Taming the Beast in Mankind – Telecommunications in the 21st Century

InterComms talked to Ecma TC32-TG22's Convenor and Swissaudec's CEO Clemens Par about the 21st century's broadcasting and communication means

Clemens Par, CEO, Swissaudec

Clemens Par pursued parallel studies in conducting at Hochschule "Mozarteum" in Salzburg and mathematics at ETH Zurich. Besides his artistic projects, e.g. as a painter, author, musician, presenter and executive producer for ARD, ORF and Schweizer Radio DRS, his scientific work focuses on inverse problems in audio engineering and invariant theory. Results have been standardized as the world's first 3D audio standard ECMA-407 in June 2014.

Clemens Par received the WIPO Award 2009 for his IP. He is CEO of Swissaudec (a Swiss enterprise specialized in 3D audio coding technologies), Expert of ISO/IEC JTC1/SC29/WG11 (MPEG), and Convenor of Ecma TC32-TG22.

ECMA-407, MPEG-H and ATSC are currently shaping the future of UHD TV up to 8K resolution and NHK 22.2 audio. A low-delay profile for ECMA-407, crafted by Swissaudec, however, may revolutionize tomorrow's mobile telecommunication and teleconferencing means – offering NHK 22.2 capabilities down to 48kb/s with a latency of one frame.

Q: Broadcasting and telecommunications industries have led parallel lives so far. You made a very firm technological statement at IMTC 20th Anniversary Forum about future technology options in this field. Could you share your vision of 21st century's multimedia scenario with our readers?

A: The very visionaries are not found in technology, they are primarily found in literature. Transgressing technology limits of one's own time has lead to the considerable oeuvre of Jules Verne; more critical approaches may, for instance, be found with Aldous Huxley. Both largely have become reality.

Unlimited visualization means and immersive audio are visionary imaginations and inventions of the past. The first 3D broadcast via telephone lines occurred in the 19th century in Paris. Three-dimensional imaging, though painted, even dates back to the 18th century.

Switching on a screen and communicating "like in reality" nowadays has become a multi-billion market. Nevertheless, broadcasting and telecommunication industries have gone separate ways, due to historic reasons: bitrate budget for telecommunication would then have compromised broadcasting quality. The stigma has remained – despite the fact that we have brilliant technologies at hand, which can equally fulfil both needs: UHD TV and advanced telecommunication.

Q: Teleconferencing rooms already go immersive. Is your stated restriction of quality still valid?

A: I have been invited to visit a well-known teleconferencing company's site with screens of several square meters. Even the room's colour was optimized for video capture, whilst cheap plastic table microphones with mono transmission captured audio. Even in modern times, people are not supposed to appropriately listen to their fellow-beings in a realistic acoustic environment. It is all visual appearance.

The fault is partly cinematic: 3D cinema sound is over-realistic and evidently reflects bad taste. In silently quoting ►

Paul Valéry: people have aesthetically grown deaf and seem to perceive no need for advanced teleconferencing audio means. We are far apart literary science fiction where people wished to explore, learn and discourse. Social media feed this innate need today, however, on technologically lowest level. Telecommunication industries, however, refrain from kissing this sleeping beauty: it's a tragic fairy tale!

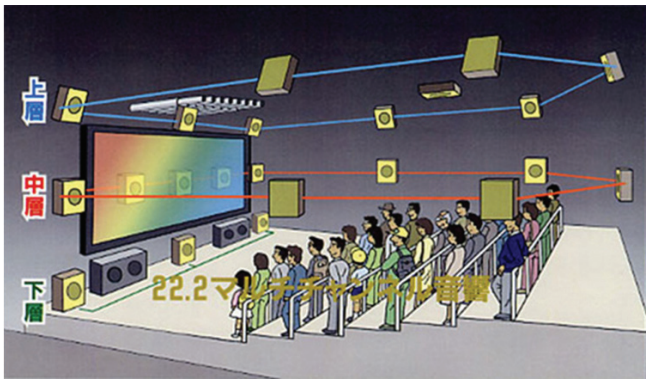


Figure 1: NHK 22.2 represents the state-of-the-art loudspeaker setup for 3D audio content production and delivery. It will be broadcasted over satellite by Japanese ARIB from 2016 onwards with AAC and is complemented by 8K HEVC video coding.

