Review of Media Sync Reference Models: Advances and Open Issues

Mario Montagud
Jack Jansen
Pablo Cesar
Centrum Wiskunde & Informatica (CWI)
Amsterdam, the Netherlands
montagud@cwi.nl
Jack.Jansen@cwi.nl
P.S.Cesar@cwi.nl

Fernando Boronat
Universitat Politècnica de València (UPV)
Grau de Gandia, Spain
fboronat@dcom.upv.es

Abstract
The advances on multimedia systems have brought new challenges and requirements for media sync. Over the years, many media sync solutions have been devised. Due to this variety, several studies have surveyed the existing solutions and proposed classification schemes or reference models for media sync. This paper claims the relevance of media sync reference models to systematically structure and synthesize this research area. Accordingly, a review of the existing reference models is provided, by examining the involved features, components and layers in each one of them. Likewise, some inconsistencies, open issues and missing components in existing reference models have been identified. Accordingly, this study reflects the need for a new modular and extensible theoretical framework or reference model to efficiently comprehend the overall media sync research area.

Author Keywords
Media Synchronization; Multi-Layer; Multi-Dimension; Reference Model; Theoretical Framework.

ACM Classification Keywords
C.2.2 [Communication Networks]: General (OSI Model);
C.2.2 [Communication Networks]: Distributed Systems.

Introduction
Media synchronization (sync hereafter) has been a key research area since the early development of (distributed) multimedia systems. Proper media sync solutions are necessary to guarantee the maintenance, during presentation or playout, of the spatial, semantical and/or temporal dependences within and between each media element in a (multi-)media system. The involved elements can include both continuous media (e.g., audio, video…) and discrete media (e.g., text, images…). Likewise, different types of media sync techniques can be distinguished, depending on the number of involved media elements,

MediaSync2015 Workshop, June 3, 2015, Brussels (Belgium)
ISBN/EAN: 978-90-5968-463-4

Media Sync Types (I)

Intra-media (aka intra-stream) **sync** is needed to maintain the original relationships between the Media Units (e.g., audio samples, video frames…) within each particular media element (e.g., audio, video, subtitles...).

Inter-media (aka inter-stream) **sync** is required to preserve the proper dependences between different media elements (e.g., lip-sync). A specific sub-type of inter-media sync is referred to as **inter-sender** (aka multi-source) **sync**, which aims to synchronize the playout of several media elements from different senders (or sources). It can also be possible that the media elements are delivered using different protocols, or even via different (e.g., broadcast and/or broadband) networks. In the latter case, this is usually referred to as **hybrid sync**.

sources (or sensors), senders, streams and receivers (see text boxes and Fig.1).

Over the years, many proprietary and standard solutions for each particular media sync type have been devised. Due to this variety, several works have surveyed the existing media sync solutions and proposed classification schemes and/or reference models to provide an overview and categorization of this research area. This paper is in this line, but it goes further. It is not our goal to survey every individual media sync solution (we rely on comprehensive works authored by other researchers and by ourselves for that), but to review the available media sync reference models. In particular, this paper examines what are the media elements, entities, components, criteria, patterns, layers and dimensions the existing media sync reference models have considered, their evolution, their common and distinct aspects, their relationships, and inconsistencies between them. Besides, this paper analyzes how much the existing reference models solve the current media sync problem space (see Fig.1).

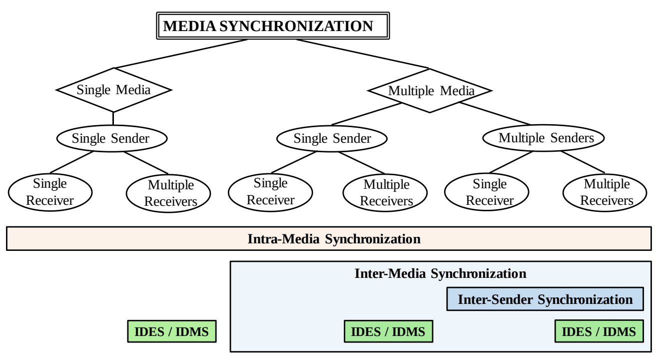

Figure 1: Overview of the Media Sync Problem Space.

