SIEMENS

An Interpolative Decoding Approach for Speech Streaming Services and Voice over IP

Tim Fingscheidt
Jesús F. Guitarte Pérez

Siemens AG Mobile Phones
München

SIEMENS

Outline

1. Introduction
2. Predictive Coding of Spectral Parameters (LSFs)
3. Receiver Model
4. Frame Erasure Concealment Algorithm
5. Simulation Results

© Information and Communication Mobile

SIEMENS

1 Introduction

- Speech decoders have to deal with lost / bad speech frames
  => Frame Erasure Concealment
- Delay constraints of some typical speech services:
  - Circuit-Switched Conversational Speech:
    - e.g. telephone conversation via fixed-network or GSM
    - tight delay constraints!
  - Packet-Switched Conversational Speech:
    - e.g. Voice-over-IP systems
    - still delay constrained, but delay jitter requires receiver buffering
  - Speech Streaming:
    - e.g. news-on-demand
    - relaxed delay constraints

Here some additional delay for frame erasure concealment is acceptable!
=> Interpolative approach!

Mobile business T. Fingscheidt, J.F. Guitarte Pérez 3 Information and Communications

SIEMENS

1 Introduction

- Interpolative approaches have proven to be promising!
  - Some recent publications:

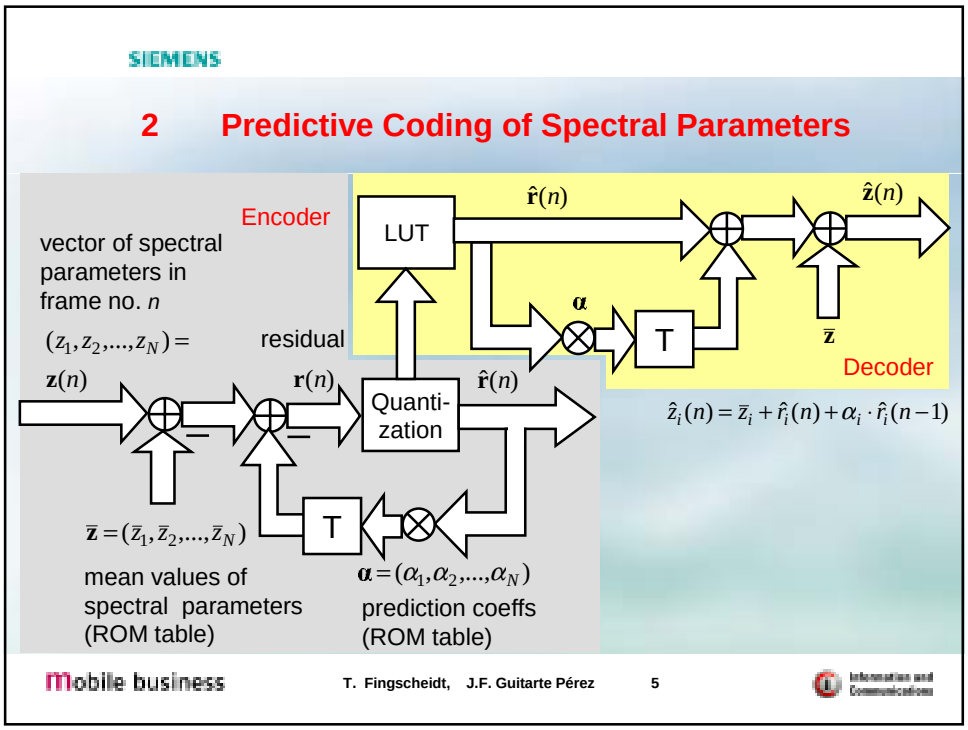
J. Wang, J. Gibson:
Parameter Interpolation to Enhance the Frame Erasure Robustness of CELP-Coders in Packet Networks, ICASSP 2001