Q: You have been a researcher for more than a decade in the field of multimedia audio with numerous discoveries, which have been normalized as the world's first 3D audio standard ECMA-407 in June 2014. Could you give our readers a little overview?

A: I am not very much an engineering guy *(laughs)*. My love has always been music, arts, pure science and its unforeseen application for the benefit of human welfare. As a professionally trained musician I have a natural affinity for beautiful soundscapes. I started as a painter with my first exhibition in Paris in 1989/90. Visual arts extend our imagination above the level of the current "calamity show" (as modern film production was called by Arnold Schoenberg's pupil David Raksin). My professor Rudolf E. Kálmán laid the foundations for my early interest in systems theory, which, though having very eminent applications, is nothing but pure thought.

My discoveries are threefold: I solved the first inverse problem in audio in a rather playful and unplanned way in 2002 when trying to find a mathematical substitute for a stereo microphone for my private studio. Inverse problems were unknown in these times in this field and, as Michael Dickreiter in his *Handbuch der Tonstudiotechnik* stated, could not even be solved.

However, Michael A. Gerzon, a very eminent mathematician at Oxford University, already had worked into this direction. My final solution eliminates frequency as a degree of freedom and stays within the realm of time and level in providing sufficient degrees of freedom. It has become the basis of my first key technology patented in 2008 and the basis of international standard ECMA-407. Models of this kind, now called inverse coding in the science community, were never applied to audio coding. So far, spatial properties had to be extracted by means of a Fourier transform and to be transmitted as opulent side information. This method is called parametric coding.

Contrarily, given an existing spatial audio signal, you may calibrate the inverse problem in such way that spatial audio may be restricted to an, occasionally transmitted, parameter set packet. For instance, a spatial data packet of ECMA-407 with NHK 22.2 requires less than 100 bytes and may last for several minutes. With parametric coding, at least 40kb/s are required in the low bitrate range with parametric coding. Given equal performance of the used base audio codec, ECMA-407 may achieve equal performance at lowest bitrates with, for instance, MPEG-H, however, with an unrestricted number of output channels up to NHK 22.2 at *all* bitrates.

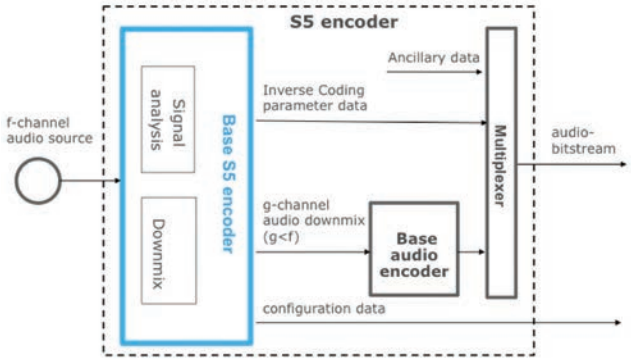


Figure 2: ECMA-407 encoder. The normative signal analysis, which may be based on invariant theory and then works in real-time, provides the lowest spatial bitrates ever achieved.

Since 2009, I have been in constant exchange with Rudolf E. Kálmán who awakened my interest in invariant theory, then a discipline of pure mathematics. David Hilbert is well known for Hilbert space or Hilbert transform. It, however, is not common knowledge that this eminent German mathematician from his thesis onwards made his foremost discoveries in invariant theory. Invariants are coefficient functions, which, like certain bacteria, essentially "survive" transforms and, as Hilbert proved in 1893, luckily form a field - arousing the scientific suspicion that such invariants existed with Gaussian (random) processes! However, these algebraic objects never were isolated.

In 1903 Grace and Young published a book on invariants, which captured my attention in 2010 with respect to apolarity behaviour (a state when invariants vanish). Apolarity proved to be key to solve this problem. This solution has been patented in 2010 for Gaussian signals. Invariants then became part of applied mathematics.

