Towards a representation model and fog-based device orchestration for audio-centric pervasive storytelling

Marcelo F. Moreno
International Audio Laboratories
Friedrich-Alexander Universität (FAU)
Erlangen, Germany
marcelo.moreno@audiolabs-erlangen.de

Nils Peters
International Audio Laboratories
Friedrich-Alexander Universität (FAU)
Erlangen, Germany
nils.peters@audiolabs-erlangen.de

Pedro H. V. de Almeida
Computer Science Department
Federal University of Juiz de Fora (UFJF)
Juiz de Fora, Brazil
phventurarodrigues@gmail.com

Carlos Pernisa Jr.
Faculty of Communication
Federal University of Juiz de Fora (UFJF)
Juiz de Fora, Brazil
carlos.pernisa@ufjf.br

***Abstract*—Internet of Things (IoT) devices, such as smart speakers and wearables, are increasingly accessible and part of people's daily lives. This opens up great new possibilities for innovative storytelling experiences, allowing new forms of interactive and truly immersive content consumption, going beyond conventional multimedia. In this context, the need for advances in the representation of pervasive storytelling is perceptible and an audio-centric approach utilizing the Internet of Sounds (IoS) can potentially better fit into people's routines because of the widespread audio capabilities in IoT devices. This work proposes a conceptual model entitled A-Presto (Audio-centric PeRvasivE STOrytelling) that aims to realize stories in a pervasive way, adapting to the users' context and available IoT devices. By modeling the specific domain of audio-centric pervasive storytelling at a high abstraction level, the proposal transparently supports the typical variability of pervasive environments, such as changes in users' location, device connectivity, power availability, and proximity between users, among others. Supported by latency experiments using a cloud-based orchestrator prototype and local IoT devices, this work proposes a fog-based runtime engine, capable of interpreting and orchestrating A-Presto storytelling instances with reduced latency.**

***Index Terms*—IoT, pervasive storytelling, audio-centric, immersion, fog-based device orchestration**

I. Introduction

The increasing affordability and widespread adoption of Internet of Things (IoT) devices have made them an inseparable part of people's daily lives. Devices like "Smart" watches, lamps, cell phones, TVs, and infrared controls are now commonplace, surrounding individuals in their day-to-day routines. These devices are commonly controlled through cloud-based personal assistants, presenting exciting possibilities for innovative experiences.

These technological advances open the door to new forms of storytelling that surpass traditional multimedia and single audio/video displays. By integrating IoT devices with cloud support, we can explore novel ways to engage audiences and create interactive narratives. This integration of smart devices and cloud services has the potential to redefine how we experience and interact with multimedia content, offering fresh opportunities for captivating storytelling.

Audio-centric interactive storytelling has been explored through the utilization of particular IoT devices, like the various stories developed as Alexa *Skills* [1]. These interactive narratives offer oral narration with character dialogues, encouraging listeners to engage by using speech. The Internet of Sounds paradigm [1] opens up new possibilities for enhanced content consumption in pervasive environments.

In this context, it becomes evident that there is a pressing need for advancements in representation and orchestration to enable pervasive storytelling. Pløhn [2] emphasizes that pervasive storytelling should seamlessly blend into the real world where the user resides, creating an immersive narrative that surrounds them. This immersive experience requires technologies, as described by Murray [3], to become transparent to the player/spectator.

Currently, the available representation model and authoring solutions that offer higher-level functionalities and personalization for pervasive storytelling have their limitations. They are often constrained by specific playback devices, restricted to user devices for presentation, and focused on domain-specific hyper/multimedia rather than addressing the representation and variabilities required for pervasive storytelling.

Indeed, considering the mobility of users and the diverse possibility of IoT devices that can be brought together both

The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS.

[1] Alexa Interactive Stories: www.amazon.com/s?rh=n%3A19353824011

individually and in dispersed environments by storytelling providers, it becomes crucial to ensure that the orchestration of the presentation occurs ubiquitously. Given the existing interconnectivity among devices, a cloud-based device orchestrator design emerges as a promising approach to achieve such ubiquity. However, it is essential to acknowledge and address the inherent challenges, including the uncertainties posed by the best-effort nature of the internet, wireless communication, mobility, energy constraints, and other related factors.

Hence, the primary objective of this study is to introduce a novel conceptual model that specifically focuses on the representation of audio-centric pervasive storytelling. This proposed model, named the "Audio-centric PeRvasivE STOrytelling" (A-Presto) model, emerged from a careful examination of the creative process in focus groups composed of researchers in the field of social communication. To ensure the validity of the structures and terminology adopted, specialists in narrative domains actively participated in the validation process.

The A-Presto model is designed to accommodate the distinctive variabilities found in pervasive environments, such as changes in location, device availability, proximity between users, and other relevant factors. Furthermore, this research introduces a fog-based presentation engine, acting as an orchestrator, responsible for interpreting the storytelling representations in A-Presto and seamlessly coordinating the presentation of storytelling components. The orchestrator operates independently of specific devices, efficiently managing them based on their availability and the authoring settings of the ongoing instance.

To validate the proposed approach, experiments were conducted with a prototype of the developed orchestrator. The results demonstrate its feasibility and underscore the importance of fog computing in enhancing the effectiveness of A-Presto.

This paper is structured as follows: Section III provides some usage scenario examples to enhance the understanding of pervasive storytelling. Section II explores related work in the field. Section IV delves into the specifics of the proposed A-Presto model. In Section V, we present the fog-based orchestrator, along with the experimental results. Finally, Section VI offers conclusions drawn from this study and outlines areas for further development of this concept.

II. Related work

The increasing use of mobile, personal, and IoT devices has facilitated the exploration of applications focused on pervasive storytelling. Augmented reality applications for pervasive narratives and games have been noteworthy in this regard. Although these initiatives do not share a common representation model for pervasive storytelling, they provide valuable insights and help identify essential requirements for the envisioned scenario.