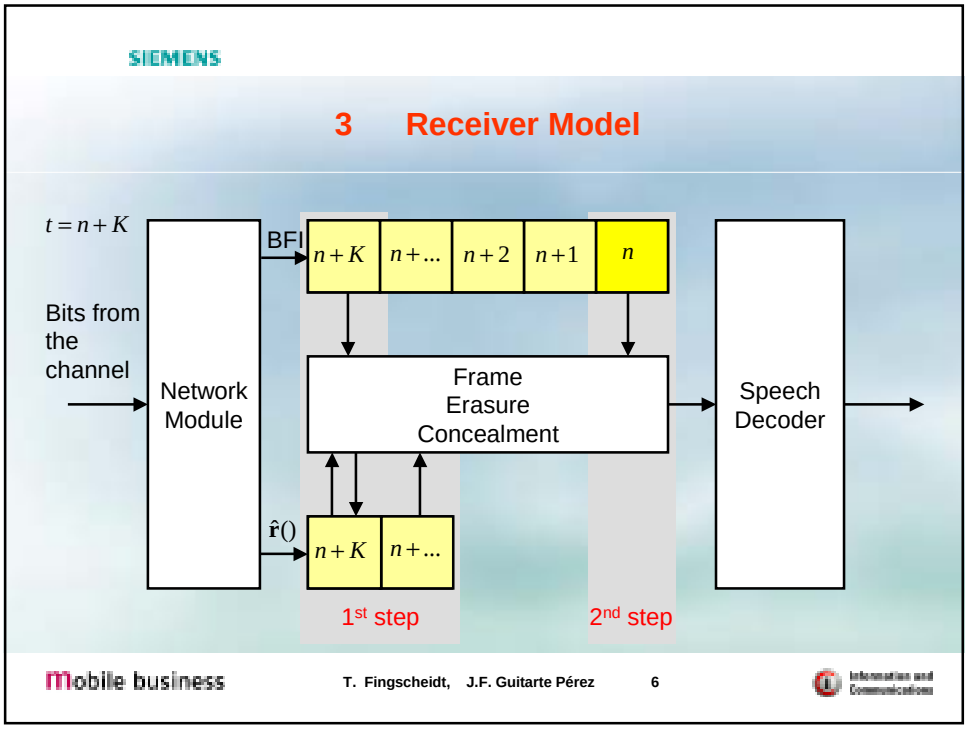
J.C. De Martin, T. Unno, V. Viswanathan:
Improved Frame Erasure Concealment for CELP-Based Coders, ICASSP 2000

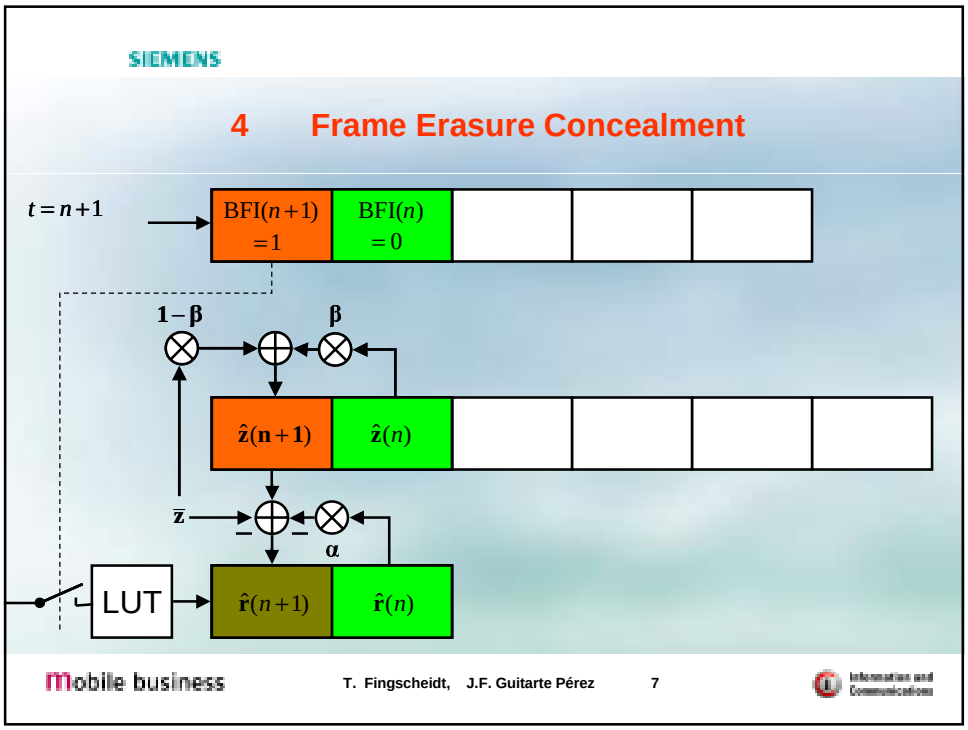
"Since it seems reasonable to expect improvements from the use of future information even in the case of predictive quantization, more work seems possible in this direction".

