

Rationalism versus Empiricism

A Crash Course in Invariant Theory and a Tribute to Rudolf E. Kálmán by Ecma TC32-TG22's Convenor and Swissaudec's CEO Clemens Par



Clemens Par, CEO, Swissaudec

CLEMENS PAR has introduced inverse problems and invariants to the world of audio coding. ECMA-407 standardizes these results as the world's first UHD 3D audio codec. Clemens Par is founder and CEO of Swissaudec, a young Swiss codec company located near Lausanne in the Canton of Vaud, highly active in international standardization at Ecma International and inside ISO/ MPEG. This publication is a tribute to professor RUDOLF E. KÁLMÁN as an ever-inspiring, foremost scientific character. Professor Kálmán received the National Medal of Science from U.S. President Barack Obama in 2008.

Systems theory is a domain well reserved by mathematicians. Whatever has swapped over into the engineering world, and particularly into the media industry, is a well-established set of formulae primarily based on statistical discoveries, with rare exceptions like the Kálmán filter, which shows "surprisingly, that the Wiener problem is the dual of the noise-free optimal regulator problem" [1]. For instance, the ingenious theorem of Bayes from 1763 or Pearson's Principal Component Analysis (PCA) from

1901, unfortunately an eugenicist, have conquered the industrial world, repeated endlessly and always looking upon phenomena in an arbitrary way [2, 3]. Fourier analysis and related methodologies like QMF filterbanks have led to endless research in video and audio analysis. No exciting news out there.

Audio, however, is a highly controversial subject, as models how human hearing occurs are closely linked to cerebral activities, which are only investigated up to a specific degree. The foremost model, conceived as a doctoral thesis in the eighties, is the so-called "Assoziationsmodell" by my friend Günther Theile [4]. See Figure 1.

Linking to mathematical models is poor and primarily based on experimental data, for instance, the famous discovery of Theile, Stoll and Link, which represents the very basis for psychoacoustic codecs, namely masking [5]. Evidently our brain is capable of recovering full information from quantized frequency response, which remarkably points towards the fundamental pattern recognition models of Rudolf Arnheim, which according to Rudolf E. Kálmán might be looked upon from an invariant principle, as will be discussed below, and seems to be steered by synaptic

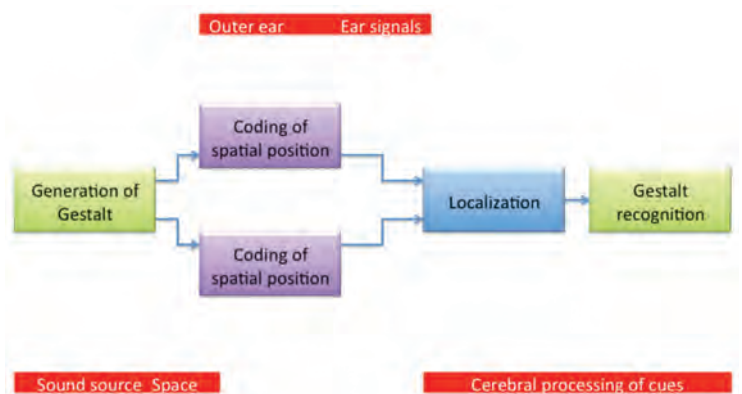


Fig. 1: "Assoziationsmodell" according to Günther Theile, describing the human auditory response to an external stimulus [4].

- ▶ activity, as indeed proofed by Nobel prize laureate Eric R. Kandel [6, 7].

The multimedia consumer is highly fixed on results, without understanding that his underlying pattern recognition is most professionally fooled. Fooling, however, is clearly not the business of pure or applied mathematics, my very discipline.

To come back to systems theory, a top-level model in audio, in analogy to physics eternal dream of the "Weltformel", would be most desirable. However, current foremost models of spatial audio, namely phantom-based models (by generating correlated signals), object-based models (by describing the three-dimensional position of a mono source in space) or scene-based models (by using spherical harmonics to model a virtual, head-related soundfield) poorly interact, and, if so, only on a theoretical basis. For instance, object-based wavefield synthesis (by generating a sound field by multiple loudspeakers and the Huygens principle) and scene-based models are mathematically identical, given an infinite number of loudspeakers. In practice they are not at all, due to spectral aliasing and other nasty side effects. Whilst object-based and scene-based models attempt to synthesize a natural sound field, phantom sources do not even occur in nature but only in the human brain. Provided that two loudspeaker signals show slight differences, sound sources miraculously appear BETWEEN the two loudspeakers. If you indeed look for them, where perceived, they are gone. The reader may easily verify this with any HiFi-setup at home, when switched to Stereo and when playing back indeed a stereo recording. Welcome to the everlastingly amazing world of psychoacoustics and pattern recognition!

