Recent Audio/Speech Coding Developments in ITU-T and future trends

Presented by Claude Lamblin, France Telecom
ITU-T SG16 Vice Chair
WP3/16 chair (Media Coding)

What ? Why ?

- Broad overview of ITU-T speech and audio coding standards
  - the achievements in recent years
  - the main trends for the next four years.
- Encourage the participation and involvement of experts from the academic communities in the development of standards

ITU-T audio/speech coding portfolio

- Significant portfolio : G.71x or G.72x series
- Various applications with different constraints → various codecs optimized with different trade-offs (quality, bit rate, complexity, robustness, delay)
- Wide range: audio bandwidths, bit rates
  - Narrow band (NB): BW=[300-3400Hz], Fs: 8 kHz, G.711 (1972)

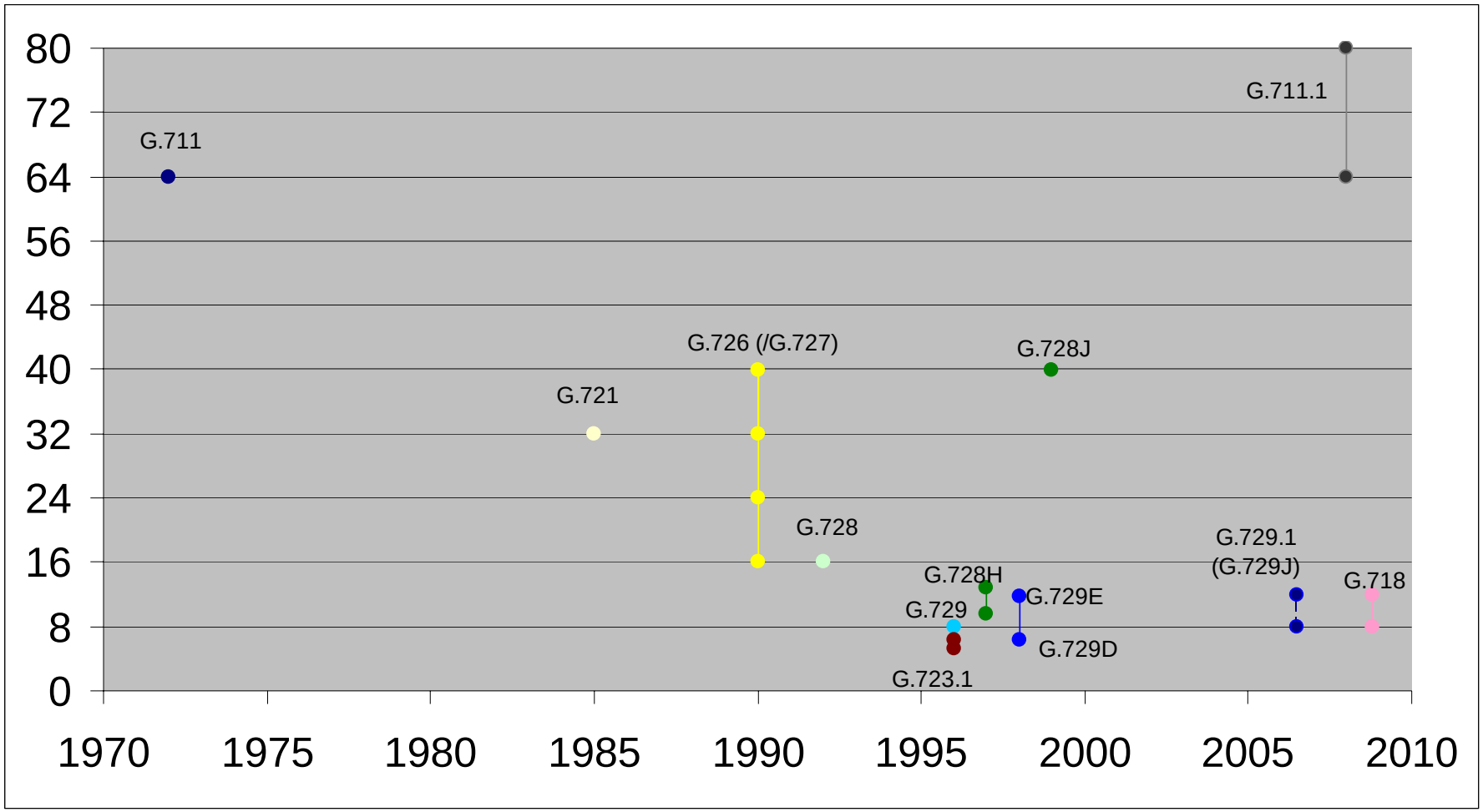
    ⇓
  - Full band (FB): BW=[20-20000Hz], Fs: 48 kHz, G.719 (06/2008)

