Scenarios for FlexCode

Harald Pobloth
Jan. 10, 2008
Outline

- FlexCode in Brief
- Scenarios in Brief
- FlexCode Advantage
- Scenarios in some Detail
- Ranking of Scenarios
- Existing Codecs
- Other 6th Framework Projects
- Summary / Conclusions
FlexCode in Brief

• Exploit
  – flexible signal models, e.g. GMMs
  – high-rate optimized quantization, e.g. SQ or LVQ
  – source-channel decoding

• to obtain coding scheme
  – adjustable to a wide range of rates
  – variable delay
  – variable protection of bit payload on channel

• that serves
  – wide range of services and scenarios
FlexCode in Brief

- Operates on the continuum of rates
  - intended optimal performance between 10 and 60 kbps
- Computational complexity independent of rate
- No storage of codebooks
  - quantization designed (high-rate approximations) according to
    - available rate
    - source distribution
  - store GMMs to represent signal statistics
    - independent of rate
- Contains advanced perceptual model
- Adapts to feedback from transmission channel
  - Protection strength
  - Rate distribution between channel- and source-coding
FlexCode in Brief

- FlexCode organized in work-packages:
  - WP1: Source coding
  - WP2: Channel coding
  - WP3: Real-world scenarios
  - WP4: Integration and demonstrator
  - WP5: Testing
  - WP6: Dissemination and standardization
- WP3 Real-world scenarios provides
  - context and constraints to FlexCode coding
  - list of service scenarios, which benefit from FlexCode
• Eight scenarios
• Ranked according to several criteria

1. Mobile Multimedia Blogging Scenario (MMBS)
   – Mobile user records or broadcasts audio-visual content
   – Different rate constraints for record and broadcast case
   – Content consumed life or downloaded from blog-server
   – Content consumed on different devices (TV, PC, mobile phone)
1. Mobile Multimedia Blogging Scenario (MMBS)

2. Multimedia Conference Scenario (MCfS)
   - Users at a variety of locations
   - 3D sound rendering enhances conferencing experience
   - Legacy equipment (mono, stereo, narrowband) support
   - Speech and music content
1. Mobile Multimedia Blogging Scenario (MMBS)
2. Multimedia Conference Scenario (MCfS)
3. Mobile Conversation Scenario (MCvS)
   – Much like current circuit switched telephony
   – Conversational
     • Delay
   – Mainly speech content
   – At least one user mobile
     • Battery, computation, bandwidth constraint
1. Mobile Multimedia Blogging Scenario (MMBS)
2. Multimedia Conference Scenario (MCfS)
3. Mobile Conversation Scenario (MCvS)
4. Internet Conversation Scenario (ICS)
   - Much like the MCvS
   - Scenario assumes no mobile devices
     • Maximum computational power, high bandwidth
   - Wireless link might be present
     • Necessity to adapt to different error patterns
Scenarios in Brief

1. Mobile Multimedia Blogging Scenario (MMBS)
2. Multimedia Conference Scenario (MCfS)
3. Mobile Conversation Scenario (MCvS)
4. Internet Conversation Scenario (ICS)
5. Multimedia On-Demand Streaming Scenario (MMSS)
   - User selects content from server
   - Stream is exclusive to user (unicast)
   - Stream is real-time
     • Some delay constraint
   - Both mobile and stationary devices considered
Scenarios in Brief

1. Mobile Multimedia Blogging Scenario (MMBS)
2. Multimedia Conference Scenario (MCfS)
3. Mobile Conversation Scenario (MCvS)
4. Internet Conversation Scenario (ICS)
5. Multimedia On-Demand Streaming Scenario (MODSS)
6. Multimedia Multicast-Streaming Scenario (MMSS)
   - Stream to several users
   - Real time
   - Both mobile and stationary terminals
1. Mobile Multimedia Blogging Scenario (MMBS)
2. Multimedia Conference Scenario (MCfS)
3. Mobile Conversation Scenario (MCvS)
4. Internet Conversation Scenario (ICS)
5. Multimedia On-Demand Streaming Scenario (MODSS)
6. Multimedia Multicast-Streaming Scenario (MMSS)
7. Multimedia Download Scenario (MDS)
   - Not real-time
     • No delay requirements
     • Out of sequence reception possible
   - Unicast
   - Mobile and stationary devices
Scenarios in Brief

