FlexCode

Project no: FP6-2002-IST-C 020023-2

Project title: FlexCode

Instrument: STREP

Thematic Priority: Information Society Technologies

D-3.1 Ordered List of Real World Service Scenarios

Due date of deliverable: 2007-03-01
Actual submission date: 2007-03-01

Start date of project: 2006-07-01 Duration: 36 Months

Organisation name of lead contractor for this deliverable: Ericsson AB

Revision: B

Project co-funded by the European Commission within the Sixth Framework Programme (2002-2006)		
Dissemination Level		
PU	Public	**X**
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

Abstract

This document provides a list of service scenarios for the FlexCode project. The service scenarios are ranked according to expected economical relevancy and the potential gain the FlexCode paradigm can add to them, compared with existing paradigms. For each scenario details like requirements in the form of, e.g., bit-rate, error-rate, delay as well as the user perspective and the used equipment and networks are described. Given the scenario description the document identifies overlaps between the scenarios and the advantages the FlexCode paradigm means to the scenarios. The performance of codecs relevant to the scenarios, which form a benchmark for the FlexCode codec is summarized.The document also provides a short outline of standardization bodies relevant for FlexCode and activities ongoing there. In addition, we describe how the FlexCode project relates to other FP6 projects and what interaction to these is possible. The two scenarios with highest ranking are identified to be the mobile conversation scenario and the multimedia on demand streaming scenario.

List of Abbreviations

3D	3 dimensional
3G	3rd generation technology
3GPP	3rd generation partnership project
AAC	Advanced audio coding
AAC-LD	Advanced audio coding - low delay
AC3	Abbreviation for Dolby Digital lossy audio compressions
ACR	Absolute category rating (listening test)
ADSL	Asymmetric digital subscriber line
AM	Amplitude modulation, radio broadcasting method
AMR	Adaptive multi-rate codec
AMR-NB	Adaptive multi-rate – narrow band speech codec
AMR-WB	Adaptive multi rate – wideband speech codec
AMR-WB+	Extension of AMR-WB
ARDOR	Adaptive rate-distortion optimised sound coder project
ARIB	Association of radio industries and businesses
ARQ	Automatic repeat request
ATIS	The alliance for telecommunications industry solutions
ATM	Asynchronous transfer mode
Bluetooth	Industrial specification for wireless personal area networks
BSAC	Bit Sliced Arithmetic Coding, an MPEG-4 standard (ISO/IEC 14496-3 subpart 4) for scalable audio coding
CCSA	China communications standards association
CPU	Central processing unit
CRC	Cyclic redundancy check
CS	Circuit switched
DANAE	Dynamic and distributed adaptation of scalable multimedia content in a context-aware environment project
DIA	Digital item adaptation
DMB	Digital multimedia broadcasting
DRM	Digital Radio Mondiale
DRM	Digital rights management
DSL	Digital subscriber line
DSLAM	Digital subscriber line access multiplexer
DTS	Digital theatre system
DVD	Digital versatile disc
eAAC+	Extension of advanced audio coding
EDGE	Enhanced data rates for GSM evolution
EGPRS	Enhanced general packet radio service
EVRC	Enhanced Variable Rate Codec
EVRC-WB	Enhanced Variable Rate Codec - Wideband
Enthrone	Integrated project in the thematic priority 'Information Society Technologies' of EU Framework programme 6 for research and development
ETSI	European telecommunications standards institute
EU	European Union
FDD	Frequency division duplex
FER	Frame error rate
FP6	6th Framework programme
FTTH	Fibre to the home
FM	Frequency modulation, radio broadcasting method
FRAND	Fair, reasonable, and non-discriminatory
G.722.1	Wideband ITU-T standard audio codec meant for low bit rate audio coding below 32 kbit/s with low complexity.

G.722.1.C	Super-wideband extension of G.722.1 meant for low bit rate audio coding below 64 kbit/s, with low complexity.
G.729	Narrow band speech codec
G.729.1	Wideband extension of G.729
GAN	Generic access network
GANC	Generic access network controller
GMSK	Gaussian minimum shift keying
GPRS	General packet radio service
GSM	Global system for mobile communications
H.263	ITU-T low bit rate video codec for video conferencing
H.264	ITU-T video codec equivalent to MPEG 4 Part 10, advanced video coding
HDTV	High definition television
HE-AAC v2	High efficiency advanced audio coding version 2 (near-identical to 3GPP eAAC+)
HiFi	High fidelity
HSPA	High-speed packet access
ICS	Internet conversation scenario
IETF	Internet engineering task force
iLBC	Internet low bit rate codec developed by Global IP Sound
IMS	Internet protocol multimedia subsystem
IP	Internet protocol
IPTV	Interactive personalized TV
iSAC	wideband adaptive speech codec developed by Global IP Sound
ISCD	Iterative source-channel decoding
ISIS	Intelligent scalability for interoperable services project
ISMA	Internet streaming media alliance
ISO/IEC	International standardization organization and International electrotechnical Commission
ITU-T	International telecommunication union – telecommunication standardization sector
ITU-T G.MMCC	ITU-T G. series multimedia convergence coder
KTH	Royal Institute of Technology in Stockholm
LDPC	Low-density parity-check code
LID	Layer independent descriptor
LTE	Long term evolution
MANET	Mobile ad-hoc networks
MBMS	Multimedia broadcast multicast service
MCfS	Multimedia conference scenario
MCvS	Mobile conversation scenario
MDS	Multimedia download scenario
MMBS	Mobile multimedia blogging scenario
MMSS	Multimedia multicast streaming scenario
MODSS	Multimedia on-demand streaming scenario
MP3	MPEG-1, audio layer 3, audio coding
MPEG	Moving picture experts group
M-PIPE	Research project taking on the problem of real-time media distribution over heterogeneous network conditions and end-user devices
MOS	Mean opinion score, numerical measure of the perceived quality of coded speech and audio material
MTSI	Multimedia telephony service for IMS
MUSHRA	Multi Stimulus test with hidden reference and anchor, methodology for subjective evaluation
NGN	Next generation networks
PAL	Phrase alternating line
PC	Personal computer

PS	Packet switched
PDA	Personal digital assistant
PHY	The physical layer of the Open systems interconnection basic reference model
PLR	Packet loss rate
PSTN	Public switched telephone network
QoS	Quality of service
QVGA	Quarter video graphics array (320x240) resolution
RFC	Request for comments documents in IETF
RTP	Real time transport protocol
SA4	System aspects 4
SSAC	Scalable speech and audio coder
SAE	System architecture evolution
SG	Study group (in ITU)
SGW	Security gateway
SIP	Session initiation protocol
SuS	Surveillance scenario
SVC	Scalable video coding
TDD	Time division duplex
TTA	Telecommunications technology association in Korea
TTC	Telecommunication technology committee, a telecommunications standards body in Japan
UDP	User datagram protocol
UDP lite	Variant of UDP that will deliver packets even if their checksum is invalid
UMA	Unlicensed mobile access
UMA	Universal media access
UMTS	Universal mobile telecommunications system
UTRA	Universal terrestrial radio access
UTRAN	UMTS terrestrial radio access network
VBR-EV	Variable bit rate embedded variable rate
VMR-WB	Variable rate multi-mode wideband speech codec
VoIP	Voice over IP
WCDMA	Wideband code division multiple access
WiFi	Wireless fidelity
WLAN	Wireless local area network
WP	Work package
WP	Working party
xDSL	see DSL

1 Introduction

This document gives an overview about the use-cases and scenarios that can benefit from the algorithms to be developed in the FlexCode project [27]. In addition, it provides a guideline for the FlexCode project about what type of flexibility is demanded by different scenarios, what the framework of the different scenarios is in terms of, e.g., technical requirements as bit-rate, error-rate, delay and so on. The document also provides a ranking of the scenarios that is based on their importance in terms of economical relevancy, feasibility and the potential gain the FlexCode paradigm can provide to the different scenarios.

One common denominator in many evolving and new service scenarios is the disconnection of a service from the supporting network. Previously, wire-line telephony was provided via dedicated circuit switched networks that interfaced with dedicated networks for mobile telephony. The Internet was yet another separate network that connected to the wire-line telephony network physically by the overlay of DSL while on the service level media or service gateways existed. These days we see the standardization of a number of enablers for ubiquitous services and all packet networks. One underlying reason for this trend lies in the transition of classical dedicated networks to all purpose packet based networks.

Both fixed and mobile networks become more and more centric to packet switched traffic. The third generation mobile network 3G and its long term evolution (LTE) focus to a large extent on packet based traffic. The data rates in mobile cellular networks are increasing, using high-speed packet access (HSPA) rates of up to 5.8Mbps on the uplink and 14Mbps on the downlink are current targets. For LTE peak data-rates of 50Mbps uplink and 100Mbps downlink are targeted. Further, mobile terminals will connect to a multitude of packet based networks, e.g., 3G networks and wireless fidelity (WiFi) networks. WiFi networks are based on IEEE802.11 standards where currently the 802.11g standard with its nominal rate of 54Mbps has the widest spread. A new standard 802.11n is currently developed with a nominal rate of 540Mbps. Another trend is the increasing popularity of ad-hoc connections and networks.

On the fixed side copper cables to homes are used more and more for DSL services. Connected to packet based core networks these DSL services lead to a conversion of circuit switched networks to packet switched networks. Lately fiber to the home (FTTH) is becoming a reality in some urban areas, with rates of the order of 100 Mbit/s in the downlink and from 10 to 100 Mbit/s in the uplink. Thus, in Europe many homes can be connected with packet switched networks that provide data-rates previously only available in core networks and local area networks.

ITU-T addresses issues that arise from this new all packet based world in the ITU-T Y specification series and a number of study groups targeting next generation networks (NGN) [15]. ITU like other institutes realized the need for other enablers in addition to the physical networks to realize and enhance services known from the legacy dedicated networks and create new ubiquitous services.

These enablers exist for example in the signaling domain where IP multimedia subsystem (IMS) is a corner stone [2]. IMS is a standardized next generation networking (NGN) architecture with a standardized implementation of the session initiation protocol (SIP) as a baseline. IMS provides essential functionality to implement network agnostic service scenarios. An indication of the range of networks and services IMS is targeting is given by the number of standardization bodies involved in IMS standardization, among them ITU, 3GPP, ETSI, and the IETF. Some of the scenarios described in this document highly depend on the availability of a system like IMS that allows the roll-out of a service over a wide range of access types. Still, the focus of the FlexCode project is on basic principles in source and channel coding. Thus, the exact implementation of the scenarios described in this document in IMS or other NGN realizations is not touched upon.

On the media delivery level other enablers like MPEG-21 digital item adaptation (DIA) [3] are essential breakthroughs in standardization. Utilizing MPEG-21 DIA a standardized way of adapting content (referred to as digital items within MPEG-21) to a variety of devices and networks is possible. Thanks to the open character of the MPEG-21 DIA standardization, the algorithms envisioned in FlexCode can serve as a valuable contribution to the DIA-Engine being at the heart of MPEG-21 DIA. The MPEG-21 DIA usage environment descriptions giving terminal capabilities, network characteristics, user characteristics and natural environment characteristics are then used to specify the operating points of the FlexCode codec. The flexibility of the FlexCode codec ensures that it can serve a maximum range of all these characteristics.

Some of the targets of FlexCode have been addressed in previous EU 6-th framework program projects. Examples of these, their relation to FlexCode, and potential usage of results from these projects are described in Sections 6.4 through 6.7. In summary, it can be said that there is still need for the development of a flexible source and channel coding system that is agnostic to the networks it operates on but still can adapt and thus deliver optimal performance to the conditions it faces.

It is to be noted that the viewpoints given in this document are based on the best knowledge of the contributing FlexCode members. They reflect the views and expectations from research and development organizations of the FlexCode industrial members. It has though to be understood that these views need not necessarily be fully aligned with the strategic marketing prospects of the respective industrial companies. In particular this is the case for the economical relevancy ranking of the various considered scenarios.

The remainder of this document is organized as follows: Section 2 gives a list of service scenarios and their high level description. These service scenarios and their technical framework are described in more detail throughout Section 3. In addition, we highlight how these scenarios can benefit from the algorithms that will be developed in FlexCode. In Section 4 the criteria for the ranking and the actual ranking of the scenarios are given. Then Section 5 provides an overview over the performance of some state of the art speech and audio codecs that serve as a performance reference for FlexCode. Section 6 describes the above mentioned relations of FlexCode to standardization bodies and other 6^{th} framework programme projects. We conclude with Section 7 summarizing the impact of the different scenarios and in particular the high-rank scenarios for the FlexCode project.

2 Service Scenarios

Below we provide a list of service scenarios and a brief textual description of the scenarios. Details like user perspective, equipment, networks, requirements etc are given in Section 3 for each scenario. It should be noted that the order of this list is not according to the ranking provided in Section 4, it rather groups the scenarios representing their functionality, i.e., person to many person communications first, followed by person to person communications again followed by download and streaming services and finally the surveillance scenario that did not fit well in any of the above categories.

2.1 Mobile Multimedia Blogging Scenario (MMBS)

It is envisioned that a user with mobile equipment generates and sends content from an experience, e.g., a concert, a sports event, a news event. The user sends the event either live or with a minimum of editing from a mobile device after recording. The content can be consumed as it is streamed by the user or it can be streamed or downloaded later on from a blog server.

In this scenario the framework is limited mainly by the sending side: The equipment has to be mobile and the data-rate and encoder complexity have to allow for real-time transmission. In the case of live transmission the rate is limited by the wireless uplink speed. A good margin to the limits of the uplink capacity should be kept since it can be assumed that the wireless network is working under considerable load due to the presence of many users at the event. The wireless network can be either a cellular 3G network or another wireless local area network. For the recording and off-line sending case the transmission speed should preferable exceed real-time. On the other hand, in this case higher data-rate due to hot-spot or WiFi connectivity can be assumed. For simplicity we assume the same rate for offline and online sending in this scenario.

2.2 Multimedia Conference Scenario (MCfS)

In this scenario it is envisioned that several users at different physical locations can setup a meeting that is perceived by all of the users similar to a face-to-face meeting. The fully featured multimedia conference scenario includes audio, visual, tactical information and file sharing, screen sharing, text chat. Not all functionality in terms of visual and tactical experience (shaking hand and alike) is currently realizable for the mass market. For the FlexCode project we restrict our scope to the audio part of the multimedia conference. Even though we expect speech to be the dominant content music should be supported to allow for, e.g., presentations that include music or even the situation where music is performed at one participating site.

For the speech and music part technology that enables 3D sound rendering is available, even though not all problems related to audio occurring in a multimedia conference scenario can be considered solved. Some of these problems fall into the fields of, e.g., multi-channel error-cancellation or sound-field analysis. Still, a reasonable demonstrator within the FlexCode project should be possible.

In this scenario, the realtime constraints are important. Thus the affordable delay is limited. Nevertheless, especially if one user is in a mobile environment, channel coding and error concealment are indispensable in order to guarantee error-free transmission. Especially if high-rate WiFi terminals are used at the receiving or sending side, channel coding becomes indispensable due to the nature of the wireless channel.

Example: Orange's livepresence (or "RealMeet room") service

The "RealMeet room" is a very-high-bandwidth IP videoconferencing system developed by Orange. It abolishes the concept of distance between interlocutors, and it allows speaking to them with eye to eye contact and getting the feeling that you can touch them. An illustration of this system is given in Figure 1.

Figure 1 Enhanced videoconferencing (RealMeet room) example.

From a technical point of view, RealMeet relies on several innovative features: real-size pictures (scale 1), scene captured from the normal position of the participants, eye-to-eye contact, spatialized sound using microphone arrays, 3D sound rendering and advanced echo cancellation, broadcast video quality.

This service, which was launched commercially in 2004, is proposed to businesses and various organizations. It is suited for work meetings and informal communications. However with the advent of very high-speed Internet (e.g. FTTH), similar high-quality video conferencing services can be envisioned in near future for the residential market as well.

Indeed broadband combined with voice over IP (VoIP) already enables a vast number of people to enjoy at least a "presence screen".

2.3 Mobile Conversation Scenario (MCvS)

This scenario describes a replacement for current circuit switched (CS) telephony. A call is setup by identifying a user on an address list and initiating a notification at the user's terminal. The address list can be either a buddy list with indication of presence or a classical telephone book like list. The content is mainly wideband speech. An advanced version of the MCvS includes multi-party calls. However, a FlexCode implementation of the scenario might omit this option. The differentiation to the Internet conversation scenario is that at least one of the users is limited by the bandwidth and computational power that is available in a mobile terminal and there is a stronger emphasis on the fact that the system has to interoperate with legacy CS networks. The channel coding algorithms that are used and described in the Multimedia Conference Scenario and the Internet Conversation Scenario need to be adapted in order to satisfy the constraints set by the limited computational power in the mobile terminals.

Currently 3GPP is finalizing a set of standards, e.g., [4], [5], providing the framework for a Multimedia Telephony service for IMS (MTSI). MTSI is closely related to the mobile conversation and the Internet conversation scenarios (see Sections 3.3 and 6.1 for more details). The MTSI provides a framework that allows for ubiquitous telephony in a standardized environment. The FlexCode codec can contribute to this environment by providing channel, speech, and audio coding techniques for the design of a codec used universally in MTSI applications.

Example: Orange's UNIK service

Orange's service called UNIK was launched on mass market in October 2006. This voice and data service is an instance of fixed/mobile convergence services combining the 2G mobile network (GSM for voice, GPRS/EDGE for data) and the WiFi home gateway network (Orange's Livebox). Customers can initiate and receive communications in the WiFi range of their home gateway and switch seamlessly to the 2G mobile network when leaving their home.

This example is a good illustration of dynamic access change during a multimedia session. In the case of UNIK, the Wifi radio access is switched to a GSM/GPRS radio access. However, more general access switching cases can be envisioned. This possibility provides strong motivation for considering flexible coding in multimedia services, even though in UNIK access switching is handled without the FlexCode paradigm.

The technology behind UNIK is referred to as Unlicensed Mobile Access (UMA), which is defined in 3GPP under the name of Generic Access Network (GAN). It allows seamless mobility with GSM/GPRS/EDGE. The network architecture of the UMA network is illustrated in Figure 2 below.

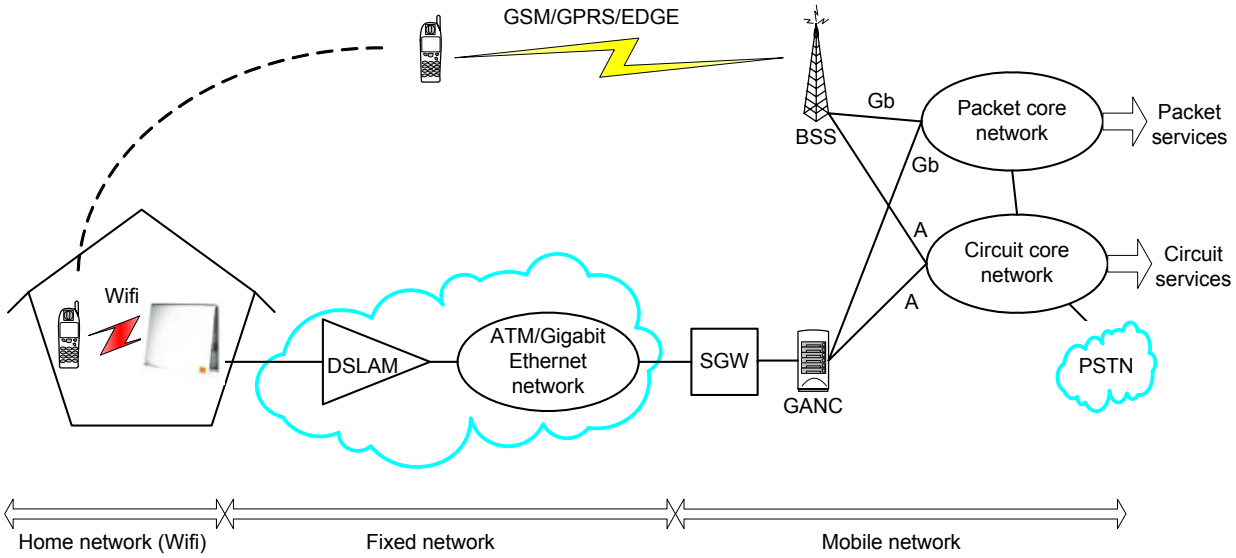


Figure 2 UMA network architecture used in UNIK

This architecture comprises three known parts: Wifi access (home gateway), ADSL access network (including a DSLAM and IP over ATM or Gigabit Ethernet), and 2G mobile network (GSM circuit and GPRS/EDGE packet). Two new equipments specific to the UMA network are introduced:

- the Generic Access Network Controller (GANC) which allows a UMA mobile to access voice services via the A interface and data services via the Gb interface

- the Security Gateway (SGW) ensuring UMA data confidentiality up to the terminal via an IPsec tunnel

The UMA streams of a UNIK client in the Wifi range are multiplexed with the other IP streams in the customer ADSL connection. Hence, an appropriate quality of service (QoS) policy is used to manage the bandwidth shared between all ADSL services.

Other similar services in Western Europe are: T-One (T-Mobile, Germany), Home Free (Telia Sonera, Denmark), Unica (TIM, Italy), BT Fusion (BT Mobile, UK).

2.4 Internet Conversation Scenario (ICS)

This scenario is similar to the mobile conversation scenario. The differences are the following:

- The address list is envisioned to be a buddy list with presence notification.

- Both users may utilize a powerful terminal (PC or alike) and have high-speed internet access. Another option is that mobile devices with WiFi connectivity are used. However, as a starting point FlexCode assumes the availability of a powerful terminal, see Section 3.4 for details.

Thus, in this scenario it is possible to show how the FlexCode paradigm performs when given maximum bandwidth and computational power.

Due to the maximum available computational power, elaborate channel coding and error concealment algorithms can be used. However, in today's VoIP systems using UDP and RTP, no source-channel coding is performed as the upper layers have no interaction to the lower layers (PHY) where the channel coding, error detection and packet rejection due to checksum failure, are performed. However, the emerging UDP lite protocol, which delivers erroneous packets to the source decoder, allows for the application of advanced error correcting algorithms and iterative source-channel decoding (ISCD). These algorithms even work if only hard decided bits are delivered by the lower layer. However, a cross-layer optimization which allows the transport of (possibly quantized) soft-information from the physical layer to the source decoder may allow even further gains if the ISCD paradigm is employed at the receiver.

The multimedia telephony service for IMS standardization by 3GPP, e.g., [4], [5] applies to this scenario in the same way as described for the mobile conversation scenario in Section 2.3.

Existing VoIP services such as Skype and Gtalk are examples of Internet conversational services.

2.5 Multimedia On-Demand Streaming Scenario (MODSS)

In this scenario a user selects combined audio and video content from a list of available content. After selection the content is streamed to the user. The user may be given the option of a pre-view consisting of a low-quality, low-rate version of the content. During pre-view the user can decide to view the high-quality version, either from the position the pre-view currently was at or from the beginning of the content. The switching between pre-view and high-quality content initiates different charging models.

The sessions in this scenario are point-to-point sessions where the server sends an individual stream to each user. This allows on-demand coding adapted to each user's network, terminal, and preferences. However, given that a single server should serve the highest possible number of users the individual complexity for each user should be kept at a minimum. Thus, it is favorable that not the entire source and channel coding process has to be repeated for each user. This is particularly important for live content where it is not possible to provision a variety of pre-stored formats.

The channel coding algorithms should be of low encoding complexity and of low to moderate decoding complexity, as the receiver may be integrated into a mobile device. However, as the receiver may also be integrated in HiFi home equipment with less demanding power saving constraints, flexible channel decoding algorithms should be used that allow different quality levels at different decoder complexities. Decoders employing iterative algorithms are perfectly suited for this case. The encoder needs no adaptation and the streams can be pre-encoded. The different receivers can use a different number of iterations to decode the signal. Battery-driven, low-cost, mobile receivers can use only a single iteration, by reducing however the error-robustness, while high quality stationary receivers can use the full number of exploitable iterations in order to guarantee the maximum possible quality of service for the given transmission channel constraints.