Bichard et al. [4] present a game that utilizes geolocation sensors (GPS), gyroscopes, cell phones, speech synthesizers, and a directional microphone. Their proposed framework establishes connections between these devices and a server, enabling the exchange of input, output, and processing data. By integrating with a Geographic Information System, the framework transforms real local information, such as traffic lights or landscapes, into virtual objects for decision-making within the programmed narrative. Players participating from the back seat of a moving car interact with virtual characters and objects using voice commands, and these virtual elements appear based on the players' location and position. The logic and coordination of the effects that constitute the gaming experience are governed by the execution of scripts developed by programmers.

The research highlighted in [5] utilize geolocation and network data mining techniques for pervasive storytelling. In 2016, the popularity of geolocation in games surged with the release of "Pokémon GO" [6] for mobile devices. Similar to the approach in [4] the players' geolocation is tracked, and virtual objects are generated based on real-world information from nearby locations, enabling interaction with these objects. Notably, this marked the rise of cloud computing adoption for executing at least a portion of the narrative logic.

The authors of [7] discuss an architecture capable of analyzing user context data and obtaining real-time information about elements in the vicinity. Moreover, the architecture incorporates data mining from connected social networks to generate a context-adapted narrative personalized to the individual's interests. This model showcases the potential of cloud applications to craft unique experiences using publicly available web and social network data, harnessed with the user's permission. As this information continually evolves with user interactions and their respective networks, it facilitates a constant adaptation of the narrative, enhancing its relevance and personalization.

In terms of speech-centric experiences, the development kit for the Alexa voice assistant [8] enables the creation of *Skills*, but it requires programming knowledge. However, Amazon also provides services like *Alexa Skills Blueprints* [9], which offer pre-designed templates for specific application types, allowing non-programmers to customize them. For storytelling, these Blueprints have features that enable simple text insertion for Text-To-Speech, native sound effects, and reactions.

The BBC R&D developed a web-based device orchestration framework to synchronize, stream, and render object-based audio content to connected smartphones, tablets, and laptops [10]. To account e.g., for the device types and their locations, the rendering of the audio assets depends on customizable rulesets. The cloud-sync corresponds to the HbbTV2.0 timeline model [11] and across participating devices, time alignments between 10-200 ms can be expected. BBC R&D also studied production aspects [12], [13] and listeners' reception [14] of non-interactive, device-orchestrated content. The latter suggests that device-orchestrated content can feel very immersive and that the overall quality can depend on the loudspeaker quality and the listening position relative to the devices.

Other efforts extend pervasive storytelling beyond computational support. For instance, [2] delves into understanding and measuring the optimal utilization of elements in pervasive storytelling to enhance user engagement and immersion. Many of these questions were initially addressed by Murray [3] even before the proliferation of devices capable of extending pervasiveness beyond controlled environments.

Currently, there is a noticeable absence of high-level models capable of effectively representing pervasive storytelling and facilitating its orchestration during presentation. However, it is encouraging to observe that existing models and languages within the domain of hyper/multimedia are being adapted and expanded to address some aspects of this specific challenge.

The Nested Context Model (NCM) version 3.0 is a conceptual model primarily designed for hypermedia [15]. It aims to represent structural concepts, events, and relationships between media, while also defining rules and operations for manipulation, synchronization, reproduction, and updating of structures and documents. In the context of pervasive storytelling, NCM does offer some elements that partially address narrative construction, including event-based executions, generic nodes, content variations, and data storage. However, because pervasive storytelling is outside of the initial scope of hyper/multimedia applications, NCM requires specific modifications, such as incorporating sensory effects, enabling storage and access to narrative history, ensuring primary device independence, and supporting cloud orchestration, among other requirements.

Indeed, NCM has been the subject of recent investigations, leading to the development of a new version of the model and its associated NCL language. For instance, Barreto et al. [16] propose multimodal interactions tailored for Interactive Digital TV, introducing support for multimodal and multi-user interaction events. Their work also allows for the identification of the specific user interacting with the Interactive Digital TV application.

To incorporate sensory effects and facilitate the authoring of multimedia applications with NCL, Rodrigues et al. [17] suggest extending the language based on the information representation architecture from the MPEG-V standard. They propose introducing new types for the "type" attribute of the <media> element to represent sensory effects without the need for additional XML tags in NCL. Additionally, to accommodate three-dimensional representations of sensory effects, they extend the attributes of the <region> element.

The authors of [18] focus on resolving synchronization-related challenges during the presentation of NCL-based multimedia applications. They propose an automatic staging operation for media objects and sensory effects, which adapts based on the type of effect, network conditions, and temporal behavior of the application. Additionally, the explicit control of preparation by the NCL document author was proposed in [19].

Other languages and models primarily targeting hyper/multimedia applications e.g., HTML5 [20], SMIL [21], or STOrM [22] also encounter difficulties in supporting pervasive storytelling. While those may undergo extensions, the need for models specifically tailored to these applications is evident, as currently, existing models are lacking.

III. Examples of Usage Scenarios

The tradition of storytelling through devices can be traced back to radio soap operas. Throughout history, narratives have revolved around a central device. As digital games evolved, the audience transformed from passive spectators into active participants, engaging with virtual environments and influencing the direction of the consumed narrative.

As media consumption technologies have progressed, the reliance on a primary device for interactive multimedia presentations has remained a common trend. However, a notable shift has occurred, with various devices now capable of connecting to the Internet, communicating with each other, and exchanging information. These devices are no longer mere extensions of a primary device; they have transformed into independent elements with the ability to access and generate information in diverse spaces, unconstrained by a single room or controlled environment.

