The Multimedia
Communications Revolution of
the 21\textsuperscript{st} Century

Lawrence Rabiner
Rutgers University and the
University of California at Santa
Barbara
Multimedia-The Perfect Storm

• Changes in the telecom environment
  – packet digital networks
  – pervasive broadband connectivity
  – ubiquitous wireless

• Changes in the multimedia environment
  – high quality codecs for text, voice, audio, image, video

• Changes in multimedia understanding
  – speech understanding goes mainstream
  – text translation capabilities
  – information retrieval and information extraction based on text, image, speech,…

Range of new services integrating processing, understanding, and networking of multimedia information
Twenty-First Century Communications
## Telecom Technology Directions

<table>
<thead>
<tr>
<th></th>
<th>20th Century</th>
<th>21st Century</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access</td>
<td>Narrowband Voice</td>
<td>Broadband Multimedia</td>
</tr>
<tr>
<td>Network</td>
<td>Circuit-Switched</td>
<td>Packet Switched Based on IP</td>
</tr>
<tr>
<td>Traffic Eng.</td>
<td>Erlang Model</td>
<td>Fractal Model</td>
</tr>
<tr>
<td>Platform</td>
<td>Intelligent Switches</td>
<td>Routers</td>
</tr>
<tr>
<td>Operations</td>
<td>People-Oriented</td>
<td>Web-Based, Automated</td>
</tr>
<tr>
<td>Devices</td>
<td>Telephone, Computer</td>
<td>PC, PDA, Universal Devices</td>
</tr>
<tr>
<td>Services</td>
<td>Simple Voice, Data</td>
<td>Panaply of Integrated Services</td>
</tr>
</tbody>
</table>

Multimedia_2007
Telecom End State

• Integrated and networked (via IP) broadband multimedia
  – data of all types (PL, FR, ATM, IP)
  – text
  – images (graphics, photos, icons)
  – audio (both speech and music)
  – video (TV grade to HDTV, video-conferencing, video email, video meeting notes)
  – virtual reality (games, sporting events, meetings)
  – searchable, browsable multimedia documents (catalogs, out-of-print books, historic documents)
  – shared reality tele-collaboration

“Ubiquitous Broadband Access to the Network”
Moore’s Law Growth in Computing Resources

[Graph showing the growth in processing power (MIPS) over the years from 1976 to 2000, with different processor models such as DSP20, DSP1, 286, 386, Pentium, DSP16, DSP16A, and DSP16110.]
Decreasing Technology Adoption Rates

Time To Reach 10 Million Customers

- Pager: 41 Years
- Fax Machine: 22 Years
- VCR: 9 Years
- Cellular Phone: 9 Years
- CD-ROM: 7 Years
- PC: 6 Years
- WWW: 4 Years
- New Browser: 1 Year

Sources: Apple, AirTouch Cellular, Info Tech and USA Today
<table>
<thead>
<tr>
<th>Activity</th>
<th>Monthly Users</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online</td>
<td>70M</td>
</tr>
<tr>
<td>Email</td>
<td>60M</td>
</tr>
<tr>
<td>News</td>
<td>35M</td>
</tr>
<tr>
<td>Weather</td>
<td>25M</td>
</tr>
<tr>
<td>Work-related Research</td>
<td>24M</td>
</tr>
<tr>
<td>Political Info</td>
<td>24M</td>
</tr>
<tr>
<td>Product Research</td>
<td>20M</td>
</tr>
<tr>
<td>Instant Messaging</td>
<td>14M</td>
</tr>
<tr>
<td>Travel Info</td>
<td>10M</td>
</tr>
<tr>
<td>Health Info</td>
<td>7M</td>
</tr>
<tr>
<td>Blog</td>
<td>5M</td>
</tr>
<tr>
<td>Share Files</td>
<td>4M</td>
</tr>
<tr>
<td>Buy Product</td>
<td>3.5M</td>
</tr>
</tbody>
</table>
Gadgets in U.S. Homes (2005)