1. Mobile Multimedia Blogging Scenario (MMBS)
2. Multimedia Conference Scenario (MCfS)
3. Mobile Conversation Scenario (MCvS)
4. Internet Conversation Scenario (ICS)
5. Multimedia On-Demand Streaming Scenario (MODSS)
6. Multimedia Multicast-Streaming Scenario (MMSS)
7. Multimedia Download Scenario (MDS)
8. Surveillance Scenario (SuS)
   - Audio surveillance
   - High compression rate while maintaining intelligibility
   - Severe noise might be present
   - Equipment should be cheap
     • Computational constraints
Scenarios in Brief

1. Mobile Multimedia Blogging Scenario (MMBS)
2. Multimedia Conference Scenario (MCfS)
3. Mobile Conversation Scenario (MCvS)
4. Internet Conversation Scenario (ICS)
5. Multimedia On-Demand Streaming Scenario (MODSS)
6. Multimedia Multicast-Streaming Scenario (MMSS)
7. Multimedia Download Scenario (MDS)
8. Surveillance Scenario (SuS)

• Others …
  – e.g. gaming, push-to-talk, voice mail
• Device contain more than one services
• Same codec can serve a magnitude of services
  ➢ This coder provides savings in
    – licensing costs
    – storage
    – implementation costs
1. Mobile Multimedia Blogging (MMBS)
   - Maximum exploitation of bottleneck upload channel
     - Upload channels of different types and at different loads
     - Codec needs to adapt to varying channel characteristics
   - Life vs. upload gives different requirements for codec
     - Delay, rate, error patterns different for life and upload cases
   - Rendering on heterogeneous devices
     - Downscaling at receiving device can save battery / computation
     - Sender can adjust if audience is known (no server upload)
2. Multimedia Conference (MCfS)
   – Content variation
     • Varying type and strength of background noise
     • Music or pure speech
     • Single or multiple speakers
   – Varying number of participants
     • Network load increases with increasing # of participants
     • FlexCode can adapt to the varying network
   – Varying number of active speakers
     • Both network load and signal characteristics change with # of active speakers
   – Participants with different terminal / network capabilities
   – Life encoding → network feedback can be used in encoder
3. **Mobile Conversation (MCvS)**
   - Different and varying network capabilities / conditions
   - Life encoding and utilization of feedback
   - Multiparty conversation
     - Can be engaged / disengaged during one session
     - Similar to MCfS
   - Environment noise varies
     - Perceptual model assures high quality speech remains even in noisy conditions
4. Internet Conversation (ICS)
   - Same as MCvS
   - More content variation likely (e.g. background music)
5. Multimedia On-Demand Streaming (MODSS)

- Varying content
- Adaptation to receiver characteristics
  - Uni-cast stream: Can be encoded to the specs of the receiver
  - Error / delay tradeoff for device at hand
- Different QoS requirements
  - Seamless change from e.g. pre-view to paid service
6. Multimedia Multicast Streaming (MMSS)
   - Similar to MODSS
   - Optimized error rate / delay tradeoff
   - Some advantages require embedded stream in MMSS
     • Adaptation to receiver
     • Encoding work load optimized
       – Not necessary to run a multitude of encoders at server
FlexCode Advantage

• Multimedia Download (MDS)
  – Adaptive error correction
    • Avoid need to re-transmit
  – Adaptation to receiving device
• Surveillance (SuS)
  – Background characteristics
  – Channel conditions
  – Rendering/storing
    • Devices can be selected to temporarily provide higher quality for active rendering / monitoring
Scenarios in some Detail

