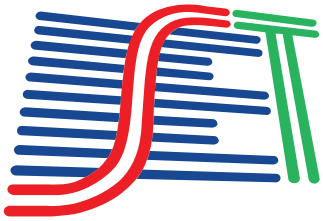
SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING

SET IJBE V. 9, 2023

ISSN print: 2446-9246
ISSN online: 2446-9432



SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING

SET IJBE V. 9, 2023

ISSN print: 2446-9246
ISSN online: 2446-9432

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING

ISSN PRINT: 2446-9246 | ISSN ONLINE: 2446-9432

International Cataloging Data in the Publication - CIP - Librarian Zoraide Gasparini CRB/9 1529

S517 SET International Journal of Broadcast Engineering —— vol.9 , (Dec. 2023). – São Paulo: Brazilian Society of Television Engineering —— SET,2023.

Annual frequency
ISSN Print 2446- 9246
ISSN online 2446- 9432
Available at: https://set.org.br/ijbe

Broadcasting – Periodic. 2. Transmission Engineering. 3. Digital TV. I. SET. II. Title.

CDD: 384.54

ANNUAL | CIRCULATION : 500 EXEMPLARY | COMMUNICATION MANAGER: TITO LIBERATO | GRAPHIC DESIGN AND DIAGRAM: SOLANGE LORENZO

EDITORIAL BOARD

CORPORATE AUTHOR AND EDITOR

SET – Brazilian Society of Television Engineering, or, in Portuguese, SET – Sociedade Brasileira de Engenharia de Televisão.

The SET, founded on March 25, 1988, is a not-for-profit technical-scientific association of engineering, technology, operations, research professionals, educational institutions and companies, that aims the dissemination of technical, operational and scientific knowledge and the improvement of the technologies of electronic audio and video media. (Statute, Article 3).

Address: Av. Mario de Andrade, 252, suite 31 - Barra Funda District - São Paulo - SP - Brazil - Postal Code: 01156-001

SET BOARD OF DIRECTORS

Deliberative Council

2023 – 2024

President: Carlos Fini

Vice-President: Claudio Eduardo Younis

OFFICE HOLDER	**SUBSTITUTE**
Carlos Fini	Carlos Cauvilla
Luiz Bellarmino Polak Padilha Claudio	David Estevam de Britto
Eduardo Younis	José Carlos Aronchi de Souza
Claudio Alberto Borgo	Luis Otavio Marchezetti
Almir Antonio Rosa	Luis Camargo
Daniela Helena Machado e Souza	José Salustiano Fagundes de Souza
Vinicius Augusto da Silva Vasconcellos	Matheus de Andrade Silva
Raymundo Costa Pinto Barros Roberto	Marcelo Santos Wance de Souza
Dias Lima Franco	Valderez de Almeida Donzelli
Emerson Weirich	Paulo Henrique Corona Viveiros de Castro
Sergio Eduardo di Santoro Bruzetti	Nelson Faria
José Eduardo Marti Cappia	Marco Tulio Nascimento
José Raimundo Lima da Cunha Marcio	Esdras Miranda de Araujo
Rogério Herman	Israel de Moraes Guratti
Cristiano Akamine	Marcelo Moreno
Rafael Duzzi de Oliveira	Fabio Ferraz
Jurandir Moreira Pitsch	Wagner Kojo

Fiscal Council

Nivelle Daou	Rafael Silviera Leal
José Chaves F. de Oliveira	Sandro Sereno
Marcos Paulo Teixeira	

Council of Former Presidents

Adilson Pontes Malta	Liliana Nakonechnyj
Carlos Eduardo Oliveira Capelão	Olímpio José Franco
Fernando Mattoso Bittencourt Filho	Roberto Dias Lima Franco
José Munhoz	

Regional representatives

North: Henrique Camargo e Eduardo Lopes
North East: Ronald Almeida e Gabriel Eskenazi
Midwest: Wender de Souza
Southeast: Geraldo Mello e Flavio Menna Barreto
South: Caio Klein, Alisson Heidemman e Caue Franzon

ABOUT THE JOURNAL

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING

The SET IJBE, (SET International Journal of Broadcast Engineering) is an open access, peer-reviewed article--at-a-time international scientific journal whose objective is to cover knowledge about communications engineering in the field of broadcasting. The SET IJBE seeks the latest and most compelling research articles and state-of-the-art technologies.

Publishing schedule and schema

On-line version – Once an article is accepted and its final version approved by the Editorial Board, it will be published immediately on-line on a one article-at-a-time basis

Printed version – Once a year, all articles accepted and published on-line over the previous twelve months will be compiled for publication in a printed version.

Types of papers

Regular (Full) Papers: Traditional and original research [from 6 to 20 pages]

Tutorial Papers: Brand new OTT (Over-The-Top) detailed implementation and fully set up state-of-the art systems [4 – 6 pages]

Letters: Short notes and consideration about current and relevant techniques, technologies and implementations involving engineering solutions [1-3 pages]

Open Access Policy

If an article is accepted for publication, it will be made available to be read and re-used under a Creative Commons Attribution (CC-BY) license.

Editorial Office:

If you require any additional information, please contact the SET IBJE (SET International Journal of Broadcast Engineering) administration staff:

Address: Av. Mario de Andrade, 252, suite 31, Postal Code:01156-001, São Paulo – SP – Brazil; E-mail: IJBE@set.org.br

Aims and Scope include, but are not limited to:

Advanced audio technology and processing
Advanced display technologies
Advanced RF Modulation Technologies
Advanced technologies and systems for emerging broadcasting applications
Applying IT Networks in Broadcast Facilities
Broadcast spectrum issues – re-packing, sharing
Cable & Satellite interconnection with terrestrial broadcasters
Cellular broadcast technologies
Communication, Networking & Broadcasting
Content Delivery Networks – CDN
Digital radio and television systems: Terrestrial, Cable, Satellite, Internet, Wireless.
Electromagnetic compatibility issues between collocated services (e.g. broadcast and LTE)
General Topics for Engineers (Math, Science & Engineering)
Hybrid receiver technology
Interactive Technology for broadcast
IP Networks management and configuration
Metadata systems and management
Mobile DTV systems (all aspects, both transmission and reception)
Mobile/dashboard technology
Next-gen broadcast platforms and standards development
Non-real time (NRT) broadcast services
Ratings technology, second screen technology and services
Secondary service system design; mitigation of interference in primary services
Securing Broadcast IT Networks
Signal Processing & Analysis
Software Defined Radio – SDR Technologies
Streaming delivery of broadcast content
Transmission, propagation, reception, re-distribution of broadcast signals AM, FM, and TV transmitter and antenna systems
Transport stream issues – ancillary services
Unlicensed device operation in TV white spaces

We wish to inform you that the activities, events and publications of the Brazilian Society of Television Engineering – SET, including this one, enjoy international support, under formal agreements, from the following international organizations. We also take this opportunity to thank them and reiterate how proud we are that they support our work.

SUMMARY

SET IJBE v. 9, 2023, 67 pages

EDITORIAL

This edition of IJBE is dedicated to Phase 3 of TV 3.0. TV 3.0 is the evolution of the Brazilian Digital Terrestrial Television System (SBTVD). Phase 3 of TV 3.0 is being carried out under the coordination of the SBTVD Forum and funded by the Brazilian Ministry of Communications through the National Research and Education Network (RNP). TV 3.0 is supported by Presidential Decree No. 11.484, which establishes guidelines for the evolution of SBTVD and the provision of radio spectrum for its implementation.

This issue features several thematic articles on aspects of the technologies adopted for TV 3.0 and candidate technologies for the physical layer.

The search for quality in the development of technologies that make digital communication systems more efficient is the focus of the researchers to whom IJBE offers the opportunity to disseminate their studies, experiments and research in the scientific and technological areas of production and distribution of information content.
We hope you enjoy these articles and feel motivated to submit them.

Best wishes
SET IJBE Editors

R&D Progress on TV 3.0 Application Coding Layer

Marcelo F. Moreno, Carlos Pernisa Júnior, Eduardo Barrere, Stanley Teixeira, Cristiane Turnes Montezano, Li-Chang Shuen, Carlos de Salles Soares Neto, Débora Christina Muchaluat-Saade, Marina Ivanov Josué, Joel A. F. dos Santos, Sérgio Colcher, Daniel de S. Moares, Derzu Omaia, Tiago Maritan Ugulino de Araújo, Guido Lemos de Souza Filho

Cite this article:

Moreno, Marcelo F.; Pernisa Júnior, Carlos; Barrere, Eduardo; Teixeira, Stanley; Montezano, Cristiane Turnes; Shuen, Li-Chang; Soares Neto, Carlos de Salles; Muchaluat-Saade, Débora Christina; Josué, Marina Ivanov; dos Santos, Joel A. F. ; Colcher, Sérgio; Moares, Daniel de S.; Omaia, Derzu; de Araújo, Tiago Maritan Ugulino; de Souza Filho, Guido Lemos; 2023. R&D Progress on TV 3.0 Application Coding Layer. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2023.1. Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING - SET IJBE V.9, 2023, Article 1, 12 p
2023 SET - Brazilian Society of Television Engineering / ISSN (Print): 2446-9246 / ISSN (Online): 2446-9432

R&D Progress on TV 3.0 Application Coding Layer

Marcelo F. Moreno (Fraunhofer IIS / UFJF),
Carlos Pernisa Júnior, Eduardo Barrere, Stanley Teixeira, Cristiane Turnes Montezano (UFJF),
Li-Chang Shuen, Carlos de Salles Soares Neto (UFMA),
Débora Christina Muchaluat-Saade, Marina Ivanov Josué, (UFF),
Joel A. F. dos Santos (CEFET-RJ),
Sérgio Colcher, Daniel de S. Moares (PUC-Rio),
Derzu Omaia, Tiago Maritan Ugulino de Araújo, Guido Lemos de Souza Filho (UFPB)

***Abstract*—This paper outlines the methodology, progress, and initial outcomes of a collaborative R&D effort by 40 researchers from six academic institutions, focused on addressing critical application-coding requirements for SBTVD Forum's TV 3.0 Project, Phase 3. The team is developing prototypes and use cases applications to validate and demonstrate TV 3.0 application coding features. Key developments include architectural changes, a persistent media player, a viewer's journey design, besides extensive research in requirements engineering, user data analysis, and novel codec support. The team also explored application coding extensibility, enhanced accessibility, immersive experiences and multimodal interaction. During SET Expo 2023, partial implementations of the prototyped use cases were showcased, highlighting the project's progress and significance. This paper provides technical details and diagrams, facilitating a thorough discussion of this innovative project.**

***Index Terms*—Application coding, TV 3.0, Application-based TV experience, Personalized TV experience, Immersive TV experience, accessibility.**

I. Introduction

Television plays a social role of immense relevance in Brazil. TV is more than a technological object in the room: it is also an important cultural artifact and an element of national integration. Therefore, any technological evolution that represents a change in the way of watching television may imply some cultural change for the society itself. Thus, the development process of SBTVD Forum's TV 3.0 Project [1], at least for its most highlighted application coding use cases and features, is being carried out not only as a technology research but also as a social study.

Brazil has been watching TV since 1950, with the inauguration of TV Tupi in São Paulo. In seven decades, Brazilian television has experienced technological, social, and content development that makes the national experience one of the richest in the world. Remarkably, the current terrestrial DTV system specifies the Ginga middleware, a national technology, as the standard for multimedia interactivity since 2007. Ginga has been proven to support a consistent evolution that made it the first Brazilian technology adopted as an international standard by ITU-T in 2009 [2] and recognized by ITU-R as an integrated broadcast-broadband system since 2017 [3].

TV 3.0 Project is currently under Phase 3[1], carrying out further tests and evaluations over the physical and video coding layers, as well as developing a reference mux/demux. Regarding the application coding layer, most of the innovative requirements established by the Call for Proposals (CfP) [4] are under study by selected Academia research groups since those requirements were not appropriately addressed in the previous phases [5]. The CfP specified 17 requirement groups for application coding, including basic aspects of backward compatibility with Ginga specifications and its implementation reuse, besides support for TV 3.0 underlying technologies. The advanced requirements include support for application-based TV experience, immersive audiovisual content, multimodal interaction, sensory effects, multi-user profiling, audience measurement, IP convergence, and extensibility, just to name a few of them.

This paper focuses on the methodology, progress, and early achievements of the Academia R&D team in addressing the high-priority application-coding requirements for TV 3.0. The team is composed of 40 researchers from 6 academic institutions, namely PUC-Rio, UFPB, UFF, UFJF, UFMA and CEFET-RJ. The work started in April 2023.

As a means of actively collaborating with the research methodology, SBTVD Forum's Technical and Market Modules jointly decided on a prioritization of requirements to determine the sequence of studies for the R&D team. In addition, the Forum's Application Coding Working Group (WG) specified initial guidelines on how to tackle each requirement based on the evaluation results from Phase 2 and the WG's expertise in standardizing/implementing digital TV middleware. Finally, the WG specified a total of 7 use cases to be prototyped, aiming at validating the R&D solutions for the prioritized requirements and publicly demonstrating the new TV 3.0 application coding features. The R&D team diligently incorporated all SBTVD Forum contributions into its methodology, allowing for consistent progress on certain requirements.

[1] TV 3.0 Project Phase 3 is funded by the Brazilian Ministry of Communications (MCom), managed by the Brazilian Network for Education and Research (RNP).

First, relevant architectural changes in the application coding layer were already proposed and agreed, based on the fundamentals of Ginga specifications for Profile-D receivers. The requirements for an application-based TV experience impose changes that include a new user interface for listing each broadcaster's initial application. In addition, a new media player is needed, capable of keeping running over application switches, regardless of whether the current audiovisual content is delivered over-the-air (OTA, broadcast) or over-the-top (OTT, broadband).

There is also significant effort in requirements engineering and social studies regarding this application-based TV experience, as mentioned before. The team is running focus groups and opinion polls with a probabilistic sample, as well as prototyping the entire viewer's journey based on the quanti/qualitative data obtained. This prototype will be further refined following the principles of Design Thinking, under discussion with a team of experts proactively assigned by RNP (Brazilian Network for Education and Research).

Another area of focus involves evaluating the features introduced by the adopted audio and video codecs, with the objective of identifying properties that can be utilized by applications and determining the necessary implementation support.

The extensibility requirement is also under study, with a focus on identifying Ginga-NCL and Ginga Common Core WebServices (Ginga CC WS) APIs as metadata so that new applications can obtain granular information about the functionality support of their interest in the receiver, thus allowing them to adapt according to the available resources.

A further R&D task has focused on the accessibility requirements, more specifically on the captioning part, where the IMSC1 standard is adopted for encoding and transmitting subtitles and sign language gloss. It uses a subset of TTML, which consists of an XML file with several possible settings for captioning, such as position, color, font, display time, synchronism, emojis, and images. In order to test and validate the forwarding of captions and glosses to mobile devices over the local network, a prototyping environment was developed, composed of a partial Ginga CC WS implementation and NCLua and HTML5 applications. New required APIs are added to the Ginga CC WS prototype, in this case for the real-time forwarding of captions and gloss in TTML format over sockets or websockets. The synchronism between the applications is performed by the Ginga CC WS, which delivers the same content, at the same time, for all socket clients, the results are being evaluated. The gloss stream is shown in the application by a 3D avatar playing sign language.

The team has also been working on the implementation of use cases related to sensory effects, immersive content and multimodal interaction. To accomplish this, the team is working on harmonizing the adopted proposals NCL 4.0 [6] and Guaraná [11]. Combined, they allow for the inclusion of sensory effects (wind, scent, light etc) in interactive multimedia applications and the execution of parts of the application on head-mounted displays connected to the TV, in a 360° scene, including 3D objects, immersive MPEG-H 3D audio [12] and traditional multimedia objects. In addition, users will be able to interact with applications using different modalities (voice, gestures etc, using input recognition devices.

In conclusion, partial implementations of prototyped use case apps were demonstrated at SET Expo 2023 in the SBTVD Forum booth. This paper is structured as follows: Section II discusses the rationale and agreed-upon changes to the application coding architecture for TV 3.0. Section III describes the application-based TV experience and a projected viewer's journey, providing context for focus group discussions. Section IV reports progress on application coding support for TV 3.0 audio/video codecs and application coding extensibility. Section V examines developments in supporting accessibility content, including second-screen delivery of captions and sign language glosses. Section VI presents achievements in immersive experience support, encompassing sensory effects, multimodal/multiuser, and virtual reality content. Finally, Section VII offers concluding remarks. This paper includes diagrams and technical details for deepening the discussion on each study of this challenging project of unparalleled opportunity.

II. Application Coding Architecture for TV 3.0

Several of the use cases designed for TV 3.0 clearly indicate the need for intensive use of multimedia applications, which, according to CfP TV 3.0 [4], will be based on extensions to Ginga specifications for Profile-D receivers (a.k.a. DTVPlay). Undoubtedly, it is through Ginga applications that broadcasters and partners will be able to leverage TV 3.0 greatest innovations, including personalization of the TV content consumption experience, segmented programming, manipulation of additional content in more immersive formats, accessibility, sensory effects, as well as new forms of interaction. In addition, it is through Ginga applications that it will be possible to build and manipulate viewer profiles that enable such personalized experiences, obviously for the viewers who consent.

It is therefore expected that Ginga applications will be running and switching constantly on TV 3.0 receivers, leading to a need to rethink the application coding support specifications. Application coding becomes no longer an accessory for broadcasters but a key element for enabling the vast majority of new TV 3.0 use cases. The CfP makes this clear through its requirements group AP6 "Enable application-oriented TV" [4]:

- AP6.1: application-oriented user experience with TV
- AP6.2: handling the presentation of all audiovisual content
- AP6.3: application switching delay (lower is better)

This represents, objectively, a paradigm shift, which in fact needs also to be discussed from the viewer's point of view, according to our study on the viewer journey possibilities, presented in Section III. Nevertheless, this evolution towards an application-oriented TV has to be reflected in the receivers' application coding support architecture. The challenge set to the TV 3.0 Project R&D team was to propose adaptations to the current Ginga architecture, according to ABNT NBR 15606-1, in order to maintain compatibility and to reuse, as much as possible, the existing implementations of TV 2.5 middleware components and subsystems.

The current TV 2.5 middleware architecture can be

set.org.br/ijbe doi: 10.18580/setijbe.2023.1 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

depicted as shown in Figure 1. In summary, this architecture demonstrates the capability for broadcasters to transmit Ginga applications via OTA that are developed using NCL/Lua or HTML5/ Javascript languages. Consequently, the Ginga-NCL [13] or Ginga-HTML5 [14] presentation engines execute these applications based on the OTA signaling rules provided. The current APIs of both presentation engines offer interesting possibilities for integrating broadcast and broadband features in D-profile receivers. These features encompass receiving broadband content through adaptive streaming, with or without DRM protection, as well as facilitating TCP and UDP communication in both client and server modes. Additionally, the architecture supports content preparation to enhance the quality of the viewing experience during the seamless transition to broadband, including the insertion of targeted advertising, among other functionalities.

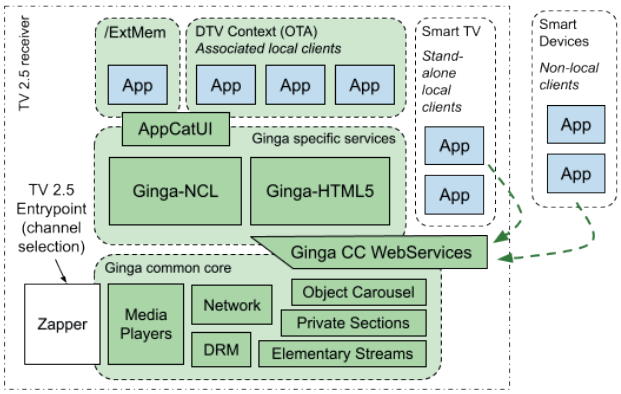


Figure 1. TV 2.5 middleware architecture.

In the case of Ginga-HTML5, since it exclusively employs W3C standardized APIs, all these resources are accessible through a decoupled API outlined in the Ginga CC WebServices specification [15]. This decoupling is achieved by utilizing a remote API that adheres to the RESTful architectural style. Consequently, this approach enables the same Ginga CC WebServices to provide TV 2.5 resources to applications beyond broadcasting, including those operating within the smart TV environment on the same receiver or on any smart device within the home network. However, access permissions must be granted by both the viewer and the broadcaster in such cases.

Even Ginga-NCL applications can utilize Ginga CC WebServices resources, particularly for use cases that encourage their integration with second-screen applications running on smart devices. For all other use cases, Ginga-NCL directly furnishes the required APIs through the NCL and Lua languages, potentially offering improved performance when accessing resources.

Starting from Profile C receivers[2], a component known as AppCatUI (Application Catalog User Interface) becomes available. This component serves the purpose of listing applications accessible within the current DTV context, allowing viewers to trigger them. It also facilitates viewers in adding and removing applications, making them persistent, and initiating their execution. Such applications can be delivered OTA with appropriate installation permissions, installed from external memory devices (/ExtMem), or downloaded from authorized repositories accessible via broadband connections.

However, despite offering these possibilities, the current TV 2.5 specifications fall short of enabling an application-oriented TV approach for several reasons. Firstly, this limitation arises because the initial entry point into the TV content consumption experience revolves around the channel abstraction, typically managed by native software responsible for channel switching, depicted in Figure 1 with the suggestive term “zapper”. Notably, this “zapper” is not an integral part of the middleware specifications and so Ginga applications have only limited control over the behavior of the zapper. When they intend to present OTT content, for example, they are required to employ an additional media player, which usually has a lifecycle closely tied to the application itself. If an attempt is made to switch to another application, the current player instance will be terminated. This underscores the necessity of incorporating a persistent media player as an essential element of the new architecture for TV 3.0 application coding.

Moreover, the application-oriented paradigm holds the potential to conceal the concept of traditional channels, presenting each broadcaster as an application capable of providing access to its complete ecosystem of content and services. To achieve this goal, it is imperative to introduce a user interface, a component of the application coding layer, that can list each broadcaster's initial application and provide access to other applications offered by broadcasters for installation or execution. The existing AppCatUI can indeed list installed Ginga applications and those available in the current DTV context, but it lacks the capability to showcase applications as primary UI for all broadcasters.

To address these specifics, the proposal, well-established between the R&D team and the SBTVD Forum for an application-oriented TV, suggests making adaptations to existing middleware architecture components. These changes aim to minimize the impact on current implementations while allowing viewers to embrace the new paradigm. The proposed architecture for coding TV 3.0 applications is depicted in Figure 2.

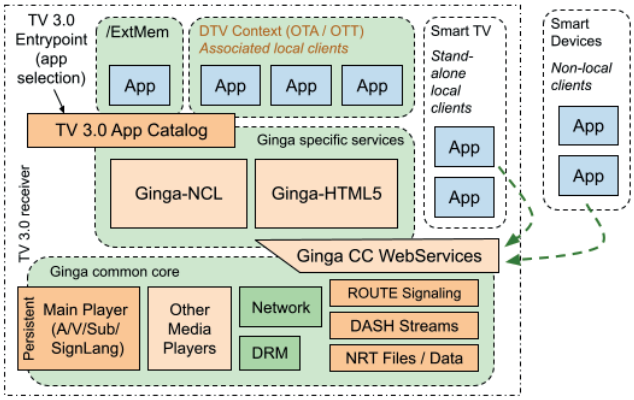


Figure 2. Proposed TV 3.0 application coding architecture.

[2] Profile C was specifically designed for receivers distributed during the analog TV signal switch-off process, representing a significant evolution addressing advanced requirements for implementing public policies and digital inclusion [16]

set.org.br/ijbe doi: 10.18580/setijbe.2023.1 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

The revamped AppCatUI evolves into a TV 3.0 Application Catalog, serving as a front-end for viewers to identify available broadcasters, explore their content ecosystems, and configure profiles and other potential options previously absent in TV receivers. In this architecture, a persistent media player enables seamless switching between OTA and OTT content within broadcasters' applications without playback interruptions. The next application can then decide whether to maintain the player with the current content, recommend new content, or even switch automatically to other content, all with the viewer's best experience in mind.

The figure also incorporates new components to access resources from other TV 3.0 system layers, notably the transport layer based on the ROUTE/DASH specification. Additional extensions to this architecture, discussed in Section VII, support integration with sensory effects, advanced interaction recognition, and virtual reality devices. It is anticipated that APIs supported by Ginga-NCL and Ginga CC WebServices will be expanded to accommodate solutions for the ongoing requirements.

To list an initial application for each broadcaster within the TV 3.0 Application Catalog, the conventional and time-consuming channel scanning process can be presented to viewers as an application discovery process. To expedite this process in TV 3.0, various options are being considered. These include the automatic instantiation of each initial application into the Application Catalog, eliminating the need for individual application downloads for each station found. Additionally, when the receiver is connected to the internet, a new discovery web service could provide a list of licensed broadcasters in the installation region. This could enable the automatic instantiation of an application, even in cases of weak signals, directing viewers to the broadcaster's linear OTT content if available. Another advantage of such a discovery service is the potential for automatic updates to the applications list, eliminating the need for repeated scanning processes as currently required for adding new TV channels.

These innovative possibilities are designed to enhance the viewer's experience and can be assessed through experimental prototyping of the entire viewer journey, even involving TV receiver aspects beyond the scope of free-to-air digital TV standards. The subsequent section outlines the initial steps in designing a possible viewer/interactor journey.

III. Application-based TV Experience

Application-oriented TV is the central concept for understanding the paradigm shift proposed for the third generation of Brazilian digital TV. The change is technological but also cultural, as it alters the way viewers traditionally relate to accessing the content offered by broadcasters. In the proposed model, channels will be replaced by applications offered by broadcasters. From this initial application, each broadcaster will be able to create its own ecosystem of internal apps and offer content both OTA and OTT, which the viewer will access depending on the existence of connectivity on their Smart TV.

This evolution makes sense in a scenario in which the predominant Smart TVs on the market already offer FAST (Free Ad-Supported Television) channels, which compete without regulation for the audience with free-to-air channels distributed by broadcast. In addition to FAST channels, there is a whole range of streaming applications that occupy the screen and, in the case of televisions connected to the Internet, monopolize viewers' choices. In fact, there are even keys dedicated to streaming services on the minimalist remote controls. In fact, on these televisions, it is increasingly difficult for viewers to find the free-to-air TV channels whose content they want to consume. The proposal for an application-oriented television paradigm attempts to resolve this issue by offering the viewers an experience in which they can easily identify the devices' native apps and the applications from free-to-air broadcasters.

To facilitate the design of a possible viewer/interactor[3] journey (see subsection III.A), we first focused on assessing the interfaces of current smart TVs and studied video streaming on digital platforms. The examination included studying the most commonly used smart TV models in the Brazilian market, based on operating systems such as Roku TV, Android TV or Google TV, WebOS, and Tizen. These platforms exhibit differences in content presentation and viewer interaction. Roku TV and Android TV prioritize application presentation, while WebOS and Tizen focus on keeping audiovisual content on the screen for extended periods, overlaying settings, menus, and other applications only when activated by the viewer. Across all systems, free–to–air TV content occupies a distinct application space, varying in colors, icons, and terminology.

In the analysis of streaming services and their interfaces, we aimed to identify familiar paths for audiovisual consumers, seeking to adapt these experiences to new interactive actions and requirements for a possible TV 3.0 Application Catalog interface. We observed that different platforms often share similarities in presenting content on their home screens, primarily dedicated to on-demand content. Tabs categorize content by genre and format, such as drama, comedy, sports, and news. Live content is typically featured within specific applications. For instance, Globoplay includes live TV content, including simultaneous broadcasts from Globo and other affiliated channels. RTVEplay prominently displays live content on its platform, and Pluto TV offers live content on its home page while organizing on-demand content in a separate tab.

Initial findings have been incorporated into a proposed viewer’s journey model, including icon arrangement on TVs, the importance of a universally recognizable identifier for free-to-air TV, and the need to carefully consider the relationship between viewers/interactors and the TV 3.0 Application Catalog interface to ensure a seamless transition for those accustomed to traditional TV.

We are also exploring the concept of a "networked time" called "Timelink" to free viewers from rigid linear TV scheduling. To establish Timelink and provide viewers with time control, an intuitive program guide using deep links in

[3] Murray [18] defines an interactor as someone who effectively interacts with content on a media device.

set.org.br/ijbe doi: 10.18580/setijbe.2023.1 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

Electronic Program Guide (EPG) metadata sent by broadcasters is crucial. While browsing the guide, viewers can access detailed information about each content item and initiate playback with a simple click. The guide can also signal which content is immediately available, considering viewer preferences and internet connectivity. This guide streamlines access to both OTA and OTT content.

In a non-linear TV landscape, a more efficient and intelligent use of remote control, particularly the colored buttons, is under investigation.

Lastly, the concept of a second screen is being reimagined as a mirror of the television screen on a separate device, a departure from the current practice of integrating the remote control within the TV interface.

A. A VIEWER'S JOURNEY PROPOSAL FOR EVALUATING THE APP-ORIENTED TV EXPERIENCE

The proposal for a viewer's journey of an application-based television, reproduced below, considers cultural, social and economic aspects of how Brazilians consume audiovisual content on free-to-air TV from the time the device is turned on for the first time to be configured until the moment the viewer chooses and watches what is being broadcasted.

An important TV 3.0 feature lies in personalizing content according to the viewer's preferences. When turning on the TV for the first time, the viewer is invited to choose the configuration language, which includes accessibility options with audio descriptions. This is followed by the possibility of defining a profile that can be shared with broadcasters, with the definition of important characteristics such as whether it is a child and what age recommendation for content is suggested. Figure 3 illustrates this viewer profile creation screen.

Figure 3. Viewer profile screen.

Instead of the channel scanning process, in TV 3.0, there is the process of discovering initial apps from broadcasters. This process is based on the geographic location of the receiver, so it will list the stations that are available in that region. Figure 4 illustrates the discovery of three broadcasters.