- DVD Player – 71%
- Mobile Phone – 71%
- Desktop Computer – 70%
- Digital Camera – 40%
- Video Game Console – 37%
- Laptop Computer – 23%
- PDA – 11%
- Digital Music Player – 10%
- Digital Video Recorder – small
Enterprise Traffic

Percent of Enterprise Traffic

- Streaming Audio
- Video Conferencing
- Streaming Video
- Terminal-Host
- Peer-to-Peer
- Voice-over-IP
- File Transfer
- Web Services
- Web Applications
- Email
- Client-Server

Three Major Telecom Trends

• VoIP replacing POTS wireline telephony
• Wireless everywhere
• Broadband access (cable, DSL, satellite, fiber) replacing narrowband access (dial-up modems)
VoIP
(Voice over IP)
VoIP-Voice on Data Networks

- IP Telephony opens huge multimedia opportunities (>8M U.S. subscribers, 1Q/2007)
  - integration of text, data, voice, image, video via IP
  - intelligent voice services (follow me, reroute calls)
  - video conferencing
  - call logs (voice recording, metadata)
  - click-to-dial
  - phone call scheduling
  - intelligent network services (caller ID, conferencing, call forwarding)
Myths about VoIP (IEEE Spectrum, March 2005)

- VoIP is free
  - broadband connection service fee; VoIP handset; local connection fees
- VoIP and POTS are the same—except for price
  - packet versus circuit switched; dumb Internet vs smart switches; smart terminal vs dumb terminal; easy to add new services vs impossible to add new services; geography independent for area codes
- QoS not a problem for VoIP
  - dropped packet issues; network delays; network jitter
- VoIP isn’t as good quality as POTS
  - need MPLS to combat QoS issues
- VoIP is just another data app
  - 800 services; life line services; real time requirements
- VoIP isn’t secure
  - uses encryption to guarantee security
- VoIP telephony = POTS telephony
  - easy to reroute calls for mobility; uses softphones for ease of porting services
U.S. VoIP Analysis-1Q07

- **Cable TV companies** have 60% market share (market share growing)
- **Vonage** and other pure play providers have 40% market share (market share falling due to patent infringement case by Verizon)
- **RBOCs** have insignificant market share
- VoIP **revenue** (estimated) of $2.6B in 2006
- **Market leaders (1Q2007):** Comcast-2.4M subs; Skype/Vonage/Cox-2.2M subs; Time-Warner Cable-~1.64M subs; Cablevision-~1.1M subs; CallWave-~0.780M subs;
VoIP Challenges

• QoS over packet networks
  – dropped packets, network delays, network jitter
• Security of VoIP calls
  – need encryption of all calls
• VoIP viewed as a data service
  – need to meet real time requirements => ATM transport over dedicated networks
  – 800 services
  – 900 lifeline services
Wireless Technology

“If you can do it on a wireline connection, the market will drive it to wireless – albeit with a different set of quality constraints”
Wireless Communicator Vision – Services and Features

- voice
- messaging – text, speech, video, IM, MM
- music – MP3, AAC, WMF
- images – photos
- text documents
- video – streamed/stored (IPTV)
- ring tones
- conferencing
- audio – streamed/stored
- micro-payments
- gaming

- web pages/web surfing
- speech recognition/synthesis
- ftp file transfers
- gps locations/tracking
- location-based services
- personalized services
- sensors/management
- software defined radio (SDR)
- Bluetooth enabled
- advanced signal processing (MIMO, smart antennas)
- roaming between 3G/4G cellular, WiFi, WiMAX

(dx) – dsp/multimedia
Major Wireless Trends

- 2.4 billion cellphones worldwide (versus 1 billion landlines); GSM 82%, CDMA 13%
- Wireless voice/messaging are dominant cellular services
- Wireless data at rates up to 1 Gbps (NTT DoCoMo lab demo using OFDM and MIMO methods)
- Seamless integration of wireless LAN technologies for ubiquitous access in homes, on the road, at the office; cellular 3G/4G, WiFi and WiMAX
- RFID, Bluetooth, UWB technologies being adopted and growing rapidly
- Wireless (data) services starting to grow (European trends):
  - text messaging-79% take rate
  - multimedia messaging-28% take rate
  - personalization (ring tones, wallpaper)-15% take rate
  - mobile video-8% take rate
U.S. Wireless Statistics

