D4.2.1	
Version	1.0
Author	NAEVATEC
Dissemination	PU
Date	27/01/1015
Status	Final

D4.2.1: Multisensory and Multi-Domain Media Element Middleware v1

Project acronym:	NUBOMEDIA
Project title:	NUBOMEDIA: an elastic Platform as a Service (PaaS) cloud for interactive social multimedia
Project duration:	2014-02-01 to 2016-09-30
Project type:	STREP
Project reference:	610576
Project web page:	http://www.nubomedia.eu
Work package	WP4
WP leader	Javier López
Deliverable nature:	Prototype
Lead editor:	Javier López
Planned delivery date	01/2015
Actual delivery date	27/01/2015
Keywords	Kurento, media server, multisensory, multidomain

The research leading to these results has been funded by the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement nº 610576

FP7 ICT-2013.1.6. Connected and Social Media

Contributors:

Javier Lopez (NAEVATEC)
Ivan Gracia (NAEVATEC)
J. A. Santos Cadenas (NAEVATEC)
Miguel París Díaz (URJC)

Internal Reviewer(s):

Luis Lopez (URJC)

Version History

Version	Date	Authors	Comments
0.1	03-01-2015	F. Javier Lopez	Initial Version
0.2	10-01-2015	Ivan Gracia	Contributions to architecture and software implementation.
1.0	22-01-2015	Ivan Gracia	Final picture added.

Table of contents

Luis Lópe
Eliminad

Luis Lópe
Eliminad

List of Figures:

Acronyms and abbreviations:

API	Application Programming Interface
AR	Augmented Reality
IMS	IP Multimedia Subsystem
IoT	Internet of Things
KMS	Kurento Media Server
RTC	Real-Time Communications
RTP	Real-time Transport Protocol
SCTP	Stream Control Transmission Protocol
VCA	Video Content Analysis
WebRTC	Web Real Time Communications

1 Executive summary

This document contains a description of the extensions to be created to the Kurento Media Server (KMS), as described in deliverable D4.1.1 providing multisensory and multi-domain capabilities. These extensions take the form of KMS modules and have two different natures:

- Multisensory extensions: This refers to the capability of managing multimedia flows containing more than audio and video information. In general, multisensory extensions are related to the integration on the multimedia stream of additional sensor data (i.e. biomedical sensors, smart-city sensors, etc.)
- Multi-domain extensions: This refers to the capability of transducing (i.e. translating) multisensory information into audio-visual information comprehensive for a human receiver. In other words, multi-domain extensions are KMS filters capable of representing sensor data, or other data, to the human receiver using some kind of Augmented Reality (AR) mechanism.

2 Introduction

2.1 State-of-the-art on multisensory multimedia

The pervasive presence of sensors and the emergence of the IoT model have added pressure from users and providers to combine sensor data with audiovisual information. This combination, which is sometimes referred as the cross media model or as the multisensory multimedia model, requires new types of technologies capable of transporting and managing arbitrary sensor data or arbitrary metadata as part of the multimedia stream, which traditionally has only contained audio and video.

In current state-of-the-art, there are two possible strategies for implementing multisensory multimedia: in-band and out-of-band.

The in-band strategy consists on extending current media formats to be able to manage sensor data into the same protocols and formats used by audiovisual data. In other words, the in-band strategy uses the same channel (i.e. band) for audio, video, sensor and other data. This approach has a number of advantages

- Synchronization is simplified as all streams may share common clock references.
- Transport is simplified as all streams share the same low level channel

But also has drawbacks:

- It requires huge modifications on the client applications using traditional audiovisual-based multimedia to support the new standards and models.

This strategy is the one used by standards such as MPEG-21, where different types of media data can be inserted as "Digial Items" into the same shared medium. It is also used for video subtitling as specified in RFC 4103 (i.e. ITU T.140), where meta-data information is transported using common RTC audiovisual transport protocols (i.e. RTP) and mechanisms.

The out-of-band strategy consists on letting untouched current (and common) audiovisual formats and use additional channels, which may use different protocols and formats, for the transport and management of sensor and metadata. This approach has a number of advantages

- It maintains compatibility with current state-of-the-art multimedia technologies, so that you don't need to modify them for supporting the integration of sensor data.
- It is more simple and efficient to implement, given that the additional channels and formats can be designed specifically adapted to the characteristics of the target data.

This model is currently being used in most IoT deployments involving multimedia communications, including many transport-agnostic XML formats (e.g. NGSI, BiTXML).

In particular, the emergence of WebRTC technologies has brought a revolution to the multisensory multimedia arena with the introduction of the WebRTC DataChannel standards. RTCDataChannels are an out-of-band technology with a number of very relevant advantages:

- They provide low-latency transports based on RTP and SCTP compatible with the low-latency ones of audiovisual data.
- They are implemented in all WWW browsers compatible with WebRTC standards.
- They have been designed for sharing the same ICE channels of the WebRTC multimedia audiovisual information, so they inherit all their properties (i.e. NAT traversal, compatibility with P2P communication models, etc.)
- The W3C has standardized a coherent API for accessing them.
- It is WWW friendly, meaning that it is fully compatible with well known data representation formats such as XML and JSON. However, it is agnostic to such formats, being possible to exchange any kind of data including binary blobs.
- There are currently hundreds of developers creating multisensory applications basing on RTCDataChannels, which guarantees its support and maturity.