- Scenarios described by:
  - Outline
  - User perspective (content, quality)
  - Equipment (user devices, middleware)
  - Networks
  - Requirements
    - Rate
    - Delay
    - Error-robustness
  - Existing codecs
  - Benefit of FlexCode to scenario
  - Standardization relations
  - Commonalities with other scenarios
<table>
<thead>
<tr>
<th>Scenario</th>
<th>User perspective</th>
<th>Equipment</th>
<th>Network</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Content</td>
<td>Quality</td>
<td>Sender</td>
</tr>
<tr>
<td>3.1 MMBS</td>
<td>Speech, audio, background noise, often simultaneously</td>
<td>High-quality acceptable on home devices. Stereo signals, some mobile devices provide mono only. Bandwidth: 16-32 kHz</td>
<td>Mobile device (phone, PDA including still camera, optionally connected to digital video camera)</td>
</tr>
<tr>
<td>3.2 MCfS</td>
<td>Speech, background noise, multiple speakers, audio</td>
<td>High-quality, Bandwidth: 8-24 kHz</td>
<td>Stationary device, mono input. (See 3.2)</td>
</tr>
<tr>
<td>3.3 MCvS</td>
<td>Speech, background noise and audio</td>
<td>High-quality, mono, Bandwidth: 8 kHz or more</td>
<td>Mobile phone</td>
</tr>
<tr>
<td>3.4 ICS</td>
<td>Mainly speech, audio should be supported</td>
<td>High-quality, mono, Bandwidth: 8 kHz or more</td>
<td>PC or WiFi phone</td>
</tr>
<tr>
<td>3.5 MODSS</td>
<td>Mixed Content (Speech, Noise, Music), Audio</td>
<td>High-quality, mono but mostly stereo, multi-channel, Bandwidth: 8-24 kHz</td>
<td>Streaming server</td>
</tr>
<tr>
<td>Scenario</td>
<td>User perspective</td>
<td>Equipment</td>
<td>Network</td>
</tr>
<tr>
<td>----------</td>
<td>-----------------</td>
<td>-----------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>Content</td>
<td>Quality</td>
<td>Sender</td>
</tr>
<tr>
<td>3.6 MMSS</td>
<td>Mixed Content (Speech, Noise, Music), Audio</td>
<td>High-quality, mono but mostly stereo or multi-channel, Bandwidth: 8-24 kHz</td>
<td>Streaming server</td>
</tr>
<tr>
<td>3.7 MDS</td>
<td>Mixed Content (Speech, Noise, Music), Audio</td>
<td>High-quality (FM-radio or DVD quality), mostly stereo or multi-channel, Bandwidth: 16-24 kHz</td>
<td>Content server</td>
</tr>
<tr>
<td>3.8 SuS</td>
<td>Speech distorted with background noise</td>
<td>Medium for monitoring, low for storing</td>
<td>Low power, los CPU hardware</td>
</tr>
<tr>
<td>Scenario</td>
<td>Rate</td>
<td>Delay</td>
<td>Error-rate</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>-------</td>
<td>------------</td>
</tr>
</tbody>
</table>
| 3.1MMBS | 40-60 kbps | Limited only by device capability ➔ a few hundred ms | Service usable at PLR > 8% | AMR-WB+, e-AAC+ | • Maximum exploitation of upload channel  
• Rendering on heterogeneous devices  
• Source / rendering device mismatch |
| 3.2 MCfS | ≈ 24-60 kbps | 200 – 400 ms end-to-end ➔ ≈ 25 ms algorithmic | Service usable at ≥ 3% PLR | AMR-WB, ITU-G.722.1, ITU-G.722.1.C, G.729.1 | • Content variation  
• Varying number of participants  
• Varying number of active speakers  
• Different network and terminal capabilities to different participants  
• Life encoding and utilization of feedback  
• Conference recording with reduced data-rate |
| 3.3 MCvS | ≈ 10-32 kbps | 100 - 300 ms end-to-end ➔ ≈ 25 – 40 ms algorithmic | ≥ 1% FER, ≥ 8% PLR if transport via Internet | AMR, AMR-WB, EVRC, VMR-WB, EVRC-WB, G.729.1 | • Different network and terminal capabilities / conditions  
• Exploitation of possible feedback  
• Multiparty conversation  
• Adaptation to environment noise |
<table>
<thead>
<tr>
<th>Scenario</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Rate</td>
</tr>
</tbody>
</table>
| 3.4 ICS  | 10-60 kbps | 200 – 400 ms end-to-end $\Rightarrow \approx 25$ ms algorithmic delay | Service usable at PLR > 8% | AMR-WB, proprietary codecs, e.g. iLBC or iSAC | • Varying network qualities  
• Content variation  
• Exploitation of feedback  
• Multiparty conversations |
| 3.5 MODSS | BW [kHz]/rate: 4 / 8-16 kbps 8 / 12-32 kbps 16/14-56 kbps 24/16-64 kbps | Limited only by device capability $\Rightarrow \text{a few hundred ms}$ | Low PLR of $\approx 1$-2% due to re-transmit | AMR, AMR-WB, AAC, HE-AAC v2, AMR-WB+, Windows Media, MPEG Surround | • Adaptation to content characteristics  
• Adaptation to receiver characteristics  
• Optimized error rate / delay tradeoff  
• Different QoS requirements |
| 3.6 MMSS | BW [kHz]/rate: 4 / 8-16 kbps 8 / 12-32 kbps 16/14-56 kbps 24/16-64 kbps | Limited only by device capability $\Rightarrow \text{a few hundred ms}$ | Service usable at PLR of $\geq 5$%, re-transmit should be avoided | AMR, AMR-WB, AAC, HE-AAC v2, AMR-WB+, Windows Media, BSAC, AAC+, MPEG Surround | • Adaptation to content characteristics  
• Adaptation to receiver characteristics  
• Optimized error rate / delay tradeoff  
• Optimized encoding work load |
| 3.7 MDS  | BW [kHz]/Rate: 16/14-56 kbps 24/16-64 kbps | Limited only by device capability $\Rightarrow \text{a few hundred ms}$ | No errors, re-transmission of lost packets | MP3, AAC, MPEG Surround, Windows Media Technologies | • Adaptive error correction  
• Adaptation to receiver characteristics |
| 3.8 SuS  | 4-10 kbps | $\leq 100$ ms to minimize device complexity | PLR up to 10% due to wireless link, re-transmit should be avoided to minimize complexity | AMR-NB, AMR-WB | • Adaptation to background characteristics  
• Adaptation to channel conditions  
• Adaptation to rendering / storing device |
Ranking of Scenarios