As a result, storytelling can now seamlessly integrate with the surroundings of the "interactor" [2], the individual both consuming and participating in the narrative experience. Pervasive storytelling can unfold without the interactor realizing how the narrative is controlling devices or gathering information. The story can adapt not only to the interactor's decisions but also to the contextual elements present in their surroundings. This marks a significant departure from conventional linear storytelling, where all available devices serve as tools, and no single local device orchestrates the storytelling.

To enable this pervasive storytelling experience, a ubiquitous orchestrator is essential, capable of seamlessly accessing any device across various contexts.

The orchestrator, operating from an undisclosed location, identifies the devices present around the interactor and utilizes them as the narrative unfolds. It seamlessly integrates the experiences envisioned and specified by the author, while also adjusting to changes in device availability and contextual variations. As a result, the interactor can seamlessly transition between environments or switch devices to consume the story, and the author can create conditional narratives tailored to the diverse contexts of the interactor's daily life.

Such variability in device availability requires a storytelling specification approach that accounts for and adapts to the potential circumstances of the interactor's context. This ensures that the author-specified experiences align with what is feasible and meaningful within the expected progression of the story.

A. Example 1: Pervasive Storytelling in a Home Environment

To illustrate the usage scenario, a fragment of a storytelling example is presented in Figure 1. In this scenario, the interactor "Anna" is at her home (1) and decides to commence a storytelling session using the voice assistant of her smartphone (2). The ubiquitous orchestrator recognizes the smart devices in Anna's home environment, including her smartphone, smart television, a networked loudspeaker

[2]This paper builds on the concept of "interactor" [3] to name the consumer, at the same time participant, of the storytelling experience.

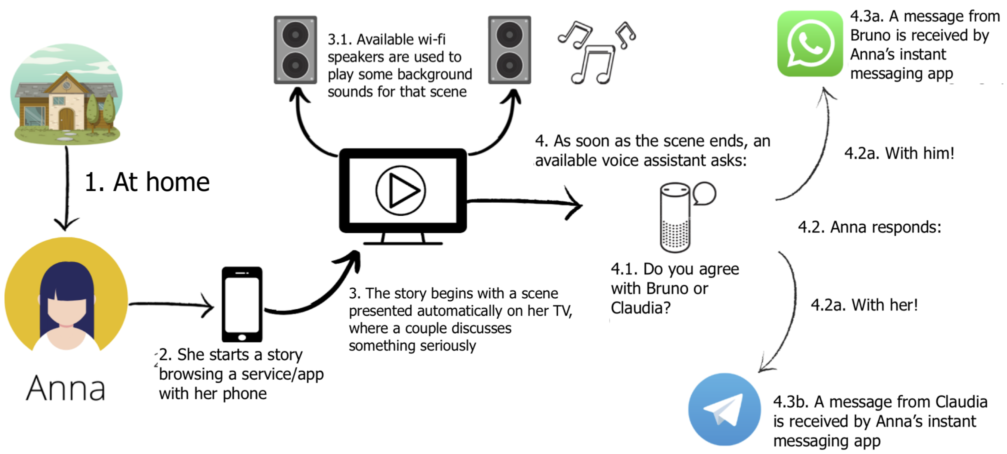

Fig. 1: Example 1: Start of pervasive storytelling in a home environment

system, and a smart speaker, all connected to the Internet. Leveraging the available devices, the storytelling adapts and initiates the presentation of a video on the television (3), using the networked speaker system to create an immersive ambiance (3.1). Following the video presentation, the domestic voice assistant on her smart speaker prompts Anna to respond verbally about the video content (4.1). The subsequent course of action will be determined based on Anna's answer (4.2), which will influence the unfolding narrative e.g., via a voice message to one of her smartphone's messaging apps which was generated by a Text-to-Speech service and ordered by the ubiquitous orchestrator. Noteworthy, Anna could have chosen to begin the story in an entirely different context, such as public transportation, allowing the storytelling to be adapted for a "mobile" experience.

B. Example 2: Pervasive Storytelling in Public

Interactions can directly involve real-world elements, adapting to the interactor's daily routine. In the scenario depicted in Figure 2, Anna decides to visit a coffee shop, and coincidentally, the orchestrator sends an alert through an application on her smartphone, inviting Anna to join her ongoing storytelling experience. The alert notifies her about an interactive kiosk at the coffee shop that plays a role in the story.

Upon reaching the kiosk, voice recognition technology identifies Anna and accesses her storytelling history. The kiosk then displays information related to the favorite drinks and foods of the characters with whom Anna has interacted the most during her storytelling journey. Subsequently, the kiosk allows her to select some of these products for consumption. The choices she makes during this interaction are stored as Anna's decision information to unlock new fragments of the storytelling experience she will encounter in the future.

IV. A-Presto Conceptual Model

The A-Presto conceptual model, introduced in this research, offers high-level abstractions that closely align with the knowledge domain of content creators. Its development involved collaborative studies with researchers from the social communication field. The entire structure and nomenclature of the model were derived from focus groups, during which these researchers engaged in open discussions about storytelling creation. By capturing their insights, terminologies, and recurrent semantics used in their creative process, the model was designed to enable easy assimilation of its entities by content creators.

The model considers that pervasive storytelling inherently relies on the occurrence of asynchronous events, given the open world in which the storytelling unfolds and the diverse range of variables that define it. Therefore, A-Presto follows an event-driven paradigm. These relationships are fulfilled by passive or actively triggered state changes in any integral part of a model instance, including conventional media, interactors, narrative progression, and environmental sensing.

With the support of a ubiquitous orchestrator, A-Presto instances do not specify a default playback device. Instead, they utilize devices linked to an interactor's profile, which could have been introduced by the interactor, the author, or the content provider. This approach ensures flexibility in device selection and enables a seamless storytelling experience, adapting to the available devices in real time during the interaction with the narrative.

A-Presto establishes a collection of structural entities that facilitate the specification of diverse routes (paths) that a storytelling can traverse. The model adopts a tree hierarchy, with each level introducing cascading circumstances essential for activating and playing its respective objects.