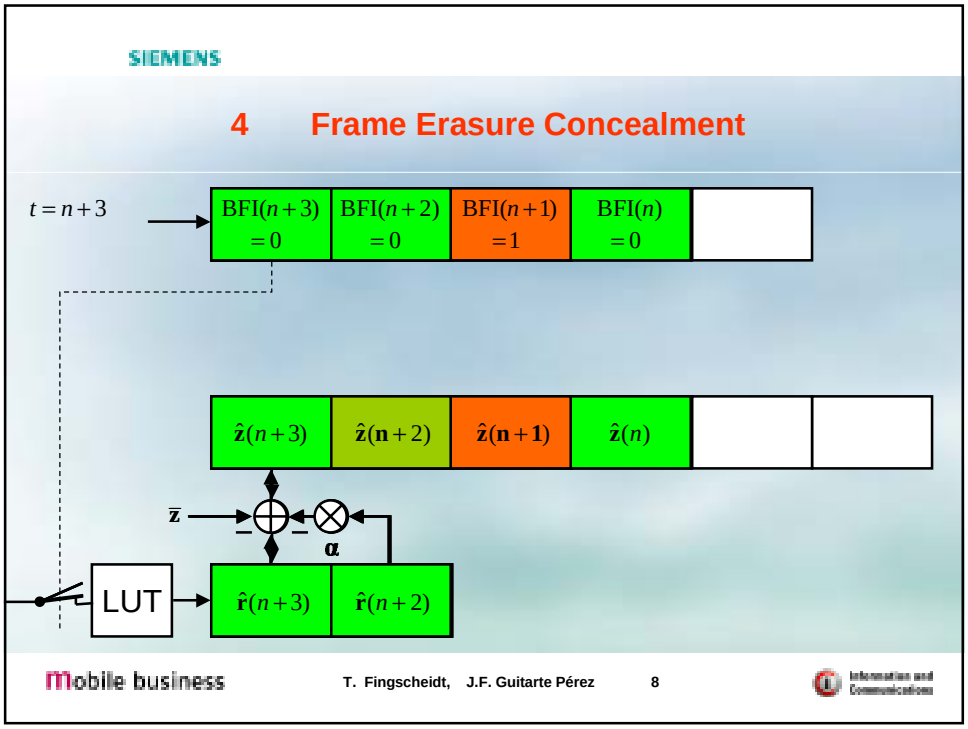
- We did such work and
  we will provide a solution to predictively coded parameters!