- Final ranking according to:
  - Economical relevancy:
    - Operator interest
    - Manufacturer interest
  - End-use interest
  - Degree of novelty:
    - How much can scenario can from FlexCode
  - Ease of implementation
    - How feasible is an implementation within FlexCode project

- 5 criteria

- Scale and ranking points
  - Very low = 0, low = 1, medium = 2, high = 3, very high = 4
### Ranking of Scenarios

#### Final ranking table

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Operator interest</th>
<th>End user interest (general showcase)</th>
<th>Manufacture interest</th>
<th>Degree of novelty (FlexCode advantage)</th>
<th>Ease of implementation</th>
<th>Ranking points</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 MMBS</td>
<td>Medium</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Medium</td>
<td>12</td>
</tr>
<tr>
<td>3.2 MCfS</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>13</td>
</tr>
<tr>
<td>3.3 MCvS</td>
<td>Very High</td>
<td>High</td>
<td>Very high</td>
<td>High</td>
<td>Low</td>
<td>15</td>
</tr>
<tr>
<td>3.4 ICS</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>High</td>
<td>Low</td>
<td>11</td>
</tr>
<tr>
<td>3.5 MODSS</td>
<td>High</td>
<td>Very high</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>15</td>
</tr>
<tr>
<td>3.6 MMSS</td>
<td>Very High</td>
<td>High</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>13</td>
</tr>
<tr>
<td>3.7 MDS</td>
<td>Medium</td>
<td>High</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>10</td>
</tr>
<tr>
<td>3.8 SuS</td>
<td>Low</td>
<td>Medium</td>
<td>Low</td>
<td>Medium</td>
<td>High</td>
<td>9</td>
</tr>
</tbody>
</table>
## Existing Codecs