On the one hand, we believe that a media sync reference model, if properly specified, can be very useful to synthesize and to systematically structure the overall media sync research area. Similarly, a review and taxonomy of the existing media sync solutions can be useful to: i) better understand this research area; ii) identify and compare the different approaches and strategies (with their strengths and weaknesses) that have been devised up to date to overcome different sync challenges; iii) analyze the evolution and latest advancements on media sync; iv) identify the necessary entities, components and features to meet the different sync demands; v) identify components and features that are not provided yet or that need further research; and vi) identify synonymous for key terms, as well as to facilitate and stimulate the use of a base and common vocabulary for media sync. Moreover, a study and categorization of the overall media sync area become convenient tasks before proceeding with the design of a specific media sync solution, as different types of media sync solutions have typically to cooperate and inter-operate in multimedia systems, probably sharing components and resources (e.g., bandwidth, delivery and feedback channels, processing and memory resources…).

On the other hand, even acknowledging the relevance of media sync reference models and the progress made in this area, our conviction is that there is no yet a reference model that (efficiently) solves the overall problem space for media sync. There are still some open issues (e.g., inconsistencies, components that have room for improvement…) and missing components. Due to this, it is necessary to re-visit and re-formulate the traditional research problems, but taking into account the recent technological advances

Media Sync Types (II)

Inter-destination media (aka inter-receiver of group) **sync** (**IDMS**) involves the simultaneous sync of the playout of the same or different media elements across different devices. The involved receivers can be either far apart (e.g., in multi-party conferencing, Social TV…) or physically close-by (e.g., in multi-screen scenarios). In the latter case, it is also typically known as **inter-device sync** (**IDES**).

and the emerging challenges and trends of the current media delivery and consumption paradigms.

Accordingly, this paper represents a first step towards the specification of a new modular and extensible theoretical framework or reference model to efficiently comprehend the overall media sync research area.

Review of Media Sync Reference Models

This section provides a review of different works (chronologically ordered) that have analyzed the state-of-the-art regarding media sync ([1-11]).

In [1], an analysis of the temporal and spatial composition of multimedia applications was presented. Accordingly, a classification model for both intra-stream and inter-stream sync, for both continuous and discrete media, was proposed. That model is composed of three sync levels (physical, system and human levels), but neither detailed description nor classification criteria are provided.

In [2], a classification scheme for media sync was presented. It is composed of the following three layers:

1) *Media layer*: it copes with intra-stream sync.

2) *Stream layer*: it copes with inter-stream sync of continuous media.

3) *Object layer*: it operates on top of the two previous layers and is responsible of offering to the multimedia application a complete and ordered multi-stream presentation, in which all media elements need to be correctly structured in time and space.

The hierarchical structure of this reference model and the abstraction level of each involved layer are shown in Fig.2. The services (i.e., the sync functionalities) provided by each layer can be accessed either directly by the multimedia application or indirectly through higher-level layers (via appropriate interfaces).

In [3], a classification of existing media sync solutions (up to 1994) was presented. It was focused on the location where the sync functionality was developed and performed: local and distributed. *Local sync techniques* are only implemented within workstations (i.e., media is captured/retrieved and consumed within single devices, without the intervention of networking equipment). *Distributed sync techniques* are used in networked environments and can also be divided into two main (sub-)approaches: i) the sync functionality is implemented at the senders and/or receivers; and ii) the sync functionality is also implemented at the involved inter-network devices. Both local and distributed sync techniques can involve a single or multiple media elements and sources.

In [4], another reference model was presented, which makes use of three design criteria to classify the existing media sync solutions. Each criterion is placed in a different orthogonal axis, such that the overall media sync problem space can be graphically systematized in a 3D cube. The criteria are:

- *Time*: whether the media sync solution makes use of global or local clocks.
- *Location*: whether the sync functionality is located at the server or at the client side.
- *Method*: the specific adjustment techniques that are used to achieve media sync.

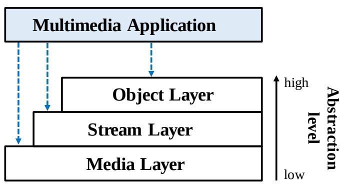

Figure 2: Three-layer Reference Model (Meyer Model) [2].

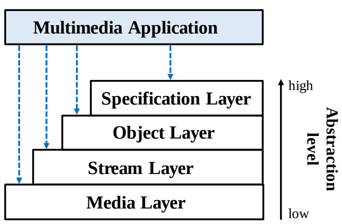

Figure 3: Four-layer Reference Model (Blakowsky Model) [5].

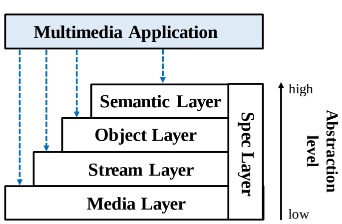

Figure 4: Five-layer Reference Model (Costa Model) [10].