From bit rate lowering to bandwidth increasing

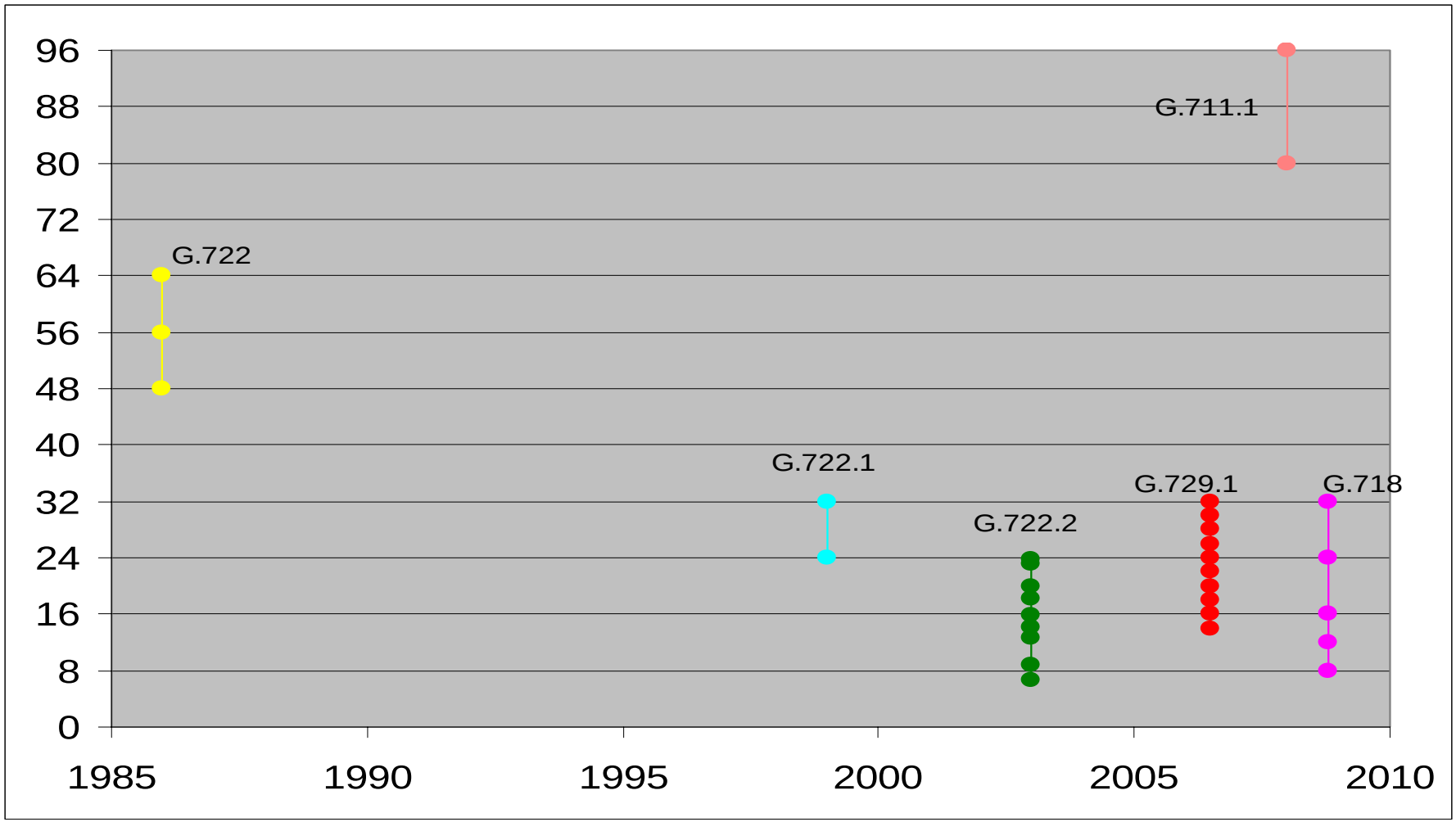
- **Objective: Quality /bit rate tradeoff**
  - Lower the bit rate / No quality degradation
  - Increase the quality / Keep the bit rate
  - Others: complexity, delay, robustness, flexibility
- **Bandwidths / Bit Rates**