Each multichannel audio signal may be easily transformed from time domain to frequency domain, i.e. the oscillation of air molecules is analyzed with respect to partial tones, which may be easily represented e.g. with Fourier or QMF analysis. From there onwards, empirical models take over, and consequently cause a nasty conglomerate of discriminators and libraries, which are legion. They all share the common disadvantage that their application has to take place in the encoder and, consequently, results need to be in the codec's bitstream, hence puffing up bandwidths.

ECMA-407 is the first codec to show the virtue that the load of the encoder is fully moved to the decoder's side. In IBC 2015's „Future Zone“ and "Technology in Action Theatre" Swissaudec, in co-operation with SES and France Télévisions, showcased an ECMA-407 NHK 22.2 satellite audio carrier, which is transported over a normal 7.1 MPEG-4 carrier with less than 2kb/s additional ECMA-407 payload - loudness and all relevant broadcaster data included! Tests led by a public broadcaster show statistically equal performance with the internationally leading, competing technologies.

Though one may argue that ECMA-407 is a phantom-based system, this standard's signals are mid-side signals, i.e. co-incident, and all complementary technologies in frequency domain are able to reconstruct the original signals with highest approximation from a given downmix (i.e. the

summing up of the original channels in order to reduce bandwidth). This implies that Swissaudec's fully-grown ECMA-407 codec transmits ALL formats, whether phantom-based, object-based or scene-based, due to the nature of used spatial loudspeaker representation techniques (i.e. mid-side signals and correlation-preserving techniques in frequency domain).

Describing NHK 22.1 by a 5.1 or 7.1 carrier, which is upmixed to the four- or threefold by less than 2kb/s payload - internally multiplexed - with statistically equal performance to foremost parametric methods, seems to be sorcery.

Let's assume for a moment that I am a magician indeed: I would need in such case NO side information at all. Apart from a bin-per-bin analysis in frequency domain, which, with zero side information, is able to recover up to 100% additional channels from a downmix with highest spatial fidelity, by simple *a priori* knowledge of the downmix, this fortunately is not the case. Evidently I must be a mathematician and not a magician.

However, I prefer being a magician as a mathematician, with a trivial mathematical trick behind. Systems theory is well established, provided enough insight of the system's INTRINSIC behaviour is given. Parametric coding may be looked upon as a very special case of such culture of thought, tailored to our hearing by empirical analysis, as already described, and by endless series of psychoacoustic tests. You might think of parametric coding in the philosophical term of empirism, extensively discussed on such level by John Locke.

There is a beautiful criticism of empirism by Immanuel Kant, who crafted the idea of *a priori* notions ("*a priori* Anschauungen oder Begriffe") in his "Critik der reinen Vernunft". Kant evidently has the merit to have discovered - on philosophical premises - the "endogenous and exogenous" invariant principles inside our cognitive system, i.e. our brain [8].

In audio coding, the general paradigm is to craft systems with highest adaption to the trained perceptual recognition patterns, which are not available to the newborn infant.

Anecdotally, at IBC 2015, I had the rare opportunity to talk to two people who were most deeply interested in ECMA-407 - due to their unfortunate position to have complete hearing loss on ONE ear only. Whilst the lady, who most willingly shared her perceptual capabilities for the sake of science, was neurally deaf since birth on her left ear, the other person was a professional sound engineer who lost hearing on his right ear due a blast occurring caused by headphones of a well-known manufacturer (which he unfortunately did not report under the premises of liability damages in due time).

Our sound engineer with hearing damage on one ear immediately caught my attention - because, when listening with headphones, he swapped left and right outputs after a certain time, which would lead to erroneous results with a person with normal hearing! This remarkable person, contrarily, activates his excellently trained brain and tries to guess the true spatial result by supplementary intellectual analysis.

► The highly intellectual, neurally deaf lady localises sound, when closing her eye, only with respect to DISTANCE, as level cues are preserved. However, her brain not being able to interpret the anatomy of her head (the step of Theile's "Localization" missing, see *Figure 1*), she CANNOT perceive localization without supplementary visual cues.

Both visitors at IBC 2015 evidently interact on a different level with the invariant principle, which evidently shapes our notion and can be recognized "endogenously" in an idealized and most elegant way in Kant's rationalistic philosophy of "*a priori* Anschauungen oder Begriffe".

As an "exogenous" invariant example, according to Kant, time is perceived by an infinite line, representing a series occurring simultaneously (representing infinity) or subsequently (representing a given time interval). Evidently the invariant is the series, which Kant perceives as outward, hence trained, notion, which in neurology is the infant engram.