- **Subscribers**
  - 207.9M 12/2005
  - 233.0M 12/2006
  - Predicted Growth to 270M subs in 2009
  - China has 400M subs 12/2005

- **Voice Revenues**
  - $119B in 2005
  - $130B in 2006
  - Predicted Growth to $180B in 2009

- **Voice Minutes**
  - 1.5 Trillion in 2005
  - 1.8 Trillion in 2006

- **SMS Messages**
  - 24.7B in last 6 months of 2004
  - 48.7B in last 6 months of 2005

- **Wireless Data Revenues**
  - $4.6B in 2004
  - $8.6B in 2005
  - 70-80% of data revenue from SMS (Short Messaging Services)

CTIA Report, 4/2006
U.S. Wireless Subscribers – 1Q2007

- AT&T/Cingular: 62.2 million subs, GSM
- Verizon: 60.7 million subs, CDMA
- Sprint/Nextel: 53.6 million subs, CDMA
- T-Mobile: ~23 million subs, GSM
- Alltel: ~9.5 million subs, CDMA
- US Cellular: ~6.5 million subs, GSM
Wireless Networking
## Growth of Wireless Networking

<table>
<thead>
<tr>
<th>Technology</th>
<th>Standard</th>
<th>Type</th>
<th>Rate</th>
<th>Range</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>3G</td>
<td>EDGE,WCDMA</td>
<td>WWAN</td>
<td>384 Kbps</td>
<td>1-5 miles</td>
<td>2 GHz</td>
</tr>
<tr>
<td>WiMax</td>
<td>802.16a</td>
<td>WMAN</td>
<td>30-75 Mbps</td>
<td>1-6 miles</td>
<td>2-6, 11 GHz</td>
</tr>
<tr>
<td>Wi-Fi</td>
<td>802.11a/b/g</td>
<td>WLAN</td>
<td>11-54 Mbps</td>
<td>&lt;300 feet</td>
<td>2.4/5 GHz</td>
</tr>
<tr>
<td>UWB</td>
<td>802.15.3</td>
<td>WPAN</td>
<td>110-480 Mbps</td>
<td>&lt;30 feet</td>
<td>4-11 GHz</td>
</tr>
<tr>
<td>Bluetooth</td>
<td>Device</td>
<td></td>
<td>&lt;720 Kbps</td>
<td>&lt;30 feet</td>
<td>2.45 GHz</td>
</tr>
</tbody>
</table>

**Key Issues:**
- line of sight
- spectrum ownership
- backhaul of traffic
Radio Technology Matrix - Mobility and Data Rates

Data Rates

- Satellite
- UMTS CDMA
- GSM
- 802.15.4
- WiMAX
- WiFi
- 802.16
- 802.20
- 802.11 a/b/g
- UWB
- 802.15.3a
Wireless Performance Challenges

- **Improved wireless devices** => more compute power, more signal power (kills batteries), size
- **More bandwidth** => more users/band, more bandwidth available to all users
- **Software radio at RF** – reduced size, reduced power, lower cost => “solves” the access air interface problem (“My terminal speaks your protocol”)
- **Better source coding** => speech, audio, images, video, gaming => effectively more capacity
- **Channel coding** => better protection against fades, dropouts and interference => better coverage in a fixed area
- **Advanced modulation/adaptive modulation** => EDGE, WCDMA, OFDM => better use of allocated frequency spectrum
- **Diversity methods** => smart antennas, time-space codes, MIMO systems => increased frequency reuse, better suppression of interference, better resource allocation, better power control, reduced noise
Broadband Technologies
Broadband/Wideband Technologies