Typical examples of multimedia on-demand services are news services with selected stories such as sport events and interviews or video on demand.

Example: Orange TV service

Orange TV is a TV on Demand (TVoD) service to be launched in 2007. Among other features, two innovative services are:

- Network Personal Video Recorder (NPTV), which can be summarized as "The TV that I want when I want". It allows to record TV streams on a shared network resource. TV programs can be paused at any time, resumed later or scanned forward or backward. Advanced video navigation also allows to automatically seek specific time segments. The user can also start back in time if he missed the beginning of a program. An example of user interface associated with this service is shown in Figure 3.

- Network Time Shifting (NTS), which is a complement to NPTV. It allows to record one or several streamed contents. The storage space is located in the network (thus alleviating the need for a local hard disk). Network storage allows sharing recordings (e.g. several customers may record the same sequence), which optimizes storage capacity.

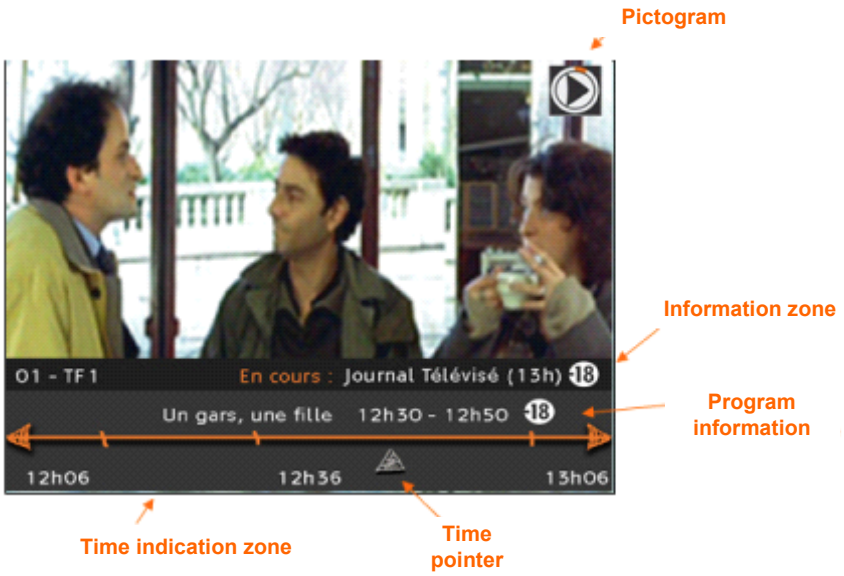


Figure 3 Example of user interface of TV on Demand (TVoD) service.

These two recording features can be controlled through a home gateway, a mobile terminal or even from a PC. All functionalities, including content visualization, may be accessed from a mobile or a PC. Video quality can vary depending on the access type (up to HDTV). Furthermore, advanced service interaction including Web 2.0 interface can be implemented, especially on high-speed access such as FTTH.

2.6 Multimedia Multicast-Streaming Scenario (MMSS)

In this scenario a server provides a data stream representing audio-visual information. Several users have the option to receive this single stream. The scenario is useful for services like mobile-TV or TV over packet networks. In fact, it is favorable in this scenario that the server can produce a single one stream for any type of user terminal or optional a multitude of hierarchical streams. In the latter case, high-quality high-bandwidth devices can subscribe to a multitude of hierarchical streams to exhibit their full potential. Mobile devices can subscribe to a subset of the streams or only to the basic stream to suit their current bandwidth situation and display setup. In case of one single stream this stream should be scalable such that network nodes and terminal devices can adapt the stream by removing the parts that exceed their capabilities. The single stream case requires network nodes that are capable of adapting the data stream. Such networks are described, e.g., in the MPEG-21 [1] standardization work or the M-PIPE project [39]. For the case of hierarchical streams the adaptation is limited to the user terminal. On the other hand, less advanced networks suffice. Multiple description coding could be used instead of truly hierarchical streams.

The channel coding in this scenario has to be highly adaptive and hierarchical. If only a subset of the stream is transmitted, the receiver should still be able to perform channel decoding. As multiple description source coding is a candidate instead of true hierarchical streams, the ISCD algorithms have to be adapted to the multiple description coding paradigm. Again, the use of iterative receiver algorithms enables the design of low-cost (low power consumption) to high-end receivers using the same algorithms, by only adapting the number of iterations used.

An example for a service involving multimedia multicast streaming is IPTV (Interactive Personalized TV) where the basic audio-visual content is sent to multiple users where it is augmented with individual information. Such individual information can include voting, messages sent between different users, personalized alerts, or personalized advertisement. This additional data content has to be protected especially well against transmission errors. In speech/audio/image/video transmission scenarios, the acceptable bit rate after channel decoding must not be equal to zero, as efficient error concealment algorithms exist. These algorithms subjectively enhance the signal quality by concealing possible bit errors. However, in data transmission scenarios, a bit error rate of almost zero should be a design target as the individual information to be transmitted must not be corrupted. Error detection algorithms have to be used alongside the correction algorithms in order to detect errors and possibly request retransmission of the data.

Example: Orange's Liveradio service
The liveradio [42] is a WiFi radio box (including batteries) that is connected to a wireless access point with high-speed Internet access, such as a home gateway (e.g.,Orange's LiveBox). With this mobile device, customers can freely listen to radio stations (including web radios), a selection of podcasts, audio books, ambiance sounds, ring tones, as well as MP3 stored on a USB player. The service – launched in Feb. 2007 - is made more interactive by the use of bookmarks and customization through a Web interface. The Liveradio box supports Windows Media, MP3 and WAV audio formats. The service is illustrated in Figure 4,

Figure 4 Liveradio (mobile radio box) using the home gateway as a Wifi access point.

The primary use of the radio box falls down in the Multimedia Multi-cast Streaming Scenario. However it also includes the Multimedia On-Demand Streaming and Download scenarios as use cases (e.g. podcasting).

2.7 Multimedia Download Scenario (MDS)

As opposed to the multimedia on-demand streaming scenario in section 2.5 in the download scenario the entire content is downloaded to the user terminal before it is rendered. This has some impact on the requirements on jitter, delay, and rate. In addition, it should be possible to render the downloaded content on different devices, i.e., playing sound on a mobile terminal should be computationally less complex than playing on 5.1 equipment. The received bit-stream should be adaptable to different device capabilities and transmission methods to the devices. Connections with mobile ad-hoc networks (MANETs), other wireless ad-hoc networks or point to point connections as, e.g., Bluetooth are likely to be exploited between the initial storage of the download and different rendering devices.

In this scenario, the delay constraints are not strict. A long acceptable delay means that powerful channel coding algorithms, e.g. Turbo codes with large interleavers or long LDPC codes, can be utilized. Another option might be the use of rateless codes, e.g. Fountain codes, or Hybrid ARQ (transmission of incremental redundancy) schemes. Such channel coding schemes allow the reduction of the throughput, as only as much redundancy is transmitted as is needed to successfully decode the block. Hybrid ARQ schemes can be realized using low-rate Turbo codes and puncturing which are envisioned to be used in FlexCode.

Typical examples of multimedia download services are music on demand (MoD) in the form of mobile music downloads, and video clips service.

2.8 Surveillance Scenario (SuS)

This scenario targets the surveillance of public places or areas as hotel lobbies, company lobbies and alike. The main challenge in this scenario is to reach high-compression rates while maintaining enough detail to recognize events and identify persons from the stored data. Most common only visual material is stored in surveillance applications. However, for certain situations it can be useful to enhance this information with audio material, especially given the relatively low data-rate of audio compared to video. For the audio material a typical challenge is the presence of severe noise at the place to be monitored.

Bad channel conditions can be expected if the surveillance equipment is installed for instance using wireless links. As the delay constraints are not as severe as in a conversational scenario, more powerful error correction might be employed. Furthermore, the sending equipment, which only performs the channel encoding, can be kept cheap as the (computationally more complex) decoding will in most cases be performed on a powerful processing unit, e.g., a workstation or a central server.

Within the FlexCode project we only consider the audio data from this scenario. We assume that there is no significant interaction on network congestion and alike between video and audio.

2.9 Other possible scenarios

The list of scenarios presented above covers a wide range of multimedia services with various characteristics that are very relevant for the FlexCode project. This list is however not exhaustive. The restriction to a selected subset of scenarios is actually justified by the focus of FlexCode on:

- Audio data (hence on audio-oriented services), and
- Services which can (a priori) benefit most from flexible joint source/channel coding.

It is worth noting that additional multimedia services that could be envisioned are often related to one or several scenarios listed above. For instance services considered as basic add-ons in IMS (e.g. Push-To-Talk, voice mail or voice network storage) share many aspects with Multimedia Mobile Blogging or Conversation Scenarios.

More advanced services, such as Multi-party Network Gaming, could be considered. Yet they often go far beyond the Flexcode considerations in terms of media type, service architecture and ease of implementation. Moreover it shares many aspects with share many aspects with the Multimedia Conferencing Scenario.

3 Detailed Description of Service Scenarios

In this section a more technical description of the scenarios is given. The descriptions of all scenarios follow a similar pattern to allow for a more easy comparison of the scenarios. The full implementations of some of the scenarios include multi-channel audio, video and other content type. Since the FlexCode project targets fundamental research in source and channel coding of speech and audio the focus in this section is on the mono speech and audio content.

Coding of multi-channel audio is typically done using parametric multi-channel codecs with a mono or in some cases a stereo codec in the core, i.e., the mono signal of the core codec is spatialized using the parametric multi-channel codec. Thus, within the FlexCode project focus is on the mono codec to provide a good basis for the multi-channel codec standardized elsewhere. The MPEG surround codec [7] is an example of a parametric multi-channel audio codec that is close to be standardized.

As outlined in the description of work [27] it is expected that many of the findings for the speech and audio source and channel coding are relevant for video coding as well. Thus, the scenarios including video are expected to gain twofold from the speech and audio codec developments; directly by the use of the techniques developed in FlexCode for the speech and audio coding and by porting the findings for speech and audio coding to video coding techniques. This work is to be done in task WP1.4 (see [27]). It appears reasonable that this task operates independent of the present document by examining the applicability of the methods developed within FlexCode for video coding.

The ordering of the scenarios in this section reflects the ordering in Section 2. The ranking is provided in Section 4.

3.1 Mobile Multimedia Blogging Scenario (MMBS)

User perspective:

- **Content:** Speech, audio, and background noise. Mostly all three of them simultaneously. In addition, video content not addressed in this scenario description.

- **Quality:** High-quality that is acceptable on home audio devices. The scenario is targeting stereo signals. However, a number of mobile recording devices will only provide mono sound. The sampling rate should be at least 16 kHz (wideband speech) or preferably 32 kHz.

Equipment:

- **Sending side:** Mobile phone or PDA including still camera, optionally coupled with a digital video camera.

- **Receiving side:** Diversity of devices ranging from mobile phone or PDA over PC to high-quality audio devices (e.g., 5.1 channel setup) and HDTV screens.

Networks

- **Sending side:** 3G WCDMA uplink radio access network or WiFi network or Bluetooth connection

- **Transport network:** Ethernet, other core networks including satellite links or microwave links, general Internet

- **Receiving side:** 3G WCDMA downlink radio access network, WiFi network, Ethernet or Bluetooth connection

- **Feedback capabilities:** It can be assumed that the downlink capability to the sending device exceeds the uplink capabilities and that the connection to the receiving device or blogg server allows for throughput sufficient for feedback purposes. Issues can arise from the fact that, in cases where the content is stored in a blogg server, feedback from the rendering user device is not available to the sending user device. In this case there are two instances of adaptation: During the transmission from the sending user device to the blogg server and during the sending from the blogg server to the rendering user device. An additional issue is the influence of potential long delays experienced when the Internet is involved. Examining the Stanford PingER statistics [35] we conclude that these delays are in the order of 25 to 200 ms.

Requirements:

- **Data-rate (for audio):** It is assumed that both video and audio are transported over a 3G HSPA enhanced uplink connection, WiFi or other WLAN. Currently the 3G network is the slowest of the networks mentioned. The data-rate for both audio and video should not exceed the rate possible on an enhanced uplink HSPA under reasonable load conditions.
  Typical data-rates for audio-visual content span a wide range depending on the application, network and rendering device. Below we list a few examples of applications and the data rates associated.
  Currently audio visual communications in WCDMA networks based on multimedia broadcast multicast service (MBMS) typically operate with 256 kbps bearer where 200 kbps are reserved for video, e.g., H.264, and 40 kbps for audio, either AMR-WB+ or e-AAC+. The remaining 16 kbps are reserved for overhead, e.g., IP headers. These rates assume the rendering on a mobile terminal with QVGA (320x240) resolution.
  For digital TV targeting the PAL standard MPEG2 codecs are more common, there a total source rate of about 4 Mbps with about 224 kbps spent for audio can be considered normal. For DVD audio-visual material MPEG2 video codecs and either MPEG2 or Dolby Digital AC3 audio codecs are commonly used. The variety of rates and channel setups is wide. For AC3 rates up to 448 kbps for a 5.1 channel setup and for MPEG2 rates up to 912 kbps for a 7.1 channel setup are standardized. In addition, many DVD players support the digital theater systems (DTS) codec.
  Table 1 summarizes the rates mentioned above. As can be seen the spread is large and for many applications the rates fall outside the range addressed in the FlexCode project (approximately 20 kbps to 60 kbps). Still, a quality as obtained using the AMR-WB+ or e-AAC+ audio codecs at a rate of 40 kbps for one channel is useful even for rendering the blogging content on home audio devices.

Table 1 Overview of rates and audio codecs for some applications with audio visual content.

Application	Total source rate [kbps]	Audio codec / rate [kbps]
MBMS over WCDMA, rendering on QVGA mobile terminals	≈240	AMR-WB+, e-AAC+ / 40
Digital TV (PAL)	≈4000	MPEG2 / ≈224
DVD	≈9000	MPEG2 / ≤912 (7.1 chan.) AC3 / ≤448 (5.1 chan.) DTS / ≤1536 (6.1 chan.)

- **Delay:** Since the blogging service is one-directional, delay is not a major concern. However, delay can be limited by the fact that it contributes to complexity and use of memory, which is limited in the mobile device. Since the Internet is involved in this scenario the network delay varies considerably. Values of a few milliseconds are as normal as several hundred milliseconds. Thus, delay and errors in forms of late packets are highly inter-related. In this scenario delay can be traded to the favor of a lower error-rate due to late packets. This additional delay does not contribute to encoder or decoder complexity but to memory requirements for the jitter buffer.

- **Error robustness:** For upload re-transmit, hybrid ARQ schemes or even rateless codes can be used. For the life case the scenario has to accept the error-rate provided by the network. Since we assume packet switched networks in this scenario the measure of interest is the packet-loss rate and the distribution of the packet loss. The number of packets lost for the decoder is the sum of the lost packets (packets never arrived) and late packets (packets that arrive when the decoder is already processing data later in time). Given that high delays at the receiving blogg server can be tolerated loss due to late packets can be neglected in this scenario. Still, the loss due to packets never arriving can be high.
  It is assumed that the link between the mobile user and the blogg server involves the Internet at some point. Further it can be assumed that this link and the wireless link generate the majority of the packet loss experienced. In [33] it is shown that the loss characteristics on the Internet are a function of the packet rate. Given that a typical audio coder emits packets in intervals of the order of 20 ms loss rates of 16% are observed [33]. In addition, it is likely that the losses are bursty. In [33] Balot found the likelihood that a packet is lost given that the previous packet was lost to be 42% for UDP packets sent every 20 ms. It is important to note that the measurements in [33] are 14 years old. Internal figures show that the loss rates in good networks are nowadays in the order of 0.1-0.01%., the Stanford PingER statistics from April 2005 show a large spread of loss rates ranging from 0.1% to approximately 3% (not considering outliers) depending on the locations the ping is exchanged between. These figures are very approximate since the characteristics of the Internet are highly varying with time and depend on the exact location and connection used. In addition, the Internet evolves such that the variety of delay and error-rates experienced broadens.
  An additional source of errors is the wireless link from the sending device. Here frame error-rates of the order of 1% to 5% are considered as normal operation values. An overview of end-to-end quality of service with focus on delay and error-rate is found in, e.g., [34].

In total the blogg service should be usable with a packet loss rate of as high as 8% (adding the 3% from the Internet to the 5% from the wireless link).

Existing codecs addressing this scenario:

As discussed the rate for the representation of audio-visual content varies widely depending on the rendering device targeted. This is particularly true for the video rate. As mentioned above, an audio rate of about 40 kbps from a codec that works well for both speech and audio and supports both mono and stereo encoding is a reasonable assumption for a scenario targeted within FlexCode. A direct match to these requirements are AMR-WB+ and 3GPP e-AAC+. Other audio codecs like, e.g., AAC can be used as well but are more appropriate for higher rate scenarios. In addition, a pure audio codec might not be the optimal choice since the bit-rate required for a given QoS is likely to be higher than for, e.g., AMR-WB+ due to the likely presence and for some situation even dominance of speech in the signal.

Potential benefit of FlexCode to scenario:

1. **Maximum exploitation of bottleneck upload channel:** As mentioned before the upload channel from the mobile device to the blogg server is likely to be a bottleneck in this scenario. Thus, it is important to exploit this channel with maximum efficiency. In addition, the bandwidth and other characteristics, e.g., error rate, of this channel are likely to be varying. The coder serving this channel is positioned on a mobile device with limitations of memory and computational complexity. In summary, it is important that one single codec can cover the varying characteristics of this upload channel with optimum performance.

2. **Rendering on heterogeneous devices:** Rendering on a variety of devices ranging from high-fidelity audio equipment to mobile phone speakers is envisioned. Since it can be assumed that the downlink capabilities to these devices will match or exceed the uplink capabilities of the source device, rate adaptation is not a major issue for the sake of fitting the data on the channel available to the rendering device. Still, in cases where the quality achieved with the complete bit-stream exceeds the rendering capabilities of a device, scaling of the stream can be advantageous to free network resources and save computational power and battery consumption.

3. **Source / rendering device mismatch:** In this scenario it is particular important to exploit the available channel with maximum efficiency. This is due to the fact that the rendering device can be of very high-quality while the source device has to be mobile and possibly connected via a cellular network. Thus, the rate available from the source device is close to the limit of what can be judged acceptable on some rendering devices. Since the FlexCode codec adapts to the upload channel it is ensured that this mismatch is kept at its minimum.

Standardization or emerging services related to this scenario:

The mixed audio and speech content expected in this scenario together with the envisioned range of networks fits well in the framework of the considerations of ITU-T SG16/WP3 Question 23 and the ITU-T G.MMCC mentioned in Section 6.2 under the condition that mobile devices are considered in the standardization. Furthermore the codec envisioned in the MPEG activity on a combined speech and audio codec falls in the range of the current scenario.

Commonalities with other scenarios:

- **Uplink similar to mobile conversation scenario:** The potential uplink capacity is equivalent to the one in the mobile conversation scenario, at least for the case of live upload outside WiFi hotspots. Still, it is likely that higher margins to the full uplink capacity are kept in the mobile conversation scenario to reduce network load. In addition the delay constraints are much more relaxed in the blogging scenario. The additional delay can be exploited for algorithmic delay in the source coder or channel coder or buffering to allow for retransmissions of lost data.

- **Encoder capability similar to mobile conversation scenario:** The principle limits of computational complexity in this scenario are the same as in the mobile conversation scenario. However, the necessity of high quality video and audio recording hardware places this scenario in high-end devices, while the mobile conversation scenario should work on mobile terminals with fewer capabilities. Moreover the delay constraints are much more relaxed in this scenario than in the mobile conversation scenario.

3.2 Multimedia Conference Scenario (MCfS)

Due to the multitude of other media types that can exist in the multimedia conference scenario, the detailed description given below is limited to audio aspects. The confined description is possible since we can assume that the sharing of network resources is less critical in this scenario than in, e.g., the mobile blogging scenario of Section 3.1. In addition, we assume that the recording equipment consists of mono devices. Thus, each participating site can be rendered at a distinct location in the 3D audio space. However, in cases where several speaker reside at one site, not every speaker can be isolated in the 3D space.

User perspective:

- **Content:** Speech, background noise, audio, multiple speakers

  In the multimedia conference scenario the use of handsets is rare. Thus, environmental noise and multiple speakers are common.

- **Quality:** Bandwidth of at least 8 kHz (wideband speech), high quality. Support for bandwidths of 16 or 24 kHz should be available in a commercial system.

Equipment:

- **Sending side:** Stationary device with mono input. For advanced systems where not only the different participating locations are rendered at different places in the audio scene but each participating person is rendered individually a multi-channel input is necessary. Since multi-channel audio is not a focus item in FlexCode we restrict the scenario to mono recording devices.

- **Receiving side:** Stationary device with multiple speaker setup, e.g., 5.1 speaker setup

- **Middleware:** Multimedia conference bridge. The functionality of this bridge can vary from providing presence information up to media transcoding and 3D rendering of the different participating parties or participants.

Networks

- **Sending side:** Ethernet connection or high-rate WiFi
- **Transport network:** general Internet, packet based core network
- **Receiving side:** Ethernet connection or high-rate WiFi
- **Feedback capabilities:** Since the conversation is bi-directional it is safe to assume that at any participating entity the network can support sending of feedback information for the received stream to the conference bridge or sender and receive feedback information for the stream it sends.
- **Data path:** In this scenario it is possible to stream the user data either directly between the user devices or via the multimedia conference bridge. These two routing options are depicted in Figure 5 and Figure 6. When streaming via the conference bridge two options for the 3D rendering are possible. One is to send the set of all mono streams to the receiving devices and perform 3D rendering there, another is to perform 3D rendering in the conference bridge. Table 2 provides an overview of the advantages and disadvantages for these options.

  We suggest the second alternative in Table 2, routing via the conference bridge but 3D rendering in the receiving devices. This not only renders mono codecs sufficient in the system, it also guarantees a simple and unified control of active speakers. The additional delay associated with this advantage is minimized and the network load is kept at a minimum. To compare the network load for transmitting a multi-channel signal the rate needed for spatial multi-channel information that is in the order of 32 kbps for a 5.1 signal [8] has to be considered. Furthermore the rate needed to transmit the core mono signal for a multi-channel signal is likely to increase since this signal consists of several speakers.

  In the suggested setup connecting legacy equipment such as mono equipment or a gateway to CS conference participants requires that these devices are able to multiplex the different mono signals themselves. In any case, this multiplexing process has to be performed either in the bridge or in the equipment, such that this question is not about additional complexity but where the required complexity is located. The CS gateway and similar devices can be integrated or moved very close to the conference bridge.

Table 2 Pros and cons for different data paths and locations for the spatialization.

Streaming via conference bridge and 3D rendering in the bridge	
Pros	**Cons**
• Unified spatialization • Lower data rate from conf. bridge to receiving devices in case of many active speakers • Only active speakers are forwarded to receivers.	• Transcoding in conf. bridge necessary • Additional delay due to transcoding • Spatialization and surround codec from conf. bridge to receivers outside FlexCode scope

Streaming via conference bridge and 3D rendering in receiving devices	
Pros	**Cons**
• Easy control of active speakers. • No transcoding functionality in conf. bridge necessary. • Mono codecs suffice in entire system. • Bridge basically provides routing. Thus, additional delay due to including bridge in traffic path minimized; no delay due to transcoding. • Only active speakers are forwarded to receivers	• Some additional delay when bridge is not in line of sight (longer path between sender and receiver). • For high number of active speakers rate to the receivers is high. However, number of active speakers should be limited.

Streaming directly between terminals, 3D rendering in receiving devices	
Pros	**Cons**
• Simple design of conference bridge. • Shortest delay of signals.	• Difficult control of active speakers. • High data-rates due to forwarding more streams than rendered (related to difficult control of active speakers). • More need for uniform implementation of receiving devices to ensure that all parties use the received streams in the same way.