The survey in [5] summarizes the media sync requirements and proposes a reference model to compare the existing intra-stream and inter-stream sync methods (up to 1996). This reference model is an evolved version of the one presented in [2]. In particular, a fourth layer, called the *Specification layer*, is added (see Fig. 3). It includes applications and tools for the creation of sync specifications. Examples of such tools are sync editors, multimedia document editors, formatting and conversion tools, authoring systems, etc. The sync specification will be used as an input to the *Object layer* for scheduling the overall presentation.

In [6], a comprehensive comparison between intra-stream and inter-stream sync solutions for continuous media was provided. Such solutions were compared in terms of the location of the sync functionality, the use or not of clock information, and the type of media (live or stored). Likewise, that study identified the control techniques used in each surveyed solution, classifying them into four categories: *common control*, *basic control*, *preventive control* and *reactive control.*

In [7], a comparative survey between intra-stream sync solutions (including playout adjustment techniques) is provided. It discusses issues related to timing information, handling of late MUs, quality evaluation metrics and adaptation to changing delay conditions.

Based on various classification criteria from [3], [4], [5] and [6], a thorough analysis and comparison between existing inter-stream sync and IDMS solutions (up to 2009) is provided in [8]. In particular, the following factors are taken into account in this taxonomy: i) the control scheme in use; ii) the use or not of global clock information; iii) if the particular media sync solution is only valid with bounded delay limits; iv) the use or not of a feedback channel; v) if the generation rate of MU is periodic or not; vi) if the solution is valid for stored or live media (or both); vii) the metadata used for media sync; and viii) the employed adjustment techniques.

The study in [9] provides a historical review of sync studies for continuous media, by also conveying the background of technological advancements (with the associated sync challenges and requirements), media sync modeling and human perception. That study claims an urgent need for the research community to further evolve and advance existing sync practices, standards and specifications. In particular, the insufficiency of the existing reference models to meet the sync demands in next-generation heterogeneous multimedia services was identified. For instance, telepresence systems, such as 3DTI (3D Tele Immersion), demand the following sync features: 1) heterogeneity of media modalities and devices; 2) support for scalable multi-party scenarios; 3) provision of the different media sync types; and 4) support for diverse multimedia applications on single platforms. However, it was claimed that the existing media sync reference models mainly focus on single dimensions, such as the location where the sync functionality is performed [3], or the type of sync demands [5], but do not support a combined interaction between these dimensions (i.e., the existing models mostly cover orthogonal dimensions). As an example, Fig.5 illustrates the relationships (i.e., the shared layers and functionalities) and limited interactions between the reference models proposed in [1], [3] and [5].

Consequently, a new multi-dimensional (i.e., multi-requirement, multi-modal, multi-layer, multi-device, multi-location and multi-activity) classification model is proposed in [9]. First, that model takes into account the scalability and heterogeneity of devices and media modalities (e.g., audio, video, haptics, sensory data…). Second, this model addresses the different types of sync demands in five hierarchical layers (i.e., multi-layer), as show in Fig.4, following the approach in [5]. The *Media Layer* in [5] is called *Intra-Stream Layer* in [9], and it copes with intra-stream sync. The *Stream layer* in [5] is divided into separate layers: *i)* the *Intra-Media Layer,* which copes with inter-stream sync of streams of the same media modality (e.g., arrays of video cameras or microphones), called *intra-media sync* in [9]; *ii)* the *Intra-Bundle Layer,* which copes with the inter-stream sync of streams of different modality (e.g., audio, video and haptics), called *intra-bundle sync* in [9]; *iii)* the *Intra-Session Layer,* which copes with inter-sender sync. Moreover, the *Intra-Session Layer* also copes with IDMS (in [9], the term *intra-session sync* is used to refer to both inter-sender sync and IDMS). The *Object layer* is not covered, as it is considered that the functionalities provided by that layer in the model in [5] are enough. Likewise, the *Specification layer* is the same as in [5]. Third, a multi-location dimension is added in order to encompass the end-to-end delivery chain (i.e., server, distribution and client sides), by extending the location-based model in [3]. The idea is to add sync control at each involved entity involved in the end-to-end delivery and sync processes. The reason is because sync skews in a specific layer occasioned in one location (e.g., delay variability when capturing and encoding media content at the server side) can be propagated to the other locations (e.g., to the network and client sides), thus having an impact on the sync performance.

