Voice over the Internet
(the basics)

Outline

- Basics about voice encoding
- Packetization trade-offs
- Architecture of basic VoIP tool
- Playback buffer (jitter buffer)
  - Adaptive playback buffers?
- How to deal with packet losses and late packets?

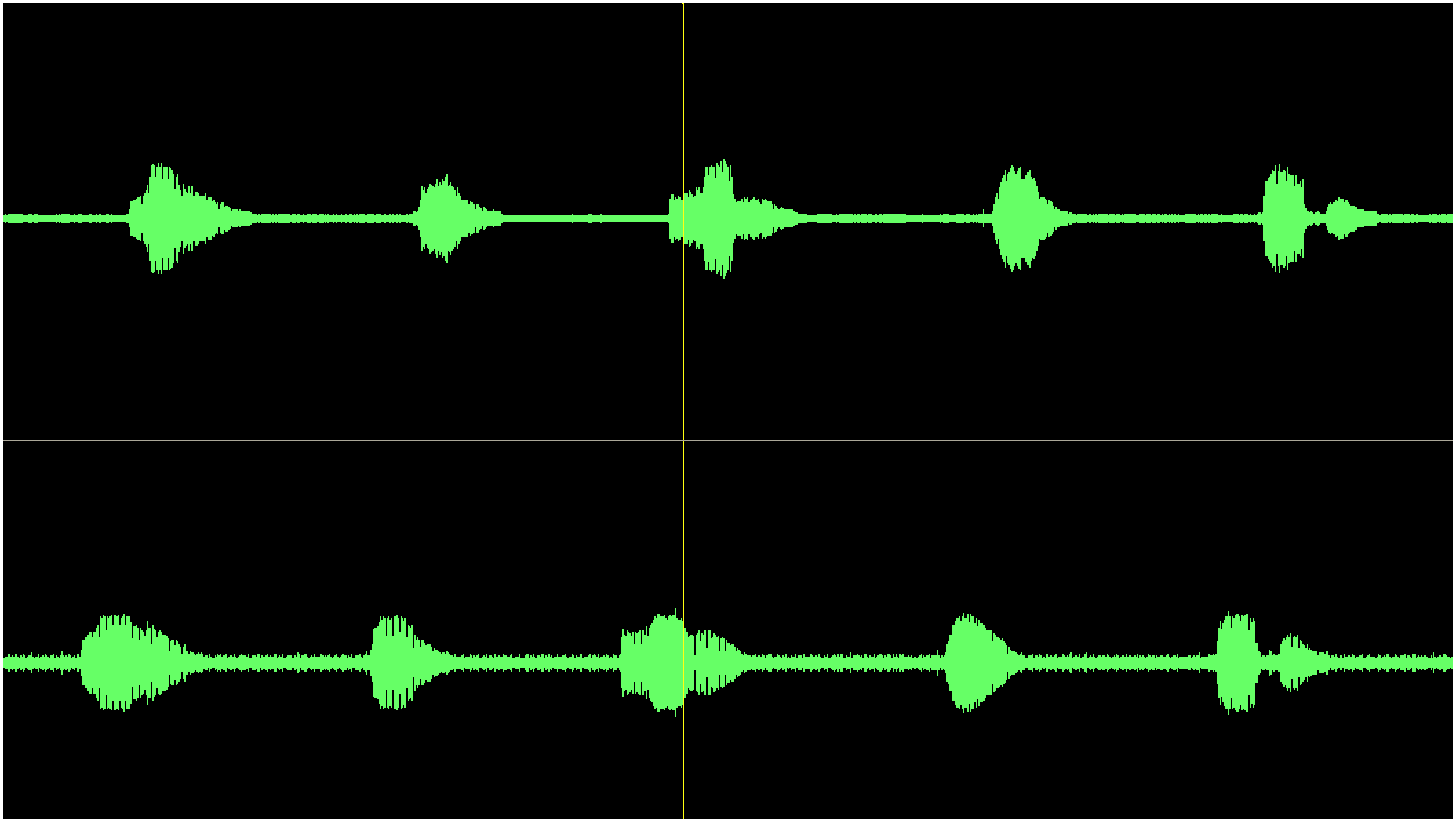
Voice over the Internet

- Includes computer2computer voice applications (like Skype, VoIPBuster, etc)
- + VoIP services
- + Telephony Routing over IP (TRIP)
- Includes “off-net” calls (calls to PSTN phones)

Reading-1

- "Voice over Internet Protocol (VoIP)" by Bur Goode, published at IEEE Proceedings, Sep'02

It all starts from an analog signal



Codecs

Codec	Algorithm	*Frame Size/ Lookahead*	Usual Rate	Comments
G.711	PCM	0.125 ms/0	64 Kb/s	Universal use
G.722		0.125 ms/1.5 ms	48, 56 or 64 Kb/s	Wideband coder
G.726	ADPCM	0.125 ms/0	32 Kb/s	High quality, low complexity
G.728	LD-CELP	0.625 ms/0	16 Kb/s	High quality in tandem; Recommended for cable
G.729(A)	CS-ACELP	10 ms/5 ms	8 Kb/s	Widespread use
G.729e	Hybrid CELP	10 ms/5 ms	11.8 Kb/s	High quality/complexity; Recommended for cable
G.723.1(6.3)	MPC-MLQ	30 ms/7.5 ms	6.3 Kb/s	Video conferencing origin
G.723.1(5.3)	ACELP	30 ms/7.5 ms	5.3 Kb/s	Video conferencing origin
IS-127	RCELP	20 ms/5ms	Var. 4.2 Kb/s avg.	
AMR	ACELP	20 ms	Var. 4.75-12.2 Kb	Compatible w. No. Amer. & Japanese digital cellular, WCDMA (not CDMA2000); Nokia IPR

How does PCM work?

- Voice spectrum extends to about 3-4KHz
- According to Nyquist's rate, a sampling frequency of 8KHz should be enough to completely reconstruct the original voice signal from the sampled signal
- PCM uses 8 bits per sample (64kbps)
- Frame size?
  - G.711 uses 125msec (too large for packet voice)
  - G.729 uses 10msec

Listen to the various codecs and judge for yourself

- http://www.data-compression.com/speech.shtml

    (look at bottom of this page)