In Figure 5, the purpose of this screen is that the viewer can easily identify where free-to-air TV stations are found on Smart TVs. This screen represents a possible harmonization with a clear indication of what are OTT streaming apps, broadcast apps and FAST channels.

Figure 4. Broadcaster scanning.

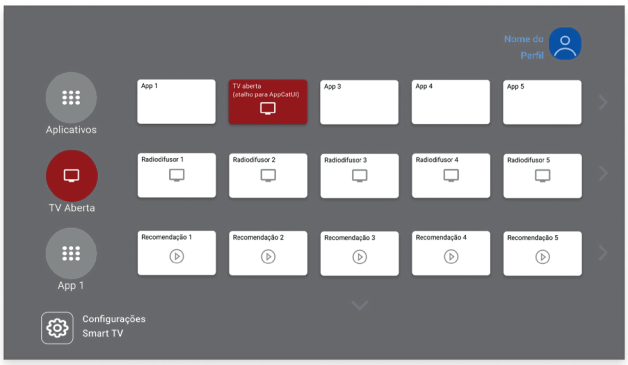


Figure 5. Example of a Smart TV Home screen harmonizing streaming and broadcast TV apps.

Figure 6 shows the TV 3.0 Application Catalog, a screen where all the initial apps for free-to-air TV broadcasters in the region are listed for the viewer. This screen corresponds to the usual broadcast TV screen, which is zapped using the remote control in the traditional way.

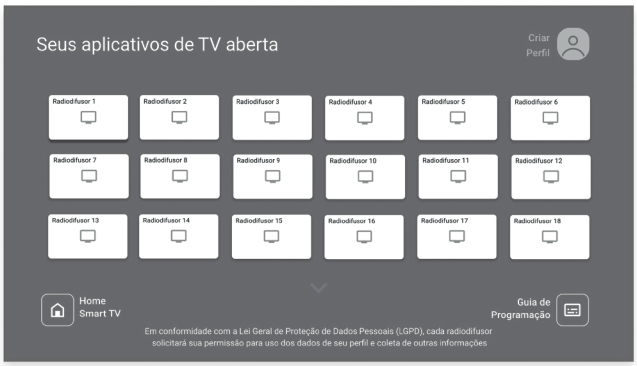


Figure 6: TV 3.0 Application Catalog environment.

Figure 7 illustrates the initial app of a broadcaster. Here, the viewer has access to traditional, linear audiovisual content that comes over the air. To watch this content, the TV receiver does not need to have an active Internet connection, and the broadcaster has some control over it, including visual identity settings to be applied to such a common initial app.

set.org.br/ijbe doi: 10.18580/setijbe.2023.1 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

Figure 7: Broadcaster initial app.

Finally, in Figure 8, we see the recommendation ecosystem of a broadcaster that suggests both OTA and OTT content. Depending on the receiver's connectivity, they will have access to a larger catalog of content.

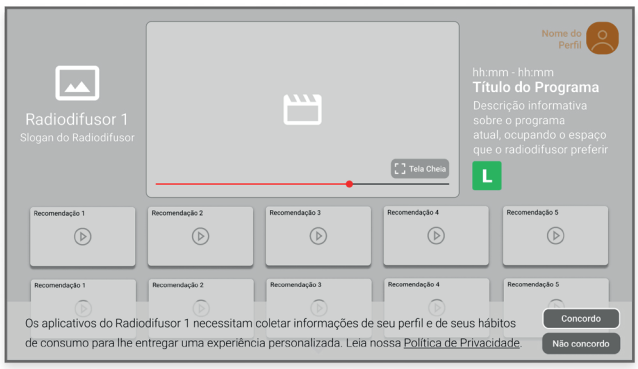


Figure 8: Broadcaster content recommendation ecosystem.

IV. Application Coding Support for TV 3.0 Audio/Video Codecs and Extensibility

Following the recommendations resulting from the tests and evaluations of TV 3.0 Project Phase 2, new video and audio codecs will be incorporated into TV 3.0. At the application layer, the R&D team has been working on how to provide support so that applications can use the advances and features of these new media. Efforts have been put into finding and integrating decoders and players for the video codecs adopted for standardization - Versatile Video Coding (VVC) ISO/IEC 23090-3, MPEG-I part 3, and Low Complexity Enhancement Video Coding (LCEVC) MPEG-5 Part 2 - which brings with it several advantages, such as better video quality with lower bit rates than its predecessors. Furthermore, the team is also committed to integrating the MPEG-H audio (ISO/IEC 23008-3) player, developed by the proponents themselves, which in addition to reproducing the new immersive audio standard, also features an interface that allows interaction with audio objects and customization of the various functionalities offered, such as changing channels to choose a track, selecting a language, etc.

Concerning the extensibility requirement, the team has focused on surveying the APIs for listing receiver properties and resources, both in Ginga-NCL and Ginga CC WS. The idea is that these APIs can be updated and harmonized according to new features introduced with the TV 3.0 project, such as version 4.0 of the NCL language. Therefore, TV 3.0 applications will be able to consult what features and functionalities are available on a receiver, and thus be able to adapt to them. This can allow even different receivers to run applications with adapted functionalities.

V. Application Coding Support for Accessibility Content

To test and evaluate the transmission and reception of captions, sign language glosses, and audio description, a prototyping environment was proposed.

This environment is centered on the Ginga CC WS server, which manages the distribution and synchronization of the accessibility media. A REST API, which is still under development, has been extended from the existing one in TV2.5. It provides routes that allow a client, external to the TV and connected to the same local network, to access this media as long as it is authenticated on the TV.

In this way, client applications can request and access the sign language gloss, subtitles, and audio description media. From this, various scenarios can be explored. In the case of hearing-impaired viewers, the glosses received can be displayed in Sign Language visual format on the user's device without overlaying the video being shown on the TV screen. In the case of subtitle display, different viewers can access subtitles in the language they prefer, allowing different people to receive different content. And the audio description client running on their cell phone can receive the audio so the user can listen to it through headphones without disturbing other viewers.

This allows content to be customized simultaneously and in a non-imposing way since each user can have their own customization on their personal device without interference from others.

Figure 9 shows this environment and demonstrates the three scenarios presented. It's possible to see the Ginga CC WS server on the TV delivering the 3 accessibility media contents to mobile devices via a Wi-Fi network. Each device receives its media and plays it according to its type. An accessibility user can view or listen to the content received on the devices.

The Ginga CC WS prototype is being implemented over node.js, and it relies on a static cyclic content to simulate a real environment.

The media to be provided were prepared in such a way that the subtitles, glosses and audio description had equivalent content. To do this, the subtitles for a given video were generated manually, followed by their translation into glosses using the VLibras translator [10]. The audio file was produced and recorded by one of the team members. To simulate a real broadcast environment in which subtitles and glosses are broadcast continuously, they were segmented into 2-second chunks, and each was stored in a different TTML file. In this way, the content is transmitted every 2 seconds to clients connected via sockets or websockets. At the end of the file transmission, Ginga CC WS restarts the cycle, transmitting the first files again.

The test application was created using HTML5 and Javascript. It offers both a desktop version suitable for TV screens and a mobile version optimized for mobile devices. Regardless of the platform it's accessed from, the application

set.org.br/ijbe doi: 10.18580/setijbe.2023.1 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

consists of three main modules: one for displaying sign language, another for showing captions, and a third for playing audio descriptions.

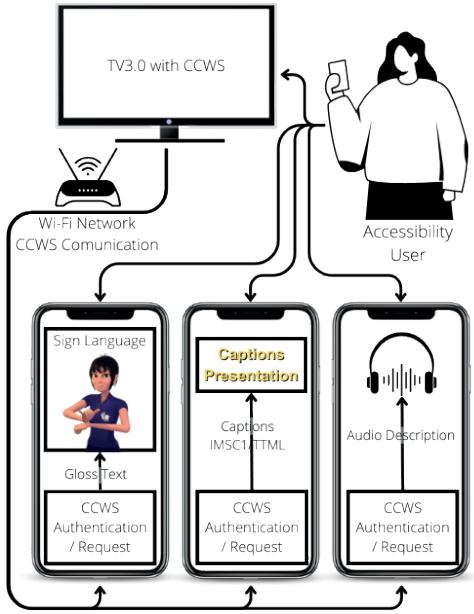


Figure 9. Accessibility prototyping environment.

The visual representation of captions in IMSC1 format is done using the open-source library imscJS [8]. This library interprets the content present in IMSC1/TTML subtitle files, allowing subtitles to be displayed in the application with the appropriate graphic formatting.

The sign language module receives the glosses from the Ginga CC WS server via a websocket. For the representation in sign language format, the application was integrated with the VLibras Widget [7], a tool that has a 3D avatar that reproduces the glosses in sign language format.

The audio description module requests the Ginga CC WS server to this media and receives the HTTP URL for the audio in DASH format [9]. The audio is then played back and the user can listen to it on their mobile device's speaker or through connected headphones. Figure 10 shows this application running during SET Expo 2023. It's possible to see the caption and sign language modules in execution at the same time on TV and on a tablet device.

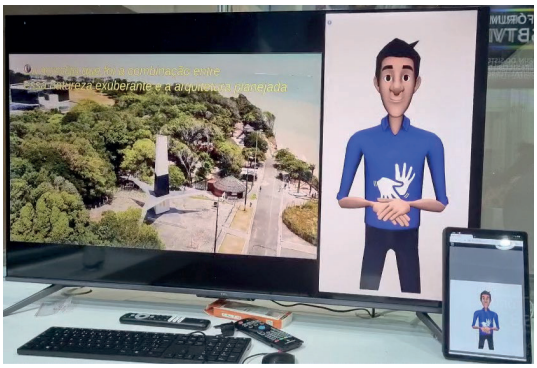


Figure 10. Accessibility application for mobile and TV.

VI. Application Coding Support for Immersive Experiences

Regarding the support for immersive experiences, the R&D team presented two use case applications during SET Expo 2023. The first focused on the execution of sensory effects and multimodal/multiuser support. The second allowed executing parts of the application on a head-mounted display (HMD) connected to the TV.

Sensory effects are used in entertainment (e.g cinema and games) to increase the user experience providing more immersion when consuming content. Aiming to provide immersive experiences in Digital TV environments, NCL 4.0 allows integrating sensory effects into interactive TV applications. In the first immersive experience use case, the R&D team specified an NCL 4.0 application that allows synchronizing light and aroma sensory effects with the audiovisual content transmitted by the broadcasters.

For the Ginga middleware to support the execution of the multisensory applications specified in NCL 4.0, it is necessary to add components capable of communicating with sensory effect renderers in the DTV receiver environment. In this way, the multisensory application will be able to activate/deactivate sensory effects and control effect presentation characteristics, such as position, the light effect color, the smell of the aroma effect, etc.

The sensory effects rendering is performed by the component named *Sensory Effect Renderer* present in Ginga common core as illustrated in Figure 11. Each *Sensory Effect Renderer* effect is associated with only one sensory effect and vice versa. The renderer defines interfaces enabling the Ginga-NCL formatter to communicate with the rendering devices and trigger actions such as starting the effect presentation, or preparing a sensory effect to guarantee temporal synchronization of the application.

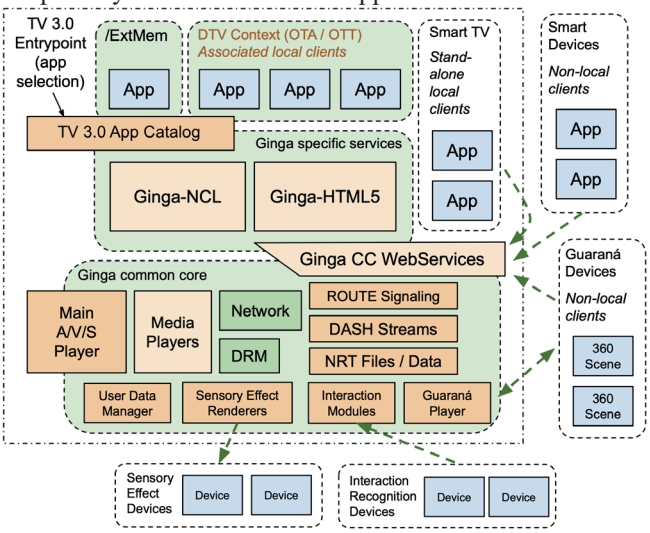


Figure 11. Ginga architecture to support immersive applications

Physical devices for rendering effects may be from different manufacturers and support different communication protocols. This communication must be implemented by the Device API, which is specific to each rendering device. The Device API must implement a set of functions such as connection to a physical device, activation/deactivation and modifying the effect intensity.

Another feature of TV 3.0 is the support for user interaction using different modes of interaction such as gestures, voice and even facial expression recognition. For example, a broadcaster can transmit an NCL 4.0 application capable of

adapting the content presented according to the viewer's facial expressions that it identifies.

User interactions with the multimedia application are managed by *Interaction Modules* present in Ginga common core. These interaction modules communicate with physical devices through predefined methods and notify the middleware when an interaction has been recognized. Additionally, the module can inform the middleware the user that performs an interaction if the recognition device is capable of identifying the user.

The first demo application showcased during SET EXPO 2023 is a travel show that presents four videos related to tourist attractions in the city of Rio de Janeiro. Initially, the application presents two videos related to the beaches of Rio de Janeiro (Ipanema and Copacabana). In both videos, a sea aroma is triggered by the application. Furthermore, a yellow light effect is presented when the sunset appears in Ipanema, and a blue light effect happens when it is a sunny morning in Copacabana. The third video presents the Botanical Garden that is synchronized with a green light effect. Additionally, the viewer can interact with the third video using gesture interaction to pause or resume it, as illustrated in Figure 12.

Figure 12. Support for gesture interaction.

Finally, the last video presents Christ the Redeemer, and a voice-based viewer interaction is asked in order to choose the last part of the show. Based on the viewer's choice, a personalized fifth video is presented. When a viewer interacts, his/her profile identification, which is already registered in the TV receiver, is recognized and shown, as presented in Figure 13. This use case also demonstrates that TV 3.0 can identify the viewer that interacts with it.

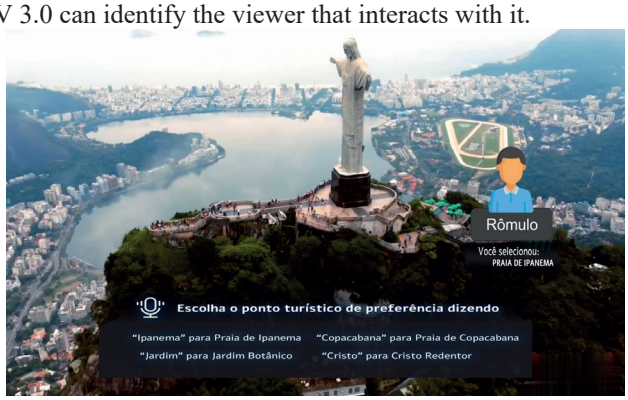


Figure 13. Support for user interaction identification.

In the use case related to the Guaraná proposal, the user starts its experience on a broadcaster's application. In that application the user has the option to watch a program where additional content is executed in an HMD. That program presents a classical music performance inside the Tiradentes Palace in Rio de Janeiro. The same content presented on the TV is available in 360º in the viewer's HMD. Together with the 360º video, the application presents photos of the palace and a video presenting the palace's architecture. Whenever the viewer turns its head in the direction of the orchestra conductor, an image describing his biography is presented. Figure 14 presents an overview of the 360º scene presented in the HMD.

Figure 14. 360º scene overview.

To implement that use case the R&D team used an implementation of the middleware Ginga Common-Core Web Services component (CC-WS) with the new API proposed for registering remote devices that will execute part of the application (in this case, a 360º scene). Once a device registers itself as a remote device, the CC-WS creates a WebSocket for the bidirectional communication between CC-WS and HMD. Once the main application (executed at the TV) starts its execution, *i.e.*, the 2D version of the orchestra presentation, at the TV, the CC-WS component transmits to the HMD the description of the 360º scene. As the presentation unfolds, the CC-WS component sends commands instructing the HMD to start/stop presenting the content in the scene. Whenever the Guaraná logo at the orchestra conductor is in the user field of view, the HMD sends to CC-WS an interaction report indicating the start of the View event of the logo. The same is performed when the logo exits the field of view, triggering the end of the View event. Once the CC-WS receives a start/stop of the View event, it replies to the HMD with a command to start/stop the conductor biography.

VII. Final Remarks

In conclusion, this paper presents a comprehensive exploration of the innovative TV 3.0 project, focusing on the R&D progress on various aspects of the application coding layer.

TV 3.0 project has ushered in a new era of television technology by prioritizing an application-oriented TV experience. Through meticulous research and development, the project has been fundamental to rethink how viewers interact with television content. The transformation of the traditional TV interface into a versatile, application-centric platform has the potential to enhance user engagement and to offer viewers a personalized control over their content

set.org.br/ijbe doi: 10.18580/setijbe.2023.1 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

consumption.

However, it's important to acknowledge that the TV 3.0 project remains a work in progress. The application coding requirements addressed in this paper are part of an ongoing journey, and solutions will continue to evolve. As the project progresses, it is expected that the remaining requirements will also be tackled with innovative solutions, further enhancing the TV 3.0 possibilities.

Extensibility plays a crucial role in the TV 3.0 project, with a commitment to evolving APIs in Ginga-NCL and Ginga CC WS. This adaptability ensures that TV 3.0 applications can seamlessly integrate with a wide range of receiver configurations, accommodating diverse user preferences and hardware capabilities. The ongoing development in this area promises even greater flexibility and compatibility in the future.

Accessibility is at the forefront of TV 3.0, with a strong emphasis on customization. The project's dedication to providing tailored captions, sign language glosses, and audio descriptions ensures that television content is inclusive and accessible to a wide audience. As the project matures, these accessibility features will continue to evolve to meet the evolving needs of viewers.

The paper also delves into the realm of immersive experiences, demonstrating TV 3.0's capability to synchronize sensory effects, support multimodal/multiuser interactions, and integrate with head-mounted displays. These developments represent a significant shift in television engagement, offering viewers interactive and captivating content experiences.

The TV 3.0 project, in general, is promoting the way for a future where television transcends its traditional confines and provides viewers with unparalleled and personalized experiences. This paper serves as a testament to the exciting possibilities and innovations that lie ahead in the realm of TV 3.0, with the understanding that the journey is ongoing, and the best is yet to come.

REFERENCES

[1] SBTVD Forum. “TV 3.0 Project”. Website. https://forumsbtvd.org.br/tv3_0

[2] ITU-T. Recommendation ITU-T H.761 (2009) “Nested Context Language (NCL) and Ginga-NCL”

[3] ITU-R. Recommendation ITU-R BT.2075-1 (2017) “Integrated broadcast-broadband system”

[4] SBTVD Forum. “Call for Proposals (CfP): TV 3.0 Project”. https://forumsbtvd.org.br/wp-content/uploads/2020/07/SBTVDTV-3-0-CfP.pdf

[5] SBTVD Forum. TV 3.0 Project - Phase 2 - Results https://forumsbtvd.org.br/tv3_0/#panel-phase2

[6] BARRETO, F. ; DE ABREU, R. S. ; JOSUE, M. I. P. ; MONTEVECCHI, E. B. B. ; VALENTIM, P. A. ; MUCHALUAT-SAADE, D. C. . Providing multimodal and multi-user interactions for digital tv applications. MULTIMEDIA TOOLS AND APPLICATIONS, v. 82, 2023. https://link.springer.com/article/10.1007/s11042-021-11847-3

[7] Universidade Federal da Paraíba (UFPB). 2023. VLibras - Governo Digital. https://vlibras.gov.br/. Online; Accessed on August 30, 2023.

[8] Pierre-Anthony Lemieux, Nigel Megitt, and Robert Bryer. 2022. ImscJS Repository. https://github.com/sandflow/imscJS. Online; Accessed on August 30, 2023.

[9] British Standards Institution. 2022. ISO/IEC 23009-1 AMD 1. Information Technology. Dynamic Adaptive Streaming Over HTTP (DASH).: Part 1. Media presentation description and segment formats. Technical Report pt. 1. https://www.iso.org/standard/83314.html

[10] Luana S. Reis, Tiago M. U. Araújo, Yuska P. C. Aguiar, Manuella A. CB Lima, and Angelina S. S. Sales. 2018. Assessment of the treatment of grammatical aspects of machine translators to Libras. XXIV Simpósio Brasileiro de Sistemas Multimídia e Web. Anais Salvador, Brasil: SBC–Sociedade Brasileira de Computação, 2018. DOI: https://doi.org/10.5753/webmedia.2018.4570.

[11] Gabriel Souza, Daniel Silva, Matheus Delgado, Renato Rodrigues, Paulo R. C. Mendes, Glauco Fiorott Amorim, Alan L. V. Guedes, and Joel dos Santos. 2020. Interactive 360-Degree Videos in Ginga-NCL Using Head-Mounted-Displays as Second Screen Devices. In Proceedings of the Brazilian Symposium on Multimedia and the Web (São Luís, Brazil) (WebMedia ’20). ACM, New York, NY, USA, 289–296.

[12] O. Major, Z. Shaban, B. Czelhan, A. Murtaza. “Immersive Audio Application Coding Proposal to the SBTVD TV 3.0 Call for Proposals”. SET International Journal of Broadcast Engineering, vol. 7, pp. 48-56 2021. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi:10.18580/setijbe.2021.4.

[13] ABNT Standard NBR 15606-2 (2023) “Digital terrestrial television - Data coding and transmission specification for digital broadcasting - Part 2: Ginga-NCL for fixed and mobile receivers - XML application language for application coding”

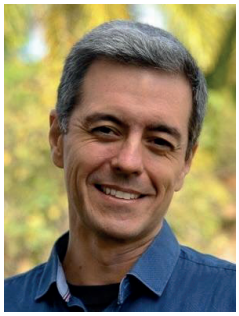
[14] ABNT Standard NBR 15606-10 “Digital terrestrial television - Data coding and transmission specification for digital broadcasting - Part 10: Ginga-HTML5 - Ginga HTML5 profile specification”

[15] ABNT Standard NBR 15606-11 “Digital terrestrial television - Data coding and transmission specification for digital broadcasting - Part 11: Ginga CC WebServices - Ginga Common Core WebServices specification”

[16] CASTRO, Cosette; BARBOSA FILHO, André. Proyecto Brasil 4D Interactividad en televisión pública. Rev. Cienc. Soc. [online]. 2016, vol.29, n.38, pp.145-159. ISSN 0797-5538.

[17] G. K. Walker, T. Stockhammer, G. Mandyam, Y. -K. Wang and C. Lo, "ROUTE/DASH IP Streaming-Based System for Delivery of Broadcast, Broadband, and Hybrid Services," in IEEE Transactions on Broadcasting, vol. 62, no. 1, pp. 328-337, March 2016, doi: 10.1109/TBC.2016.2515539.

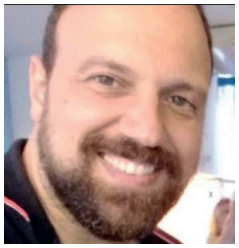
[18] Murray, Janet H. (1997). Hamlet on the Holodeck: The Future of Narrative in Cyberspace. New York: Simon & Schuster. ISBN 068482723-9.



Marcelo F. Moreno has been an associate professor in the Department of Computer Science at the Federal University of Juiz de Fora (UFJF) since 2011. He holds a master's degree (2002) and a doctorate (2008) in Computer Science from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). In 2022/23, he held a Visiting Professor chair at the International Audio Laboratories Erlangen, a joint institution of Friedrich-Alexander Universität (FAU) and Fraunhofer IIS, Germany. He is co-chair of the Task Group on Media Coding in the Focus Group on Metaverse organized by the International Telecommunication Union (ITU-T). He is a co-editor of ITU-T Recommendation H.761 "NCL and Ginga-NCL", as well as other ITU-T recommendations and technical papers. Since 2015, he has been the coordinator of the Application Coding Working Group of the Brazilian Digital TV System (SBTVD) Forum. His areas of interest include Operating Systems, Computer Networks, and Distributed Systems, with an emphasis on Multimedia communication, architecture, and modeling. He is currently a CNPq Technological Development Productivity Fellow (DT-2).

Carlos Pernisa Júnior has been a full professor at the Department of Professional Techniques and Strategic Content at the Faculty of Communication at the Federal University of Juiz de Fora (UFJF). He has a degree in Social Communication - Journalism, from the Federal University of Juiz de Fora (1990), a master's degree (1995) and a doctorate (2000) in Communication and Culture from the School of Communication of the Federal University of Rio de Janeiro (UFRJ). Member of the permanent body of the Postgraduate Program in Communication at the Federal University of Juiz de Fora (UFJF). He is leader of the “Digital Media Laboratory” Research Group (UFJF / CNPq). He has experience in the area of Communication, with an emphasis on Digital Media, working mainly on the following topics: communication, digital media, journalism, cinema and image.

Eduardo Barrére has been an associate professor in the Department of Computer Science at the Federal University of Juiz de Fora (UFJF) since 2009. He holds a master's degree (1997) in Computer Science form the Federal University of São Carlos (UFSCar) and a doctorate (2008) in Systems and Computing Engineering from the Federal University of Rio de Janeiro (UFRJ). He is the coordinator of the Computer Application and Innovation Laboratory (LApIC) at UFJF. Develops research in the areas of computer networks, Digital TV and multimedia.



Stanley Cunha Teixeira has been a research associate at the Digital Media Laboratory (LMD) of the Federal University of Juiz de Fora (UFJF). He holds a Ph.D. in "Intelligence Technologies and Digital Design" from the Pontifical Catholic University of São Paulo (PUC-SP) and a master's degree in "Aesthetics, Networks, and Technoculture" from UFJF. Additionally, he is specialized in "TV, Cinema, and Digital Media" from UFJF and holds a bachelor's degree in Social Communication (Journalism) from UFJF. He is a member of the research group on Digital Technologies in Education (TEDE) at the Federal University of Paraná. Currently, his focus is on research and development (R&D) of new information and communication technologies (ICTs) for building interactive, immersive, assistive, and inclusive experiences

Cristiane Turnes Montezano is a journalist at the Federal University of Juiz de Fora (UFJF). She is a Master and Doctoral student in the Postgraduate Programme in Communication of the Federal University of Juiz de Fora (PPGCOM/UFJF). Her studies focus on audiovisual in the digital environment, with research on streaming content, TV flows with the digital environment and TV 3.0. Member of the Digital Media Laboratory (LMD) Research Group.

Li-Chang Shuen has been an Associate Professor in the Department of Social Communication at the Federal University of Maranhão since 2008. Permanent professor of the Post-Graduate Program in Communication - Professional Master's Degree - at UFMA. She holds a degree in Social Communication - Journalism from the Federal University of Maranhão (2002) and a master's degree in Communication from the Federal University of Pernambuco (2005), with a PhD in Social Sciences from the Center for Research and Graduate Studies on the Americas of the University of Brasília (2013) and a postdoctoral internship in Political Science at the Institute of Social and Political Studies (IESP) of the State University of Rio de Janeiro (2017). Coordinator of the Integrated Laboratory of Research and Journalistic Practices of the Department of Social Communication at UFMA. PhD student in Computer Science at the Federal University of Maranhão. She has experience in the area of Communication, with emphasis on Journalism, Technology and Political Communication, working mainly on the following topics: television, journalism, politics, communication and technology, public opinion.

Débora Christina Muchaluat-Saade is a full Professor at the Department of Computer Science of Fluminense Federal University (UFF). She holds a Computer Engineering bachelor's degree, MsC and PhD in Computer Science from PUC-Rio. She is currently the vice-coordinator of the Special Committee on Multimedia and the Web (CE-WebMedia) of the SBC (Brazilian Computing Society) and she was the coordinator of the Special Committee on Computing Applied to Healthcare (CE-CAS) from 2017 to 2019. She is the founder of the MídiaCom Research Lab (www.midiacom.uff.br) and one of the lab deans. She is a member of the council and technical committee of the Brazilian Digital TV Forum. She has been contributing to the development of NCL (Nested Context Language) and Ginga-NCL, used in the Brazilian Digital TV Standard (ABNT NBR 15606-2) and IPTV services (ITU-T H.761). Her research interests are multimedia, mulsemedia, computer networks, wireless networks, smart grids, IoT, interactive digital TV and digital healthcare.

set.org.br/ijbe doi: 10.18580/setijbe.2023.1 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

Marina Josué received the Ph.D. degree in Computer Science from the Fluminense Federal University (UFF), in 2021. She is currently a postdoctoral researcher in the postgraduate program in Computer Science at UFF actuanting on e-health and digital TV projects. Her main research interests include multimedia systems, computer networks, IPTV and Digital TV.

Joel dos Santos is professor at the Computer Science Department at Cefet/RJ, teaching at the Technical High-school, undergraduate and graduate courses in computer science. He currently coordinates the Postgraduate Program in Computer Science (2021-2025). He has a degree from Universidade Federal Fluminense (UFF) in Telecommunications Engineering (2009), with a sandwich in the period 2007-2008 at Universität Ulm, a Master's degree in Computer Science (2012) and a PhD in Computer Science (2016), with a sandwich in the period 2014-2015 at Inria-Grenoble. He has worked in the Multimedia area since 2006 in several projects related to digital TV, authoring and validating multimedia applications and multisensory multimedia applications. He was vice-coordinator (2020/21 biennium) and coordinator (2022/23 biennium) of the Special Committee on Multimedia and the Web (CE-WebMedia) of the SBC (Brazilian Computing Society). Within the scope of Cefet/RJ, he is leader of the Multimedia Research Group.