Finally, a fourth *application-dependent* or *activity-dependent* dimension is added to characterize the heterogeneity of applications and performed activities on the human perception. This dimension is relevant because it is not appropriate to use a single media sync reference model to represent all possible use cases or applications, as the requirements on temporal sync, and the sync reference (being this reference a specific device, site, stream, media type, and/or participant) to be selected are largely application-dependent. Likewise, each multimedia application should be able to adaptively select the most proper sync references, based upon the functionality of the performed activities, the network and end-systems conditions, and on the users' requirements or interests.

The orthogonal dimensions, with their hierarchical structure (if any), of this media sync reference model are illustrated in Fig.6.

.

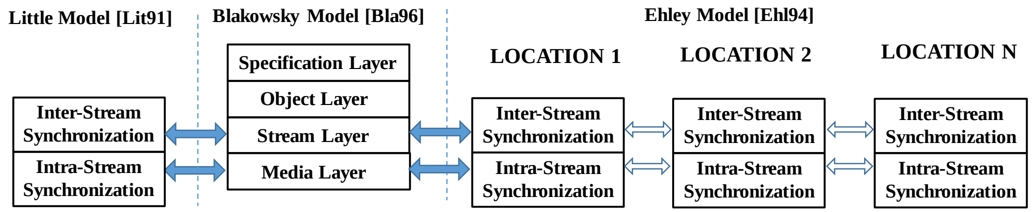

Figure 5: Interaction between Media Sync Reference Models (re-drawn from [9])

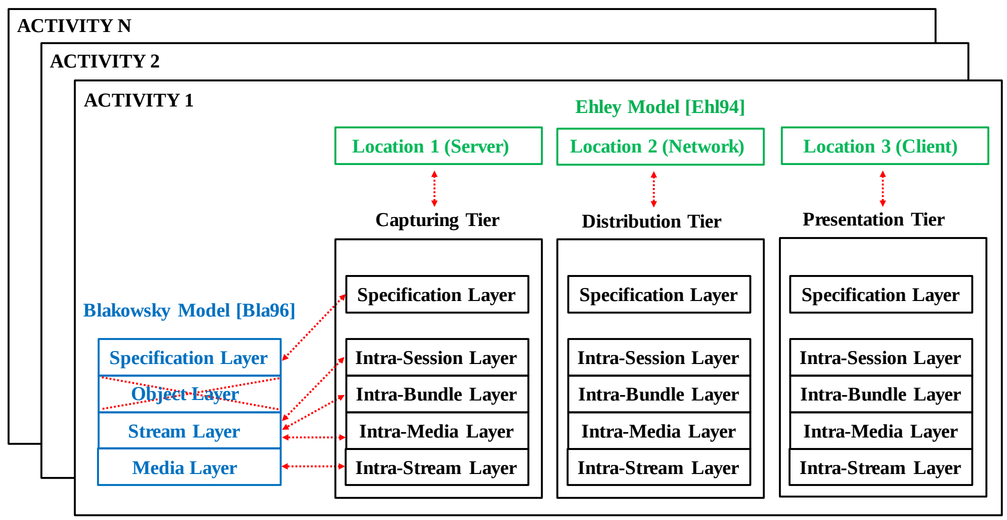

Figure 6: Multi-dimensional Media Sync Reference Model in [10].

The study in [10] provides a systematic literature review and mapping study on TV content sync. That study surveys the existing media sync solutions focused on the TV area, classifying them in terms of: types of involved devices, types of media content, types of sync techniques, targeted applications or scenarios, and evaluation methodologies. The following sync specific aspects are considered to classify the existing solutions: protocols, algorithms, delivery channels, specification methods, architectural schemes, allowable asynchrony levels and evaluation metrics. Likewise, the four-layer model in [5] is slightly modified and extended in [10] (see Fig.4). First, the *Specification layer* is relocated. It is argued that this layer should not be an isolated layer, but it must be bound to the other layers, since all of them need their own sync specification. Second, a fifth layer, called *Semantic layer*, is added on top of the *Object layer*, which has to cope with IDMS, content-based sync and contextual information (e.g., cross media, mash-ups…). According to [10], the *Semantic* layer is responsible of communication, search, retrieval and interpretation of media content and playout timings. It has to take into account the semantic relationships between the involved types of content being consumed, as well as how to access, generate and/or consume extra related content. This layer is essential to enable advanced sync-sensitive services, such as personalization, interactive services, multi-screen settings, etc.

Finally, the review and taxonomy of IDMS solutions from [8] is updated and extended in [11], by taking into account other relevant factors, such as the involved media types, the targeted application, and the evaluation methodology and metrics employed in each solution.

