Visual-communication technology has existed since the 1960s in various forms, including ISDN video-conferencing, internet-based desktop solutions, and immersive telepresence rooms. The technology is well established among enterprises, yet for several reasons, it has not been adopted on a mass-market scale.

Prerequisites for the mass-market adoption of visual-communication are good audio and video (AV) quality and optimum user experience at an appropriate cost. Trials have shown that users begin to appreciate visual-communication services once AV quality reaches a certain level. Although solutions that provide an acceptable level of quality are available, the equipment remains prohibitively expensive, preventing large-scale adoption.

The processing capacity required for HD video and high-quality audio is already present in many consumer products, including mobile phones. The cost of peripherals, such as displays, cameras and microphones, has declined significantly in a very short time. The convenience of components that are cheaper and commercially available has led to the development of high-quality visual-communication end points.

Interoperability and reliability are additional obstacles that have prevented the technology from being adopted for mainstream use. In order for visual-communication technology to be taken up by the mass market, users need to be able to connect to other users via completely reliable connections, irrespective of the network and platform they are using. User demands are beginning to drive development, and include desired functionality such as seamless switching to and from visual communication from other media, including voice.

The inclusion of all criteria—AV quality, reasonable cost of providing adequate AV quality, interoperability and reliability—in a system, calls for standardization that forms the basis of affordable, non-proprietary, high-quality visual-communication services. This article outlines the proof of concept (PoC) deployment at Ericsson of an evolved multimedia telephony service based on the established 3GPP IMS standard. The deployment shows that IMS can support a visual-communication system that meets the requirements for audio and visual quality, interoperability, reliability and cost, thus facilitating the entry of visual-communication services into the mass market.

**IMS platform**
IMS is the standard framework for IP-based multimedia services and was consequently chosen as the platform for the visual-communication PoC deployment. IMS supports some basic functions that are not yet supported by other proprietary systems. Such functions include a global addressing scheme, QoS control of access and transport, and standardized inter-working principles. GSMA has adopted an IMS-based VoLTE solution for voice and messaging, and has recently begun work on a complementary IMS profile for visual-communication services over LTE mobile access. IMS is ideally positioned to support the deployment of visual-communication services on mobile broadband devices such as smartphones and tablets—a factor

### Terms and abbreviations

<table>
<thead>
<tr>
<th>3GPP</th>
<th>3rd Generation Partnership Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>AV</td>
<td>audio and video</td>
</tr>
<tr>
<td>AVC</td>
<td>advanced video-coding</td>
</tr>
<tr>
<td>CCAS</td>
<td>conference control application server</td>
</tr>
<tr>
<td>DSL</td>
<td>digital subscriber line</td>
</tr>
<tr>
<td>EVS</td>
<td>Enhanced Voice Services</td>
</tr>
<tr>
<td>GSMA</td>
<td>Global System for Mobile</td>
</tr>
<tr>
<td>Communications Association</td>
<td></td>
</tr>
<tr>
<td>GUI</td>
<td>graphical user interface</td>
</tr>
<tr>
<td>HD</td>
<td>high-definition</td>
</tr>
<tr>
<td>HSPA</td>
<td>High-Speed Packet Access</td>
</tr>
<tr>
<td>IMS</td>
<td>IP Multimedia System</td>
</tr>
<tr>
<td>IP</td>
<td>Internet Protocol</td>
</tr>
<tr>
<td>IPTV</td>
<td>IP Television</td>
</tr>
<tr>
<td>ISDN</td>
<td>integrated services digital network</td>
</tr>
<tr>
<td>ITU</td>
<td>International Telecommunication Union</td>
</tr>
<tr>
<td>LTE</td>
<td>Long Term Evolution</td>
</tr>
<tr>
<td>MCU</td>
<td>multipoint conference unit</td>
</tr>
<tr>
<td>MRFP</td>
<td>media resource function processor</td>
</tr>
<tr>
<td>NNI</td>
<td>Network to Network Interface</td>
</tr>
<tr>
<td>PoC</td>
<td>proof of concept</td>
</tr>
<tr>
<td>QoS</td>
<td>Quality of Service</td>
</tr>
<tr>
<td>SD</td>
<td>standard-definition</td>
</tr>
<tr>
<td>SIP</td>
<td>Session Initiation Protocol</td>
</tr>
<tr>
<td>SVC</td>
<td>scalable video codec</td>
</tr>
<tr>
<td>VoLTE</td>
<td>voice-over-LTE</td>
</tr>
<tr>
<td>UNI</td>
<td>User to Network interface</td>
</tr>
</tbody>
</table>
that is instrumental in securing mass-market adoption of this technology.

