Victor Steinberg

Video Standards

Signals, Formats and Interfaces

Part 9

Audio & Video, Tight Friendship

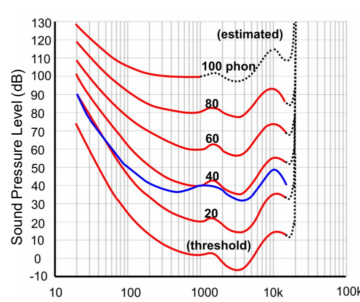
www.videoq.com

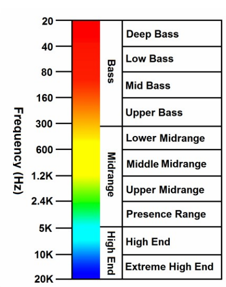
Background Information

The word “video” in Latin means “I see”; the word “audio” means “I hear”.
Picture has its beauty, but only sound can spell out a story. In other words, video image requires **accompanying sound**.
Many years ago the honorable members of French ***Académie des sciences*** unanimously decided to **ignore** any propositions on **3 subjects**:
• perpetuum mobile (*perpetual motion*),
• flying apparatus heavier than air,
• and … *method of sound conservation and reproduction*.
So, sound recording and transmission technologies are much older than video technologies. In a way, they are also more advanced, e.g. audio scientists and audio engineers used **dual logarithmic scale** long before modern UHDTV HDR video histograms.

The **equal loudness curves** diagram illustrates:

- Very large (≈ 100 dB, i.e. 100,000 times on a linear scale) range of **Sound Pressure Levels** (and **phon** units), perceived by the human ear: from the **hearing threshold** at the bottom to the **pain limit** at the top
- **Comfortable Loudness Range** of voice and music is in the middle of the diagram, normal conversation level is 40 dB ~ 60 dB, and, as we all know, this range depends on the **ambient noise** conditions
- The **Audible Frequency Range** is about 20 Hz ~ 20 kHz, and this range is fully covered by modern AV (Audio-Video) production tools
- **Maximum Sensitivity Range** (aka Midrange) is about 350 Hz ~ 5 kHz





Audio Signal Formats & Interfaces

Analogue audio professional interfaces use mostly **XLR** connectors and **shielded twisted pair cables**.
The line impedance varies from 110 Ohm to 600 Ohm and there is an issue of balanced ⇔ unbalanced line type transformation.

Digital audio format **AES3** was jointly developed by the Audio Engineering Society (**AES**) and the European Broadcasting Union (**EBU**).
AES3 signal can carry **two** PCM (uncompressed) channels, audio data up to 24 bit @192 kHz. The AES/EBU standard was first published in 1985.
AES3 has been incorporated into the International standard **IEC 60958**, and is also available in a consumer-grade variant known as SPDIF.

Digital AES3 signal can be transmitted via various physical interfaces:

• **Balanced AES3**: IEC 60958 Type I 3-pin **XLR**, **RCA** (aka "phono"), **CAT 5 STP**, or multi-channel **D25** type connector (*Warning: various pin-outs),* 110 Ohm impedance, 2~7 Vpp at source, max distance 180~500 m,

• **Coaxial AES3-id**: 75 Ohm **BNC** connector, **compatible with digital video** infrastructure, 1 Vpp signal, max distance 500 m,

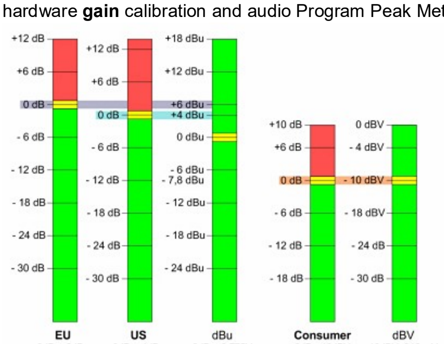
• **Optical TOSLINK** (from **Tos**hiba **Link**), mostly for consumer applications, the effective range of plastic optical cables is limited to 5~10 m,

• **Coaxial SPDIF** (Sony Philips Digital Interconnect Format), RCA or BNC connectors.

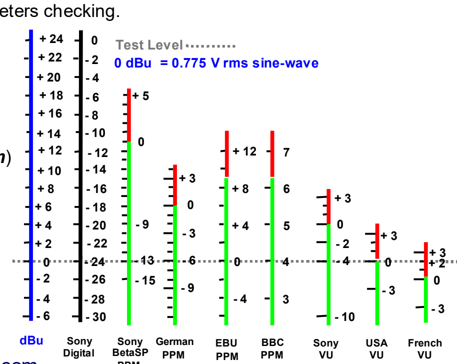
AES3 is the preferred interface format for professional applications, there is a noticeable trend towards the **AES3-id** format via **BNC** connectors.

