

Multimedia Information and Authoring for Personalized Media Networks

Insook Choi^{1,*}, Robin Bargar²

Abstract

Personalized media includes user-targeted and user-generated content (UGC) exchanged through social media and interactive applications. The increased consumption of UGC presents challenges and opportunities to multimedia information systems. We work towards modeling a deep structure for content networks. To gain insights, a hybrid practice with Media Framework (MF) is presented for network creation of personalized media, which leverages the authoring methodology with user-generated semantics. The system's vertical integration allows users to audition their personalized media networks in the context of a global system network. A navigation scheme with dynamic GUI shifts the interaction paradigm for content query and sharing. MF adopts a multimodal architecture anticipating emerging use cases and genres. To model diversification of platforms, information processing is robust across multiple technology configurations. Physical and virtual networks are integrated with distributed services and transactions, IoT, and semantic networks representing media content. MF applies spatiotemporal and semantic signal processing to differentiate action responsiveness and information responsiveness. The extension of multimedia information processing into authoring enables generating interactive and impermanent media on computationally enabled devices. The outcome of this integrated approach with presented methodologies demonstrates a paradigmatic shift of the concept of UGC as personalized media network, which is dynamical and evolvable.

Key Words: Multimedia Information and Authoring System, Multimodal Architecture, Personalization, Semantic Network

I. INTRODUCTION

A media work produced for consumers is likely to be packaged as a genre. In new media practice, the currency of a genre tends to be coupled with social adaptation of a technology configuration conceived as a platform. An example is the CD-ROM. Once a promising genre, CD-ROM based interactive storytelling faded away in the mid 1990s as web based technology configurations were adopted [1]. Similarly in text-based communication, email is replaced by text messaging and social text services such as WhatsApp, KakaoTalk and WeChat as mobile supplants desktop Internet access [2]. For the foreseeable future we expect platform transience to accelerate and diversify rather than converge to a dominant medium. Short-lived genres and under-defined use practices around emerging platforms merit human creativity and present a challenge to industry to support sustainable multimedia information systems and authoring practices.

With web 2.0, ICT research touches on topics such as sharing or on demand economy [3-6], platform economy [7], and collaborative consumption [8]. While scholars such as Benkler [9-11] emphasize the importance of

collective intelligence from a productive point of view, most works on these topics focus on new patterns of consumption and related market opportunities. The motivation of the present research is to test and prototype a collaborative and productive system while defining what it means to be a deeper information structure, beyond the level of structure of technology configurations, and how to facilitate the creative act as a sustainable practice. To begin with, we need to understand a pattern of emergence of a genre, especially around personalized media. We look at the emergence of its transactional structure, which leverages technology configurations such as web services. For deep information structure we apply semantic data to persist across variable technology configurations. We identify personalization and responsiveness of media experience as two factors in one intricate circuit of foreseeable user demands and expectations, and we present architecture to test the system requirements to meet them.

1.1. Two Classes of Responsiveness

The 21st century continues to see increasing challenges to multimedia information systems as everyday consumers expect highly responsive media experiences delivered to

Manuscript received September 11, 2017; Revised September 15, 2017; Accepted September 21, 2017 (ID No. JMIS-2017-0031)

Corresponding Author (*): Insook Choi, MediaCityUK, U. Salford, Room 2.08C, Salford M50 2HE, UK, +44 161 295 6207

¹School of Arts and Media, University of Salford, Greater Manchester, UK, insook@insookchoi.com

²University of Salford, MediaCityUK, Greater Manchester, UK, r.l.bargar@salford.ac.uk

their everyday devices. As devices evolve so do users' expectations, demanding greater responsiveness for more intimate media experiences. In use practice, different devices provide conditions for different expectations of functions and responsiveness. To facilitate quality user experience, the technology affordance for responsive media is broadly engineered by GPU processing, lossless compression, capacitive sensing, and feature detection combined with Web 3.0, HTML5, and 4G, 5G and beyond with locally-high bandwidth. Multiple layers of structured information processing are required to intimately handle end point user experiences, which in turn, generate further expectations to be translated into system requirements to meet the expectations. In the meantime, the definition of devices goes through transformation in a larger time frame. A canonical example is a cell phone, initially designed for voice communications with mobility, later becoming a multifunctional device highly responsive to users for accessing both local and remote multimedia information.

This is where a new perspective is needed for integrated information processing to account for both what users are doing and what users are expecting as respondents to system performance. With the multimodal dimension, we introduce a distinction between two classes of responsiveness: action responsiveness and information responsiveness. *Action responsiveness* relates to a perceived immediacy and requires a process that responds to what users are doing. This loosely couples with users' enhanced sensorimotor experience. *Information responsiveness* relates to a perceived relevance and requires a system to process both contextual information as well as personalized data to respond to what users are expecting. This loosely couples with users' enhanced cognitive experience. These two classes draw a distinction—for where in devices, physical or online sites, and when in time windows of user actions and expectations, and how in scheduling—to deliver responsiveness. On this note, the distinction between user action and expectation must be factored into an integrated approach to enable quality media experiences. An integrated approach needs dedicated experimentation to test the feasibility and methodology of co-processing multimedia and multimodal information. For this rationale we include section 4.3, the discussion of dedicated experimentation with IoT devices provided to physical theater professionals, as an intensive case enhancing the paper's main subject of multimedia information and authoring. IoT combined with human performance yields excellent opportunities in a playful experimental setting to generate cases to set benchmarks and test required functionality for multimodal architecture.