The model comprises four distinct structural entities designed to provide organization and logic for presenting media, sensory effects, and inputs throughout the narrative path: *Pervasive Storytelling*, *Episode*, *Experience*, and *Sequence*. Each instance of these entities is governed by its unique set of circumstances, defined through the *Circumstances* entity, discussed in Subsection IV-E. An overview of the model's hierarchical structure is illustrated in Figure 3, depicting the interrelationships among the structural entities.

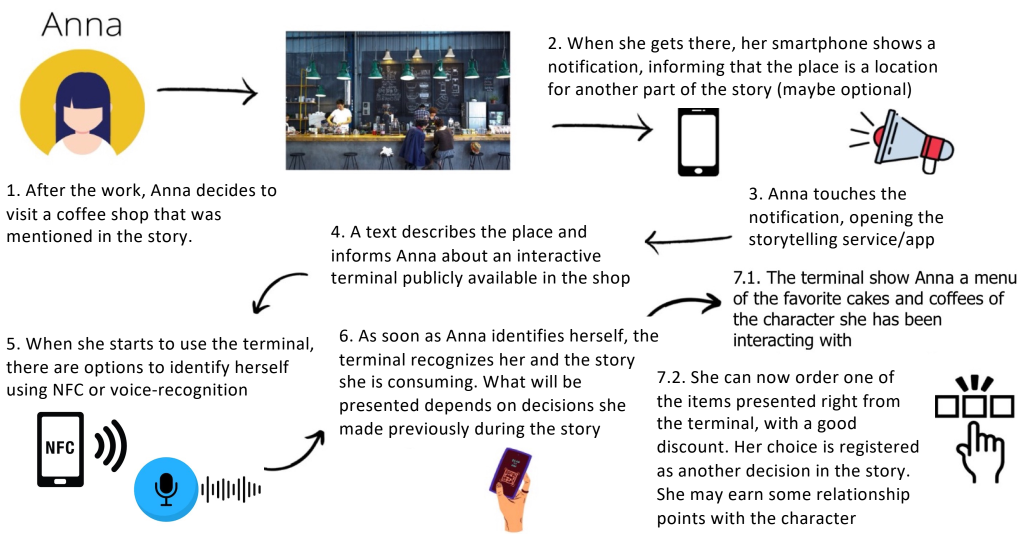

Fig. 2: Example 2: Continuation of pervasive storytelling in public

A. *Pervasive Storytelling entity*

The *Pervasive Storytelling* structural entity represents the entirety of a storytelling composition. It serves as the foundational root, housing metadata pertinent to the story, including its title, genre, and more. This entity includes properties for story identification, such as id, name, class, and other attributes, enabling referencing, modification, or access by third parties. Additionally, it may encompass crucial metadata for use by the orchestrator, content search tools, or for introducing the story before its actual telling commences.

An instance of *Pervasive Storytelling* is composed of one or more instances of *Episode*. As multiple stories can interrelate, the narrative path followed by an interactor in a *Pervasive Storytelling* is recorded in a history accessible by other stories that the same interactor or their network of interactors might consume. (This access is, of course, subject to the necessary permissions e.g., for privacy preservation.) As a result, references to other integral parts of the narrative path, namely *Episode, Experience,* and *Sequence*, are also included in the history.

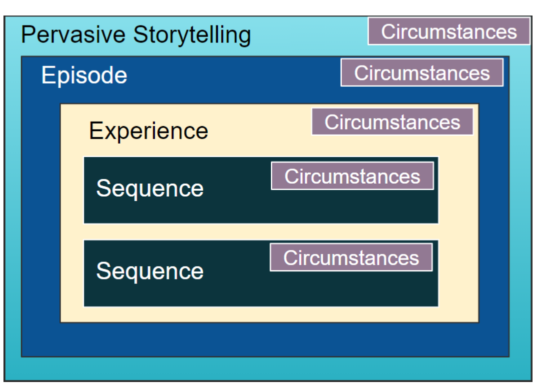

Fig. 3: Overview of A-Presto structural entities (nested layout)
.

Being the root entity of the model, *Pervasive Storytelling* enables defining conditions (*Circumstances*) for initiating storytelling for an interested interactor. This allows for the application of restrictions, such as usage rights based on location, age ratings, or even requiring the interactor to have previously consumed another specific story, among other possible criteria.

B. *Episode entity*

The *Episode* entity is a direct child of *Pervasive Storytelling*. An *Episode* carries the semantics of a "narrative unit" similar to episodes in a television series, where each episode represents a self-contained portion of the entire storytelling. It includes content description metadata to ensure accurate identification by the interactor.

Within an *Episode*, the interactor is exposed to instances of *Experience*. These experiences shape the interactive encounters and content presented to the interactor throughout the episode. Similar to *Pervasive Storytelling*, *Episode* allows for individual settings of security, legal, parental, and availability controls for each specific episode.

Regarding availability control, *Episodes* can utilize circumstances to impose certain requirements, such as a previous minimum consumption progress or a desired interaction profile, before being accessible to the interactor. Consequently, episodes can be delivered according to the interactor's progress in previous episodes of the same or other stories. This mechanism offers a dynamic and progressive narrative experience tailored to each interactor's unique journey.

C. *Experience entity*

An *Episode* is composed of *Experiences*, represented as instances of the entity *Experience*. In an analogous manner, experiences can be likened to scenes in an audiovisual production. However, they differ in that experiences are dependent on asynchronous circumstances and do not require predefined durations. Each *Experience* comprises instances of the entity *Sequence*, which run in parallel with one another.