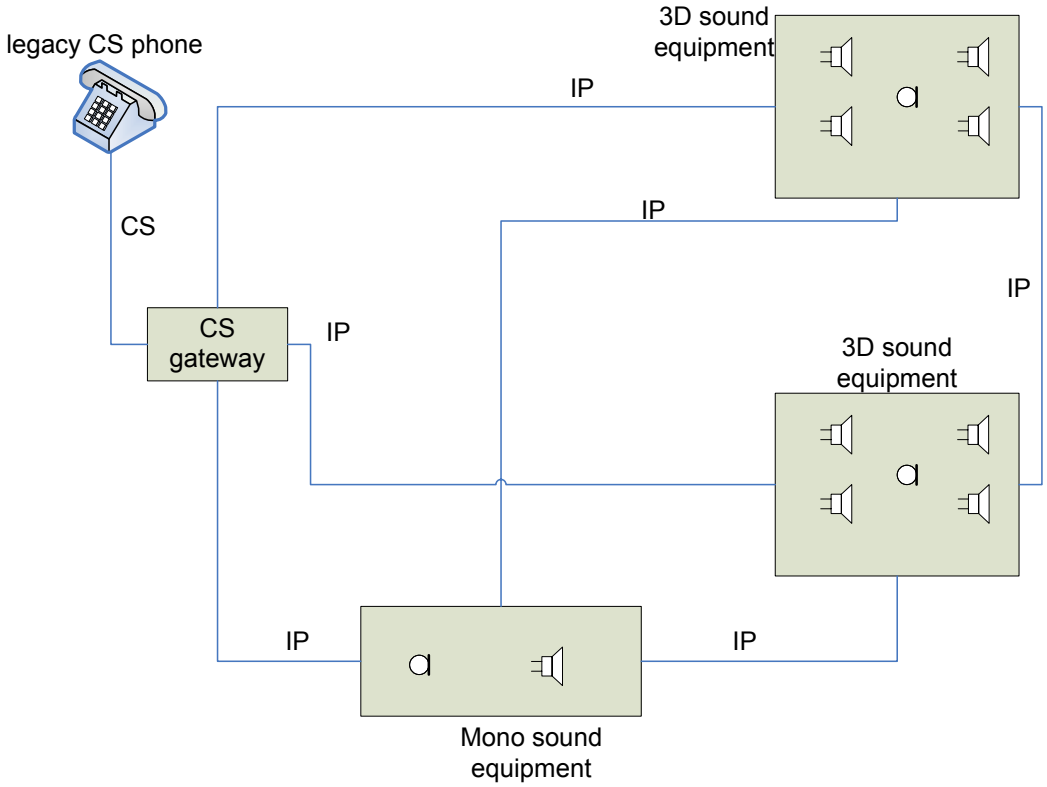


Figure 5 All interconnected conference system.

Requirements:

- **Data-rate (for audio):** In the following discussion we assume the suggested setup with routing via the conference bridge and 3D rendering in the receiving devices. In this case all the non-spatialized streams have to be sent to the receiver. In cases where the audio content consists of speech, data rates for a high quality speech codec in the order of 24 kbps should suffice. However, if multiple speakers are observed, high background noise conditions occur or audio signals are to be transmitted the conference experience should not degrade. Thus, the rate has to adapt to the content and is expected to increase up to 60 kbps for some signals. In addition, the possible presence of music means that the performance of a pure speech codec is not sufficient for this scenario.

- **Delay:** Preferably less than 200 ms, no more than 400 ms [6] end-to-end delay including jitter buffer, network delay, algorithmic delay other access delay. For the case of non-optimal general Internet connections average delays of 100 ms and more should be considered normal. Given the delays on the transport network and other access delays and the conference bridge operation the codec delay should be kept at a minimum and not exceed the AMR-WB algorithmic delay of 25 ms. If the content is dominated by music the delay could possibly be increased since less interaction can be expected in that case. It should be noted that simultaneous performance of music at several locations is outside the scope of this scenario since that case requires very short delays.

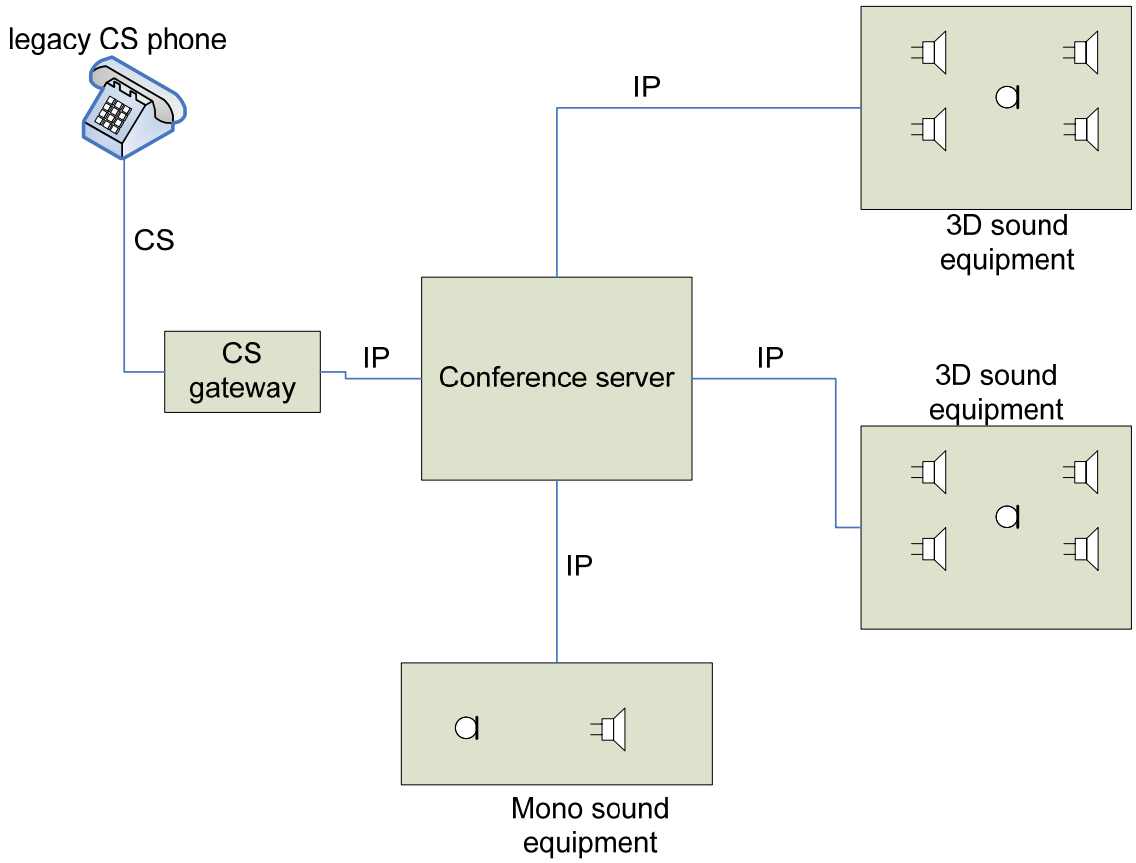


Figure 6 Data flow utilizing a conference server.

Error robustness: What is stated about error patterns in the Internet in Section 3.1 is valid for this scenario as well. For the FlexCode case we assume a good Internet connection in this scenario (e.g., close proximity to main network nodes) since this scenario is anticipated for companies and not for residential usage. Thus, the error-rates are to be expected lower than in, e.g., the Internet conversation scenario. In addition, WiFi connectivity is considered. Consequently, error rates can still be expected to be several percent, especially in situations where WiFi connectivity is used at the end user equipment. Since low-delay is important in this scenario additional packet loss due to late arrivals will be experienced. Dedicated packet based core networks of telecom operators should lead to better performance. In summary, error-rates in the order of 3-5% should be handled in this scenario.

As the content to be transmitted is mainly considered to be speech, audio and video, the error-rate can be reduced using source-channel decoding. Furthermore, a non-zero error rate is acceptable, as residual bit errors might be concealed in the source decoder or will be corrected in the iterative source-channel decoder. If the focus of the scenario is widened to high-rate WiFi networks at the sending or receiving side, high performance channel coding algorithms with a low coding rate can be used to combat the high error rates that might occur during bad channel conditions. Furthermore, if physical layer WiFi chipsets that allow to pass the soft-information after channel decoding to the higher layers are available, more elaborate source-channel decoding algorithms using soft-information are possible.

Existing codecs addressing this scenario:

Since the 3D rendering is performed in the receiving devices all wideband speech coders, e.g. AMR-WB (ITU-T G.722.2, 3GPP TS26.171), are addressing this scenario. However, these codecs typically perform relatively poor for music signals and a multimedia conference shall support these signals. In this respect speech and audio coders as, e.g., AMR-WB+ [9], and e-AAC+ [10] are interesting for this scenario. However, both these codecs do not satisfy the delay requirements for this scenario. Codecs with delays acceptable for this scenario and speech and audio capabilities that are sufficient are ITU G.722.1 and G.722.1.C. Conferencing is also a primary application of ITU-T G.729.1.

Potential benefit of FlexCode to scenario:

The following characteristics of the scenario lead to a potential benefit of using the FlexCode codec:

1. **Content variation:** The FlexCode codec should adapt to the content type (dominated by speech but music should not lead to performance breakdown) this adaptation should be automated and for the pure speech operation low delays are mandatory, for music delay or bit-rate increase can be treated against each other. In addition, background noise and multiple speakers are commonly observed in the conference situation. This flexibility and automated adaptation is not present in current codecs.

2. **Varying number of participants:** The number of participants varies for each conference and even during the conference. While the number of active speakers at each time instance might be influenced only slightly by the total number of conference participants the network load can be influenced. The FlexCode codec can adapt to the network load. However, using multicast from the conference bridge to the participants this problem is reduced.

3. **Varying number of active speakers:** The number of streams delivered to the receiving devices varies with the number of active speakers. Even though most of the time only one speaker is active and the allowed number of active speakers should be limited to three or four, the system has to work reliably and with good quality of service even for the case when the number of speakers reaches the maximum. The FlexCode codec can adapt the data-stream to the network load resulting from the number of active speakers.

4. **Different network and terminal capabilities to different participants:** Both the network connection and the used equipment can vary between participants. Using an advanced speaker setup, e.g., 5.1 setup, ensures the full functionality of the 3D conferencing, some users can be connected with mono speakers and others even with legacy circuit switched equipment. The FlexCode codec should address all these equipment. As mentioned before multichannel audio is outside the FlexCode scope and is covered by, e.g., MPEG surround.

5. **Life encoding and utilization of feedback can be accommodated:** In this scenario the source is encoded and consumed in real-time. Thus, an encoding for the circumstances (network, terminals and alike) at hand is possible. In addition, the feedback collected from all networks and terminals involved can be distributed to the encoding device. It should be noted that one single data stream is encoded for a number of terminals. Thus, it is necessary to adapt the stream such that an overall optimum is achieved or the worst case is covered. Another option is the adaptation of the stream in the conference bridge. This should be possible with minimum computational effort.

6. **Conference recording:** One feature of conferencing systems is recording for archive purposes. Using FlexCode the recording with lower quality and less 3D rendering and a resulting low-rate should be effortless.

Standardization or emerging services related to this scenario:

G.711EV in ITU-T Q.10/16

Commonalities with other scenarios:

- **Network similar to Internet conversation:** In case the general Internet is used as transport network the delays and error-rates experienced are similar to these parameters in ICS. The difference in this scenario is the routing via the conference bridge. Given that the bridge does not perform transcoding and that the network provisioning is sufficient this routing should not result in a major source of error or delay. Thus, delay, error-rate, and to some extend bit-rate are similar to the ICS.

- **Content type similar to Internet conversation scenario:** In both scenarios speech is envisioned to be the main content while audio should be supported as well. A difference is the frequent appearance of background noise and multiple speakers in the present scenario.

- **Computational resources similar to Internet conversation scenario:** The form factor of the hardware is similar to the one used in the Internet conversation scenario. The differences are as follows: In this scenario dedicated hardware is envisioned; in this scenario several instances of the decoder have to be running in the receiver and the 3D rendering has to be performed as well. Thus, even though in both cases high-performance hardware with little battery restrictions can be assumed in the multimedia conference scenario there will be less room for the FlexCode codec on this hardware.

3.3 Mobile Conversation Scenario (MCvS)

User perspective:

- **Content:** Speech, background noise, audio. In addition, video content not addressed in this scenario description.

- **Quality:** High quality, for speech meaning at least a sampling rate of 16kHz (wideband speech). Most of the conversational devices support mono signals which are the focus of the FlexCode project. However, in commercial versions multi channel audio should also be considered in the audio rendering. There even artificial localization and mixing of different speakers in space in teleconference type service should be considered. This content is eliminated from this scenario to differentiate it towards the MCfS. Narrowband speech with a sampling frequency of 8 kHz has to be considered for compatibility with legacy CS systems.

Equipment:

- **Sending side:** Mobile phone
- **Receiving side:** Mobile or fixed phone

Both sending and receiving sides involve devices requiring low complexity algorithms, which hardens the task of the codec addressing this scenario. However, given the timeframe of the project and the evolution of computational resources, two parallel solutions could be proposed, one taking into account the full potential of the coding/ decoding method and the second one being complexity cost constrained.

Networks

- **Sending side:** 3G uplink radio access network, WiFi
- **Transport network:** PSTN/NGN, core networks, general Internet
- **Receiving side:** 3G downlink radio access network, PSTN/NGN for fixed receiver, WiFi
- **Handover and protocol issues:** handover between different networks, mixing of several input channels in case of teleconference
- **Feedback capabilities:** Adaptation to channel and source characteristics as well as user requirements

Requirements:

- **Data-rate (for audio):** The bit-rate overlaps with the Multimedia Conference Scenario in Section 3.2. However, in the MCfS higher rendering quality of multi-speaker and audio signals is considered and less stringent network conditions can be expected. Thus, the rate range for this scenario is between 10 kbps and 32 kbps.
- **Delay:** Since the scenario is conversation oriented, the delay is crucial to the quality of the communication. Requirement for mouth to ear one way delay is between 100ms and 300ms [6]. This leads to an algorithmic delay of about 25-40 ms.

- **Error rate:** Tolerable error rate depends on the utilized codec, concealment methods and error statistics. In classical CS services typically, 1% frame error rate is acceptable. Especially in mobile scenarios, the wireless transmission links can cause severe distortions resulting in high error rates. In addition, cases where the general Internet is involved can cause high packet loss rates. Thus, the scenario should address packet loss rates of at least up to 8%. Better channel coding algorithms can reduce the packet losses. If the erroneous bitstream and eventually reliability information on these bits can be delivered to the source-decoder, near-capacity iterative joint-source channel decoding becomes possible.

Existing codecs addressing this scenario: AMR and AMR-WB for 3GPP, EVRC, VMR-WB and EVRC-WB for 3GPP2

ITU-T G.727 is used in DECT cordless phones. Next Generation DECT will support ITU-T G.727, G.711 for narrowband speech. G.722, G.729.1 for wideband speech and MPEG AAC-LD for superwideband speech.

Potential benefit of FlexCode to scenario:

1. **Different network capabilities / conditions**: The network connection can vary between participants and during one session.

2. **Life encoding and utilization of feedback**: Same as multimedia conference scenario. One difference is that here we assume encoding tailored for the channel and device capabilities / conditions of one receiver.

3. **Multiparty conversation**: In case the scenario is extended to allow two or more participants in the conversation, bandwidth should be distributed between one or more channels and the FlexCode codec assures that this is done seamlessly. In addition, adaptation to several receivers is necessary.

4. **Adaptation to environment noise conditions**: The perceptual model in the FlexCode codes ensures that the perceived overall high quality of speech is preserved even in fluctuating environmental noise conditions as the speaker moves without considerable increase in bitrate.

Standardization or emerging services related to this scenario:

The 3GPP is currently in the process of finalizing the multimedia telephony service for IMS (MTSI). This set of standards provides a framework for the implementation of the mobile conversation scenario. Since the MTSI is ubiquitous it faces a wide range of network capabilities and conditions, a variety of user terminals, guaranteed QoS levels and so forth. The techniques that will be developed in the FlexCode project can support the wide range of conditions faced in the MTSI. The mobile conversation scenario is a scenario illustrating these conditions well.

Commonalities with other scenarios:

- **Content** similar to Internet conversation and multimedia conference scenarios.

- **Delay** similar to Internet conversation and multimedia conference scenarios.

- **Computational requirements** similar to mobile blogging requirements in the lower complexity case.

3.4 Internet Conversation Scenario (ICS)

The Internet Conversation does not differ much from the typical mobile conversation scenario in Section 3.3. In fact, both scenarios can be part of the same fixed/mobile converged service. The requirements for the content, QoS and user experience are the same. The equipment and connectivity does not necessarily include mobile terminals or wireless networks. In cases where no mobile terminals and wireless connections are present, the complexity requirements may be relaxed compared to mobile conversation. In addition, the rate may vary more flexible when necessary. However, Internet conversation does not explicitly mean PC based terminals.

User perspective:

- **Content:** Mainly speech. Music can occur and should be supported. Situations where participants share music as part of the conversation or play a piece of music from radio or HiFi equipment to share it with the other participant.
- **Quality:** High-quality wideband, 8 kHz bandwidth or super-wideband, 16 kHz bandwidth

Equipment:

- **Sending side:** Personal computer or WiFi phone
- **Receiving side:** Personal computer or WiFi phone

Even though the scenario potentially includes WiFi phones we assume that computational complexity is of minor importance in this scenario. This is to be able to show the full potential of the paradigm developed in FlexCode. In case the complexity exceeds the capabilities of a mobile device a further step is to reduce the complexity and evaluate and minimize the performance loss due to this step.

Networks

- **Sending side:** xDSL, optic fibers and WiFi or Ethernet
- **Transport network:** general Internet
- **Receiving side:** xDSL, optic fibers and WiFi or Ethernet
- **Feedback capabilities:** Given that the conversation is bi-directional and that the rate of the feedback information can be considered low relative to the rate for audio the sending and receiving of feedback should be possible.

Requirements:

- **Data-rate:** To deliver high-quality wideband speech current codecs operate in the region of 24 kbps, see Section 5. In case the content is dominated by audio signals the rate can increase up to 60 kbps.

- **Delay:** Preferably less than 200 ms, no more than 400 ms [6] end-to-end delay including jitter buffer, network delay, algorithmic delay other access delay. For the case of non-optimal general Internet connections average delays of approximately 100 ms for the network should be considered normal. Given these delays on the transport network, the delays due to WiFi access and other access delays the codec delay should be kept at a minimum and not exceed the AMR-WB algorithmic delay of 25 ms. For pure music content the delay requirements might be relaxed.

- **Error robustness:** What is stated in previous sections about error patterns in the Internet is valid for this scenario as well. Additional transmission errors for the case of WiFi connectivity of the user equipment have to be considered.

  Delay is a critical parameter in this scenario. Thus, the packet loss rate is influenced by packet loss due to late arrivals, given good jitter buffer strategies late-loss can be kept below 1%.

  In summary, the upper limit for what can be considered a “normal” network does exceed packet loss rates of 3%. The error rate is mainly determined by the packet loss of the Internet and WiFi connections. We envision that this scenario can address PLRs of 8% or higher. Eventually, bit errors on (possibly) mobile intermediate transmission links will cause packet losses in traditional UDP transmissions, even if only single bit errors have occurred which could easily be correct in a source-channel decoder. However, new standards might not discard such packets and allow the source-channel decoder at the end-user terminal to correct those errors.

Existing codecs addressing this scenario:

There exist proprietary codecs that are claimed to specifically address the conditions faced in VoIP over the general Internet, i.e. high packet loss and varying transport delay. In most cases no detailed information about these codecs is published, an exception is the iLBC codec [41] standardized at the IETF (RFC 3951, 3952).

For codecs optimized for high quality at low bitrate with moderate packet loss rate (i.e. the codecs used for mobile telephony) there are other means to improve the robustness to high packet loss rates and other conditions faced in the Internet. An example are redundancy schemes supported by the RTP payload formats. The AMR codec family is very well suited for VoIP applications. Its bitrate flexibility enables adding redundancy to combat high packet loss conditions without increasing the total bitrate. Figure 7 shows the result from a subjective listening test illustrating that AMR-NB can perform well in high packet loss conditions. Recently AMR codecs are standardized by 3GPP for VoIP usage in [37].

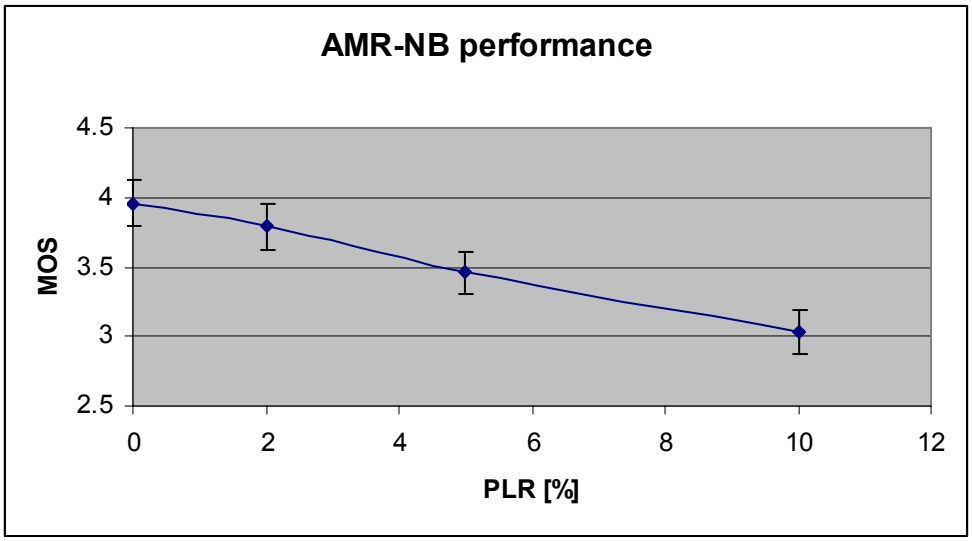


Figure 7 AMR-NB at a gross rate (not including IP overhead) of ≤15 kbps under packet loss conditions. The gross rate consists of AMR-NB encoder rate added with simple redundancy schemes for packet loss conditions. Source: Ericsson AB

Potential benefit of FlexCode to scenario:

1. **Varying network qualities:** Since the Internet is a highly varying medium in terms of throughput, delay, error-rate, jitter, error-patterns high advantages are expected and have already been proven for a source and channel codec adapting to the conditions at hand. When considering the usage of WiFi connectivity even higher variations of these parameters have to be handled.

   Due to the varying network conditions some codecs designed for VoIP usage balance the compromise between performance under packet loss conditions and coding efficiency in error-free conditions towards better performance under packet loss conditions. This can be compensated for by higher bit-rates such that good quality is achieved in the error-free case. The FlexCode codec should not have a pre-set trade-off in this respect but adapt to the channel conditions observed. This should help to optimize the rate / quality trade off under any observed channel condition to minimize network load.

2. **Content variation:** The FlexCode codec will adapt to the content type (dominated by speech but music should not lead to performance breakdown). For pure speech operation low delays are mandatory, for music delay or bit-rate increase can be treated against each other. This flexibility and automated adaptation is not present in current codecs.

3. **Exploitation of feedback:** The main focus of this scenario is the conversation between two users. In this case the encoder can adapt exploiting feedback from the receiver about network conditions and user equipment.

4. **Multiparty conversations:** If the scenario is extended to allow two or more participants in the conversation advantages similar to what is described for the MCfS (Section 3.2) and the MCvS (Section 3.3) can be obtained.

Standardization or emerging services related to this scenario:

The multimedia telephony service for IMS (MTSI) currently standardized in the 3GPP provides a framework for the implementation of the Internet conversation scenario. Basically, the arguments for MTSI given in Section 3.3 are valid for the Internet conversation scenario as well.

Commonalities with other scenarios:

- **Network similar to multimedia conference scenario:** See Section 3.2
- **Content type similar to mobile conversation scenario:** In both scenarios speech is envisioned to be the main content while audio should be supported as well.
- **Computational resources similar to multimedia conference scenario:** In both scenarios focus is on high-performance hardware. In this scenario the main hardware platform is a generic PC while in the conference scenario dedicated high-performance hardware is envisioned. In a second step mobile equipment should be supported in this scenario setting some limit to computations.