Sérgio Colcher is a full-time Professor at the Computer Science Department of the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). He completed his Computer Engineering undergraduate studies in 1990 at PUC-Rio and holds the M.Sc. and Ph.D. degrees in Computer Science from the the same University (1993 and 1999, respectively). He has also been for a Post-Doctorate at ISIMA (Institut Supérieur D'Informatique, de Modélisation et des Applications) – Université Blaise Pascal, Clermont-Ferrand, France, during the year of 2003. Dr. Colcher worked as a hardware development engineer at COBRA (a Brazilian industry) between 1990 and 1991, and as a Researcher (trainee) at the IBM-Rio Scientific Center, between 1992 and 1995 (during his graduate studies). Prof. Colcher is the author of the best-known networking technology textbook in Portuguese, "Redes de Computadores: das LANs, MANs e WANs as Redes ATM" (Elsevier, 1995). This book was nominated, in 1996, for the “Jabuti Prize” (Science and Technology Category), one of the most important awards in Brazilian literature. Dr. Colcher is also the author of a book on VoIP (Elsevier, 2005). In 2012, Prof. Colcher was awarded the prize for "Personality of the Year (Academic Sector Category)" from the Brazilian Information Technology Industry Association (ASSESPRO). In 2014, he was the academic chair of the 20th biennial ITS (International Telecommunications Society) Conference and an invited keynote speaker at the same conference. Prof. Colcher currently holds a grant as a researcher from a project with the U.S. Air Force Office of Scientific Research (AFOSR).

Daniel de S. Moraes is a PhD student in Computer Science at the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). He also holds Masters (2019) and Bachelor (2016) degrees in Computer Science from the Federal University of Maranhão (UFMA), with sandwich undergraduate period (2013-2014) at the National University of Ireland Maynooth by the Science without Borders program. He has worked in several projects in Digital TV and Multimedia in the Laboratory of Advanced Web Systems at UFMA, from 2011 to 2019. Currently, he is a researcher at the TeleMídia Laboratory - PUC-Rio since 2019, working mainly on the following topics: Artificial Intelligence applied to Multimedia Systems, Digital TV, Multimedia Applications, Authoring Tools and Multimedia Document Engineering.

Derzu Omaia holds a degree in Computer Science from the Federal University of Paraíba (UFPB) and a Master's degree in Informatics from the Graduate Program in Informatics (PPGI) at UFPB. He is currently a PhD student at the Federal University of Pernambuco (UFPE) and a Professor at the Federal University of Paraíba (UFPB). He has experience in Computer Science, with an emphasis on Digital TV (ginga middleware, interactive applications, multiplexing), Digital Image Processing and Computer Vision.

Tiago Maritan Ugulino de Araújo has an undergraduate and Master's in Computer Science from the Federal University of Paraíba (UFPB), and a PhD in Computer Engineering from the Federal University of Rio Grande do Norte (UFRN). He is an Associate Professor and Vice Director in the Informatics Center at UFPB and a researcher at the Digital Video Applications (LAViD). He coordinates the Suite VLibras project (vlibras.gov.br), an open platform for machine translation of Brazilian Sign Language (Libras) content for Digital TV, Web, Cinema, and mobile devices, installed on more than 120,000 websites, including the Federal Government, Federal Chamber, Senate Federal. He has experience in the areas of Accessibility in Computer Systems, Assistive Technology, Distributed Multimedia Systems, and Digital TV, working mainly on the following topics: machine translation for LIBRAS, automatic generation of audio description, and Digital TV.

Guido Lemos is a full professor in the Computer Systems Department of the Informatics Center at the Federal University of Paraíba (UFPB) and holds a PhD in Informatics from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). He coordinates LAVID (Center for Research and Extension in Digital Video Applications) where he develops research on the following topics: digital television, digital cinema, distributed multimedia applications, video distribution networks, distributed artistic performances, accessibility, information security, fake news, telehealth and blockchain applications. He has worked on the development of the Ginga middleware, published as ITU-T and ITU-R recommendations. Other research results include the development

set.org.br/ijbe doi: 10.18580/setijbe.2023.1 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

of a 4K 3D video storage, transmission and display system called Fogo Player, the development of video servers for live and on-demand transmission, called DLive and DVod, which were used in RNP's Digital Video Network and in USP-SP's IPTV service, the VLibras accessibility software; the development of technologies for the registration, validation and preservation of Digital Diplomas based on blockchain, which will be used in 270 Brazilian public universities; and finally, the development of the V4H health video system, which uses digital signature, blockchain registration and preservation technologies to add security to the use of videos generated during consultations. He is also a member of the Deliberative Council of the Brazilian Digital Television System Forum and a guest of Ancine's Technical Chamber for Accessibility.

Received in 2023-06-06 I Approved in 2023-07-08

set.org.br/ijbe doi: 10.18580/setijbe.2023.1 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

An Overview of Audio Technologies, Immersion and Personalization Features envisaged for the TV3.0

Regis Rossi Alves Faria
Almir Antônio Rosa,
Eduardo Mendes,
Ana Amélia Benedito Silva,
Douglas Henrique Siqueira Abreu,
Henrique Rozena

Cite this article:

Faria, Regis Rossi Alves; Rosa, Almir Antônio; Mendes, Eduardo; Silva, Ana Amélia Benedito; Abreu, Douglas Henrique Siqueira; Rozena, Henrique; 2023. An Overview of Audio Technologies, Immersion and Personalization Features envisaged for the TV3.0. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2023.2. Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING - SET IJBE V.9, 2023, Article 2, 15 p
2023 SET - Brazilian Society of Television Engineering / ISSN (Print): 2446-9246 / ISSN (Online): 2446-9432

An Overview of Audio Technologies, Immersion and Personalization Features envisaged for the TV3.0

Regis Rossi Alves Faria, Almir Antônio Rosa, Eduardo Mendes,
Ana Amélia Benedito Silva, Douglas Henrique Siqueira Abreu, Henrique Rozena

***Abstract*—In 2021 the Forum of the Brazilian Digital Terrestrial Television System (SBTVD) accomplished the phase 2 of the TV3.0 project, consisting of a series of tests of the technologies proposed for this next generation of TVD system in Brazil. The tests were conducted by research groups of Brazilian universities. Particularly referring to the audio coding layer of the system, we carried out at the University of São Paulo 13 groups of tests, as prescribed in a public Call for Proposals (CfP), and could assess the technologies capabilities and versatility in providing a series of new features for the next generation of audio for the digital broadcasting system. This paper summarizes the main results of this testing and evaluation phase, and brings an overview of the stimulating new features that content producers and audience would have available to create and consume immersive and personalized services at home.**

***Index Terms*— Next-generation Audio, Immersive Audio, Audio Coding, Audio Personalization, TV3.0**

I. Introduction

THIS paper aims at presenting an overview of the results of the evaluations and tests carried out on the candidate technologies proposed for the audio coding layer of the next-generation of DTV (TV 3.0 system), detaching their distinguished features on audio immersivity and content personalization. Two of them were tested in laboratory and could have their features evaluated and their compliance verified against the requirements defined in the public Call for Proposals (CfP) issued by the Forum of the Brazilian Digital Terrestrial Television System (SBTVD) [1]. The methodology included the execution of 13 tests groups, applied to each of the technology systems, and employing a common set of audio test content material, so as to permit assessment of performance and conformance issues based on a common ground.

The paper is to be organized as follows: first we present the 13 tests groups, their main features and technical requirements. Then, we present an overview of the technologies considered, and the audio test content employed in the tests. Following, the test cases performed in the laboratory are described, illustrating the facilities, employed setups, operating steps and conditions. Next, considering the fulfilled features and requirements, we give a glance on the potential of the new services expected in terms of audio immersion and content personalization control capabilities that the sound in the TV3.0 will offer, and finish with some pertinent discussions for the present and future.

II. Tests groups

The tests groups concerning the audio coding layer performed in the Testing and Evaluation phase of the TV3.0 SBTVD project were defined in the Phase 2 Call for Proposals (CfP) document first issued on December, 2020 by the Forum SBTVD and revised on March 15, 2021 [1]. The document describes 13 groups of Test Cases (TC) and the respective requirements (AC) for each Test Case, detailing the minimum technical specifications to be fulfilled. The tests were numbered from 1 to 13.

In this section we describe the Test Cases (TC), their requirements (AC), and the main features addressed to be evaluated, as specified in the CfP.

1. Test 1 (Immersive audio)

Test 1 addressed three TCs. In TC1.1, the requirements to the tested system were to demonstrate its ability to present audio in the specified channel mode, and to present the audio objects rendered together to various output setups (e.g., 2.0, 5.1, and 5.1+4H channels), in accordance with the target bitrate for this TC.

In TC1.2, the requirement was to demonstrate the system's ability to present scene-based HOA (Higher-Order Ambisonics) content in accordance with the target bitrate for this TC.

2. Test 2 (Interactivity and personalization)

Test 2 consisted of seven TCs. For TC2.1, the requirement "AC2.1: Language Selection" addressed the system's ability to allow end-users to select between multiple audio languages based on user interaction or automatic language selection (e.g., the receiver's preferred audio settings).

For TC2.2, the requirement AC2.1 (Selection of different preselections) targeted the system's ability to allow (or not) end-users to select between different preselections created in

This work was supported in part by CNPq, the Brazilian National Council for Scientific and Technological Development.

Regis Faria (regis@usp.br), Almir Almas (alalmas@usp.br), Ana Amélia B. Silva (aamelia@usp.br) and Eduardo Mendes (edusm@usp.br) are with the University of São Paulo.

Douglas Abreu (d229998@dac.unicamp.br) is with the School of Electrical and Computing Engineering, Universidade Estadual de Campinas, Campinas, Brazil.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

the production. TC2.3 evaluated the system's ability to switch between multiple commentators (e.g., during a sports event the user at home could switch between the usual commentator and the premium commentator or local team commentator).

For the Test Case TC2.4 the requirement (AC2.1: Display of textual labels) concerned the system's ability to display to the end-users correct textual labels for all audio objects that allow interactivity options and preselections as created in production. Test Case TC2.5 evaluated the system's ability to enable the end-users to interact with any audio object and adjust the object level at the end-user's device according to broadcaster settings in production (requirement AC2.2: Audio object loudness interactivity changing the level relative to the background). The user should be able to increase/decrease the object level (relative to the background) inside a range of min/max gain specified by the broadcaster, which might differ for each object.

The TC2.6 evaluated the requirement "AC2.3: Audio object interactivity, changing the object position". The test consisted of demonstrating the system's ability to enable the end-users to interact with any object and to adjust the sound object position at the end-user's device, according to broadcaster settings.

The last TC in this group – TC2.7 – addressed the requirement "AC2.4: Enable Interactivity when using external sound reproduction systems". It aimed to demonstrate the system's ability to enable interactivity when using external sound reproduction devices (e.g. soundbar/AVR, home theaters). The tests should demonstrate the system's ability to enable the interactivity options on the main receiving device (e.g., TV/STB) while the immersive sound is reproduced by the external sound reproduction device.

3. *Test 3 (Audio description)*

Test 3 consisted of four Test Cases addressing the selection and use of audio description content, delivered as an additional audio object with associated metadata.

For the TC3.1, the tested requirements were "AC3.1 and AC3.2: Audio description in the same stream as the main audio". It aimed to demonstrate the system's ability to enable audio description delivered in the same stream as the main audio (e.g., a single stream containing the main audio mix and alternative mix with audio description).

Test Case TC3.2 (requirement "AC3.3 Part 1: Audio description delivered as an additional audio object") consisted in demonstrating the system's ability to enable the audio description service, when available, and Test Case TC3.3 (requirement "AC3.3 Part 2: Audio description delivered as additional audio objects and language selection") aimed to demonstrate the end user's ability to enable/disable audio description available in multiple languages.

For the Test Case TC3.4, the requirement was: "AC3.3 Part 3: Audio description delivered as additional audio objects and spatial separation of main dialog and audio description". This test should demonstrate the system's ability to enable/disable audio description and spatially separate the main dialog and the audio description for better speech intelligibility".

4. *Test 4 (Audio emergency warning information)*

The test 4 consisted of only one TC (TC4.1) which should demonstrate how the audio system can deliver emergency warning information audio content. The test concerned showing what metadata is carried in the audio bitstream and how other applications could access or process this metadata to achieve the same result (e.g. some sort of API specification).

5. *Test 5 (Flexible audio playback configuration)*

For Test 5 (Flexible audio playback configuration), only one TC addressing the requirements "AC5.1 and AC5.2". The test should demonstrate the system's ability to decode and render the same content using multiple audio playback configurations and systems, including TV loudspeakers, soundbars, home theaters (immersive and 5.1 AVRs), and binaural.

6. *Test 6 (Consistent loudness)*

For Test 6, three TCs were conducted. TC6.1 addressed the requirement "AC6.1: Loudness Normalization Test - Programs" and the test should demonstrate the system's ability to achieve the target loudness level across multiple programs, i.e., to evaluate the effectiveness of solutions for guaranteeing a consistent loudness experience, without undesired (and sometimes exaggerated) volume changes between different programs.

TC6.2 addressed the requirement AC6.2 (Loudness Normalization Test for Preselections) and should demonstrate the ability to preserve the target loudness level across multiple preselections inside the same program. TC6.3 (requirement AC6.2: Loudness Compensation Test) should demonstrate the ability to preserve the target loudness level after user interaction (e.g., if the user increases the level of the dialog the overall loudness shall not increase).

7. *Test 7 (Seamless configuration changes and Audio/Video alignment)*

For the Test 7, seven TCs were evaluated. Test Case TC7.1 addressed the requirement "AC7.1: Seamless configuration changes" and considered demonstrating the system's ability to seamlessly play back content during configuration changes. Configuration changes between available configurations could include, for instance, combinations between 2.0, 5.1, and 5.1+4H output formats.

TC7.2 (requirement "AC7.2: Seamless content playback during user interaction") should demonstrate the ability to seamlessly playback content during user interaction, such as changes between different audio languages or preselections, increasing or decreasing the level of various audio objects, without audio drop-outs or glitches.

TC7.3 (requirement "AC7.3 Part 1: Seamless content playback during changes in production") should demonstrate the system's ability to seamlessly playback content during changes in production during a live broadcast. This test employed a special live setup, with equipment and software to permit a live edition of the broadcasting settings. Typical changes in a live broadcast should be tested, including:

a. Change the audio scene (objects, preselections, etc.);
b. Enable/disable dialogs in multiple languages;

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

c. Enable/disable Audio Description in multiple languages;

d. Enable/disable interactivity options for one or more preselections;

e. Change the interactivity options (min/max gain and position values) for one or more objects;

f. Change the textual labels for one or more objects or preselections.

TC7.4 (requirement AC7.3 Part 2: Seamless content playback during changes in production using a contribution feed) should demonstrate the system's ability to seamlessly playback during changes in production in a live broadcast scenario.

TC7.5 (requirement "AC7.4 Part 1: Seamless Ad-Insertion") should demonstrate the system's ability to enable seamless advertisement insertion at any time instance, e.g., switch between the main feed authored live (e.g. a content playout 1) and an additional feed containing a pre-authored advertisement break (e.g. a content playout 2).

TC7.6 (requirement "AC7.4 Part 2: User selection persistency after the Ad-break") should demonstrate the ability to preserve the user interaction settings after the ad-break, e.g., if the user selects, before the ad- break, the English language (EN) and increases the dialog level with 7 dB, after the ad-break the content will start with the exact same settings.

TC7.7 (requirement "AC7.4 Part 3: Hybrid Delivery") should demonstrate the system's ability to synchronize and replace the main soundtrack delivered via broadcast for an alternative audio signal delivered via broadband (Internet link).

8. Test 8 (Audio coding efficiency)

For Test 8 (Audio coding efficiency), the CfP document specified that "the proponent's documentation provided on the Quality Assessment Reports should provide the data to analyze the audio coding efficiency". Therefore, this TC focused on the analysis of available technical assessments and previous studies, such as subjective tests, conducted by third parties [2-6], mainly to evaluate the system's ability to deliver the minimum MUSHRA (ITU-R BS.1534-3 Multiple Stimuli with Hidden Reference and Anchor, a subjective test methodology) quality scores for several audio formats and target bitrates

9. Test 9 (Audio End to end latency)

Test 9 considered two requirements to evaluate the system's ability to provide live audio with minimum end-to-end latency.

The requirement AC9.1 was specified to be tested during the execution of Test Case TC1.1, and considered if the proponent's system was able to encode and decode according to the requirements AC1.1.1, AC1.1.2, and AC1.1.3.

The requirement AC9.2 was not to be analyzed with a feature test, and should be verified through the analysis of the proponent's documentation provided in the Document Analysis phase. As specified in the CfP, the delay (latency) of each module of the real-time test setup should be documented, including the audio and video encoding delay, additional video buffering (if any) before the video encoder, audio decoding and rendering delay, transcoding to a different format delay, and final decoding delay in the external sound reproduction system.

10. Test 10 (Audio/Video synchronization)

Test 10 consisted of only one TC. TC10.1 addressed the requirement "AC10.1" targeted to demonstrate the system's ability to perform adequate A/V synchronization.

11. Test 11 (New immersive audio services)

For Test 11, the single Test Case TC11.1 addressed the requirement "AC11.1" and the test task was to verify the system's ability to perform playback of audio demonstrating one or more of those applications: VR / AR / XR / 3DoF (Degree of Freedom) / 6DoF. A video codec should be chosen by the proponent to be used in this test.

12. Test 12 (Interoperability with different distribution platforms)

Test 12 consisted of a single TC (TC12.1) whose requirements were AC12.1 and AC13.1. The test aimed to demonstrate the system's ability to send multiple audio contents over two or more communications channels.

13. Test 13 (Audio scalability and extensibility)

For Test 13 the CfP specified two requirements: "AC13.1" (for scalability) and "AC13.2" (for extensibility). The use test addressed the system's ability to enable scalability (e.g. to enhance the over-the-air audio experience with additional Internet-delivered audio content, such as new sports commentator options) and extensibility (e.g. support new settings and/or features in the future, in a backward-compatible way).

III. TECHNOLOGIES IN CONSIDERATION

Three international audio coding standards responded to the Call for Proposals (CfP) of technologies and were accepted as candidates to supply their systems for the audio component/layer of the TV3.0 system.

This section provides a concise introduction to the three audio coding technologies taken into consideration in the SBTVD TV3.0 testing and evaluation phase: the AVSA; Dolby Atmos (AC-4); and MPEG-H.

A. AVSA system

The acronym AVSA stands for AVS-Audio, the IEEE 1857.8-2020 Standard for Second Generation Audio Coding, also known as the audio stream that matches AVS2 audio standard [7]. This standard defines a set of tools for the compression, decompression, and packaging of multimedia data, aimed at efficient transmission and storage over the Internet. This standard, an evolution of the IEEE Std. 1857.2-2013, provides a flexible configuration of compression parameters to deliver an improved Quality of Experience (QoE). This includes bitrates ranging from 16 kb/s to 192 kb/s per channel, and supports up to 128 channels for audio signals with a sampling frequency from 8 kHz to 192 kHz and quantization resolutions of 8, 16, and 24 bits.

It presents a defined set of audio encoding tools for the transmission and decoding of recorded music, voice,

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

environmental sounds, and instrumentals. The framework of AVS2 (Audio Video Coding Standard 2) is divided into two profiles: the base channel encoding profile (base_profile) and the 3D audio object encoding profile (3D_profile). The 3D object encoding includes object audio data and metadata, allowing for spatial configuration, movement localization, and additional descriptive information such as acoustic properties, directional cues, and volume levels.

AVS2 distinguishes itself by employing adaptive bit rate control and advanced psychoacoustic models to achieve high compression efficiency without compromising sound quality. It offers encoding options for both channel signal and audio object, allowing a flexible configuration between 128 sound objects and 128 channel signals. Furthermore, the GA (General Audio) encoding technology provides multiple encoding options that share a common core module, thus achieving high efficiency and low algorithmic latency.

When compared to its predecessor, AVS2 has improved the degree of 3D audio encoding, achieving a greater compression efficiency and sound quality, and saving up to 50% in bit rate. This improvement makes the standard especially useful for applications and services that include audio accompanying video, TV audio systems, digital audio storage, audio broadcasting, and communication. Hence, the standard is suitable for high-resolution digital broadcasting, digital storage media, broadband wireless multimedia communications, broadband Internet media streaming, digital cinema, and video surveillance.

In the context of the IEEE 1857.8-2020 standard, the data flow of the AVSA system is delineated in several steps, ranging from audio input to encoded audio output. A schematic for AVSA is shown in Fig. 1. Specifically, in addition to the AVS2 audio encoding profiles, the General Audio (GA) Encoding Technology also forms part of its ecosystem. AVSA is currently in use in broadcast services in China.

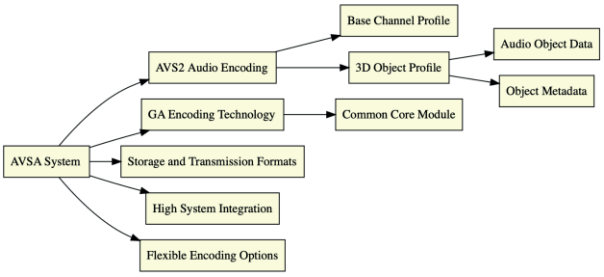


Fig. 1: Description diagram of the AVSA system as defined by the IEEE 1857.8-2020 standard.

B. AC-4 (Dolby) system

The Dolby AC-4 Audio Codec is defined in the technical specification ETSI TS 103 190 (Digital Audio Compression AC-4 Standard) published by the European Telecommunications Standards Institute (ETSI) [8], and represents an innovation in encoding efficiency and application versatility over the previous AC-3 format. It proposes several advancements in audio encoding and also offers flexibility and efficiency for a broad range of applications and output devices. As specified in ETSI TS 103 190-2 (Part 2: Immersive and personalized audio), this codec is designed to support a diverse range of content types, including legacy channel-based, object-oriented immersive, and customized audio [9]. Additionally, AC-4 is compatible with the ATSC (Advanced Television System Committee) 3.0 audio system, leveraging several specific features to enhance the quality and efficiency of audio transmission.

One of the highlighted features of AC-4 is the A/V frame alignment, aimed at mitigating complications associated with the synchronization of multimedia content at segmentation points. When enabled, this feature claims significant simplification on splicing workflows and transcoding to or from formats that utilize video-based frame alignment, such as HD-SDI (Serial Digital Interface, introduced by the Society of Motion Picture and Television Engineers - SMPTE).

The AC-4 system also introduces improvements in dialogue intelligibility through a user-controlled dialogue enhancement feature. Moreover, AC-4 incorporates support for the Extensible Metadata Delivery Format (EMDF), as defined in ETSI TS 103 190-1 (Part 1: Channel based coding). This format allows the transmission of third-party metadata and application data in AC-4 bitstreams, offering a framework for the inclusion of additional user data.

Regarding control of dynamic range and volume, AC-4 adheres to global standards, incorporating an extensive set of volume metadata and Dynamic Range Control (DRC) that are compliant with international norms, including ATSC A/85 [10]. This allows for flexible DRC implementation, adaptable to a wide range of device profiles and user application scenarios.

Another advancement of AC-4 is its volume verification engine, which ensures the accuracy of the transmitted volume metadata. This feature, combined with a real-time volume leveler, can be activated to guarantee audio output consistency.

The AC-4 system offers an array of innovative features for efficient encoding and application versatility, as depicted in Fig. 2. It includes features like A/V frame alignment and DRC, among others. AC-4 has been selected as the Next-Generation Audio (NGA) audio format for the United States, Canada and Mexico, formalized in the ATSC A/300 document.

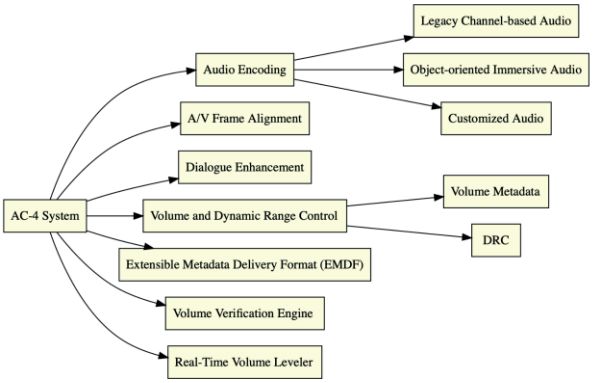


Fig. 2: Description diagram of the Dolby AC-4 system, as specified in ETSI TS 103 190-2.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

C. *ISO/IEC MPEG-H 3D Audio system*

The MPEG-H audio system represents an advanced paradigm in NGA systems. This innovation is formalized as an open international standard under ISO/IEC 23008-3, also known as MPEG-H 3D Audio [11][12]. One of the standout features of this audio system is its declared ability to heighten realism, allowing sound to come not only from the sides but also from above and below the listener. This multi-dimensional experience is further enriched by interactivity features that enable viewers to customize their auditory experience by choosing from various predefined audio presentations—referred to as Presets—or making manual adjustments to audio elements.

The interactive potential is particularly evident in the system's integrated renderer and the advanced management of dynamic range and volume, which optimize content playback according to the capabilities of the playback device. This facilitates seamless audio content delivery across a variety of devices, ranging from headphones to high-quality speaker systems, making it a versatile tool in content creation.

The roots of MPEG-H trace back to the Moving Pictures Experts Group (MPEG), a joint initiative of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC), also responsible for the introduction of the popular MP3, DVD and MP4 multimedia formats. This audio compression standard evolved through a competitive and collaborative process involving experts in audio coding technology. MPEG-H audio encapsulates an intricate combination of highly efficient encoding technologies, the ability to represent audio in three different formats – channel-based, object-based, and scene-based – and advanced volume and dynamic range control (DRC).

Another notable element in the MPEG-H audio system is its classification into profiles and levels. Profiles are essentially a subset of available tools tailored for specific applications. For instance, the High profile incorporates all features and is essentially a theoretical construct. Low complexity and Baseline profiles are more targeted, with the former including additional encoding tools for specific applications such as Virtual Reality (VR) or Augmented Reality (AR). Levels introduce additional parameters that allow for finer tuning of these tools.

MPEG-H has gained international acceptance and has been included in many standards, such as ATSC 3.0, TTA in South Korea, and SBTVD in Brazil, as well as in 3GPP for 360° video streaming services over 5G (5th generation of mobile network).

The MPEG-H 3D Audio system offers an elaborate combination of functions, as illustrated in Fig. 3, whose features are determined by the profiles and levels used. In addition to being available on digital TV receivers, MPEG-H decoding and rendering is currently available in a variety of equipment, including AVR's and soundbar systems.

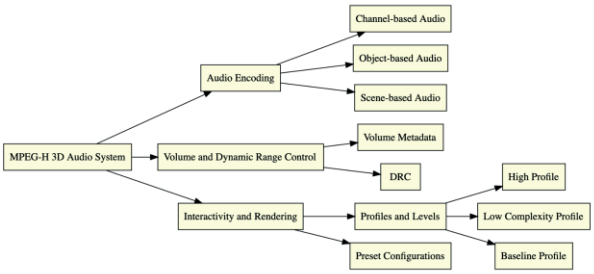


Fig. 3: Description diagram of the MPEG-H 3D Audio system, in accordance with the international ISO/IEC 23008-3 standard.

IV. AUDIO TEST CONTENT

The mandatory tests used audio test content items provided by four different sources, and included audio (wav) files, video (mov) files, and metadata (xml) files describing the audio program organization within the (sound items) payload. After verifying the provided material, 24 (twenty four) audio test content items were validated to be used in the test cases.

The audio test content items consisted of a rich set of types covering several channel setup organizations, program combinations, audio material in different languages, audio-description and emergency warning information. Twelve (12) program file types were defined in the CfP, with different program structures, number of channels, sound content, and bitrates.

In this section we present the 12 test content types used, describing the kind of content of specific items (e.g. what was in the scene, types of sound included) and describing the file type organization in terms of their ADM (Audio Definition Model) structure.

A. *An overview of the Audio Definition Model (ADM)*

The ITU Audio Definition Model (ADM) is a more recent standard proposed in 2019 for audio (scene and file) description in the Recommendation ITU-R BS.2076 [13]. The scheme has been connected to so-called "Next Generation Audio" (NGA) and found support in the broadcast community. In its construct, the audio elements are grouped within the file according to a hierarchy of audioProgrammes, audioContents and audioObjects, as shown in Fig. 4.

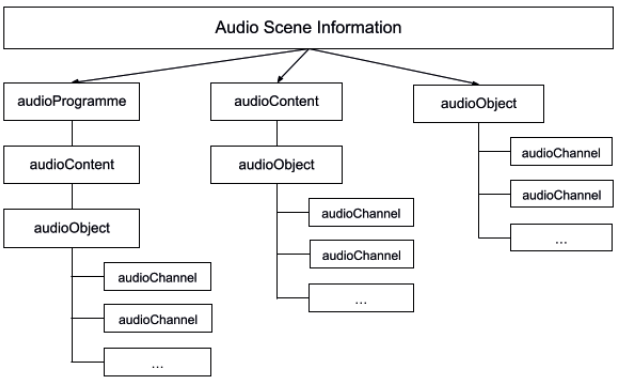


Fig. 4: ADM file structure, according to ITU-R BS.2076.

Audio programmes and audio contents have attributes such as name and language. Audio programmes bring all the audio contents together: they combine all the contents to make the

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

complete ‘mix’.

An audioProgramme may contain, for instance, an audioContent for ‘narrator’ and another one for ‘background music’. Or, in another example, an audioProgramme may contain an audioContent for English speakers, called ‘dialogue-en’, together with a ‘backgroundMusic’ content, and another audioProgramme may be prepared for Portuguese speakers, which contains a ‘dialogue-port’ audioContent and the same ‘backgroundMusic’.

The objects are effectively sound source elements, they will have for example attributes such as azimuth, elevation and distance to describe the location of the sound in the scene.

The ADM model is divided into two sections: the content part, and the format part. The content part describes what is contained in the audio, describing things such as the language of any dialogue, the loudness, etc. The format part describes the technical nature of the audio (e.g. the formats of the tracks and/or streams) so it can be decoded/rendered correctly.

Several types of audio are possible, for example: a conventional track (e.g. front-left track); a HOA component; a group of channels, and so on. The type of audio stream will define which channels are inside. There may be audio channels associated to DirectSpeakers (which will then associate a specific speakerLabel), or to a HOA pack, a Matrix, or a Binaural set [13].