1.2. Extended Perspective on Media Network

The term “media network” carries multiple meanings across domains. From an engineering perspective a media network is a technology infrastructure for encoding media content as time-domain signal and transmitting the content to remote recipients. From a broadcasting perspective a media network is a group of geographically distributed providers who produce and distribute media content. From a social media perspective a media network describes a group of people who share media resources peer to peer, either by transmitting media files or by sharing network addresses that connect to remote files or media streams. Extending these perspectives, we describe a *personalized media network* as a *structure of media content linked to other media content*. Precursors of this extension are numerous. In 1944 Vannevar Bush proposed networks of linked documents in “memex” [12]; in the 1960's Project Xanadu [13] described a hypertext machine language to create nonlinear paths through documents [14, 15]. In 1956 semantic networks were introduced [16] and examples were implemented in the early 1960's [17, 18]. HTTP was first implemented in 1989 [19, 20] and the semantic web was proposed in 2001 [21]. These models call upon machine text analysis to generate semantic connections between documents based on automated content classification and recognition. Beyond text, audiovisual media present challenges to this approach, despite the advanced state of the art in automated feature recognition and classification of media content for consumer use cases. For example the Apple iPhone™ Photos™ app [22] automatically sorts photographs into rudimentary collections such as “places,” “people,” “selfies” and “moments” by combining visual object recognition with time, date and geolocation data. However, most techniques work reliably only when performed on a pool of samples that are pre-trained or well defined in terms of semantic classification. In reality, to yield meaningful information from associations between media contents there is a significant limitation in the success rate of automated recognition of objects and event features.

Industry applications show that semantic classification for large samples of media contents requires human intervention to identify associative meanings that emerge between recognized visual features. Netflix and Facebook curate, cleanse and recommend content at large scale by employing large teams for brute-force human analysis of content [23, 24], including text analysis as well as digital media. This indicates that semantic classification of large-scale media contents requires large teams and dedicated vocabulary that is human-monitored and supplemented with data about readers [25]. From this we can surmise that, for automated creation of media networks, depending solely on machine recommendation is insufficient if we

expect to generate personalized and responsive media content. Further, the meaning of a personalized media network is nested in a larger frame of media networks, which extend beyond the particulars of a single semantic classification; otherwise there would be no personalization. This leads to the necessity of an intermediary step using multimedia information and authoring that facilitates a hybrid of personalized and shared media networks. To implement such authoring system, the architecture presented in this paper facilitates vertical integration between 1) individual authors' perspectives and their constructed media networks, and 2) a system's global graph as an objectified and collective media network (see Fig. 1 in Section III, and Figs. 2 and 5 in Section IV).

1.3. Introduction to an Authoring Paradigm for Personalized Media Networks

Situating individual authors' work in a collective framework is a new paradigm introduced in this paper. Two overarching research and development questions were: 1) *Is it feasible to implement an authoring system to transform creative practice to be more collaborative, subjected to and relevant to information processing without compromising individuals' creativity?* 2) *Is it feasible to harvest semantics using a process that is native to the multimedia information and authoring system in one framework?* For the first question, the feasibility has been largely contested through project based learning exercises as reported in [26] and continues to be addressed through small scale productions (See Section V). Furthermore, recent trends in uses and gratification over media contents show how three usages are interdependent—to acquire information, to be connected, and to self express—and how greater gratification can be derived from use and reuse of user generated media contents [27]. The motivation for user generated media practice is already high: the system design objectives address collaborative methodologies, ease of use, or giving more user control over the ease of use. The second question yields overall design directives for system architecture and implementation.

The present authoring system facilitates users in a process of creating networks of media contents connected by metadata. The authoring process begins by creating or selecting a set of media contents, then making interpretations of the media contents and attributes. The interpretations provide descriptions that are attached to media content as semantic metadata. The metadata enables media contents to be connected by semantic association, in turn used to create personalized media networks. The framework for authoring networks of media contents enables users to interact with the networks to generate

responsive multimodal experiences.

Examining terminology such as “rich media”, “hyper media”, or “multimedia”, we note that descriptive terminologies carry subtle technological implications, anchored on user experiences of media responsiveness. In parallel to the previous distinctions made between action responsiveness and information responsiveness from users experience requirements, we generalize media signal processing requirements into two domain classes: *spatiotemporal* and *semantic*. This is to provide a unifying principle with clear differentiation of the two ends' requirements. Spatiotemporal and semantic structures characterize the “where” and the “how” of media responsiveness measured from users' end points. In turn these indicate when and how to suffice users' expectations with system performance in terms of technological requirements. Multimedia presentation with spatiotemporal and semantic signal processing combined with interactive display induces multimodal experiences. Authoring multimodal experiences requires a technology platform to support multiple media types and their related modes of user interaction.