Analog Audio Levels: dB, dBu & dBV Scales

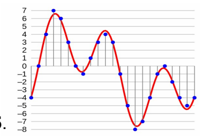
The **dB** value by definition is a **ratio** of any **two** values converted to the logarithmic scale,
i.e. the dB value means something only in combination with some explicit or implicit (assumed) **reference value**.
The **dBu** is a dB value of any analog audio **rms** (*root mean square*) **voltage** referenced to the **nominal analog rms voltage 0.775 V**,
thus 0 dBu means 0.775 V rms (**1 mW** power on **600 Ohm** load). For a **sinusoidal** signal 0 dBu also designates **2.19 Vpp** (*peak-to peak*).
On **balanced** analog line the **differential** voltage can go up to +24 dBu, max Vpp = 0.775 x (1024/20) x 2 = 24.57 V.
For the **unbalanced** analog line the audio signal **headroom** is much smaller, voltage can go up to +4 dBu, max Vpp = 1.23 V.
The seldom used **dBV** unit is similar to dBu, the only difference is the reference level of **1 V rms**.
There is no audio equivalent of Reference White video level. Instead audio engineers use "**test**" or "**alignment**" level, which serves **only** for the hardware **gain** calibration and audio Program Peak Meters (PPM) or Volume Unit (VU) Meters checking.



These two diagrams illustrate
the complete chaos
(*you may call it **creative freedom***)
ruling in the analog audio realm.



Analog ⇔ Digital, dBu ⇔ dBFS Levels Mapping

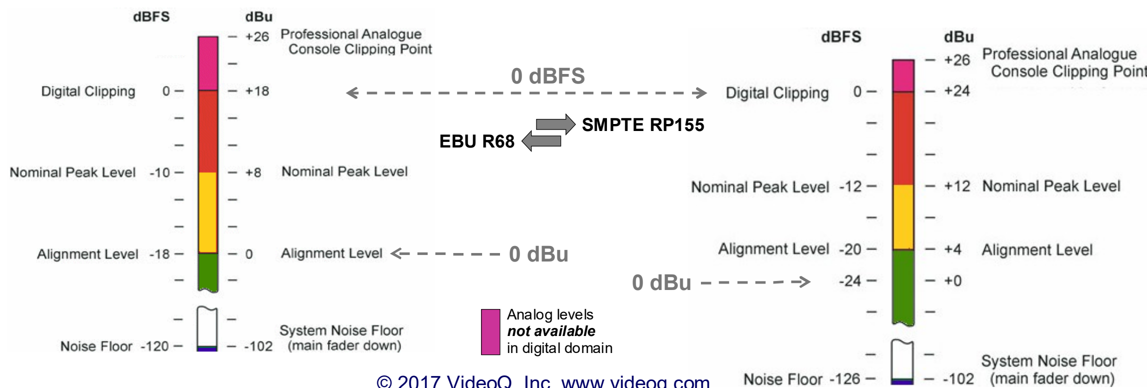


Decibels relative to full scale **dBFS** is a unit of measurement for amplitude levels in digital systems,
which have a **defined maximum peak level**, e.g. for 24-bit digital audio the max magnitude value is 2^{23} - 1.
Unlike the digital video, there no standard scheme for mapping the analog audio levels to digital audio levels.

Currently, there are **two** standards mapping 0 dBu level to the **different points** on dBFS scale: **EBU R68** and **SMPTE RP155**.

EBU R68, used in many countries, is mapping the alignment level of **0 dBu** to **-18 dBFS**.

SMPTE RP155 is mapping the US installations alignment level of **+4 dBu** to **-20 dBFS**, so **0 dBu** is mapped to **-24 dBFS** (**6 dB lower** than EBU!).

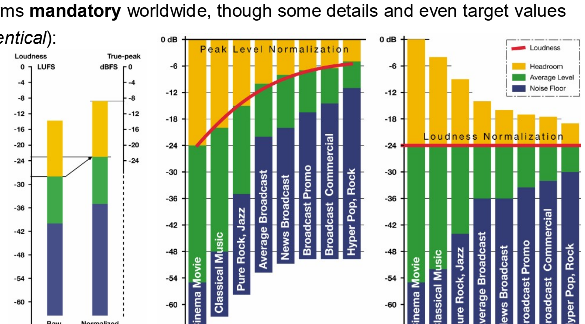


Integrated Loudness, Loudness Range, True Peak

In acoustics, loudness is the **subjective perception** of sound pressure. With application to AV content the most important factor is **relative** loudness of different **segments** of **rendered** sounds **timeline**. Loudness control & monitoring is needed in production, post production and distribution of AV content.