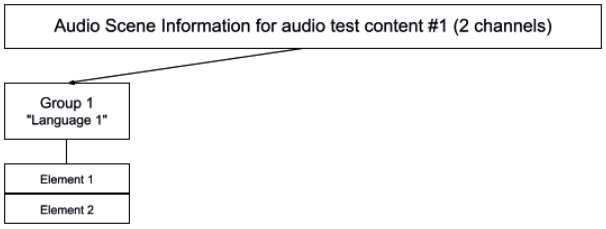
B. Tested audio content types

The following paragraphs present the 12 file types used in the tests, with examples of specific validated items of each type. The presentation of each file type begins by numbering the programs/contents included, with a brief description of their content, and information on the number of channels and bitrate of the file.

The rationale behind the segregation and/or grouping of elements throughout distinct programs, in different multichannel sets, is in facilitating their selection and access for experience personalization. It should be noted that Test Cases therefore derive from what each content organization can offer, and the content selection and personalization that might be possible, with each type, depends on that organization

1. *1: Stereo Mix (Language 1) - (2 Channels / 48kbps)*

This file type has the following ADM structure:

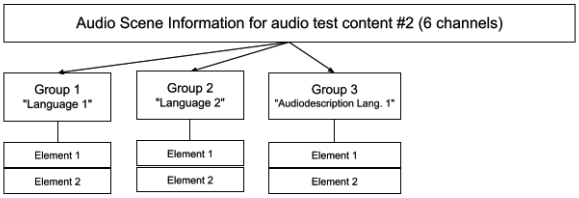


Three specific test items were used:

- 1.1 Aphorism on nature. Content: Nature documentary featuring a Portuguese narrator discussing the importance of nature. It includes ambient sounds (e.g. running water, footsteps, birds, bats) and background music.
- 1.2 Mountain bike. Content: GoPro-style biking trail recording, capturing the biker's breathing, the wheel against the ground, and wind sounds.
- 1.3 Trains passing by. Content: Urban subway footage with varying low and high sound levels.

2. *1: Stereo mix (Language 1) + 2: Stereo mix (Language 2) + 3: Stereo Audio Description (Language 1) - (6 channels / 144kbps)*

This file type has the following ADM structure:

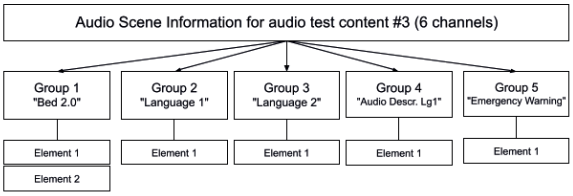


Two specific test items were used:

- 2.1 Aphorism on nature. Content: Two stereo mix contents with different narrators (one in Portuguese, one in English) backed by environment sounds (steps in a cave, trees being hit by the wind and a torch) and classic music playing as background soundtrack. One stereo audio description content (in Portuguese).
- 2.2 Phoenix - German & French. Content: European post-war movie with two stereo mixes (French and German dialogue), and background natural sounds like birds chirping and gravel footsteps.

3. *1: Channel Bed 2.0 + 2: Language 1 Mono + 3: Language 2 Mono + 4: Audio Description Mono (Language 1) + 5: Emergency Warning Information Mono - (6 channels / 192kbps)*

This file type has the following ADM structure:



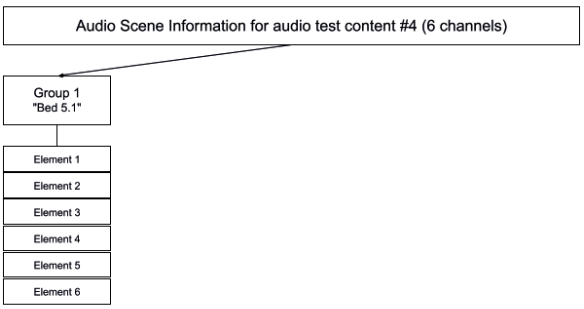
Two specific test items were used:

- 3.1 Aphorism on nature. Content: Similar to 2.1 with nature documentary content and three narrators.
- 3.2 4ever. Content: Short video clip for a French television company, featuring a city bell, classical music, and multiple languages.

4. *Channel Bed 5.1 - (6 channels / 144kbps)*

This file type has the following ADM structure:

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

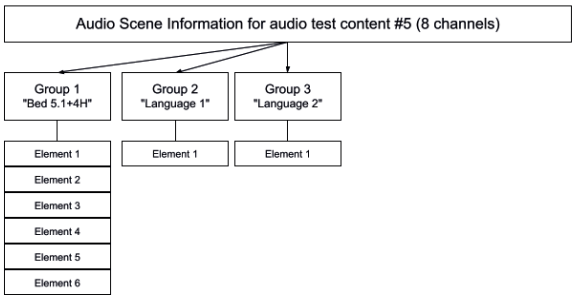


Three specific test items were used:

- 4.1 Aphorism on nature. Content: Nature documentary with Portuguese narration. Features similar ambient sounds as prior examples.
- 4.2 Mountain bike. Content: GoPro-style biking trail recording, similar to content 1.2..
- 4.3 Record's report. Content: Five-minute news report, in Portuguese with background music.

5. *1: Channel Bed 5.1 + 2: Language 1 Mono + 3: Language 2 Mono - (8 channels / 240kbps)*

This file type has the following ADM structure:

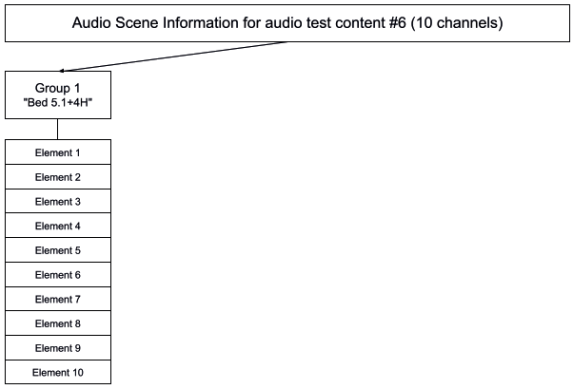


Two specific test items were used:

- 5.1 Aphorism on nature. Content: Nature documentary with dual narration in Portuguese and English translation, featuring similar ambient sounds.
- 5.2 One day in berlin. Content: Daily life in the city during summer with Portuguese and German narrations.

6. *Channel Bed 5.1+4H - (10 channels / 256kbps)*

This file type has the following ADM structure:

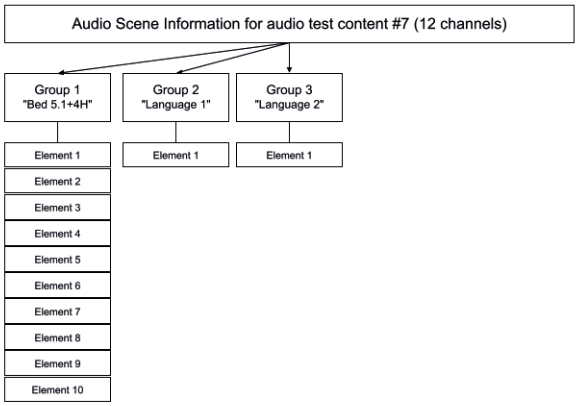


Three specific test items were used:

- 6.1 Aphorism on nature. Content: Features ambient nature sounds without narration.
- 6.2 Mountain bike. Content: GoPro-style biking trail recording.
- 6.3 Eurovision Sweden. Content: Audio from the Eurovision Israel 2019 event featuring piano, opening music, a singer, and a crowd.

7. *1: Channel Bed 5.1+4H + 2: Language 1 Mono + 3: Language 2 Mono - (12 channels / 352kbps)*

This file type has the following ADM structure:

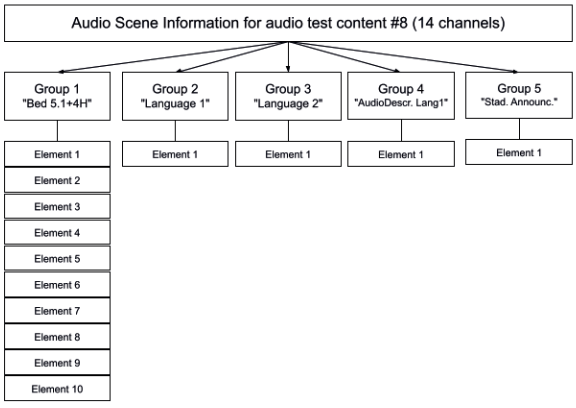


Two specific test items were used:

- 7.1 Aphorism on nature. Content: Nature documentary with ambient sounds.
- 7.2 Le Mans Astray. Content: First half features a high-speed car and traffic sounds; the second half transitions to a nature documentary.

8. *1: Channel Bed 5.1+4H + 2: Language 1 Mono + 3: Language 2 Mono + 4: Mix Stereo (Language 1) + 5: Mix Mono (Language 2) - (15 channels / 448kbps)*

This file type has the following ADM structure:



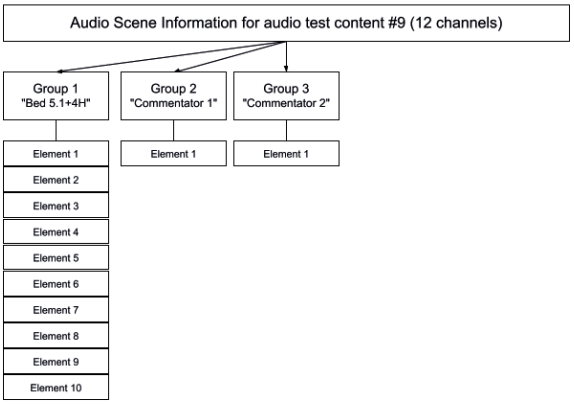
Two specific test items were used:

- 8.1 Aphorism on nature. Nature documentary with ambient sounds.
- 8.2 European championship Berlin/Glasgow 2018. Audio from a hurdle race championship featuring opening music, narration, and crowd sounds.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

9. *1: Channel Bed 5.1+4H + 2: Commentator 1 Mono + 3: Commentator 2 Mono - (12 channels / 352kbps)*

This file type has the following ADM structure:

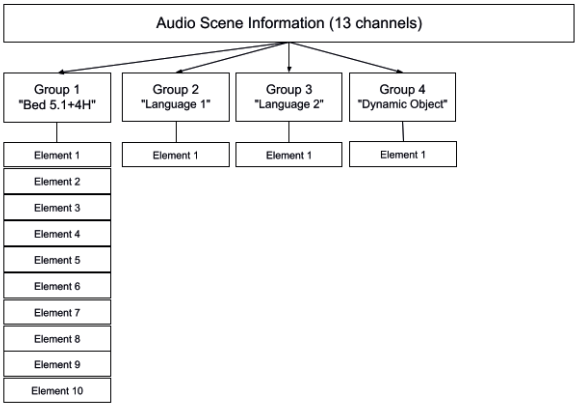


Two specific test items were used:

- 9.1 Rio de Janeiro carnival. Content: Samba school performance featuring two singers and instruments.
- 9.2 Carnaval 2020 Rio de Janeiro. Content: Audio from a Carnival Parade in 2020, featuring samba music.

10. *1: Channel Bed 5.1+4H + 2: Language 1 Mono + 3: Language 2 Mono + 4: Dynamic object Mono - (13 channels / 400kbps)*

This file type has the following ADM structure:

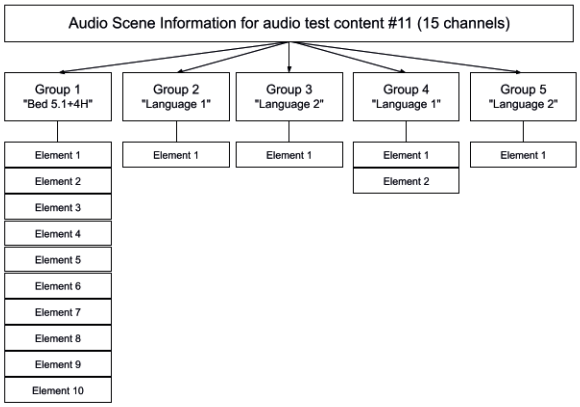


One specific test item was used:

- 10.1 Le Mans Astray (5.1+4H). First half features a high-speed car; the second half is a nature documentary.

11. *1: Channel Bed 5.1+4H + 2: Language 1 Mono + 3: Language 2 Mono + 4: Mix Stereo (Language 1) + 5: Mix Mono (Language 2) - (15 channels / 448kbps)*

This file type has the following ADM structure:



Two specific test items were used:

- 11.1 Aphorism on nature. Content: Nature documentary with ambient sounds.
- 11.2 European song contest Lisbon 2018. Content: features crowd sounds, music, and special effects.

Fig. 5 illustrates the audio test file 11.1, showing its 15 tracks.

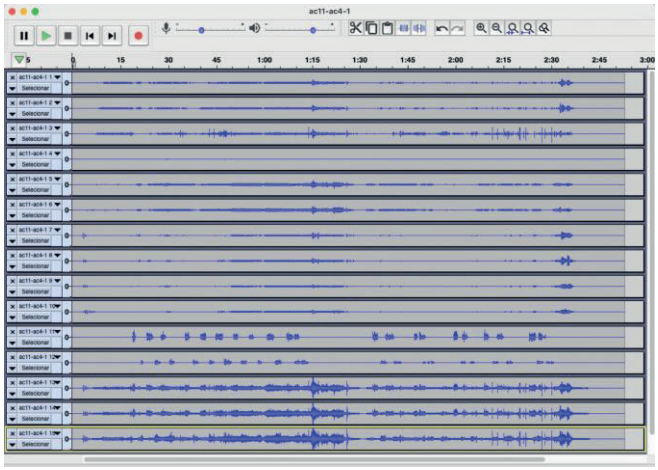
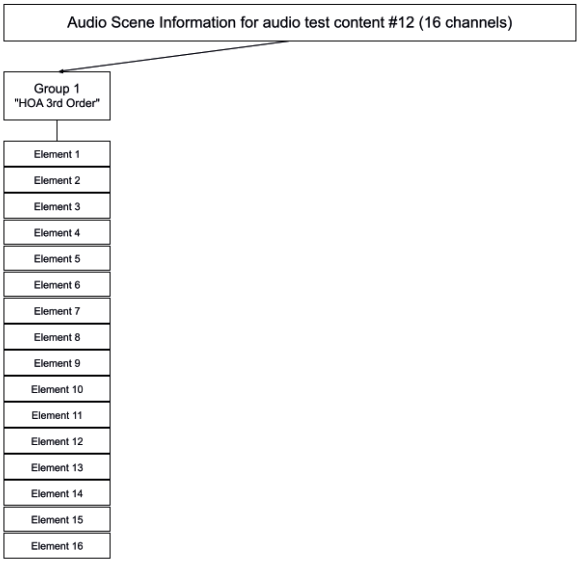


Fig. 5: Example of a file of type #11 with 15 channels.

12. *1: HOA (3rd order) - (16 channels / 320kbps)*

This file type has the following ADM structure:



https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

One specific test item was used:

- 12.1 Clouds of franconia. Content: Soundtrack introduction with guitar and drum (HOA formatted with ACN channel ordering, SN3D normalization).

V. Assessment and Testing Methodologies

The testing and evaluation phase 2 lasted six months (from July to December 2021) and was aimed at assessing the candidates' technologies in order to verify their compliance to the specified requirements and evaluating their performance in laboratory tests, conducted with a set of audio test content and conditions as close as possible to expected operation conditions at future broadcasting service.

As specified in the CfP methodology, the first evaluation stage was to analyze each candidate's technology documentation and technical specifications to verify theoretically if they fulfilled the mandatory requirements. All three candidate technologies properly submitted their documentation which were evaluated at this stage, also benefiting from additional documentation provided by standardization institutions and known previous studies [2-6].

Following this stage, the next one was to conduct laboratory tests with physical hardware and software provided by the candidates, in order to assess the actual functioning of the systems and the fulfillment of the requirements under operating conditions. In this stage, only 2 of the candidates submitted equipment and software and took part in the tests.

A. Laboratory setups

In this section we briefly describe the laboratory setups, presenting the studio and test room environments, delivery interfaces and equipment employed. Particularly of interest, is the audition test room prepared for simulating a common domestic audition environment, where, in addition to the TV set loudspeakers, we used an external 5.1+4H loudspeaker setup and a soundbar system to reproduce 2D/3D spatial sound-fields in the room (see Fig. 6). We also tested the immersive and functional capabilities using conventional stereo (2.0) reproduction and binaural setups using earphones, covering all required output layouts.

Fig. 6: Auditory test room and sound systems used in the tests.

Two scenarios for tests were considered, employing two different setup modes, as required by the CfP [1]: (1) real-time encoding/decoding setup, and (2) non-real-time encoding/decoding setup.

The real-time setup emulates a typical broadcast scenario, where the broadcast feed is authored in one location (e.g., event location or studio) and provided over a contribution link to the broadcast center where it is monitored and re-authored if needed, prior to delivery (over the air or over Internet). For this setup we employed two different locations: a simulated studio/broadcast facility, where it was possible to author metadata and prepare the broadcast feed; and the test room, where the end-users could experience the programs using appropriate receivers with decoders and renderer equipment and software. The live delivery between these facilities employed an IP-based link. Fig. 7 illustrates the real-time setup, as specified in the CfP [1].

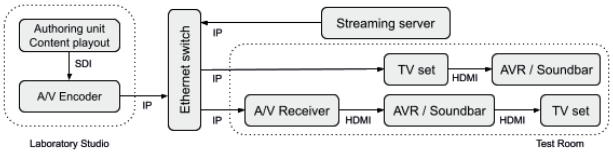


Fig. 7: Studio and Test room system setup for the real-time encoding/delivering/decoding.

The non-real-time setup, the transport layer between the audio encoder and the A/V receiver was specified to use an MP4 container format. All test content should be available as MP4 files stored in the receivers. For an end-user's setup with an A/V receiver, it should use an HDMI interface to feed the external sound system and demonstrate the audio system capabilities. The figure 8 illustrates the non-real time setup, as specified by the CfP.

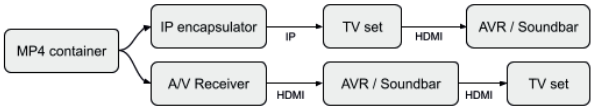


Fig. 8: Test room system setup for the non-real-time encoding/decoding. Sources in MP4 could be delivered to a TV set and external sound systems using IP and HDMI interfaces.

B. Test Cases performed

By conveniently breaking down the 13 test groups into a set of 46 test cases, and considering an average of (at least) 2 audio test content items (of 12 distinct types) used per TC, plus 2 technologies to evaluate, a minimum of 184 rounds of laboratory tests was required. Table I shows the expanded structure of the 46 Test Cases.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

TABLE I – EXPANDED TEST CASES

TC1.1.1 (immersive audio), AC1.1.1, audio 1 (2.0), real-time
TC1.1.2 (immersive audio), AC1.1.2, audio 4 (5.1), real-time
TC1.1.3 (immersive audio), AC1.1.3, audio 6 (5.1+4H), real-time
TC1.1.4 (immersive audio), AC1.2, audio 9, real-time
TC1.2 (immersive audio), AC1.3, HOA, audio 12, non-real-time *
TC2.1 (interactivity and personalization), AC2.1 language sel, audio 3/5/8, real-time
TC2.2 (interactivity and personalization), AC2.1 preselec, audio 8/11, real-time
TC2.3 (interactivity and personalization), AC2.1, coment switch, audio 9, real-time
TC2.4 (interactivity and personalization), AC2.1, labels display, audio 8/9/11, real-time **(AC08)**
TC2.4 (interactivity and personalization), AC2.1, labels display, audio 8/9/11, real-time **(AC09)**
TC2.4 (interactivity and personalization), AC2.1, labels display, audio 8/9/11, real-time **(AC11)**
TC2.5 (interactivity and personalization), AC2.2, loudness int, audio 8/9, real-time **(AC08)**
TC2.5 (interactivity and personalization), AC2.2, loudness int, audio 8/9, real-time **(AC09)**
TC2.6 (interactivity and personalization), AC2.3, object position, audio 8/10, real-time **(AC08)**
TC2.6 (interactivity and personalization), AC2.3, object position, audio 8/10, real-time **(AC10)**
TC2.7 (interactivity and personalization), AC2.4, int on external, audio 8/9/10, real-time **(AC08)**
TC2.7 (interactivity and personalization), AC2.4, int on external, audio 8/9/10, real-time **(AC09)**
TC2.7 (interactivity and personalization), AC2.4, int on external, audio 8/9/10, real-time **(AC10)**
TC3.1 (Audio description), AC3.1/AC3.2, AD in same stream/alternate mix, audio 2, real-time
TC3.2 (Audio description), AC3.3 P1, AD in additional audio, audio 3/8, real-time
TC3.3 (Audio description), AC3.3 P2, AD in addit audio and language, audio 3/8, real-time
TC3.4 (Audio description), AC3.3 P3, AD in addit audio and spatial, audio 3/8, real-time
TC4.1 (Audio emergency), AC4.1, audio 3, real-time
TC5.1 (Flexible audio playback), AC5.1/AC5.2, audio 8/11, real-time **(AC08)**
TC5.1 (Flexible audio playback), AC5.1/AC5.2, audio 8/11, real-time **(AC11)**
TC6.1 (Consistent loudness), AC6.1, loudness norm program, audio 8/11, non-real-time *
TC6.2 (Consistent loudness), AC6.2, loudness norm preselections, audio 8/11, non-real-time *
TC6.3 (Consistent loudness), AC6.2, loudness comp, audio 8/11, non-real-time *
TC7.1 (Seamless config changes), AC7.1, seamless config changes, audio 1/4/6/8, real-time
TC7.2 (Seamless config changes), AC7.2, seaml. playback user interact, audio 8/11, real-time
TC7.3 (Seamless config changes), AC7.3 P1, seaml. playback changes in product., audio 8/9/11, real-time
TC7.4 (Seamless configuration changes), AC7.3 P2, seaml. plbck chngs in prod. w/ feed, audio 8/11, real-time
TC7.5 (Seamless configuration changes), AC7.4 P1, seamless ad-Insertion, audio 1/11, real-time
TC7.6 (Seamless configuration changes), AC7.4 P2, user select persist. after ad-break, audio 1/11, real-time
TC7.7 (Seamless configuration changes), AC7.4 P3, hybrid delivery, audio 3, real-time
TC8 (Audio coding efficiency), AC8.1 kbps @ MOS 4 / MUSHRA > 80
TC9.1.1 (Latency), AC9.1 during TC1.1, AC.9.2 latency (ms)
TC9.2 (Latency), AC9.1 during TC1.1, AC.9.2 latency (ms)
TC10 (A/V Sync), AC10.1, adequate A/V sync, audio 3/5/7/9, real-time **(AC03)**
TC10 (A/V Sync), AC10.1, adequate A/V sync, audio 3/5/7/9, real-time **(AC05)**
TC10 (A/V Sync), AC10.1, adequate A/V sync, audio 3/5/7/9, real-time **(AC07)**
TC10 (A/V Sync), AC10.1, adequate A/V sync, audio 3/5/7/9, real-time **(AC09)**
TC11.1 (New immersive audio services), AC11.1, VR / AR / XR / 3DoF / 6DoF
TC12.1 (Interoperability), AC12.1 and AC13.1 w/ different distrib. plats, audio 8, real-time
TC13 (Scalability/Extensibility), AC13.1/AC13.2, during TC12

The actual number of test rounds executed, however, was nearly 10 times as much, considering that there were several audio test files of each type; rounds were done by different testers/evaluators; several real-time tests required changing attributes' values in real-time authoring; and that tests were repeated in order to verify the result consistency for a range of attributes' values and reproducibility of the operations.

For each test case, the respective requirements were evaluated and the results of the tests were to produce a classification in three possible outputs:

- "Fulfilled", if the proponent's system was fully able to satisfy the requirements;
- "Partially Fulfilled", if the proponent's system fails to satisfy part of the requirements;
- "Not Fulfilled", if the proponent's system fails to satisfy the requirements.

The next paragraphs describe how the TCs sessions were organized and conducted for each test group.

Test 1 (Immersive audio):

This test was divided in two parts, one aimed at channel-based setups (three TCs to test the system's ability to present audio in the specified channel modes 2.0, 5.1, and 5.1+4H channels and target bitrates during real-time set up) and one aimed at a scene-based setup (one TC to demonstrate the ability to present scene-based HOA content in the expected target bitrate of 20 kbps per HOA-channel, therefore a total HOA bitrate equal to 320 kbps) through non-real-time setup.

In both parts, the videos (with the audio content) were played continuously and they were heard through the AVR system and soundbar to evaluate the overall 3D experience.

Test 2 (Interactivity and personalization):

This test was divided into seven TCs. The first one examined the ability to select between multiple audio languages based on user interaction or automatic language selection through real-time set up. The system should allow authoring the metadata in the studio, enabling the desired personalization options. Then, during playback, these options should be displayed on the receiver side. This test also investigated the capability of the audio system in maintaining the chosen language when restarting the receiver.

The second TC evaluated the capability of the audio system to display all preselections authored in production; the capability of the end-user to manually switch between different preselections during live playback; and the capability of the audio system to correctly render the preselections on the receiver side.

The third TC examined the ability of end-users to switch between multiple available commentators, and evaluated the display on the receiver side of all available commentators authored in the production through real-time set up.

The fourth TC consisted in real-time authoring metadata at the studio for several audio elements (e.g., Dialogs, Commentators, Stadium Announcers) and several preselections (e.g., Main mix, Dialog+, Stadium). The authored signal was then encoded and in the test room it was evaluated the capability to display the textual labels for all preselections and audio objects allowing user interactivity as well as the capability of the audio system to correctly render the preselections

The fifth TC consisted of demonstrating the ability to interact with any audio object and adjust the level through real-time set up. It evaluated if the user was able to increase or decrease the object level (relative to the background) inside the range specified by the broadcaster during live playback.

The sixth TC evaluated if the end-user was able to move audio elements inside an area (space) specified by the broadcaster in real-time setup.

The seventh TC tested the interactivity options when using external sound reproduction devices through real-time set up (e.g., Soundbar/AVR). This TC tested if the end-user could interact with the audio scene (e.g., change an object level or position, change the preset) during live playback, and evaluated the immersive experience reproduced on the external sound device (e.g., objects moved in the 3D space should be perceived at specific positions according to the user interaction).

Test 3 (Audio description):

This test was divided into four TCs. The first two consisted of demonstrating the system's ability to display the available audio description in multiple languages (as authored in production) and to enable and to switch between audio description elements during live playback through real-time set up.

The third TC examined the ability to enable/disable audio description in multiple languages, as authored in production and the capability of the audio system to start the playback of the audio description in the authored stream.

The fourth TC evaluated the system's ability to enable/disable audio description and spatially separate the

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

main dialog and the audio description through real-time setup. It evaluated the capability of the audio system to correctly reproduce the main dialog and the audio description at the desired locations in each preselection during live playback according to the metadata authored in production.

Test 4 (Audio emergency warning information):

This test consisted of demonstrating audio emergency warning information presentation through real-time set up. During this test, it was verified the continuous playback of the content before, during, and after its delivery, and the capability of the audio system to signal the Emergency Information in the authoring system and the flexibility to control it (i.e., if the audio object should be active in all preselections or a dedicated preselection, should mute the main dialog or playback over the main dialog).

Test 5 (Flexible audio playback configuration):

This test evaluated the system's ability to present the same content on TV loudspeakers, soundbar, AVR connected to a 5.1 and 5.1+4H loudspeaker setups as well on headphones in real-time set up.

Test 6 (Consistent loudness):

This test was divided into three parts. The first one evaluated the ability to achieve the target loudness level across multiple programs through non-real-time set up. The audio content was decoded using the proponent software audio decoder to three different target loudness levels: -31, -24, and -16 LKFS and the loudness consistency across multiple test items was evaluated. The program loudness was measured according to ITU-R BS.1770-4 with a tolerance of +/-3 dB, using the FFMPEG tool. From the FFMPEG tool output results, the loudness measurement was the value of the parameter labeled as "Integrated Loudness" (I), in LUFS units (equal to LKFS as defined in Rec. ITU-R BS.1770).

The second part examined the system's ability to preserve the target loudness level across multiple preselections through non-real-time set up. The test demanded re-authoring the content using the proponent authoring tool and adding more preselections. After that, the audio was encoded and decoded to the target loudness levels: -31, -24, and -16 LKFS using the proponent encoder and the loudness consistency was measured according to ITU-R BS.1770-4 with a tolerance of +/-3 dB, using FFMPEG tool.

The third part evaluated the preservation of the target loudness level after user interaction. The test demanded re-authoring the content using the proponent authoring tool and changing the minimum and maximum gain interactivity options for several dialog objects to at least +/- 10 dB. After that, the audio content was encoded and MP4 files mixing audio and video streams were created. The MP4 files were playbacked using the proponent video player and the increase of level of dialog objects was evaluated as well the overall perceived loudness before and after the user interaction.

Test 7 (Seamless configuration changes and Audio/Video alignment):

This test was divided into seven TCs. The first one tested the system's ability to seamless playback content during configuration changes through real-time set up. The expected result was a continuous and seamless playback during all configuration changes.

The second test examined the seamless playback content during user interaction through real-time set up. It evaluated the occurrence of audio drop-outs or glitches in real-time playback during changes between different audio languages or preselections, increasing or decreasing the level of various audio objects.

The third test consisted of demonstrating the system's ability to seamlessly playback content during changes in production during a live broadcast through real-time set up. All typical changes in a live broadcast were tested.

The fourth test examined the seamless playback during changes in production in a live broadcast scenario through real-time set up. For this test, the pre-recorded output of the proponent's authoring system was used as production content in the broadcaster studio. Using the authoring system in the studio, the metadata was re-authored, and potential errors in the original authoring were treated. The ability of enabling/disabling interactivity for one and more preselections and one and more audio objects were evaluated as well as the interactivity options (min/max gain and position values) and the capability to seamlessly playback the content and correctly display the user interaction options while making the changes in live production.

The fifth test analyzed the system's performance during advertisement insertion. Using a clean SDI switch, two different contests were switch in order to evaluate the capability of the audio system to display on the receiver side the various interactivity options corresponding to each configuration and seamlessly update the user interface at each ad-insertion and evaluate continuously and seamlessly playback the content during the ad-insertion.