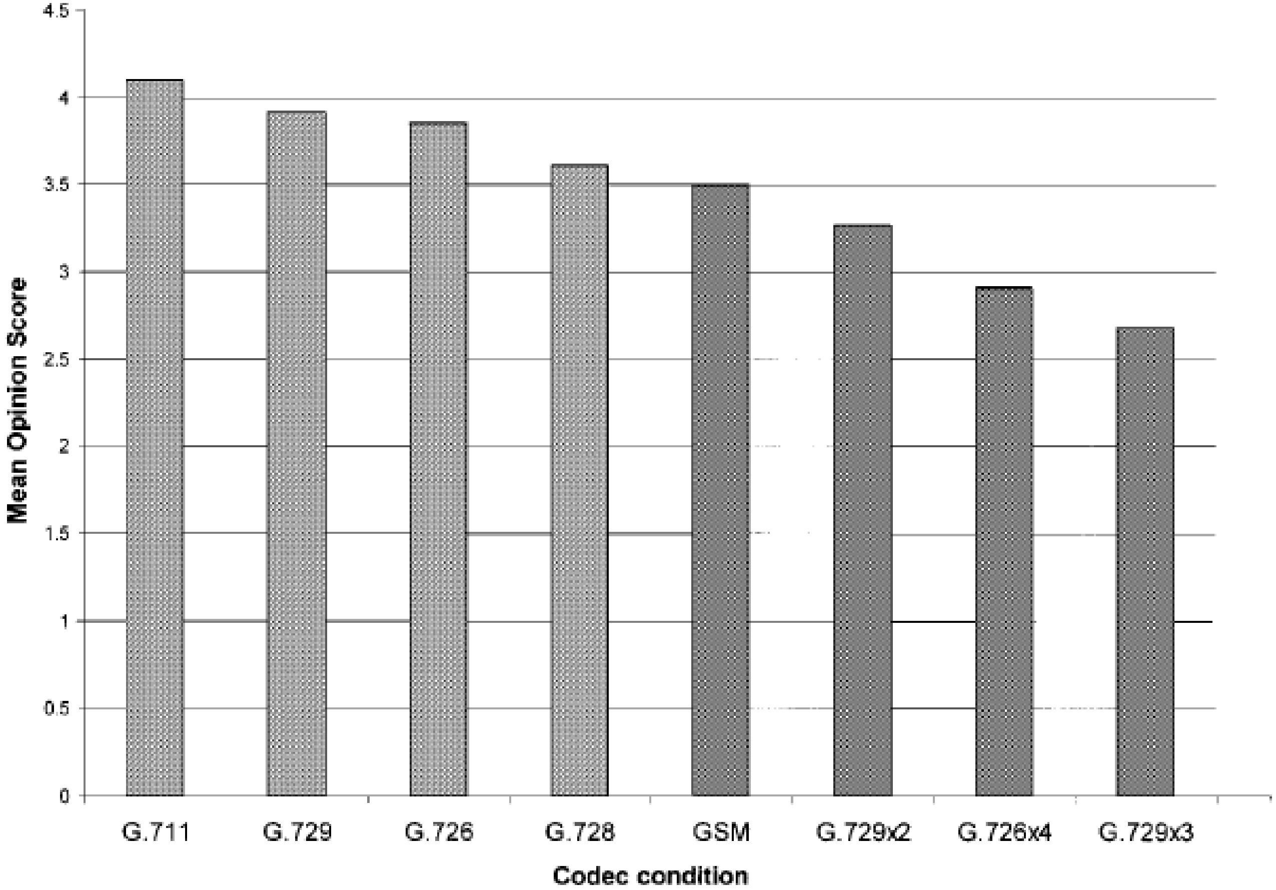
Popular recent codecs for VoIP

- See GlobalIPSound

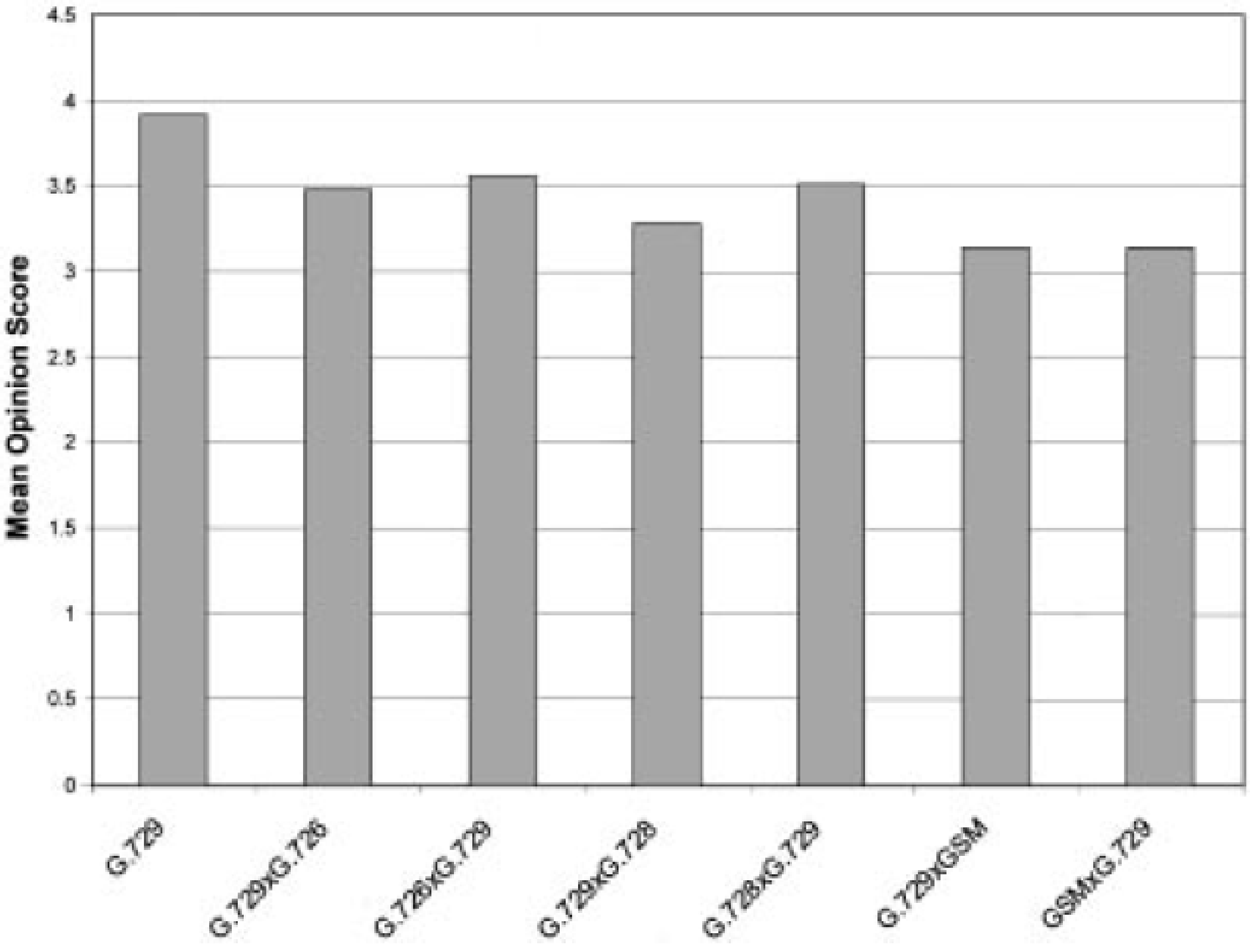
  (http://www.gipscorp.com/products/demos.php)
  - Wide band codecs (50-8,000 Hz)
  - iLBC (packetization: 20 and 30 msec, bitrate: 15.2 kbps and 13.3 kbps)
    - Free, open-source
    - No error propagation when lost frame (problem with LPC)
  - iSAC (proprietary – best codec currently?)
    - PACKET SIZE Adaptive, 30 - 60 ms
    - BIT RATE Adaptive and variable, range 10 - 32 kbps
    - SAMPLING RATE 16 kHz
    - AUDIO BANDWIDTH 8 kHz

MOS scores

- Also look at the effect of "codec concatenation" (e.g., G.729*3)



Effects of transcoding



Packetization tradeoffs

- R: encoding rate (bps)
- H: header size per packet (bits)
  - E.g., 40B for RTP/UDP/IP packet
- S: packetization period or sample duration (sec)
- BW: voice transmission requirement
  - BW = R + H/S
  - How can you decrease BW?
  - Lower R means more complex codec, more correlations across successive packets
  - Higher S means more delay at sender and larger sensitivity to packet losses