Widely used **Peak Programme Meter** (PPM) is described in **IEC 60268-10**. **ITU BS.645** recommends three PPM reference levels: **Measurement Level** (ML), **Alignment Level** (AL) and **Permitted Maximum Level** (PML).
Traditional methods, such as the PPM and VU metering have been combined with (*not replaced by*) the measurement of:
Momentary (**ML**, 400 ms), **Short Term** (**SL**, 3 s), **Integrated** (**IL**, from start to stop) **Loudness** in LUFS/LKFS units, plus **True-peak** (**TP**) in dBTP.
The measurement methodology was set by **EBU R128** "Loudness normalisation and permitted maximum of audio signals" (*see also supplementing R128s1 and EBU Tech 3341~ 3344*) and **ITU-R BS.1770** "Algorithms to measure audio programme loudness and true-peak audio level" standards.
Later, national standards in many countries made compliance with these norms **mandatory** worldwide, though some details and even target values may be slightly different (*since ITU BS.1770-2, LKFS and LUFS units are identical*):

- R128 ⇒ **IL**: **-23 LUFS ± 0.5 LU**, **Max TP: -1 dBTP, Max SL: -18 LUFS**
- USA ATSC A/85 ⇒ **IL: -24 LKFS ± 1 LU, Max TP**: **-2 dBTP**



Diagrams – courtesy of Thomas Lund, TC Electronic

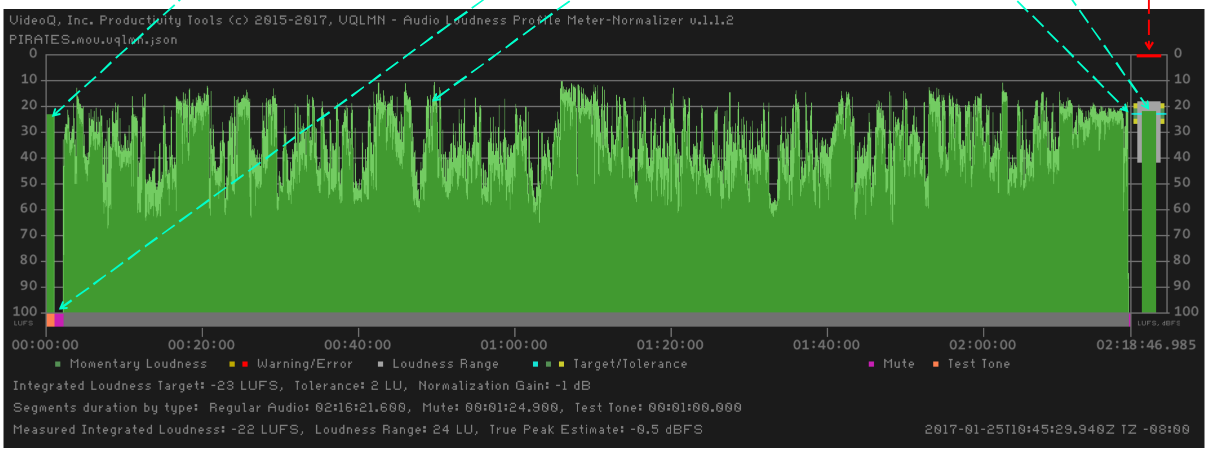
Loudness Profiling – VideoQ Media Ambit ™

The **Momentary Loudness Timeline Profile** looks good, the **Integrated Loudness** value is within specs, but **two segments** of the **Profile** and **one parameter** (TP level) are out of specs:
60s long -23 dBFS Test Tone at timeline start, then 60s long Mute, then Regular Audio, and **True Peak** value is too high.

