Bruce Devlin

Standards for Audio

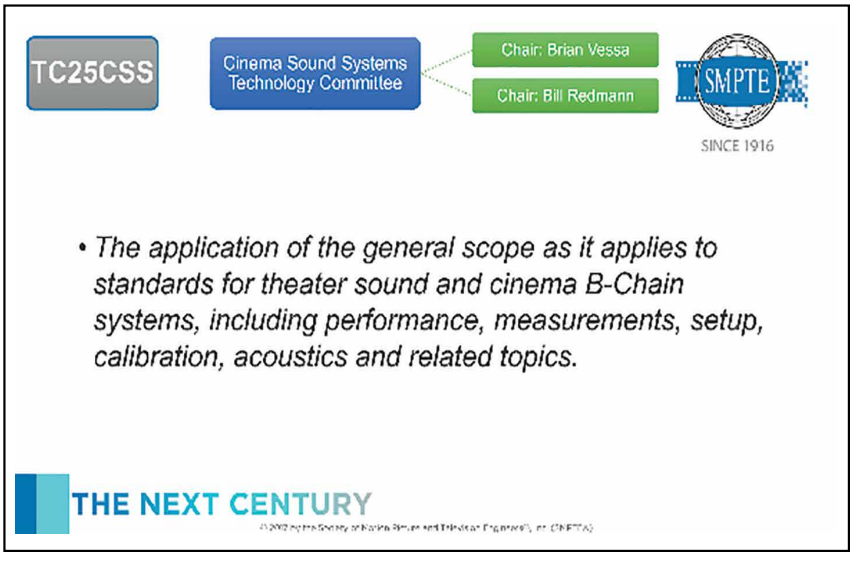
"Don't forget about the audio!" was probably the second pearl of wisdom given to me at the BBC, when I started my career in Broadcast way back in time. The first pearl of wisdom was, "*Before you pick up a heavy, expensive, fragile monitor; make sure that you have somewhere to put it down*." Even though both the comments were delivered by my boss while he was walking from the room chuckling with importance, they both stuck because he was right.

In the SMPTE standards world, we have most definitely not forgotten about the audio. Projects ranging from Immersive Audio Bitstreams to the seemingly mundane labeling of audio components cover both the live carriage of audio over IP networks to the storage of audio for mastering and distribution. The media world is becoming more integrated, automated, and systems orientated, which means that SMPTE cannot act in isolation; so it is reassuring to know that the Society has very close ties with user communities and audio expert groups such as the Audio Engineering Society (AES).

Within the SMPTE Standards Community, there is a group dedicated to cinema sound, TC25CSS. Two years ago, SMPTE published the first-ever standard for a digital wideband pink noise signal to be used for sound system calibration (ST 2095-1), complete with supplied 48K and 96K pink noise audio files and pseudocode for generating this pink noise in a device. In 2017, SMPTE published the first-ever recommended practice for calibrating a cinema sound system (ST 2096-1) and an RP for maintaining this calibration (ST 2096-2). Following on from these successes, we are in the process of planning new work for NextGen cinema sound systems.

TC25CSS has a working group dedicated to the interoperability of immersive sound systems in Digital Cinema. Immersive audio is unique in that it has the added dimension of height. The work covers the areas of D-Cinema that require standardization to achieve interoperability. The group is creating a number of engineering documents, including standardizing a single object-based distribution file format and related protocols for interoperable playback into a variety of theatrical speaker configurations. This is important for several reasons: a single bitstream format maximizes the potential for the standard to be adopted around the world, increases the chances of interoperability between equipment and software vendors, and simplifies mixing and distribution. The group is very close to completing its draft of the Immersive Audio Bitstream specification ST 2098-2.

