

Introduction to Speech/Voice Coding

Jerry D. Gibson

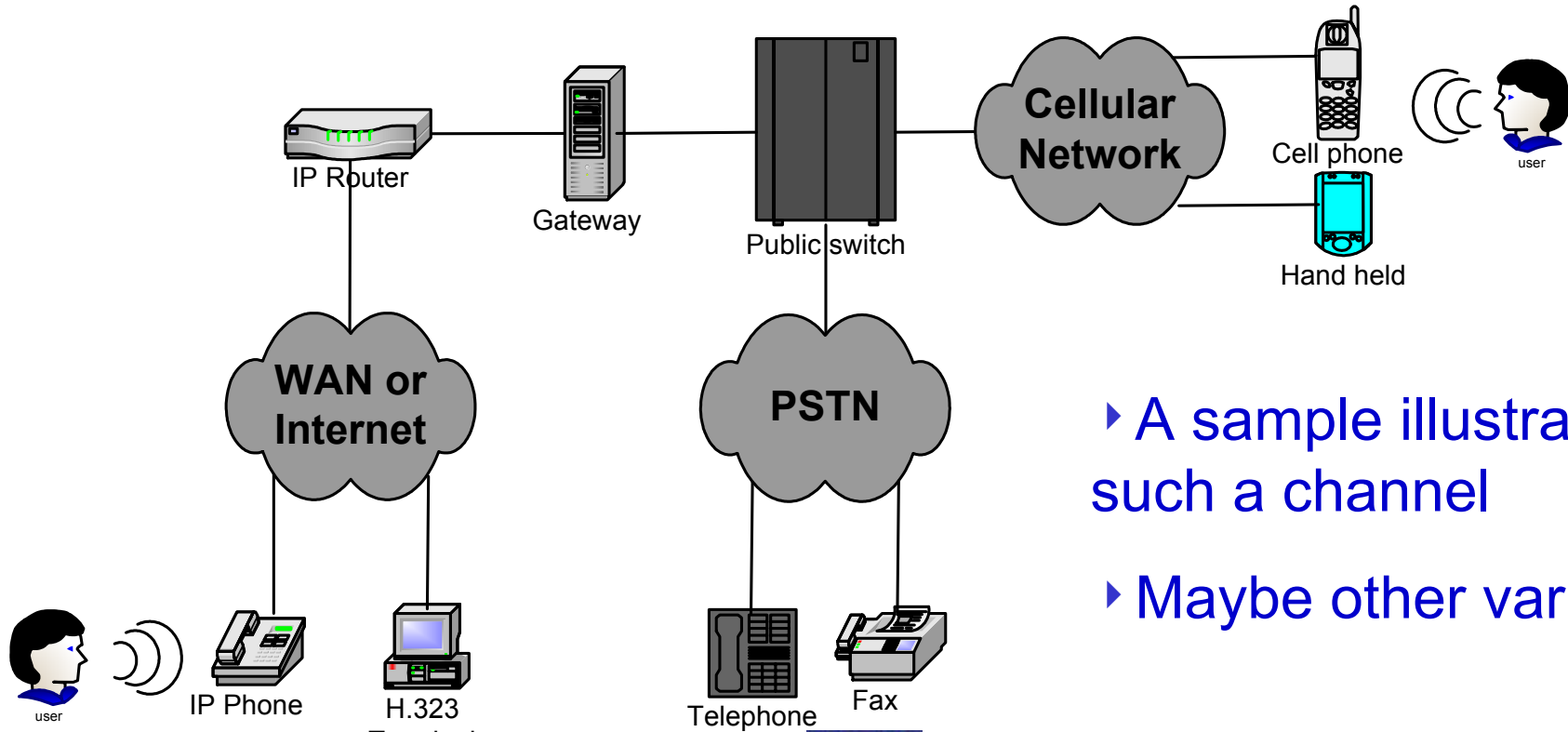
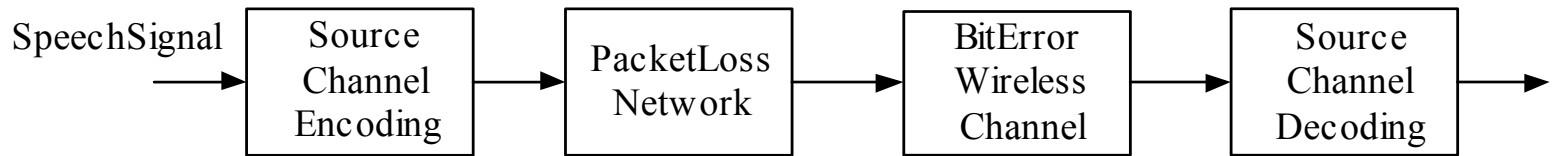
January 21, 2005