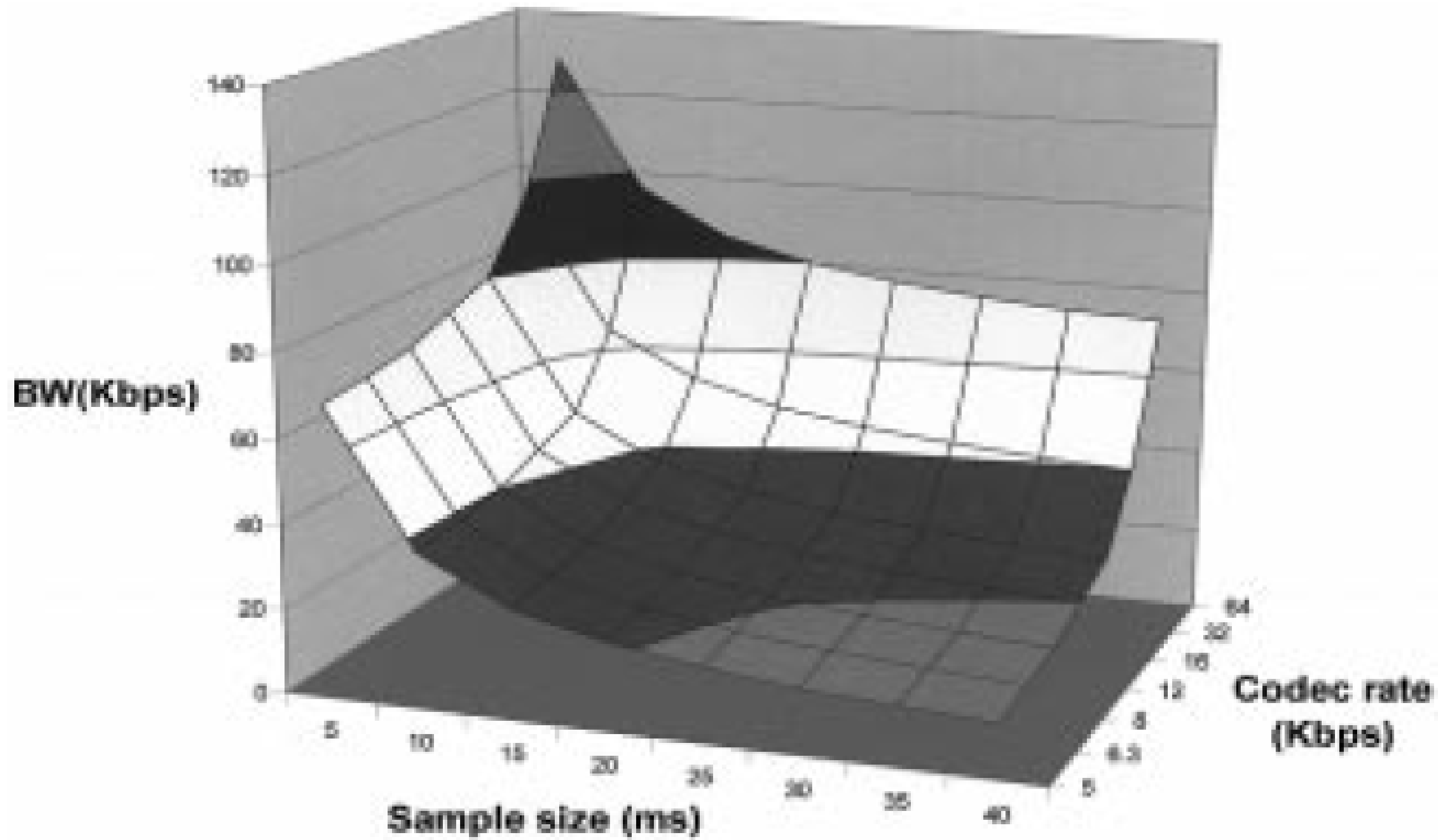


Fig. 5. The varying bands, from top to bottom, represent the following VoIP bandwidth requirements (40-byte headers): 120–140, 100–120, 80–100, 60–80, 40–60, 20–40, and 0–20.

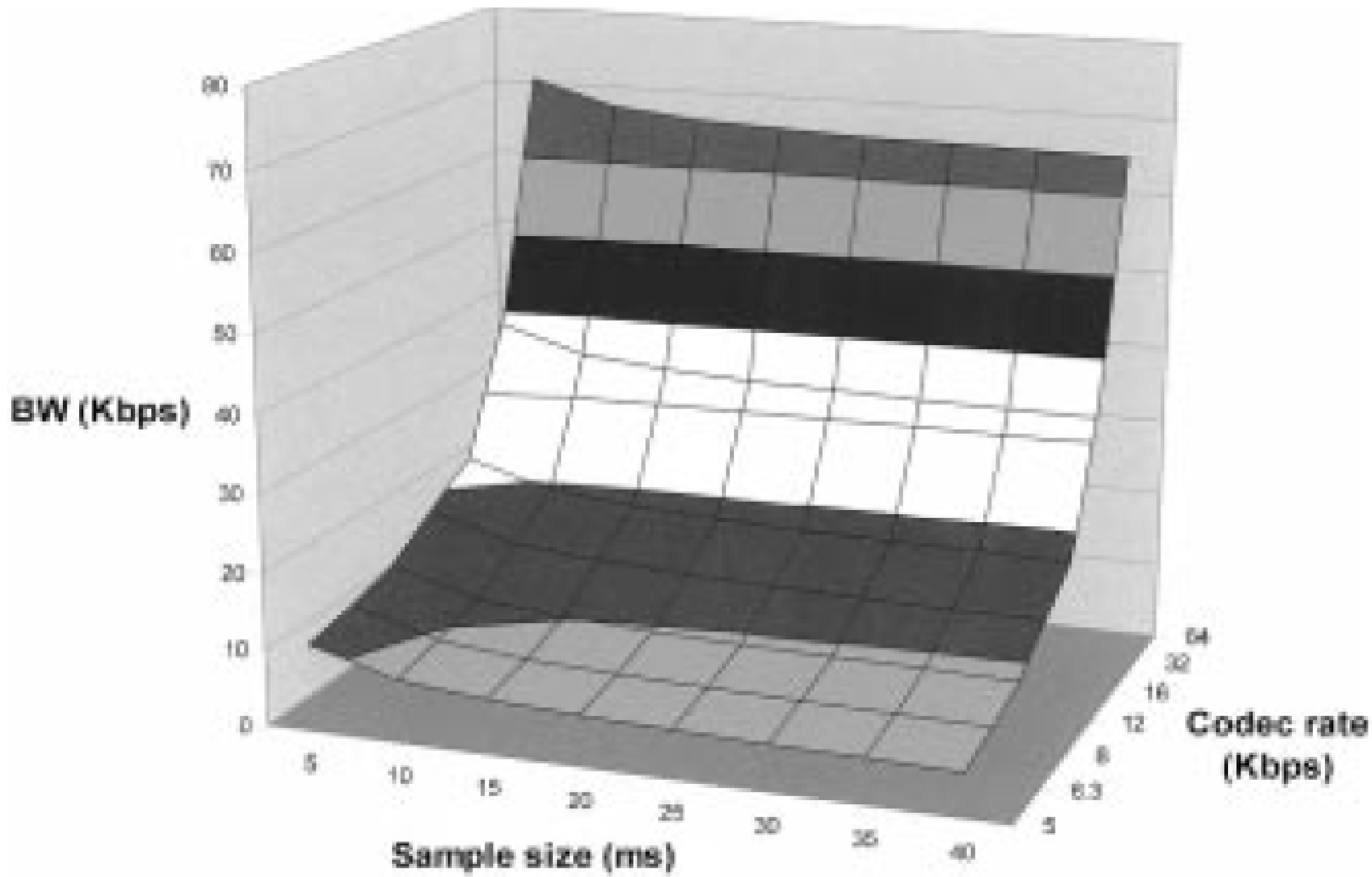


Fig. 6. From top to bottom, varying bands represent the following VoIP bandwidth requirements (4-byte headers): 70–80, 60–70, 50–60, 40–50, 30–40, 20–30, 10–20, 0–10.