Creation of the bitstream is not enough to achieve interoperability. There needs to be an Immersive Audio Metadata specification that provides key definitions of the parameters that are integral to immersive audio. Immersive Audio Channels and Soundfield Groups need to be defined and labeled for easy interchange. Finally, the



Digital Object Identifier 10.5594/JMI.2018.2795440
Date of publication: 27 February 2018

bitstream needs to be decoded and rendered into various speaker configurations, which vary by sound system and theater. To do this effectively, renderer expectations need to be specified for the behavior of the renderer and the expected end results, depending on the actual speaker layout available. Specific tests need to be developed to ensure that immersive audio renderers meet these expectations. All of this work is covered by TC25CSS and will be available when complete in the SMPTE 2098 document suite.

TC21DC is following on to these 25CSS standards and is creating standards for the packaging and playback of immersive audio. Specification of a new Immersive Audio Track File is under way, as are new constraints created to accommodate immersive audio Digital Cinema Packages. Synchronization and data communication in immersive sound systems are being standardized so different pieces of equipment can talk and work together.

The result of these combined efforts will allow for the interoperable creation and playback of immersive audio using any toolset, and immersive sound system that complies with the standards, which is what D-Cinema is all about!

While the new and exciting ways of delivering audio are easy headline grabbers, we shouldn't forget about some of the more mundane but commercially significant areas of change. In 2018, we are still in the midst of migration toward a multiplatform distribution and publishing model of moving content to the consumer. It might be surprising to many that something as simple as using a label to say "this audio is French" is *not* in widespread use today. Historically, this might be blamed on working practices inherited from videotape with only four audio channels and no metadata. It might be blamed on the lack of standards, but the reality is that there should be a compelling commercial reason to spend money upstream inserting the metadata so that the benefits of using that metadata downstream can be realized.

My personal belief is that system-level standards like SMPTE's IMF (ST 2067) tip the balance in that economic equation. Within that standard, there are no conventional channel numbers or layout. The playback of audio from a stereo pair or 5.1 or 7.1 group is controlled entirely with metadata and part of that metadata covers language and usage. A new standard is under way to specify the carriage of immersive audio in IMF, which is based on the cinema work noted above.

The need to create much of the metadata is well understood and agreed, such as "L and R" for left and right channels. New definitions are in progress to describe Immersive Audio and speaker layouts, and many other definitions can be inherited from Internet standards such as RFC 5646 for language codes. Note that in this technical metadata, we use the English Language terms and it is up to a vendor to translate the terms for display in a local language.

Not all of this metadata has an existing common vocabulary, so SMPTE's 21DC, 25CSS, 30MR, and 35PM committees all contribute to work with their own special skills. This work highlights the differences in terminology and working practice from around the world. For example, the Audio Metadata field *ContentKind* indicates the use of a particular audio mix. Terms like *Primary* and *Voice Over* might be widespread, but not terms to describe access services. In the U.S., the term *Descriptive Video Service* is used to describe a service designed for visually impaired audiences that combine primary program content with narration content, describing the selected action in the picture. In the U.K., the narrative for the visually impaired is delivered as a part of the main service and is described as *Audio Description*. Although similar, these two terms describe different services. It's up to the great work of the technology committees to figure out when to merge terms, add terms, duplicate terms, or ignore terms and I wish them the best of luck in this area, where there are no clear right or wrong answers.

SMPTE has also recently created standards for audio mapping and carriage in MXF and SDI (for example, ST 381-4) and is working closely with AES on the carriage of AES3 and AES67 using the ST 2059 protocol.

The world of audio is very much alive and well and pervasive throughout SMPTE's work. I am personally looking forward to the next couple of years of developments and am grateful to Brian Vessa, whose in-depth knowledge in this area contributed to this article.

UPCOMING TECHNOLOGY COMMITTEE (TC) MEETINGS

12–15 March
SMPTE HQ,
White Plains, NY, USA

18–21 June
Ryerson University
Toronto, Ontario, Canada

19–22 September
EBU
Geneva, Switzerland (following IBC)

3–6 December
Dolby Laboratories,
San Francisco, CA, USA