Quality	Fs (Hz)	BW (kHz)	Rates (kbit/s)
NB	8000	0.3-3.4	80 (G.711.1) → 5.3 (G.723.1)
WB	16000	0.05-7	96 (G.711.1) → 6.6 (G.722.2)
SWB	32000	0.05-14	24,32,48 (G.722.1 C)
FB	48000	0.02-20	32 → 128 (G.719)

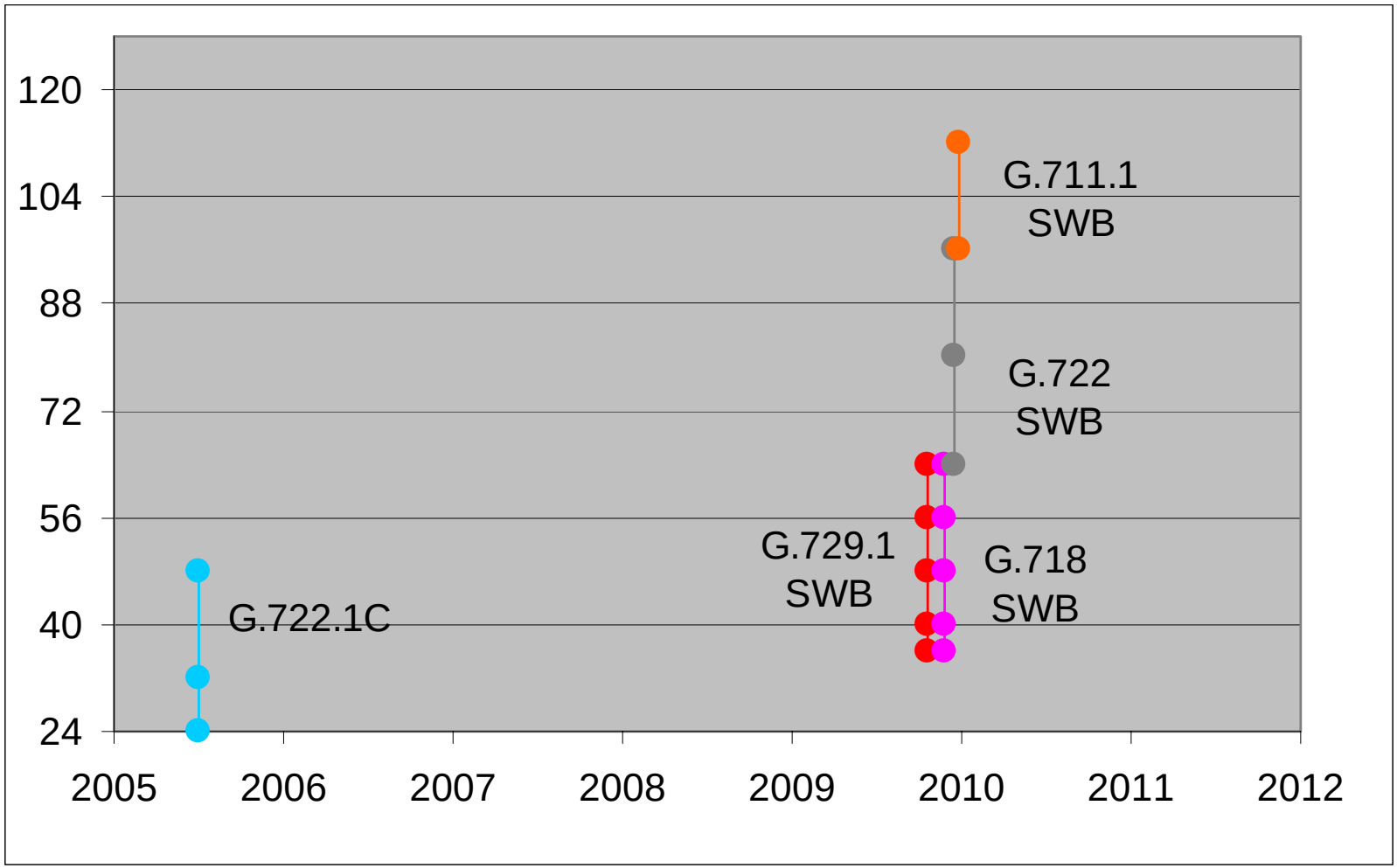
ITU-T Narrowband Standards



ITU-T Wideband Standards



ITU-T Superwideband Standards



ITU-T fullband Standard: G.719

Low-complexity full-band audio coding for high-quality conversational applications

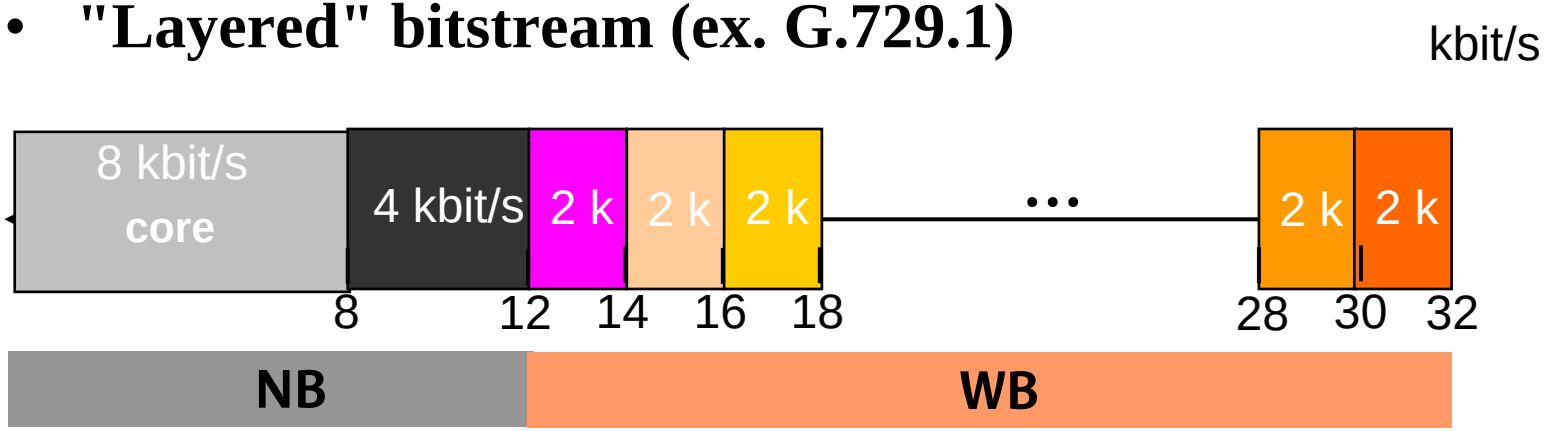
- **1st fullband ITU-T audio coding standard (06/2008)**
- Applications: High quality A/V conferencing (telepresence), *streaming*
- Multirate: 32 kbit/s up to 128 kbit/s
- BW=[20Hz, 20000Hz], Fs= 48 kHz
- Low latency: 40 ms (20 ms frame)
- Low complexity: 15 WMOPS up to 21 WMOPS
- Quality: *cf " Fullband Conversational Codec: What Testing Methodology?", C. Quinquis & al, EUSIPCO-2008*
- Coding scheme: transform coding with adaptive time-resolution, adaptive bit-allocation and lattice vector quantization

