**3GPP TSG-WG SA4 Meeting #118-e *S4-220468r02***

**, 6-14 April 2022**

|  |
| --- |
| *CR-Form-v12.2* |
| **Pseudo CHANGE REQUEST** |
|  |
|  | **26.805** | **CR** |  | **rev** | **-** | **Current version:** | **1.0.1** |  |
|  |
| *For* [***HE******LP***](http://www.3gpp.org/3G_Specs/CRs.htm#_blank)*on using this form: comprehensive instructions can be found at* [*http://www.3gpp.org/Change-Requests*](http://www.3gpp.org/Change-Requests)*.* |
|  |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| ***Proposed change affects:*** | UICC apps |  | ME | **X** | Radio Access Network |  | Core Network | **X** |

|  |
| --- |
|  |
| ***Title:***  | [FS\_NPN4AVProd]: KI6 Interfacing Audio Channels |
|  |  |
| ***Source to WG:*** | Sennheiser, Ericsson LM |
| ***Source to TSG:*** | S4 |
|  |  |
| ***Work item code:*** | FS\_NPN4AVProd |  | ***Date:*** | 31/03/2022 |
|  |  |  |  |  |
| ***Category:*** | B |  | ***Release:*** | Rel-17 |
|  | *Use one of the following categories:****F*** *(correction)****A*** *(mirror corresponding to a change in an earlier release)****B*** *(addition of feature),* ***C*** *(functional modification of feature)****D*** *(editorial modification)*Detailed explanations of the above categories canbe found in 3GPP [TR 21.900](http://www.3gpp.org/ftp/Specs/html-info/21900.htm). | *Use one of the following releases:Rel-8 (Release 8)Rel-9 (Release 9)Rel-10 (Release 10)Rel-11 (Release 11)…Rel-16 (Release 16)Rel-17 (Release 17)Rel-18 (Release 18)Rel-19 (Release 19)* |
|  |  |
| ***Reason for change:*** |  |
|  |  |
| ***Summary of change:*** | The Key issues #6 is extended with additional information around existing audio solutions. |
|  |  |
| ***Consequences if not approved:*** |  |
|  |  |
| ***Clauses affected:*** |  |
|  |  |
|  | **Y** | **N** |  |  |
| ***Other specs*** |  | **X** |  Other core specifications  | TS/TR ... CR ...  |
| ***affected:*** |  | **X** |  Test specifications | TS/TR ... CR ...  |
| ***(show related CRs)*** |  | **X** |  O&M Specifications | TS/TR ... CR ...  |
|  |  |
| ***Other comments:*** |  |
|  |  |
| ***This CR's revision history:*** |  |

\*\*\*\* First Change \*\*\*\*

## 6.7 Key Issue #6: Interfacing Audio Channels

Editor’s Note: This clause should describe implications on protocol usage, when a predefined number of audio channels (as in MADI or SDI) is allocated, independently on its needs. In SDI, always 32 audio channels are allocated. Unused audio channels are “muted”. See ST 299 for more details. (https://tech.ebu.ch/docs/techreports/tr002.pdf)

* 1. Are muted audio channels used for other purposes in SDI / MADI, which should be considered for 5G deployments?
	In MADI the channel have an active flag. If the channel is inactive there is no need to transfer the data within the 5G system, but it should be ready to transport the audio stream without noticeable delay.
	2. Is it needed to send audio frames with “many null payload bytes“? What is the practice in ST 2110, which also supports separated A & V?
	In ST 2110-30 the audio is AES67 encoded. The payload is given by the SDP description but can also be derived from the packet stream itself. Inactive channels with null bytes are not allocated in payload. So packet size depends on channels within payload.
	3. Would all audio channel perceive same quality/QoS? Or can some audio channels require low latency while other audio channels are “embedded with video”?
	Since each audio channel stream is announced via SDP in AES67, one could create multiple streams with different quality levels. Dynamically changing the quality level during a production is usually not done. Different quality levels should co-exists in parallel but separated within the streams.

Editor’s Note: This clause should describe the possibility of configuring audio channels on a need basis.

### 6.7.1 Circuit-switched audio transmission

The Multiple Audio Digital Interface (MADI) [38] and the Serial Digital Interface (SDI) [35][36] embed audio channels together with video channels onto the same physical medium. Multiple Audio Digital Interface (MADI) [38] supporting [serial digital transmission](https://en.wikipedia.org/wiki/Serial_transmission) over [coaxial cable](https://en.wikipedia.org/wiki/Coaxial_cable) or optical [fibre](https://en.wikipedia.org/wiki/Fibre-optic) lines of 28, 56, 32, or 64 channels; and [sampling rates](https://en.wikipedia.org/wiki/Sampling_rate) to 96 kHz and beyond with an [audio bit depth](https://en.wikipedia.org/wiki/Audio_bit_depth) of up to 24 bits per channel. Where encapsulated audio and video are used then fewer channels are likely to be deployed. As a minimum, this should consist of two audio channels.

5G System resources are shared among devices and radio resources should preferably not be allocated and left idle. This key issue should study, how in particular audio channels are allocated in existing media productions and how 5G based media productions can interwork with existing media productions, when a more dynamic allocation of audio channels is used on 5G Systems.

Audio may be carried as an encapsulated signal multiplexed with video and data, or as a separate set of streams. For tier one or audio-only productions, the audio is treated as separate discrete streams per channel. For tier two and three productions and contribution workflows, it may be desirable to carry audio and video multiplexed with the video.

A channel is usually a mono signal. An audio channel can be considered as:

- *Active* or *inactive:* Not all channels (allocated in MADI or SDI) may be required for all applications so it should be possible to describe a channel as either active or inactive so as to make more efficient use of available bandwidth.