The sixth test examined the system's ability to preserve the user interaction settings after the ad-break. After re-authoring metadata in the authoring system, a clean SDI switch was used to switch between contents. After one minute, the contents were switched and verified the ability to preserve the user selections after the ad-break.

The seventh test consisted of demonstrating the system's ability to synchronize during the replacement of the main soundtrack delivered via broadcast for an alternative audio signal delivered via broadband through real-time set up. In the beginning of the test, the content was encoded offline and prepared as multiple ISOBMFF streams ready for DASH streaming from the Streaming Server containing a Channel Bed 2.0 program and different objects. In the test room, it was evaluated the ability: to synchronize the multiple streams received live; to display the options available (e.g. the playing starts always with stream 1 but options from stream 2 and 3 shall be displayed); to switch to additional languages coming from IP chain 2 (Streaming Server) and to switch back to the main language if the IP chain 2 is disconnected (e.g., stopped from the Streaming Server).

Test 8 (Audio coding efficiency):

This test examined the proponent's documentation provided on the Quality Assessment Reports, without laboratory tests.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

Test 9 (Audio End to end latency):

This test was performed during the execution of Test Case TC1.1, and its requirement AC9.1 was classified correlated to the requirements AC1.1.1, AC1.1.2, and AC1.1.3.

Test 10 (Audio/Video synchronization)

This test examined the A/V synchronization through real-time set up. It evaluated the downmix and rendering when the content was played back over a 5.1 setup and a stereo setup and the overall 3D experience when the content was played back over an external sound system.

Test 11 (New immersive audio services)

This test examined the ability of the system to perform playback of audio in VR / AR / XR applications, with 3DoF. or 6DoF. It evaluated: features of the codec; the readiness for real-time coding/decoding; the readiness of delivery of the format; and how the application works and could manipulate the audio codec stream. The test verified the demonstration of 3D audio VR and 3DoF support (none of the proponents included a 6DoF example) and was partially fulfilled, as the provided schemes did not permit the evaluation of the capability of real-time encoding and delivering/decoding in a streaming/broadcasting fashion. For the time being, it demonstrated, however, the capability of offline encoding and decoding on a cell phone application.

Test 12 (Interoperability with different distribution platforms)

This test analyzed the system's ability to send multiple audio contents over two or more communications channels through real-time set up. It needed multiple ISOBMFF streams ready for DASH streaming from the Streaming Server to be created in order to verify the abilities to: synchronize multiple streams received live; display the options available; switch to additional languages coming from IP chain 2 (Streaming Server) and the ability to switch back to the main language if the IP chain 2 is disconnected (e.g., stopped from the Streaming Server).

Test 13 (Audio scalability and extensibility):

This test was performed during the execution of Test Case 12.1, verifying its requirement AC13.1. Concerning the requirement AC13.2, it was not analyzed with a feature test but through the analysis of the proponent's documentation provided in the Document Analysis phase.

VI. Results and discussions

The results obtained in the conducted test cases provided a rich showcase of the actual capabilities delivered by the current versions of the candidate technologies and their potential to deliver a set of new services for audio presentation in a variety of program types.

A final and comprehensive report of the results – indicating the fulfillment classification for each requirement and for all tested technologies – was prepared by our team, and then published publicly by the Forum SBTVD [14].

This section presents an overview of the audio personalization and immersion features that were evaluated and would be possible in the next generation of digital TV; communicates the decisions made by the technical Forum regarding the official adoption of technologies; and brings some discussions on present limitations, future prospects, aesthetics challenges, technical issues, and ongoing activities towards the implementation of the new TV3.0.

A. Audio personalization and immersion features

The personalization of audio program presentation – in terms of enabling the user to control sound selection, positioning, distance, loudness, equalization and effects – has being thought for television markets even before the first generation of DTV in 2007, when ISO/IEC issued the call for technology proposals for MPEG SAOC (Spatial Audio Object Coding). The concept of authoring presets for final users' selection and personalization, for instance, had already been addressed by Lee et al in 2006, at that time using the MPEG-4 BIFS (BInary Format for Scenes) 3D audio scene description tool in the MPEG-4 Systems [15].

The ISO/IEC call for proposals for developing the MPEG-H 3D Audio standard was issued 10 years ago (Jan. 2013), targeting application scenarios such as personal home theaters, handheld smartphones, 3D video, telepresence rooms, cloud-based gaming, and accurate sonic localization for audio-only program listening (e.g. virtual concert halls), seen as potential markets for the future. For broadcast application, a special attention was given to cover demands for transmitting several groups of audio content and making them accessible independently from each other, such as multi-language announcer and commentator voices, environment sounds, and different sound mix presets.

Sport events and karaoke were typical program examples, where the user should be able to select and to enable or not several sounds out of a set of options (e.g. turning on/off the stadium announcements and audience sound; selecting the preferred language and commentator; selecting specific instruments, vocal option, or presets with different instrumentation) and to adjust localization and loudness levels for the sounds. Keeping timbre, sound localization and envelopment were also requirements for MPEG-H, as well as the capability of downmixing the number of channels and rendering the spatial sound to a lower hierarchy of loudspeaker setup, with 10.1 (ten full channels plus a low frequency one) or 8.1 channels.

The TV3.0 personalization capabilities will include the selection and activation of available sound programs and the control of loudness (prominence) levels. End-users shall be able to switch between components (e.g. alternative mix substreams or audio objects), to adjust their loudness, and to enable services such as audio description and emergency warning information.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

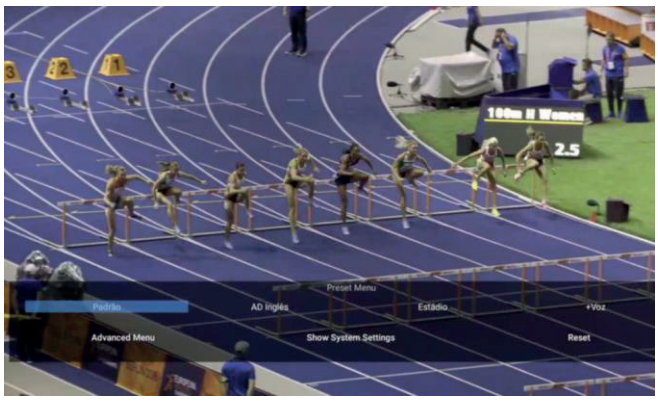


Fig. 9: Menu for selection of programs/contents and personalization of sound system settings.

Sound personalization possibilities illustrated in Fig. 9 include:

- change of audio program (from the available options delivered)
- change language of announcer / commentator
- activation of the audio description feature (or not)
- change playback configuration (e.g. output system such as soundbar, external 5.1+4H system or embedded stereo loudspeakers of the TV set)
- choose which sound elements (objects) available in the programs the end-user wants (or not) to listen
- modify the presence level of specific sounds (loudness adjustment), e.g. turning up speech or lowering the background environment sound
- reset for the default configurations set by the broadcaster

Considering the immersion features, the next generation will enable a flexible selection and seamless switch between available audio playback configurations, such as 2.0, 5.1, and 5.1+4H channel-based layouts (home-theaters) or binaural renderization (for earphones). It shall also permit at the end-user device to control the sound panorama and to adjust sound object position in the listening area, according to available broadcaster's settings and within the range conceived by the program content producer.

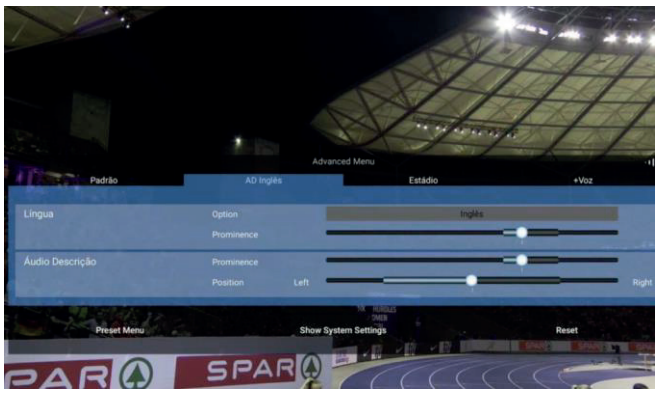


Fig. 10: Menu for audio description and language selection, loudness (prominence) and sound spatial positioning.

Immersion personalization possibilities illustrated in Fig. 10 include:

- define the positioning of specific sounds in the selected program around the user (panorama setup: left-right within the range set by the broadcaster)
- define the prominence of the specific selected sounds
- reset the configurations back to the broadcaster's default settings.

The Test Cases verified some possible immersive personalization capabilities, such as the end-user ability to move audio elements at the end-user's device inside an area specified by the broadcaster (e.g., min/max position interactivity values), and have shown that this operable area may be set differently for each object. However, the present end-user's interfaces for changing the spatial scene were restricted to a simple panorama (left-right) adjustment, controlled with the remote control keys. Although not available in the actual tested interactive interfaces, it would be possible to change the sound object positioning not only on the planar situation (2D) but also on its elevation (for 3D playback).

Considering the choices for loudspeaker systems and possible enhancements in immersive experience, it remains open space for further investigations. The determination of the immersion level delivered by a certain system could be a valuable tool both for manufacturers – in calibrating and assessing the level of perceived spatial immersion achieved, and for final users – in optimizing their listening setups. Faria has proposed in 2005 a discrete 6-level scale for immersion degree assessment, starting from 0 (no spatial information) up to 6 (3D coherent spatial impression) [16], but no practical usage of this or similar scales has been tracked so far.

Considering the choices for multichannel playback layouts, we comment on some interesting points. Television sets shall continue offering the 2.0 (stereo) output, but for taking advantage of 2D and 3D experience the end-user must use an external multi-loudspeaker sound system, which will receive the audio stream from the DTV receptor typically using a HDMI cable, and then will distribute the output signals for each of its loudspeakers. An AVR unit might be one choice, in which case it should be considered if it handles MPEG-H decoding and rendering functions in an integrated manner (which has been the tested situation using a commercial native MPEG-H AV receiver).

Alternatively, some DTV devices might include native multichannel loudspeakers support. A flat HDTV television set prototype with over a hundred micro-loudspeakers embedded onto its frames has been tested by NHK in 2012, aiming to enhance frontal spatialization using a wave-field synthesis scheme [17]. However, such types of loudspeaker arrays have not shown popularity like newer soundbar systems – a single multichannel audio receiver with an (embedded) array of loudspeakers. A study from Ando (2011) has pointed out that a minimum of 8 channels would be required for providing sound localization and envelopment to satisfactorily recreate a spatial impression at home using loudspeakers [18]. Employing layouts with loudspeaker at 3 height levels, that outcome converges to support configurations beyond the regular planar ITU 5.1 surround

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

scheme, such as unusual 8.1, 9.1, and 10.1 layouts (which mixes loudspeakers in the middle, upper, lower and top positions in different azimuths and elevations) and the more recent 5.1+4H (10 loudspeakers) proposed setup.

Notwithstanding, the popularity of multiple loudspeakers in home environments is not borne out by history, being binaural/stereo output or using a single speaker array the most likely successful choices for 3D audio consumption. Even so, current soundbars' ability to provide height effects are still limited. A sense of envelopment from above is noticeable, but the accuracy of depth and elevation positioning is less clear. It is worth recalling that the studies by Ando [18] evaluated loudspeaker layouts with 3 height levels, a requirement not explored in the present tests. Finally, improvements are expected in the definition of the elevation dimension for speaker arrays.

B. Official deliberations of technologies to adopt

Although both final tested candidate technologies have demonstrated fair sound quality – and less complexity/greater ease of use in particular test cases and functionalities – after weighting the complete fulfillment of all CfP requirements, the tests results were evaluated by the Forum SBTVD and a final deliberation pointed towards the formal adoption of the international open-standard MPEG-H 3D Audio for the future TV3.0 Over-the-Air (OTA) and Internet distribution (OTT) – ISO/IEC Standard 23008-3, now in its 3rd. edition, revised in 2022 [11]. It also decided to maintain the E-AC3 and AAC (MPEG-4 Advanced Audio Coding) audio formats currently supported in TV2.5 for the distribution of alternative content over the Internet, including optional Dolby AC-4 support.

Most use scenarios highlight the capability for the program producers to assemble different setups of audio presentations (which could be seen as "presets") and to deliver them in different sets of channels and programs. The systems take the advantage of being compliant to NGA program description schemes, such as with the xml-based ITU Audio Definition Model (ADM) metadata format [13], which conveys programs descriptions and how they are organized and could be selected and presented at the receiver (player) side.

C. Discussions on challenges, prospects on aesthetics and technical issues

The TV3.0 system opens and welcomes a broad avenue for novel aesthetics and ways to explore the system's resources to code and deliver interactive and personalizable content. There are challenges, however, on the current content production chain, not considering yet the benefits and impacts on the production and post-production methods, and the unexplored creative potential for artists in inventing new forms of programs, both to take advantage of the new features offered by the technology and to engage the users in active participation in the program experience. Some system's capabilities, such as the "emergency warning delivery" could be, for example, explored by TV3.0 application coding experts to incorporate new features into the future specification.

The ISO/IEC 23090-4 MPEG-I Immersive Audio forthcoming standard – currently under development – is expected (possibly in a shorter time-to-market than usual) to further expand the possibilities for immersive program creation, delivery, consumption and personalization, especially for Virtual and Augmented Reality audio presentations. MPEG-I will permit 6DoF experience (6 Degrees of Freedom) i.e. translation {x,y,z} and rotation {yaw,pitch,roll} directed by the physical movement of the user in the space. Several enhancements in terms of immersive experience are expected, such as better directivity perception and positioning of sound sources, ambience and reverberation, and new rendering technologies.

The new technology also brings novelties in the transport layer, employing novel and alternative methods to the conventional MPEG TS (Transport Stream) container format used for transmission and storage of media. MPEG-H content can be packetized in MPEG-H Audio Stream (MHAS) format, or encapsulated into ISOBMFF files (which we have tested in the present study using the non-real-time encoding/decoding setup). Besides MPEG TS, delivery could be implemented over the ROUTE/DASH protocol (Real-Time Object Delivery over Unidirectional Transport/ Dynamic Adaptive Streaming over HTTP) which we have tested in the present study using the real-time encoding/ decoding setup.

Finally, as incorporated in the mandatory tested requirements, our studies also unveiled the possibilities for binaural rendering for Virtual and Augmented Reality applications, and scene-based audio formats, such as HOA, which are establishing new landmarks for film and communications in the 21st century.

Scene-based streams carry full sound field representations, agnostic of the final required loudspeaker feeds to recreate them at playback, but are sensitive to channel bandwidth and compressive artifacts. MPEG-H (and future MPEG-I) use several techniques for spatial compression of HOA signals. Bleidt et al (2017) states that broadcast quality (scoring a minimum of 80 MUSHRA points) of HOA content up to the sixth order (which sums up to 49 HOA coefficients) is achieved at bitrates as low as 300 kb/s and transparent quality transmission (minimum of 90 MUSHRA points) are achieved at 500 kb/s independent of the HOA order [19]. In a standard SDI environment, however, limited to 16 channels, MPEG-H is capable of delivering a full set of 15 audio channels plus 1 control channel (for metadata), which means a (pure) HOA signal up to the third order (16 channels).

VII. Final Remarks

The effectiveness of personalizing a sound scene is an enterprise dependent not only on the technology of coding, transmitting and rendering, but also on the content motivation and interactivity aspects. One has to observe that while current technologies offer the possibility of defining specific loudness and 3D positioning for a sound in a scene, the final user or spectator ability to control these attributes are limited by the user interface, which can, at last, encourage or demotivate spectators in the task of interacting with the scene or setting it up. Intuitive means to select sounds and pinpoint their desired positions in the space by voice command or gestural tracking should be much more taken for granted than operating over a remote control.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

Besides the user interface, one has also to consider the program's ability to entice the spectator to participate in its flow by selecting options and settings. Content providers and producers are ultimately responsible for the creativity and innovation in the next generation of attractive and popular interactions.

These features open a new and vast avenue to be explored by program content producers, artists, and broadcasters, as they can assemble and deliver diverse audio contents in the same stream, for alternative presentation options, and also permit the user to select or not some sound items, languages and accessibility services, to modify their spatial distribution around, and hence to personalize the content and the immersive aspects of the listened experience.

Concerning the MPEG-I forthcoming standard, its compatibility with the MPEG-H architecture, codecs, bitstream and transport mechanisms is expected to facilitate the eventual integration of its new features into the next TVD realm, but that is a future addressable target, not subject to further speculation at the moment.

By the time we conclude this article, the TV3.0 project is undergoing. Still deliberations are expected on other layers, as necessary developments and tests are in progress (for instance in the applications coding layer, which also employs MPEG-H). Working Groups (GTs) are currently addressing technology issues and needs (such as the necessary developments in authoring tools and innovative personalizable and interactive programs for additional tests) and challenges in establishing regulatory policies for Internet OTT (Over the Top) and broadcast OTA (Over the Air) delivery.

REFERENCES

[1] Brazilian Digital Terrestrial TV Forum, CfP Phase 2 / Testing and Evaluation: TV 3.0 Project (15 March 2021), available at the TV3.0 Project website at https://forumsbtvd.org.br/tv3_0/

[2] ATSC 3.0 Audio Testing Report, Doc. S34-2B-048r7 (12 August 2015).

[3] ATSC Standard: A/342:2021 Part 1, Audio Common Elements (Doc. A/342:2021 Part 1) (9 March 2021)

[4] ATSC Standard: A/342:2021 Part 2, AC-4 System (Doc. A/342:2021 Part 2)" (10 March 2021).

[5] Advanced Television Systems Committee, "ATSC 3.0 Audio Testing Report, Doc. S34-2B-048r7", Washington, USA, August 2015.

[6] MPEG-H 3D Audio Verification Test Report, ISO/IEC JTC1/SC29/WG11 MPEG2017/N16584 (January 2017, Geneva)

[7] IEEE Standard for Second Generation Audio Coding, IEEE Std. 1857.8-2020, pp.1-470, 25 Nov. 2020, doi: 10.1109/IEEESTD.2020.9271961.

[8] ETSI TS 103 190 Digital Audio Compression (AC-4) Standard, version 1.1.1, 2014. Available at https://www.etsi.org/deliver/etsi_ts/103100_103199/103190/01.01.01_60/ts_103190v010101p.pdf

[9] ETSI TS 103 190-2 Digital Audio Compression (AC-4) Standard; Part 2: Immersive and personalized audio, version 1.2.1, 2018. Available at https://www.etsi.org/deliver/etsi_ts/103100_103199/10319002/01.02.01_60/ts_10319002v010201p.pdf.

[10] Advanced Television Systems Committee, "Techniques for Establishing and Maintaining Audio Loudness for Digital Television," Doc. A/85:2013, Washington, D.C., Mar. 12, 2013. Corrigendum No. 1, "SPL," approved Feb. 11, 2021.

[11] Information Technology—High Efficiency Coding and Media Delivery in Heterogeneous Environments—Part 3: 3D Audio, Standard ISO/IEC 23008-3:2022, 3rd ed., 2022. Available at https://www.iso.org/standard/83525.html.

[12] Y. Grewe, A. Murtaza, S. Meltzer. "MPEG-H Audio System for SBTVD TV 3.0 Call for Proposals", *SET International Journal of Broadcast Engineering*, v. 7, 2021.

[13] Recommendation ITU-R BS.2076-2 (10/2019), Audio Definition Model, https://www.itu.int/rec/R-REC-BS.2076/en.

[14] R. R. A. Faria, A. A. Rosa, E. Mendes, A. A. B. Silva, D. H. S. Abreu, S. D. Costa, H. F. Rozena, G. K. F. Komatzu, "Testing and Evaluation Report: TV 3.0 Project – Audio Coding", Brazilian Digital Terrestrial Television System Forum, University of São Paulo, Dec. 3, 2021. Available at https://forumsbtvd.org.br/tv3_0/#panel-phase2.

[15] D. Jang, T. Lee, Y. Lee, & Yoo, J. H. (2006, October). A personalized preset-based audio system for interactive service. In Audio Engineering Society Convention 121. Audio Engineering Society Available: http://www.aes.org/e-lib/browse.cfm?elib=13738

[16] R. R. A. Faria, M. K. Zuffo and J. A. Zuffo (2005, July). Improving spatial perception through sound field simulation in VR. In IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2005. IEEE, doi: 10.1109/VECIMS.2005.1567573.

[17] Okubo, H., Sugimoto, T., Oishi, S., & Ando, A. (2012, October). A method for reproducing frontal sound field of 22.2 multichannel sound utilizing a loudspeaker array frame. In Audio Engineering Society Convention 133. Audio Engineering Society. Available: http://www.aes.org/e-lib/browse.cfm?elib=16456.

[18] A. Ando, "Conversion of Multichannel Sound Signal Maintaining Physical Properties of Sound in Reproduced Sound Field," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1467-1475, Aug. 2011, doi: 10.1109/TASL.2010.2092429.

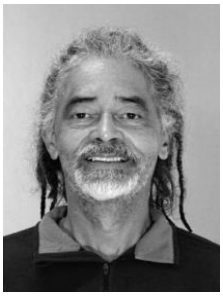
[19] R. L. Bleidt, D. Sen, A. Niedermeier, B. Czelhan, ... & M. Y. Kim, "Development of the MPEG‐H TV audio system for ATSC 3.0", *IEEE Transactions on broadcasting*, 63(1), pp. 202‐236, 2017, doi: 10.1109/TBC.2017.2661258.

Regis Rossi A. Faria received the B.S. degree in electrical engineering from the Federal University of Minas Gerais (UFMG), Brazil, in 1990 and the M.S. degree and Ph.D. degree in electrical engineering from the University of São Paulo (USP), Brazil, respectively in 1997 and 2005.

In 2004 he was a researcher at USP collaborating to the development of the first generation of the Brazilian digital television system (SBTVD) in the audio layer. He has coordinated the testing and evaluation of audio technologies for the SBTVD TV3.0 project phase 2 in 2021, being a current collaborator with the SBTVD technical Forum.

Dr. Faria is an Associate Professor with the University of São Paulo, where he coordinates the Laboratory of Audio and Music Technology (LATM) at the School of Arts, Sciences and Humanities (EACH), and has research interest in spatial audio, sound and music computing, audio engineering, immersive aesthetics and production technologies. He is member of the Audio Engineering Society and of the Brazilian Computer Society.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2



Almir Almas is PhD in Communication and Semiotics by the Pontifical Catholic University of Sao Paulo and an Associate Professor of the Department of Film, Radio and Television at the School of Communications and Arts of the University of São Paulo, and Researcher of the Program of Postgraduate Studies in Media and Audiovisual Processes, where he is the General Coordinator of the Research Group LabArteMídia (Laboratory of Art, Media and Digital Technologies) and Obted (Brazilian Observatory of Digital Television and Technological Convergence).

He is currently Visiting Professor and Researcher (Support by FAPESP BPE) at the Faculty of Humanities and Social Sciences and School of English and Modern Languages at Oxford Brookes University. Author of 'Televisão digital terrestre: sistemas, padrões e modelos' (Digital terrestrial television: systems, standards and models), among other books and articles.

Dr. Almas is Filmmaker/Videoartist/VJ, and Artist of the Cobaia Art Collective and Formigueiro. He is a Member of the Board of the Brazilian Society of Television Engineering (SET) and Member of the Brazilian Digital Terrestrial Television System Forum (FORUM SBTVD).

Eduardo Santos Mendes was born in São Paulo, Brazil. He received the B.S. and M.S. degrees in Film from USP (São Paulo University) and the Ph.D. degree in Arts/Film Sound in 2000 from USP.

He is a sound supervisor since 1984 and professor at USP since 1990 where he taught in the Bachelor's Degree in Film, Radio and TV and Film and Video courses. At the moment, he teaches in the Bachelor's Degree in Audiovisual Course. He is a member and advisor in the Postgraduate Program in the Audiovisual Media and Processes (USP) since 2002. Dr. Mendes was also a guest teacher in schools in Mexico and Belgium. In his work as a sound supervisor, he collaborated with main Brazilian directors such as Carlos Adriano, Carlos Reichenbach, Lina Chamie, Tata Amaral and Walter Hugo Khouri.

Dr. Mendes is part of the board of directors of CIBA/CILECT, was awarded in Brazilian and International festivals such as Brasilia, FestRio and Havana (Cuba). He is a researcher of audiovisual technology and narrative, focusing on sound.

Ana Amelia Benedito Silva is a Brazilian Professor of the Postgraduate Program in Modeling Complex Systems and the Undergraduate Course in Information Systems at the School of Arts, Sciences and Humanities at the University of São Paulo. She achieved her bachelor's, master's, and PhD degrees in Electrical Engineering at the University of São Paulo.

Ana Amelia Benedito Silva has experience in the field of studies of rhythmic phenomena, with an emphasis on sleep, and the application of mathematical and statistical models to the analysis of empirical data mainly linked to the area of biology and health. Themes of activity: sleep, sleep-wake cycle, shift worker sleep, biological rhythms, biostatistics, mathematical modeling, series analysis temporal, malnutrition, exercise physiology.

Douglas H. S. Abreu was born in Lavras - MG, Brazil. He earned his bachelor's degree in Information Systems from the Federal University of Lavras, Brazil, in 2015, and his master's degree in Systems Engineering and Automation from the same university in 2017. Currently, he is a professor at the Polytechnic School of PUC Campinas and a doctoral candidate in Electrical Engineering and Computing at UNICAMP.

Throughout his career, Douglas has focused on Machine Learning, audio and acoustics, and recently cybersecurity, working on notable projects such as mapping critical infrastructures and their cyber vulnerabilities in the Brazilian electrical sector, and testing and evaluating audio technologies for Brazilian TV 3.0. He is currently involved in the Technological Training Program in Artificial Intelligence (AI) promoted by CPQD and PUC-Campinas with the support of the Ministry of Science, Technology, and Innovations (MCTI).

Prof. Abreu is a member of the Laboratory of Acoustics of Communications at FEEC/UNICAMP (LAC-Unicamp).

Henrique F. Rozena was born in São Paulo, capital of the state of São Paulo, Brazil, in 2000. He graduated from high school at Escola de Aplicação and continues his studies at the FATEC State Technology college, within the digital media design course.

In 2021 he was an assistant fellow in the research and testing project for audiovisual equipment in the TV 3.0 Project, and has since then helping to conduct the X-Reality events at the University of São Paulo where the project is conducted.

Received in 2023-06-08 I Approved in 2023-07-08

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.2 Web Link: https://dx.doi.org/10.18580/setijbe.2023.2

The Use of Artificial Intelligence Enabling Scalable Audio Description on Brazilian Television: A Workflow Proposal

Luiz F. Kruszielski
Pedro H. L. Leite
Pedro Bravo
Marcelo Lemmer
Edmundo Hoyle

Kruszielski, Luiz F.; Leite, Pedro H. L.; Bravo, Pedro; Lemmer, Marcelo and Hoyle, Edmundo; 2023. The Use of Artificial Intelligence Enabling Scalable Audio Description on Brazilian Television: A Workflow Proposal. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2023.3. Web Link: https://dx.doi.org/10.18580/setijbe.2023.3

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING - SET IJBE V.9, 2023, Article 3, 5 p
2023 SET - Brazilian Society of Television Engineering / ISSN (Print): 2446-9246 / ISSN (Online): 2446-9432

The Use of Artificial Intelligence Enabling Scalable Audio Description on Brazilian Television: A Workflow Proposal

Luiz F. Kruszielski, Pedro H. L. Leite, Pedro Bravo, Marcelo Lemmer, Edmundo Hoyle

***Abstract*—Recently, Artificial Intelligence (AI) technologies have been gaining ground in various areas of knowledge, significantly impacting many academic and business spheres. One application that can benefit from AI is the inclusion of people with disabilities in audiovisual content, where the scaling capacity of certain processes can bring new accessibility opportunities. In this work, we show what a traditional workflow of an audio description for dramaturgy audiovisual content looks like, and from there, we propose a new workflow for generating audio description audios for visually impaired people using synthetic voice created with Artificial Intelligence models. The proposed workflow simplifies and considerably reduces production time and costs, besides allowing the generation of audios on a larger scale compared to a traditional workflow, enabling a broader reach of the target audience. It also allows multiple people to work simultaneously on the same project while preserving sound identity through the synthetic voice and standardized mixing. With this proposal, we believe that accessibility on Brazilian television can be expanded to serve a much larger audience.**

***Index Terms*—Audio Description, Artificial Intelligence, Voice Synthesis, Accessibility,**

I. Introduction

Audio description(AD) is the most important tool for enabling inclusion of people with visual impairments, but its implementation is not always feasible due to the complexities of the workflow and scalability. In AD, elements and actions determined as important in a scene are narrated by a voice, providing the possibility of understanding the plot and narrative context without the need for visual presence.

According to Joel Snyder, "Audio Description provides narration of the visual elements - action, costumes, settings, and the like - of theater, television/film, museum exhibitions, and other events. The technique allows patrons who are blind or have low vision the opportunity to experience arts events more completely - the visual is made verbal. AD is a kind of literary art form, a type of poetry. Using words that are succinct, vivid, and imaginative, describers try to convey the visual image to people who are blind or have low vision."[1] Most audio-visual content does not have AD, and this deprives the entire population with visual impairment or low vision of this content. The cultural implications caused by this type of impediment are enormous. This restriction not only prevents the delivery of entertainment content but also enables cultural integration with the general public, as we understand that much of the television content is part of the daily life of Brazilians. Enabling the consumption of this content by a part of the population that, otherwise, would not have access, creates a common culture, and this is a way to create inclusion.

In this article, we are presenting the workflow being implemented in the production of audio description for series and daily drama series (novela) of Grupo Globo's programming. This process incorporates synthetic voice generation and automatic mixing techniques. We believe that the proposed workflow has the capacity to reach a larger number of people with visual impairments, providing access to this resource that is often unavailable. The work will be divided into five distinct sections. Initially, we will address the current panorama of audio description and the challenges it presents. In the second section, we will detail the technique employed to create synthetic voice. The third section will address the conventional audio description production process, in addition to introducing the workflow proposed by this study. In the fourth section, we will share the challenges and the advantages we identified when using the new workflow. Finally, in the last section, we will present our conclusions.