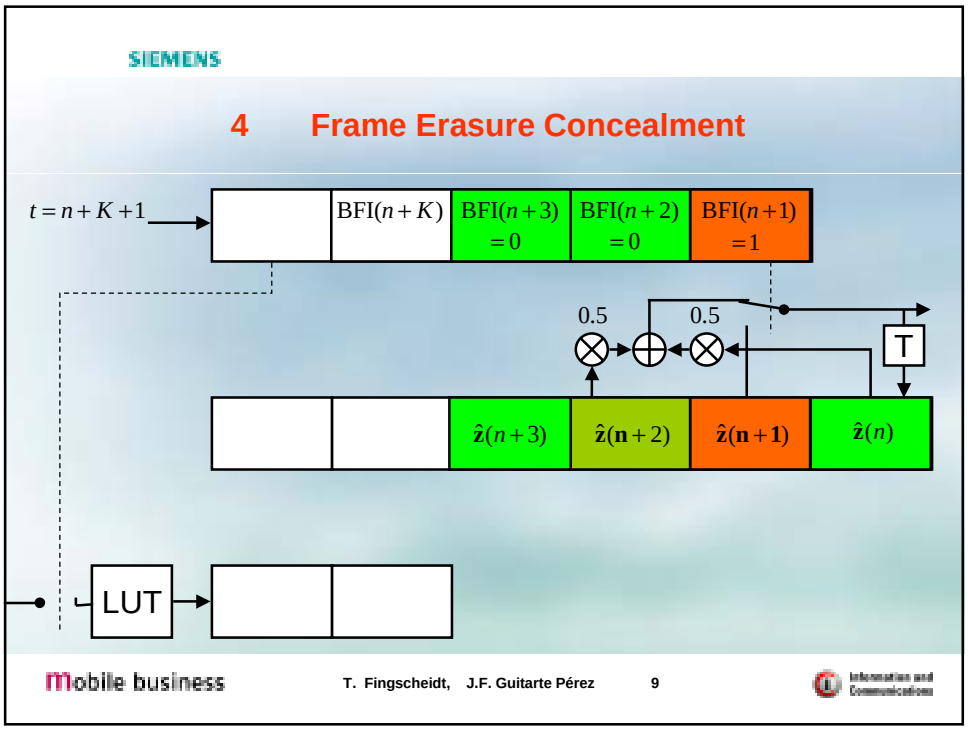
Mobile business T. Fingscheidt, J.F. Guitarte Pérez 4 Information and Communications

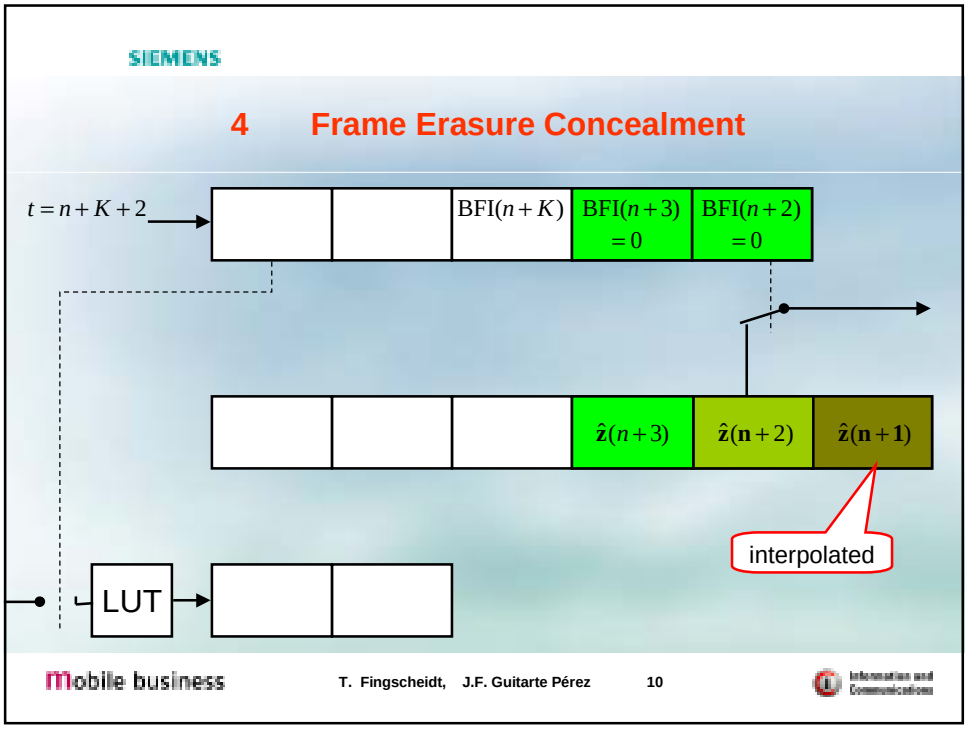

SIEMENS
2 Predictive Coding of Spectral Parameters
Encoder
vector of spectral parameters in frame no. n
$(z_1, z_2, ..., z_N) =$
$\mathbf{z}(n)$
residual
$\mathbf{r}(n)$
LUT
Quanti-zation
$\hat{\mathbf{r}}(n)$
$\hat{\mathbf{z}}(n)$
$\boldsymbol{\alpha}$
T
$\bar{\mathbf{z}}$
Decoder
$\hat{z}_i(n) = \bar{z}_i + \hat{r}_i(n) + \alpha_i \cdot \hat{r}_i(n-1)$
$\bar{\mathbf{z}} = (\bar{z}_1, \bar{z}_2, ..., \bar{z}_N)$
mean values of spectral parameters (ROM table)
$\boldsymbol{\alpha} = (\alpha_1, \alpha_2, ..., \alpha_N)$
prediction coeffs (ROM table)
Mobile business
T. Fingscheidt, J.F. Guitarte Pérez
5