3.5 Multimedia On-Demand Streaming Scenario (MODSS)

User perspective:

- **Content:** Mixed Content (Speech, Noise, Music), Audio. Stereo and multichannel configurations (e.g. 5.1 channels) are very likely. Video should be addressed as well in this scenario.
- **Quality:** Depending on content. For audio content such as short news reports over slow links at least good narrowband quality. For previews (of movies, etc.), at least good AM radio quality. In general, high-quality sound (FM radio or DVD quality). The sampling rate should be at least 8 kHz in specific speech-oriented cases, 16 or 24 kHz for previews, 32, 44.1 or 48 kHz in the general case.

Equipment:

- **Sending side:** Streaming server.
- **Receiving side:** Diversity of terminals, e.g., mobile phones, Wifi devices, audio and video players, HDTV screens.

Networks

The sender is considered to be part of the transport network.

- **Transport network:** Core networks, general Internet
- **Receiving side:** Low to high-speed downlink radio access, xDSL, optic fibers.

Requirements:

- **Data-rate (for audio, mono signals):** depending on input audio sampling rate: 8-16 kbit/s for 8 kHz case, 12-32 kbit/s for 16 kHz, 14-56 kbit/s for 32 kHz and 16-64 kbit/s for 48 kHz. If the material is not only audio, but also video, bit rates for audio may be more limited. In case of multichannel audio, rates for the spatial information should be similar to the rates used in MPEG Surround.

- **Delay:** Since the streaming service is one-directional, delay is not a major concern. However, near-constant jitter is desirable for good quality of service and delay should be limited, especially for mobile devices, as it contributes to complexity and use of memory.

- **Error rate:** Possibility to retransmit lost packets. However, delay associated with retransmission should be acceptable for mobile devices with limited buffering capability. In particular in the case of a mobile receiver, the error rate / delay trade off should be properly optimized.

  The error rate is mainly determined by the end-user who can be either connected to the streaming server by an Internet connection at home or use a wireless link at home or even use a mobile terminal connected to mobile radio networks. Power error correction schemes can be used due to the large allowable delay. Furthermore, hybrid ARQ schemes can be used in order to reduce the throughput: in good channel conditions, only a little amount of data needs to be transmitted while in bad channel conditions, an incremental amount of redundancy can be transmitted.

Existing codecs addressing this scenario:

AMR, AAC, 3GPP e-AAC+, AMR-WB+, Windows Media Technologies, MPEG Surround

Potential benefit of FlexCode to scenario:

The following characteristics of the scenario lead to a potential benefit of using the FlexCode codec:

1. **Adaptation to content characteristics**: The input format may vary depending on the material.

2. **Adaptation to receiving characteristics:** Coding flexibility will allow to access multimedia content seamlessly alleviating the end-user from any unnecessary complexity. The adaptation to terminal capabilities should also be facilitated.

3. **Optimized error rate / delay tradeoff :** In general low error rates could be achieved by using longer delays. The related buffering could however be prohibitive on some mobile devices. The Flexcode concept can solve this problem by optimizing properly the robustness / delay trade-off (especially on wireless links).

4. **Different QoS requirements:** The FlexCode codec should allow for seamless switching between different QoS requirements due to, e.g., different charging models. An example is the switching from preview to full quality view.

Standardization or emerging services related to this scenario:

MPEG scalable audio, MPEG-21/DIA content adaptation engine, ISMA streaming engine

Commonalities with other scenarios:

- **Downlink and equipment similar to Multimedia Multicast-Streaming Scenario:** The variety of content type, terminals and downlink capacity is similar to the one in the multicast-streaming conversation scenario. The relaxed delay can be exploited for algorithmic delay in the source coder or channel coder or buffering to allow for retransmissions of lost packets.

- **Signals similar to multimedia multicast streaming scenario and multimedia download scenario.**

3.6 Multimedia Multicast-Streaming Scenario (MMSS)

User perspective:

- **Content:** Mixed Content (Speech, Noise, Music), Audio. Stereo and multichannel configurations are very likely (5.1 and other). Video should be addressed as well in this scenario.

- **Quality:** Depending on content. For audio content such as short news reports over slow links at least good narrowband quality. For previews (of movies, etc.), at least good AM radio quality. In general, high-quality sound (FM radio or DVD quality). The sampling rate should be at least 8 kHz in specific speech-oriented cases, 16 or 24 kHz for previews, 32, 44.1 or 48 kHz in the general case.

Equipment:

- **Sending side:** Streaming server.

- **Receiving side:** Diversity of terminals, e.g., mobile phones, audio and video players, HDTV screens.

Networks

The sender is considered to be part of the transport network.

- **Transport network:** core networks, general Internet

- **Receiving side:** LAN, low to high-speed downlink radio access, xDSL, optic fibers

Requirements:

Streaming tends to stress the network as constant rate and jitter are desirable for good quality of service. To reduce network loads and target a variety of networks and terminals, mid-rate codecs are more attractive. This scenario is especially well suited for embedded media coders (with hierarchical bitstream).

- **Data-rate (for audio):** Depending on input audio sampling rate: 8-16 kbit/s for 8 kHz case, 12-32 kbit/s for 16 kHz, 14-56 kbit/s for 32 kHz and 16-64 kbit/s for 48 kHz. If the material is not only audio, but also video, bit rates for audio will be more limited. In case of multichannel audio, rates for the spatial information should be similar to the rates used in MPEG Surround.

- **Delay:** Since the streaming service is one-directional, delay is not a major concern. However, delay can be limited to few hundred milliseconds in case of interactive services. In addition the receiving side may have little computing power and memory which puts a strong constraint on buffering (especially if video is used)..

- **Error rate:** Relatively high error rates might be acceptable in the case of audio/video transmission as those residual errors might be concealed by error concealing algorithms present in the source decoder. However, it might be beneficial to use powerful (iterative) source-channel decoding in order to increase the overall perceived audio/video quality and to eliminate the need for error concealment as the ISCD algorithms considered incorporate implicit error concealment. In the case of, e.g., IPTV, this scenario specifies additional binary content (the personalized information) to be transmitted. This additional content must be protected extremely well, as no bit errors are acceptable. Hybrid ARQ schemes might be a candidate to solve this problem.

Existing codecs addressing this scenario:

AMR, AAC, HE-AAC v2, AMR-WB+, Windows Media Technologies, MPEG Surround

BSAC used in Digital Multimedia Broadcasting (DMB) applications, AAC+ in DRM

Potential benefit of FlexCode to scenario:

1. **Adaptation to content characteristics**: The input format may vary depending on the material.

2. **Adaptation to receiving characteristics:** Coding flexibility will allow to access multimedia content seamlessly alleviating the end-user from any unnecessary complexity. The adaptation to terminal capabilities should also be facilitated.

3. **Optimized error rate / delay tradeoff :** In general low error rates could be achieved by using longer delays. The related buffering could however be prohibitive on some mobile devices. The Flexcode concept can solve this problem by optimizing properly the robustness / delay trade-off (especially on wireless links).

4. **Optimized encoding work load**: The number of encoders in the streaming server is minimized and does not depend on the number of connected clients. This assumes that the media encoder has a specific coding format suitable for this kind of service, e.g. embedded (or hierarchical) coding or multiple description coding.

Standardization or emerging services related to this scenario:

MPEG scalable audio, MPEG-21/DIA content adaptation engine, ISMA streaming engine, MBMS in UMTS

Commonalities with other scenarios:

- **Downlink and equipment similar to Multimedia On-Demand Streaming Scenario:** The variety of content type, terminals and downlink capacity is similar to the one in the multicast-streaming conversation scenario. The relaxed delay can be exploited for algorithmic delay in the source coder or channel coder or buffering to allow for retransmissions of lost packets.

- **Signals similar to multimedia on demand streaming scenario and multimedia download scenario.**

3.7 Multimedia Download Scenario (MDS)

User perspective:

- **Content:** Mixed Content (Speech, Noise, Music), Audio. Stereo and multichannel configurations are very likely (5.1 and other). Video should be addressed as well in this scenario.

- **Quality:** Depending on content. For commercial services, most likely high-quality sound (FM radio or DVD quality). The sampling rate should then be at least 32, 44.1 or 48 kHz.

Equipment:

- **Sending side:** Content server on general Internet or operator network.

- **Receiving side:** Diversity of terminals, e.g., mobile phones, solid-state audio and video players.

Networks

- **Transport network:** core and access networks (FTTH, xDSL, 3G…) , general Internet

- **Receiving side:** Ethernet, ad-hoc network (Wifi, Bluetooth…)

Requirements:

- **Data-rate (for audio):** Variable bit-rate coding is possible. High-efficiency coding techniques are desirable to reduce the download time. For the constant bit rate case, expected values are in the range 14-56 kbit/s for 32 kHz and 16-64 kbit/s for 48 kHz sampled signals. In case of multichannel audio, rates for the spatial information should be similar to the rates used in MPEG Surround.

- **Delay:** Since the download service is one-directional and implies an offline playback, delay is not constrained. The coding algorithm should enable features such as preview as soon as the download started.

- **Error rate:** No errors. Retransmission of lost packets is possible. As large delays can be accepted in this scenario, powerful error correcting codes using, e.g. large interleavers in the case of Turbo codes, can be utilized. Such codes are able to operate very close to the Shannon limit. The decoded content at the receiving side shall preferably be absolutely error-free. If very powerful, low-rate channel codes are used during very good transmission conditions, data rate and bandwidth are wasted. To avoid this waste, either adaptive error correcting codes which change their rate according to the current channel conditions can be used or rateless channel coding schemes like, e.g., digital fountain codes, can be used. Another rateless approach which might be a candidate (and could easily be incorporated in existing Turbo coding schemes) is a Hybrid ARQ scheme. In such schemes, additional redundancy is sent until the receiver is able to perform error-free decoding (which can be guaranteed using error detection codes, e.g. CRC checks).

Existing codecs addressing this scenario (for audio):

MP3, AAC, MPEG Surround, Windows Media Technologies

Potential benefit of FlexCode to scenario:

1. **Adaptive error correction**: Advanced techniques such as Turbo codes or Hybrid ARQ can be implemented to avoid retransmission and to reduce the download time.

2. **Adaptation to receiving characteristics:** The adaptation to terminal capabilities should also be facilitated by the flexibility introduced by the coding format developed in Flexcode.

Standardization or emerging services related to this scenario:

MPEG scalable audio

Commonalities with other scenarios:

- **Content type and quality similar to Multimedia Multicast Streaming Scenario:** In general users will expect high quality content, including stereo or multichannel material.

- **Functionality very close to Multimedia On-Demand Streaming Scenario:** The main difference is that content can be viewed progressively with the on-demand streaming, while content is first downloaded entirely in the download scenario.

3.8 Surveillance Scenario (SuS)

User perspective:

- **Content:** Mainly speech distorted with background noise.

- **Quality:** Two levels are of interest: For monitoring the highest quality the equipment and channels provide is desirable, the main limitations are costs. For storage intelligibility is the main measure, i.e., the rate should be minimized to allow intelligibility of the coded source. Pre-processing for noise reduction and other speech enhancements are part of this scenario. Narrowband (4 kHz bandwidth) signals are sufficient.

Equipment:

- **Sending side:** Dedicated low-power, low CPU capability hardware.
- **Receiving side:** Monitoring device, e.g. a PC or dedicated hardware and digital storage device without decoding.

Networks

- **Sending side:** Wireless connectivity via, e.g., WiFi or Bluetooth
- **Transport network:** Packet based dedicated network
- **Receiving side:** Ethernet connection

Requirements:

- **Data-rate:** Since cost and efficiency are the main properties this scenario strives for low rates in the order of 4-10 kbps.
- **Delay:** Since this service is uni-directional delay is not a major concern. However, the permissible complexity and storage is low.
- **Error rate:** The error characteristics in this scenario are dominated by the wireless connectivity of the sending device. It can be expected that packet loss rates in the order of 0-10% occur. The transport network can be considered very reliable.

  As the delay constraints are not as severe as in a conversational scenario, more powerful error correction might be employed. Furthermore, the sending equipment, which only performs the channel encoding, can be kept cheap as the (computationally more complex) decoding will in most cases be performed on a powerful processing unit, e.g., a workstation or a central server. However, there is a severe complexity constraint due to the desire to accommodate a high number of sending devices on one receiving processing unit.

Existing codecs addressing this scenario:

All low-rate speech codecs potentially address this scenario, e.g. AMR-NB, AMR-WB. An additional requirement here is the robustness to background noise.

Potential benefit of FlexCode to scenario:

- **Adaptation to background characteristics:** The FlexCode codec's operation should adapt to the amount and characteristics of the background noise. For this scenario part of the flexibility could consist of adaptation of the pre-processing.
- **Adaptation to channel conditions:** The characteristics of the wireless link from the sending device will be varying for different device locations. The FlexCode codec should adapt automatically to these characteristics.
- **Adaptation to rendering / storing device:** It should be possible to obtain better quality and higher-rates for sending devices actively monitored than for the ones that are just stored.

Commonalities with other scenarios:

This scenario is rather unique in its target and operation conditions compared to the other scenarios listed in this document. One common factor with the MMBS and the MCfS is the focus on speech signals with strong background noise. Nonetheless, in the SuS the background noise can be expected to be most severe.

3.9 Details Overview

Table 3 and Table 4 provide an overview over some of the aspects of the scenarios. Further elaborations are found in Sections 3.1 through 3.8.

Table 3 Overview of the scenarios, user perspective, equipment and network.

Scenario	User perspective		Equipment		Network		
	Content	Quality	Sender	Receiver	Sender	Transport	Receiver
3.1 MMBS	Speech, audio, background noise, often simultaneously	High-quality acceptable on home devices. Stereo signals, some mobile devices provide mono only. Bandwidth: 16-32 kHz	Mobile device (phone, PDA including still camera, optionally connected to digital video camera)	Diversity of devices: Mobile, PC, high-quality audio devices (e.g., 5.1 channel setup)	3G WCDMA uplink radio access, WiFi	Packet core networks, Internet	3G WCDMA downlink radio access, WiFi or Ethernet
3.2 MCfS	Speech, background noise, multiple speakers, audio	High-quality, Bandwidth: 8-24 kHz	Stationary device, mono input. (See 3.2)	Stationary device with multiple speaker setup, e.g., 5.1.	Ethernet or high-rate WiFi	Packet core networks, Internet	Ethernet or high-rate WiFi
3.3 MCvS	Speech, background noise and audio	High-quality, mono, Bandwidth: 8 kHz or more	Mobile phone	Mobile or fixed phone	3G uplink radio access, WiFi	PSTN/NGN, core networks, general Internet	3G downlink radio access, PSTN/NGN for fixed receiver, WiFi
3.4 ICS	Mainly speech, audio should be supported	High-quality, mono, Bandwidth: 8 kHz or more	PC or WiFi phone	PC or WiFi phone	xDSL, optic fibers, Ethernet or WiFi	Internet	xDSL, optic fibers, Ethernet or WiFi
3.5 MODSS	Mixed Content (Speech, Noise, Music), Audio	High-quality, mono but mostly stereo, multi-channel, Bandwidth: 8-24 kHz	Streaming server	Diversity of devices: Mobile (phone, WiFi), dedicated audio / video players, HDTV screens		Packet core networks, Internet	Low to high-speed downlink radio access, xDSL, optic fibers

Scenario	User perspective		Equipment		Network		
	Content	Quality	Sender	Receiver	Sender	Transport	Receiver
3.6 MMSS	Mixed Content (Speech, Noise, Music), Audio	High-quality, mono but mostly stereo or multi-channel, Bandwidth: 8-24 kHz	Streaming server	Diversity of devices: Mobile (phone, WiFi), dedicated audio / video players, HDTV screens		Packet core networks, Internet	Low to high-speed downlink radio access, xDSL, optic fibers
3.7 MDS	Mixed Content (Speech, Noise, Music), Audio	High-quality (FM-radio or DVD quality), mostly stereo or multi-channel, Bandwidth: 16-24 kHz	Content server	Diversity of devices: Mobile (phone, WiFi), dedicated audio / video players		Core and access networks (FTTH, xDSL, 3G, ...), Internet	Ethernet, WiFi, ad-hoc network, e.g., Bluetooth
3.8 SuS	Speech distorted with background noise	Medium for monitoring, low for storing	Low power, los CPU hardware	Monitoring device (PC, dedicated hardware), storing device	WiFi, Bluetooth	Packet based core network	Ethernet connection

Table 4 Overview of scenarios in terms of rate, delay, error-rate, existing codecs and advantages from the FlexCode codec to the scenario.

Scenario	Requirements			Existing Codecs	FlexCode advantage
	Rate	Delay	Error-rate		
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

Scenario	Requirements			Existing Codecs	FlexCode advantage
	Rate	Delay	Error-rate		
3.4 ICS	10-60 kbps	200 – 400 ms end-to-end ➔ ≈ 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 ➔ a few hundred ms	Low PLR of ≈ 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 ➔ a few hundred ms	Service usable at PLR of ≥ 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 ➔ 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	≤ 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

4 Scenario Ranking

The goal of the present document is to provide a list of real world scenarios, a ranking of these, and to select the two most relevant scenarios [27]. Given the ranking, the list becomes an ordered list of service scenarios. The ranking of a scenario is defined based on several criteria that are specified in the present section. In the previous sections the order of the scenarios was not according to these criteria but according to the types of scenarios.

4.1 Ranking Criteria

The factors contributing to the scenario selection are manifold, expressing views of the operator, consumer, regulator, as well as the developer. By developer we understand the parties involved in the realization of this project. The following paragraphs list some requirements that help in the selection and ranking of the scenarios.

The consumer point of view is expressed through the type of services, perceived quality of services and the price attached to them:

A. Consumer needs/demands/requirements.

Several requirements issued from the operator and regulator point of view are addressed within different standardization bodies and projects [38], [39], [40]. These, combined with practical aspects that will be further presented, will help define the criteria used for ranking the scenarios.

The definition of the NGN as “a packet-based network able to provide telecommunication services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies” [15] points to the following requirements:

B. Use of IP: The use of packet based networks is seen as a means of convergence both within NGN and IMS.

C. Independence of service related and transport related technologies

The heterogeneity of the networks, terminals and, ultimately, people who consume and interact with the information presented to them gave birth to the standardization efforts in MPEG 21 related to the digital item adaptation (DIA). This work relates to the adaptation to different variable aspects involved in information communication. The standard specifies only the tools that assist the adaptation process, and adaptation engines should be provided elsewhere. This is the place where FlexCode could contribute by taking into account the following items:

D. Adaptation to input/output device capabilities: Aspects such as the variability of audio input and output devices in terms of signal quality should be taken into account at the encoder and decoder sides and provide means of seamless adaptation. In addition the variability of storage and computational capabilities should be considered.

E. Adaptation to network capabilities: To improve transmission efficiency adaptation mechanisms at the coding/decoding ends to e.g. lower delivery bandwidth or lower network capacity should be available.

F. Adaptation to network conditions: Similarly to F, in view of the transmission efficiency, adaptation to variable network conditions such as available bandwidth, error rate, or delay characteristics should be provided.

G. Adaptation to user characteristics and preferences: The variability of user preferences in terms of audio material presentation, or user characteristics such as possible audio impairments, should also be taken into account.

H. Adaptation to natural environment characteristics: Mechanisms of adaptation to different noise levels or noise frequency spectrum, both as characteristics of the input audio environment as well as of the output audio environment should enhance the quality of the relevant audio material.

I. Quality improvement in VoIP (NGN). Given the convergence trend toward Internet as communications universal network, a great importance is given to VoIP systems. Today's VoIP systems provide good quality only at relatively high bit rates, and higher quality at lower bitrates is still a demand which could create even new equipment and service market.

J. Voice + Audio + Video convergence (NGN). From FlexCode point of view it implies single voice and audio codec.

K. Voice fixed-mobile convergence/substitution (Unik and alike systems), see section 2.3.

Practical aspects which should be taken into account when defining the criteria are related to the availability of resources allocated to the project:

L. Feasibility : Practical aspects, such as the existence of related know-how within the project participants, the time framework should be taken into account for the realization of a fruitful collaboration..

Additional general aspects should also be considered:

M. Backward compatibility: This is an issue related to resulting the FlexCode codec, which might be considered in the planning, but it is not really a requirement.

Table 5 summarizes requirements B through L for all scenarios. All the criteria related to the needs (consumer/operator) are obviously subject to the feasibility (requirement L); therefore the feasibility is put on a different dimension of Table 5. While the requirements B through L are not exactly criteria for selecting a scenario or another, they will guide at the detailed description of the scenarios toward meeting the present and future communications demands. Requirement A is missing in Table 5 since it is used as a ranking criterion in Table 6.

The values for the criteria B through K in Table 5 are varying between 0 (not needed), 1 (possible use), and 2 (vital need). The values for each scenario are derived from the detailed description of the scenario in Section 3. Most of the requirements in Table 5 are related to adaptation capabilities, which are some of the main advantages brought by FlexCode, a scenario where the requirements are "vital" would be deemed to be more interesting.

While Table 5 gives an overview of the requirements and their feasibility, the ranking of the scenarios is performed according to the following selection criteria:

- **Operator interest:** How interesting is the scenario for operators in terms of service offer and differentiation, expected revenue and market positions.

- **End-user interest:** How much advantage and interest have consumers from the improvement and realization of the scenario.

- **Manufacturer interest:** How interesting is the scenario for equipment manufacturers in terms of expected revenue and market positions.

- **Degree of novelty** with respect to the existing solutions of the scenario: How much can the scenario gain from the FlexCode paradigm.

- **Ease of implementation:** Is it feasible to create an implementation that reflects the conditions of the scenario and its requirements within the FlexCode project.

The above criteria are varied between very low (0), low (1), medium (2), high (3) and very high (4) in Table 6. The values in parenthesis denote the number of ranking points associated. To obtain the values in Table 6, both the detailed description in Section 3 and internal information from the FlexCode industrial partners are used. It should be noted that the values in Table 6 do not necessary reflect the strategic marketing prospects of the FlexCode industrial partners. They reflect joined estimates of the research organizations of the partners.

Table 5 Feasibility of requirements B-K within different scenarios (0: not needed, 1: possible use, 2: vital need)

	B Use of IP	C Indep. of service and transport	D Adapt. to I/O device capab.	E Adapt. to network capab.	F Adapt. to network cond.	G Adapt. to user charact.	H Adapt. to natural env.	I Quality improve. in VoIP	J Voice/ Audio converg.	K Voice fixed/ mobile conv.
L: Feasibility	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
3.1 MMBS	1	1	1	1	1	2	2	1	1	1
3.2 MCfS	1	2	2	2	2	1	2	1	1	1
3.3 MCvS	**1**	**1**	**2**	**2**	**2**	**1**	**1**	**1**	**1**	**2**
3.4 ICS	2	2	2	2	2	1	1	2	1	2
3.5 MODSS	**1**	**2**	**2**	**2**	**2**	**2**	**0**	**1**	**1**	**1**
3.6 MMSS	1	2	2	2	2	2	0	1	1	1
3.7 MDS	1	1	1	0	0	2	0	0	1	1
3.8 SuS	1	0	1	1	1	0	2	1	1	0

Table 6 Scenario ranking. Ranking points: Very low = 0, low = 1, medium = 2, high = 3, very high = 4

	Operator interest	End user interest (general showcase)	Manufacturer interest	Degree of novelty (FlexCode advantage)	Ease of implementation	Ranking points
3.1 MMBS	Medium	High	High	Medium	Medium	12
3.2 MCfS	High	High	Medium	High	Medium	13
3.3 MCvS	**Very High**	**High**	**Very high**	**High**	**Low**	**15**
3.4 ICS	Medium	High	Medium	High	Low	11
3.5 MODSS	**High**	**Very high**	**High**	**High**	**Medium**	**15**
3.6 MMSS	Very High	High	High	Medium	Low	13
3.7 MDS	Medium	High	Medium	Low	Medium	10
3.8 SuS	Low	Medium	Low	Medium	High	9

For the operator and manufacturer interests the grading in Table 6 are motivated as follows. (The ranking X / Y below gives the operator / manufacturer scores.)

- **MMBS** Medium / High: The usage of multimedia mobile blogging will increase the demand for high-end terminals and boost the utilization and demand of high-speed cellular networks. Operators will have some interest in providing such an innovative service that is a logical continuation of the Internet blogging trend.