Our invariant-driven inverse encoder within the framework of ECMA-407 makes use of these results. Invariants are much faster than is statistical analysis; they require very little known data. This is why our ECMA-407 implementations work in real-time and make equally way for UHD TV and telecommunication industries.

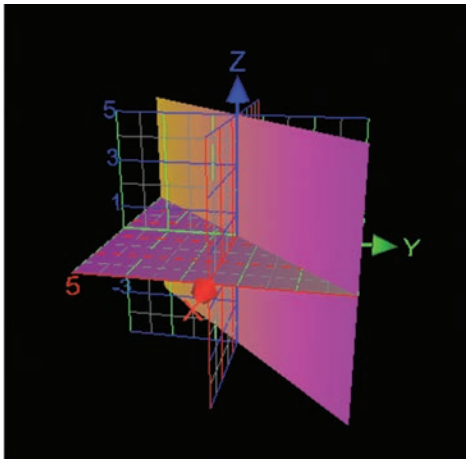


Figure 3: A rotating vertical plain is the answer to a one-century-old problem: its associated invariants represent a sound real-time alternative to statistical analysis.

Inverse coding works in time domain and requires almost no computational effort. My third scientific goal was to find an equivalent to this technology in the Fourier field, with lowest possible computational complexity and minimal latency. My research was likewise successful in this field, which means that up to the double number of output channels can be created from a downmix. Contrarily to parametric coding in frequency domain, this method requires no side information at all, which makes it fully conformant with international standard ECMA-407. Computation essentially is related to going from time domain to frequency domain and back. When using this technology, latency can be restricted to one frame, as with most speech coders.

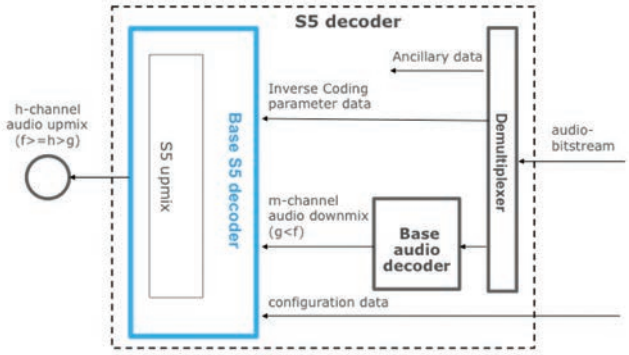


Figure 4: ECMA-407 decoder. Its normative S5 upmix performs in real-time, according to the conveyed inverse coding parameter data.

Q: You are the Convenor of Ecma TC32-TG22, and likewise an MPEG Audio Expert since January 2012, primarily working in the field of UHD TV audio transport. In August 2014, your company launched the first ECMA-507 UHD TV test carrier in co-operation with France Télévisions and SES Astra. What is your personal forecast regarding UHD TV transmission?

A: MPEG-H 3D audio has been primarily been driven by Japanese broadcaster NHK. We already know 8K recording and HEVC coding equipment complemented with NHK 22.2 audio and AAC, after NHK's years' long scientific promotion of MPEG-H 3D audio under Kimio Hamasaki. Contrarily, MPEG-H and ATSC both know pre-defined proprietary reference quality encoders, which, unlike ECMA-407, are not compliant with AAC or HE-AAC. Japanese ARIB has announced its first UHD TV test satellite broadcasts for the Olympic games. It may be expected that UHD TV sets, which are capable to *render* 3D audio, will be launched by the Japanese and South Korean industry about this time.

Rendering NHK 22.2 is far from being trivial, particularly on loudspeakers, as wavefield synthesis or cross-talk cancellation have to take place inside the device itself – no average consumer will ever set up a complicated, properly measured NHK 22.2 listening room! Virtualization means like wavefield synthesis and crosstalk cancellation both work decently; loudspeakers can then be directly mounted within the UHD TV's frame. NHK demonstrated such a system next to our booth in IBC 2014's "Future Zone".

Real NHK 22.2 loudspeaker reproduction only happens in the studio and in the laboratory. MPEG-H may be expected to become international standard in 2015, while ATSC standardization is still on its way. Dolby, DTS and Barco have already launched 3D sound in cinema and have stirred public awareness of this subject. Broadcasters currently explore 3D audio production and transport.