A. Current Scenario of Audio Description:

Data from IBGE in 2010 indicate that Brazil has a population of about 6.5 million people with high or severe visual impairment [2]. This data is corroborated by the National Health Survey (PNS) of 2019[3], which shows that 3.4% (3.978 million) of the population has visual impairment. It is important to emphasize that audio description is not only important for people with total vision loss, as those with partial and severe loss can also benefit from this technique. Other groups, such as people with intellectual disabilities and learning disorders, can also take advantage of audio description "...as it is a second sensory channel to be used for faster comprehension of visual information [4]."

Audio description for open television began in 1982 by the American network PBS, where it was simultaneously transmitted on television and the audio description was transmitted on FM Radio [4]. In Brazil, except for some film

L. F. Kruszielski is with Grupo Globo, Rio de Janeiro, RJ, Brazil (e-mail: luiz.fk@g.globo).
Pedro H. L. Leite. is with Grupo Globo, Rio de Janeiro, RJ, Brazil (e-mail: pedro.hleite@g.globo).
Edmundo Hoyle is with the Grupo Globo, Rio de Janeiro, RJ, Brazil (e-mail: edmundo.hoyle@g.globo).
Marcelo Lemmer is with the Grupo Globo, Rio de Janeiro, RJ, Brazil (e-mail: marcelo.edward@g.globo).

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.3 Web Link: https://dx.doi.org/10.18580/setijbe.2023.3

content, the production of audio description for dramaturgical content is very scarce. In recent years, this production has started to grow, but it is still far from covering most of the television content. This is partly due to the dynamics that exist in the process of producing television dramaturgical content. In the Globo group, most of the content produced daily by dramaturgy is made in the "open work" format, that is, the daily drama is not recorded with all its chapters finalized and is written according to the repercussion of the chapters that are being broadcast. This means that the delivery of episodes is very close to the exhibition, which can significantly complicate the construction process of AD since it can only be executed with the completed material. The public demand for this technology in this type of dramaturgy is not new. In 2005, for example, a group of visually impaired people wrote an open letter requesting that the soap opera América, which featured a visually impaired character, be produced with audio description [4]."

By the no. 188 ordinance from the Brazilian agency of telecommunications, Anatel, (2010) [5], it is mandatory for all open television networks to broadcast a minimum of 20 hours per week of content with AD. Partly, the material produced today for AD is performed in auditorium programs, being live audio description. In this process, a person narrates in real-time what is happening in the program. This type of use of AD is relatively simpler than when applied to dramaturgy content and eventual mistakes tends to be admitted. This is because it is in a relatively controlled environment where the variation of the content displayed on the screen is significantly less complex than in drama. The program takes place entirely in the same environment, with a group of participants and pre-defined agendas without major changes. In dramaturgy, the environment where a scene takes place can be very different from the next, requiring the audio describer, that is, the one who is visualizing and narrating in loco, to verbally present that scenario in each situation, or an introduction of various characters. Due to the complexity, to create audio description for daily dramaturgy, it is necessary that the content is locked, that is, there is no more alteration in the editing or sound design of it, so that AD script can be made and subsequently narrated. More about this workflow is detailed in section III. One of the difficulties in creating AD content for daily dramaturgy is working with open works, that is, works that do not have the script written from start to finish, and can be modified according to the response from the audience. This condition means that the time between what is produced and what is broadcast can be greatly reduced. Another complicating factor is that the audiovisual product can be changed according to external factors, such as the availability of the broadcast schedule on the day or the sale of commercial breaks, and which occur even closer to the broadcast. An automated tool that allows anyone to edit audio description content is extremely useful and necessary for this type of situation in order to have AD.

II. Use of Synthetic Voice to Generate Audio Description:

The recent developments in synthetic voice using AI plays a key role in allowing the possibility of using an scalable voice for AD, avoiding the need of recording procedures for each content. In this section we describe the method proposed for creating a voice that sounds natural and is not perceived as machine generated content. Also, the neutral intonation required for AD fits the current data availability for voice synthesis, specially in Brazilian Portuguese [6].

For audio description, the technique used to synthesize voice is Text-to-Speech (TTS), where the text is the input to a computer algorithm that generates speech as output. Modern TTS systems are created using machine learning architectures, specifically deep learning. These structures can capture semantic and linguistic relationships between the words in the text and can generate pronunciation and intonation consistent with the textual intent.

To build these structures, some training steps are required, during which pairs of text and corresponding audio signals are presented to the models in batches. The training process involves an initial stage with a neutral voice using the data available in [6]. After obtaining a model with a neutral voice, there is a second stage related to transfer learning for the target voice, which has a smaller amount of recorded hours available. In this final stage, fine-tuning of timbre (acoustic properties of the voice) and prosody (rhythm, stress and intonation of the speech) is performed for the individualities of the voice that will carry out the audio description. The learning architectures used were those described in works [7-8] (Tacotron2 and Multiband-MelGAN, respectively) using an open-source code implementation[1]. The Tacotron2 model, whose block diagram can be seen in Figure 1, uses convolutional [9] and bidirectional LSTM recurrent networks (which use future and past context in sequences) [10] with attention mechanisms to decode the text and relate it to psychoacoustic characteristics, which are implicitly modeled by the network weights. These characteristics are subsequently transformed into a mel-spectrogram by convolutional layers and linear projections. Since mel-spectrograms do not contain phase information and have a bandwidth narrower than desired, a subsequent step of generating the waveform in time is still necessary, where the second model used in this work, described below, comes into play.

The neural vocoder Multiband-MelGAN is a machine learning model based on generative networks whose intuition is to leverage the characteristics of generating faithful samples of Generative Adversarial Networks (GANs) [11] to reconstruct waveform time-amplitude signals from mel spectrograms. Multiband-MelGAN is an extension to the original MelGAN [12], which uses convolutional networks in its generator and a multi-scale audio discriminator, using downsampling techniques. In the Multiband case, processing is done in subbands, creating and joining signals in several individual frequency bands, instead of a single full-band signal as in the case of simple MelGAN. This division in

[1] https://github.com/TensorSpeech/TensorFlowTTS

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.3 Web Link: https://dx.doi.org/10.18580/setijbe.2023.3

processing may help the network learn which parameters are important for each frequency band, consistent with the predictions of psychoacoustic models that show that our auditory apparatus excites frequency bands differently and that our perception is also heterogeneous in this sense.

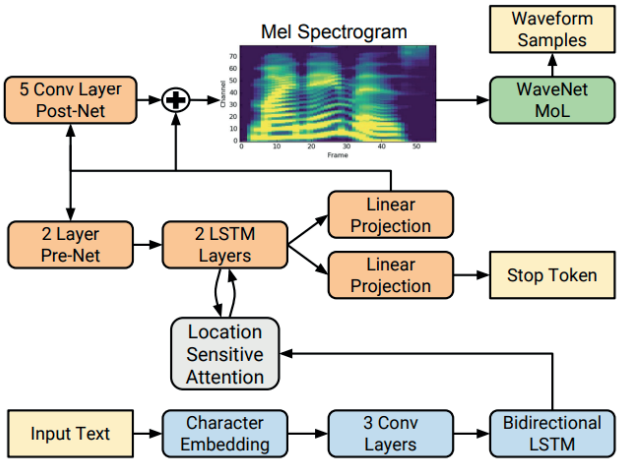


Figure 1. Block Diagram of Tacotron 2 [6].

III. Audio Description workflow:

The AD technical process can be described in four stages (Fig.2):

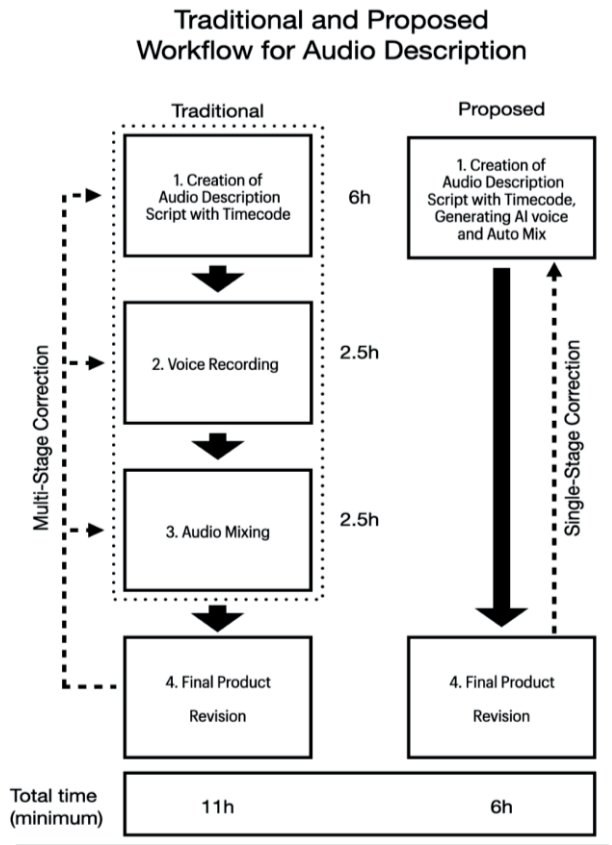


Figure 2. Traditional and proposed workflow for Audio description. The two processes are showed parallel in a vertical disposition. The dotted line arrow shows how the process of the revision can make influence in the different stages and cause a cascade correction effect.

1- Script Creation: A document containing the information to be narrated in the audio description is created by an audio describer, based on the finished audiovisual content. In this stage, to create the speech that narrates the scene, the audio describer already considers the space between existing voices in the audiovisual content, so that the narration does not overlap any dialogue. Then, a narration script is created from the approximate timecode already marked during the scriptwriting process. The minimum estimated time to adequately complete this stage for a 40-minute program is about 6 hours.

2- Audio Recording: The narrator must have a clear and pleasant voice and must be able to read the script text naturally and neutrally. It is also necessary to have high-quality recording equipment, minimally compatible with a broadcast standard. This includes a recording booth with adequate acoustics and quality microphones. For content with multiple episodes, it is important to use the same voice to maintain narrative unity. The voice is perceived by the viewer as a character, and changing it generates a strangeness similar to changing an actor for the same role during a program. The minimum estimated time to adequately complete this stage for a 40-minute program is about two and a half hours.

3- Audio Mixing: The audio recorded by the narrator is mixed with the original video. The audio editor must adjust the volume of the description audio so that it does not interfere with the original video sound. This primarily involves avoiding overlap between the audio description narration and dialogues originally contained in the video. For this, the audio editor can often use the artifice of speeding up the narration to fit it into a specific time frame. If the editor feels that the result was not satisfactory, they can request a re-recording or a text modification to fit that time. Ambient sounds and sound effects are also important for understanding the scene, as they are an integral part of the actions and help the viewer understand what type of environment a scene occurs in. The minimum estimated time to adequately complete this stage for a 40-minute program is about two and a half hours.

4- Evaluation of the Final Content: A team of technical evaluators verifies, among other things, the quality of the audios and the mix. It is worth mentioning that at all stages, the participation of a consultant is extremely important for a good understanding of the product. This consultant necessarily needs to be someone with visual impairment and usually not only accompanies the final product but can also follow and interfere in other stages of the process. All the times shown here do not consider eventual changes and corrections in the previous stages, where it is often necessary to redo the entire process to change a section.

The time estimation for each stage was based on the experience of producing AD of drama content made by Grupo Globo. This time refers to an estimate of the minimum necessary to deliver a product in suitable conditions. However, this value can increase significantly according to the complexity of each work and unforeseen events that may occur in a production.

The proposed process would modify the traditional workflow by altering stages 2 and 3. It would be executed by the scriptwriter themselves during stage 1, where the script would be inputted into a program, and the voice generation

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.5 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

and mixing would be immediately done, which can be tested and validated by the scriptwriter.

IV. Challenges and Advantages of Using the New Workflow

In stage 2, recording the narration would be replaced by generating synthetic voice. It is important to maintain the same requirements at this stage of the traditional method - a neutral, clear, pleasant voice with optimal recording quality. One of the biggest challenges and the most important aspect at this stage to be successful with a synthetic voice, is that at no point can this voice be perceived as artificial - the viewer must interpret it as a natural, non-robotic voice with the absence of audible artifacts. A failure in this aspect could lead the viewer to a break in the immersion of the storytelling. The perception of something "strange" in the voice can draw the viewer's attention to the voice itself and not to the story being told. This would also occur with failure in the way of speaking words correctly. With the model presented in session 2, we believe we have enough quality to meet these criteria. An advantage that synthetic voice can bring at this stage is vocal continuity, as the voice remains the same regardless of content production volume or content duration. This allows the audio description to be made by different people, avoiding possible changes in the voice that the viewer is used to hear.

Regarding the mixing process in stage 3, it is also important to maintain the characteristics required in the traditional workflow. A drama mix have an extensive dynamic sound level variation, that is a significant challenge in creating a mixing system. An automatic mixing system was created that kept the AD narrator voice intelligible in varied sound dynamic levels environments. In this system, it is possible to hear sound effects and ambiences in soft sound level scenes, and still have the AD voice at an intelligible volume in moments where high intensity level music is occurring. The AD narration was placed at a volume similar to the dialogue level, and the program could interpret the dynamics of current events to act only where and when necessary. In order to avoid overlap of in content dialog with AD, the scriptwriter validates the AD voice duration as the script is created, making the necessary changes to the text, fitting the speech excerpts within the dialog gaps.

This would significantly reduce the time needed to produce an AD, considering that stages 2 and 3 happen instantly. It also allows different people to work on the same project, such as two AD scriptwriters doing different sections of the same chapter, enabling an even greater reduction in the time it takes to create an AD.

Another advantage that emerges from the proposed flow is regarding eventual changes and corrections. Changes at an advanced stage of the pipeline required backward corrections and could provoke duplicated work at all previous phases, which can be quite complex and time-consuming. When the adjustment is made with a single-stage setup, the result and the testing of this adjustment is immediate, transforming a multi-stage process with several people into a single interaction. This also allows easier small corrections, where tasks such as dividing a program into different blocks with no text change could be made directly by the video or audio editor.

V. Conclusion:

In this work, we evidenced how a significant portion of the population can benefit from audio description. We discussed about some of the existing difficulties for large-scale implementation of it in a television production flow. We mapped the technical stages necessary to perform an audio description. For each stage, the necessary requirements were raised to perform quality audio description. From this understanding, we propose a method that reduces implementation time, complexity, and therefore the production cost. In this way, the use of artificial intelligence becomes an important tool that can speed up and simplify the AD process. The time to make a quality audio description often conflicts with the agility needed in a production line of television drama content. We believe that the model proposed here can lead to a more comprehensive and efficient implementation of audio description. The demand for this type of content exists and is not always met. Our proposal can not only help fulfill a relevant social role for the media production but also enable the integration of a significant portion of the population that still has no independent access to some kinds of content (such as daily drama). With this, we can deliver programs to a more diverse public and thus potentially increase the quality of life and social inclusion of millions of people.

References

1. J. Snyder, (2022, Aug) “Fundamentals of audio Description” Available: https://adp.acb.org/adi/ADA%20Fundamentals.doc.pdf
2. MEC (Brazilian ministry of education) “Data reafirma os direitos das pessoas com deficiência visual ” Available: http://portal.mec.gov.br/component/tags/tag/deficiencia-visual
3. Agência de notícias do IBGE (Brazilian statistic and Geography News Agency) “PNS 2019: país tem 17,3 milhões de pessoas com algum tipo de deficiência” Available: https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/31445-pns-2019-pais-tem-17-3-milhoes-de-pessoas-com-algum-tipo-de-deficiencia
4. L.M. Villa, P. Romeu. (2010). *Audiodescrição : transformando imagens em palavras.* (1st ed.) [Online]. Available: https://www.ufrgs.br/comacesso/wp-content/uploads/2019/01/Audiodescric%CC%A7a%CC%83o-Transformando-Imagens-em-Palavras.pdf
5. Anatel: Portaria 188 de 2010 - *Dispõe sobre audiodescrição e estabelece novos prazos de implementação*. Available: https://informacoes.anatel.gov.br/legislacao/normas-do-mc/443-portaria-18.x
6. P. H. L. Leite, E. Hoyle, Á. Antelo, L. F. Kruszielski, and L. W. P. Biscainho, “A corpus of neutral voice speech in Brazilian Portuguese,” in *Computational Processing of the Portuguese Language*, Fortaleza, 2022, pp. 344–352.
7. J. Shen, R. Pang, R. J. Weiss, et al., “Natural TTS synthesis by conditioning wavenet on MEL spectrogram predictions,” in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* Calgary, 2018, pp. 4779–4783.
8. G. Yang, S. Yang, K. Liu, P. Fang, W. Chen, and L. Xie, “Multi-band melgan: Faster waveform generation for high-quality text-to-speech,” *2021 IEEE Spoken Laguage Technology Workshop (SLT)*, Shenzen, 2020 pp. 492–498.
9. I. Goodfellow, Y. Bengio e A. Courville, “Convolutional Networks” in *Deep Learning Book,* 1st ed. Cambridge, 2016.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.5 Web Link: https://dx.doi.org/10.18580/setijbe.2023.1

10. I. Goodfellow, Y. Bengio e A. Courville, “Sequence Modeling: Recurrent and Recursive Nets,” in *Deep Learning Book,* 1st ed. Cambridge, 2016.
11. I. Goodfellow, J. Pouget-Abadie, M. Mirza et al., “Generative Adversarial Nets,” in Advances in Neural Information Processing Systems,vol. 27, Montreal, 2014.
12. K. Kumar, R. Kumar, T. de Boissiere et al., “MelGAN: Generative Adversarial Networks for Conditional Waveform Synthesis,” em *Advances in Neural Information Processing Systems*, vol. 32, Vancouver, 2019.

Luiz F. Kruszielski is graduated in Music-Sound Production at UFPR (Curitiba, Brazil 2004) and have a master (Tokyo, Japan 2011) and a doctor degree in Sound and the Environment at Tokyo University of the Arts (Tokyo, Japan 2013). He worked as a professional sound designer since 2003, and from 2013, he works at Globo TV Network (Rio de Janeiro, Brazil) as a researcher for sound technologies, later becoming a Sound Producer, where he was the technical responsible of the sound for more than 10 drama series and telenovelas for a total of more than 400 episodes. He currently work as an Innovation Specialist in the same institution.

Pedro H. L. Leite is graduated in Electronics and Computer Engineering (BSc.) at UFRJ (Federal University of Rio de Janeiro, 2021). He currently works as an innovation researcher at Grupo Globo and is a masters student at the Audio Processing Group at the Signals, Multimedia and Telecommunications lab in UFRJ (GPA/SMT-UFRJ). His main research interests are audio/speech processing and artificial intelligence.

Pedro Bravo is an undergraduate student in Control and Automation Engineering at UFRJ. Engaged in research in the field of Biomedical Engineering, focusing on electromechanical systems and biological signals for myoelectric prosthetics. Currently working in the Innovation department at Globo, focusing on projects related to programming and machine learning.

Edmundo Hoyle received his BSc. in Physics at National University of Trujillo (Peru) in 2004 and his DSc at the Federal University of Rio de Janeiro (Brasil) with specialization in Image Processing in 2013. Currently, he works as researcher at Grupo Globo and his main research interests are image processing, computer vision and Artificial Intelligence.

Marcelo Lemmer graduated in Music Education from UFRJ, worked as a music producer and since 2017 has been working as an audio description consultant for theater plays, short films, live musical events, and series. Currently, works in the area of audio/audio description production at Rede Globo.

Received in 2023-06-08 I Approved in 2023-07-08

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.3 Web Link: https://dx.doi.org/10.18580/setijbe.2023.3

A Python Tool to Predict Wireless Network Signals in Indoor Environments using Neural Networks

Breno Batista Nascimento Silva
Edson Tafeli C.Santos

Silva, Breno Batista Nascimento; Santos, Edson Tafeli C.; 2023. A Python Tool to Predict Wireless Network Signals in Indoor Environments using Neural Networks. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2023.4. Web Link: https://dx.doi.org/10.18580/setijbe.2023.4

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING - SET IJBE V.9, 2023, Article 4, 5 p
2023 SET - Brazilian Society of Television Engineering / ISSN (Print): 2446-9246 / ISSN (Online): 2446-9432

A Python Tool to Predict Wireless Network Signals in Indoor Environments using NeuralNetworks

Breno Batista Nascimento Silva and Edson Tafeli C.Santos

***Abstract* — The use of neural networks proved to be effective in creating more accurate predictive models compared to traditional approaches. The Python tool developed made it possible to train and adjust these models based on the information collected, taking into account factors such as the physical structure of the site, obstacles present and building materials. The results obtained during the research indicated significant improvements in prediction accuracy compared to conventional methods. This suggests great potential for the practical use of the tool in real-world scenarios, such as the planning and optimisation of indoor wireless networks, contributing to more stable and reliable connectivity indoors. The aim of this work was to create a Python-based tool that uses neural networks to predict wireless network signals in indoor environments. The innovative approach, which combines mapping and field measurements, demonstrated an increase in the accuracy of predictions, promoting advances in the efficiency and reliability of wireless networks in indoor spaces.**

***Index Terms* — Signal prediction , propagation model , neural networks , perceptron's .**

I. INTRODUCTION

The widespread use and availability of wireless networks has revolutionized modern connectivity, enabling seamless data transmission and interaction between devices. However, the complex challenges associated with signal propagation in indoor environments, characterised by obstacles, interference, and signal attenuation, represent obstacles to ensuring consistent and reliable performance of wireless networks. This work addresses these challenges by introducing an innovative tool created using the Python programming language. The tool uses neural networks to improve the prediction and accuracy of wireless network signals, based on measured data for mapping and training on real-world measurements.

This contribution focus on a comparative analyses of path loss propagation models in indoor industrial environments at 2.4 GHz Industrial, Scientific and Medical (ISM) band and 5.0 GHz, Wi-Fi.

This work was supported in part by the CAPES/Mack-Pesquisa Program at U.P.M,which financed equipment and scholarships for the authors.

Breno Batista Nascimento Silva is an undergraduate student in Electrical Engineering at the U.P.Mackenzie School of Engineering (E.E) -São Paulo/SP-Brazil. (batistabreno03@gmail.com)

The main objective of this article was to develop a tool based on an empirical propagation model to provide first- order coverage prediction results in indoor environmentsusing low-cost tools. The starting point of this work is the one-slope model for training the neural network that was implemented.

II. PATH LOSS PROPAGATION MODELS

Radio frequency signals are the main mechanism for propagating information. The basic model of radio propagation is based on the transmitter and receiver and the transmission medium. Propagation in confined media occurs when the electromagnetic wave passes through a material medium in a closed environment, thus limiting a region of space where multiple reflections of signal infractions can occur.

A. One – Slope Model

The path loos in dB is given by

$$L_{dB} = L_{0,dB} + 10n\log(d) \quad (1)$$

where L0,dB is the path loss obtained at distance of 1.0 m from the transmitter and path loss exponent n is determined

TABLE I – L(D0) FOR VARIOUS VALUES AND FREQUENCIES

L(d0) (dB)	
Frequency (MHz)	L0
900	31.5
1900	38.0
2400	40.2
4000	44.5
5300	46.9

E.T.C.Santos works as a professor of Electrical Engineering at U.P.Mackenzie in the Digital TV laboratory of the School of Engineering (E.E) -São Paulo/SP-Brazil. (edson.santos@mackenzie.br).

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.4 Web Link: https://dx.doi.org/10.18580/setijbe.2023.4

experimentally using a linear interpolation procedure [1].

III. NEURAL NETWORKS

Artificial Neural Networks (ANNs) are data structures based on the functioning of the human brain, it is a bio-inspired computational model, this data structure is made up of artificial neurons, which are inspired by natural neurons. The brain is a highly complex, non-linear and parallel computer (information processing system). It has the ability to organize its structural constituents, known as neurons, in such a way as to carry out certain processing (e.g. pattern recognition, perception and motor control) much faster than the fastest existing computer[2].

Neural networks have a network of artificial neurons that are interconnected, and through Learning Algorithms, simulate the decision-making capacity of the human brain. A neural network is a massively parallelized processor made up of simple processing units that have the natural propensity to store experiential knowledge and make it available for use.

It resembles the brain in two respects:

a) Knowledge is acquired by the network from its environment through a learning process.
b) Connection strengths between neurons, known as synaptic weights, are used to store the acquired knowledge.

The learning process of ANNs is one of the important qualities of these structures. The term "learning" corresponds to the process of adjusting the network's free parameters through a mechanism of presenting environmental stimuli, known as input or training patterns (or data): Stimulus -> adaptation -> new network behaviour.

There are basically three learning paradigms:

Supervised learning: also known as teacher learning, in which the teacher has knowledge of the environment and provides the desired input-response example set. Training is done using the error correction learning rule.

i. Unsupervised learning: there is no supervisor to evaluate the network's performance in relation to the input data. No error measure is used to feed back to the network. They generally employ a competitive learning algorithm (the network's output neurons compete to become active, with a single neuron winning the competition).

ii. Reinforcement learning: there is no direct interaction with a supervisor or specific model of the environment. Generally, the only information available is a scalar value that indicates the quality of the ANN's performance. During the learning process, the network tests some actions (outputs) and receives a reinforcement signal (stimulus) from the environment that allows it to evaluate the quality of its action.

IV. METHODOLOGY

The procedures of this study were structured in different stages, with the aim of evaluating the effectiveness of the Multilayer Perceptron Artificial Neural Network (MLP ANN) in predicting Wi-Fi signal loss in indoor environments. The stages were outlined as follows:

A. Creation and Training of the RNA-MLP

At this stage, the RNA-MLP was created and configured using the neurolab library in Python. The data collected, containing information on distance and Wi-Fi signal loss, was used to train the neural network. The structure of the RNA-MLP consisted of input layers, one or more intermediate layers and an output layer. Training was carried out using the gradient descent algorithm to adjust the network's synaptic weights, minimizing the prediction error.

```
#!pip install neurolab
import numpy as np
import matplotlib.pyplot as plt
import neurolab as nl

#Montar o Google Drive no Colab (caso os dados estejam no Google Drive)
from google.colab import drive
drive.mount('/content/drive')

#Caminho para o arquivo .txt
file_path = '/content/TesteSalaPrincAzimute(1).txt'

#Aquisição de Dados a partir de um arquivo .txt
mat = np.loadtxt(file_path)
vetDist = mat[:, 0].reshape(-1, 1)  # Dist em Metros
vetPerda = mat[:, 1].reshape(-1, 1)  # Perda em dbm
freq = 2412  # Frequencia em Mhz
```

Fig. 1 Data import via txt file in python Collab.

The code extract in question describes the creation, configuration, training and evaluation of an Artificial Neural Network (ANN) using the "neurolab" library in Python. The neural network is designed to predict Wi-Fi signal losses based on the distances between measurement points and the corresponding signal losses.

B. Creation of the Neural Network

The neural network is initialised using the newff() function from the "neurolab" library. In this case, the network is configured with an input layer, an intermediate layer with 24 neurons and an output layer with 1 neuron. The minimum and maximum distance ranges (vetDist) are supplied as input to the network's input layer.

```
#Criação de uma nova Rede Neural
redeneural = nl.net.newff([[np.min(vetDist), np.max(vetDist)]], [24, 1], [nl.trans.LogSig(), nl.trans.PureLin()])
```

Fig. 2 Create Neural network

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.4 Web Link: https://dx.doi.org/10.18580/setijbe.2023.4

C. *Definition of Neural Network Properties*

The neural network is configured to use the gradient descent training algorithm (train_gd) and the sum of squares error function (SSE()) to evaluate the prediction error. In addition, the neural network is initialised.

```
#Definição das propriedades da Rede Neural
redeneural.trainf = nl.train.train_gd
redeneural.errorf = nl.error.SSE()
redeneural.init()
```

Fig. 3 Definition of the Neural Network's properties.

D. *Neural Network Training*

The training stage is carried out using the neural network's train() method. In this case, the distance data (vetDist) and the corresponding signal losses (vetLoss) are used to train the neural network. Training is conducted for a specific number of epochs (in this case, 50,000 epochs) and the error value is displayed periodically (every 1x10e-25).

```
#Treinamento da Rede Neural
error = redeneural.train(vetDist, vetPerda, epochs=50000, show=1*10**-25)
```

Fig. 4 Neural Network Training

E. *Comparison to Field Tests*

At this stage, the results obtained by the RNA-MLP were compared with real data collected through field tests. The field tests involved directly measuring Wi-Fi signal strength at different distances within the environment. The comparison sought to verify the ability of the MLP-NRNA to accurately predict signal loss compared to practical observations.

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
The maximum number of train epochs is reached
Valores Medidos de Perda:
[-50. -48. -56. -55.]
Valores Calculados a partir da Rede de Perdas:
[-50.  -48.  -55.5 -55.5]
```

Fig. 5 Values collected in field measurements.

F. *Comparison with the OneSlope Model*

This stage involved comparing the results obtained by the RNA-MLP with the values calculated using the OneSlope propagation model. The OneSlope model is an analytical approach to estimating signal loss in indoor environments. The aim of the comparison was to assess the ability of the RNA-MLP to overcome the limitations of the traditional analytical model and provide more accurate predictions.