Included in IMS are standardized definitions for interworking principles on the UNI and NNI interfaces, and IMS also provides a well-defined integration point, which can be leveraged to achieve interconnectivity with existing proprietary systems.

**Proof-of-concept implementation**

**Audio**

Source-signal bandwidth has a greater impact than anything else on audio quality. The human voice spans a frequency range of at least 50Hz-12,000Hz, and human hearing spans a similar or even larger range. Full reproduction of the human voice requires a sample rate of 32kHz or 48kHz.

Spatial listening in a telecommunication environment can be emulated using stereo capture and rendering. For audio-only clients, the additional spatial information that comes with stereo rendering improves the listener’s ability to determine who is speaking. For video clients, it further enhances the experience because there is a tighter correspondence between the video image and the audio stream. In other words, the voice of the person speaking appears to come from the direction of the screen displaying the video of that person. Stereo further improves the perceived audio quality in rooms with less-than-ideal acoustics.

Ericsson’s PoC IMS deployment uses stereo-audio based on G.719 encoding (an ITU standard audio codec). This recently standardized codec delivers a 48kHz sample rate at affordable bit rates and provides high-quality audio to fixed-access network users. Ongoing standardization activities for a new 3GPP codec, known as Enhanced Voice Services (EVS), will provide similar levels of audio quality at bit rates suitable for mobile-access networks.

Audio quality in real-time communication is affected by several other factors, including the acoustic environment, the electroacoustic hardware, speech enhancement, and transmission losses. End points should be designed to take each of these factors into consideration to ensure good audio quality. Subsequent network processing cannot compensate for poor-quality audio, frequency responses or echo control. The end points used in the PoC implementation include state-of-the-art speech enhancement functions such as full-stereo echo-cancelling, noise reduction and time-scaling jitter-buffers.

**Video**

Adding media to complement voice on an IMS-based system is a fairly straightforward process. The technology’s flexibility allows a voice call between two video-capable terminals to be transformed into a voice-and-video call by simply adding video media to the call. Similarly, a voice-only terminal can reject video media in an incoming multimedia call, and the call remains voice-only.

Video communication involves many of the same challenges as voice, such as firewall traversal, jitter and end-to-end latency. Video, however, is more sensitive to packet loss due to high codec compression. There is a trade-off between bandwidth and quality; determining the minimum acceptable bandwidth with respect to quality for various usages of video is an important consideration to ensure a positive experience for the user.

Ericsson’s PoC IMS deployment uses the H.264 advanced video-coding (AVC) video compression standard. This codec is widely used in areas as diverse as digital television, Blu-ray, YouTube and 3GPP mobile communication systems. AVC supports video quality ranging from thumbnails at rates of a few kbps to cinema-screen-sized full HD at several hundred Mbps. Ericsson Research contributed to the standardization of AVC, and recently optimized an internal version to support real-time simultaneous encoding and decoding of full-motion HD-quality video on a PC.

As with most other video-coding standards, AVC generates a video stream with a variable bit rate. However, the output rate can be controlled through an encoder-internal feedback loop, the algorithm for which is not standardized. Consequently, Ericsson’s PoC implementation includes several different types of rate-control algorithm, ranging from IPTV to low-delay conversational. The AVC standard allows considerable freedom to set encoding parameters, making it possible to trade processing...
requirements for video quality at a given bit rate, as illustrated in Figure 1.

The video codec parameters for the PoC implementation were selected using ITU Standard P.910, Subjective Video Quality Assessment Methods for Multimedia Applications. These assessment methods are used to evaluate multimedia applications to select algorithms and assess quality during an audiovisual connection.

Tests were conducted in native and up-scaled resolutions, and video quality was assessed as a function of resolution, bit rate and codec parameters. Professional high-quality video recordings from typical conferencing situations were used as source material for the tests.

Usability
An essential part of the PoC system design process were usability studies. The conclusion of these studies was that high media-quality must be complemented with ease of use if visual communication is to gain widespread popularity.

The influence of quality and ease of use on the PoC trials is exemplified in the way pre-booked multi-party video calls are handled. At startup, all client-related booking information is retrieved from the central booking system. Booking information is continuously updated as long as the client is registered. Booked meetings are displayed in the GUI, and the user can join the desired meeting by selecting it from the list displayed.