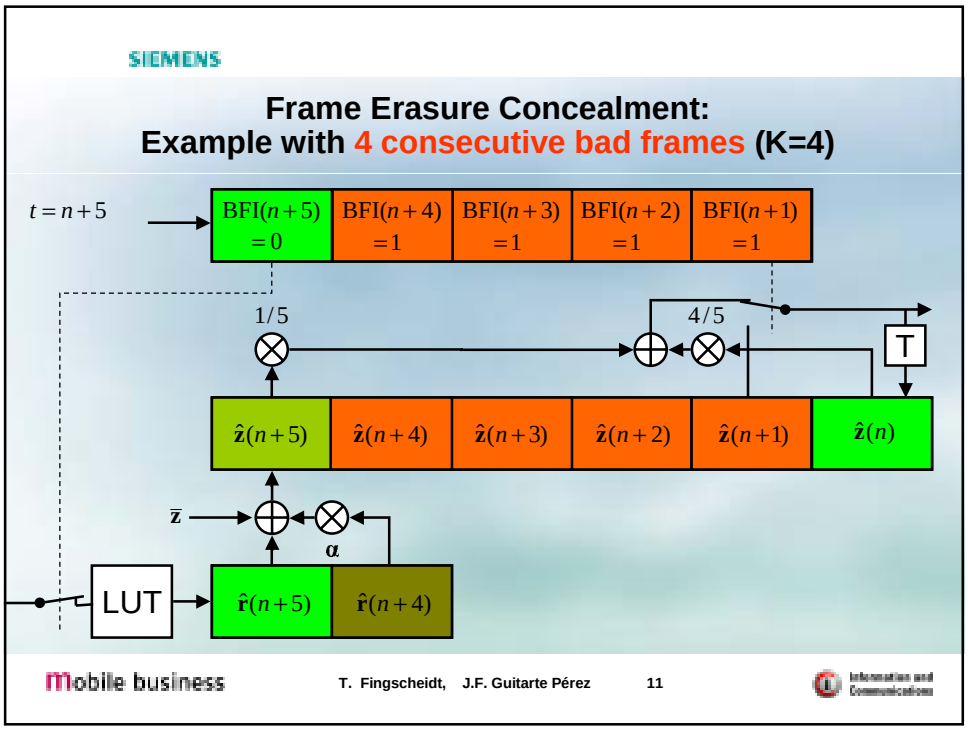

SIEMENS
3 Receiver Model
$t = n + K$
Bits from the channel
Network Module
BFI
$n+K$
$n+...$
$n+2$
$n+1$
n
Frame Erasure Concealment
Speech Decoder
$\hat{\mathbf{r}}()$
$n+K$
$n+...$
1st step
2nd step
Mobile business
T. Fingscheidt, J.F. Guitarte Pérez
6