Control Mechanisms for Packet Audio in the Internet

Jean-Chrysostome Bolot   Andrés Vega-García
INRIA
B. P. 93
06902 Sophia-Antipolis Cedex
France
{bolot, avega}@sophia.inria.fr

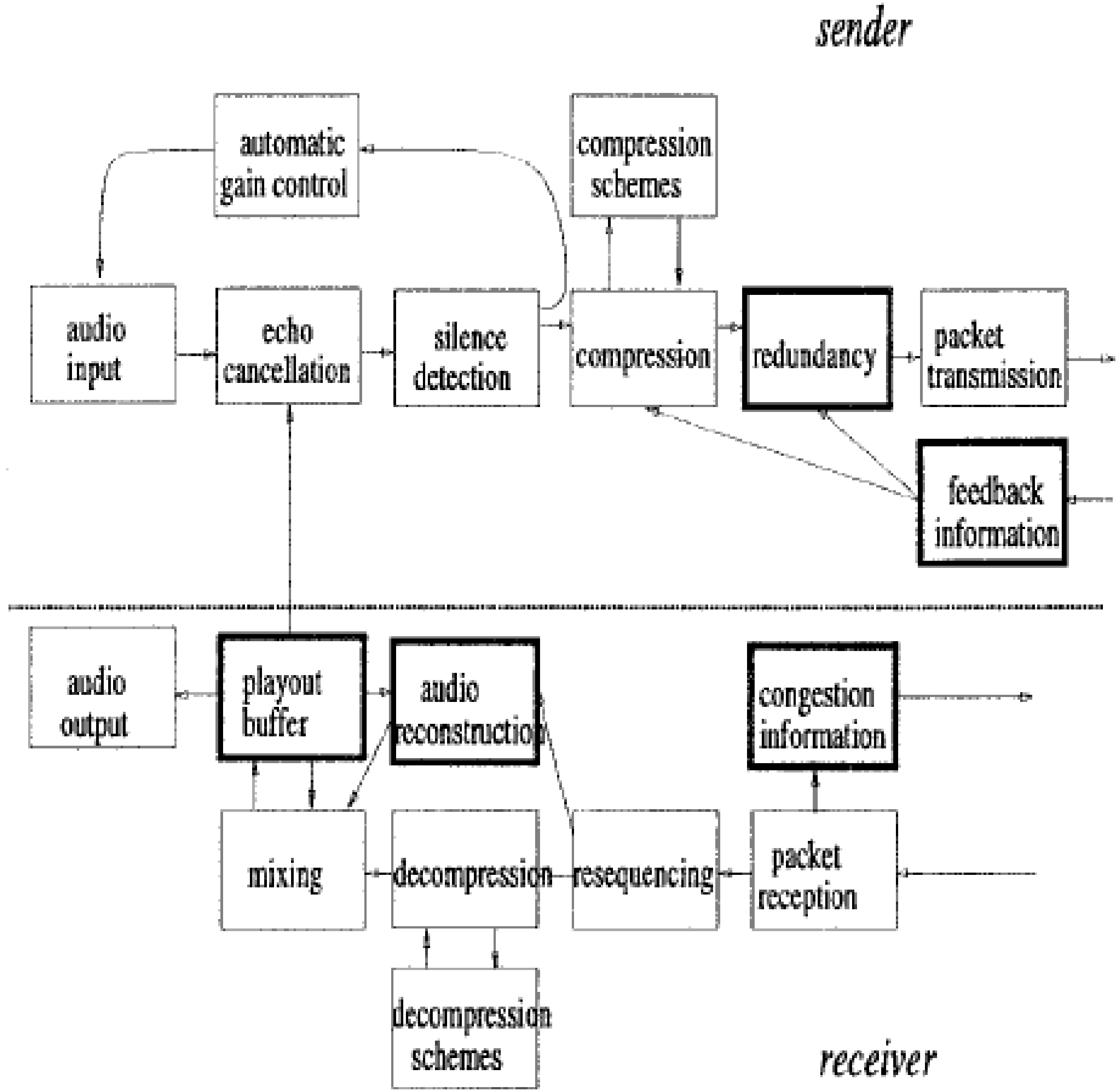


Figure 1: Structure of the audio tool

Network effects

- One-way delay between sender/receiver
  - Includes encoding, packetization, transmission, propagation, queueing, jitter compensation, decoding
  - Typically, acceptable if < 150msec for domestic calls and < 400msec for international
    - Depends on call's interactivity
  - What can we do to reduce packet delay?

Network effects (cont')

- Packet losses
  - Low-bitrate codecs are very sensitive to packet losses (why?)
  - Should we do retransmissions?
  - Should we do Forward-Error-Correction?
  - Or just, packet loss concealment? How?
- Delay variation or jitter
  - Jitter compensation buffer at receiver
  - How large should this buffer be?
  - Losing vs discarding packets
  - Delay budget calculations
- Insufficient network capacity
  - Rate adaptation (use multiple codecs)

Adaptive Playout Mechanisms for Packetized Audio Applications in Wide-Area Networks*

Ramachandran Ramjee, Jim Kurose, Don Towsley
Department of Computer Science
Univ. of Massachusetts
{ramjee,kurose,towsley}@cs.umass.edu