- *Muted* or *unmuted*: An active channel may be temporary muted where it may be required but the UE is not transmitting any data.

- *Silent:* A silent channel is active and unmuted but with a low-level audio signal. This may be used to provide atomospherhic or spot effects.

In MADI [38] and SDI [35][36] there is a mode bit to indicate whether a channel is active or not.

The first four bits of a MADI subframe are used for frame synchronization and to indicate the state. If a channel is set to inactive, all payload data must be set to zero. Thus, an individual audio channel cannot be taken out of the multiplex compared to a packet-based transmission approach.

Also, MADI is synchronized in itself: it does not lock to the audio sampling frequency. This indicates an additional need for audio clock synchronicity via a word clock.

In audio production networks a word clock is a dedicated (physical) distributed timing signal. It is the one master clock providing the sampling frequency for all audio processing devices.

Editor’s Note: It should be checked, whether there is a DVB or SMPTE threshold definition for “silence”.

Communication channels are usually speech-only and of a lower quality than main programme audio but do require low-latency solutions. There is also a requirement for one-to-many solutions so that a director can speak to multiple end users at the same time.

SDI (Serial Digital Interface) [35][36] is a family of standards widely used in the media production domain to transport uncompressed video signals. Various SDI interface (SD-SDI, HD-SDI, 3G-SDI, 6G-SDI, 12G-SDI and 24G-SDI) are available to support from standard definition up to UHD video resolutions.

SDI can carry also embedded audio.

3G-SDI, known as the 3Gbit/s interface, defined different mapping levels (A, B-DL, B-DS) for the carriage of 1080-line image formats and associated ancillary data. With respect to the audio, 3G-SDI may contain up to 16 audio channels or 32 if dual-link applications are considered or SMPTE ST 299-2 is used.

NOTE: 3G-SDI and later supports 32 channels but in practice it is limited to 16 channels as it is rare to find products that support more than 16 channels. In fact many products only support 8 channels. For transporting higher than 16 channels of Higher Order Ambisonics (HOA) over SDI, refer to the HOA Transport Format (HTF) as defined in ETSI TS 103 589 [81].

Each audio channel in SDI (and in MADI) is AES3-encoded, removing the need for data conversion between both interfaces. Since the audio is data embedded in the ancillary data, the synchronisation is relative to the video signal. The timing information as well as additional metadata (i.e. sample rate, bit depth, channel status) are transmitted in additional audio control packets.

In Tier one scenarios, in general, the audio signals come from the microphones installed in the studio/location (and not from the cameras) while in Tier two and Tier three productions, especially for contribution links, embedded audio is transmitted multiplexed with the video. When the audio is embedded, MPEG-2 Transport Stream might be used over RTP/UDP/IP instead of native RTP carriage. For ST 2110-30 scenarios, six conformance levels are defined [40]. Level A is the only mandatory conformance level to be supported by all compliant equipment and is defined as follows:

- Linear 24-bit PCM encoding.

- 48 kHz sampling frequency (media clock).

- 1 to 8 channels per stream.

- 1 ms packet time (48 audio samples per channel in each packet).

### 6.7.2 Packet-based audio channels

#### 6.7.2.1 Introduction

Audio protocols have evolved towards packet-based technologies to be compliant with standard technologies and infrastructure such as Ethernet and IP. Various protocols have been developed in the industry. Depending on the operating layer one can distinguish between *audio-over-Ethernet* at Layer 2 of the protocol stack (clause 6.7.2.2) and *audio-over-IP* at Layer 3 (clause 6.7.2.3).

#### 6.7.2.2 Audio-over-Ethernet

Layer-2 systems with a proprietary data link protocol. Sometimes compliant with Ethernet framing to utilize existing infrastructure and components, thereby reducing cost. Operated within different topologies, mostly star or tree.

Examples: CobraNet, EtherSound, Soundgrid.

#### 6.7.2.3 Audio-over-IP

Layer-3 protocols based on IP. Often using derivates of UDP at Layer 4 with additional supporting proprietary protocols for solving device discovery, synchronisation, management and configuration, etc.

Examples: DANTE (Digital Audio Networking Through Ethernet), WheatNet, Livewire, Q-LAN, Ravenna.

While DANTE is the most prominent example, native synchronisation depends on underlying PTP-v1, while PTP-v2 support was added later [82]. Payload definition varies per protocol, and so different solution are thus not compatible with each other. Routing and network pathing via the IP layer enables coexistence with other IP-based services, but may depend on managed infrastructure to guarantee QoS.

Interoperability was later introduced via standardization of AES67 [57] as the common denominator of all packet-based audio transport technologies. AES67 defines a profile that all audio-over-IP solutions must provide to be compliant:

- Synchronization via PTPv2, as specified in IEEE 1588-2008 [22].

- 48 kHz audio sampling rate.

- 16- or 24-bit linear PCM encoding ("L16/L24").

- 1 to 8 channels per stream.

- 48 samples per packet.

- IP/UDP/RTP header with unicast and multicast support.

- DiffServ for network QoS using ToS labelling in the IP packet header.

- SDP for announcing sessions.

AES67 is thus compliant with "level A" in ST 2110-30 [3].

### 6.7.3 Solutions

### 6.7.4 Discussion

\*\*\*\* Next Change \*\*\*\*

## 2 References

[81] ETSI TS 103 589: "Higher Order Ambisonics (HOA) Transport Format".

[82] Offical Audinate DANTE FAQ: <https://www.audinate.com/learning/faqs/clocking>

\*\*\*\* Next Change \*\*\*\*

## 3.3 Abbreviations

HOA Higher Order Ambisonics

HTF HOA Transport Format

\*\*\*\* Last Change \*\*\*\*