Discussion

After the review of the existing media sync reference models, some open issues and missing components can be reflected.

First, the lack of consistent, rigorous and comprehensive layering policies has been noticed. Some of the existing reference models differ in the number of layers, as well as on the location and functionality of each one of them. Besides, clearer and more complete descriptions of the functionalities offered by each layer are necessary.

Second, the dependences and interactions between the proposed layers and dimensions are not sufficiently specified. It is important to specify inter-layer and inter-dimension policies, since the key issue is that interaction and exchange of services between them can (efficiently) happen.

Third, the introduction of the Semantic Layer is a good point, but it may be better to add it as a sub-layer of the Specification Layer. Besides, the IDMS functionality must not be placed at this layer, because the semantic relationships are relevant to enable the different forms of media sync, not only for IDMS. Furthermore, we also think it is not a good idea to place the IDMS functionality at the Intra-Session Layer, together with the inter-sender sync functionality, as proposed in [10]. According to this, the most appropriate location for the IDMS and IDES functionalities needs to be analyzed.

Fourth, the Specification Layer has been mainly targeted for indicating temporal relationships, but weak support for providing semantical and spatial relationships is provided.

Finally, a key missing aspect in the existing reference models is the support of "user-level" sync. The users are the most important and central "components" (i.e., the mainstay) of multimedia systems. Therefore, contemporary media sync solutions need to take into account important "user-level" aspects, such as their needs, preferences, interests, attention, presence, perceptual issues as well as contextual aspects. Likewise, in interactive scenarios, the participants' roles, the conversation dynamics, as well as social and psychological aspects need to be considered to enable truly socially-aware or context-aware multimedia systems.

Given these open issues and missing components, we will put our efforts towards the specification of a modular and extensible reference model or theoretical framework to efficiently comprehend the overall media sync research area. We believe it will contribute to consolidate the advances on media sync and to drive the future research on this area.

Acknowledgements

The work by Mario Montagud has been carried out during the tenure of an ERCIM 'Alain Bensoussan' Fellowship Programme. UPV work has been funded, partially, by the "Fondo Europeo de Desarrollo Regional (FEDER)" and the Spanish Ministry of Economy and Competitiveness, under its R&D&I Support Program, in project with reference TEC2013-45492-R.

References

1. T.D.C. Little and A. Ghafoor, "Spatio-temporal composition of distributed multimedia objects for value-added networks", IEEE Computer, vol. 24, no. 10, pp. 42–50, October 1991.
2. T. Meyer, W. Effelsberg, R. Steinmetz, "A taxonomy on multimedia synchronization", Fourth Workshop on Future Trends of Distributed Computing Systems, pp. 97-103, September 1993.
3. L. Ehley, B. Furht, M. Ilyas, "Evaluation of multimedia synchronization techniques", International Conference on Multimedia Computing and Systems, pp. 514-519, May 1994.
4. D. Köhler, H. Müller, "Multimedia playout synchronization using buffer level control, Second International Workshop on Advanced Teleservices and High-Speed Communication Architectures, pp.167–180, Heidelberg (Germany), Sept. 1994
5. G. Blakowski, R. Steinmetz, "A media synchronization survey: reference model, specification, and case studies", IEEE JSAC, 14(1), pp. 5-35, January 1996.
6. Y. Ishibashi, S. Tasaka, "A comparative survey of synchronization algorithms for continuous media in network environments", IEEE LCN 2000, pp.337-348, Florida (USA), November 2000.
7. N. Laoutaris, and I. Stavrakakis, "Intrastream synchronization for continuous media streams: a survey of playout schedulers", IEEE Network Magazine, 16 (3), 30-40, 2002.
8. F. Boronat, J. Lloret, and M. García, "Multimedia group and inter-stream synchronization techniques: A comparative study", Information Systems, 34(1), pp. 108-131, March 2009.
9. Z. Huang, K. Nahrstedt, R. Steinmetz, "Evolution of temporal multimedia synchronization principles: A historical viewpoint", ACM TOMCCAP. 9, 1s, Article 34, 23 pages, October 2013.
10. R. M. Costa, C. A. S. Santos, "Systematic Review of Multiple Contents Synchronization in Interactive Television Scenario", ISRN Communications and Networking, Volume 2014, Article ID 127142, pp. 1-17, February 2014.
11. M. Montagud, Design, Development and Evaluation of an Adaptive and Standardized RTCP-Based IDMS Solution, PhD thesis, Co-supervisors: F. Boronat and P. Cesar, Universitat Politècnica de València (UPV), Spain, March 2015.