Henning Schulzrinne
GMD-Fokus
Berlin, Germany
hgs@fokus.gmd.de

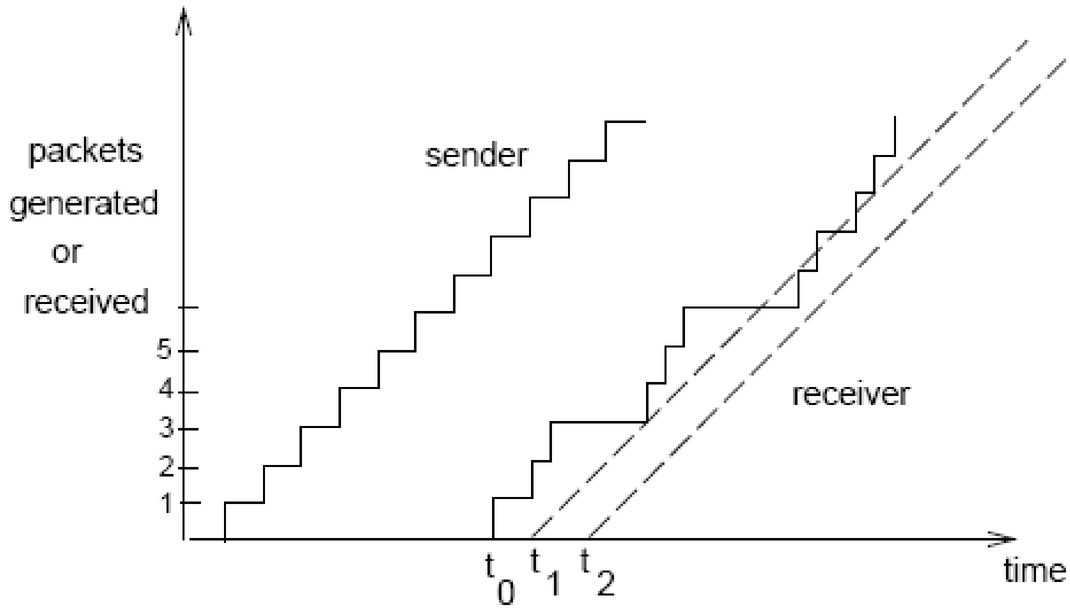


Figure 1: Generation and reconstruction of packetized voice

Delay budget

Delay Source (G.729)	On-net Budget (ms)
Device Sample Capture	0.1
Encoding Delay (Algorithmic Delay + Processing Delay)	17.5
Packetization/ Depacketization Delay	20
Move to Output Queue/Queue Delay	0.5
Access (up) Link Transmission Delay	10
Backbone Network Transmission Delay	Dnw
Access (down) Link Transmission Delay	10
Input Queue to Application	0.5
Jitter Buffer	60
Decoder Processing Delay	2
Device Playout Delay	0.5
Total	121.1 + Dnw