Flexibility: Multirate codecs

- **Quality ↑ with bit rate ↑ and audio bandwidth ↑**
- Heterogeneous accesses
- Congestion control / Differentiated QoS
- Adaptation to channel errors
- **Rate chosen at the encoder for each frame**
  - NB: G.726 (40/32/24/16), G.728 (40/16/12.8/9.6), G.729 (11.8/8/6.4+DTX), G.723.1(6.3/5.3+DTX)
  - WB: G.722.1 (24/32), G.722.2 (6.6/8.85/12.65/ 14.25/15.85/18.25/19.85/23.05/23.85+DTX)
  - SWB: G.722.1 C (24/32/48)
  - FB: G.719 (32, 36, 40, 44, 48, 52, 56, 60, 64, 68, 72, 76, 80, 84, 88, 96, 104, 112, 120, 128)
- **Embedded schemes: layered audio coding**

Embedded Schemes

- **"Layered" bitstream (ex. G.729.1)**



- **Higher flexibility**
  - "on the fly" bit rate selection by simple truncation of the bit stream by any component of the communication chain
  - Easy adaptation to any service requirements and interconnected networks/terminals
  - No out band signaling, no multiple codec negotiation, no transcoding

"Old" ITU-T Embedded coders

- *bit rate scalability only*
- *no audio bandwidth scalability*

G.7xx	BW	Rates (kbit/s)	date
G.711	NB	64, 56, …	1972
G.727	NB	40,32, 24,16	1990
G.722	WB	64, 56, 48	1988

New ITU-T Embedded coders

Bit rate and audio bandwidth scalability NB-WB

- G.729.1 (06/2006): 8/12/14/16/18/.../28/30/32 kbit/s
  - *"ITU-T G.729.1: An 8-32 kbit/s Scalable Coder Interoperable with G.729 for Wideband Telephony and Voice over IP", S. Ragot & al, ICASSP-2007*
  - *"Pre-Echo Reduction in the ITU-T G.729.1 Embedded Coder", B. Kövesi & al, EUSIPCO-2008*
- G.711.1 (03/2008): 64/80/96 kbit/s
  - *"G.711.1: A Wideband Extension to ITU-T G.711", Y. Hiwasaki & al, EUSIPCO-2008*
  - *"Noise Shaping in an ITU-T G.711-Interoperable Embedded Codec", J. Lapierre & al, EUSIPCO-2008*
- G.718 (09/2008): 8/12/16/24/32 kbit/s
  - *"ITU-T EV-VBR: A Robust 8-32 kbit/s Scalable Coder For Error Prone Telecommunications Channels", T. Vaillancourt & al, EUSIPCO-2008*

Future ITU-T Embedded coders

Bit rate, audio bandwidth (NB-WB-SWB) and channel (mono/stereo) scalability

- SWB ext. G.729.1 and G.718: fall 2009
  - Mono SWB enhancement layers and stereo WB/SWB layers on top of G.729.1 and G.718 (at 32 kbit/s)
    - Gross bit rates: 36, 40, 48, 56, 64 kbit/s
  - 2 candidate consortia in the ongoing ($\rightarrow$ 10/2008) qualification phase:
    - Ericsson, France Telecom, Motorola, Nokia, Panasonic, Texas Instrument, VoiceAge
    - ETRI, Huawei
- SWB ext. G.711.1 and G.722: end 2009
  - Mono SWB enhancement layers and stereo WB/SWB layers on top of G.711.1 (wb modes at 80 & 96 kbit/s) and G.722 (at 56 & 64 kbit/s)
  - 10 potential candidates: Dolby, ETRI, France Telecom, Fraunhofer, Huawei, NTT, Panasonic, Philips, Siemens, VoiceAge