- **DSL** – Digital Subscriber Line (140M world-wide, 20.2M U.S., 4Q2005)
- **Cable modems** (70M world-wide, 29M U.S., 2Q2006)
- **Fiber-to-the-X** (FTTX) (6M homes passed in 2006, estimated 600K-1M subs in 2006)
- **Satellite**
- **High speed data service** (private line or switched data, e.g., ATM, FR, IP)
Worldwide Broadband

Total Broadband By Technology (4Q_2006)

- DSL, 66%
- CM, 22%
- FTTx, 11%
- Other, 1%

Source: Point Topic Ltd
Worldwide Growth in Broadband

Q4-2005: 52M cable; 18M FTTx; 140M DSL => 210M broadband
Homes Passed by FTTH (North America)

FTTH Realities (North America)

• Fiber passes 6 million homes in North America and is being marketed to more than 5 million customers (3Q 2006)
• Verizon aggressively pursuing FiOS Services to homes (investing $23B over next several years)
• 4.4 million homes passed by FiOS services (3Q 2006); service marketed to 3.1 million customers
• 348,000 FiOS customers in 4Q 2006 (12% of available market)
  – 141,000 new customers in 1Q2007
  – loss of 553,000 residential phone lines in 2Q 2006
• FiOS is Verizon’s best hope of competing with cable companies; ‘triple play’ of telephony, data (Internet access) and video services
• AT&T providing fiber-to-the-node (FTTN) (U-verse) in 2007 providing 30-60 Mbps to the home within 3,000 feet from the fiber serving node; 20,000 subs in 1Q2007
Multimedia Technologies
Signal Processing in Multimedia Systems

• **compression and coding** of the multimedia signals; standards-based, proprietary

• **organizing, storing, and retrieving** multimedia signals; streaming, layering, QOS issues

• **accessing** multimedia signals by matching the user to the machine; GUI, spoken language interface (SLI), media conversion, agents

• **searching** multimedia archives and databases (based on machine intelligence); text, image, speech

• **browsing** multimedia archives and documents (based on human intelligence); text, image, audio, video
Technology Assumptions

• multimedia processing is a lot more than compression and coding
• multimedia applications need to be standards-based
• handling (delivery, display) of multimedia signals is crucial
• the user interface is critical to usability of most applications
• a multimedia experience is shared between people and machines
Compression and Coding of Multimedia Signals (Standards-Based)
## Compression of Multimedia Signals

<table>
<thead>
<tr>
<th>Speech/Audio Type</th>
<th>Frequency Range</th>
<th>Sampling Rate</th>
<th>Bits/Sample</th>
<th>Uncompressed Bitrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Narrowband Speech</td>
<td>200-3200 Hz</td>
<td>8 kHz</td>
<td>16</td>
<td>128 kb/s</td>
</tr>
<tr>
<td>Wideband Speech</td>
<td>50-7000 Hz</td>
<td>16 kHz</td>
<td>16</td>
<td>256 kb/s</td>
</tr>
<tr>
<td>CD Audio</td>
<td>20-20000 Hz</td>
<td>44.1 kHz</td>
<td>16 x 2 channels</td>
<td>1.41 Mb/s</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Image Type</th>
<th>Pixels per Frame</th>
<th>Bits/Pixel</th>
<th>Uncompressed Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAX</td>
<td>1700 x 2200</td>
<td>1</td>
<td>3.74 Mb</td>
</tr>
<tr>
<td>VGA</td>
<td>640 x 480</td>
<td>8</td>
<td>2.46 Mb</td>
</tr>
<tr>
<td>XVGA</td>
<td>1024 x 768</td>
<td>24</td>
<td>18.87 Mb</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Video Type</th>
<th>Pixels per Frame</th>
<th>Image Aspect Ratio</th>
<th>Frames per Second</th>
<th>Bits/Pixel</th>
<th>Uncompressed Bitrate</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTSC</td>
<td>480 x 483</td>
<td>4:3</td>
<td>29.97</td>
<td>16</td>
<td>111.2 Mb/s</td>
</tr>
<tr>
<td>PAL</td>
<td>576 x 576</td>
<td>4:3</td>
<td>25</td>
<td>16</td>
<td>132.7 Mb/s</td>
</tr>
<tr>
<td>CIF</td>
<td>352 x 288</td>
<td>4:3</td>
<td>14.98</td>
<td>12</td>
<td>18.2 Mb/s</td>
</tr>
<tr>
<td>QCIF</td>
<td>176 x 144</td>
<td>4:3</td>
<td>9.99</td>
<td>12</td>
<td>3.0 Mb/s</td>
</tr>
<tr>
<td>HDTV</td>
<td>1280 x 720</td>
<td>16:9</td>
<td>59.94</td>
<td>12</td>
<td>622.9 Mb/s</td>
</tr>
<tr>
<td>HDTV</td>
<td>1920 x 1080</td>
<td>16:9</td>
<td>29.97</td>
<td>12</td>
<td>745.7 Mb/s</td>
</tr>
</tbody>
</table>
Bandwidths for Speech and Audio Signals