Hence, the adoption of RTCDataChannels does not require any modification of the audiovisual capabilities of the clients, just the addition of an additional channel for the communication and management of the desired data and meta-data.

2.2 State-of-the-art on multi-domain multimedia

The representation of sensor data in a human comprehensive way through Augmented Reality (AR) techniques is quite in the mood in the last few years, and the scientific and engineering bibliography contains many references related to technologies and use cases for it [1-4]. In particular, the emergence of new AR wearable devices such as glasses or helmets is bringing a new gold era to these types of technologies.

There are different approaches for facing this technological problem, however, in general all of them are based on implementing the following methodology:

- First, data is captured by some kind of sensor either remotely or at the device of the end-user. In the former case, a specific protocol is required for the transport of the sensor data and, eventually, of the media data, as described above in this document.
- Second, some kind of AR object is created representing data in a human comprehensive way. This AR object may have different natures including:
    - Text showing the data in numeric or verbal format.
    - 2D graphs, which may include gauges, pie-charts, bar-charts, line-graphs, etc.

    - 3D objects, which may represent the status of the sensor through some kind of reality (e.g. a heat-beating, a car running, etc.)
- Third, the AR object is rendered on top of the video stream visualized by the end-user either floating at some specific coordinates either hooked through some kind of marker or marker-less mechanism.

3 Objectives

The objective of this deliverable is to design, implement, integrate, specify and document the appropriate multisensory and multi-domain technologies suitable to be used in NUBOMEDIA and accordingly to NUBOMEDIA partners' requirements. The execution of this objective requires the fulfillment of a number of sub-objectives, which include:

- To determine the better technological approach for providing multisensory transport for NUBOMEDIA.
- To design the appropriate software architecture for implementing such transport onto NUBOMEDIA client and server platforms.
- To create the appropriate implementation as KMS modules and endpoints.
- To determine the better technological approach for providing multi-domain capabilities for NUBOMEDIA.
- To design the appropriate software architecture for implementing such capabilities into the NUBOMEDIA infrastructure.
- To implement the architecture through the creation of the necessary KMS modules.
- To integrate into such modules the associated AR capabilities providing sensor data transducing and translation.

4 Implementation strategy

4.1 Implementing multisensory media elements

As we have discussed in sections above, for the implementation of NUBOMEDIA multisensory capabilities we have two options: in-band and out-of-band mechanism. The former has the advantage of being more compact and solving better synchronization issues (e.g. rendering of the sensor data synchronized with the audiovisual data). However, they have a clear an important disadvantage: lack of universal support. Currently, very few platforms and applications support such types of formats and implementing them for all NUBOMEDIA target platforms is out of the reach of a small project like NUBOMEDIA.

Hence, the best option for us is to make use of out-of-band mechanism, so that NUBOMEDIA multisensory technologies can coexist with state-of-the-art video technologies (i.e. players, recorders, endpoints, etc.) Following the discussion above, the most reasonable strategy to follow is to provide support for WebRTC (RTC) DataChannels and inherit all the advantages of standardized (although standards are still under development, they are maturing quite fast) WWW technologies.

This implementation strategy requires integrating RTCDataChannel capabilities at KMS. For achieving this, we propose the following methodology:

- To integrate RTCDataChannel support into KMS WebRtcEndpoint. This integration must expose, as a minimum, SCTP transport capabilities.
- To expose RTCDataChannel capabilities through a simple to use API accessible from the WebRtcEndpoint. This API needs to make possible to receive and send arbitrary data based on both textual and binary formats.

4.2 Implementing multi-domain media elements

In the case of multi-domain media elements, or transducing multisensory information into audio-visual information, the media server's elements need to be able to send and receive data from and to other media elements. The data sent could then be overlaid on a media stream, being passed through the media element that is receiving this data.

5 Software architecture

5.1 Multisensory media elements

The WebRTC standard defines a mechanism to exchange not only media, but also data. It is a feature called Data Channels, that is to be implemented in the short-medium term. This element is basically a Session Control Transmission Protocol (SCTP) connection encrypted using DTLS. In consonance with the existing structure of the WebRtcEndpoint, the approach is to create two GStreamer elements (KmsSctpDec and KmsSctpEnc) that will marshall and unmarshall SCTP packets. These two elements will be then connected to the existing GstDtlsSrtpDec and GstDtlsSrtpEnc, where the packets will be encrypted or decrypted.

The following diagram shows the internal structure of the WebRtcEndpoint, where the new GStreamer elements would be placed.