The booking function for meetings has been integrated with e-mail and calendar systems through the development of a plug-in for Microsoft Outlook. This plug-in allows users to add video to a regular meeting simply by clicking a button in the calendar application. Other aspects of the PoC user interface, such as the address book and call-handling procedures, were designed after consideration of the results of the usability studies.

Architecture
As a consequence of Ericsson’s PoC trial results, two additional functions were added to IMS to extend its capabilities: a conference control application server (CCAS) and a media resource function
processor (MRFP). These functions provide a simple mechanism to manage clients participating in a visual-communication session.

A visual-communication session can be a scheduled as a multi-party (conference) call that each client connects to, or an ad hoc multi-party call initiated by one client. The CCAS instructs the MRFP, which is the IMS term for a multipoint conference unit (MCU), to create a voice and video bridge for the clients, while the MRFP ensures that all participating clients have audio and visual contact.

For multi-party calls, there are several ways of distributing media to clients, and each method has its advantages and disadvantages. In the distributed conference method, each client sends all relevant media to all other clients. All communication is sent directly between peers and requires no specific media support in the network, minimizing delays. The main disadvantage of this approach is that the total bandwidth consumed by the call rises exponentially with each additional client. Furthermore, unless a common codec with a shared configuration is used, each client must encode the media individually to each peer, placing a significant computational load on the sender.

An alternative method – centralized conference – lets each client make a point-to-point call to an MRFP in the network. Bandwidth demands increase linearly with each additional client, and provided that the MRFP can support a large set of codecs and configurations, each sender can use their preferred media encoding. The disadvantages of this approach are:

- participants have less freedom to compose and present media;
- the complexity of the MRFP can become significant;
- media is often delayed while being processed; and
- media processing can cause quality degradation due to the transcoding introduced when composing, resulting in a mosaic video image of the participants.

The IMS PoC implementation is based on a combination of centralized and distributed methods, exploiting the advantages of each method to the greatest possible extent. For audio, the MRFP terminates the media plane and mixes the audio signals from all active participants, in the same way as it would during a regular centralized conference. For video, however, the media is switched to the furthest possible extent without re-encoding it, as shown in Figure 3.

To facilitate video switching without the need for transcoding in the MRFP, a number of common formats have been adopted. HD-capable clients send three video streams – HD, SD, and thumbnail – to the MRFP. Standard-definition clients send two streams: SD and thumbnail. In this way, clients with reduced processing capacity or limited-access bandwidth can just send thumbnails. The MRFP relays the appropriate streams to the other clients based on speech activity or manual user choice.

In the PoC system, a client receives one high-resolution stream – HD or SD, depending on the capabilities of the client – and several thumbnail streams. The client can then present full-screen video of the active participant with thumbnail overlays of the non-active participants. The system permits any combination of video images to be negotiated between the client and the MRFP.

Because each client sends multiple video streams, the screen layout of each participant can be tailored, while at the same time reducing the transmission delay and avoiding any loss of quality due to transcoding in the MRFP. A similar solution could be implemented using the H.264 scalable video codec (SVC); however, with the conditions at hand – video streams with significant differences in resolution and a requirement for efficient PC implementation of the codec – the AVC solution is a more advantageous one.

The PoC implementation includes two end points: a client intended for fixed installation in conference rooms and a personal client for desktop or laptop use. Both types are implemented as soft clients on standard PCs and use affordable off-the-shelf consumer electronics for AV delivery.

**Interoperability**

An essential part of building a visual-communication service for mass-market deployment is interoperability. Mobile telephony would not have achieved the market success it enjoys today if it had been based on non-compatible, proprietary protocols. In contrast, the visual-communication market continues to be dominated by numerous solutions that are mutually incompatible, creating an acute need for a single standard.
Several major players in the market have already initiated standardization activities. The use of IMS, however, solves this problem. As an open standard, IMS is available to all vendors, and unlike many proprietary solutions, it can be implemented on systems from mobile devices to telepresence rooms, and on different network architectures such as DSL, HSPA and LTE.

IMS provides standardized user-to-network and network-to-network interfaces that allow operators to create networks that support multimedia features beyond messaging. IMS meets user demands for seamless switching to and from video communication from other media such as voice in a generic and interoperable way. Ericsson’s IMS deployment uses a standardized SIP interface, which is becoming increasingly popular for enterprise video systems.

Reliability – the operator’s role
Operators have a key role to play in visual communication. Because real-time visual communication requires audio and video synchronization, it is vital that this service is provided on a high-quality core network with low packet loss and low latency. For mass-market breakthrough of visual communication, Reliability and trust are key issues; and operators are uniquely positioned to offer this service. By providing reliable and secure visual-communication services, operators can reduce subscriber churn from internet-based competition while driving IP traffic into networks where it can be controlled and monetized.