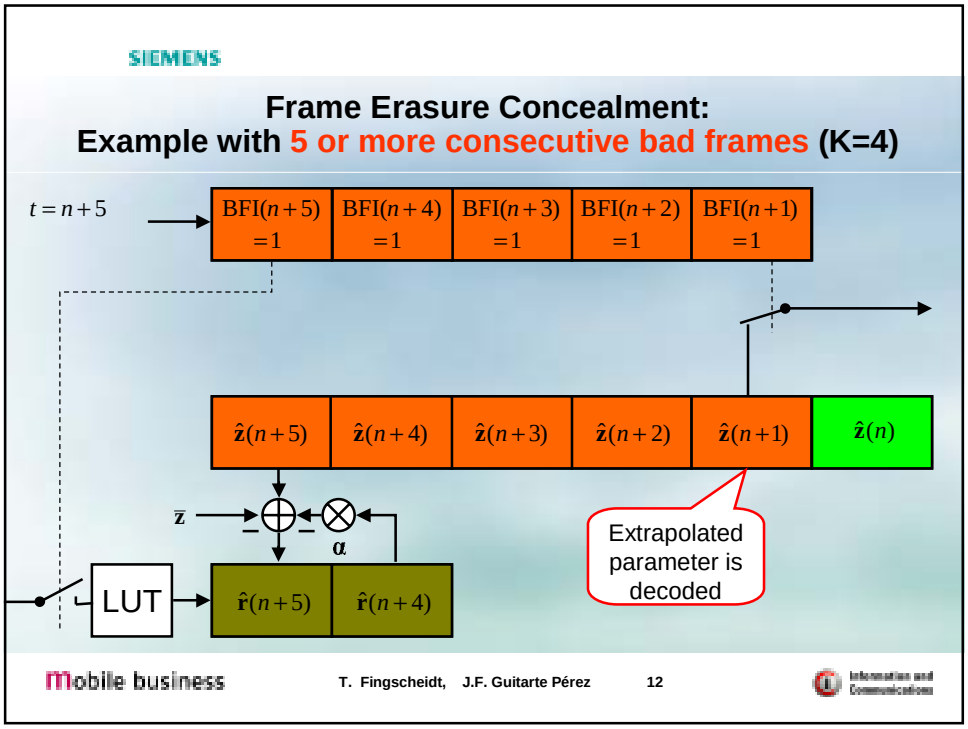

SIEMENS
4 Frame Erasure Concealment
t = n+1
BFI(n+1) = 1
BFI(n) = 0
1 − β
β
ẑ(n+1)
ẑ(n)
z̄
α
LUT
r̂(n+1)
r̂(n)
Mobile business
T. Fingscheidt, J.F. Guitarte Pérez
7


SIEMENS
4 Frame Erasure Concealment
t = n+3
BFI(n+3) = 0
BFI(n+2) = 0
BFI(n+1) = 1
BFI(n) = 0
ẑ(n+3)
ẑ(n+2)
ẑ(n+1)
ẑ(n)
z̄
α
LUT
r̂(n+3)
r̂(n+2)
Mobile business
T. Fingscheidt, J.F. Guitarte Pérez
8


SIEMENS
4 Frame Erasure Concealment
$t = n + K + 1$
BFI$(n+K)$
BFI$(n+3)$ $= 0$
BFI$(n+2)$ $= 0$
BFI$(n+1)$ $= 1$
0.5
0.5
T
$\hat{\mathbf{z}}(n+3)$
$\hat{\mathbf{z}}(\mathbf{n}+2)$
$\hat{\mathbf{z}}(\mathbf{n}+\mathbf{1})$
$\hat{\mathbf{z}}(n)$
LUT
Mobile business
T. Fingscheidt, J.F. Guitarte Pérez
9


SIEMENS
4 Frame Erasure Concealment
$t = n + K + 2$
BFI$(n+K)$
BFI$(n+3)$ $= 0$
BFI$(n+2)$ $= 0$
$\hat{\mathbf{z}}(n+3)$
$\hat{\mathbf{z}}(\mathbf{n}+2)$
$\hat{\mathbf{z}}(\mathbf{n}+\mathbf{1})$
interpolated
LUT
Mobile business
T. Fingscheidt, J.F. Guitarte Pérez
10


SIEMENS
Frame Erasure Concealment:
Example with 4 consecutive bad frames (K=4)
t = n+5
BFI(n+5) = 0
BFI(n+4) = 1
BFI(n+3) = 1
BFI(n+2) = 1
BFI(n+1) = 1
1/5
4/5
T
ẑ(n+5)
ẑ(n+4)
ẑ(n+3)
ẑ(n+2)
ẑ(n+1)
ẑ(n)
z̄
α
LUT
r̂(n+5)
r̂(n+4)
Mobile business
T. Fingscheidt, J.F. Guitarte Pérez
11