Applications of Speech Coding

- Wireline Telephony
- Videoconferencing
- Digital Cellular
- IP Telephony
- Voice Mail
- Speech Storage

One Example: Voice Transmission over Tandem Networks



- ▶ A sample illustration of such a channel
- ▶ Maybe other variations.

January 21, 2005 Terminal



Uncompressed Bit Rates for Speech and Audio

<u>Source</u>	<u>Bandwidth (Hz)</u>	<u>Sampling Rate</u>	<u>Bits Per Sample</u>	<u>Bit Rate</u>
Telephone Speech	200-3400	8 ks/s	12	96 kb/s
Wideband Speech	50-7000	16 ks/s	14	224 kb/s
Wideband Audio (2 Channels)	20-20000	44.1 ks/s	16/ch	1.412 Mb/s (2 channels)

Original

Classical music

Music+voice

8KHz



16KHz



Voice Coding Issues

- Classification by Input Voice Bandwidth
 - Narrowband: 200-3400 Hz (telephony)
 - Wideband: 50-7000 Hz
 - Audio: 20-20000 Hz, not used for voice
- Quality (naturalness) and Intelligibility
- Delay
- Complexity

What is Acceptable Delay?

- One way transmission time (processing and propagation delay)
- 0 to 150 ms: Acceptable for most user applications
- 150 to 400 ms: Acceptable depending upon the transmission time impact
- Above 400 ms: Unacceptable for general network planning purposes.

Subjective Quality Assessment

- Absolute Category Rating (ACR) Tests (ITU-T Recommendation P.800)
 - Listening Quality Assessment
 - Listeners rate the speech as Excellent (5), Good (4), Fair (3), Poor (2), and Bad (1)
 - The arithmetic mean is reported as the Mean Opinion Score (MOS)
 - Varies with each test and across languages

Speech-Layer Objective Model

- Perceptual Evaluation of Speech Quality (PESQ) ITU-T P.862
 - Full reference—requires the input test speech as a reference
 - Allows evaluation of discontinuous types of degradation, such as packet losses in VoIP and bit errors in digital cellular
 - Some problems in implementation reported so an applications guide will be published, P.862.2

Speech Coding Standards








- Narrowband speech
 - GSM-AMR, G.729, G.723, G.728, IS-127(EVRC), IS-96(QCELP), IS-95(VSELP)
 - G.711(PCM), G.721(ADPCM), G.726(ADPCM)
 - LPC-10, MELP,...
- Wideband speech
 - G.722 (ADPCM)
 - G.722.1 (Transform)
 - AMR-WB (CELP)
- Wideband audio
 - MPEG-1,2,4
 - Philips PASC
 - Sony ATRAC
 - DOLBY AC-3

Narrowband Speech Coding Performance

- Original Sample
- G.711 μ -Law PCM 64 Kbps
- G.726 ADPCM 16, 24, 32 and 40 Kbps
- G.723.1
 - ACELP 5.3 Kbps
 - MP-MLQ 6.3 Kbps
- G.729 ACELP 8 Kbps
- MPEG-4










Coded Classical Music

8KHz:       
G.711 G.726(32K) G.729(8K) G.723.1(5.3K) NB_AMR(12.2K, 6.7K, 4.75K)

16KHz:

      
G.722(64K, 48K) G.722.1(32K, 24K) WB_AMR(23.85K, 12.65K, 6.60K)

Coded Music+Voice

8KHz:       
G.711 G.726(32K) G.729(8K) G.723.1(5.3K) NB_AMR(12.2K, 6.7K, 4.75K)

16KHz:

      
G.722(64K, 48K) G.722.1(32K, 24K) WB_AMR(23.85K, 12.65K, 6.60K)

Packet Loss Concealment

G.729 without Packet Loss



G.729 with 20% Packet Loss



G.729 with Packet Loss Concealment



G.711 without Packet Loss



G.711 with burst of 6 frames lost



G.711 with new error concealment



Networks for Voice Communications

- Public Switched Telephone Network (PSTN)
- Digital Cellular
- Voice over IP (VoIP)
- Voice over Wireless Local Area Networks (Voice over Wi-Fi)
- All developed independently

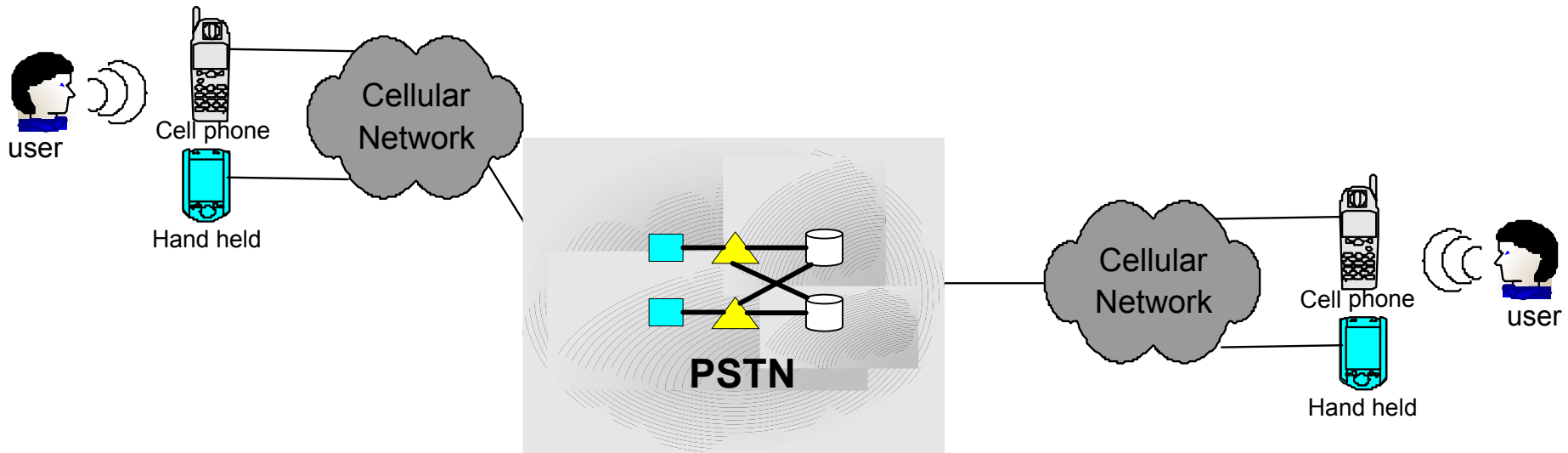
Other Issues in Multimedia Over Networks

- Tandem Coding
- Background Impairments
- Transcoding
- Delay
- Complexity
- Power Consumption