The goal of this project is to develop cross-platform architecture for authoring and generating multimodal experiences, that can host diverse configurations of devices and network services defined as coherent multimedia information systems. Therefor in Section II, prior to proceeding to technical discussion, we first analyze transaction structure in network-based models of media platforms in the current media industry landscape. Transaction structure has a deep implication how media is produced because the production anticipates the form of dissemination. By constructing a table, this section surveys contemporary models of user-media transactions supported by media network services, and examines how these models contribute to a paradigm for authoring personalized media networks. Section III presents the Media Framework platform, reviews system interactivity requirements, and architecture. Sections IV and V demonstrate the vertical integration from individual users and collaborative authoring outcomes to the system's global media display with multiple levels of granularity and access points in a graph structure. Section IV focuses on the software designs for authoring and illustrates test cases. Examples include collaborative production of interactive documentary scenes and media controlled by IoT systems. Section V narrows down to focus on interface paradigms both for authoring and for end users' interaction, based on criteria for personalization and responsiveness. Section VI summarizes implications and outlines future directions.

II. TRANSACTION STRUCTURE IN NETWORK-BASED SERVICE MODELS FOR MEDIA PLATFORMS

Within a limited capacity, recent technology platforms enable a fusion of multimedia authoring and social media, which reduces a gap between the roles of content providers and consumers. This fusion implies certain direction towards future multimedia information systems and authoring practice. Four factors are associated with this implication: factor one – technology configuration in a form of emerging platforms; factor two – concept shift by new use case and through use practice; factor three – emergence of a genre as an exchange value; factor four – transactional structure of the network services model of a media platform.

2.1. Evolving Media Landscape

New generations of network services entail a shift in the relationship between media creators and consumers. In part this shift represents new models of media distribution and acquisition. An indicative example is the emergence of video platforms for on-demand streaming. These platforms have greatly increased the choices available to viewers and have also provided new modes of sharing video both privately and commercially. In 2005 YouTube introduced a “channel” strategy [28] that shifts the use-based meaning of “channel” from a historical definition to a new definition. Historically a channel was a frequency

band for terrestrial signal transmission, with scheduled content presentations controlled by a broadcasting company. Today a channel is a web address controlled by a content provider uploading media for viewers’ access at their choice. This example illustrates relationship of three factors (above): technology configuration, concept shift and transaction. Combining the technology of a video sharing platform with formalization of the concept shift of “channel,” all viewers are potential content providers, a role that increases their personalization of media transactions.

The collective impact of new models and services also entails the development of new media genres. A genre denotes a structure for media content and style of its presentation. Traditionally, a genre comes to social recognition after a long trajectory of collective practice of a kind, until it establishes a known practice such as a novel or a movie, an opera aria or a rock-‘n-roll song. Therefor a genre assumes a kind of social contract between a media maker and an audience—it defines audience expectations of the style a media work adopts.

To observe the impact of web services on new media genres we can consider the example of on-demand content delivery. New media producers for video on-demand can use web technology to measure social transactions in terms of video download or stream requests. A consumer video download requires a personal commitment that is greater than a consumer watching a broadcast. In many cases a download requires a consumer provide an identity and online address, whereas broadcast consumers are

Table 1. Transaction Structure in Network-based Service Models of Media Platforms (adapted from [26])

Media Service Model		Media Transaction Type							
Service Model	Commodity Examples	On-Demand	Peer-to-Peer	User Control	Provider Control	One-to-One	One-to-Many	Push	Pull
Broadcast	Cable TV Satellite TV				✓		✓	✓	
WebSite	NYTimes.com	✓		✓			✓		✓
Media CDN	YouTube Netflix Soundcloud iTunes	✓		✓		✓			✓
Social Media	Facebook LinkedIn	✓	✓	✓		✓			✓
Subscriber “Following”	Twitter Snapchat Instagram		✓		✓		✓	✓	
Media Recommendation	Spotify Tidal	✓			✓	✓		✓	
Subscriber “Channels”	YouTube	✓		✓		✓			✓

technically anonymous. To evoke in consumers a sense of personal relationship, many on-demand video producers adopt a style of presentation and video editing that is more personal and less formal than broadcast. The technical quality and stylistic differences between broadcast and online video are often attributed as differences of high budget compared to low budget production values. However, these differences may originate from the video producer's intent to connect with an audience using a personal style that is informal and conveys a sense of individuality, extending the peer-to-peer experience from the web interface to the video message. The personal style represents an economy of expression in terms of media technique and presentation. Expanding on this example, as media audiences are becoming familiar with request-based and peer-based access to media, and undertake exchanges of media, the personalization of media becomes an expected attribute of media genres. In this sense the new modes of media transaction based on advanced network services can redefine the roles of consumers and providers.

2.2. Personalization and New Media Credibility

Media credibility is essential to its transaction value. Credibility is a quality attributed by a consumer or a perceiver [29] ascribing to the validity of information sources and reliability of technical production pathway from producer to consumer. Personalization of online media transactions can lend credibility to online providers as trusted media sources. On-demand media transactions are often hosted in social media, and media providers as well as recipients tend to personalize these transactions. Personalization introduces new criteria that online viewers use to assess media credibility for entertainment value and for factual accuracy [30]. Online personalization implies a persona, which may function in online transactions similar to the broadcast role of a professional media personality. In contrast to broadcast presenting, online personalization may increase credibility by disavowing the formalities of large-scale broadcast infrastructure, rejecting signature corporate persona of broadcast presenters.