Compact Disc

FM-Radio

AM-Radio

Telephone

Frequency in Hz

10 20 50 200 3400 7000 15000 20000
Telephone Bandwidth Speech Coding Demo

- 64 kb/s G.711 Mu-Law PCM
- 32 kb/s G.726 ADPCM
- 16 kb/s G.728 LD-CELP
- 8 kb/s G.729 CS-ACELP
- 4.8 kb/s CELP
- 2.4 kb/s LPC10(e)
Speech Coder Quality

- GOOD (4)
- FAIR (3)
- POOR (2)
- BAD (1)

BIT RATE (kb/s)

- G.723.1
- G.729
- IS-127
- G.728
- G.726
- G.711

1980 1990 2000

MELP 1995 FS1016 IS54

FS1015
Wideband Speech Coding Demo

- 128 kb/s, 3.2 kHz telephone bandwidth
- 256 kb/s, 7 kHz bandwidth original
- 64 kb/s, G.722 2-band SBC
- 32 kb/s, LD-CELP
- 16 kb/s, BE-CELP

Male and female talkers
Audio Coding Standards

- **MP3** – layer 3 of MPEG-1 audio coding for MPEG1 video coding standard for movies on CDROM
- **AAC** – audio coding for MPEG2 video coding standard for high quality movies on DVD
- **AAC+** -- audio coding for MPEG4 video coding standard (includes Spectral Band Replication and Parametric Stereo)
- **AAZ** – scalable audio standard as part of MPEG4 audio SLS (scalable-to-lossless standard)
Audio Coding Demo

• Female Vocal Solo
  – original and coded (unspecified order)
  – 48 kbps, 64 kbps, 80 kbps, original
  – can you tell the sequence?

  Actual Sequence: 80 kbps, original, 64 kbps, 48 kbps

• Orchestra
  – original and coded (unspecified order)
  – 64 kbps, 128 kbps, original
  – can you tell the sequence?

  Actual Sequence: 64 kbps, 128 kbps, original
Image Coding Principles

• spatial redundancy
  – repeated patterns
  – image correlations in space
  – spectral correlations

• temporal redundancy
  – repeated objects in video sequence
  – predictable moves of objects--horizontal, vertical, fade-in and out, zoom, pan

• take advantage of human visual system
  – perceptual masking of intensity, color, texture, time sequence
  – regions of interest (ROI)
Generic Image Coding Algorithm

- Short-Term Analysis
  - Intensity
  - Texture
  - Motion

- JND Estimation
  - Just-Noticeable Distortion Profile

- Adaptive Coding
  - Constant Quality
  - Constant Bit Rate

Bit Rate Control
Monochrome Image Coding

8 bit

0.5 bit

0.33 bit

0.25 bit
Color Image Coding

24 bit  1 bit
0.5 bit  0.25 bit
Image Coding Standards: Continuous Tone

• JPEG
  – processing in blocks
  – DCT spectral analysis of blocks
  – psychophysically based scalar quantization
  – entropy encoding