However, neither the American nor the European markets are prepared for the broad launching of 3D broadcasts. The same situation is valid for South Korea, another technology driving market. Given the fact that UHD TV has been announced by ARIB to be broadcasted regularly from 2020 onwards, UHD TV may expected to be launched first in South Korea, followed by Europe and by North America.

Swissaudec's current focus with its ECMA-407 implementations, however, lies on webcasts and telecommunications, supported by its automatic 2D to 3D upmix as a substitute to genuine content production. (Similar technologies are likewise applied in 2D with HD TV for the automatic conversion of stereo to 5.1 Surround.)

Invariant theory plays a key role in 2D to 3D conversion; it allows us to determine the most suitable spatial model for the upmix in real-time up to NHK 22.2.

UHD TV is a multi-billion market, which still requires sustainable investments from the side of industry. We may anticipate this huge leap in multimedia content delivery by catering 3D audio content up to NHK 22.2 over the Internet.

Q: There is no common awareness of 3D audio rendering techniques. Could you give a short explanation how 3D audio may be consumed on a mobile device?

A: Primary consumption is via binaural means. With Smartphones, tablets and computers, we have experienced the "headphone boom" and the rising of brands like Beats or Sennheiser. Binaural 3D audio was already commercially launched in the seventies with little commercial success – the "Walkman" simply had not yet been invented! Markets are evidently 'out of phase': broad 3D audio consumption could have been made possible twenty years earlier!

Binaural 3D audio technologies are technologically nothing exciting. Every mammal brain is used to adapt its perception of localisation to differences in time, amplitude and frequency caused by its head's exterior anatomy. If you record sound with a dummy head you anticipate what my friend Günther Theile calls *inverse Filterung*: when such signals are played back on headphones the said differences in time, amplitude and frequency automatically represent localisation cues to the human brain.

Interestingly enough, such artificial head signals sound very bad when being played back on stereo loudspeakers. *Inverse Filterung* in the brain includes an equalization step for restituting the original frequency response. No sound processor currently equals such amazing capability! In the strict sense this likewise imposes an inverse problem. As Günther Theile proved, the brain can be fooled with respect to frequency response and its complementary localization cues.

If I now record sweeps in an NHK 22.2 laboratory for each loudspeaker with a dummy head, I may subsequently convolve each channel of a 3D audio signal with such measurements. All of a sudden, my headphones become an NHK 22.2 listening environment! This is what 3D rendering on mobile devices is all about.

Figure 5: Immersive 3D audio solutions on mobile devices. Swissaudec in parallel addresses the multi-billion UHD TV and telecommunications markets.

Q: What is your market prognosis regarding immersive audio for mobile devices in the near future?

A: Mobile devices currently face severely declining sales, see, for instance, Samsung's alerting revenue forecasts for Smartphones, due to market saturation. The consumer now has a sufficient level of functionalities and third party applications at hand. Unique selling propositions on highest technological level therefore are of growing importance.

The foremost USP evidently is a severely enriched personal experience, which is synonymous with "immersive audio". Due to screen size constraints, visual immersion on a smart device is wishful thinking unless complemented by complex wearable hardware. Contrarily, creating an immersive experience via headphones is easy!

A boom in 3D audio may therefore be expected in parallel with UHD TV, which stirs public awareness for immersion. UHD TV unconsciously educates to interpret complex audio localization cues together with video. The same visual cues, now catered by the rather small smart device's screen, all of a sudden become virtual reality through "trained" imagination – they are all stirred by immersive 3D sound!

Swissaudec expects the smart device segment, with or without wearable hardware, to become the primordial market segment for 3D audio from 2016 onwards. Its ECMA-407 implementations are in the sweetspot with respect to available bit budget from 48kb/s to 128kb/s (currently YouTube is consumed at an average of 48kb/s with HE-AAC in stereo).