Target IL: -23 LUFS

Actual IL: -22.0 LUFS

True-peak: -0.5 dBTP

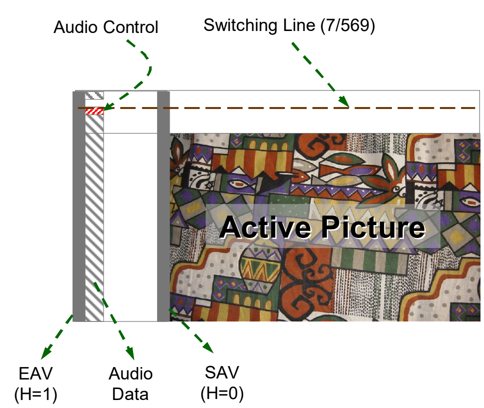


AES3, AES10, Embedded Audio

AES10 aka Multi-channel Audio Digital Interface (MADI) is an AES3 compatible standard interface that carries **up to 64 channels** of digital audio.
Up to 16 AES3 channels (stereo pairs) can be transmitted together with the video stream **via SDI interface** as defined by **ITU-R BT.1120**.
Audio data **embedding** and **de-embedding** provide for more flexibility in joint, separate or hybrid AV processing workflow optimization.
Big advantage of the embedded audio is the **automatic compensation of video delay**; audio data latency is always the same as video latency.

Channels **1 to 4**	Channels **5 to 8**	Channels **9 to 12**	Channels **13 to 16**
↓	↓	↓	↓
Group 1	Group 2	Group 3	Group 4

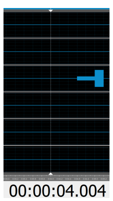
- Audio is embedded **in groups**, e.g. one group contains **2.0 x 4**, **5.1** or **7.1** signals, all channels in a group have the same **sample rate**, **phase** and **synchronicity**
- Audio is carried in the **horizontal blanking** period
- Audio data are carried in the **Cb/Cr** data stream
- Audio control packet is carried in the **Y** data stream (once per field, 2 lines after switch point)
- Audio is inserted immediately after the line CRC words
- No audio is inserted on the line after the switch point



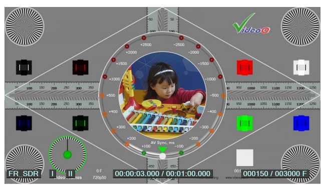
Audio ⇔ Video Synchronization

Well known **AV Sync** (aka Lip Sync) problems are often caused by the video & audio data processing in two separate workflows.
Also, significant (and even variable) video & audio delays may be added by video synchronizers and compression codecs.
The **EBU Recommendation R37** “The relative timing of the sound and vision components of a television signal” states that end-to-end audio/video sync should be within **-40 ms** and **+60 ms** (audio before / after video, respectively) and that each stage should be within **-5 ms** and **+15 ms**.
The **ATSC** recommends **-15 ms** and **+45 ms**.
ITU reported the thresholds for AV Sync error detectability as **-45 ms** and **+125 ms**.
For film production, acceptable lip sync error limit is considered to be **± 22 ms**.
There are two quite different types of AV errors measurements:

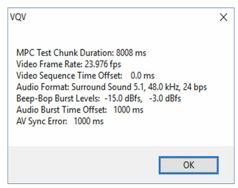
- Measurement of the **A & V propagation delays** (latencies) within the AV processing workflow; the difference between audio latency and video latency is the processing chain AV Sync error. Such measurement usually requires artificial test signals, e.g. audio burst (beep) plus video flash.
- Measurement or assessment of actual **AV content**, i.e. **correlating** the rendered sound with lips/mouth image activity.

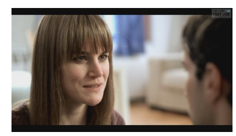


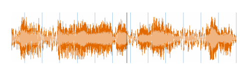
+



=







About This Presentation

Produced by	**Josef Marc**
Written by	**Victor Steinberg, PhD**
Narrated by	**Josef Marc**
Conceived by	**Roderick Snell**
Technical consulting by	**Maxim Levkov**

Based on the book
"Video Standards: Signals, Formats and Interfaces" by Victor Steinberg
Published by Snell & Wilcox

For further reading we recommend wikipedia.org

About VideoQ

Company History

- Founded in 2005
- Formed by an Engineering Awards winning team sharing between them decades of global video technology.
- VideoQ is a renowned player in calibration and benchmarking of video processors, transcoders and displays, providing tools and technologies instantly revealing artifacts, problems and deficiencies, thus raising the bar in productivity and video quality experience.
- VideoQ products and services cover all aspects of video processing and quality assurance - from visual picture quality estimation and quality control to fully automated processing, utilizing advanced VideoQ algorithms and robotic video quality analyzers, including latest UHD and HDR developments.

Operations

- Headquarters in Sunnyvale, CA, USA
- Software developers in Silicon Valley and worldwide
- Distributors and partners in several countries
- Sales & support offices in USA, UK