```
import numpy as np

def oneSlop(Ld0, n, dist):
    t = len(dist)
    L = np.zeros(t)

    for i in range(t):
        L[i] = Ld0 + 10 * n * np.log10(dist[i])

    return L
```

Fig. 6 One Slope function.

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
The maximum number of train epochs is reached
Valores Medidos de Perda:
[-50. -48. -56. -55.]
Valores Calculados a partir da Rede de Perdas:
[-50.  -48.  -55.5 -55.5]
peOneSlop [-48.76165386 -49.14388397 -49.51140923 -49.86531926 -50.20658692
 -50.53608435 -50.85459647 -51.16283218 -51.46143394 -51.75098575
 -52.03202009 -52.30502377 -52.57044302 -52.82868789 -53.08013605
 -53.3251361  -53.56401051 -53.79705813 -54.02455646 -54.24676365
 -54.46392024 -54.67625073 -54.88396496 -55.08725941 -55.28631825
 -55.48131438 -55.67241032 -55.85975904 -56.04350467 -56.2237832
 -56.40072305 -56.57444563 -56.74506585 -56.91269258 -57.07742905
 -57.23937325 -57.39861824 -57.55525255 -57.70936038 -57.86102196
 -58.01031372]
```

Fig. 7 Results of models.

G. *Error evaluation*

In this step, the average errors and squared errors for the MLP-NRNA predictions were calculated in relation to the actual results and the values estimated by the OneSlope model. This evaluation quantified the performance of the MLP-NRNA in terms of its ability to accurately predict Wi-Fi signal losses.

```
#Cálculo dos Erros
def ermedio(y_true, y_pred):
    return np.mean(np.abs(y_true - y_pred))

def ermquad(y_true, y_pred):
    return np.mean((y_true - y_pred) ** 2)

#Cálculo do Erro Médio e Erro Quadrático para a Rede Neural
ErroMedioRN = ermedio(vetPerda, perdasRN)
ErroQuadRN = ermquad(vetPerda, perdasRN)

#Cálculo do Erro Médio e Erro Quadrático para o OneSlop
ErroMedioOS = ermedio(vetPerda, peOneSlop)
ErroQuadOS = ermquad(vetPerda, peOneSlop)
```

Fig. 8 Calculation of the average error and squared error for the models used.

H. *Construction of graphs for data presentation*

In order to better visualise the models used, graphs were drawn in Python, using the matplotlib.pyplot library. This library has a wide variety of graphs and in this research the point graph was used, based on the distance and losses measured.

```
plt.plot(vetDist , vetPerda,  'r', linewidth=5)
plt.plot(vetDist,perdasRN, 'b', linewidth=2)
plt.plot(d, peOneSlop ,'g', linewidth=3)
#plt.plot(vetDist, pfriss, 'black',linewidth=4)


plt.legend(['Valores Reais', 'Rede Neural', 'OneSlope', 'Friss'])
plt.title('Distância x Perdas')
plt.xlabel('Distância (m)')
plt.ylabel('Perdas (dBm)')

plt.show()
```

Fig. 9 Code developed for creating graphs and presenting results.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.4 Web Link: https://dx.doi.org/10.18580/setijbe.2023.4

V. NUMERICAL AND MEASUREMENT RESULTS

The test environment was Building 6 of the School of Engineering, which has three floors of classrooms. The measurements were carried out on the 5 GHz Wi-Fi signal.

The Table II contains information on 5G signal loss measurements carried out in different rooms, where the crucial input for the prediction is the distance between the router and the measurement point.

TABLE II – MEASUREMENTS TAKEN TO TRAIN THE NEURAL NETWORK

BUILDING 6/ 3rd FLOOR						
Environment	Heights (m)	Length (m)	Depth (m)	Distância Roteador – Ponto (m)	Losses measured in Notebook(dB)	Losses measured in Mobile (dB)
302	3.65	7.51	11.1	1	-31	-45
311	3.96	6.85	6.13	20	-87	-89
312	3.95	8.44	6.07	16.7	-89	-90
313/315	3.98	6.3	12.7	13.5	-86	-76
314	3.96	8.45	6.6	13	-85	-85
304	3.98	7.16	8.8	-	-	-
305	4.02	6.15	8.4	-	-	-
306	4.01	8.84	8.47	-	-	-
307	4.01	6.17	6.45	-	-	-
308	4.16	9.91	6.44	-	-	-
309	4.03	7.1	7.81	-	-	-

The measurement procedure was carried out as follows :

Room: The number of the room or environment where the measurements were taken.

Distance Router - Point: The distance in some standard of measurement (such as metres) between the router (signal source) and the measurement point inside the room.

Notebook: The amount of signal loss in decibels when measured with a notebook.

Mobile phone: The amount of signal loss in decibels when measured with a mobile phone.

As seen in the Table II the signal is no longer detected by the notebook's measurement software or by the mobile phone from room 314. since rooms 311, 312, 313/315 and 314 are geographically close to room 302. where the 5G signal transmitter is located. Now let's understand how this data can be used to train an Artificial Neural Network (ANN):

a) ANN input: The distance between the router and the measurement point (Router - Point Distance) will be used as the input for the ANN. This means that the ANN will learn the relationship between distance and signal loss.

b) Desired ANN output: The desired ANN output is the predicted signal loss. You can choose to use the measurements made with the notebook (Notebook) or with the mobile phone (Mobile) as the target output for training.

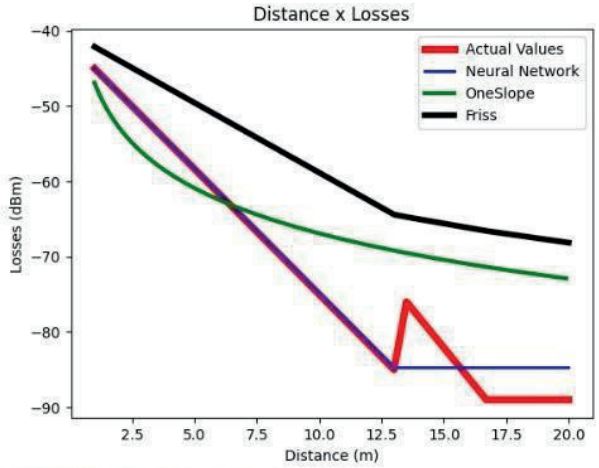
c) Data preparation: The distance will be normalised. making it compatible with the requirements of the ANN. This normalisation ensures that all the inputs are in the same range of values, which helps with training.

d) Creating the ANN: The ANN will be designed with an architecture that includes an input layer (corresponding to the distance), one or more hidden layers and an output layer.

e) Training the ANN: It was trained with the normalised distances as input and the signal losses (Notebook or Mobile) as target output. During training, the ANN will adjust its weights to minimise the error between the predictions and the actual values.

f) Evaluation and Adjustment: Performance is evaluated using error metrics such as the Mean Absolute Error (MAE) or the Mean Squared Error (MSE). calculated with the neural network predictions and the actual losses. Adjustments to the hyperparameters can be made based on these results.

g) Using the trained neural network: After training, the neural network can be used to predict signal losses in new rooms based on the distances between the routers and the measurement points. This is useful for estimating the quality of the 5G signal at different distances.




```
Mean Neural Network Error: 3.50000002174927
Neural Network Quadratic Error: 22.549999999554664
OneSlop Average Error  : 19.478772908747395
OneSlop Quadratic Error: 440.97505160924834
```

Fig 10 . Losses measured on the 3rd floor.

The Figure 10 shows the values obtained from measurements made on a notebook and a mobile phone to obtain reference values for training the neural network at the same observation point. The neural network is being trained with the values presented and in relation to the reference model.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.4 Web Link: https://dx.doi.org/10.18580/setijbe.2023.4

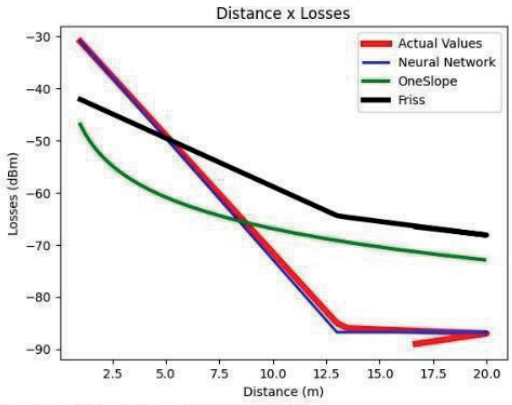


Fig 11 . Losses measured on the 3rd floor.

The Figure 11 shows the values obtained and the convergence of the measurements for the trained neural network. There is a loss of reference for the most distant values.

VI. CONCLUSION

This work developed a tool based on the Python programming language, in the COLAB programming environment, to predict signals in wireless networks in indoor environments, using neural networks as a method of improving the accuracy of predictions. The main focus was on dealing with the challenges of signal propagation indoors, where obstacles, interference and attenuation can cause significant variations in wireless network signals.

The main objective of the project was to face these challenges by combining data obtained through detailed mapping of the indoor environment and real measurements of signal strength. The use of neural networks has demonstrated effectiveness in creating more accurate predictive models compared to traditional approaches. The developed Python tool allowed training and adjusting these models based on the collected information, considering the physical structure of the environment, present obstacles, and construction materials. The results obtained indicated significant improvements in the accuracy of forecasts compared to conventional methods, showing a great potential for the application of the tool in real-world scenarios. This includes planning and optimizing wireless networks in indoor environments, contributing to more stable and reliable connectivity. The project employed the methodology of creating and training an Artificial Neural Network to predict Wi-Fi signal losses, comparing the results of the neural network with field measurements and a traditional analytical model (OneSlope model). Comparison with real measurements and the analytical model revealed the ability of RNA-MLP to overcome the limitations of the traditional model and provide more accurate predictions, especially in scenarios with signal obstructions.

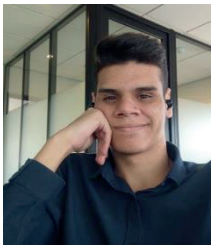
ACKNOWLEDGMENT

We would like to thank the CAPES/MackPesquisa fund for supporting the scholarship and the research. We would also like to thank the engineering school for providing their laboratories and rooms for the tests. We would like to thank our colleagues at the U.P.M. Digital TV Laboratory for their support and access to the measurement equipment.

REFERENCES

[1] K. Pahlavan and A. H. Levesque, Wireless Information Networks 2thed., Ed. Wiley, Chichester, England, 2005, page(s): 1 – 4.

[2] HAYKIN, Simon. Redes Neurais: Princípios e prática. Porto AlegreRS:Bookman, 2001.



Breno Batista Nascimento Silva was born in the city of Sousa-PB-Brazil on 14 December 2000. He graduated from the Electronics Technician course at SENAI- Guarulhos-SP-Brazil, where he joined at the age of 15. He is currently in the final year of his bachelor's degree in electrical engineering, which he entered in 2019 at Mackenzie Presbyterian University. He has a scholarship from the Mackenzie Presbyterian Institute. He was a student on a scientific initiation scholarship (PIBIC) from CAPES/Mackpesquisa from 2022 to 2023. His research and interests are in Artificial Neural Networks (ANNs) using the Python language. He currently works as an Accounting Measurement Analyst at the Energy Commercialization Chamber (CCEE).

Edson Tafeli Carneiro dos Santos obtained a bachelor's degree of Electrical Engineering, in 1993, at the FEI Industrial Engineering College, located in São Bernardo do Campo, São Paulo, Brazil. He acquired his Master's degree in Electrical Engineering at Universidade Mackenzie, São Paulo, Brazil, in 2007; and his PhD from the same institution in 2015. He was an assistant professor at the Mauá School of Engineering between 2012 and 2016. He has been working at the School of Engineering at Mackenzie Presbiterian University since 1999 and is currently an assistant professor and researcher at the Digital TV Laboratory. His research focuses on Electromagnetic Theory, Microwaves, Wave Propagation and Antennas, working on the following topics: Broadband Antennas, Dual Polarisation Antennas, Wireless Sensor Networks, Plasma and Electromagnetic Simulation.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.4 Web Link: https://dx.doi.org/10.18580/setijbe.2023.4

Amplifying In-Vehicle DTV Entertainment: ATSC 3.0 Broadcast Signal Relay via WiFi Gateway

Sungjun Ahn
Yongsuk Kim
Sung-Ik Park

Cite this article:

Ahn, Sungjun; Kim, Yongsuk; Park, Sung-Ik; 2023. Amplifying In-Vehicle DTV Entertainment: ATSC 3.0 Broadcast Signal Relay via WiFi Gateway. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2023.5. Web Link: https://dx.doi.org/10.18580/setijbe.2023.5

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING - SET IJBE V.9, 2023, Short Paper 1, 2 p
2023 SET - Brazilian Society of Television Engineering / ISSN (Print): 2446-9246 / ISSN (Online): 2446-9432

Amplifying In-Vehicle DTV Entertainment: ATSC 3.0 Broadcast Signal Relay via WiFi Gateway

Sungjun Ahn, *Member, IEEE,* Yongsuk Kim, and Sung-Ik Park, *Fellow, IEEE*

***Abstract*—This paper presents the relayed distribution of ATSC 3.0 broadcast signals to mobile users in moving vehicles. The gateway relay featured in this work seamlessly converts received ATSC 3.0 signals into a WiFi interface. This proposal exhibits the use of an ATSC 3.0-to-WiFi gateway to amplify broadcasting media in automotive, allowing personalized experience on individual seat positions.**

***Index Terms*—ATSC 3.0, in-vehicle entertainment, mobile broadcasting, ATSC 3.0-to-WiFi gateway.**

I. Introduction

In recent years, the demand for in-vehicle entertainment and connectivity has surged, fueled by the rising prevalence of self-driving technology and the ubiquitous consumption of digital media on mobile devices. The community's approach, as a response, has first focused on developing technologies that facilitate direct-to-vehicle (D2V) content delivery. The major concern in this development has been building sufficient reliability to cope with dynamic channel situations. Notably, the use of multi-antenna diversity [1] and broadcast-broadband cooperation based on dual connectivity [2] have been proposed as solutions.

Within dynamic automotive environments, it is known as quite demanding to serve rich video content to every passenger. Considering the cellular networks these days, various physical obstacles and traffic problems incur frequent streaming interruptions. Moreover, from the user's view, it is also demanding to rely on paid data channels for streaming huge amounts of video data during a long journey on the road. Advanced Television Systems Committee (ATSC) has long remarked on such issues and has made careful efforts to support vehicle broadcasting from the very first stage of developing a new standard, ATSC 3.0. The broadcasting-based solutions such as [1] have hence been highlighted for this use case.

In fact, specific ideas to serve each individual passenger's device have been less identified so far. Such sort of detail has been recognized as the next step after building the D2V connectivity. Nonetheless, since the D2V supply is actually being embodied in the real world, it is no more a future work to hold off. This paper, in this context, introduces a feasible solution to build a bridge from air ATSC 3.0 signals to the end devices inside a vehicle.

Particularly, this is a showcasing of a WiFi gateway operating as an ATSC 3.0 relay with interface conversion ability. The presented gateway system seamlessly captures the received broadcast signals and converts them into a format compatible with WiFi-enabled devices, ensuring a smooth and uninterrupted streaming experience. Accordingly, the individual users at the seat are allowed to enjoy content on their own personalized displays in convenient positions. This paper presents the architecture design of the ATSC 3.0-to-WiFi gateway and its actual use in automotive systems. With the advent of this gateway system, the momentum of D2V broadcasting will be amplified, as it allows passengers to access a diverse array of ATSC 3.0 broadcast content on their personalized devices and displays while on the move.

II. Design and The Use for Mobile Broadcasting

Fig. 1 illustrates the concept of the ATSC 3.0-to-WiFi gateway. Detailed design and applications will be included in the final manuscript.

The benefits extend beyond entertainment alone. The gateway system also opens up possibilities for educational content delivery, emergency broadcasts, and location-specific information dissemination to enhance the overall in-vehicle experience.

III. Conclusion

This paper addressed the challenge of delivering ATSC 3.0 broadcast signals to mobile users within a moving vehicle, leveraging the concept of a gateway system that converts these signals into a WiFi interface. The presented gateway system is a compact solution to provide terrestrial broadcast content to personal mobile devices, acting as a bridge between the ATSC 3.0 over the air and the WiFi network within the vehicle. Penetration loss, cabling burden, and position-dependent accessibility problems are hence resolved, thereby offering an enjoyable media experience condition. With the assistance of this vehicle gateway system, digital terrestrial broadcasting will be pleasantly embraced into infotainment systems in automotive, and would subsequently propel the expansion of D2V opportunities beyond media entertainment.

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (RS-2023-00224660, Development of Receiver Chip for ATSC 3.0 Mobile Broadcast).

Sungjun Ahn and Sung-Ik Park are with the Media Research Division, Electronics and Telecommunications Research Institute (ETRI), 218 Gajeong-ro, Yuseong-gu, Daejeon, 34129, South Korea (e-mail: {sjahn, psi76}@etri.re.kr).

Yongsuk Kim is with LowaSIS, Inc., Geeongdae-ro 17-gil, Buk-gu, Daegu, 41566, South Korea (e-mail: yskim@lowasis.com).

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.5 Web Link: https://dx.doi.org/10.18580/setijbe.2023.5

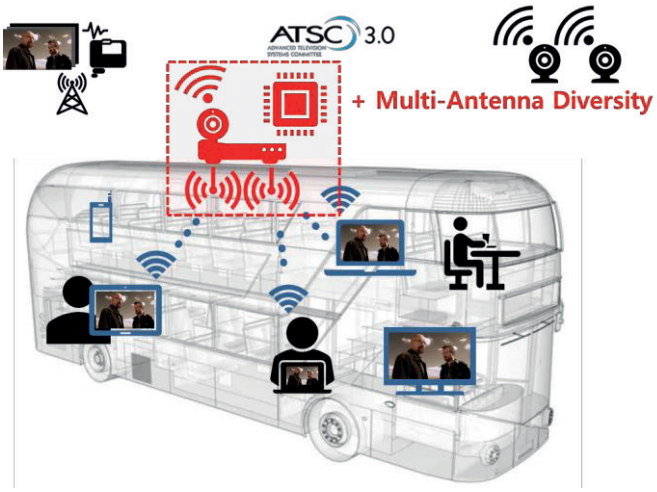


Fig. 1. Conceptual description of the ATSC 3.0-to-WiFi gateway mounted on vehicle: Mass transportation case.

REFERENCES

[1] S. Ahn *et al*., "Multi-antenna diversity gain in terrestrial broadcasting receivers on vehicles: A coverage probability perspective," *ETRI Journal*, 2021.

[2] S. Ahn *et al*., "Cooperation between LDM-based terrestrial broadcast and broadband unicast: On scalable video streaming applications," *IEEE Trans. Broadcast*., vol. 67, no. 1, pp. 2–22, Mar. 2021.

Received in 2023-06-07 I Approved in 2023-08-04

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.5 Web Link: https://dx.doi.org/10.18580/setijbe.2023.5

Features and Applications of ATSC 3.0 Transmitter Identification (TxID)

Bo-mi Lim
Sunhyoung Kwon
Sungjun Ahn
Sung-Ik Park
Namho Hur

Cite this article:

Lim, ABo-mi; Kwon, Sunhyoung; Ahn, Sungjun; Park, Sung-Ik; Hur, Namho; 2023. Features and Applications of ATSC 3.0 Transmitter Identification (TxID). SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2023.6. Web Link: https://dx.doi.org/10.18580/setijbe.2023.6

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING - SET IJBE V.9, 2023, Short Paper 2, 2 p
2023 SET - Brazilian Society of Television Engineering / ISSN (Print): 2446-9246 / ISSN (Online): 2446-9432

Features and Applications of ATSC 3.0 Transmitter Identification (TxID)

Bo-mi Lim, Sunhyoung Kwon, Sungjun Ahn, Sung-Ik Park, and Namho Hur

***Abstract*—Transmitter identification (TxID) is uniquely assigned to each transmitter to identify and control the transmitters in Advanced Television Systems Committee (ATSC) 3.0 broadcast networks, especially on a single frequency network (SFN). A transmitter also generates the TxID signal in addition to the ATSC 3.0 host signal but combines them, resulting in interfering with each other. This paper summarizes the TxID technique in ATSC 3.0 physical layer standard, including the detection performance, influences on the host ATSC 3.0 signal, and applications.**

***Index Terms*—ATSC 3.0, TxID, transmitter identification, SFN**

I. Introduction

BY adopting orthogonal frequency division multiplexing (OFDM) for the second digital broadcast, the broadcasters can actively apply a single frequency network (SFN) under insufficient frequency bands. To construct a more efficient SFN, the Advanced Television Systems Committee (ATSC) 3.0 standard supports the centralized transmitter control system based on the broadcast gateway and transmitter identification (TxID) [1]-[3]. TxID is a unique value to identify each transmitter in the nationwide broadcast area. Also, transmitters generate TxID signals to differentiate signals from which transmitters come on the receiver side. This paper briefly introduces the TxID technique in ATSC 3.0 and considers the detection performance of the TxID signal in addition to applications.

II. Transmitter Identification (TxID)

A. Features

TxID is a unique value between 0 to 8191 to identify the individual transmitter in ATSC 3.0 broadcast coverage areas. Therefore, a broadcast gateway controls the emission time, carrier offset, and the multiple input single input (MISO) filter of each transmitter in addition to passing common broadcast streams. Transmitters can generate the TxID signal depending on their own TxID values apart from the ATSC 3.0 signal, as shown in Fig. 1. As a result, the receiver can separate the received signals that appear as a superposition of transmit signals from multiple transmitters like multipath delayed fading channel and analyze their relative delays and amplitudes. The transmitter generates the unique binary Gold sequence with a length of 8192 depending on the assigned

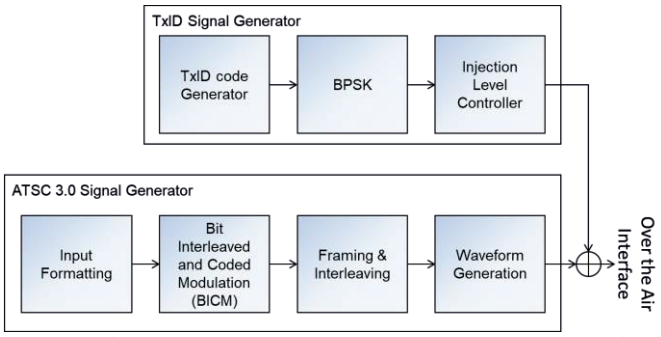


Fig. 1. Block diagram of ATSC 3.0 transmitter employing channel bonding.

TxID value and modulates the sequence as binary phase shift keying (BPSK). The modulated signal is time-synchronously injected on the first preamble symbol of the ATSC 3.0 host signal with 9 to 45 dB lower power. Therefore, two signals interfere with each other.

While the TxID is much lower than the ATSC 3.0 preamble signal, it can be detectable due to enough processing gain resulting from a direct sequence spread spectrum (DSSS) with an 8192 symbol length. The theoretical processing gain of the TxID signal is about 39 dB [4]. In addition, as the TxID signal is always aligned with the first preamble symbol, it is repeatedly injected according to the size of fast Fourier transform (FFT). For example, in the case of 32K FFT size, the TxID signal is continuously delivered four times. Therefore, the processing gain increases by 3 dB when the FFT size doubles. In [4], the authors dealt with the detection performance of the TxID signals and proposed the detection schemes. Since the TxID signal is the same over the transmission period, the detection performance may enhance after ensemble averages of receiver signals. Also, removing the preamble signal from the received signal significantly improves the reception performance.

The TxID signal also interferes with the ATSC 3.0 preamble signal. As the injection level gets larger, the TxID signal can be easily detected without an advanced detection algorithm, but the preamble deteriorates more. Therefore, there is a trade-off between the detection performance of the TxID signal and the preamble signal. In [5] and [6], the impact of the TxID signal on the preamble detection is considered both theoretically and practically with respect to the injection levels of the TxID signal and the protection modes of Layer 1 (L1) signaling data, L1-Basic and L1-Detail, conveyed in the preamble symbol. Protection modes

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2017-0-00081, Development of Transmission Technology for Ultra High Quality UHD).

Bo-mi Lim, Sunhyoung Kwon, Sungjun Ahn, Sung-Ik Park, and Namho Hur are with the Media Research Division, Electronics and Telecommunications Research Institute (ETRI), 218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700 South Korea (e-mail: {blim_vrossi46, shkwon, sjahn, psi76, namho}@etri.re.kr).

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.6 Web Link: https://dx.doi.org/10.18580/setijbe.2023.6

1 and 2, generally used in ATSC 3.0 broadcasting system, are less influenced by the TxID signal. Therefore, the TxID signal might be provided all the time, not just when the broadcasters build their broadcast networks.

B. Applications

TxID signal enriches the broadcast service coverage, avoiding the deep nulls caused by multiple transmitting signals simultaneously arriving on the receiver side. In Seoul metropolitan areas [7], [8] and Jeju areas [9], the broadcasters examined the coverage areas with poor reception performance and enough signal strength based on the TxID signal analysis. By adjusting transmit time delays among transmitters, the reception performance might be improved. Even more, TxID signals enable channel modeling under a rich scattered fading environment [10]. By decoding TxID, the receiver can recognize which broadcast service areas it belongs to, so the broadcaster can provide the local service application like a geo-targeted advertisement.

III. Conclusion

This paper considered the TxID technique in the ATSC 3.0 physical layer standard. Though the TxID signal may interfere with the host ATSC 3.0 signal, it serves processing gain. It enables the broadcaster to manage the broadcast network in a centralized way, orchestrate transmitters, and provide localization services.

References

[1] ATSC Standard: A/322, Physical Layer Protocol, document A/322:2023, Advanced Television System Committee, Washington, DC, USA, March 2023.

[2] ATSC Standard: A/324, Scheduler/Studio to Transmitter Link, document A/324:2023, Advanced Television System Committee, Washington, DC, USA, March 2023.

[3] S.-I. Park *et al.,* "ATSC 3.0 Transmitter Identification Signals and Applications," *IEEE Trans. Broadcast*., vol. 63, no. 1, pp. 240-249, March 2017.

[4] S. Kwon *et al.,* "Detection Schemes for ATSC 3.0 Transmitter Identification in Single Frequency Network," *IEEE Trans. Broadcast*., vol. 66, no. 2, pp. 229-240, June 2020.

[5] B.-m. Lim *et al., "*Laboratory Test Analysis of TxID Impact into ATSC 3.0 Preamble," *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Valencia, Spain, June 2018, pp. 1-3.

[6] B.-m. Lim *et al.,* "Performance Evaluation of ATSC 3.0 Preamble for TxID Signal-Injected Use Cases," *2022 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Bilbao, Spain, June 2022, pp. 1-5.

[7] J. Lee *et al*., "Transmitter Identification Signal Detection Algorithm for ATSC 3.0 Single Frequency Networks," *IEEE Trans. Broadcast*., vol. 66, no. 3, pp. 737-743, Sep. 2020.

[8] S. Jeon *et al*., "Field Trial Results for ATSC 3.0 TxID Transmission and Detection in Single Frequency Network of Seoul," *2018 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting (BMSB)*, Valencia, Spain, 2018, pp. 1-4.

[9] B. Lim *et al*., "Field Evaluation of Transmit Diversity Code Filter Sets in ATSC 3.0 Single Frequency Networks," *IEEE Trans. Broadcast*., vol. 68, no. 1, pp. 191-201, March 2022.

[10] S. Ahn *et al*., "Characterization and Modeling of UHF Wireless Channel in Terrestrial SFN Environments: Urban Fading Profiles," in *IEEE Transactions on Broadcasting*, vol. 68, no. 4, pp. 803-818, Dec. 2022.

Received in 2023-06-07 I Approved in 2023-08-04

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.6 Web Link: https://dx.doi.org/10.18580/setijbe.2023.6

Roads of MIMO Broadcasting: An Overview of Variant Technologies

Sung-Ik Park
Bo-mi Lim
Hoiyoon Jung
Namho Hur
Sungjun Ahn

Cite this article:

Park, Sung-Ik; Lim, Bo-mi; Jung, Hoiyoon; Hur, Namho; Ahn, Sungjun; 2023. Roads of MIMO Broadcasting: An Overview of Variant Technologies. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2023.7. Web Link: https://dx.doi.org/10.18580/setijbe.2023.7

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING - SET IJBE V.9, 2023, Short Paper 3, 2 p
2023 SET - Brazilian Society of Television Engineering / ISSN (Print): 2446-9246 / ISSN (Online): 2446-9432

Roads of MIMO Broadcasting: An Overview of Variant Technologies

Sung-Ik Park, *Fellow, IEEE,* Bo-mi Lim, *Member, IEEE,* Hoiyoon Jung, *Member, IEEE,* Namho Hur, *Member, IEEE,* and Sungjun Ahn, *Member, IEEE*

***Abstract*—This paper outlines futuristic ATSC 3.0 multiple-input multiple-output (MIMO) broadcasting technologies in terms of three variants: 1) Frequency reuse-1 MIMO, 2) Backward-compatible MIMO, and 3) Channel-bonded MIMO. Through a brief discussion of their principles, features, and use cases, this paper sheds light on the diverse paths of MIMO broadcasting paved for the future of digital broadcasting systems. Furthermore, the paper discusses the implications and potential advancements of these technologies, emphasizing their role in achieving higher data rates and improved flexibility.**

***Index Terms*—MIMO broadcasting, ATSC 3.0, reuse-1 MIMO, backward compatible MIMO, channel bonding.**

I. Introduction

LATELY, the actual services of the new digital terrestrial broadcasting standard, Advanced Television Systems Committee (ATSC) 3.0, have been commenced in several countries [1]. As a start-off for this new-generation broadcasting ecosystem, the launches in the United States and South Korea have departed from a generic single-input single-output (SISO) topology [2]. Since such baseline deployment has been brought into reality, pivoting from this milestone, the broadcast community is promptly preparing for the next step.

Distributing multiple-input multiple-output (MIMO) is recognized as one of the possible directions. The standard suite of ATSC 3.0 has already included MIMO operations and defined the concrete system chain [3]. This inclusion has primarily been for increased data capacity, where it doubles the transmission channel in a naïve sense. The ATSC 3.0 MIMO physically relies on cross-polarization, and stationary environments with well-pivoted directional antennas will be its primary target use case [4].

As mentioned, the first aim of building such MIMO technology has been at capacity doubling, which will bring more rich media quality or a diversified array of content [5]. However, the world is encountering greatly divergent local situations, which seek different values or are constrained differently by unique states of affairs. Such diversity, as a consequence, necessitates dedicated system evolutions into variant forms.