Summary
An evolved multimedia telephony service using IMS has already been deployed, using commercially available, affordable PC and audiovisual equipment. The PoC deployment started in 2010, and there are now more than 40 multimedia conference rooms installed in Ericsson facilities worldwide. Feedback indicates a high level of user satisfaction, and numerous requests for additional installations have been received.

The success of the project shows that an IMS-based system meets each of the criteria – sound quality, video quality, interoperability and easy switching of media – that users require from a visual-communication system, with operators supplying the remaining piece of the puzzle: reliability.

The IMS standard is ideally suited to support the expansion of the visual-communication market from the enterprise to the consumer sector. Similarly, visual communication can become the mass-market service that IMS requires to become widely adopted by operators – and that, will result in satisfied users who can finally enjoy the service quality that will make visual communication mainstream.

The next step
Ericsson’s PoC has demonstrated the possibilities offered by an IMS-based visual-communication solution. It is clear that the potential for an operator-provided, high-quality mass-market service for video meetings is great.

This conclusion has led to the ongoing work of bringing ideas from the PoC to the market. With close cooperation between two of Ericsson’s business units BU Networks and BU Multimedia, a commercial solution is currently being defined.

Bo Burman
Current working as a senior specialist in video telephony at Ericsson Research Multimedia Technologies. Since joining Ericsson in 1996, he has worked with video compression and various aspects of mobile video communication in multiple research and business unit projects, including standardization in 3GPP, ITU-T, IETF and MPEG. He holds an M.Sc. in computer technology and engineering with a focus on telematics from the Institute of Technology at Linköping University, Sweden.

Anders Eriksson
An expert in audio media processing at Ericsson Research. Since joining Ericsson in 1994 he has been active in research, standardization, and implementation of speech signal processing aspects of the voice service. He holds an M.Sc. in applied physics and electrical engineering from Linköping University, Sweden and a Ph.D. in automatic control from Uppsala University, Sweden.

Andreas Bergkvist
is a strategic product manager in Product Line IMS with a focus on short- and long-term strategy development for securing a competitive Ericsson offering in the video communication area. He is also solution manager for the VisualCom program. He joined Ericsson in 2005 on the R&D Global Graduate Program and has since had various roles, primarily in the development organization for the Ericsson Session Border Gateway, including project management, system management and technical product management. He holds an M.Sc. in engineering physics from Uppsala University, Sweden.
Klaus Schneider

joined Ericsson in 1991 and is currently working as program director for the VisualCom program with a focus on product planning and industry relations. He has held various positions at Ericsson R&D in development, project management, and line management, including head of a product development unit with responsibility for introduction, deployment and support of Ericsson’s Core and IMS product range. He holds an M.Sc. in mathematics from the Technical University (RWTH) of Aachen, Germany.

Håkan Djuphammar

is vice president of Architecture & Portfolio at Group Function Technology & Portfolio Management, and is responsible for overall architecture and portfolio issues for the Ericsson Group. He holds an M.Sc. in telecommunications from Imperial College, London, UK and a B.Sc. in electrical engineering from Chalmers University, Gothenburg, Sweden. Previously, he was vice president, Product Line Management within Ericsson’s Business Unit CDMA Mobile Systems. He has held various positions at Ericsson since 1991, including 3G product management and PDC system management. Before joining Ericsson, he was co-founder and president of Abstract Electronics AB.

25 years ago

In 1986, Ericsson Review published an article on the frequency planning of digital radio-relay networks. Similar to the article in this issue microwave capacity evolution, it mentions that the task of the network planning engineer is to select radio frequencies and antenna types in such a way that the influence of interfering signals is within the margins of the planning objectives for the overall performance of the radio circuit.

50 years ago

In 1961, an article was published about extending telephone plants with regard to value of subscribers time. While this article essentially covers the cost and dimensioning aspects of plant expansion, it is based on figures measuring the inconvenience to subscribers as a result of congested networks. The article mentions subscribers’ demands in terms of accessibility and intelligibility of conversations and offers a calculation method for determining the number of switches needed. Interestingly the calculation concludes that the greatest factor is the value of subscribers’ time.

75 years ago

In 1936, Ericsson Review marked the 60 years since Lars Magnus began the activities that would lead to the creation of the Ericsson Group. An article depicting the evolution of Ericsson’s telephone instruments between 1878 and 1935 shows how tastes and raw-material technology have shaped the telephone.