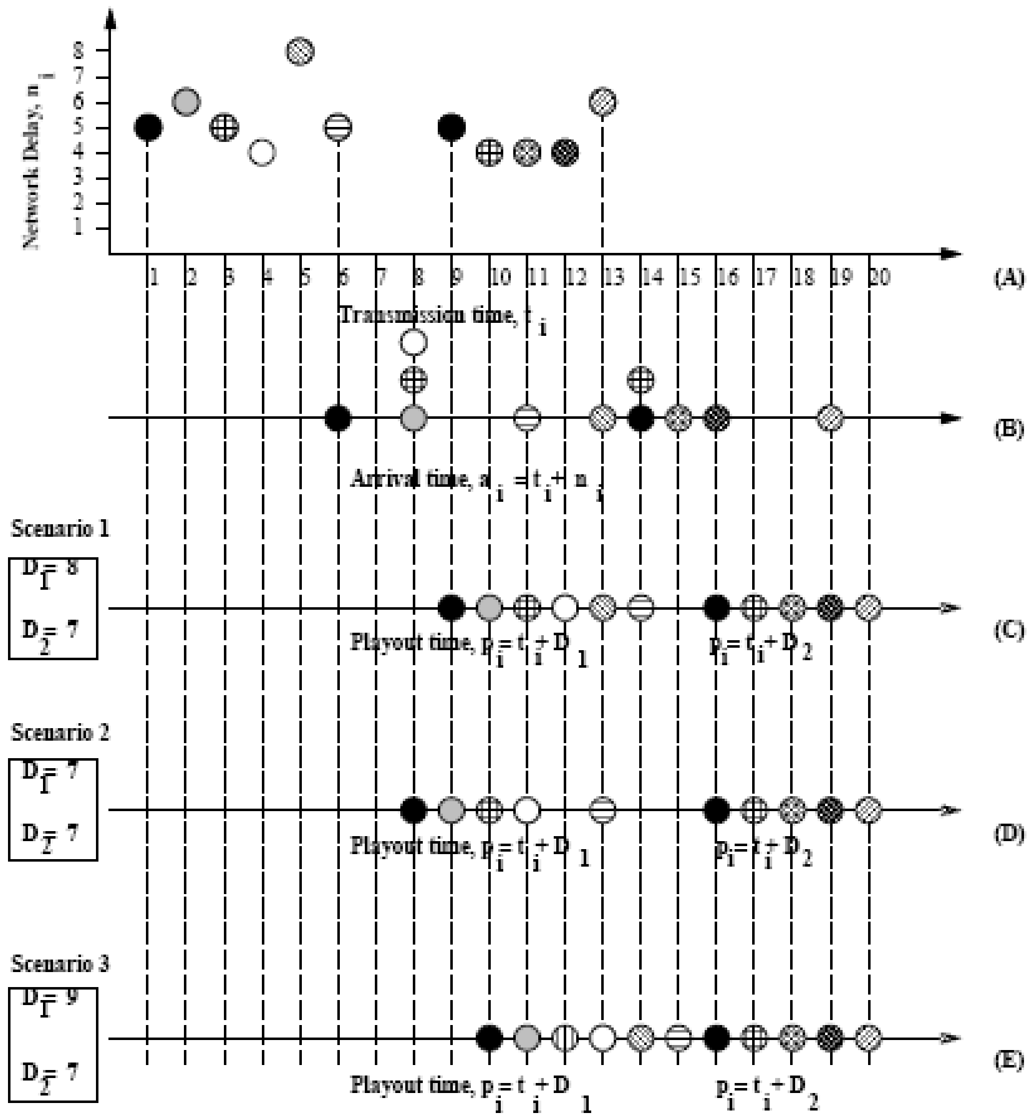


Figure 3: Example illustrating playout mechanisms

A Survey of Packet Loss Recovery Techniques for Streaming Audio

Colin Perkins, Orion Hodson, and Vicky Hardman
University College London

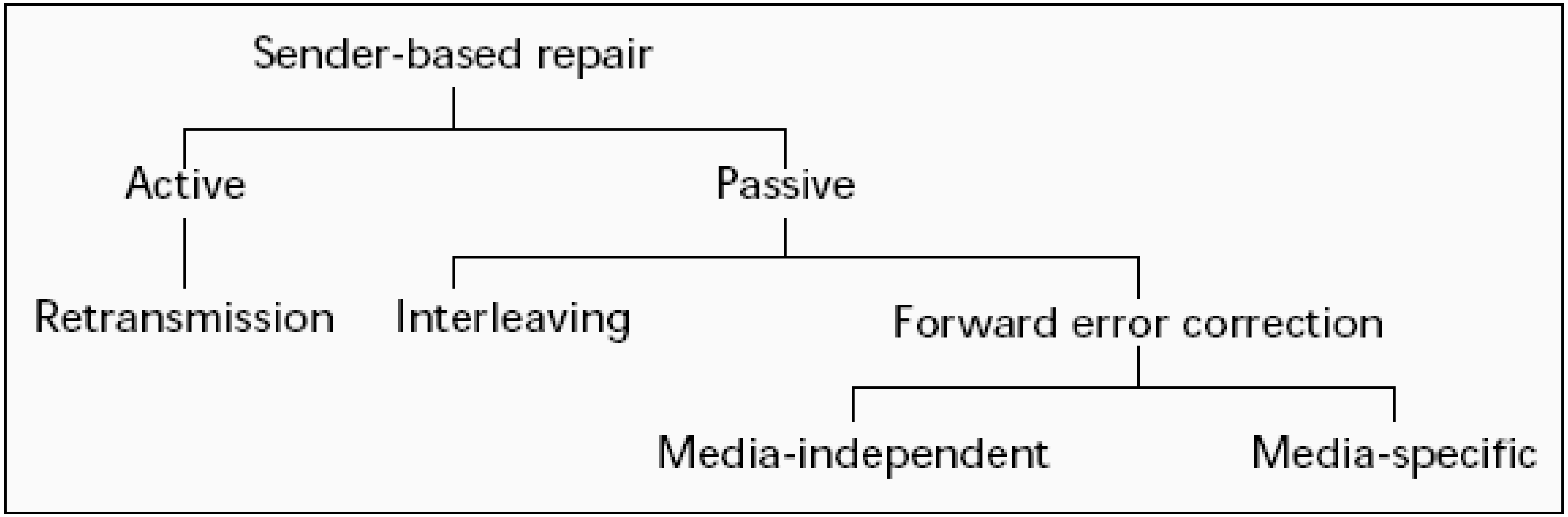


Figure 3. *A taxonomy of sender-based repair techniques.*