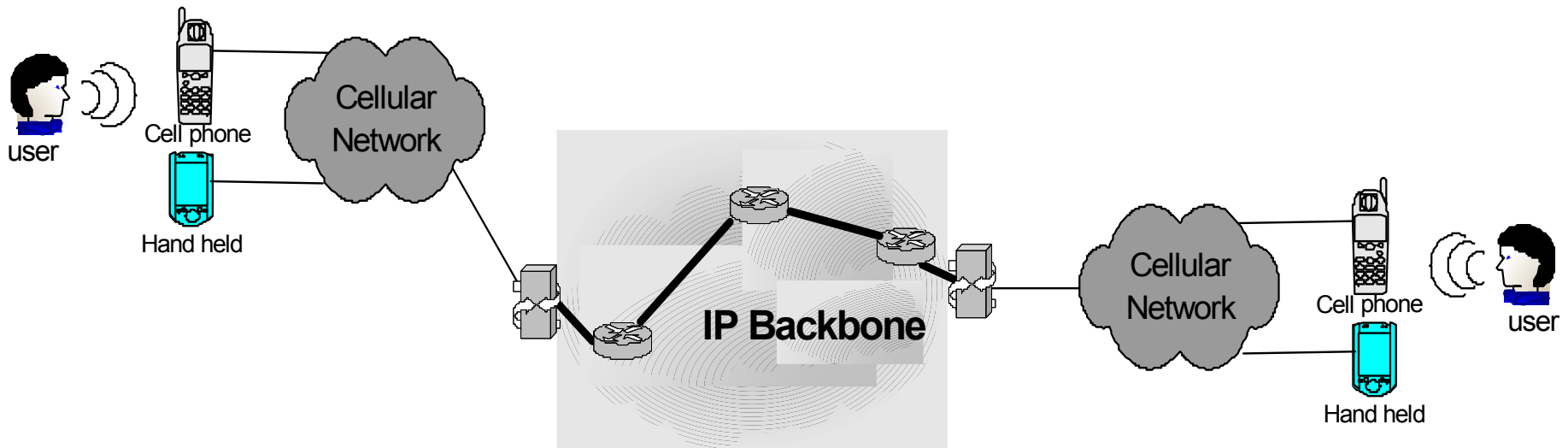
More Issues

- Packetization
- Error Concealment
- Scalability
 - SNR
 - Spatial
 - Temporal
 - Bandwidth

Tandem Digital Cellular with PSTN



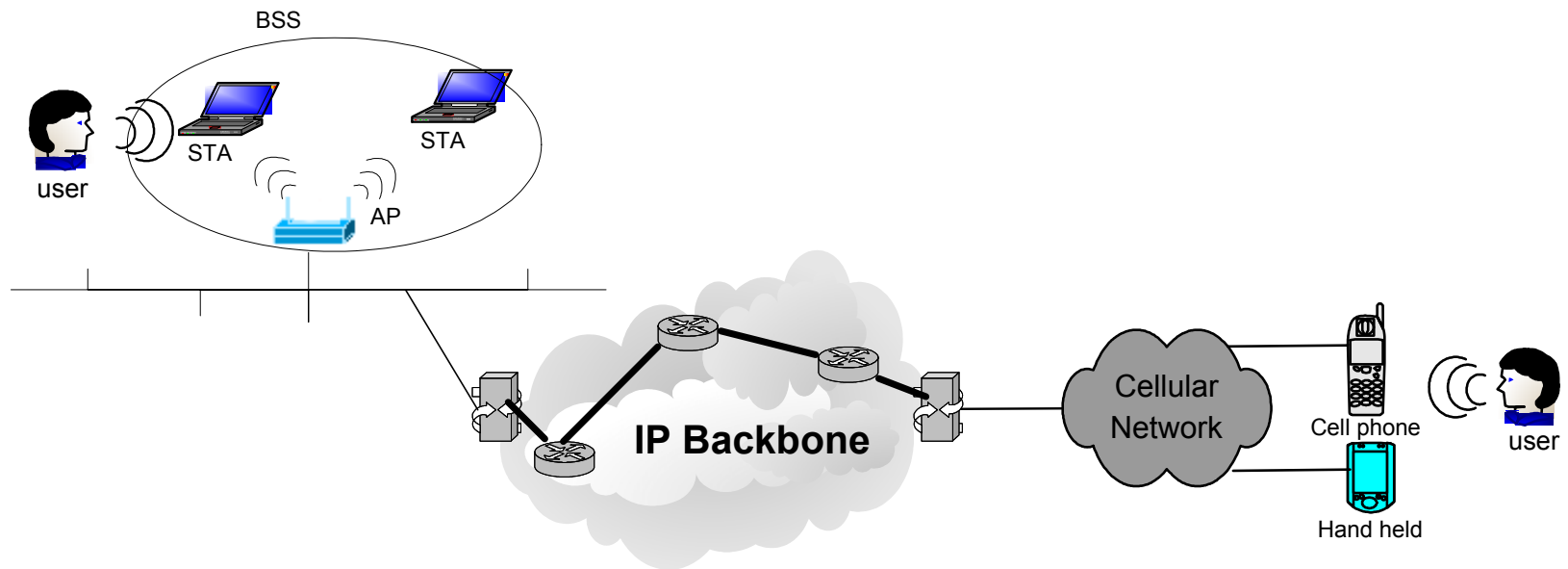
Tandem Digital Cellular and VoIP



January 21, 2005



Tandem Digital Cellular, VoIP, and Voice over Wi-Fi



The Public Switched Telephone Network

- Reliable, dedicated connection
- Reliable access
- All-digital in U. S. except for local loops
- Time division multiplexed transmission in the backbone
- Primary voice codec is G.711 (log-PCM) at 64 kbits/sec per voice channel