- **MCfS** High / Medium: The conference equipment market is important but does not have a similar volume as the markets addressed by some of the other scenarios. Operators will keep an important interest in providing conferencing applications for businesses, especially with enriched sound (e.g. 3D sound). With the advent of (very) high-speed links, including FTTH, similar services will certainly be offered on mass market.

- **MCvS** Very High **/** Very high: Mobile and fixed conversation are the killer applications in today's telecom industry and can be expected to remain one of the (if not the) most important applications even in the future. Consequently revenue from equipment utilized in these applications can be considered very high. The expected transition towards wideband quality (instead of narrowband quality) for speech will certainly stimulate this market for operators.

- **ICS** Medium / Medium: The revenue for pure Internet conversation equipment such as WiFi phones is uncertain. Most equipment utilized for Internet conversation (PC, broadband connection) is already present at most end-users. Still, it is important for manufacturers to understand and be part of this market to be able to foresee and influence its effect on other segments. The quality of service is a key issue for an operator and clients, and pure Internet conversation will not satisfy most quality expectations. Mobile conversation is by far more important for operators. However operator clients will use this service as part of a multiple-play Internet offer.

- **MODSS** High / High: Services as mobile TV and music streaming become ever more popular and customized to clients. Consequently service need, terminal demand and high-rate cellular network utilization will increase.

- **MMSS** Very High **/** High: The argument is similar as for the MODSS, with main focus on the mobile TV multicast application.

- **MDS** Medium / Medium: Download of multimedia content to a large extend happens via PC and broadband connections, however download will also take place on mobile devices with high-speed links. The user interface and storage capability provided by PCs or high-capacity devices is likely to make them the platforms of choice for many users. Even considering that for, e.g., rural places a trend towards replacement of DSL or other fixed line broadband with high-speed cellular network access is observed the equipment sales resulting from this scenario are judged as moderate. This service will be offered by operators, it also has a limited added value, but it has an attractive value / cost ratio.

- **SuS** Low / Low: Audio-based surveillance will be a limited market for both operators and manufacturers.

The scenarios with highlighted text in both Table 5 and Table 6 are the highest ranked scenarios with respect to the considered requirements and selection criteria. Using Table 6 the list of scenarios could be further ordered with decreasing ranking points. However, we abandon this option since the criteria used for ranking are of very different nature and their relative weighting is arguable. For the real-time implementation the ease of implementation and the degree of novelty (FlexCode advantage) are very important. In this respect the multimedia on demand streaming scenario (MODSS) is a good choice. The mobile conversation scenario (MCvS) appears somewhat harder to implement but its other parameters are high enough to make it one of the most interesting scenarios for the FlexCode project.

4.2 Highest ranked scenarios

As mentioned before, one of the main goals with this document is to judge the ranking of the scenarios presented and select the highest ranked scenarios in the Flexcode project. In line with [27] and to keep a reasonable number of scenarios for consideration and focus, we select the two highest ranked scenarios. Based on the presented figures in Table 5 and Table 6, the two most relevant scenarios are the MCvS and the MODSS.

The requirements and operating conditions for source and channel coding for the other scenarios do overlap to some extend with the two selected ones. The MCvS, the Internet conversation scenario (ICS), and the multimedia conference scenario (MCfS) are closely related (see Table 3 and Table 4). In fact, the main differentiating factor of the ICS are the increased terminal capabilities; the potential presence of circuit switched networks is the most prominent differentiating factor of the MCvS; and the focus on multi-part conversations differentiates the MCfS. However, as mentioned in Section 3.4, when considering WiFi phones the terminal capabilities are limited in the ICS as well. In practice a majority of the traffic of the MCvS will be routed via core-networks much more reliable than the general Internet. However, considering that the MCvS includes the case where the traffic is routed via the general Internet and ignoring the higher demands on the support of music in the ICS, the codec requirements resulting from the ICS represent a subset of the requirements of the MCvS. Since multipart conversations are potentially possible in the MCvS some of the functionality of the MCfS is included in the MCvS. It should be noted that at this point we do not consider multipart conversations as part of the FlexCode implementation of the MCvS.

The MODSS addresses one directional communications with some focus on high quality audio signals. This is true for the multimedia multicast scenario (MMSS) and the multimedia download scenario (MDS) as well. The differences of these scenarios are however slightly larger as for the conversational scenarios. For example is the implementation effort higher for the MMSS and encoding adapted to one specific receiver is not possible in the MMSS. MDS implementation is less involved, the expected advantage from FlexCode on the other hand is low.

5 State of the Art Coder Properties

As stated in [27] the performance of the codec designed within FlexCode is supposed to be similar to the performance of state of the art speech and audio codecs. Thus, the bit-rate and error-rate versus perceived quality properties are not described for each scenario in sections 2 and 3, instead the performance for some of the state of the art codecs addressing the scenarios is summarized in this section. Section 3 contanis details about what codec addresses which scenario.

We review the performance of several types of codecs:

- Multi-rate speech codec for conversational purposes, ITU-T G.722.2 / 3GPP TS26.171 adaptive multi-rate wideband codec (AMR-WB) [14]

- Multi-rate speech / audio codecs for conferencing and streaming applications, ITU-T G.722.1 Main Body / Annex C

- Embedded speech codec for conversational purposes, ITU-T G.729.1

- Speech and audio codecs for non-conversational purposes, 3GPP enhanced AAC plus (e-AAC+) [10] and extended AMR-WB (AMR-WB+) [9].

- Audio codec for non-conversational purposes advanced audio coding (MPEG-4 AAC) main profile.

- Embedded audio coding for non-conversational purposes, MPEG-4 bit-sliced arithmetic coding (BSAC)

A summary of the properties of these codecs is found in Table 7. The performance of the conversational speech codec AMR-WB, the non-conversational speech and audio codecs AMR-WB+ and e-AAC+, and the AAC audio codec are characterized in [29], [28], and [25], respectively. Characterization test results can be found in [22] for G.729.1, in [23] and [24] for G.722.1 Annex C. The performance of MPEG4 BSAC is evaluated in [25].

It should be noted that the codecs listed here are highly optimized and tuned for their respective operation conditions. To achieve a comparable performance with the FlexCode codec is a very ambitious task due to a number of reasons. One is the fact that the FlexCode codec's operation conditions are much wider than for most conventional codecs. Another reason is that codec optimization is an engineering task that requires time and manpower. The FlexCode project has to divide its time and manpower between the research of new methods as outlined in [27] and this optimization. However, focus will be on the development of new methods such that performance gaps that can be overcome with further tuning of the codec should be accepted in the resulting FlexCode codec.

Table 7 Codecs addressing the scenarios

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

We summarize the performance of the codecs considered both under error-free conditions and for erroneous channels. There are several considerations that make the comparison of the different results difficult. For audio codecs standardized in MPEG the encoder is not specified, thus different encoder implementations lead to different results. In addition, MPEG codecs are highly configurable such that the performance at a given bit-rate can vary to a large extend dependent on the codec configuration. The differences in equipment, material, number and experience of listeners from one test to another make the following statement from [28] valid when comparing the results from different characterization tests:
"*In the reporting of subjective test results, it is generally agreed that comparisons of results are valid only for conditions conducted within the same experiment. It is not valid, for example, to directly or statistically compare subjective test results for one codec across two bit-rates when those results have been obtained from different experiments.*" Still, gathering a number of different tests can be informative.

A further difficulty in the direct comparison of the test results are the different test methodologies used for different types of codecs and source material. While in speech coding mean opinion scores (MOS) are common, e-AAC+ and AMR-WB+ are characterized with the more recent MUlti Stimulus test with Hidden Reference and Anchor (MUSHRA) [31]. For the characterization of audio coders methods standardized in ITU-R [32] that include an absolute category rating (ACR) with scroes similar to the MOS used to be common. Recently MUSHRA procedures become more common even in this field.

Describing codec performance under error conditions is not a trivial task since the random nature of channel errors complicates the comparison of experiments from different characterization studies. In addition, the channel models used and their associated statistics vary from characterization study to characterization study. Furthermore, these tests often include interleaving and other techniques to reduce the effects of packet loss. The different usage of these techniques further obscures the possibility to compare results.

Generally, a distinction has to be made between circuit switched (CS) and packet switched (PS) networks. In traditional PS networks, channel coding is performed on the lowest layer. An error detection algorithm detects possible transmission errors. In the case of an error, the packet is either re-requested (in non realtime environments) or marked as lost. The source decoder has to deal with lost packets and take appropriate measure to recover the lost information. For this reason, the packet loss rate (PLR) is a performance measure commonly used in PS networks. On the other hand, in CS networks, it is common that the channel decoded (possibly erroneous) bitstream is delivered to the source decoder which itself has to handle possible biterrors. This situation is advantageous for the application of (iterative) joint source-channel decoding. Single bit errors causing a packet loss in PS networks can be corrected in the source-channel decoder in CS networks, which usually increases the QoS.

However, in PS networks, new technologies like UDP lite allow the source decoder to access the (possibly) erroneous bitstream and enable technologies like iterative joint-source channel decoding at the receiver. Furthermore, new cross-layer optimization aspects might provide so-called soft information, i.e., reliability information on the single bits, to the source-channel decoder, which furthermore enhances the decoder performance and thus the QoS.

5.1 Characterization of AMR-WB (3GPP 26.171, ITU-T G.722.2)

The AMR-WB codec is standardized both by ITU and the 3GPP. We show an excerpt of the results of the 3GPP characterization [29]. For error-free channels Figure 8 and Figure 9 summarize the AMR-WB performance compared to G.722 and G.722.1.

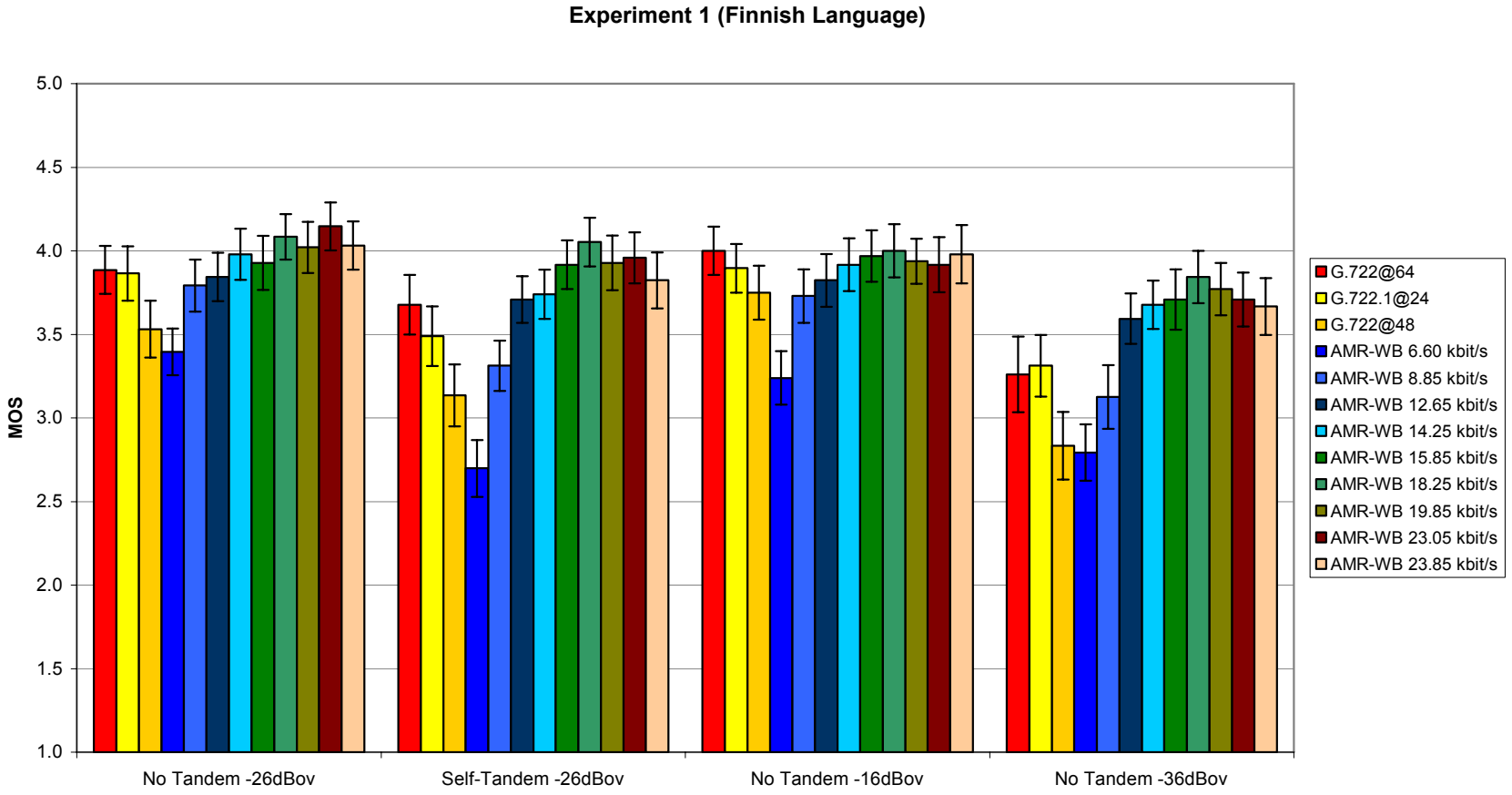


Figure 8 AMR-WB characterization test result from 3GPP with different input levels and self tandeming. Finnish language. The bars indicate the 95% confidence intervals. From [29].

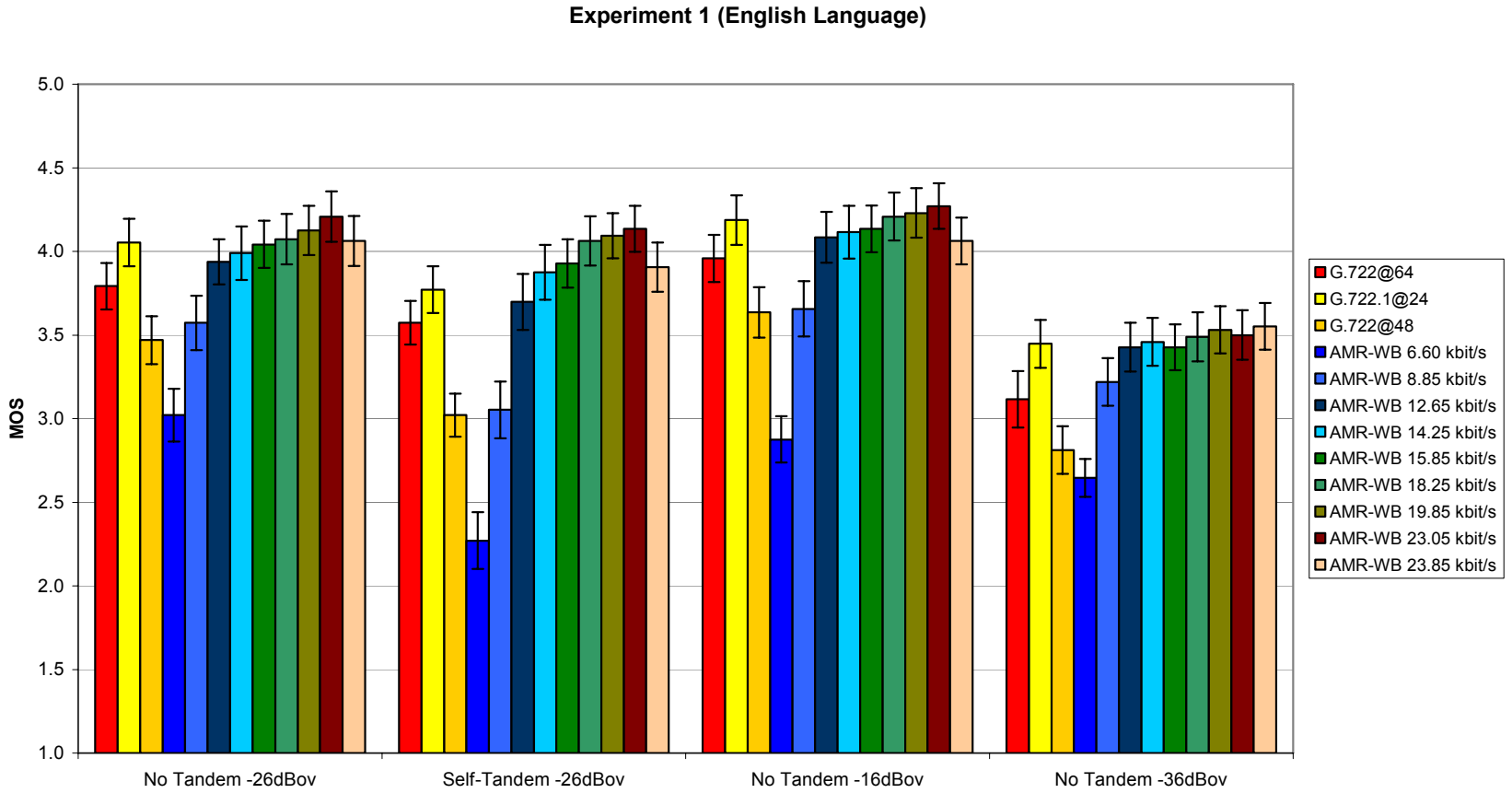
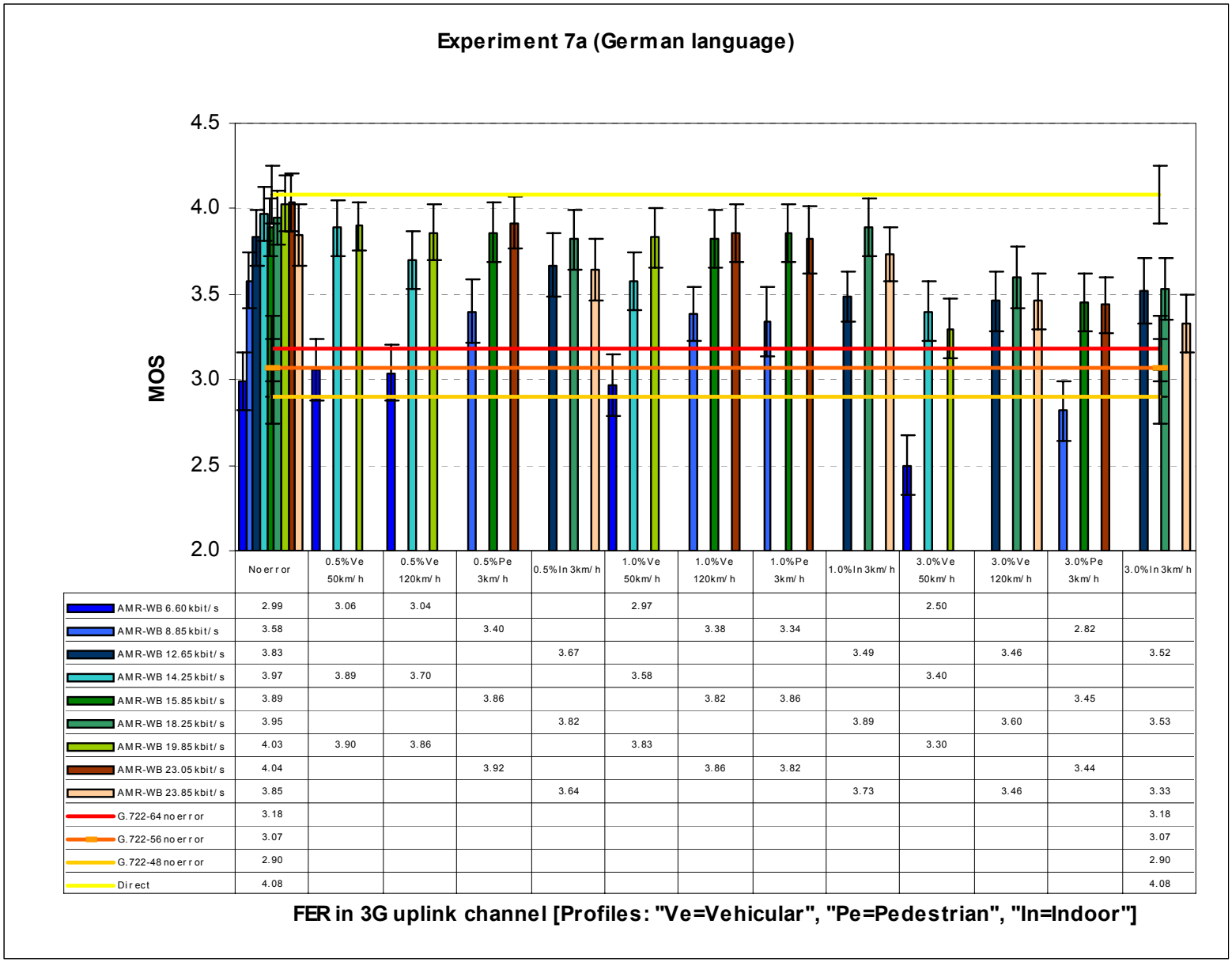
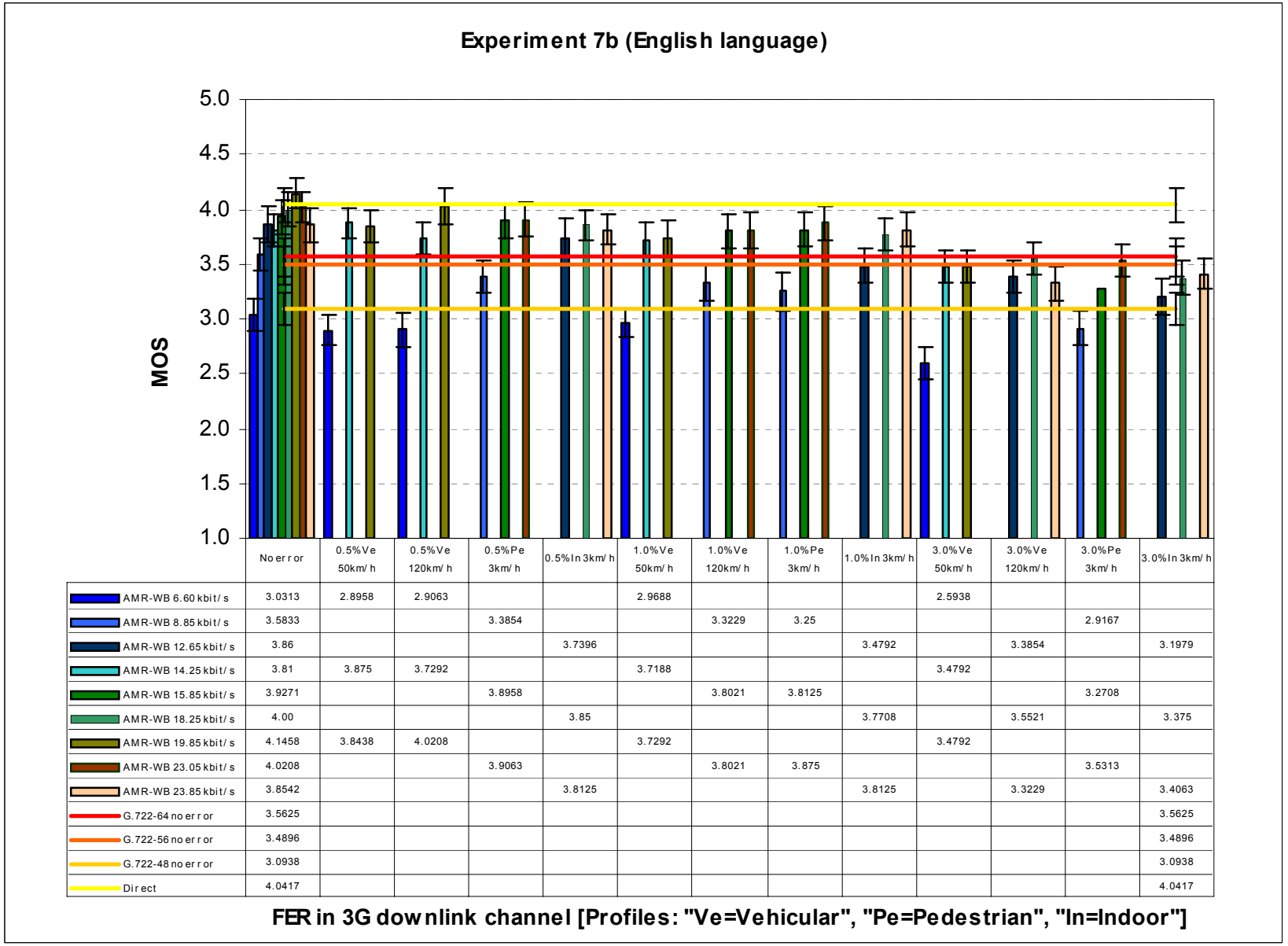


Figure 9 AMR-WB characterization test result from 3GPP with different input levels and self tandeming. English language. The bars indicate the 95% confidence intervals. From [29].