Stereo, however, is not immersive – as it presents erroneous cues to the human brain: sound sources are localized *in* the head – a common, however, fully unnatural experience. Do you expect the conductor together with his big orchestra perform Bruckner *in* your basal ganglions? This currently happens in six billion devices equipped with AAC and HE-AAC.

The only standard compliant with these codecs is ECMA-407 - capable of making immersive audio *immediately* happen on these six billion devices!

Q: You have primarily answered in terms of future UHD TV. However, what is the direct impact of ECMA-407 for the telecommunication industry?

A: Setting up an audio group in an IT standardization body like Ecma International, known for optical storage or ECMAScript (better known as JavaScript), would have been a bold enterprise 10 years ago. Now the world has grown smaller with smart devices with sufficient memory and high computational power and has turned into an amazing multimedia world where communication perfectly fits in!

Communication, however, is a conservative market: for the reasons already stated, we still live in the mono stone age of highly optimized speech coders, which are fully agnostic of ambient "non-verbal" communication in the broad sense.

The essential lesson is taught by shared video content on social networks – what people wish to express by means of images and sounds goes far beyond verbal expression truncated by speech coders. The future of teleconferencing is via smart devices sharing a virtual and yet real environment over the globe with intelligent and secure communication systems.

In my opinion, security plays a key role – no company or professional organization wishes to share confidential information to a hidden public in the wires, and the added value with respect to virtual reality only remains a long-term business case if security is properly addressed.

ECMA-407 is perfectly compliant with virtual reality communications, as lowest delay may be achieved with highly demanding formats like NHK 22.2. You may now ask the question why such a complex format may be interesting for telecommunications – according to Günther Theile,

these channels are primordially perceived as *point sources*. You may thus allocate multiple speakers to their precise position in space – with current speech coders they are all unnaturally sitting *in* your basal ganglions!

There is a precise psychoacoustic reason for advanced telecommunication means – either for transporting "non-verbal" ambiance or for creating a natural interlocutory environment with teleconferencing.

Q: Is your technology proposal to adapt 3D codecs to future enhanced communication environments and to quit the world of speech coders?

A: ECMA-407 has been designed this way – mostly by public broadcasters who wish this codec to serve the needs of UHD TV. It simultaneously is a telecommunication means because of its low complexity, low delay capabilities. An ECMA-407 decoder currently requires 33.7 MOPS per second in C++ for UHD TV and can be seamlessly switched to low delay for telecommunications. It's an all-in-one tool for the television set or settop box, for the tablet and for the Smartphone.

Didn't you ever wish to talk to people on the screen in your living room simply via the Internet and probably even share multimedia content you are just watching? There is a desperate need for a dynamic Facebook substitute in multimedia – which curiously enough has been never adequately met by industry! I have been leading a study on this segment together with other well-known manufacturers and researchers whilst standardizing ECMA-407 as an open and base-coder-agnostic concept. Continuous one-way communication is never appreciated when alternatives are at hand.

My kids spend more time in social networks than I would ever do in my lifetime. As a grown-up you may think of this phenomenon in terms of decadence. The truth is that grown-ups have only been conditioned to one-way communication through their television sets, which populated homes from the fifties onwards. I personally grew up without television set, only with my books, my friends and my pets. It seems that I am neither conditioned for one-way communication nor for social networks over the Internet (*laughs*).

Q: What is your technical conclusion for the design of a future telecommunication system?

A: High video resolution on a smart device is already feasible. When looking for a perfect microphone, you already find it in the MEMS world: omnidirectional, capable of beamforming and with excellent sound quality. Local transmission and optical fibres complement such a system in combination with a central processor, which is capable of creating a soundscape from the multiple worlds of interlocutors.

Interlocutor 1 is in the New York City subway and the hammering of a passing train can be heard in the background. Interlocutor 2 talks to Interlocutor 1 whilst somebody is playing piano in the background. Interlocutor 3 is lying on the beach, and all enjoy the waves' background ambiance. This is communication on highest level, which broadcasting industry perfectly knows to combine in a three-dimensional environment – we have broadcasted such stunning content by France Télévisions over satellite! The same is valid for images.