SIEMENS
Frame Erasure Concealment:
Example with 5 or more consecutive bad frames (K=4)
t = n+5
BFI(n+5) = 1
BFI(n+4) = 1
BFI(n+3) = 1
BFI(n+2) = 1
BFI(n+1) = 1
ẑ(n+5)
ẑ(n+4)
ẑ(n+3)
ẑ(n+2)
ẑ(n+1)
ẑ(n)
z̄
α
Extrapolated parameter is decoded
LUT
r̂(n+5)
r̂(n+4)
Mobile business
T. Fingscheidt, J.F. Guitarte Pérez
12

SIEMENS

4 Frame Erasure Concealment: Summary

2 steps are done at each frame time instant (indices *i* omitted), here: *t=n+K*:

1) 1st step with frame *t=n+K* (i.e. the just received frame):

BFI(n+K) = 0: $\hat{z}(n+K) = \bar{z} + \hat{r}(n+K) + \alpha \cdot \hat{r}(n+K-1)$ usual prediction

BFI(n+K) = 1: $\hat{z}(n+K) = (1-\beta) \cdot \bar{z} + \beta \cdot \hat{z}(n+K-1)$ extrapolation

$\hat{r}(n+K) = \hat{z}(n+K) - \bar{z} - \alpha \cdot \hat{r}(n+K-1)$ update of residual

2) 2nd step with frame *t=n* (i.e. the frame to be decoded now):

BFI(n) = 0: Decode $\hat{z}(n)$

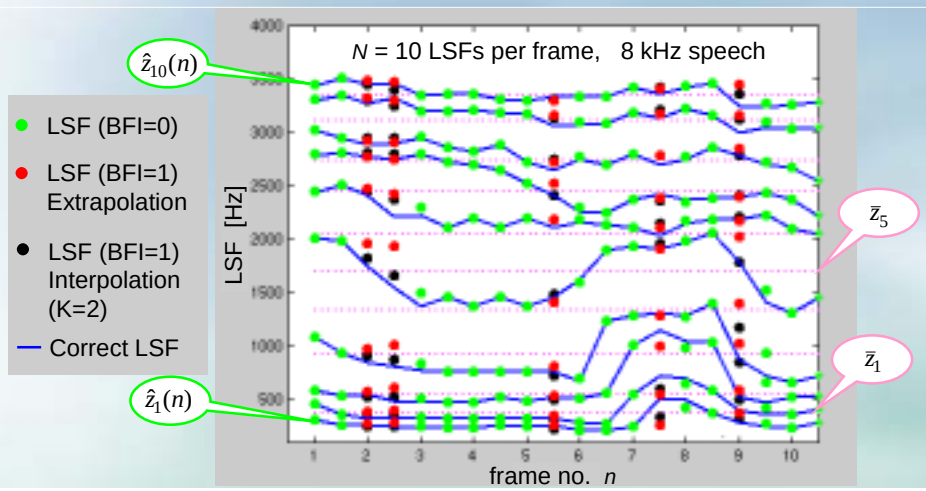
BFI(n) = 1: $\hat{z}(n) = \frac{m-n}{m-n+1} \cdot [\hat{z}(n-1) - \hat{z}(m)] + \hat{z}(m)$ linear interpolation

t=m is the first good frame after *t=n*

Mobile business T. Fingscheidt, J.F. Guitarte Pérez 13 Information and Communications

SIEMENS

5 Simulation Results: Spectral Coefficients (LSFs)



Mobile business T. Fingscheidt, J.F. Guitarte Pérez 14 Information and Communications

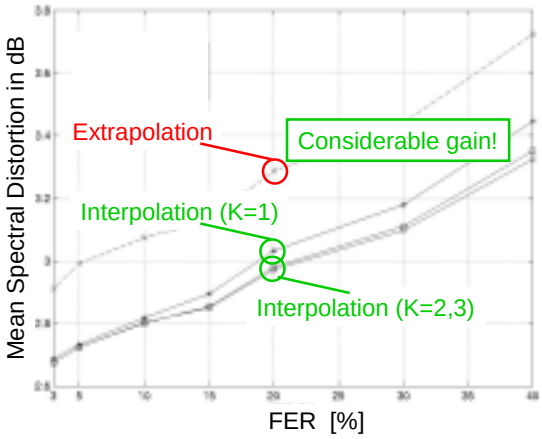
SIEMENS

5 Simulation Results: Spectral Coefficients

Parameters:
Spectral coefficients

Quality Measure:
Mean spectral distortion
(only in bad and 1st good frames)