On-demand network services support three types of credibility enhancement: *transactional symmetry*, *access control*, and *participation as a media producer*. 1) Transactional symmetry: peer-to-peer transaction enhances personalization by representing 1:1 symmetry of sender and receiver, implying a bi-directional, individual mode of communication. A media consumer's assessment of the credibility of media resources is supported by peer-to-peer symmetry and bi-directionality [31]. 2) Access control: On-demand media enables a consumer to control media access, and the sense of control provides further

context to attribute credibility to acquired media resources. [32]. 3) Participation as a media producer: Participatory internet-based media transactions increase audience familiarity with media as an exchangeable resource. Credibility is enhanced where an on-demand media consumer is empowered to perform as a media provider. Media assets retrieved through on-demand and social media sources can be retransmitted as part of a new media message or new channel. When on-demand media resources are repurposed in new media transmissions, the peer-to-peer exchange provides a context for credibility of the media source. A consumer who re-sends a media resource underscores that item's credibility for themselves and for their social network of recipients.

2.3. Hybrid Transactions

New media genres based upon symmetrical transactions stand in contrast to broadcast media where transaction asymmetry is interpreted as a trusted source. Broadcast is a one-to-many paradigm where a provider limits access to media resources by controlling production, marketing and scheduling. The "gatekeeper" [30] generates credibility by establishing the asymmetry of studio production and unilateral control of distribution. High quality production values underscore the credibility of the broadcast source. However, the increasing popularity of on-demand media and the shift of audiences away from broadcast media indicate the possibility that for many consumers, media credibility generated by peer-to-peer network symmetry may be equivalent to media credibility generated by broadcast network asymmetry [33].

Table 1 shows how different models of media services integrate different transactions. Media services models are shown in relationship to media transactions types. Several models combine peer-to-peer and broadcast paradigms; their popularity indicates that media consumers recognize hybrid credibility [34]. Hybrid examples include push technology subscription services such as Twitter, Snapchat, and Instagram, and recommendation systems such as Spotify and Pandora. In a hybrid model a subscriber or "follower" requests access to a media service, while a provider maintains broadcast-style media access by scheduling transmissions. This hybrid credibility integrates symmetrical exchange of peer-to-peer and personalized transactions, with asymmetrical exchange of one-to-many services-style transactions. Section III presents a cross-platform authoring framework and architecture designed for hybrid combinations of network-based services transactions to develop new modes of media production and distribution.

III. A MEDIA FRAMEWORK

Media Framework (MF) is a research platform that provides a multimodal architecture to test and model future media genres that are dynamic and responsive with processing modules that are distributed yet highly integrated. MF is network-based and designed to facilitate asset management, remote collaboration and lightweight programmability specified by diverse workflow requirements. With native integration of distributed media resources and interactive display devices, MF is designed to develop dynamic content that is procedurally generated, anticipating the need for new workflows that integrate network-based media production and distribution. Use scenarios of this framework include interactive media creation and presentation, asset management and curating, authoring distributed multimedia, workflows for collaborative production, ambient media display, and interactive multimodal performance. The scope of system architecture encompasses cloud computing, social media integration, content delivery network and services (CDNS), distributed or dedicated applications, sensors and IoT networks, and interactive control and display devices. A recent trend in online multimedia fuses the use of cloud storage, cloud services, streaming, and social media with lightweight multimedia authoring capability. Many QoS problems such as described in [35] are now solved by cloud services. This landscape suggests that with minimal coding capacity users can be capable of multimedia authoring that extends from lightweight authoring to professional quality production. For this reason, the core developmental strategy of MF focuses on a communication layer to enable a set of distributed processes to perform as a coordinated medium rather than as an aggregate of remote transactions. Along with this strategy, most interface design problems in MF focus on increased system interactivity for authoring and auditioning. In the section below, we simplify the description of interactivity into seven dimensions from user-centered perspectives.

3.1 Seven Dimensions of Interactivity for Distributed Multimedia Authoring Systems

In [26], seven dimensions of system responsiveness requirements are identified for distributed multimedia authoring: 1) Interpersonal Interaction, 2) Human to Media Resource Interaction, 3) Resource Type Interaction, 4) Multiple Resource Channel Interaction, 5) Multimodal Interaction, 6) Curatorial Interaction, and 7) Collaborative Documentation Interaction. These are further elaborated as follows in order to draw out implications for system

implementation.