<table>
<thead>
<tr>
<th>Type</th>
<th>Rate [kbps]</th>
<th>Delay [ms]</th>
<th>Bandwidth [kHz]</th>
<th>Codecs considered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech, conversational</td>
<td>6.6-23.85</td>
<td>25</td>
<td>0.05-7</td>
<td>AMR-WB</td>
</tr>
<tr>
<td></td>
<td>8-32</td>
<td>≈ 48</td>
<td>0.05-4 and 0.05-7</td>
<td>G.729.1</td>
</tr>
<tr>
<td>Speech &amp; Audio, conferencing and streaming</td>
<td>24, 32</td>
<td>40</td>
<td>0.05-7</td>
<td>G.722.1</td>
</tr>
<tr>
<td></td>
<td>24, 32, 48</td>
<td>40</td>
<td>0.05-14</td>
<td>G.722.1 Annex C</td>
</tr>
<tr>
<td>Speech &amp; Audio non-conversational</td>
<td>6-36 (mono)</td>
<td>100-200</td>
<td>Varying with bit-rate from 6.2 to 19</td>
<td>AMR-WB+</td>
</tr>
<tr>
<td></td>
<td>7-48 (stereo)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10-44 (mono)</td>
<td></td>
<td>Varying with bit-rate from 10 to 17</td>
<td>3GPP e-AAC+</td>
</tr>
<tr>
<td></td>
<td>16-52 (stereo)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audio non-conversational</td>
<td>Typically between 6 and 64 for mono (12-128 for stereo)</td>
<td>Depending on bit rate</td>
<td>Up to 0.02-20 Depending on bit rate</td>
<td>AAC</td>
</tr>
<tr>
<td></td>
<td>Depending on bit rate</td>
<td></td>
<td></td>
<td>BSAC</td>
</tr>
</tbody>
</table>

### Codecs
- AMR-WB
- G.729.1
- G.722.1
- G.722.1 Annex C
- AMR-WB+
- 3GPP e-AAC+
- AAC
- BSAC
Standardization Bodies

• 3GPP
  – Focus on SA4
  – Multimedia Telephony Service for IMS (MTSI)

• ITU
  – Focus on SG 16 WP 3
    • Q23: New speech and audio codec (ITU-T G.MMCC)
    • Q10: Maintainance / extensions of existing codecs
    • Q9: VBR-EV codec development

• MPEG
  – Focus on exploration work on Speech and Audio Coding
• Enthrone
  – Streaming and download for mass-market
  – Relies on:
    • MPEG-21
    • Universal multimedia access (UMA)
  – Scalable codecs needed
  – Choice of audio codec still open
• ISIS / DANAE
  – Multimedia content search and delivery
    • User interaction
  – Multimedia adaptation
    • Context adaptation (MPEG-21 DIA)
    • DRM
    • Scalable codecs
Other 6th FP projects

• ARDOR
  – Adaptive rate-distortion optimised sound coder
  – Rate-distortion controlled combination of different coding techniques
  – High-quality audio
  – Overlap in targets and tools with FlexCode
  – FlexCode:
    • More complete approach:
      – Channel coding included
      – Covering wider range of scenarios
      – More scenario oriented
    • Further utilization of high-rate theory in source codec:
      – Eliminate need of storing CB tables
    • Targets lower rates
Other 6th FP projects

• M-Pipe
  – Cross layer optimization
    • Layer independent descriptor
  – Requires specific network structure
  – FlexCode: Audio source coding more fundamental
  – FlexCode: Channel coding closer to source coding
Summary / Conclusions

- Eight scenarios described
  - Most scenarios focus on packet switched networks
  - Advantage of FlexCode for scenarios identified
    - Flexible codec gives intrinsic advantage for scenarios
- Flexible codec can serve several scenarios
- Easy adaptation to channel / equipment necessary
  - Target device not known when encoding
  - Content servers serve large number of users
- Two highest ranked scenarios
  - Mobile conversation scenario
  - Multimedia on demand streaming scenario
Summary / Conclusions

- Overlap with other scenarios identified
  - MCvS large overlap with ICS and MCfS
    - ICS more computational power
    - MCfS focus on multi-party conversation, architecture different
  - MODS some overlap with MMSS and MDS
    - MMSS different architecture (multicast vs. unicast)
    - MDS different requirements (unlimited re-transmit, delay)

- Performance of benchmark codecs shown

- Perspective of FlexCode
  - Standardization bodies
  - Other 6th FP projects

- Full document at: http://www.flexcode.rwth-aachen.de/materials.html