$\mathbf{c} = [1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0.9 \quad 0.8 \quad 0.7 \quad 0.1 \quad 0.01],$$

(⟵)