- 1) **Interpersonal Interaction** characterizes communications, both lateral (peer to peer) and hierarchical (project overseer with task delegation). At the baseline, communication tools such as internal live camera feed or external conferencing tools can facilitate interpersonal interaction. However, for collaborative authoring, this interaction goes beyond simple communications. Technical requirements include differentiation of interpersonal roles in collaboration, based on the project profile and participants' contributions. Interpersonal Interaction implicates system level definition for access control to read/write and document annotation, which are deeply related to the 7th requirement.
- 2) **Human to Media Resource (H2MR) Interaction** characterizes how authors and audiences access and treat or navigate media contents. This includes one-to-many relationships between a user and media resources, and many-to-many resource access among multiple users, creating an m to n to p relationship. H2MR Interaction also implicates system level definition for access control to media in a shared repository. In H2MR Interaction, the sense of immediacy and flow from query and access is most important to facilitate, which generates requirements for when and where storage, memory, processing and bandwidth are prioritized.
- 3) **Resource Type Interaction** characterizes how many media types are engaged concurrently and how they are accessed. Most traditional media genres use a single screen and a single frame for content within the screen space. Most genres limit how audiences expect the types of media that are concurrently accessed. Contemporary media consumers practice multi-frame and multi-screen engagement, and different screens bring different modes of interaction with multiple media types such as 2D, 3D, hyperlinked and dynamic interactive formats. System support for Resource Type Interaction can enable authoring to coordinate diverse media types including support to integrate media across several interactive devices. Multiple media types can be accessed by single query for combined interactive retrieval.
- 4) **Multiple Resource Channel Interaction** characterizes the number of concurrent resource channels—such as websites, repositories, or live signal applications—that are linked and feeding into the workflow, and how these are combined for interactive retrieval. Contemporary media consumers practice parallel access to multiple channels, for example opening multiple web pages and surfing

them, which provides ad hoc one-at-a-time interaction. System support for Multiple Resource Channel Interaction can manage multiple source channels directly feeding simultaneously to the production line. Authoring requirements can be used to optimize channel management to support synchronized interaction with multiple resources. This is significantly different from a web surfing paradigm.

- 5) **Multimodal Interaction** characterizes the user’s concurrent engagement with audio, visual, tactile, and ambient signals. Presently most devices dedicate concurrent interaction to a single media source and do not enable simultaneous interactions with multiple modality types. System support for concurrent interaction across multiple modality types will enable authoring for collective interaction.
- 6) **Curatorial Interaction** characterizes authoring requirements for social engagement in creative QA

such as team critique or instructors’ assessment or peer review, both synchronous and asynchronous. A simple form is the “Like” rating system of Facebook, while collaborator QA is well structured and relates to project specific resources. Curatorial Interaction implicates integration of semantic data as QA feedback into the authoring process.

- 7) **Collaborative Documentation Interaction** characterizes dynamic project documentation that integrates multiple authors in phases. Authors adopt roles defined in Interpersonal Interaction (requirement 1) and may participate in multiple phases of discourse. Structural requirements include support for timescale of Collaborative Documentation in phases such as pre- to post- production, distribution, audience responses, and contributions of new media in productions that involve multiple versions.

Table 2. Dependencies shared between Interaction and System Function: The Up Arrow indicates Interaction depends on Function. The Left Arrow indicates function depends on interaction. The shading Light to Dark indicates the level of dependency, lightest being minimum and darkest being maximum dependency, on a scale of three shades of grey.

Dependency Table ↑ Interaction depends on Function ← Function depends on Interaction Shading represents Dependency (Light to Dark = Minimum to Maximum) System Function type	Interaction type			
	Human to Media Resource Interaction	Resource Type Interaction	Multiple Resource Channel Interaction	Multimodal Interaction
Media Display Scheduling	↑	↑	←	↑
Scene Graph Traversal	↑	↑	←	↑
Semantic tagging	↑	↑		↑
Media Resources Access	↑	↑	←	↑
Media Resources Control	↑	↑		
Query Processing	↑	↑		↑
Resource Retrieval	↑	↑	↑	↑
Display Device Management	↑	↑		↑
Distributed Storage Access			↑	
Local Storage Access	↑	↑	↑	↑
Remote Process Virtualization	↑	↑	↑	
Concurrent Control Processing	↑	↑		↑
Workflow type				
Resource Navigation	←	←	←	←
Concurrent Interaction	←	←	←	←
Authoring	←	←	↑	↑
Multimodal Experience	←	←	←	←
Collaboration	←	←	←	←

Four dimensions of system responsiveness are selected from the above to illustrate mutual dependencies of experience design and system function. Table 2 shows a predominant dependency of interaction type upon system function type, but in Workflows the system functions exhibit dependency on interactions. The degree of dependency ranges from Requirement (essential for functionality) to Optimization (valuable for responsiveness but not required for baseline functionality).

3.2 Media Framework Architecture

Networks of media resources are implemented using networks of media devices and services to generate interactive display experiences. Interoperability of networked devices and services generates interactivity in the seven classes discussed in section III.3.1. To support integrated authoring across media types in multiple combinations, Media Framework provides a research platform for prototyping configurations of networked interoperability, and for authoring interactive content adopting a bespoke network configuration as an integrated

multimedia platform. In the current MF implementation web browsers are adapted for production interfaces as thin clients, in ways that, in spite of the lightweight front-end application, the interface provides access to high quality media resources. For real-time, a proxy technique allows rapid manipulation of media for project prototyping, which can be displayed with high resolution in a completed media presentation. Media resources comprise heterogeneous media types, by which we mean, the media signals are channeled through and outsourced from multiple types of services, software and hardware, and devices and sensors. They are managed with uniform representation for authoring and previewing through interactive control and display. In this context, semantic computing is applied to authoring, which enables a dynamic authoring paradigm that replaces traditional nonlinear editing [36]. The leading design strategy is a common representation of a dynamic project space with a single workspace supporting heterogeneous media resource selection and display, agnostic of project platforms or media types or channels that handle concurrent workflow.