• JPEG-2000 (improved image quality over JPEG)
  – uses wavelet technology for improved image quality
  – modern architecture and standard
  – downloadable software
  – handles broad range of conditions

• Motion JPEG-2000 (MJ2)
  – no inter-frame coding
  – used for video clips on digital still camera
  – frame-based video recording and editing
  – both lossless and lossy compression
Image Coding: Continuous Tone

- JPEG-2000/Motion JPEG-2000 used for:
  - digital photography
  - medical imaging (lossless compression)
  - document imaging
  - surveillance
  - satellite imagery
  - FAX
# JPEG Performance

<table>
<thead>
<tr>
<th>Bits/Pixel</th>
<th>Quality</th>
<th>Compression Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \geq 2 )</td>
<td>Indistinguishable</td>
<td>8-to-1</td>
</tr>
<tr>
<td>1.5</td>
<td>Excellent</td>
<td>10.7-to-1</td>
</tr>
<tr>
<td>0.75</td>
<td>Very Good</td>
<td>21.4-to-1</td>
</tr>
<tr>
<td>0.5</td>
<td>Good</td>
<td>32-to-1</td>
</tr>
<tr>
<td>0.25</td>
<td>Fair</td>
<td>64-to-1</td>
</tr>
</tbody>
</table>
JPEG Image Comparisons

Uncompressed
378 KiB
1:1

JPEG JFIF
11.2 KiB
1:33.65
JPEG q 30

JPEG 2000
11.2 KiB
1:33.65
Video Coding

- Video teleconferencing: H.261-H.264 standards
- Movie storage on CDROM: MPEG-1, 1.2 Mbps for video, 256 kbps for audio
- Broadcast video on DVD: MPEG-2, 2-15 Mbps for video/audio
- Low bit rate video telephony to HDTV broadcast: MPEG-4, 16 kbps (video telephony) to 500 Mbps (broadcast HDTV)
Multimedia Coding

- How much compression can be achieved for text, voice, audio, image, video
  - **lossless compression** – well understood theoretical limits
  - **lossy compression** – perceptual minimum rate (perceptual entropy of source); Region of Interest (RoI) coding – put bits only where noticed
Multimedia Processing Issues

• Lossless versus lossy coding
  – **Moore’s law** – doubles memory and processing every 18 months → less compression needed
  – **embedded devices** – smaller footprints, less memory and processing for any given task
  – **networks** – fading, packet loss, jitter, bit errors ⇒ need error protection bits ⇒ less coding bits available
Multimedia Processing Issues

- **scalable coding** – needed for converged networks with shared content displayed on vastly different end devices
- **adaptive coding** – need to adjust coding to widely varying source/channel conditions (congestion, jitter, delay, wireless fades, noise)
- **sensor networks** – ultimately using billions of small, inexpensive networked sensors for monitoring air, water, temperature, pressure, purity, chemical composition
Multimedia Understanding
Systems
Multimedia Understanding Tasks

• Text-based systems:
  – search (data mining, Information Retrieval (IR), Information Extraction (IE))-Google,…
  – language identification
  – language translation

• Speech-based systems:
  – natural language understanding
  – speaker verification
  – spoken language identification
  – spoken language translation
  – data mining (both IR and IE) from voice queries
Multimedia Understanding Tasks

• Audio-based systems:
  – detect music genre
  – detect artist and group
  – determine music identity by matching short sections
  – align music and lyrics automatically
  – note and chord recognition
  – score transcription
Multimedia Understanding Tasks

• Image-based systems:
  – image class identification
  – scene analysis
  – event detection/object detection
  – people detection
  – face recognition
  – data mining

• Video-based systems:
  – activity detection
  – activity identification
  – people tracking/object tracking
  – object identification and motion detection
  – summarization of content
  – content-based sampling
Text/Audio/Image/Video
Understanding Challenges