Other activities in ITU-T audio group

- Maintenance of existing ITU-T speech and audio coding Rec.:

  G.711, G.712, G.711.1, G.718, G.719, G.722, G.722.1, G.722.2, G.723.1, G.726, G.727, G.728, G.729, G.729.1, G.191, G.192
- Related functionalities:
  - Voice/Sound Activity detection
  - Discontinuous transmission (DTX) / Comfort Noise Generation (CNG)
  - Packet Loss Concealment (PLC)
  - Lossless Compression (LLC)
- Media Coding Summary Database (MCSD) and toolbox for coding
- Software Tool Library (STL): G.191

Voice Activity Detection

- Voice Activity Detection / Discontinuous transmission /Comfort Noise Generation
- G.729 VAD/DTX/CNG
  - G.729 Annex B (1996) optimized for DSVD applications (Digital Simultaneous Voice and Data)
  - G.729 Appendices II & III (08/2005): VAD optimized for VoIP systems to provide different quality/bandwidth efficiency tradeoffs
- Generic Sound Activity Detector (GSAD): 2006 → 2009
  - voice → rich audio signal (music, information tones, …)
- G.729.1 Annex C (06/2008): DTX/CNG scheme
  - *"On the ITU-T G.729.1 Silence Compression Scheme", P. Setiawan & al, EUSIPCO-2008*

Packet Loss Concealment

- Wireless or IP systems → Packet losses
- Requirement: Robustness against Frame erasures
- Recent standards (since G.729 in 1995)
  - PLC procedure standardized with the main Recommendation
- "Older" standards, PLC added later
  - G.711 (1972): PLC in Appendix I (1999)
  - G.728 (1992): PLC in Annex I (1999)
  - G.722 (1988): PLC in Appendices III & IV (11/2006) (on request from ETSI TC DECT)

Lossless Compression

Reduce bit rate while keeping bit exactness (no quality degradation)

- G.711 LLC: lossless coding of G.711 coded speech
- Proposed in 06/2007; to be completed in 02/2009
- Terms of reference (ToR):
  - Use cases
    - VoIP Trunk Gateway Applications, Narrowband conferencing
    - FAX or modem, DTMF, Transport of bit stream (TFO)
    - Transcoded speech (tandem with lower bit rate NB standards : AMR, G.72x, …)
  - Design constraints (Requirements)
    - Frame length: 5, 10, 20, 30, 40 ms (in samples: 40/80/160/240/320)
    - Complexity: < 2 WMOPS (40-sample frame);
    - Memory: RAM < 6 kbytes, DROM < 16 kbytes; PROM < 5 k Basic Operations
    - Compression efficiency per use case (to be assessed with various databases covering several use cases)
- 7 potential candidates: Cisco, Ericsson, Huawei, Nokia, NTT, Qualcomm, Texas Instruments

Media Coding Summary Data base

Media: Audio, Video, still image, graphic, character
SDOs: ITU-T, 3GPP, 3GPP2, ISO/IEC MPEG, ETSI, ARIB, ..

- Items for an audio coding standard:
  - Formal name, Nickname
  - Technology, Speech Model?
  - Audio Bandwidth, Sample Rate, Frame Length
  - Bitrate(s), Embedded Scalability?
  - VAD/DTX/CNG? PLC?
  - Complexity (fixed/floating point)
  - Memory: RAM, PROM, DROM
  - Software (enc/dec), Fixed/Floating
  - Primary Applications
  - IPR Status, Contact point

Toolbox for content coding (for IPTV): 3 parts : audio / video /text (accessibility)