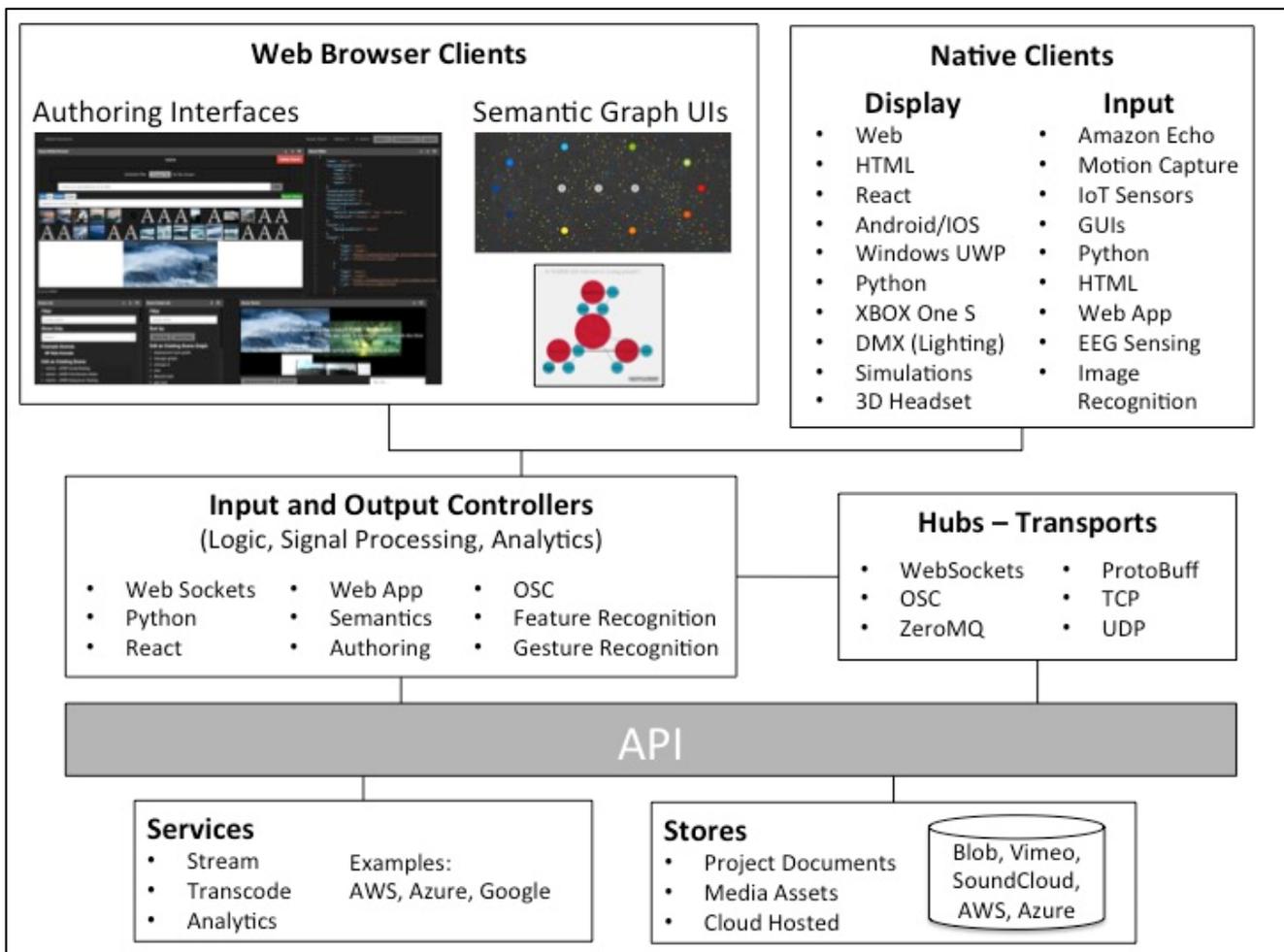


Fig. 1. Media Framework System Architecture

MF architecture for prototyping personalized media networks is illustrated in Figure 1. First designed and tested with web interfaces and browser-based media displays, the system extends to support native interfaces and IoT devices for data acquisition applied to media control, and for dedicated systems actuation such as lighting control using the DMX protocol. The API provides functionality through Hubs, Stores and Services, both local or cloud based. A Hub is a transport layer that implements client-agnostic logic for baseline MF functionality. Hub support includes web sockets with authentication, UDP and TCP as well as communication with dedicated protocols such as game engine controls. Controllers enable lightweight clients by implementing heavier processing in function classes that support multiple clients of similar type. Controllers can maximize node interoperability by providing dedicated code to support specific behaviors required by specialized classes of devices. The API defines clients as Input Controllers that accept commands or messages, and Output Controllers that send media data streams or control data streams to procedural media display clients. Controller implementations in Python and React are supported. Control and display transmissions are asynchronous unless dedicated synchronous clients are implemented. Commercial web services include Azure, AWS, Vimeo and SoundCloud. IoT implementations include Amazon Echo, motion capture, EEG capture, and Arduino-based sensors for touch, sound, light and acceleration. The MF Hub logic applies a common authoring and scheduling framework to data and media transmitted among all nodes.