The author of the pervasive storytelling utilizes experiences to outline the steps in the interactor's journey within an episode, some of which may be optional. The interactor's progress through experiences recorded in their history enables the establishment of new links with other experiences, either within the current storytelling or in subsequent ones. This dynamic enables the creation of diverse experiences in the same episode, based on the interactor's actions, previous decisions, and accumulated history. Part of the idealization of the experience is inspired by the concept of kinesthetic and emotional body experimentation [23]. It grants the author the ability to craft unique and personalized experiences for the interactor, enhancing immersion and engagement.

The model's capability to capture, store, and make available data from the interactor (refer to Subsection IV-G) is instrumental in generating these dynamic experiences. It ensures that the author can adapt and shape the narrative path based on the interactor's interactions, leading to a more fluid and personalized storytelling experience.

D. Sequence entity

Each sequence represented as an instance of the *Sequence* entity, consists of media, sensory effects, and decision points presented one after the other. *Sequences* operate similarly to SMIL's *seq* composition [21], existing in parallel within the same experience. However, there is a crucial distinction: different synchronization behaviors can be defined concerning other sequences within the same *Experience*.

In an *Experience*, a group of sequences may share the same synchronous *timeline*, allowing for concurrent playback. Alternatively, similar to SMIL's *excl*, only one sequence from the group may be played at a given time, ensuring exclusive presentation. Moreover, it is also possible to specify that only the first sequence that has its circumstances satisfied will be played, promoting selective playback.

Implicitly, each sequence's duration is defined by the total duration of all its internal objects, provided that all objects have explicitly or implicitly defined durations. If any media, sensory effect, or decision point within a sequence lacks an implicit time (e.g., a voice recording to be listened to, a smoke effect, or an interactive question to be answered), its duration must be explicitly specified. Additionally, it is feasible to explicitly define a maximum duration for a sequence, limiting its time until closure.

Furthermore, for each sequence, it is possible to specify the preferred type(s) of devices for playback (*deviceType*). For instance, this feature enables defining that lighting effects should always occur on RGB lighting devices, ensuring compatibility and enhancing the intended experience across various devices.

E. Circumstances entity

Circumstances play a vital role in defining when actions are executed on structural entities. These circumstances constitute a logical set of rules created by the author, serving as the basis for controlling the reproduction of structural entities.

Various sources can provide information for determining circumstances, including logs with state transition events, current playback states, properties of the *Aura*, and interactor-specific details. Additionally, circumstances can check for the availability of devices and employ relational expressions and logical operators to compose logical sentences. When a circumstance and its corresponding action are defined for an element, the orchestrator assesses whether the logic described is true or false. Once confirmed as true, the associated action is triggered.

For instance, consider the following example: If a circumstance states "Age of registered interactor is greater than 18 years old" and is linked to the action of starting an experience, it would restrict the experience to be accessible only to adults. Adding another circumstance like "Voice assistant device is available" would narrow the accessibility further, allowing only adults who are close to their voice assistants to start the experience. Lastly, incorporating the circumstance "Interactor pathway in Aura includes full experience X" would introduce a condition where only adults near their voice assistants, who have already undergone experience X, can initiate the specified experience (for the meaning of Aura see Section IV-G). The versatility of circumstances allows for precise control and customization of the storytelling experience based on various factors and conditions.

F. Media, Sensory Effect, and Decision Point entities

The model adopts a representation-centric approach, focusing on the objects to be displayed or obtained rather than the specific devices that perform the actions. The output entities include Conventional Media (*Media*) and Sensory Effects (*Sensory Effect*), while the input entity comprises Decision Points (*Decision Point*). These entities, along with the structural entities, are depicted in Figure 4.

Conventional media, well-established in the audiovisual domain, are straightforward to comprehend. Sensory effects, on the other hand, encompass stimuli that target other senses, such as taste, smell, touch, and unconventional audiovisual representations.

When reproducing an output entity, the orchestrator must always play it on the first available and compatible device. The orchestrator determines how to represent the defined effect or media, depending on the interactor's preferences, the author's specifications, and device availability. The author can specify preferred device categories for reproducing the outputs if desired.

Decision points facilitate the correlation between data acquisition from interaction or sensing and the subsequent actions triggered upon data capture. Each decision point contains a set of Options (*Options*), where each option incorporates logical expressions that define the necessary data to be obtained, expected values for the option to be true, and the action to be triggered if the criteria are met. The decision point can use interactor-specific information and designate from

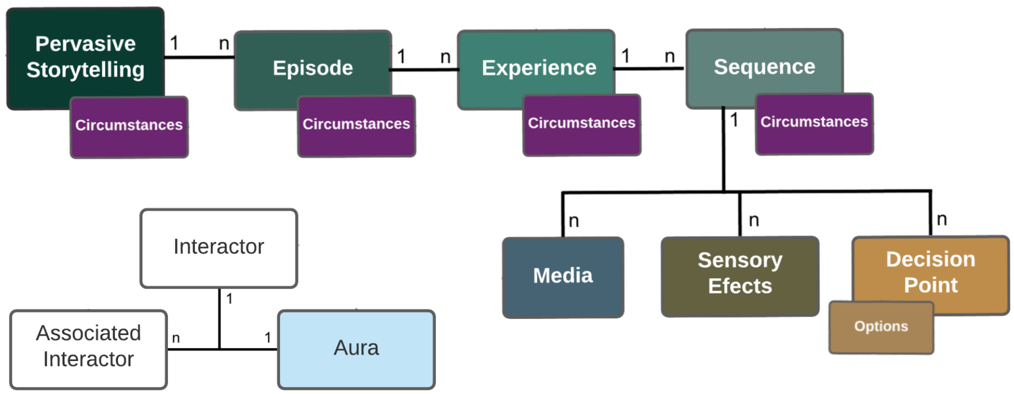

Fig. 4: A-Presto main entities - associative diagram

which profile the interaction is anticipated. To handle scenarios where responses may be absent due to pervasive environment variabilities, a decision point should specify the course of action if no options are satisfied within a defined time limit.