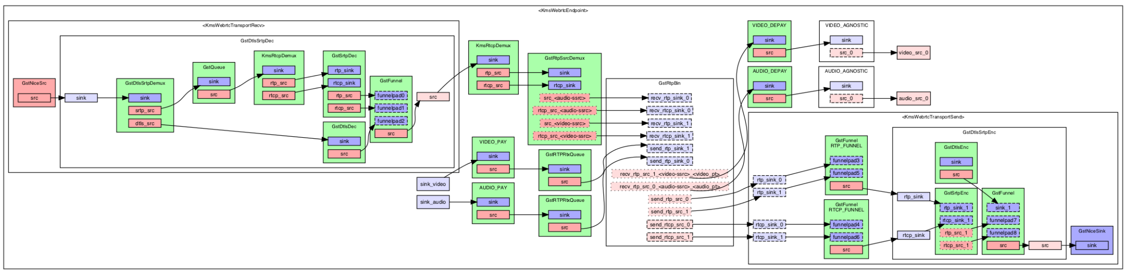

Figure 1. WebRtcEndpoint internal structure expressed in term of GStreamer media elements.

5.2 Multi-domain media elements

In order to support data flow, it is needed to provide KmsElement with a data channel. These modifications are depicted in the connection graph in the following figure.

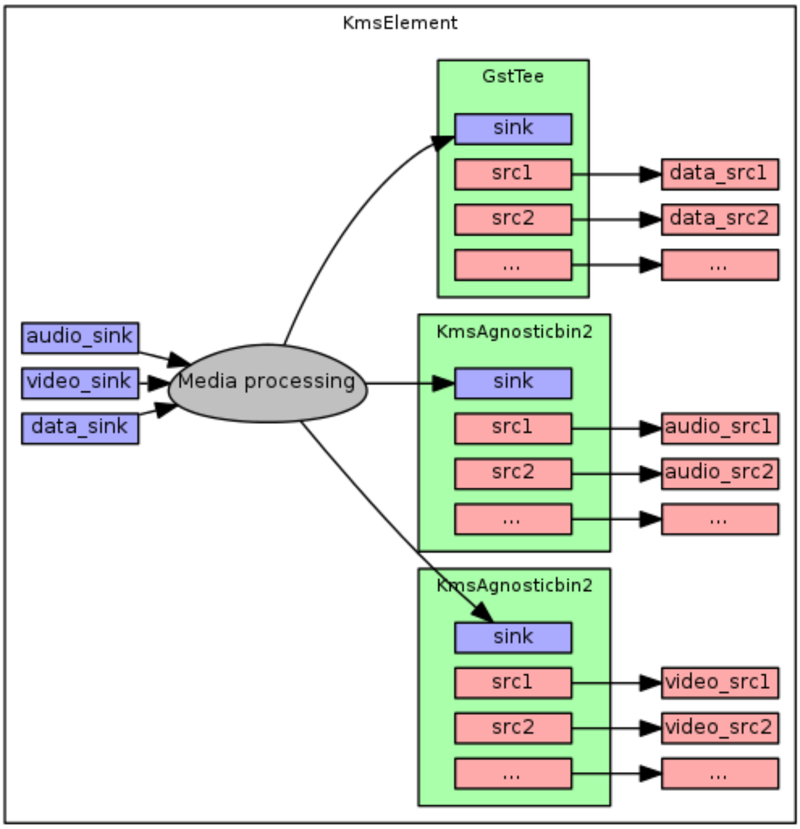

Figure 2. KmsElement connection graph expressed in term of GStreamer media elements.

There will be only one kind of data by default, meaning that each buffer will be considered as an independent message. It is the receiver of the buffer the one in charge of providing semantics to each message, be it a remote receiver or another internal GStreamer element. In this way, data are just non-changing streams for GStreamer, so they don't require any special negotiation.

As for synchronization, buffers in data channels are marked with a pts, analogous to the ones that audio and video buffers have. This mark allows an element to synchronize in case it is required. For this purpose, the receiver element will be responsible of keeping the packages of a certain type during a variable amount of time, until it is considered that no new buffers of the type to synchronize with arrive. For instance, video packages are retained for a couple of milliseconds, in case a data package arrives and synchronization with this data package is required.

This mechanism does not guarantee synchronization in every circumstance, because if the media that is being awaited takes too long to arrive, there could be a desynchronization. In this case, each element will decide if the data is shown even if it did arrive late, or if it is critical to show data in perfect synchronization.

6 Implementation status (NUBOMEDIA Release 3)

Implementation efforts of multisensory and multi-domain media elements are planned o start during NUBOMEDIA Release 4.

7 References

[1] Payton, David W., et al. "Pheromone robotics." Intelligent Systems and Smart Manufacturing. International Society for Optics and Photonics, 2001.
[2] Narzt, Wolfgang, et al. "Augmented reality navigation systems." Universal Access in the Information Society 4.3 (2006): 177-187.
[3] Vazquez-Alvarez, Yolanda, Ian Oakley, and Stephen A. Brewster. "Auditory display design for exploration in mobile audio-augmented reality." Personal and Ubiquitous computing 16.8 (2012): 987-999.
[4] Akinbiyi, Takintope, et al. "Dynamic augmented reality for sensory substitution in robot-assisted surgical systems." Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE. IEEE, 2006.