3.2.1. Relationship to W3C Multimodal Recommendations

The WWW Consortium Multimodal Interfaces Working Group (MMI) [37] developed recommendations for Multimodal Architecture and Interfaces [38]. MMI identifies requirements relevant to MF, which include multimodal display and interaction, components distribution, device diversity, normative use of network and cloud services, and asynchronous event-driven run time operation. The MF architecture is compliant with MMI recommendations and exhibits a number of structures parallel to the MMI architecture. All MMI layers are functionally present in MF, with differences in layer design. MMI locates all endpoint devices in a Presentation layer, whereas MF uses a Client layer for endpoint devices. MMI centralizes control of endpoint devices by Interaction Managers located in the Presentation layer, whereas MF introduces a Controller for each client. MF Controllers link clients to the API and also provide enhanced functional logic and dedicated MF client support. MF Controllers fulfill functions of MMI

Interaction Managers to execute run-time communications with Transports and Sessions, also intake and route stored data as needed. As in Figure 1, this is because the Controllers directly communicate with Hubs-Transports. MF Controllers and Hubs-Transports share sessions analogous to the MMI Session layer. While maintaining clients as lightweight as possible, the use of client Controllers enables normalized functional programming for many different kinds of clients.

Overall MMI recommendations focus on multimodal control of third party devices but are not designed for producing media content. Whereas, MF is dedicated to author, display and control media content. MF Authoring is hallmark capacity beyond MMI scope, a detailed multi-function authoring workflow applied across all device configurations. MF authoring defines deep structured media content represented as persistent multimedia information independent of technology configurations (see section I). MF distinguishes Authoring run-time as an interpreted workflow using special Controller and Client components that are not used at Display run-time, whereas MMI markup is rudimentary authoring not run-time enabled, using XML in a Web framework. MF utilizes browsers for convenience and ubiquity but does not use web markup as a core data type, and can operate without browsers.

IV. SOFTWARE DESIGN METHODOLOGIES FOR AUTHORIZING PERSONALIZED MEDIA NETWORKS

The software design for the MF platform utilizes several methodologies: 1) metadata is used to curate semantic association: this is described below, 2) authoring structure is used for vertical integration of semantic scope with encapsulation, discussed in 4.1, and 3) configuration prototyping is used for multiple modalities of interaction with case driven projects: this is discussed in 4.2 and 4.3.

The MF platform uses metadata to develop and apply personalized media networks. The metadata carries semantic functions to describes qualities and contents of a media resource. Semantic associations from one media resource to another create a media resource network. Associations may be based upon many types of relationship that provide narrative context. A media network is generated during an authoring process and is presented as interactive media during a display process. The authoring process includes structuring semantics, and the selection of display procedures, display priorities and interactive display attributes. Most display attributes apply across multiple media types while some attributes are

determined by media type. The display process generates dynamic media sequences by traversing the media network through semantic associations, while applying display procedures to the associated media. Interaction data from users provides context and temporal dynamics that influence the automated resource selection and display timing.

Multimodal media can be managed in a single semantic network and displayed across multiple devices in coordination of a user's actions. Semantic data supports multimodal authoring by providing descriptions that are agnostic to media type. Semantic data structures are developed as part of the authoring process and this process is a required step in the creation of personalized media networks. The flexibility of semantic data enables rapid organization during authoring, but semantic data presents challenges for maintaining structured vocabulary across media resources and among multiple users. A well-structured semantic framework is essential for developing and sharing personalized media networks, but shared vocabulary is difficult to develop and maintain, and burdensome to involve as part of an authoring process. To support the creation of semantic structure during a rapid authoring cycle, MF encapsulates the scope of semantic data associated with media resources. By limiting scope we reduce the need to consult a master structured vocabulary, and we increase the ease of use for applying semantic data in intuitive ways.

4.1 Scenes, Semantic Scope and Media Network Structure

We define a *Scene* as set of media resources with semantic data and display rules. A scene limits the scope for binding semantic data and spatiotemporal attributes to a set of media resources. A scene contains a set of media resources and associated *tags* that assign semantic metadata to each resource. Tags are resident with the scene and are not embedded as attributes of media resource files. A media resource does not automatically carry tags from one scene to another. Within a scene a *theme* defines a set of associated tags, and controls the aggregate display of all media resources that share those tags. A theme is scope limited to the scene where it is declared and will be interpreted as a separate entity if declared in other scenes. A scene also defines a set of display procedures that apply to all resources in the scene, independent of themes. Each media type in the scene interprets the scene's display procedures appropriate to its type; many display attributes apply across all media types.

A scene encapsulates a semantic vocabulary limiting consistency requirements within scene-level authoring.

Above the scene level, a *world* graph is defined as a network of semantic associations and path planning across multiple scenes. Two classes of semantic connections assert media network structure at world level. A *scene group* designates a higher-level semantic unit for layering multiple scenes. A *trope* designates a set of themes drawn from multiple scenes to form semantic associations between scenes. Metadata for scene groups and tropes maintains semantic scope at world level, independent of semantic scope encapsulated at scene level. Users collaborate in naming scene groups and tropes to connect their scenes making a world graph. This introduces a method of shared authoring for interactive social media based on personalized media networks.