In this paper, we introduce the evolutions of MIMO broadcasting technology on the ATSC 3.0 basis, also enlightening the particular need, use cases, and the background behind them. This report starts from a *reuse-1 MIMO,* which is currently a special interest of the Brazilian broadcasting community, and continues with *backward compatible (B-Comp) MIMO* and *channel-bonded (CB) MIMO* that have emerged from other contexts. Essential characteristics are discovered, leading to a comprehensive understanding of these technologies.

II. Current Status of ATSC 3.0 MIMO

Having the basic SISO-form ATSC 3.0 deployed in the real world, broadcasters have subsequently started preparing to bring ATSC 3.0 MIMO to the earth.

The broadcasters in South Korea are envisaging two possibilities for ATSC 3.0 MIMO: (*i*) A way more enriched ultra-high-definition (UHD) video service with 8K resolution [6], [7]; and (*ii*) an integral of multiple 4K UHD programs in the same frequency channel, where each program is from the different service provider. Principally, South Korea pursues high-quality and enriched videos more than other features. The latter imagination (*ii*) is conceived as appealing to the practitioners because it can create new business opportunities and stimulate the network operator's role.

III. Emerging Technologies Based on ATSC 3.0 MIMO

A. Reuse-1 MIMO

Reuse-1 MIMO is a topology allowing the coexistence of plural, different, uncoordinated MIMO service signals in the same single radio frequency. Shortly speaking, multiple different service providers here share the same frequency channel [8], [9]. The powerful error protection capability of ATSC 3.0 enables this system, allowing the receiver to decode the desired signal successfully from a noisy mixture of plural MIMO service signals.

Brazilian broadcasting enablers are especially interested in this topology, particularly concerned with Brazil's spectrum circumstance. The new Brazilian broadcasting standard project, so-called *TV 3.0*, announced the mandate of reuse-1 operability on a MIMO basis.

This measure was to create additional data capacity while coping with an oversaturated spectrum issue. Brazil's radio spectrum dedicated to terrestrial broadcasting is devastatingly saturated since there are so many on-air programs ongoing simultaneously. To resolve this problem, Brazil is attempting

This work was supported by Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (2022-0-00923, Development of Transceiver Technology for Terrestrial 8K Media Broadcast).

Sung-Ik Park, Bo-mi Lim, Hoiyoon Jung, Namho Hur, and Sungjun Ahn are with the Media Research Division, Electronics and Telecommunications Research Institute (ETRI), 218 Gajeong-ro, Yuseong-gu, Daejeon, 305-700 South Korea (e-mail: {psi76, blim_vrossi46, junghy, namho, sjahn}@etri.re.kr).

to build a totally new, MIMO-based network foundation that would be not backward compatible.

The receiver at an arbitrary spot can tune to the desired service signal by pivoting the receive antenna properly, aligning it to the desired source's direction, cutting off the undesired signals by leaving them somewhere off the beam. Combined with the robust channel coding of ATSC 3.0, this assessment facilitates reuse-1 MIMO network even though many service providers are sharing the frequency, making the signal space *crowded*.

1) Single frequency network (SFN) with MIMO broadcasting

MIMO SFN could be considered a counterpart of reuse-1 MIMO, while single-frequency channel transmissions underlie both technologies. MIMO SFN lets clusters of towers transmit the same MIMO signal with centralized coordination, whereas the reuse-1 MIMO gives a mixture of different MIMO service signals.

B. B-Comp MIMO

The concept B-Comp MIMO has emerged from the countries that have already commenced ATSC 3.0 SISO services. This is considered a lubricating technology that assists a soft transition from SISO to MIMO ecosystem, or a spectrally efficient platform to embrace diverse target device-ends in the same frequency channel [10]-[13].

Specifically, B-Comp MIMO is a co-transmission of SISO and MIMO signals [10]. To this end, the physical layer multiplexing between them can rely on time division multiplexing (TDM) or layered division multiplexing (LDM). For example, B-Comp MIMO can harness the benefits of MIMO technology, serving dedicated MIMO terminals equipped with dual-polarized antennas, while serving SISO-based (physically constrained) mobile terminals and legacy television sets simultaneously in the same physical layer frame. As is designed, both types of end-terminals operate without any conflict.

C. CB MIMO

In terms of capacity amplification, CB MIMO goes one step further than the original ATSC 3.0 MIMO. CB MIMO utilizes two, consecutive or non-consecutive frequency channels along with leveraging cross-polarized MIMO technology at the same time [14]. This is, in essence, an integration of channel bonding and MIMO both defined in ATSC 3.0 physical layer.

By employing parallel transmission paths, ATSC 3.0-based CB MIMO is expected to provide up to about 200 Mbps data capacity.

IV. Conclusion

This paper introduced several variants of ATSC 3.0 MIMO to summarize the evolution of MIMO broadcasting technology. Reuse-1 MIMO, B-Comp MIMO, and CB MIMO were investigated, whose target markets deviate in different directions. For each technology, we exhibited the backgrounds and features. The implications and potential advancements of these technologies were discussed, emphasizing their role in achieving higher data rates and improved flexibility.

References

[1] S. Ahn *et al*., "Characterization and modeling of UHF wireless channel in terrestrial SFN environments: Urban fading profiles," *IEEE Trans. Broadcast*., vol. 68, no. 4, pp. 803-818, Dec. 2022.

[2] S.-I. Park *et al*., "Performance analysis of all modulation and code combinations in ATSC 3.0 physical layer protocol," *IEEE Trans. Broadcast*., vol. 65, no. 2, pp. 197-210, Jun. 2019.

[3] D. Gomez-Barquero *et al*., "MIMO for ATSC 3.0," *IEEE Trans. Broadcast*., vol. 62, no. 1, pp. 298-305, Mar. 2016.

[4] E. Garro *et al*., "Layered division multiplexing with co-located multiple-input multiple-output schemes," *IEEE Trans. Broadcast*., vol. 66, no. 1, pp. 9-20, Mar. 2020.

[5] S. Ahn *et al*., "ATSC 3.0 for future broadcasting: Features and extensibility," *SET Int. J. Broadcast Eng*., vol. 6, pp. 21-36, Dec. 2020.

[6] H. Jung *et al*., "Feasibility verification of ATSC 3.0 MIMO system for 8K-UHD terrestrial broadcasting," *IEEE Trans. Broadcast.*, vol. 67, no. 4, pp. 909-916, Dec. 2021.

[7] S. Ahn *et al*., "Converged distribution of 5G Media: Opportunities of overlaid broadcast and emerging applications over dual connectivity," *IEEE Trans. Broadcast.*, vol. 68, no. 2, pp. 501-516, Jun. 2022.

[8] Y. Wu *et al.*, "Cloud transmission: A new spectrum reuse friendly digital terrestrial broadcasting transmission system," *IEEE Trans. Broadcast*., vol. 58, no. 3, pp. 329-337, Sept. 2012.

[9] J. Montalban *et al.*, "Cloud transmission: System performance and application scenarios," *IEEE Trans. Broadcast.*, vol. 60, no. 2, pp. 170-184, Jun. 2014.

[10] J. Kang *et al*., "Feasibility of backward compatible MIMO broadcasting: Issues in SISO-MIMO coexistence," *IEEE Trans. Broadcast*., vol. 69, no. 2, pp.589-609, Jun. 2023.

[11] Y. Wu *et al*., "Inter-tower communications network signal structure, and interference analysis for terrestrial broadcasting and datacasting," *IEEE Trans. Broadcast*., vol. 69, no. 2, pp. 610-616, Jun. 2023.

[12] Z. H. Hong *et al.*, "Implementation of wireless backhaul and inter-tower communications with MIMO in ATSC 3.0," *IEEE Trans. Broadcast*., vol. 69, no. 2, pp. 579-588, Jun. 2023.

[13] E. Iradier *et al*., "Guest editorial special Issue on inter-tower communications and networks," *IEEE Trans. Broadcast*., vol. 69, no. 2, pp. 553-559, Jun. 2023.

[14] L. Stadelmeier *et al*., "Channel bonding for ATSC 3.0," *IEEE Trans. Broadcast*., vol. 62, no. 1, pp.289-297, Mar 2016.

Received in 2023-06-07 I Approved in 2023-08-04

System Verification of Advanced ISDB-T Gateway

Kohei Kambara

Cite this article:

Kambara, Kohei; 2023. System Verification of Advanced ISDB-T. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2023.8. Web Link: https://dx.doi.org/10.18580/setijbe.2023.8

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING - SET IJBE V.9, 2023, Short Paper 4, 2 p
2023 SET - Brazilian Society of Television Engineering / ISSN (Print): 2446-9246 / ISSN (Online): 2446-9432

System Verification of Advanced ISDB-T

Kohei Kambara

***Abstract*—The Advanced Integrated Services Digital Broadcasting-Terrestrial (Advanced ISDB-T) is a next generation digital terrestrial broadcasting system. From 2019, various experiments have been conducted and the fundamental performances were verified. In 2022, as for the total system verification test, the large-scale verification tests were conducted in 4 large cities in Japan. The verification test with using actual hardware including transmitter station showed the feasibility of the system.**

***Index Terms*—Digital Terrestrial Broadcasting, Advanced ISDB-T,**

I. Introduction

The first-generation digital terrestrial television broadcasting using the Integrated Services Digital Broadcasting-Terrestrial (ISDB-T) system[1] was launched in 2003 in Japan. Since then, ISDB-T system was adopted in 20 countries including Japan. More than 20 years have passed since ISDB-T system was developed, and during these years, there are various evolutions on broadcasting technologies. Ultra-high definition television (UHDTV) satellite broadcasting services using the Integrated Services Digital Broadcasting for Satellite, 3rd generation (ISDB-S3) system [2, 3] were launched in Japan in 2018, and UHDTV has become much popular recently.

II. Advanced ISDB-T System

In order to improve the user experience and for the efficient use of terrestrial broadcasting frequency band, we are currently developing the advanced terrestrial broadcasting system for the next generation of digital terrestrial television broadcasting. For the physical layer we have developed the transmitting system Advanced Integrated Services Digital Broadcasting-Terrestrial (Advanced ISDB-T) [4, 5]. With inheriting the features of ISDB-T such as hierarchical transmission and partial reception and so on, Advanced ISDB-T has improved the transmitting performance and functions. With utilizing the latest technologies, Advanced ISDB-T improved the capacity for 1.7 times larger than ISDB-T. The main key for improving the capacity was the adoption of low-density parity-check (LDPC) codes [6] and non-uniform constellations (NUC) [7]. Availability of larger FFT size such as 16k and 32k, or higher carrier modulation such as 256QAM, 1024QAM, 4096QAM and, the expanded signal bandwidth from 5.57 MHz to 5.83 MHz have also contributed to the increasement of the capacity.

The research and development for the next generation digital terrestrial broadcasting was not only limited to the physical layer. We have also conducted the research and development of transport layer and video/audio coding. The internet protocol (IP) based transport layer was intended to realize high level harmonization between broadcast and broadband. The system enables provision of integrated broadcast and broadband services, such as multi-view video, content replacement and augmented reality (AR)/virtual reality (VR) in TV programs. To verify the integrated broadcast and broadband services, we have developed an all-IP software-based integrated master control system that outputs signals to transmission stations and broadband networks. For the video coding, the system adopted Versatile Video Coding (VVC) which is the latest video coding standard that enables high efficiency and multiple functions. For the audio cording the system utilized Moving Picture Expert Group (MPEG)-H 3D Audio (3DA).

III. Total Verification Tests

From 2019 to 2022, The experimental transmitter stations are constructed in 4 large cities in Japan and various transmitting experiments were conducted. To verify the system in total, we have conducted the end-to-end verification test with using actual equipment including the experimental transmitter stations in 2022. Fig.1 shows the block diagram of the verification test. Fig.2 shows the equipment of the receiving site. With using the hierarchical transmission, two UHDTV services for fixed reception and two HDTV services for mobile reception within 6-MHz bandwidth of UHF band was demonstrated. The video and audio content were encoded/decoded with VVC and MPEG-H 3DA real-time encoder/decoder. The integrated broadcast and broadband services were also verified with using a signal via a broadcast and broadband network.

IV. Conclusion

The end-to-end verification test using actual equipment of the Advanced ISDB-T was conducted in for large cities in Japan. The test was successfully done which shows the feasibility of the Advanced ISDB-T system in total.

Acknowledgment

Part of this research was conducted under a contract from the Association for Promotion of Advanced Broadcasting Services (A-PAB) as part of its project commissioned by the Technical Examination Services Concerning Frequency Crowding of the Ministry of Internal Affairs and Communications, titled “Survey and Studies on Technical Measures for Effective Use of Broadcasting Frequencies (Survey and Studies for New Broadcasting Services).”

set.org.br/ijbe doi: 10.18580/setijbe.2023.8 Web Link: https://dx.doi.org/10.18580/setijbe.2023.8

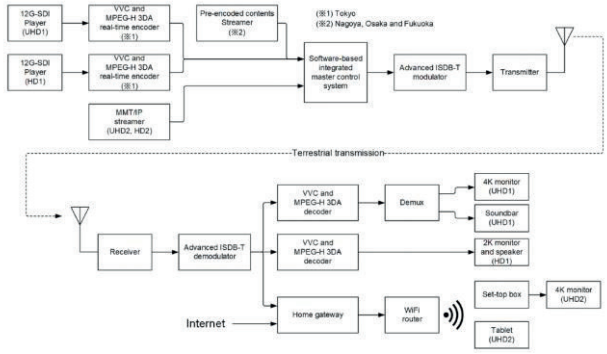


Fig. 1. Block diagram of the total verification test

Fig. 2. Equipment of the receiver site

REFERENCES

[1] ARIB standard: "Transmission system for digital terrestrial television broadcasting," ARIB STD-B31 Version 2.2 (Mar. 2014)

[2] ARIB standard: "Transmission system for advanced wide band digital satellite broadcasting," ARIB STD-B44 Version 2.0 (Jul.2014)

[3] Recommendation ITU-R: "Transmission system for UHDTV satellite broadcasting," BO.2098-0 (Dec. 2016)

[4] M. Nakamura et al.: "A study on the transmission system for an advanced ISDB-T," in *Proc. 14th IEEE Int. Symp. on BMSB*,4A-2 (Jun. 2019)

[5] N. Shirai et al.: "Laboratory experiments and large-scale field trials for evaluating the advanced ISDB-T," in *Proc. 14th IEEE Int. Symp. on BMSB*, 4A-4 (Jun. 2019)

[6] S. Asakura et al.: "FPGA-based performance evaluation of FEC codes for an Advanced ISDB-T," *ITE Trans. MTA*, 9, 3, pp. 180-187 (Jul. 2020)

[7] S. Asakura et al.: "Transmission performance evaluation of an Advanced ISDB-T -Non-uniform constellation performance in an echo channel-," *ITE Tech*. Rep., 44, 16, pp. 21-24 (Jul. 2020), in Japanese

Kohei Kambara
received B.E. and M.E. degrees in electrical and computer engineering from Yokohama National University, Kanagawa, Japan, in 1999 and 2001. He joined NHK in 2001. Since 2019, he has been working for NHK STRL. He is a senior research engineer of the Advanced Transmission Systems Research Division and is engaged in the development of next-generation terrestrial broadcasting systems for UHDT

Received in 2023-06-07 I Approved in 2023-08-04

Globo's Ultimate Operational Challenge: a creative workflow editing in cloud

Priscila David
Ariza Bertelli

Cite this article:

David, Priscila; Bertelli, Ariza; 2023. Globo's Ultimate Operational Challenge: a creative workflow editing in cloud. SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING. ISSN Print: 2446-9246 ISSN Online: 2446-9432. doi: 10.18580/setijbe.2023.9. Web Link: https://dx.doi.org/10.18580/setijbe.2023.9

SET INTERNATIONAL JOURNAL OF BROADCAST ENGINEERING - SET IJBE V.9, 2023, Short Paper 9, 5 p
2023 SET - Brazilian Society of Television Engineering / ISSN (Print): 2446-9246 / ISSN (Online): 2446-9432

Globo's Ultimate Operational Challenge: a creative workflow editing in cloud

Priscila David, Product Owner Specialist, Ariza Bertelli, Media Solutions Analyst

***Abstract*—In 2022, the famous and epic soap opera "Chocolate com Pimenta" in Brazil became a good surprise for TV Globo: the post-production chain accomplished a** simple, **productive, and economical workflow. A cloud based, remote and collaborative editing produced the entertainment content in an innovative way. Globo, a free-to-air television network, saw in this path an excellent opportunity to thrive technologically and to offer spectators a unique experience. Globo's objective was to provide a special edition of the soap opera through Globo Play, an online video on demand platform, and by Open TV. Having the team in charge located at Post-Production Center of Estúdios Globo, the material was edited collaboratively in HD (XDCAM codec) directly connected to cloud. The process was successfully achieved and has helped to maintain Globo into the future of technological innovations. And besides, the business model of this unique approach was very attractive for Globo, solving the market need for a cloud-native solution, notably when the Post-Production Center deals with big files and assets routinely, enabling the increase of the content production to a next level (cost benefit). Globo interprets this initiative not as a mere technological advance, but also as a step taken toward a more concrete operational level. Through all the effortful work, Globo, its customers, and everybody involved in this project could come together to cheer and look forward for a brand-new efficient future.**

***Index Terms*—Post-Production, Globo, Globo Play, cloud, innovations, collaboratively, workflow, technological advance, efficient, future.**

I. Introduction

Cloud architectures are increasingly gaining space and notoriety in Media Tech companies due to significant advantages, such as collaboration, flexibility, and efficiency. Here are some cloud differentials:

A. Accessible

Through connectivity with the large internet network, cloud applications can be accessed anywhere and at any time, essentially enabling remote work.

B. Collaborative

Cloud native applications enable collaboration between users worldwide. Different workgroups can access and edit files simultaneously, enabling agile and efficient results.

C. Scalable

The cloud infrastructure is very broad, compared to specific servers for content storage. It is robust and stable to handle a significant amount of data traffic and is highly consistent.

D. Resilient

Cloud providers have automatic storage backup and disaster recovery routines, which greatly mitigate the risk of data loss in the cloud environment. In addition, cloud architectures benefit from periodic and continuous infrastructure and security robustness updates.

E. Efficient

The cloud context brings a new business model, where you only pay for the resources used, the environment works 24/7 without the need for a dedicated operational team.

Cloud architecture is evolving significantly around the world given the strong enterprise adoption, greatly driven by the growth of cloud service providers such as Amazon Web Services, Microsoft Azure, Google Cloud Platform, and IBM. More and more new datacenters are created fulfilling technical requirements such as low latency and high availability. Besides, Digital Transformation is an essential factor in cloud adherence, enabling agility, innovation, and efficiency since the constantly changing market demands. As more companies and users identify the benefits of cloud architecture, its growth tends to increase substantially in the coming years.

TV Globo is one of the largest television stations in Brazil and Latin America, in addition to being a Media Tech company, which combines elements of technology and media to provide innovative solutions and services related to the media industry. The company offers a distinguished programming, which includes soap operas, series, entertainment programs, journalism, sports, and movies. The broadcaster is also recognized for producing and exporting Brazilian soap operas to several countries.

Globo has invested in technological innovation over the last few years to improve its production, transmission, and interaction with the public. Experimenting with cloud architecture could not be different for Globo to further boost its innovation context.

II. Globo company

TV Globo has its content broadcasted in most of Brazilian territory through its five networked main stations together with allied companies. Its high-quality standard has contributed for Globo to establish itself as a leader content producer whose prior mission is to delight people around the

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.9 Web Link: https://dx.doi.org/10.18580/setijbe.2023.9

world with the best quality. Its path toward becoming a pioneer is due to its history being so linked to the history of Brazilian TV.

In 2015, Globo proudly released Globo Play that would embody the new range of opportunities aiming to break boundaries imposed by the fast pace of life brought by modern times. Globo understands that in the current context, watching TV in the traditional way might be challenging, which calls for the need for innovation, that is, the development of new ways of delivering content to spectators on varied platforms. Globo Play was born from the desire to provide the audience flexibility and mobility. Globo Play is an OTT and VOD digital platform, which, through GPS and IP locations services, offers video content to viewers on their smart TVs, mobile phones, tablets, and personal computers with access to internet. Additionally, Globo Play is compatible with the ultimate 4k and HDR technology.

III. Post-production pipeline

Post-Production is a super important step in the content production chain. It is the process where the raw material is edited to generate the finished material, also known as the desired final product, ready to be exhibited to the public.

In the post-production, processes such as video editing, sound and audio mixing, color grading, visual effects and artistic and technical review of the content produced are carried out.

This phase builds the content storytelling to be delivered to spectators and represents a key role while ensuring its technical quality, correcting possible errors that may have occurred during the recording.

In the technical context, post-production also takes care of media management along the content production path, such as video, audio, and image files, to facilitate access and efficient location during the editing process.

These are the main sub-steps of the Post-Production:

A. Ingest

It all starts with importing the media, captured during the recording, into the editing system. The Ingest stage is the entry point for content in Post-Production.

B. Content Storage

As important as the Ingest of the material, is the place where it will be saved for the editing step. The cloud architecture takes part in a fundamental role in this step, as the content is a very important material. High availability and resiliency storage are fundamentals in the post-production process. With the advancement of Cloud Journey, it was understood that cloud storage is an interesting model and that it makes a lot of sense as mentioned above.

C. Scene Editing

It is the step where the story line is originated. The editors select the best parts of the raw material, the most relevant clips and arranging the chronology of the facts.

D. Audio Editing

As significant as the video, is the audio treatment. During the Post-Production, a special attention is given to sound quality, such as: audio level, listening comfort, mixing dialogue with soundtrack or ambient noise.

E. Visual Effects

This is the part of the chain where visual arts or animated graphics are inserted, involving the creation and addition of these elements to the video, including multi-layer compositing, motion tacking, 3D modeling, animation, etc.

F. Finalization

It is one of the last processes of the Post-Production, responsible for joining/finalizing all the previous steps mentioned before. After creating the initial sequence, the editors refine the editing, always improving the narrative and the audio and video set.

G. Quality Control

In the last step of the Post-Production process, the content will be in the appropriate format for Distribution and Exhibition (Open TV, Pay TV, Streaming) and thoroughly reviewed. The story will be thrilling millions of people with an excellent audiovisual technical quality.

It goes without saying that all the previous mentioned steps can overlap and be repeated a few times throughout the Post-Production process as editors refine audio-visual content, meeting creative and technical requirements.

IV. Problematic

As we know, editing content in the cloud has gained a lot of popularity around the world, given the availability of taking advantage of shared resources in a scalable structure, capacity to store and process large volumes of data, which eliminates the need for local storage. Globo was looking for a new simple, efficient and innovative Post-Production workflow, where it would not have to burden traditional on-premises storages, and at the same time, would not impact the operational experience of editors, especially in terms of speed. The new solution needed to maintain the robustness, resiliency, low latency, and security of the usual on-premises structure.

In addition, the new workflow needed to have an interesting business model for Globo, where the budget consumed by each share of this storage location could be financially passed on to the respective Production that contracted the service (“pay as you go”). In other words, Globo was looking for an approach of technical and business innovation.

In this way, it was understood that the cloud architecture associated with the Storage as a Service model would be an interesting case to be explored and experimented by Globo to

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.9 Web Link: https://dx.doi.org/10.18580/setijbe.2023.9

confirm the technical, operational, and financial viability to be practiced.

V. WORKFLOW PROPOSAL FOR CASE "CHOCOLATE COM PIMENTA"

"Chocolate com Pimenta" is an epic soap opera in Brazil, it has been shown a few times by Globo and is always a record audience. In this way, this would be the key product to take advantage of this new workflow.

Figure 1: Globo's Soap Opera

In a search of a solution, where it was possible to send files to the cloud in a high speed and the possibility of obtaining them with low latency to a local infrastructure, whenever required, in addition to the requested synergy with the editing software, we explored LucidLink application and created the following workflow:

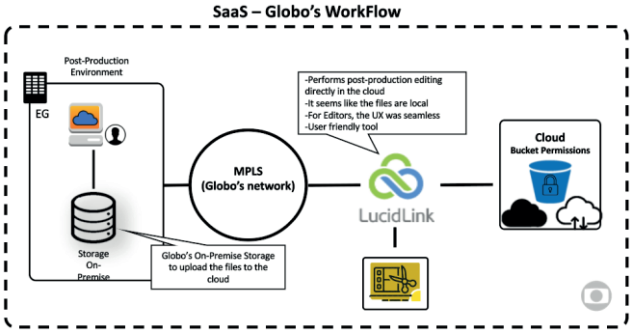


Figure 2: SaaS - Globo's Workflow

The content is sent to the cloud in an innovative way: the file is divided into small packs of approximately 256K, and these are sent in parallel, making everything faster (of course, the network directly influences this upload).

For example, a 1GB file would be divided into approximately 4 files of 256K, and these sent in parallel, would arrive in the cloud faster, as shown in the illustrations below:

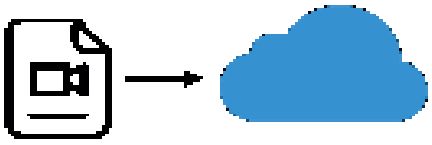


Figure 3: Normal Sent

Figure 4: Pack Sent

The same process occurs when downloading the file, which makes this operation much faster than usual.

Furthermore, the software is mounted on a computer as a local drive, which is very close to the on-premises environment, and can be used with Linux, MAC and Windows, facilitating the entire editing workflow.

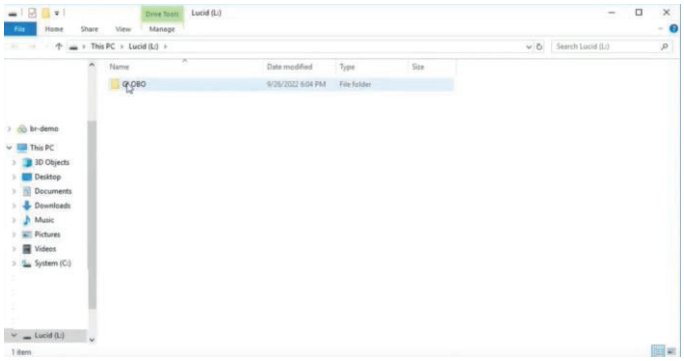


Figure 5: LucidLink drive in Windows

The software has several ways to be accessed, and an interface capable of mapping the number of files that are present in the cloud bucket, in addition to track the remaining uploaded files that are still being sent to the cloud.

As soon as the file is uploaded to the cloud, it is already possible to verify it from another machine and a different user.

It also provides permissions for folders within the drive, where the administrator can change the access for each user, with specific privileges.

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.9 Web Link: https://dx.doi.org/10.18580/setijbe.2023.9

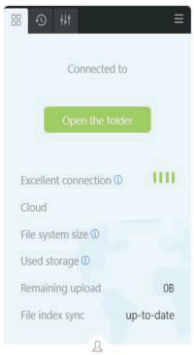


Figure 6: LucidLink Interface

Another great differential of this workflow includes its business model, which has compatibility with different cloud providers, where you pay for the storage used at the peak of the month. The cloud was also part of our study with LucidLink, as it allows us to send and receive files at high rates.

Other differential is the functionality of the local cache that enables operational fluidity in terms of editing.

The tests of this workflow were made for some chapters of the special edition of "Chocolate com Pimenta" in XDCAM 50 and edited in a professional edition software (installed in a local machine). During the test, users did not see any differences between the new workflow and the on-premises structure, which was extremely similar to what was done previously with on premise storage.

Figure 7: Special Edition

VI. Conclusion

We illustrate the conclusion of this article with a comment of one of Globo's editor: "This here is the ease that the editor needs to work [...] it works perfectly, it is looking like the local drive". The implementation of the workflow was extremely successful and the feedback from the operational team from Post-Production was very positive. Users are getting a smooth and efficient experience, as if they were working in a local workflow.

The architecture also allowed the material in the cloud to be collaborative between users, regardless of their geographic location, with high availability storage and the reliability, security, and integrity of the files.

This new workflow optimized internal processes, generating cost savings, and boosting the efficiency and quality of audiovisual production. All this reinforces the importance of the cloud for the industry and the journey toward an increasingly collaborative and technologically advanced future.

Acknowledgment

The authors would like to express our sincere gratitude to the people below for their invaluable guidance and support during the path of this research. Their experience and encouragement were fundamental for the elaboration of this study:

Adonias Nobrega de Melo Junior
Erik Haier Maia
Heloisa Dina Felix Lima Bezerra
Isabelle Ferreira Pesset
José Almeida Junior
Kaique Rodrigues de Amorim
Marcelo Vale Fontana
Marcos Luiz Morais Santos
Rebeca Pereira Borges de Souza
Roberto Tavares de Menezes Filho
Thiago de Carvalho Barreiros

Thank you all for your contributions, making this research a reality.

References

[1] Lucid Link Work Flow https://www.lucidlink.com/workflow. Accessed: 2023-07-19

[2] Globo Group Company https://grupoglobo.globo.com/. Accessed: 2023-07-19

Authors

Priscila David was born in Rio de Janeiro, Brazil in 1989. She graduated as a Telecommunications Engineer Bachelor, holds an MBA in Strategic People Management. She was the author of the Poster "4K and 4K-HDR VOD in Rio's 2016 Olympic Games" published by IBC in 2017.
She has been working at Globo for 17 years. In the last two years, she has been the Product Owner of Post-Production projects in the Media Solutions area at Globo.

Ariza Bertelli was born in Minas Gerais, Brazil in 2000. She is majoring in Electrical Engineering with an emphasis on Robotics and Industrial Automation. She was a member of IEEE from 2019 to 2021. She received the award for the 3 best education project at RNR: "Talking with the hands" in 2020. She has been working at Globo for 1 year as

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.9 Web Link: https://dx.doi.org/10.18580/setijbe.2023.9

a Media Solutions intern, and in June 2023, as a Media Solution Analyst.

Received in 2023-06-06 I Approved in 2023-08-04

https://set.org.br/ijbe doi: 10.18580/setijbe.2023.9 Web Link: https://dx.doi.org/10.18580/setijbe.2023.9