It is important to note that this trio of input and output entities does not have circumstances and must always be included within a sequence. Each entity has specific attributes tailored to its type of representation. As this work does not introduce a language based on the model, it does not define the mandatory properties of each media, sensory effect, or decision point. The development of a language, preferably visual for this purpose, is left to future works.

G. Interactor and Aura entities

To store interactor information, including personal data and preferences that may influence the storytelling experience, the entity *Interactor* is defined. Each interactor is associated with an *Aura* entity, which accumulates the contextual knowledge of their respective pervasive storytelling instances in which they engage. As illustrated in Figure 4, the *Aura* operates similarly to a *Karma* system found in games, where the player's recorded actions and attitudes influence subsequent events and actions. Consequently, the interactor carries with them the consequences of their choices, impacting the narrative progression. The data stored in the *Aura* can be used for logical validations within *Circumstances* and *Decision Points*. The *Aura* retains the entire path taken by the interactor in a pervasive storytelling instance, along with the states of the structural entities' instances.

Interactor properties facilitate the registration of personal information, preferences, and associations with profiles, including any physical and/or sensory disabilities. This allows the option of routes that consider accessibility and inclusion, making the experience more tailored to individual needs. Furthermore, interactor properties can define values for parameterizing abstract and personalized sensations, such as the temperature corresponding to feelings of cold and heat for the interactor. These personalized data enable the author to adapt the sensory effects and conventional media experiences accordingly.

The *Associated Interactor* entity facilitates cross-referencing interactor data throughout the storytelling process. This allows multiple interactors to join in a social circle and engage in a shared pervasive storytelling experience. For example, during the consumption of a story by one interactor, if another friendly interactor initiates a communal story supporting multiple participants, they can engage in pervasive storytelling together. The data for each interactor is recorded individually in their respective Aura, while their information and devices are interconnected when necessary, enhancing the sense of shared immersion and interaction among participants.

H. Presentation state machine

The A-Presto model includes a presentation state machine associated with each of its structural entities. This state machine governs the transitions between different states, considering the influence of circumstances and the variability of pervasive environments. The presentation state machine extends the states commonly found in hypermedia models like NCM [15], introducing new states and transitions to accommodate the unique requirements of pervasive storytelling. Figure 5 illustrates the presentation state machine.

Initially, an instance of a structural entity is in the *Enabled* state. In this state, its circumstances are actively observed, and if the logical test of the entity's circumstance evaluates to true, it transitions to the *Occurring* state, and the presentation of the element starts.

The author has the flexibility to define other initial states, such as *Disabled*. In the *Disabled* state, the element's circumstances are not evaluated, and it remains inactive, preventing its reproduction until explicitly enabled, such as at a decision point.

Once an element is in the *Occurring* state, it remains active, and its circumstances are continuously evaluated. If the circumstances become false, the element and its descendants are paused, and the state transitions to *Locked*. The *Locked* state is solely managed by the orchestrator and is used to temporarily halt presentation when circumstances become invalid due to factors like device disconnection, mobility, or environmental changes. Once the circumstances become valid

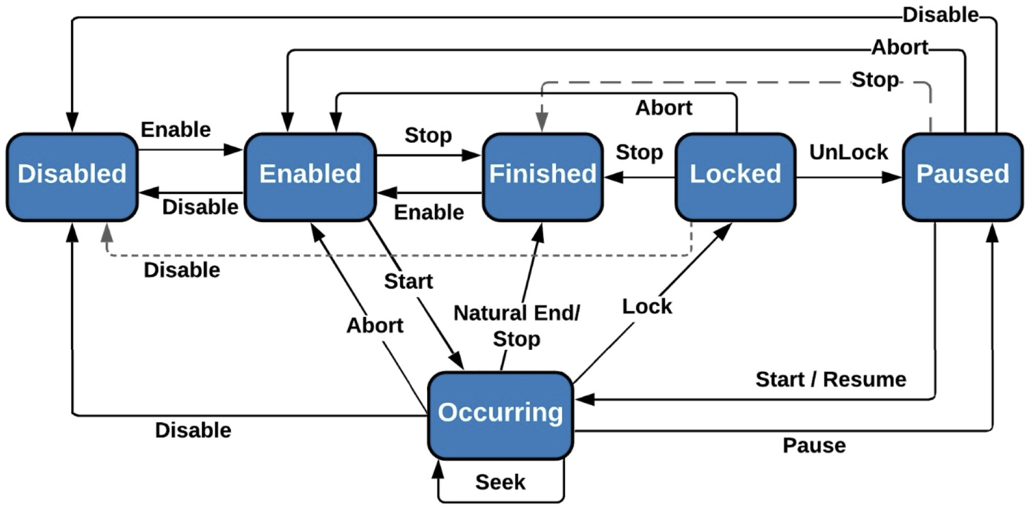

Fig. 5: A-Presto presentation state machine

again, the element transitions to the *Paused* state, allowing the interactor to resume presentation as desired while keeping its circumstances under constant evaluation until the element is closed.

The *Finished* state is reached when a structural entity has completed its presentation, i.e., all its children are disabled or have reached their natural end.

In the *Paused* state, the element and its descendants are temporarily paused, awaiting an external action, a decision point, or specific circumstances to resume presentation. The *Paused* state enables more interactive and dynamic storytelling experiences.

The introduction of the *Locked* state is unique to the pervasive storytelling context, where uncertainties in device connections and environmental variations demand a more refined control of the presentation. It provides a mechanism to handle potential disruptions in the presentation process due to the variability of pervasive environments.

The A-Presto presentation state machine ensures a flexible and adaptable storytelling experience, enabling smooth transitions between different states and accommodating the dynamic nature of pervasive storytelling.