Characteristics of the PSTN

- Low bit error rate--less than 10^{-9}
- Low delay voice codec (G.711)
- Low complexity voice codec (G.711)
- High quality voice—called toll quality (MOS>4.0 for G.711)
- Other ITU standardized codecs are available but G.711 is the universal solution

Voice Codecs for the PSTN

Codec	Rate kbits/s	MO S	Complexity (MIPS)	Frame Size/Look Ahead (ms)
G.711	64	4.0+	$\ll 1$	0.125
G.721/ 726	32	~ 4.0	1.25	0.125
G.728	16	3.9	30	0.625
G.729	8	4.0	20	10/5
G.729 A	8	4.0	12	10/5
G.723. 1	5.3/6.3	3.7/ 3.9	11	30/7.5

PSTN Codec Tandem Performance

Voice Codec	Mean Opinion Score (MOS)
G.711x 4	>3.50
G.726x 4	2.91
G.729x 2	3.27
G.729x 3	2.68
G.726+ G.729	3.56
G.729+ G.726	3.48

Voice Over Internet Protocol (VoIP)

- Targeted to compete with the PSTN
- The PSTN uses dedicated circuit-switched connections (inefficient use of bandwidth)
- Packet switching breaks information into payloads that are then placed in packets
- Packets have headers that contain information on destination, routing, control, and management

VoIP (continued)

- Headers require additional bandwidth and processing
- Packet payloads may be one or more frames of encoded voice data
 - Few frames per packet reduce latency but increases overhead inefficiency due to headers
 - More frames per packet reduces header inefficiency but increases delay

VoIP (continued)

- Retransmissions are undesirable because of possible multiple nodes in the path so TCP/IP is not used
- RTP/UDP/IP is used but headers are long, so header compression may be employed

Properties of VoIP Codecs

Codec	Relevant Properties
G.711	Low delay, toll quality, low complexity, higher rate
G.729	Toll quality, acceptable delay, low rate, acceptable complexity
G.723.1	Low rate, acceptable quality, relatively high delay
G.722	Wideband speech, low delay, low complexity, higher rate

User Datagram Protocol (UDP)

- No handshaking—connectionless
- No guarantee of data delivery
- No guarantee of the order of data delivery
- No congestion control
- No guarantee on delay

Typical Header for RTP/UDP/IP Voice Transmission

- IP Header 20 Bytes
- UDP Header 8 Bytes
- RTP Header 12 Bytes
- Example for G.711 Voice
 - 10 ms of G.711 voice data = 640 bits or 80 bytes
 - Header overhead of 40 bytes of RTP/UDP/IP Header
 - Packet efficiency of 67%
- Use longer data packet payloads (latency)
- Use header compression (40 bytes to 4 bytes)

Sources of Latency

- Codec Encoding Delay
 - Specified by G.114 as one frame delay before processing and one frame delay for processing for a total delay of 2 frames + look ahead
- One frame delay at wireline interface to synchronize rates, or
- Multiple frame delay for packetization
- Jitter Buffer Delay

Packet Loss Concealment (PLC)

- Packets may be lost due to buffer overflows, excess traffic, or corrupted packets
- Common PLC Methods
 - Silence insertion
 - Repeat last good packet with attenuation
 - Interpolation or voice synthesis
- Must be matched to the voice codec
- PLC must be optimized to maintain voice quality

Voice Over IP Summary

- Targeted to compete with the PSTN
- Reality
 - High delay and large delay variability
 - Occasionally long periods of packet loss
- Other Important Issues
 - Packet loss concealment
 - Jitter buffers
 - Evaluating user satisfaction