Software Tools Library (STL): G.191

for speech and audio coding standardization

- Common set of tools:
  - host lab sessions emulating real usage conditions:
    - input signal conditioning (level and/or sampling frequency adjustment, addition of background noises and of reverberation, input terminal characteristics filtering), transmission conditions (frame erasures, bit errors, bitstream truncation), output conditioning (…), reference processing (MNRU, bandwidth limitation, "old" ITU-T coders, …)
  - ITU-T audio codecs specification and performance evaluation:
    - bit-exact fixed point C code using set of basic operators (simulate DSP)
- Releases: first: 1992, …, **last: 2005, future: 2009**
  - Basic operators: revision of 16/32-bit operators weights, new control flow operators, alternative set (40-bit acc.)
  - Channel errors: error patterns and statistics for packet-based networks (IP) and wireless networks, embedded structure
  - Tools for wider bandwidths processing (SWB, FB): terminal characteristic filters, reverberation tool, stereo processing
  - ITU-T coders: G.722 (PLC option), G.728

ITU-T speech and audio coding: standardization process

cf. "Fullband Conversational Codec: What Testing Methodology?", C. Quinquis, P.Usai, EUSIPCO-2008

- Terms of Reference (ToRs) *and time schedule*
  - Applications
  - Requirements and objectives
    - Sampling frequency, audio bandwidth(s)
    - Frame length, algorithmic delay
    - Bit rate(s)
    - Quality performance on various conditions (speech, music, noisy environment, errors, reverberant ?, MCU)
    - Complexity: computational, memory
- Qualification phase: test of subset of requirements, floating point C-code
- Selection or optimization/characterization phase: test of extensive set of requirements (and objectives), fixed point C-code

Performance Assessment

- **Quality (c/o ITU-T SG12 "Performance and quality of service")**
  - Subjective: ACR, DCR, CCR (ITU-T P.800), Ref-A-Bx2 (ITU-R BS.1116 & BS.1285), MUSHRA (ITU-R BS.1534)

  (cf. "Fullband Conversational Codec: What Testing Methodology?", C. Quinquis & al, EUSIPCO-2008)
  - Objective: ITU-T P.862.2 ("WB-PESQ"), ITU-R BS.1387 (PEAQ)
- **Complexity**
  - ITU-T audio codecs specification: bit-exact fixed point C code using library of basic operators (simulate DSP)
  - Weights of basic operators and control flow operators
  - Memory
    - Data ROM and static /scratch RAM: in 16-bit kword
    - Program ROM : basic operators and function calls counter

Conclusion (1)

Why be involved in ITU-T ?

- Truly global and not-discriminatory standards
- Working together for consensus decisions
- Very flexible to start new initiatives
- Fast & transparent procedures
- *Looking towards the **standards of the future** cooperating with Academia and R&D institutions*

Conclusion (2)

How to be involved in ITU-T audio & speech coding work?

- How to get ITU-T Speech and audio coding Recommendations
  - **Free download** (text + C-code) **:** http://www.itu.int/rec/T-REC-G/e

  (G-series: **Transmission systems and media, digital systems and networks)**

- How to follow ongoing ITU-T audio coding work ?
  - Email reflector: wp3audio@yahoogroups.com

- How to participate ITU-T audio coding work (or any other ITU-T work)?
  - ITU Membership (fee): http://www.itu.int/members/index.html
    - **Member States (191), Sector Members (569), Associates (155)**
  - Or Invited experts (free)

Conclusion (3)

Speech & Audio Coding Context

- **Universal Multimedia Access**
  - Various networks interconnected
  - Heterogeneous terminals /Different accesses
- **Networks interoperability**
  - Multiple incompatible coding standards
  - Adaptation networks, accesses, terminals
- **Enhance quality, flexibility and robustness**
  - Bandwidths ↑ (NB → HiFi) ; channels ↑ (mono → stereo → 3D)
  - Scalable (bit rates, bandwidths, channels)
  - "Universal": speech (clean, noisy), music, mixed content, Error recovery
- **Targeted applications in ITU-T :**
  - **1) conversational :** Conventional (PSTN, CME), Packetised voice (e.g. VoIP), AV multimedia, 3G and future wireless (4G, WiFi)
  - *2) Non Conversational: Multimedia streaming, Storage, …*

Thank you

Muchas Gracias

Merci

Questions ?