	No error	0.5%Ve 50km/h	0.5%Ve 120km/h	0.5%Pe 3km/h	0.5%In 3km/h	1.0%Ve 50km/h	1.0%Ve 120km/h	1.0%Pe 3km/h	1.0%In 3km/h	3.0%Ve 50km/h	3.0%Ve 120km/h	3.0%Pe 3km/h	3.0%In 3km/h
AMR-WB 6.60 kbit/s	2.99	3.06	3.04			2.97				2.50			
AMR-WB 8.85 kbit/s	3.58			3.40			3.38	3.34				2.82	
AMR-WB 12.65 kbit/s	3.83				3.67				3.49		3.46		3.52
AMR-WB 14.25 kbit/s	3.97	3.89	3.70			3.58				3.40			
AMR-WB 15.85 kbit/s	3.89			3.86			3.82	3.86				3.45	
AMR-WB 18.25 kbit/s	3.95				3.82				3.89		3.60		3.53
AMR-WB 19.85 kbit/s	4.03	3.90	3.86			3.83				3.30			
AMR-WB 23.05 kbit/s	4.04			3.92			3.86	3.82				3.44	
AMR-WB 23.85 kbit/s	3.85				3.64				3.73		3.46		3.33
G.722-64 no error	3.18												3.18
G.722-56 no error	3.07												3.07
G.722-48 no error	2.90												2.90
Direct	4.08												4.08

Figure 10 AMR-WB performance in erroneous 3G uplink channels. German language. From [29].



	No error	0.5%Ve 50km/h	0.5%Ve 120km/h	0.5%Pe 3km/h	0.5%In 3km/h	1.0%Ve 50km/h	1.0%Ve 120km/h	1.0%Pe 3km/h	1.0%In 3km/h	3.0%Ve 50km/h	3.0%Ve 120km/h	3.0%Pe 3km/h	3.0%In 3km/h
AMR-WB 6.60 kbit/s	3.0313	2.8958	2.9063			2.9688				2.5938			
AMR-WB 8.85 kbit/s	3.5833			3.3854			3.3229	3.25				2.9167	
AMR-WB 12.65 kbit/s	3.86				3.7396				3.4792		3.3854		3.1979
AMR-WB 14.25 kbit/s	3.81	3.875	3.7292			3.7188				3.4792			
AMR-WB 15.85 kbit/s	3.9271			3.8958			3.8021	3.8125				3.2708	
AMR-WB 18.25 kbit/s	4.00				3.85				3.7708		3.5521		3.375
AMR-WB 19.85 kbit/s	4.1458	3.8438	4.0208			3.7292				3.4792			
AMR-WB 23.05 kbit/s	4.0208			3.9063			3.8021	3.875				3.5313	
AMR-WB 23.85 kbit/s	3.8542				3.8125				3.8125		3.3229		3.4063
G.722-64 no error	3.5625												3.5625
G.722-56 no error	3.4896												3.4896
G.722-48 no error	3.0938												3.0938
Direct	4.0417												4.0417

Figure 11 AMR-WB performance in erroneous 3G downlink channels. English language. From [29].

The characterization of AMR-WB under error-conditions [29] considers several channels. GSM Gaussian Minimum Shift Keying (GMSK) and 3G uplink and downlink channels in different profiles were considered. We show the results for the 3G channels in Figure 10 and Figure 11. In these conditions the loss is measured in % frame error rate (FER). The figures show that the obtained score varies slightly depending on the channel profile indicating that the distribution and not only the frequency of the loss is a parameter. However, this variation falls within the 95% confidence intervals in most cases.

The degradation of the MOS score as a function of frame error rate in Figure 10 and Figure 11 is hard to compare to, e.g., the degradation of the MUSHRA scores in Figure 20 and Figure 21. The grading scale used is different (MOS and MUSHRA), the bandwidths of the signals are different, and it is not possible to compare the % frame error-rate to the % packet loss rate. The reason for the latter is that for packet switched systems techniques like interleafing and similar can be used while in the circuit switched systems unequal error protection of different bits in the stream is utilized. It should be noted that a frame error is experienced only when bits in the highest protection class can not be recovered.

5.2 Characterization of ITU-T G.729.1, G722.1 and G.722.1 C

The characterization test results of ITU-T G.729.1 [22] are summarized in Figure 12, Figure 13, and Figure 14 for narrowband clean speech, wideband clean speech and wideband music signals, respectively. These results are expressed in terms of MOS, with a 95% confidence interval (around 0.1 MOS). The performance of reference coders with the same bandwidth is also presented. In particular the figures show comparisons between ITU-T G.729.1, G.722.1, and AMR-WB. All coders considered here are implemented in fixed-point arithmetic.

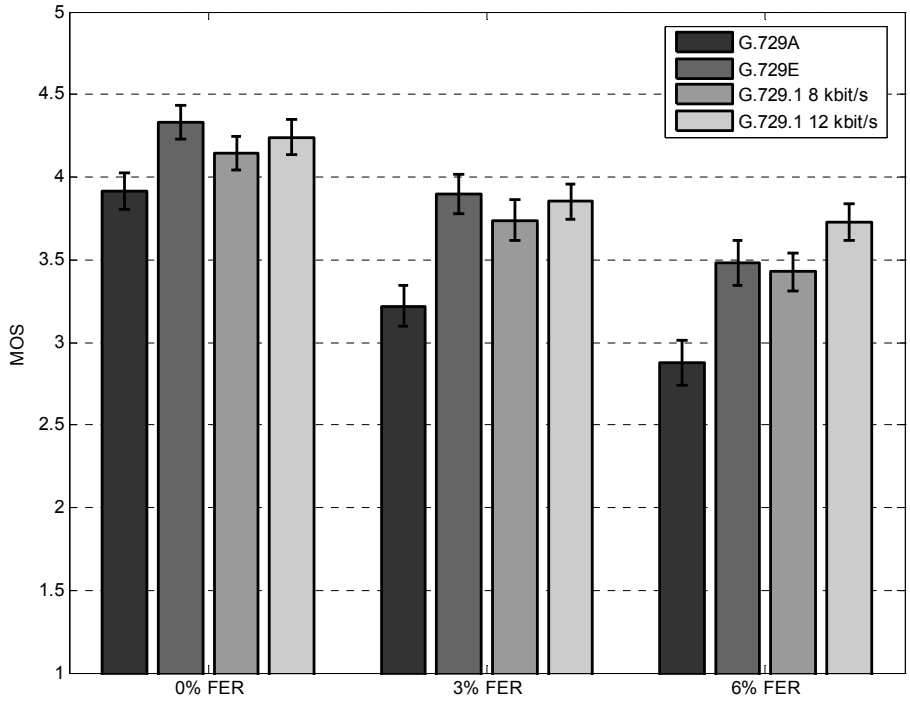


Figure 12 Performance of G.729.1 for narrowband clean speech signals (mono) and different frame error rates (0%, 3%, 6%) with G.729A and G.729E references (Source: France Telecom) – nominal level -26 dBov.

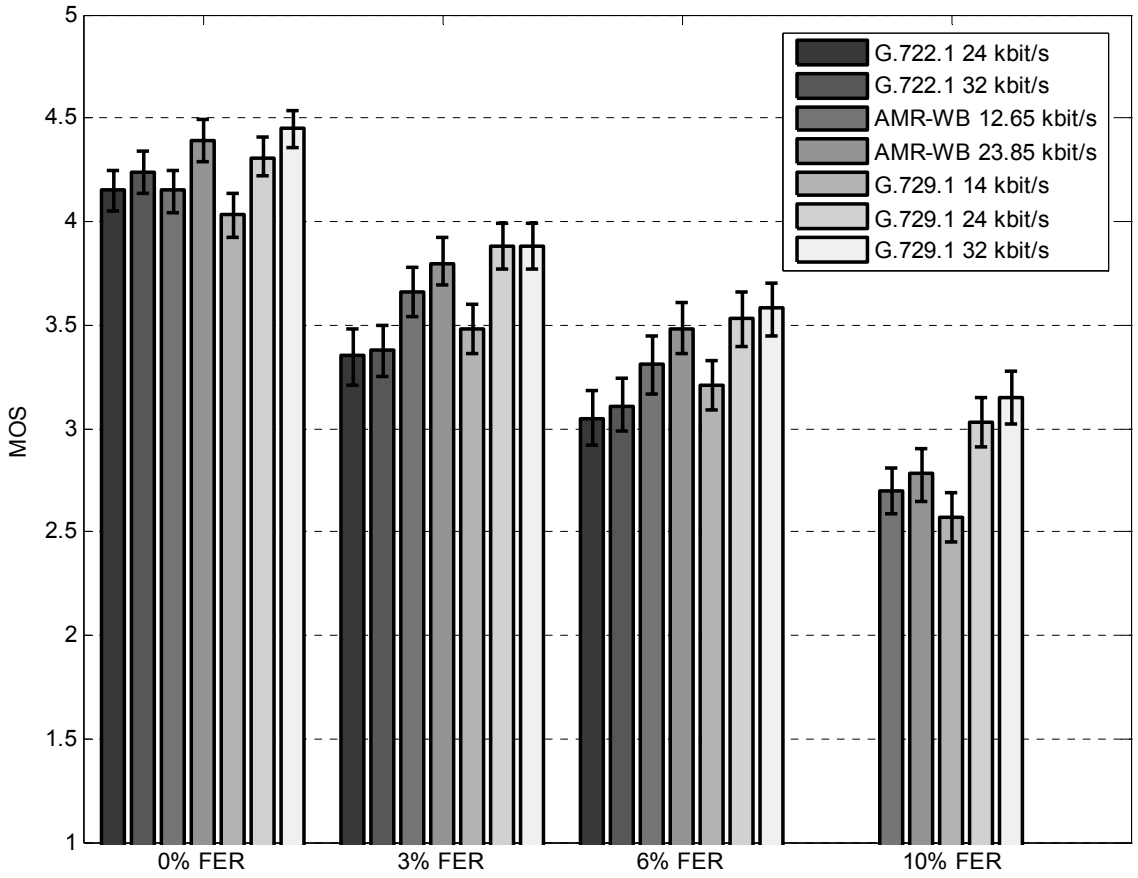


Figure 13 Performance of G.729.1 for wideband clean speech signals (mono) and different frame error rates (0%, 3%, 6%, 10%) with G.722.1 and AMR-WB references – G.722.1 not tested in 10% FER condition, nominal level -26 dBov (Source: Dynastat).

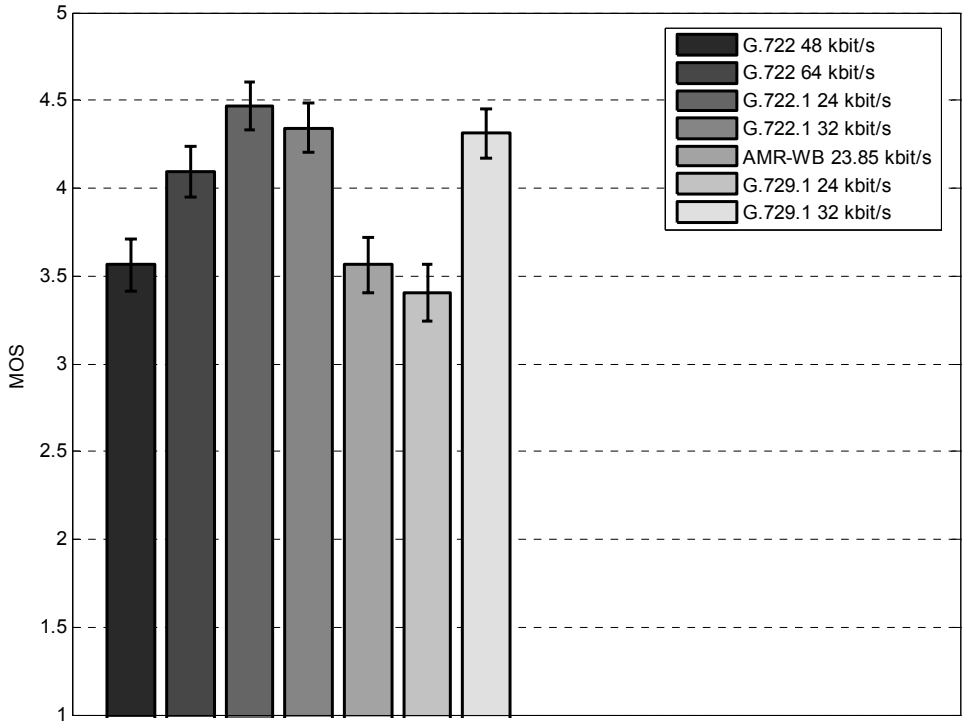


Figure 14 Performance of G.729.1 for wideband music signals (mono) and no frame error with G.722, G.722.1 and AMR-WB references - nominal level -26 dBov (Source: France Telecom).

Excerpts of the characterization test results of ITU-T G.722.1C are presented in Figure 15 and Figure 16, for superwideband clean speech and music signals, respectively. G.722.1 is compared with a coder of similar delay, MPEG4 AAC-LD, and with two non-conversational coders, 3GPP AMR-WB+ and e-AAC+.

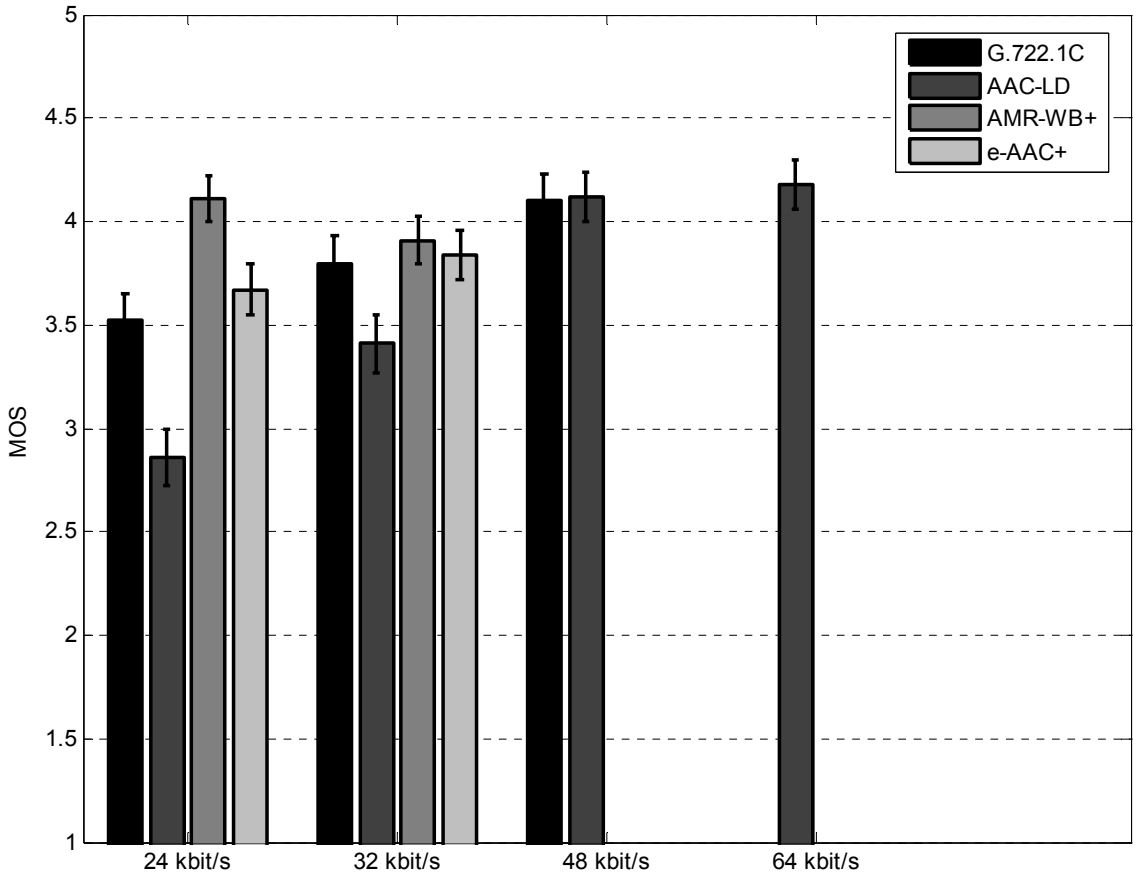


Figure 15 Performance of G.722.1C for wideband clean speech signals (mono) with AAC-LD, AMR-WB+ and e-AAC+ references, nominal level -26 dBov (Source: France Telecom)

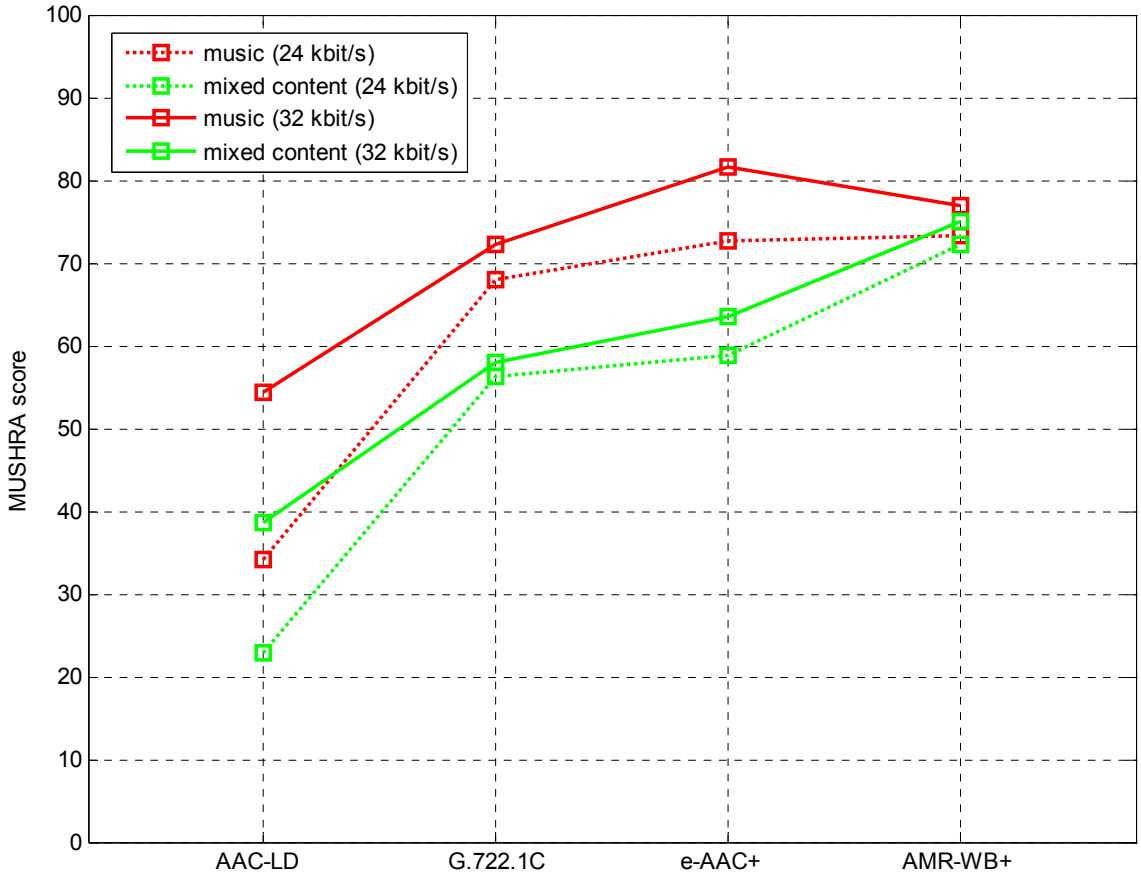


Figure 16 Performance of G.722.1C for wideband music and mixed content (mono) with AAC-LD, AMR-WB+ and e-AAC+ references (Source: France Telecom)

5.3 Characterization of 3GPP audio codecs (AMR-WB+, 3GPP e-AAC+)

The two 3GPP audio codecs' performance as tested in [28] is shown in Figure 17 for mono signals and lower rates (9.75 kbps to 20 kbps) and in Figure 18 for stereo signals and rates between 14.25 kbps and 28 kbps. We only summarize the characterization test results from [28] since the performance of the codec algorithms used throughout the selection tests (also included in [28]) appear to considerably deviate from the performance of the finally standardized codecs. In addition, informal tests with these two codecs in conjunction with the H.263 and H.264 video codecs have been performed within Ericsson AB. The results describe the quality of service using 3GPP Rel-6 audio and video codecs and are summarized in Figure 19.

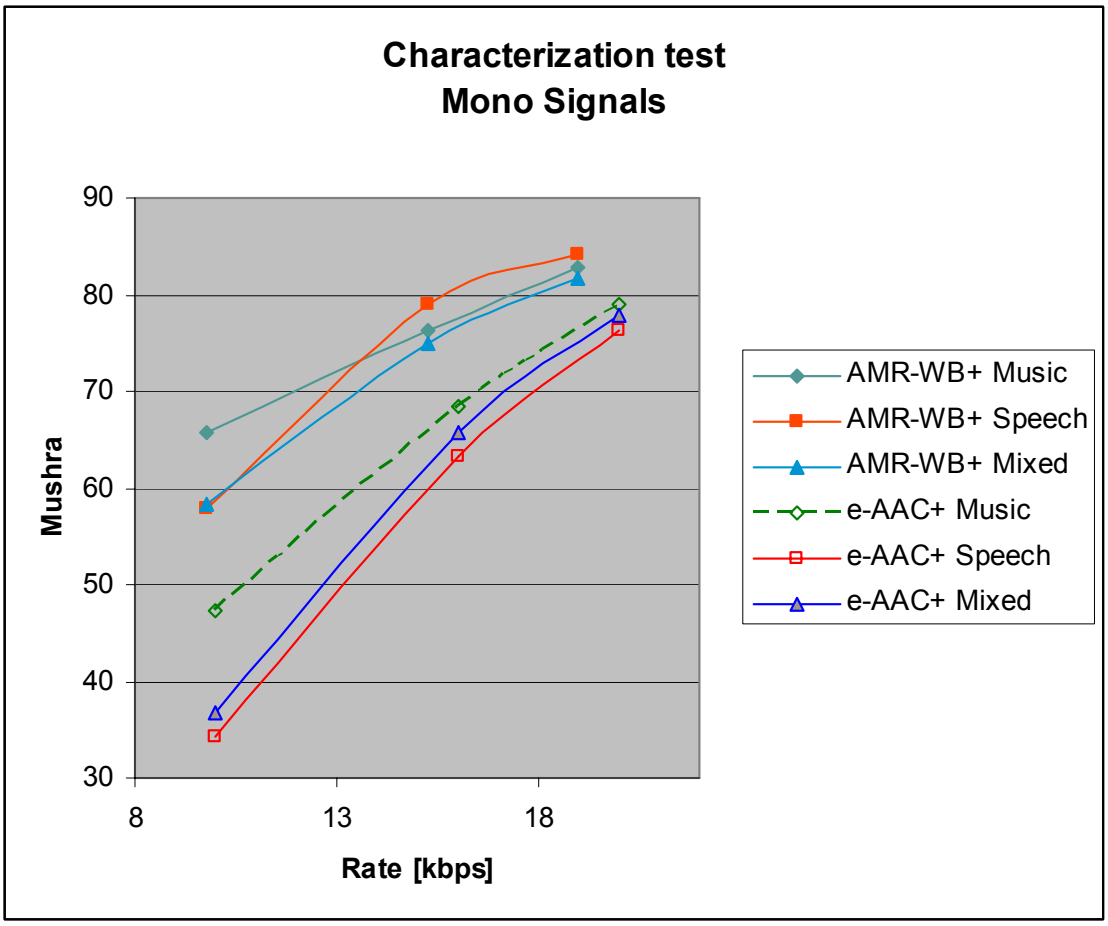


Figure 17 Performance of 3GPP audio codecs for mono signals from 3GPP characterization test.