Channel:
Random frame loss



Mobile business T. Fingscheidt, J.F. Guitarte Pérez 15 Information and Communications

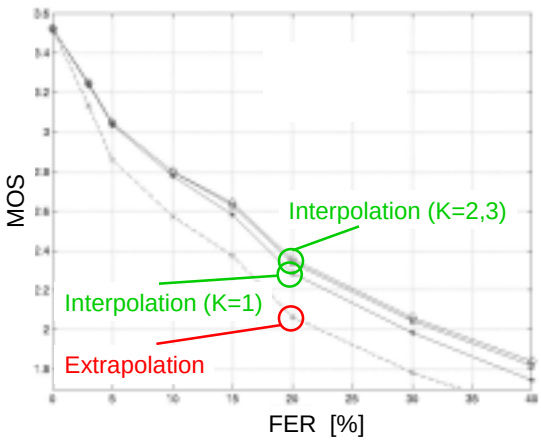
SIEMENS

5 Simulation Results: All Parameters of AMR @ 10.2 kbs

Parameters:
all

Quality Measure:
Mean opinion score
(MOS)

Channel:
Random frame loss



Mobile business T. Fingscheidt, J.F. Guitarte Pérez 16 Information and Communications

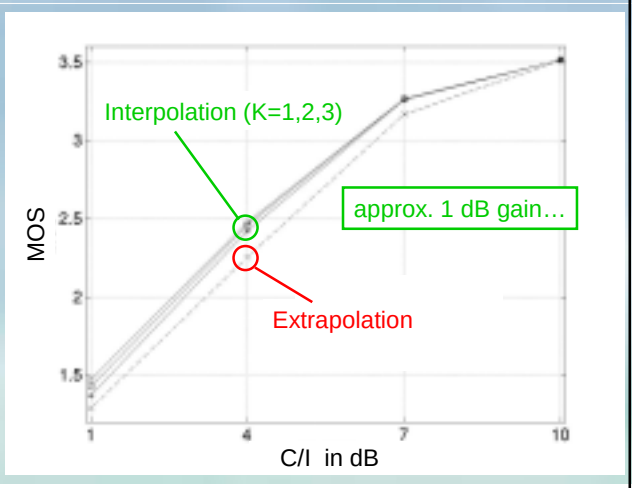
SIEMENS

5 GSM Simulation Results: All parameters

Parameters:
all

Quality Measure:
Mean opinion score (MOS)

Channel:
GSM Full-rate channel @ 3 km/h, IFH



SIEMENS

Conclusions

- VoIP & Streaming:
  It is possible / necessary to introduce a decoder buffer and thus add. delay
- Interpolation instead of extrapolation yields better results!
- Interpolation of predictively coded parameters …
  - … is possible with 1 frame delay only!
  - … yields most of the gain already with 1 frame delay!
- With very simple means …
  - … a 1dB gain is possible in circuit-switched GSM FR channel
  - … considerable gains are possible in random frame loss channels

© Information and Communication Mobile

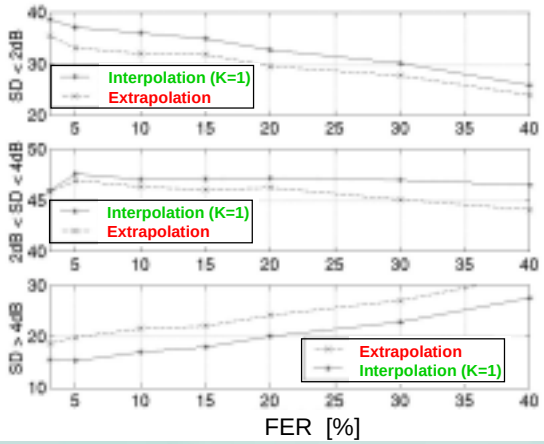
SIEMENS

5 Simulation Results: Spectral Coefficients (LSFs)

Parameters:
LSF coefficients

Quality Measure:
Outliers of spectral distortion (SD)
(only in bad and 1st good frames)

Channel:
Random frame loss



© Information and Communication Mobile