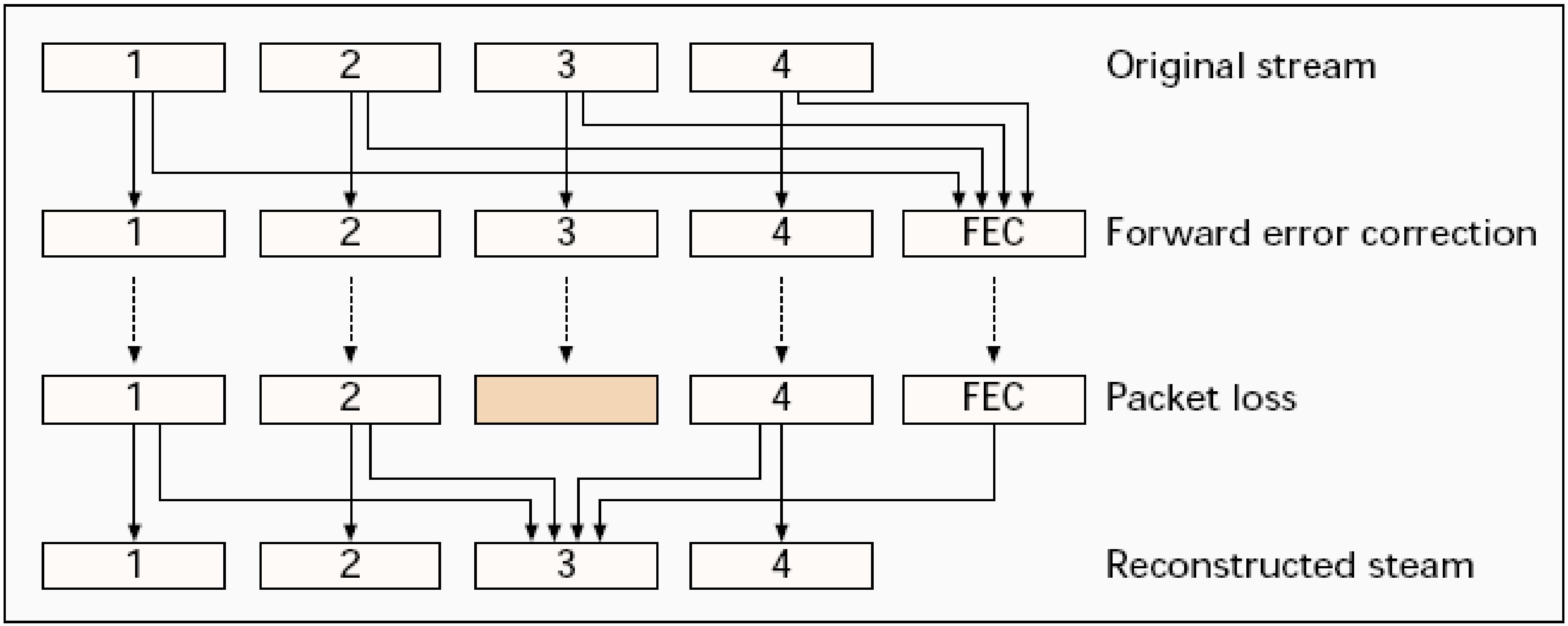


Figure 4. *Repair using parity FEC.*

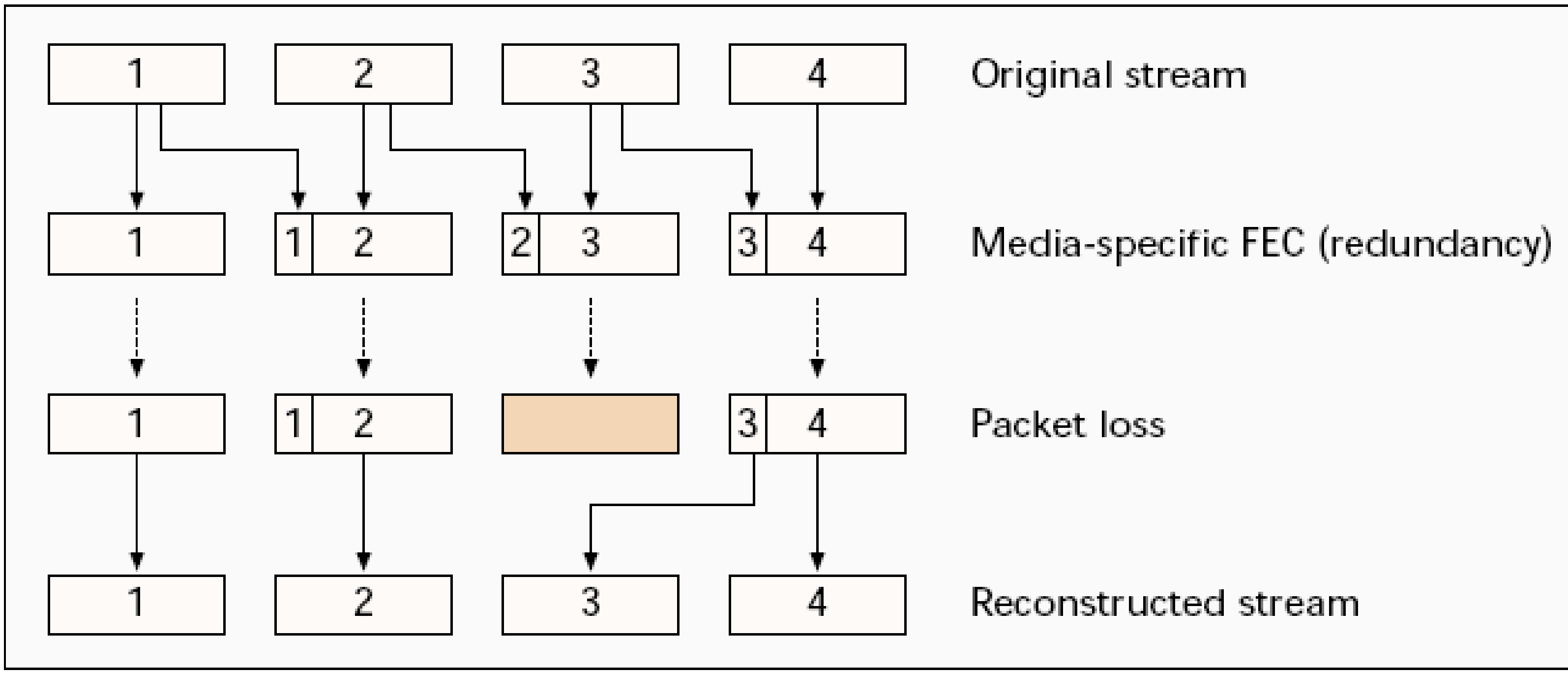


Figure 5. *Repair using media-specific FEC.*

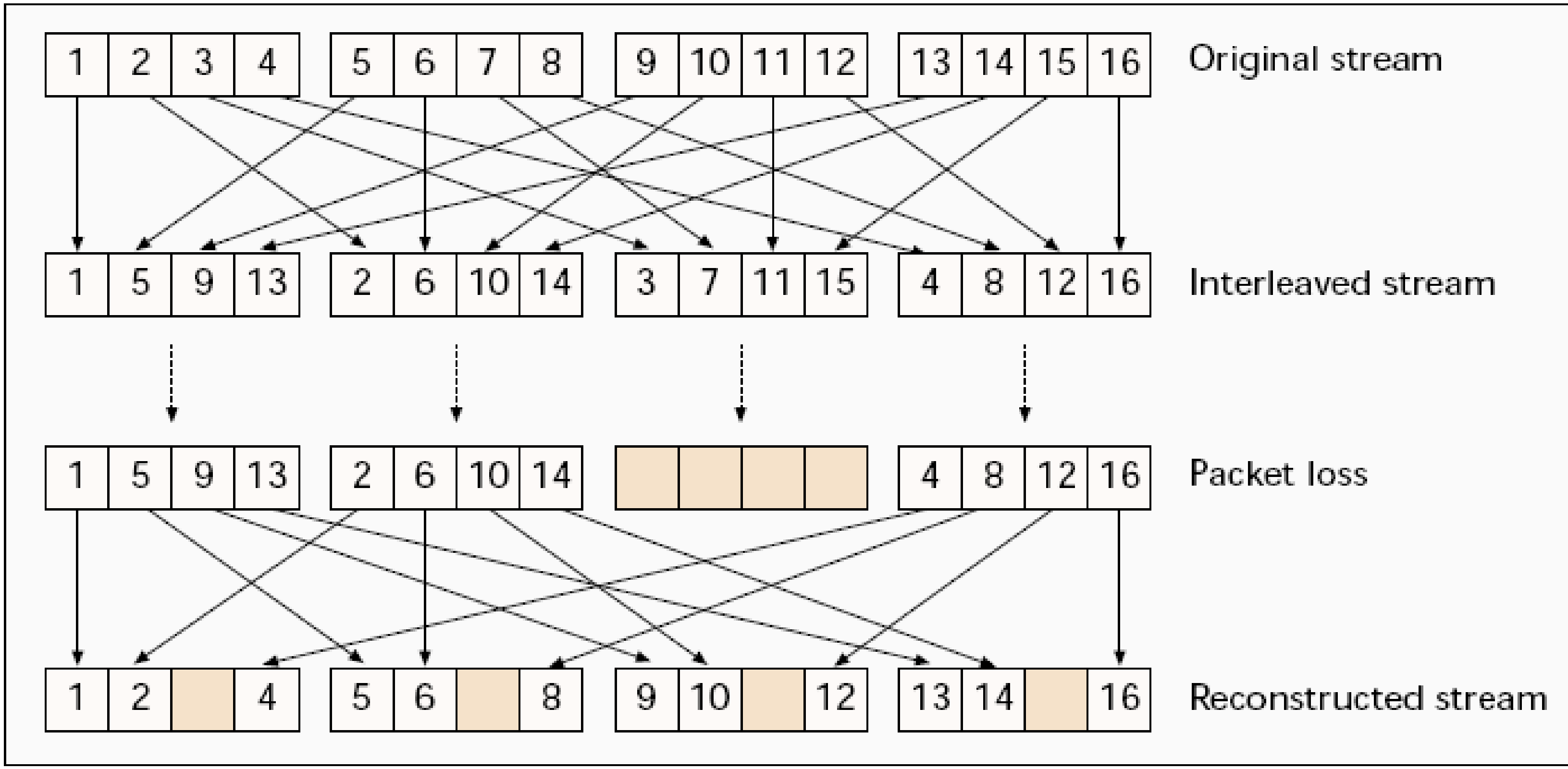


Figure 6. *Interleaving units across multiple packets.*

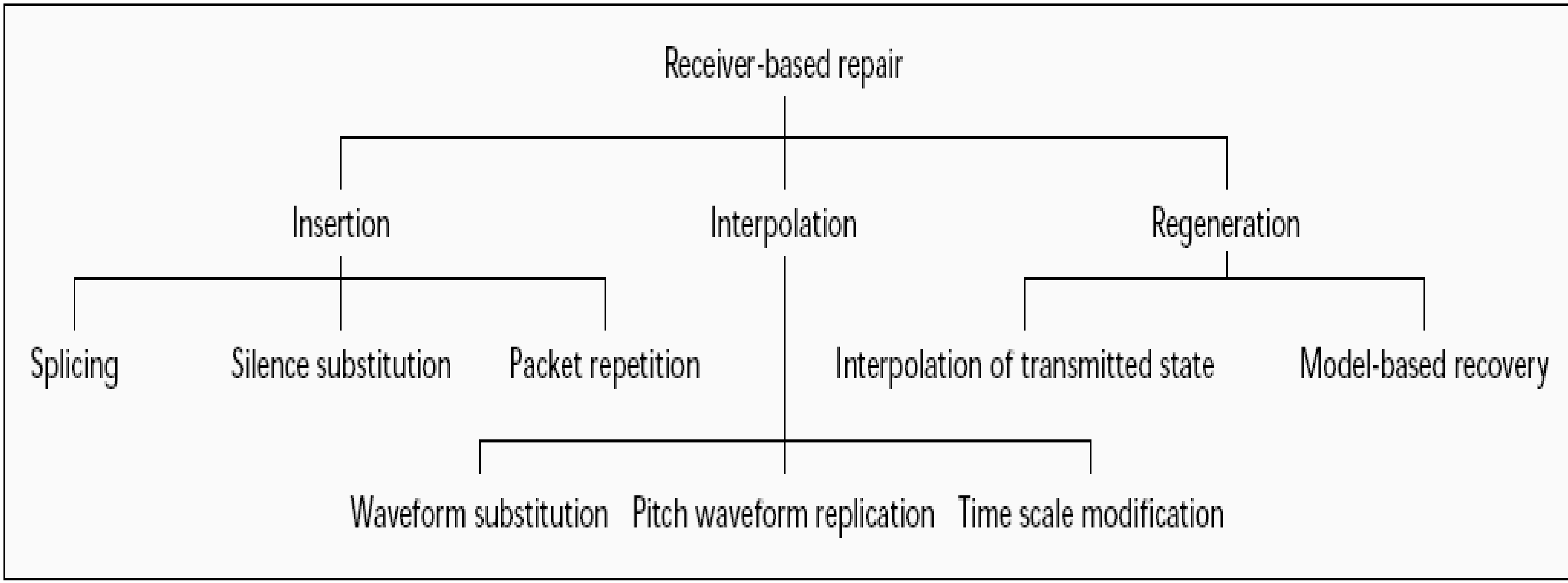


Figure 7. *A taxonomy of error concealment techniques.*