• Go from **Information Retrieval** types of data mining (almost exclusively from text) to Information Extraction; i.e., tell me the relations between search terms and the real world

• Enable all modes of Information Retrieval (and ultimately Information Extraction) from text, speech, image, video inputs and outputs, e.g., speech query input, text/speech/image/video output

• Take media understanding to the next level, including media syntax and semantics; e.g.,
  – **Input:** get me video of the bears and lions
  – **Output:** football games versus zoo or jungle pictures

• Utilize a lot more dialogue technology at the User Interface
Some Typical Multimedia (and Multimodal) Systems
Multimedia Applications

• Text-to-Speech User Interfaces:
  – auditory interfaces (for speech-based dialogues with machines)
  – visual interfaces (for display-based dialogues with machines)
• Natural Language Speech User Interfaces:
  – speech recognition interfaces (for command-and-control and dialogues with machines)
  – voice mining of text/voice/image/video
• Digital Library of Images/Videos for Storage, Browsing and Searching:
  – CYBRARY using DJVU image compression
  – Pictorial Transcripts of video content
• Multimodal User Interfaces:
  – speech and pointing (via PDA) for accident reports
  – speech and pointing (via tablet) for finding places
Capturing Multimedia (Linguistic) Intelligence via ASR

- Large vocabulary speech recognition (>200,000 words)
- Speaker independent system
- Real-time recognition, very low latency
- Word accuracy of 82% (DARPA HUB-4 test)
DVL-Digital Video Libraries

(Behzad Shahraray, AT&T Labs)
Digital Video Library (DVL)

- DVL is a digital video management system that enables the content-based retrieval and intelligent browsing of video information
  - applies text, image, video, audio, and speech processing techniques to organize, index, and condense video information
  - generates multiple representations of the video contents to enable the delivery of information over a range of information appliances and a wide range of available bandwidth.
Automated Extraction of the Semantic Hierarchy of News

Broadcast News Categories

Table of Contents

News Summary

Story 1
Story 2
...

Anchor

Detailed News Reporting

News

Commercials

Audio

Video

Text

Content-based Searching and Browsing

Topic detection and categorization

Story extraction by text processing

Multi-Modal Segmentation

Categorization using audio and video and text

Linear Retrieval

Broadcast news programs: across multiple media; linear in time; flat structure.

This is the broadcast content transcribed by human. It is used to illustrate the construction of semantics using automated techniques based on multimedia.
Video Segmentation Example

- **Video Shot Boundary Detection**
  - finds edit points

- **Content Based sampling**
  - shot Boundary Detection
  - inter-shot motion-based segmentation

- **Representative Image selection to serve as a visual index or for use in creating compact representations for efficient browsing**
Telecom Futures

- **Terabit optical networks** go into commercial use, making bandwidth virtually inexhaustible and low cost
- **Software defined radios** redefine wireless interoperability between wireless devices and wireless network protocols
- Networks with **billions of wireless sensors** provide steady flow of information about air, water, mechanical stress, traffic, locations, etc.
- **Fiber-to-the-home** becomes the standard for broadband communication changing the nature and the use of the Internet in the home
- **Telecommunications** becomes the norm for conducting day-to-day business, using high quality video conferencing to conduct ‘face-to-face’ business—globally
- **High definition images and videos** redefine the consumer electronics area with 3D television in high definition formats becoming routine
- **Lifelike computer graphic displays** change the nature of learning, especially in teaching concepts that are more graphical and visual than usual
- **Computer mesh networks** learn to interoperate flawlessly, providing a seamless world-wide platform for sharing computing and storage resources as well as applications
Summary

• The **Multimedia Communications Revolution** integrates and merges key concepts in **computing, communications, and networking**
  – it will continue to change the way we work, relax, learn, play, and communicate
  – many challenges ahead, especially in understanding of multimedia
  – signal processing will play a major role in reaching the desired end state \( \rightarrow \) ubiquitous access to people and information, anywhere, anytime, anyplace