V. A-Presto Fog-based device orchestration

The A-Presto model relies on an orchestrator responsible for controlling and executing all elements of an A-Presto instance in accordance with the structure defined in Section IV and the playback logic outlined in Section IV-H. The orchestrator can be controlled externally through HTTP-REST requests. To execute actions defined in the presentation state machine (Figure 5), such as *start*, *pause*, and *abort*, an API route was created for the *Presentation Engine* component, capable of processing these requests in the cloud, simulating the playback control tools that an interactor would have.

The API developed alongside the orchestrator also facilitates runtime changes to interactor data or storytelling instances, as well as forcing state changes for testing purposes. The orchestrator adheres to the defined transition rules, even when forced to perform specific actions, such as aborting an element in the *Enabled* state, demonstrating its robustness and adherence to the presentation state machine.

When the orchestrator is called by an interactor to start an episode, it retrieves the corresponding storytelling instance from the interactor's *Aura* layer and passes it to an initialization function. This function checks the circumstances of the storytelling instance to validate if the story can proceed. If the conditions are met, the selected *Pervasive Storytelling* and *Episode* are started and registered in a new *thread* of the engine, which continuously observes changes for future evaluation. The engine uses an observer pattern to monitor changes in the *Aura*, the presentation state machine, and the registered devices. Each *thread* created observes all these elements, and if any change occurs, it calls the evaluation function for the respective registered element.

The evaluation function checks the state of the parent element initially, as the parent's occurrence is required for child elements to occur. For example, an element in the *Enabled* state, when enabled by a decision point, checks if the parent is occurring and then evaluates its circumstances for occurrence.

While an element is occurring (*Occurring*), the engine continuously checks the state of the parent and its own circumstances to determine if it should remain occurring. If specific circumstances for pausing (*toPause*) or finishing (*toFinish*) are defined, they are also evaluated.

The engine also monitors the states of the children, to observe if they have completed, ensuring that their end is reached. When an element is disabled or has completed its presentation, it is removed and needs to be instantiated again if re-enabled.

The enablement process allows a decision point to request an element to be enabled, whether it is currently disabled (*Disabled*) or has reached its natural end (*Finished*), as described in Section IV-H.

A. Latency experiments with an A-Presto prototype

Experiments were conducted to measure the latency between request and execution using an A-Presto orchestrator prototype. four devices are used: two smart lights, a TV, and a smart plug.

The tests involve creating a storytelling instance with a sequence of ten sensory effect elements, with each effect executed every 5 seconds. The latency is measured by recording the time between sending the request to the device and the actual execution of the effect, using a camera capturing at 120 frames per second.

The results are presented in Figure 6 and show that devices controlled by the manufacturer's cloud service (Tuya brand devices: Socket, TV, and Lamp) have an average latency of approximately 1.99 sec, 2.26 sec, and 2.11 sec, respectively. The YeeLight device, controlled via a local network edge computing controller, shows an average latency of less than one second, even with a first, odd measurement of 3.67 sec. Without this outlier measurement, the latency is 0.53±0.06 sec.

Device: Power plug				
Test #	Frames	Time (s)	Avg.	∂ (S.D.)
1	270	2,2500	1,99s	0,23s
2	263	2,1917		
3	253	2,1083		
4	273	2,2750		
5	251	2,0917		
6	224	1,8667		
7	200	1,6667		
8	200	1,6667		
9	226	1,8833		
10	225	1,8750		

Device: Bulb (Tuya)				
Test #	Frames	Time (s)	Avg.	∂ (S.D.)
1	265	2,2083	2,11s	0,10s
2	264	2,2000		
3	250	2,0833		
4	244	2,0333		
5	252	2,1000		
6	261	2,1750		
7	238	1,9833		
8	275	2,2917		
9	246	2,0500		
10	236	1,9667		

Device: TV (universal IR control)				
Test #	Frames	Time (s)	Avg.	∂ (S.D.)
1	346	2,8833	2,26s	0,31s
2	263	2,1917		
3	277	2,3083		
4	237	1,9750		
5	283	2,3583		
6	288	2,4000		
7	260	2,1667		
8	300	2,5000		
9	206	1,7167		
10	251	2,0917		

Device: Bulb (YeeLight)				
Test #	Frames	Time (s)	Avg.	∂ (S.D.)
1	440	3,6667	0,85s	0,99s
2	64	0,5333		
3	56	0,4667		
4	80	0,6667		
5	61	0,5083		
6	66	0,5500		
7	64	0,5333		
8	68	0,5667		
9	66	0,5500		
10	54	0,4500		

Fig. 6: Latency between orchestrator request and the response of four devices. Red color: largest delay. Blue color: smallest delay

B. The Benefit of a fog-based architecture

The experiments in the previous section highlighted the challenge of cloud-based services to timely control local devices, which is especially critical for audio-centric applications. A fog-based orchestrator that can directly access local IoT devices and can interface with remote cloud-based services has the potential to noticeably reduce latency and improve the Quality of Service (QoS) in pervasive storytelling scenarios. This will consequently also enhance the Quality of experience (QoE). Furthermore, because less data will be sent to the cloud, the interactor's privacy can be better protected.

VI. Final Remarks

In conclusion, the exploration of pervasive storytelling and its integration with emerging technologies has opened up exciting possibilities for interactive and immersive narrative experiences. Throughout this work, we have delved into various aspects of pervasive storytelling, including its definition, challenges, and potential solutions. The concept of pervasive storytelling revolves around creating narratives that seamlessly blend with the real world, leveraging technologies like IoS devices, geolocation sensors, and cloud computing to provide dynamic and contextually relevant experiences for users.

The proposed A-Presto model aims at the definition of high-level abstractions that facilitate the creation and orchestration of pervasive storytelling experiences. A-Presto's focus on audio-centric storytelling, combined with the consideration of asynchronous events and the introduction of novel entities like the Aura, demonstrates its potential to advance the field. Furthermore, the discussion on the Orchestrator highlights the importance of incorporating fog computing to reduce latency and enhance interactivity, crucial for delivering smooth and engaging experiences in pervasive environments.