Automatized computer processing, based on prior artistic knowledge, creates this virtual world, Jules Verne has been dreaming of in the 19th century. As for audio, we know how to transport such an entire scene at bitrates as low as 48kb/s with ECMA-407, and HEVC likewise points towards the future of video compression and transport. My conclusion is that the future is here right now but that conservative business models in telecommunication impede true innovation. We would be most happy to provide such experience with ECMA-407 already in 2015!

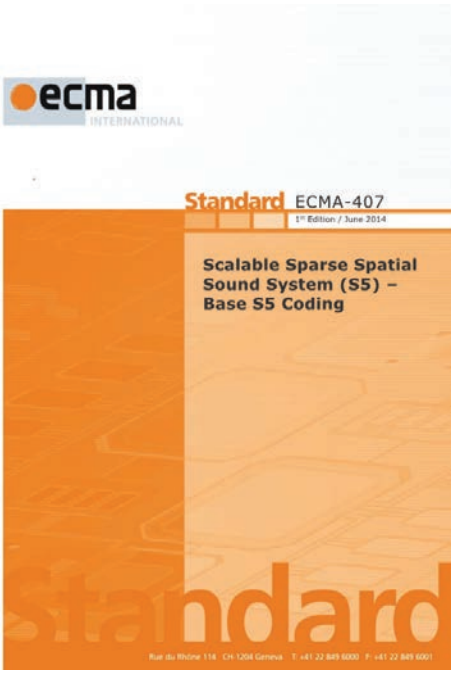
Q: You will be present in NAB Labs Futures Park at the 2015 NAB Show in Las Vegas with an UHD TV premiere for ECMA-407 in North America. Do you see a realistic chance to reconcile the broadcasting and telecommunications worlds on short terms?

A: Visionaries hate shortsighted reality. We evidently will likewise showcase our low delay ECMA-407decoder with UHD TV, and it is our hope that the same number of engineers will be aware of this technology revolution as was the case in IBC 2014's „Future Zone": we welcomed more than 1'500 visitors! Some of them were active in telecommunications, and it is my hope that they will spread the word for a visionary interactive multimedia concept, which will help people around the globe to understand their world in a better way.

Figure 6: ECMA-407 in IBC 2014's prestigious "Future Zone" with an ECMA-407 satellite test carrier and ECMA-407 on mobile devices.

If, for instance, such multimedia communication were enabled between Africa, Europe and North America, people in the first world would have been confronted with the rude consequences of Ebola. Human empathy is developed through our sensory capabilities, which are highly restricted by current telecommunication means. If such communication were possible between Ukraine and the whole world, including Russia, the observed mutual political isolation would probably never have happened.

In such context, true technology progress is desirable. This has been Jules Verne's initial vision. This has been the foremost hope and despair of Paul Valéry, the very root of his cultural pessimism. He knew about the beast in mankind, only tamed by beauty or truth. Telecommunications may serve the spreading of both: beauty and truth!



ECMA-407 is the world's first 3D audio standard approved in June 2014. It is primarily based on inverse problems, as formulated by Russian-Armenian astrophysicist V. Hambardzumyan in 1929, which enjoy high popularity in fields like physics or tomography with three scientific journals. In audio coding, inverse problems severely reduce the amount of spatial data, which needs to be conveyed.

ECMA-407 describes a scalable multichannel coding system for spatial audio data compression, which can be applied to provide 3D audio experience with little overhead. Such system may incorporate a wide range of state-of-the-art audio codecs like AAC, HE-AAC, Ogg or USAC. By using an audio codec, which may offer encapsulation capacity for external data, the entire ECMA-407 bitstream may be carried within the audio coder stream with little overhead and maintain a compatible bit stream syntax. ECMA-407 thus becomes "invisible" even with highly complex 3D audio formats like NHK 22.2.

ECMA-407 specifies the base S5 encoder and decoder in terms of configuration data, downmix, inverse coding parameter data and upmix and provides reference and guidance on how to incorporate further components. See http://www.ecma-international.org/publications/standards/Ecma-407.htm.

For more information please visit:
www.ecma-international.org/publications/standards/Ecma-407.htm
www.swissaudec.com