Mathematics is functioning in a similar way. In a Gaussian, hence random, signal a multitude of mathematical objects occur at given times randomly, e.g. in terms of topology, analysis and algebra. Most of them are useless, because they cannot be observed CONTEXTUALLY. They are like beautiful unique flowers in our beautiful Swiss mountainside near St. Moritz coming and going. The only known objects, which may be observed, regardless ANY contextual notion, are algebraic invariants, ingeniously described as an algebraic field for the first time by David Hilbert in 1893 [9]. A field means that algebraic invariants form a closed system, which allows algebraic manipulations in a GENERALIZED context.

All of a sudden, probability has disappeared and instead a hidden context is revealed in a given data set, which can be easily compared to another data set. The foremost advantage, however, is that the context is not due to intrinsic knowledge but due to ABSTRACT behaviour - the very approach Kant took in crafting his philosophy.

This is why invariants in audio at the same time, for instance, describe "Space" AND "Localization" in *Figure 1*, as a synonymous mathematical "engram". Parametric coding, contrarily, requires TWO models, each with endless model extractions and extensive databases.

The final question always is what gives the better result: empirical or rational solutions. For this very reason, I ran a statistical ECMA-407 encoder against an invariant-driven one. And magic indeed happened: though only taking one frame of shortest length, the invariant encoder gave a PRECISE coding result in real time, whilst the statistical encoder had to run eight minutes data and remained with an OK coding result.

If multimedia engineering ever should discover invariants, our world will change!

Why did this not happen earlier?

I wish to pay my homage to the very person who averted me to the problem of invariant isolation with Gaussian processes, which led to my humble results by applying the apolarity principle:

When Rudolf E. Kálmán discovered the fundamental

solution to the Wiener problem, where "the objective is to obtain the specification of a linear dynamic system (Wiener filter), which accomplishes the prediction, separation, or detection of a random signal", no computers were available to a large public, for which Kálmán's results would have been fit – though Kálmán precisely described the computational side to the Wiener problem from a highly modern perspective [1]. Kálmán made his most eminent tribute to science in the "Transactions of the ASME-Journal of Basic Engineering" in 1960. Only NASA, having appropriate computer infrastructure indeed at hand, discovered the importance of this jewel for its Apollo program, developed by Stanley F. Schmidt e.a.

In a world of utilitarianism, without application science seems to be idle. Visiting Rudolf E. Kálmán and his lovely wife together with my partner Melanie Angélique Grümmer a few days ago, professor Kálmán and I discussed sarcastically "applied fundamental science", as propagated today by universities. The whole article recalls this unique conversation.

Fundamental science is rational, applied sciences are empirical. So what is the in-between?

I justify my existence as a rationalist by thought, and CONSEQUENTLY by empiric results. Industry would be well advised to encourage fundamental research. However, when it comes to the question to replace candles by Edison's bulbs, the candle light makers indeed are not amused. Hence their preference for applied sciences, where revolution remains intrinsically limited and consequently under full economic control. Mental revolutions are cheap and everlasting outside a gimmick industry - illegally shedding its electronic waste all over Africa.

There is no such silent voice on earth as is the voice of reason.

For more information please visit:

www.swissaudec.com

1. R. E. Kálmán. *A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME - Journal of Basic Engineering*, 82, 1960.
2. T. Bayes, R. Price. *An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. In Phil. Trans.* 53, pp. 370-418, 1763.
3. K. Pearson. *On Lines and Planes of Closest Fit to Systems of Points in Space. In Philosophical Magazine* 2, pp. 559-572, 1901.
4. G. Theile. *On the Localisation in the Superimposed Soundfield. Technische Universität Berlin*, 1980.
5. G. Stoll, G. Theile, M. Link. *Low Bit-rate Coding of High-quality Audio Signals. AES Preprint* 2432, 3/1987.
6. R. Arnheim. *Gestalt Psychology and Artistic Form. In L. L. Whyte (ed.). Aspects of Form. A Symposium on Form in Nature and Art*, pp. 196-208. 2nd ed. London, 1968.
7. E. Pavlopoulos, P. Trifilieff, V. Chevalyere, L. Fioriti, S. Zairis, A. Pagano, G. Malleret, E. R. Kandel. *Neutralized1 activates CPEB3: a function for nonproteolytic ubiquitin in synaptic plasticity and memory storage. Elsevier*, 2011.
8. I. Kant. *Critik der reinen Vernunft. Riga*, 1781.
9. D. Hilbert. *Über die vollen Invariantensysteme. Mathematische Annalen Bd. 42*, 1893.