As pervasive storytelling continues to evolve, challenges related to privacy, data security, and cross-platform compatibility will require careful consideration. Additionally, efforts to develop user-friendly authoring tools and standardized languages will be essential to democratize the creation process and foster broader adoption of pervasive storytelling across different domains.

Acknowledgment

This work has been supported by UFJF's High-Speed Integrated Research Network (RePesq) https://www.repesq.ufjf.br.

References

[1] L. Turchet, M. Lagrange, C. Rottondi, G. Fazekas, N. Peters, J. Østergaard, F. Font, T. Bäckström, and C. Fischione, "The internet of sounds: Convergent trends, insights and future directions," *IEEE Internet of Things Journal*, 2023.

[2] T. Pløhn, S. Louchart, and T. Aalberg, "Dynamic pervasive storytelling in long lasting learning games," *Electronic Journal of e-Learning*, pp. 193–205, 2015.

[3] J. H. Murray, *Hamlet on the holodeck: The future of narrative in cyberspace*. New York, NY: Simon and Schuster, 1997.

[4] R. H. R. Harper, M. Rauterberg, and M. Combetto, "Backseat playgrounds: pervasive storytelling in vast location based games," in *Proceedings of the 5th International Conference on Entertainment Computing*, ser. Lecture Notes in Computer Science. Berlin: Springer, 2006, pp. 117–122.

[5] M. O. Jewell and C. J. Hooper, "Linking locations: Storytelling with pervasive technology," in *Narrative and Hypertext Workshop at Hypertext'11*, June 2011.

[6] Niantic. (2016) Pokemon go. [Online]. Available: https://pokemongolive.com/en/

[7] F. Colace, M. Casillo, S. Lemma, M. Lombardi, and M. De Santo, "A context-aware approach for a collaborative, pervasive and adaptive digital storytelling," in *Proceedings of the 22nd International Conference on Distributed Multimedia Systems*, 2016, pp. 165–174.

[8] Amazon. (2021) Alexa skills kit - what are alexa skills? [Online]. Available: https://developer.amazon.com/en-US/alexa/alexa-skills-kit

[9] ——. (2021) Create alexa skills in minutes. [Online]. Available: https://blueprints.amazon.com

[10] K. Hentschel and J. Francombe, "Framework for web delivery of immersive audio experiences using device orchestration," in *Adjunct Proceedings of ACM TVX*, Manchester UK, 2019.

[11] ETSI TS 102 796, "Hybrid Broadcast Broadband TV (HbbTV) 2.0.3 Specification," European Broadcasting Union, Geneva, 2020.

[12] D. Geary, J. Francombe, K. Hentschel, and D. Murphy, "Design dimensions of co-located multi-device audio experiences," *Applied Sciences*, vol. 12, no. 15, p. 7512, 2022.

[13] K. Hentschel and J. Francombe, "Exploring audio device orchestration in workshops with audio professionals," in *Audio Engineering Society Convention 148*. Audio Engineering Society, 2020.

[14] J. Francombe, J. Woodcock, R. J. Hughes, R. Mason, A. Franck, C. Pike, T. Brookes, W. J. Davies, P. J. Jackson, T. J. Cox *et al.*, "Qualitative evaluation of media device orchestration for immersive spatial audio reproduction," *Journal of the Audio Engineering Society*, vol. 66, no. 6, pp. 414–429, 2018.

[15] L. F. G. Soares and R. F. Rodrigues, "Nested context model 3.0: Part 1–NCM core," Departamento de Informática PUC-Rio, Rio de Janeeiro, Monografias do Departamento de Informática da PUC-Rio 12, 2005.

[16] F. Barreto, R. S. de Abreu, E. B. B. Montevecchi, M. I. P. Josué, P. A. Valentim, and D. C. Muchaluat-Saade, "Extending ginga-ncl to specify multimodal interactions with multiple users," in *Proceedings of the Brazilian Symposium on Multimedia and the Web*, ser. WebMedia '20. New York, NY, USA: ACM, 2020, p. 281–288.

[17] R. O. Rodrigues, M. I. P. Josué, R. S. Abreu, G. F. Amorim, D. C. Muchaluat-Saade, and J. A. F. d. Santos, "A proposal for supporting sensory effect rendering in ginga-ncl," in *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, ser. WebMedia '19. New York, NY, USA: ACM, 2019, p. 273–280.

[18] M. Ivanov, M. F. Moreno, and D. C. Muchaluat-Saade, "Automatic preparation of media objects in multimedia applications," in *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, ser. WebMedia '19. New York, NY, USA: ACM, 2019, p. 289–296.

[19] M. Josué, D. Muchaluat-Saade, and M. F. Moreno, "Preparation of media object presentation and sensory effect rendering in mulsemedia applications," in *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web*, ser. WebMedia '18. New York, NY, USA: ACM, 2018, p. 45–52.

[20] W3C, "Html 5.2," World Wide Web Consortium, W3C Recommendation, 2017. [Online]. Available: https://www.w3.org/TR/2017/REC-html52-20171214/

[21] ——, "Synchronized multimedia integration language (SMIL 3.0)," World Wide Web Consortium, W3C Recommendation, 2008. [Online]. Available: http://www.w3.org/TR/2008/REC-SMIL3-20081201/

[22] M. A. Freesz, L. Yung, and M. F. Moreno, "Storm: A hypermedia authoring model for interactive digital out-of-home media," ser. WebMedia '17. New York, NY, USA: ACM, 2017, p. 41–48. [Online]. Available: https://doi.org/10.1145/3126858.3126889

[23] B. D. Junqueira, "Competência emocional e competência informacional: a aquisição qualificada de competências a partir da narração de histórias," Master's thesis, Faculdade de Biblioteconomia e Comunicação, 2010.