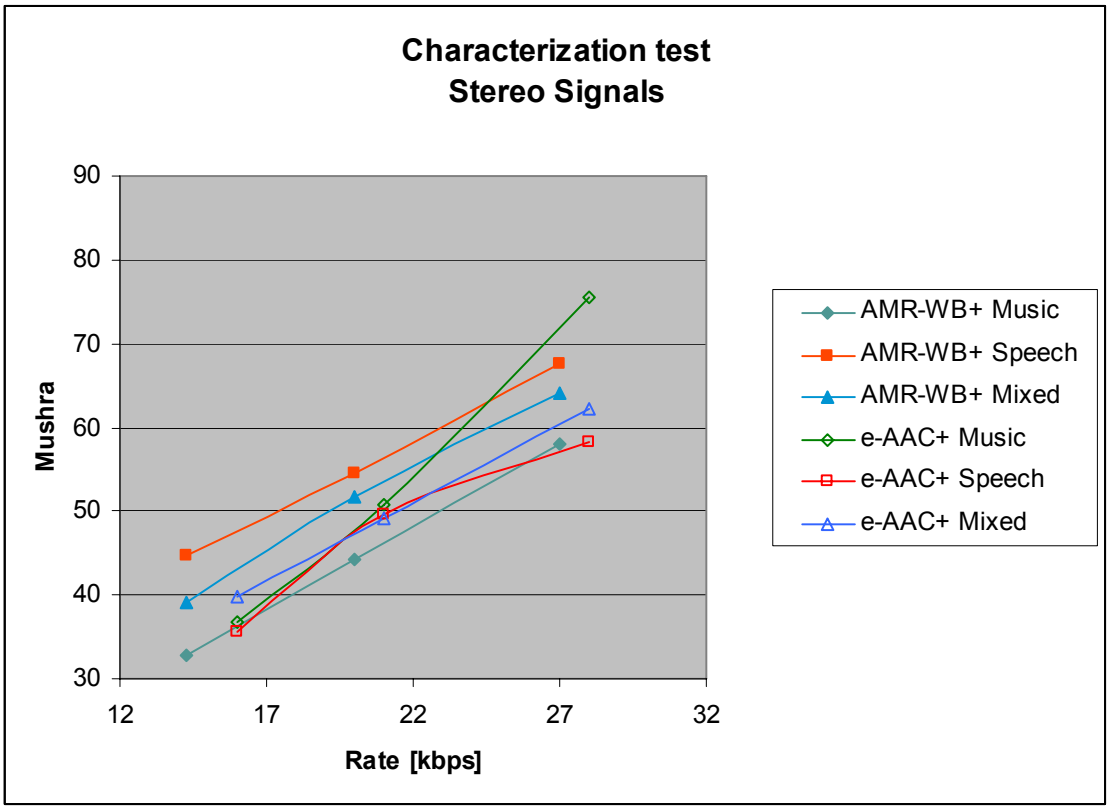


Figure 18 Performance of 3GPP audio codecs for stereo signals from 3GPP characterization test.

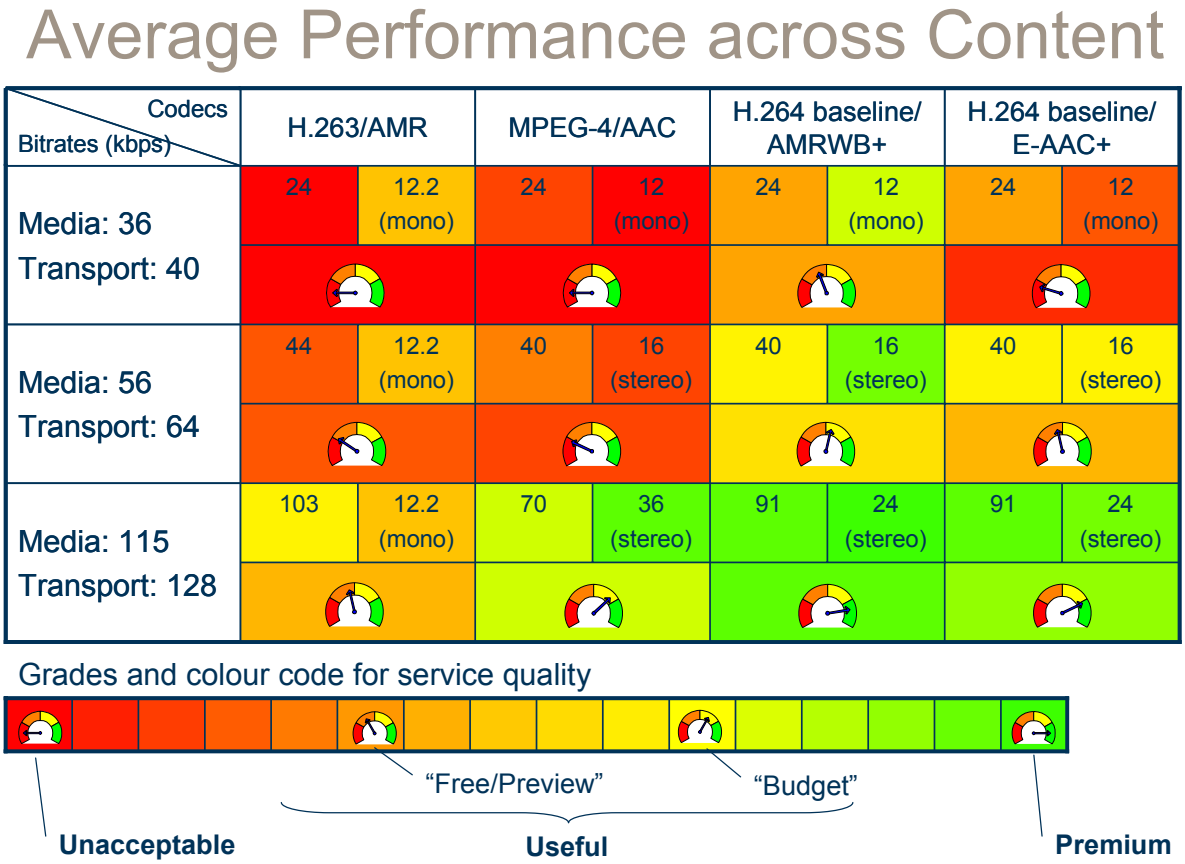


Figure 19 Results of informal tests performed at Ericsson AB for AMR-NB, AAC, AMR-WB+, and e-AAC+ audio codecs together with H.263 and H.264 video codecs. The upper left panels in each row show the video rate and service quality and the upper right panels show the audio rate and service quality. The lower wide panels in each row show the overall service quality for the given setup.

Below we summarize the material available from the characterization of e-AAC+ and AMR-WB+ [28] operating on erroneous channels. For the 3GPP audio codecs two wireless packet switched channel types were considered during the characterization tests, these are EGPRS (Enhanced General Packet Radio Service) and UTRAN (UMTS Terrestrial Radio Access Network). For mono signals only EGPRS simulations were performed, these are summarized in Figure 20. The stereo results for the EGPRS simulations are shown in Figure 21. These figures suggest that the MUSHRA points the two codecs obtain reduce by approximately a factor two for a packet loss rate of 5%.

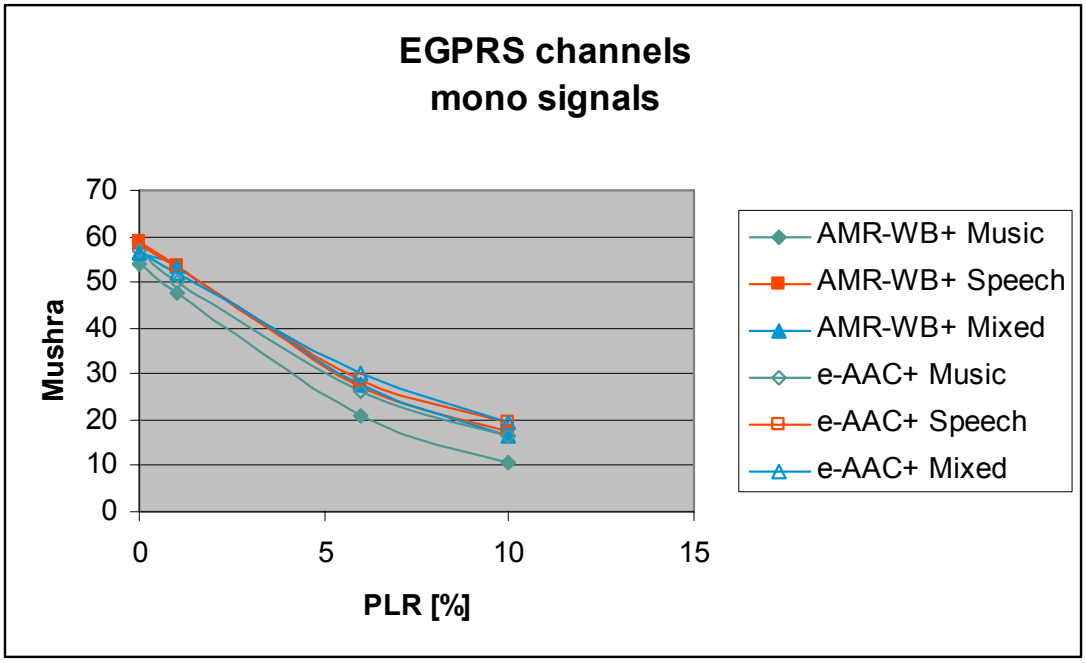


Figure 20 AMR-WB+ and e-AAC+ performance under packet loss conditions for mono signals as found in 3GPP characterization tests. AMR-WB+ is run at a gross rate (including packetization overhead) of 16 kbps and e-AAC+ at a gross rate of 20 kbps.

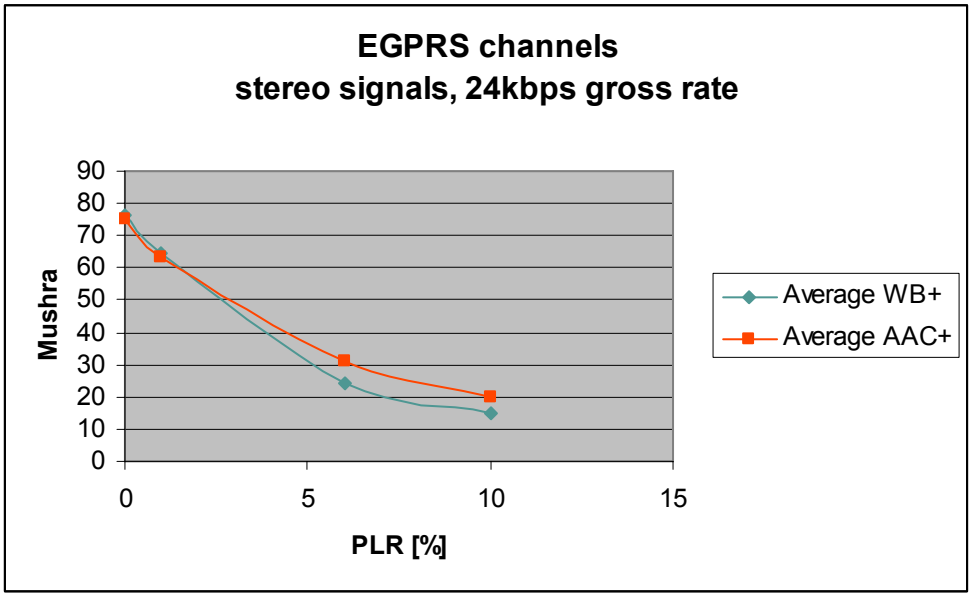


Figure 21 AMR-WB+ and e-AAC+ performance under packet loss conditions for stereo signals as found in 3GPP characterization tests. Both codecs are run at a gross rate (including packetization overhead) of 24 kbps. These results show the average for all signal types.

In our opinion the test results for UTRAN channels in [28] do not provide a good comparison between the two codecs under test. In one experiment the gross rate used is very different for the two codecs under test (20 kbps for AMR-WB+ and 32 kbps for e-AAC+) and in the other test the gross rate is equal for the two codecs but the interleafing strategy is very different. This is noted also in [28].

5.4 Characterization of MPEG codecs (AAC, BSAC)

MPEG test results derived from [25] are presented in Figure 22. These results use the ITU-R BS.1284 quality scale and include 95% confidence intervals. (Error resilient) BSAC is compared with AAC (main profile). Conclusions from [25] are that BSAC at 96 kbit/s is equivalent to AAC main at 96 kbit/s, that the performance of BSAC increases monotonically, and that BSAC at 64 kbit/s does not perform as well as AAC at 64 kbits (due to the overhead of bit-rate scalability).

While the 3GPP and ITU encoders are standardized, the MPEG audio encoders are not standardized since MPEG follows the philosophy that the decoder is standardized and every encoder that can serve the standardized decoder is conformant with the MPEG standard. Thus, the performance of different implementations of the MPEG audio encoders can differ.

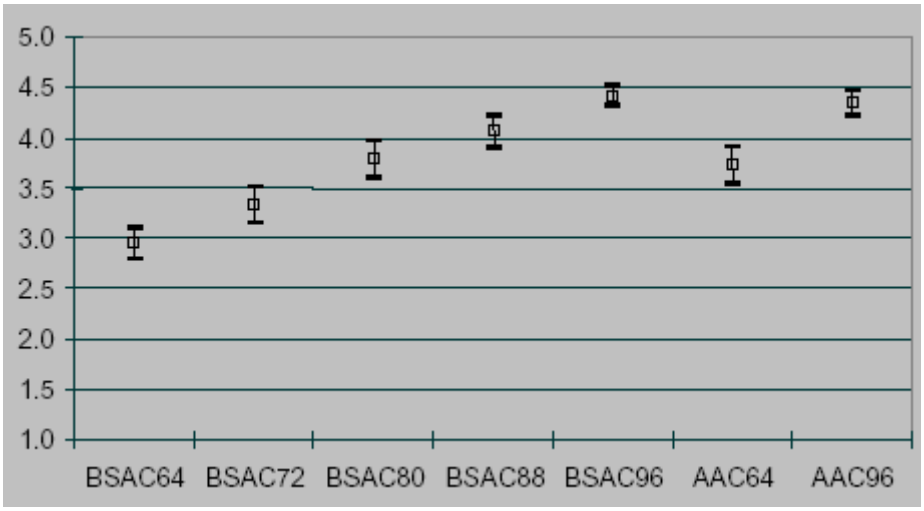


Figure 22 Results of MPEG tests evaluating BSAC and AAC for stereo signals with full bandwidth (48 kHz sampling rate).

6 Standardization Prospects and Links with other Projects

To highlight possible dissemination and put the FlexCode project into perspective, this section lists some potential standardization bodies, activities in these, and projects of the sixth framework programme (FP6) relevant for FlexCode. The standardization bodies treated are the 3rd generation partnership project (3GPP), the international telecommunication union (ITU), and the moving pictures expert group (MPEG).The FP6 projects mentioned are Enthrone, M-Pipe, ARDOR, and DANAE.

6.1 Third Generation Partnership Project (3GPP)

The 3rd Generation Partnership Project (3GPP) is a collaboration project among a number of telecommunications standards bodies, the “Organizational Partners” which are the European ETSI and other regional organizations such as ARIB, CCSA, ATIS, TTA, and TTC. The scope of 3GPP is to produce globally applicable Technical Specifications and Technical Reports for a 3rd Generation Mobile System based on evolved GSM core networks and the radio access technologies that they support (i.e., Universal Terrestrial Radio Access (UTRA) both Frequency Division Duplex (FDD) and Time Division Duplex (TDD) modes) as well as their future evolutions (UTRA-UTRAN Long Term Evolution (LTE) and 3GPP System Architecture Evolution (SAE)). Also within the scope is the maintenance and development of the Global System for Mobile communication (GSM) Technical Specifications and Technical Reports including evolved radio access technologies (e.g. General Packet Radio Service (GPRS) and Enhanced Data rates for GSM Evolution (EDGE)), which was transferred from ETSI.

The technical work within the 3GPP is carried out by various groups and subgroups. Codec responsible is the group SA4 (System Aspects 4 (Codec)). SA4 specifies the codecs for the various media services within 3GPP, their proper use and the related protocols enabling the services. Within this group codecs like AMR, AMR-WB, AMR-WB+ and e-AAC+ have been standardized. Most recently the group has specified the Multimedia Telephony Service for IMS (MTSI) which constitutes the first fully functional end-to-end media layer specification for real-time services over IP.

If Flexcode delivers technology suitable for mobile applications, SA4 will be the target group within which a 3GPP standardization might be possible. However, unlike other standardization bodies such as ITU or MPEG, 3GPP does not standardize a codec per se. Rather, firstly, a service must be specified imposing particular requirements, which cannot be met by existing codecs. Then, in turn, a new codec standardization effort may be initiated, usually in form of a work item.

6.2 International Telecommunication Union (ITU)

The ITU has a tradition of standardizing speech codecs in their ITU-T G. series of standards. Currently the ITU telecommunications standardization sector (ITU-T) is studying new codecs in study group 16 working party 3 (ITU-T SG16/WP3). A possible launch of a new WP3 codec question targeting a new speech and audio codec is still in its infancy, with discussions in WP3/Question 23 about the need and possible terms of reference and requirements for a new speech and audio coder. The potential outcome of this work is sometimes referred to as a multi-media convergence coder (ITU-T G.MMCC) where the convergence may refer to a converged speech and audio coder. The primary target of the envisioned codec are conversational services over Next-Generation Networks (NGN) [15] and at this point it is not clear whether mobile devices and streaming applications will be outside the scope of the new codec. A request for further input to SG16/23 is found in [16]. In addition, ITU handles maintenance and extensions of existing standards in WP3, question 10 and standardization of a new scalable speech coder with additional requirements on audio coding capabilities is ongoing in WP3, question 9 VBR-EV (variable bit-rate embedded variable rate).

The status and nature of this standardization effort make it a potential dissemination platform for FlexCode. The envisioned usage of the codec in heterogeneous networks calls for a flexible coder where the self configuring concept of FlexCode fits very well. However, at this point it is not known if the exploration work in the direction of a new combined speech and audio coder will be further pursued. All three industrial partners in FlexCode actively participate in ITU-T SG 16. Thus, the project has direct access to this standardization work and can react directly if potential for dissemination becomes evident.

6.3 Moving Pictures Experts Group (MPEG)

The MPEG is a working group of the International Standardization Organisation and International Electrotechnical Commission (ISO/IEC) partnership. Among all MPEG activities the one most likely to be relevant for FlexCode is the exploration work on scalable speech and audio coding. A milestone was the call for information document [17] in January 2005. The interest for this exploration has been varying and the responses to [17] did not lead to a call for proposals. Currently MPEG is re-iterating the process and defined a new workplan for the exploration of speech and audio coding [18] that targets low bit-rates (≤24kbps per channel). This workplan is meant to flourish the work on requirements and targeted signals for a possible new speech and audio coder within MPEG. The initial focus on scalability of the resulting coder is degrading while coding efficiency appears to be of increasing importance to the experts within MPEG.

The MPEG-21 standard provides a framework that is interesting for the FlexCode project. In particular the MPEG-21 digital item adaptation (DIA) is interesting for FlexCode. Utilizing MPEG-21 DIA a standardized way of adapting content (referred to as digital items within MPEG-21) to a variety of devices and networks is possible. Thanks to the open character of the MPEG-21 DIA standardization, the algorithms envisioned in FlexCode can serve as a valuable contribution to the DIA-Engine being at the heart of MPEG-21 DIA.

MPEG standardization meetings are attended by all three industrial partners of the FlexCode project. Thus, the possibility to monitor and take influence on the standardization is similar to the situation for ITU-T SG16.

6.4 Enthrone 6th Framework Programme Project

The Enthrone project [19] studies how to successfully provide audio-visual services to mass market, focusing on media distribution across heterogeneous networks and reception at various user terminals. The goal of this project is to propose an integrated management of resources to support end-to-end quality of service (QoS) over heterogeneous networks and terminals, with content from diverse sources and originating in different formats. In practice, Enthrone relies on the MPEG-21 standard and especially the concept of universal multimedia access (UMA[1]). The QoS of compressed media streams is controlled and adapted during its transmission based on MPEG-21 Digital Item Adaptation (DIA). The architecture used in Enthrone is summarized in Figure 23 below.

[1] Please note that the acronym UMA is used for unlicensed mobile access in other sections of this document.

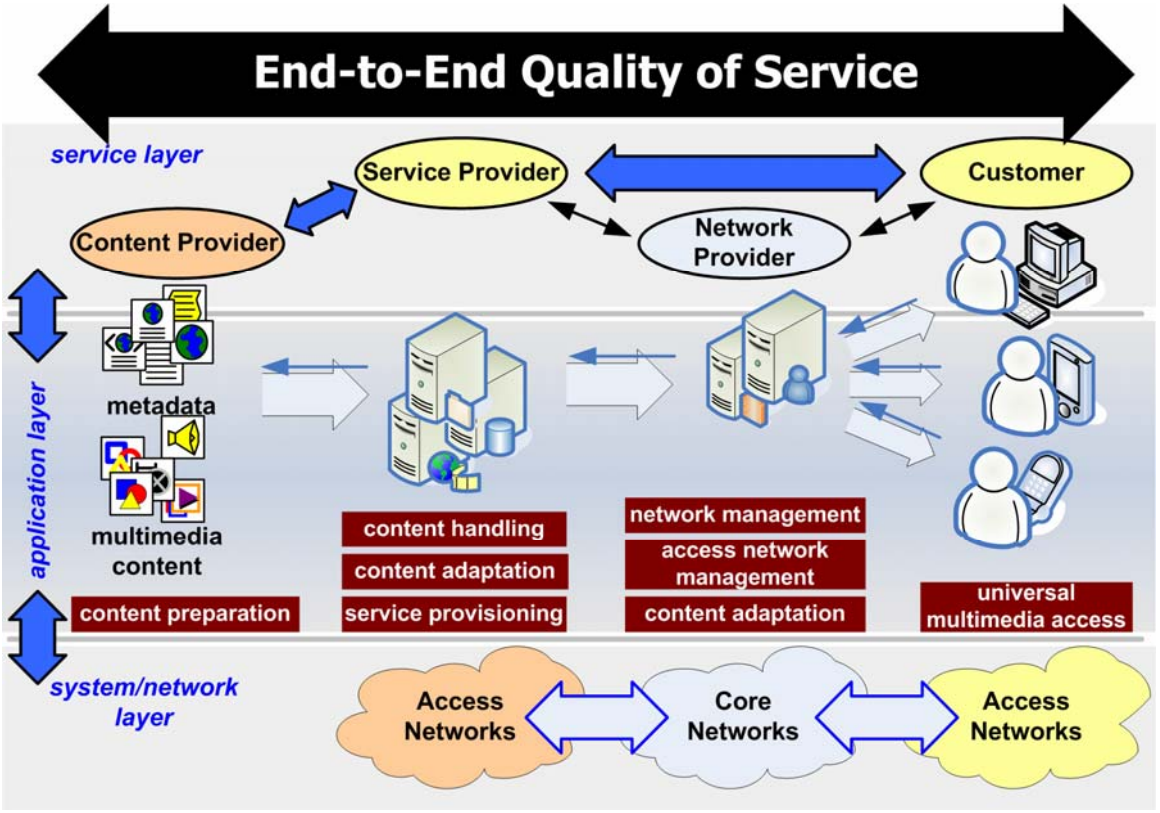


Figure 23 Overview of the Enthrone architecture.

As shown in Figure 23, the Enthrone application scenario corresponds to a subset of the scenarios considered within Flexcode. Enthrone is related to the Multimedia On-Demand Streaming, Multimedia Multicast Streaming, and Multimedia Download scenarios described in Section 2 of this document.