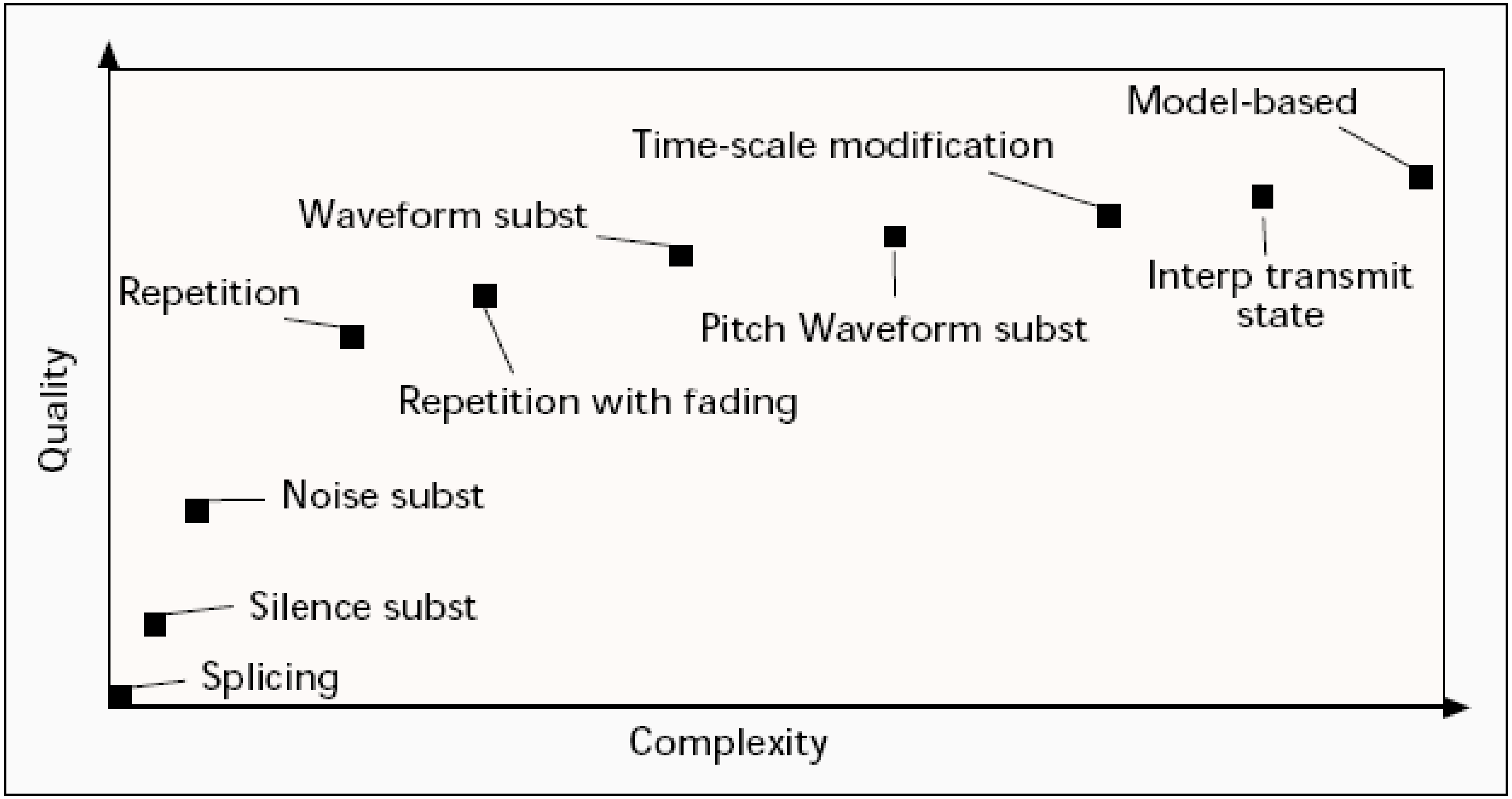
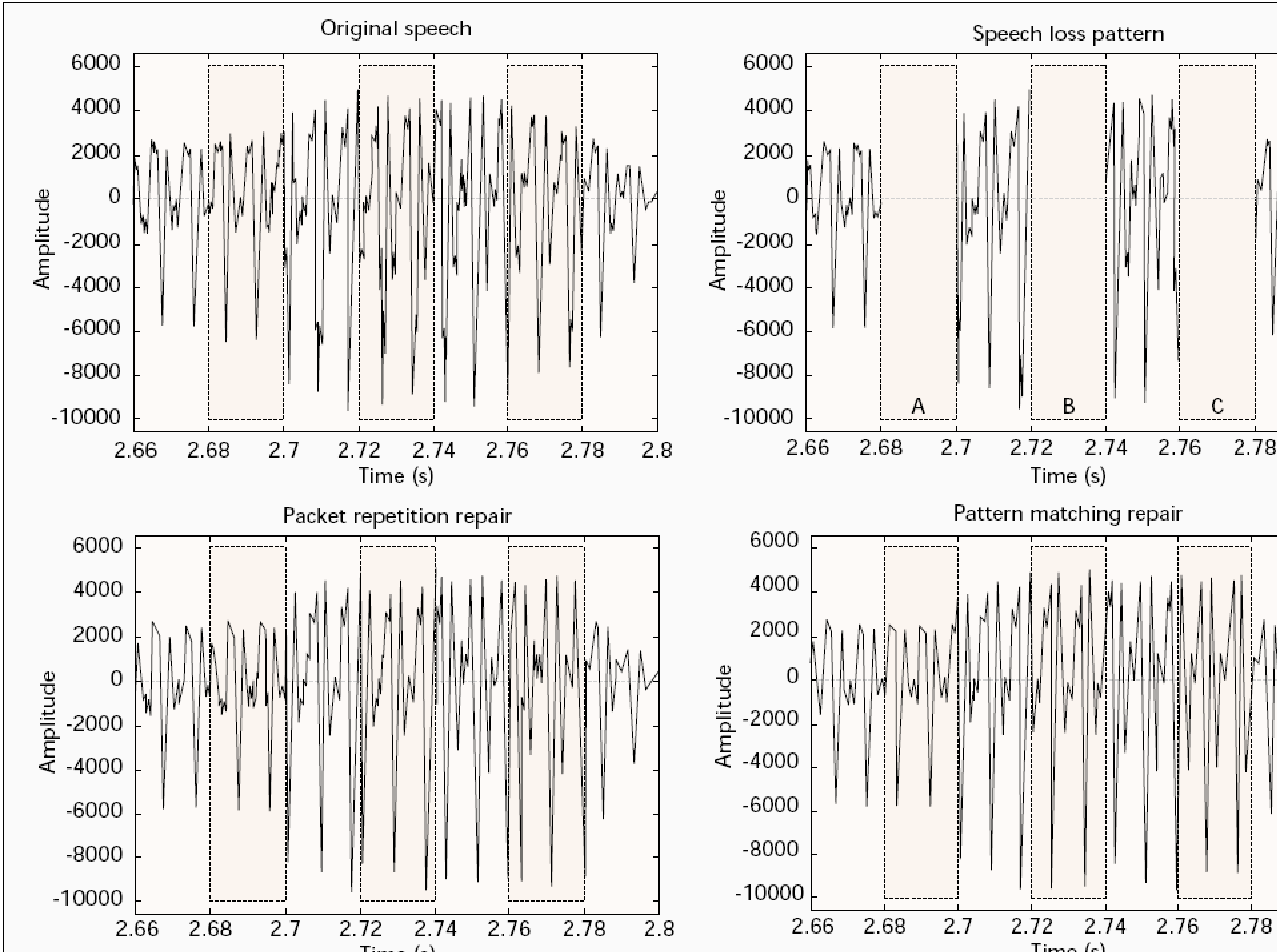


Figure 8. *Rough quality/complexity trade-off for error concealment.*


Original speech
Speech loss pattern
Packet repetition repair
Pattern matching repair
Amplitude
Time (s)
A
B
C
6000
4000
2000
0
-2000
-4000
-6000
-8000
-10000
2.66
2.68
2.7
2.72
2.74
2.76
2.78
2.8