One aspect of Enthrone is content handling, which boils down to two questions:

- How to represent content in a QoS-sensitive way – this is done using *scalable coding* (which is also referred to as embedded or hierarchical coding).
- How to implement and use scalability - part of this consists in correlating encoding characteristics, network capabilities and terminal heterogeneity for end-to-end QoS provisioning.

This approach of content handling in Enthrone provides a generic adaptation to QoS requirements. Enthrone considers the MPEG-4 SVC codec [20] for video coding . The choice of the audio codec is still open.

Compared to Enthrone, FlexCode is dealing mainly with content compression and representation. It should be noted that Enthrone may benefit from the source coding developments of FlexCode, since the audio codec to be used is not yet defined. This possible collaboration would require to liaise between projects, define audio coding requirements within ENTHRONE and align the FlexCode and Enthrone time schedules accordingly.

6.5 ISIS and DANAE 6th Framework Programme Projects

The ISIS and DANAE project deal with the adaptation, delivery and search of multimedia content.

The Intelligent Scalability for Interoperable Services (ISIS) project [21] – which ended in 2004 - studied service scenarios of multimedia content search and delivery (video, audio, 2D/3D graphics). The main focus was on helping users during their interaction with complex services (e.g. search in a large database of available content), especially on mobile terminals, by means of personalization and customization agents (client or server). One important aspect dealt with context adaptation to tailor the quality and type of service provided (using MPEG-21 DIA and context detection based on user's past actions). Besides, this project considered scalable content representation formats and reused MPEG-4 BSAC for audio coding. The underlying service architecture is summarized in Figure 24 below.

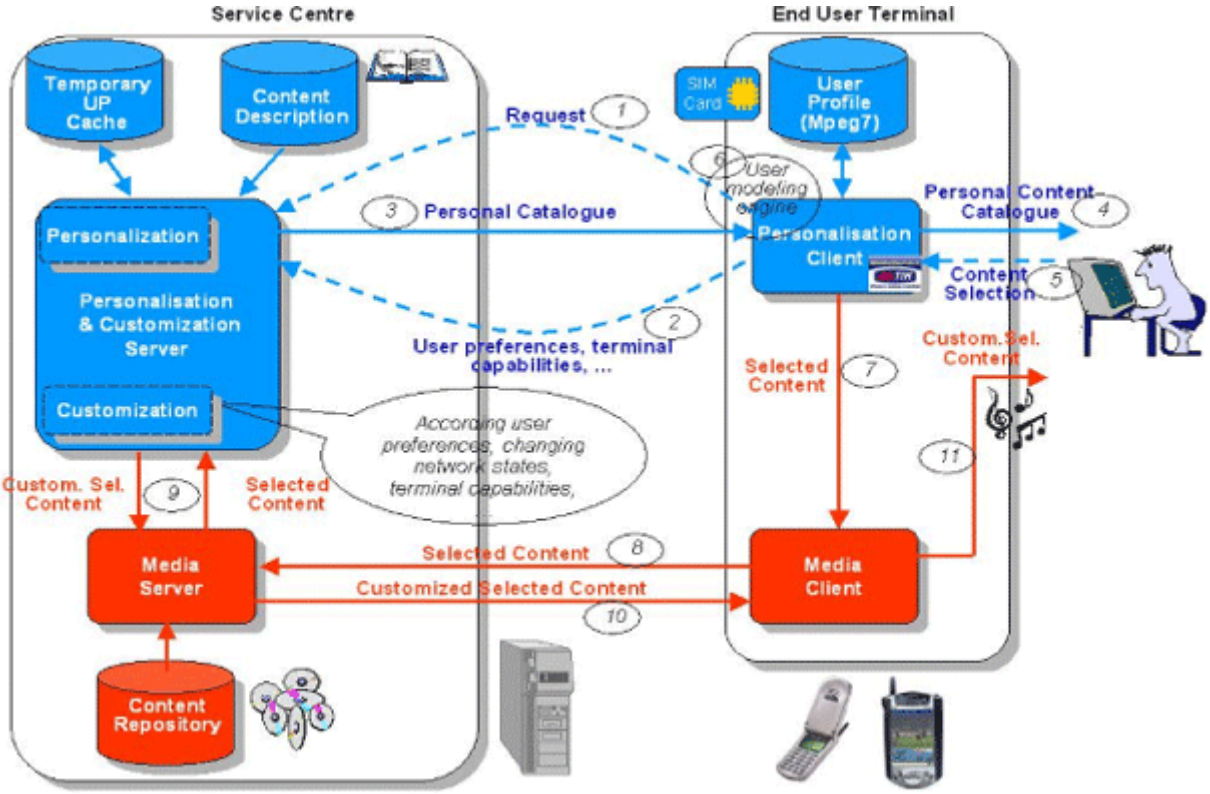


Figure 24 Summary of ISIS architecture.

The Dynamic and distributed Adaptation of scalable multimedia content on a context-Aware Environment (DANAE) project [26] – which ended in 2006 - pursued research started in the ISIS project. The DANAE project focused on two technical parts:

- Advanced MPEG-21 architecture (client, server, proxy) with end-to-end QoS support, personalization, Digital Item Adaptation (DIA), and Digital Rights Managements (DRM).

- MPEG scalable codecs with emphasis on video coding - MPEG-4 BSAC was used for audio coding, and many media types were considered (audio, video, 2D graphics, 2D/3D virtual characters).

The application scenario considered in DANAE is the streaming of rich media content. It is summarized in Figure 25 below.

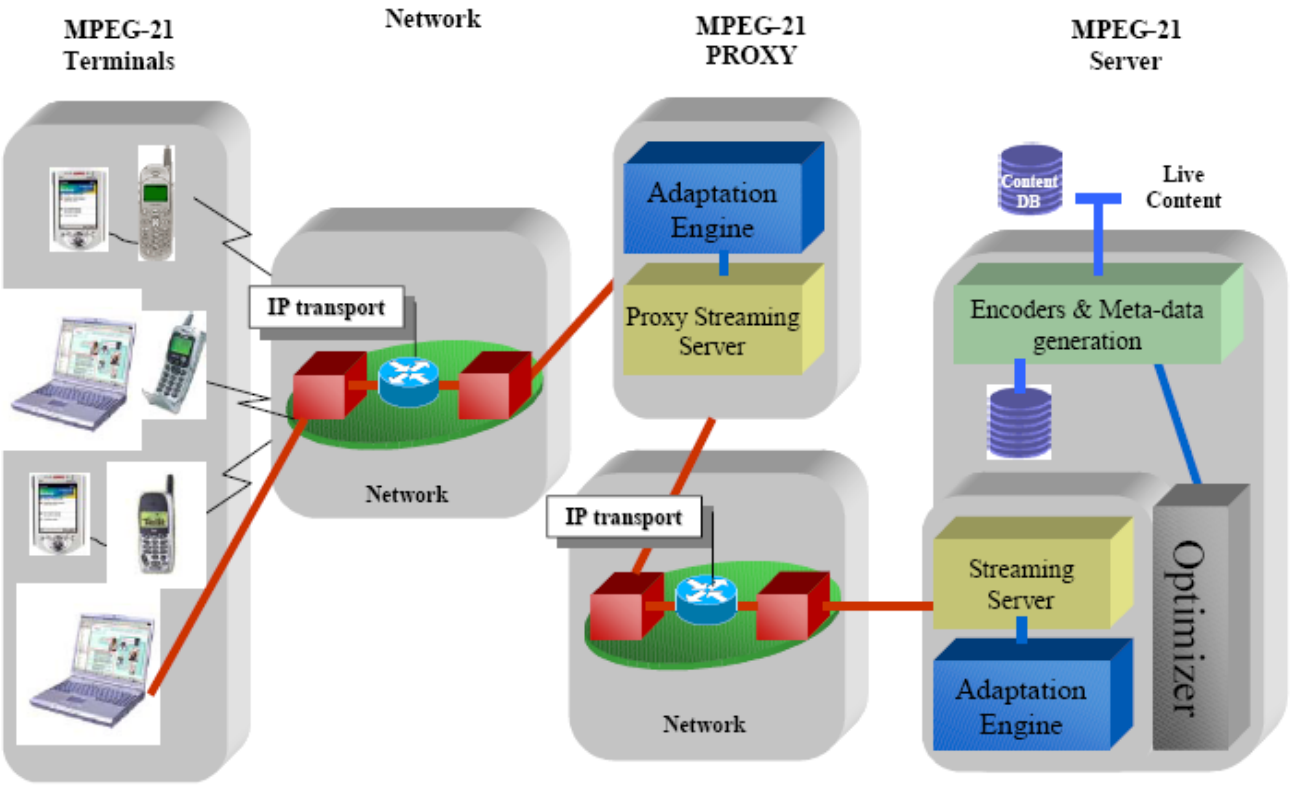


Figure 25 Summary of DANAE architecture.

Some of the FlexCode scenarios fit well in the scenario work already defined in the ISIS and DANAE project. The FlexCode codec can be integrated in the architecture of the ISIS and DANAE projects as the media adaptation is based on MPEG-21 in both projects.

6.6 ARDOR 6th Framework Programme Project

The Adaptive Rate-Distortion Optimised sound codeR (ARDOR) project focused on universal audio coding. The main objective was to develop a codec that can be used for many different applications such as telephony, Internet radio, or solid-state audio players. Hence, a single codec can be used in a variety of scenarios instead of a large collection of codecs. The ARDOR coding approach is summarized in Figure 26 below. Several coding techniques are combined (namely, sinusoidal coding, transform coding, CELP coding) and selected based on a rate-distortion optimization.

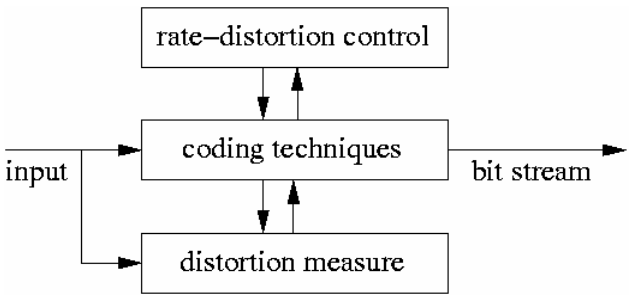


Figure 26 High-level description of the ARDOR codec.

This ARDOR codec was developed to adapt to the time-varying characteristics of the input signal, to user preferences, and to application-imposed constraints or time-varying network-imposed constraints on coding attributes such as bit rate, and quality. There is indeed a clear connection between ARDOR and Flexcode with respect to source coding objectives. It is worth noting that two ARDOR contributors, KTH and France Telecom, participate in Flexcode.

While some of the tools used in ARDOR are expected to be found in the FlexCode codec as well, the FlexCode approach goes further. FlexCode includes for example channel coding and high-rate rate-distortion theory is used in conjunction with signal models (e.g., Gaussian mixture models) to allow for flexible quantization and large codebooks optimized online without the need to store codebook tables.

6.7 M-Pipe 6th Framework Programme Project

The M-Pipe project focuses on the utilization of cross-layer communication for media delivery in scenarios with heterogeneous networks and terminal capabilities. The basic idea is to provide a simple signalling of media-stream properties to all network layers such that local adaptation and optimization of the stream can be performed in individual network nodes. For this purpose the layer independent descriptor (LID) was introduced. The LID describes the properties of the media data contained in a data packet and guides the network nodes on how to handle the data in terms of, e.g., packet drops, truncation, error protection. The LID is agnostic to the media type and codec used and included in the packet header represented by a LID label to reduce overhead.

Within M-Pipe research on all network layers has been performed starting from modulation schemes, unequal error protection and turbo codes over router queue design to scalable source coders for audio and video. All research was performed with the local adaptation and optimization guided by the LID in mind. While the LID enhanced media-streams can be transported via legacy networks the full advantage of the M-Pipe approach can only be achieved when LID enabled network nodes are present. Within the FlexCode project concepts that do not require certain capabilities of the network are envisioned. In FlexCode adaptation of source and channel coding is governed by session setup and feedback information at the sending end and local adaptation at the receiving end.

There is potential for the FlexCode project to consider the findings on scalable source coding and flexible channel coding within the M-Pipe project. Specifically the scalable speech and audio coder (SSAC) and the work on turbo codes and LDPC within M-Pipe can give valuable input to the FlexCode project. However, since the system developed in FlexCode does not assume additional functionality of the network nodes the resulting solution will differ from the M-Pipe solution. The exchange with the M-Pipe project is facilitated by the fact that Ericsson AB is a partner in both the M-Pipe and the FlexCode project. However, to which extent data can be exchanged by the projects is governed by the respective Consortium Agreements as well as decisions made by the projects.

7 Summary and Conclusions

This document contains a list of service scenarios that can benefit from the FlexCode paradigm. For each scenario the end-user perspective, equipment and network requirements, as well as coding requirements (bitrate, coding fidelity, delay, and computational complexity) are presented. In addition, the scenario descriptions contain a list of the most suitable existing coders and discussion on the potential benefit of FlexCode for the particular scenario. The final ranking of the scenarios is given in Table 6.

As noted in [27] the performance of the codec designed in the FlexCode project is supposed to be similar to the performance of state-of-the-art codecs. Thus, this document also summarizes the performance of the existing codecs found relevant for the different scenarios in Section 5. The figures of merit for the different codecs were gathered mainly from the characterization and verification tests performed by the respective standardization bodies. In addition, the FlexCode partners provided internal material. The codecs' performances are characterized by means of large scale listening tests, under various conditions (error conditions, type of input signal, bitrate and delay constraints, etc.). Such a characterization provides an important performance benchmark for the FlexCode codec.

To put the listed scenarios in perspective and to describe possible dissemination and sources of cooperation Section 6 lists some relevant standardization bodies and other FP6 projects. The standardization bodies mentioned are ITU, 3GPP, and MPEG. Further, the FP6 projects Enthrone, M-Pipe, ARDOR, and DANAE and their relation to FlexCode are described. The discussion is focused on potential overlap and collaboration that can be mutually beneficial. We found that a number of current and recent standardization activities are interesting for FlexCode. Some of them since results of the project could be disseminated to these activities, e.g., the ITU-T G.MMCC activity or the MPEG exploration work on speech and audio coding. MPEG-21 DIA is another potential target for FlexCode technology.

Other standardization activities are interesting since they specify the environment for a number of the scenarios listed in this document. The 3GPP and ETSI activities on multimedia telephony service for IMS (MTSI) are important for, e.g., the mobile conversation scenario and the Internet conversation scenario. The ITU activities on next generation networks (NGN) provide other frameworks for many of the listed scenarios. One architecture falling into the NGN scope is the IP multimedia subsystem (IMS) which is a potential framework for almost all scenarios mentioned in this document.

Most scenarios in this document are operating on packet based networks, many of them on the general Internet. In addition, wireless networks, such as WiFi and 3G networks, are considered. These networks are error-prone and channel coding is indispensable in particular when the general Internet and wireless connection of the end user equipment are involved. Channel coding has traditionally been considered as part of the lower layers in packet switched networks, with the packet loss notification being the only information transferred to the application layer. With the horizon of the UDP-lite protocol, standardized by the IETF, the previous consideration has to be revised. Partially damaged payloads can be delivered to the application which opens for many traditional channel coding tools like, e.g., unequal error protection to the packet switched world.

In most of the described scenarios a clear advantage of a flexible coder is identified, this can be due to different user equipment, different connections, varying network conditions, and so on. Thus, in many cases a data stream individual to each user is desirable. However, it is also observed that individual encoding towards each user of a service is either impossible or leads to unreasonable computational load at the content source for a number of scenarios. For, e.g., the mobile blogging scenario not all users are known and connected when the content is sent from the mobile terminal to the blogg server. Thus, individual encoding at the content source is not possible and adaptation in the blogging server is necessary. Another example are the multimedia streaming and download scenarios in sections 2.5, 2.6, and 2.7 and where the fact that several thousand clients can be expected to be connected to one content server leads to very high computational loads. Also for the multimedia conference scenario encoding for each user is difficult since this would mean that the sender has to run a high number of encoders in parallel. Thus, it is important to make the adaptation to individual user's equipment and channel conditions an extremely lightweight process. Examples for such lightweight techniques in source

coding are embedded coding, multiple description coding, or the provisioning of hierarchical streams among which a connected terminal can choose.

In Section 4 we provide a ranking of the scenarios in several dimensions, including economical relevancy for operators and manufacturers, feasibility, end-user aspects, and the advantage FlexCode is expected to inject into the scenario. Based on these criteria we identify the two most relevant scenarios. These are found to be the mobile conversation scenario and the multimedia on demand streaming scenario. An overview of the different aspects and resulting codec requirements is given in Table 3 and Table 4. The full ranking order of the scenarios is found in Table 6.

Utilizing Table 3 and Table 4 we note in Section 4.2 that with the flexibility required to cover the mobile conversation scenario, we partly cover the Internet conversation scenario and the multimedia conference scenario. For the streaming and download scenarios there is overlap as well, e.g., the signal types and equipment targeted overlaps. The multimedia multicast streaming scenario is the second most interesting out of this group of scenarios. The multicast architecture differs from the unicast architecture of the multimedia on demand streaming scenario and makes the multicast scenario more difficult to implement.

At the later stage of the project the selected two most relevant scenarios will be used as a base for a real-time implementation demonstration. In this document an attempt to balance between the broad scope of services (including all eventualities that can occur, including gateways and interoperability with a large amount of legacy equipment) and particular session setups of services (where one particular connection is described) is made. One of the reasons was to allow for a real-time demonstration that is not too remote from the scenario described. Still, the exact implementation of the real-time demonstration might be adapted to results found later in the project and to implementation practicalities.

The listed scenarios and the derived properties show that a codec with the flexibility as the one to be developed in the FlexCode project can not only cover a number of scenarios with a single codec. Such a codec also offers a vast number of advantages to one single scenario, such as seamless adaptation to channel setup and conditions, support of a variety of content, utilization of feedback information, and adaptation to rendering devices.

References

[1] ISO/IETC JTC1/SC 29/WG 11/ N5231, "*MPEG-21 Overview v.5*", October 2002, http://www.chiariglione.org/mpeg/standards/mpeg-21/mpeg-21.htm

[2] G. Camarillo, M. A. Garcia-Martin, "*The 3G IP Multimedia Subsystem, Merging the Internet and the cellular worlds*", Wileys 2004

[3] ISO/IEC 21000-7:2004, "*Information Technology- Multimedia Framework – Part 7: Digital Item Adaptation"*

[4] 3GPP TS 22.173, "*Multimedia Telephony Service and supplementary services; Stage 1*"

[5] 3GPP TS 24.173, "*IMS Multimedia telephone service and supplementary services; Stage 3"*

[6] ITU-T Recommendation G.114: "*One-way transmission time*" , May 2003

[7] J. Breebaart, J. Herre, C. Faller, J. Rödén, F. Myburg, S. Disch, H. Purnhagen, G. Hotho, M. Neusinger, K. Kjörling, W. Oomen, *"MPEG Spatial Audio Coding / MPEG Surround: Overview and Current Status",* AES 119th convention, October 2005

[8] ISO/IEC JTC 1/SC 29/WG 11/N7138, "*Report on MPEG Spatial Audio Coding RM0 Listening Tests*"

[9] 3GPP TS 26.290 V6.3.0, "*Audio codec processing functions; Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions*", June 2005

[10] 3GPP TS 26.401 V7.0.0, "*General audio codec audio processing functions; Enhanced aacPlus general audio codec; General description*", June 2006

[11] 3GPP TS 26.936 V6.1.0, "*Performance characterization of 3GPP audio codecs*", March 2006

[12] ISO/IEC13818-7, "*Information technology -- Generic coding of moving pictures and associated audio information -- Part 7: Advanced Audio Coding (AAC)*", January 2006

[13] ISO/IEC 14496-3, 3rd edition, "*Information technology; Coding of audio-visual objects; Part 3: Audio*", HE-AAC profile, July 2005

[14] 3GPP TS 26.171, "Speech codec speech processing functions; Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description", October 2006

[15] ITU-T Y.2001 "*General overview of NGN*", December 2004

[16] ITU-T TD214 WP3/16, "*DRAFT LS to SG11, SG12 and SG13 on next generation speech and audio coding*"

[17] ISO/IEC JTC 1/SC 29/WG 11N7040, "*Call for Information on Scalable Speech and Audio Coding*"

[18] ISO/IEC JTC1/SC29/WG11, MPEG2006/N8331, "*Workplan for Exploration of Speech and Audio Coding*"

[19] Enthrone web site (phase 1): www.**enthrone**.org/

[20] Draft ISO/IEC 14496-10, "*Joint Draft 9 of SVC Amendment*", January 2007, http://ftp3.itu.ch/av-arch/jvt-site/2007_01_Marrakech/JVT-V201.zip

[21] ISIS web site: http://isis.rd.francetelecom.com/

[22] ITU-T SG16 Temporary Document, "*Executive Summary of G729.1 Characterisation step 2– Experiments 1, 2 & 3*", TD-258-GEN/16 Attachment 2, Nov. 2006 (Source: Q7/12 Rapporteurs)

[23] ITU-T SG16 Temporary Document, "*Characterisation test results of the 14khz low-complexity audio coding algorithm at 24, 32, and 48 kbps extension to ITU-T G.722.1: phase 1*", TD WP3/16 AC-05-29, Strasbourg, April 2005 (Source: Q10/16 Rapporteur)

[24] ITU-T SG16 Temporary Document, "*Characterisation test results of the 14khz low-complexity audio coding algorithm at 24, 32, and 48 kbps extension to ITU-T G.722.1: phase 2*", TD WP3/16 AC-05-30, Strasbourg, April 2005 (Source: Q10/16 Rapporteur)

[25] ISO/IEC JTC1/SC29/WG11, "*Report on the MPEG-4 Version 2 Audio Verification Test*", N3075, Dec. 1999

[26] DANAE web site: http://danae.rd.francetelecom.com/

[27] Sixth Framework Programme, Project FP6-2002-IST-C 020023-2 "*Annex I: Description of Work*"

[28] 3GPP TR 26.936V6.1.0 "*Performance characterization of 3GPP audio codecs*", March 2006

[29] 3GPP TR26.976V6.0.0 "*Performance characterization of the Adaptive MultiRate Wideband (AMR-WB) speech codec*", December 2004

[30] ISO/IEC/ IEC JTC 1/SC 29/WG 11N3075 "*Report on the MPEG-4 Audio Version 2 Verification Test*", December 1999

[31] ITU-R BS.1534-1 "*Method for the subjective assessment of intermediate quality level of coding systems*", January 2003

[32] ITU-R BS.1284-1 "*General methods for the subjective assessment of sound quality*" December 2003

[33] J. C. Bolot, "*Characterizing End-to-End Packet Delay and Loss in the Internet*", J. High-Speed Networks, vol. **2**(3), p. 305-323, Dec. 1993

[34] R. Cuny and A. Lakaniemi, "*VoiP in 3G networks: an end-to-end quality of service analysis*", Proc. IEEE Vehicular Techn. Conf., p. 930 – 934, April 2003

[35] IEPM (Internet end-to-end performance monitoring), SLAC (Stanford linear accelerator) homepage: http://www-iepm.slac.stanford.edu/monitoring/

[36] M-Pipe demonstrator: http://www.ist-mpipe.org/demonstrator/demonstrator.html

[37] 3GPP TS 26.114, "*IP Multimedia Subsystem (IMS); Multimedia Telephony; Media handling and interaction (Release 7)*"

[38] A. Vetro and C. Timmerer *"Digital Item Adaptation: Overview of Standardization and Research Activities"*, IEEE Transactions on Multimedia, vol. 7, No. 3 June 2005

[39] F. Kalleitner, S. Håkansson, I. Gódor and Á. Kovács *"M-Pipe- A Novel Media Delivery Framework"* , EuMob'06, European Symposium on Mobile Media Delivery, September 20, 2006, Alghero, Italy

[40] The Future of Voice - ITU Workshop, 15-16 January, Geneva

[41] Global IP Solutions homepage: http://www.gipscorp.com/high-quality-codecs/index.php

[42] Liveradio portal: http://liveradio.orange.fr/