Media resources are displayed using the procedural methods of the scenes where they are declared. When themes and scenes are accessed at world level they maintain procedural integrity inherited from scene level. From world level multiple scenes may be concurrently displayed to create media layers. The logic for concurrent scene layers and scene transitions is defined above world level in a *score*, which declares global temporal structure and logical priority for scene and trope processing. The scene, world graph and score collectively ensure state representation of all media events.

4.2 Long Form and Short Form Media

Traversing a media network generates a flow of media resources as a series of automated real-time edits. Media resources can be presented as layers in a mixed and distributed presentation space, including concurrent sound tracks or visual frames coordinated on multiple devices. The authoring process determines layout and framing controls for single and multiple sound tracks, audio-visual layers, still and moving images, and 2D and 3D display devices, all configured in the score (see section IV.4.1). Formal structure of resulting interactive media is not defined by a fixed timeline; formal structure is generated by one or more display sequences defining procedural traversal of a media network. To describe media form we account for the extent of the media network in terms of ratio and distribution of connections to resources and to display modalities. Procedural constraints for interaction contribute to the formal structure by defining paths to traverse semantic links.

In practice we designed a project-based learning application, Global Digital City (GDC), generating a network of media resources created by over 60 students from Asia, Europe and North America. The students produced an extensive series of micro-documentaries exploring 10 cities: Seoul, Kuala Lumpur, Beijing, Hong

Kong, Dalian, Chengdu, Shenyang, Panjin, Chicago, and Manchester. The students then collaborated to connect their cities using networks of scene groups and tropes, which required students to identify and name shared thematic associations. The resulting media network is very large and currently contains 274 scenes, 589 themes, 1625 text objects, 3246 images, 605 videos and 288 audio objects, requiring about 50GB of Azure storage. The scenes adopt non-uniform procedural styles for presentation. The resulting GDC is a long form media work, defined partly by the extent of material and number of media network connections, and partly by the even distribution of content and wide range of presentation styles across the extent of the network. GDC succeeded in creating a large-scale multi-point navigation interface for nonlinear content (section V). Highly personalized scenes can be retrieved during navigation at a global level, as well as retrieving scenes that share commonplace themes.

As we worked with the extensive structure and wished to move beyond an initial exploration mode, we noted that a world graph structure of many nodes does not present clear entry and exit points to define narrative sequences. Interpreting the GDC work requires extensive exploration, as there are no introductory or summative sections. While the GDC network generates a long form media presentation, the formal structure of the network is an aggregate of small units—each city comprising a set of independent micro-documentaries. The form becomes too long for a single interactive session. This excess is enhanced by flatness of hierarchy in the world graph, and the resulting even distribution of semantic associations and multiple procedural display styles. From this analysis we identified requirements to differentiate techniques for short form, and to provide methods for embedding short form within long form. In terms of broader applications, the GDC network implementation yields insights toward new interface and navigation forms for multimedia blogging and tweeting.

A following project, prepared for the UK's Nations and Regions Media conference (NARM) [39] tested the development of short form. Here university students created media to illustrate their preferred modes of media access, discussing where, when and why they consume media. Students' preferences draw sharp contrasts to legacy broadcast models and traditional commissioning models of the BBC. As a design principle for the NARM production, constraints were applied in the planning stage to limit the range of media resources and number of semantic entities. These constraints focused students' development on the creation of a world graph based on five tropes representing their key questions and positions. Media resources created by students include data graphics,

video interviews with other students, "headline" texts, and views of students' preferred modes of media consuming. By curating assets and generating semantic associations selectively, and by using a limited range of procedural styles for presentation, the NARM project presents short form semantic pathways with clear points of departure and arrival. The result is that students pose questions and float teasers that take the place of in-depth answers, captured in students' anecdotal musing and informal statements discussing their viewing habits and program preferences. In terms of interactive experience, NARM encourages short form browsing whereas GDC demands long form "binge" viewing to grasp its scope. The NARM project demonstrates an approach to short form, and opens up the challenge to embed short form in long form. In terms of broader applications, the NARM implementation yields promising insights for future backpack journalism [40].

4.3 Ambient Form

A third project demonstrates a prototype of interactive ambient media. Ambient media [41] is displayed as a design element of a built environment to provide information observed peripherally or in passing. A requirement for ambient design is to provide information in media that will not receive an observer's dedicated attention. Short Form and Long Form define media networks to be traversed by users' attentive exploration. In contrast, Ambient Form defines media networks to be traversed by users' actions performed for other purposes. Ambient media interaction paradigms are designed to respond to unencumbered actions when observers are not focused on controlling media.

For prototyping ambient media we worked with professional theatre performers who are trained in physical movement performance. This process was informed by extensive prior work on human-machine performance systems [42-45]. While previous works used self-contained sensors and actuators, in this work IoT sensing brings a new chapter to multimodal and multimedia information processing. IoT devices are configured to capture data of performers' actions. Performers adopt a persona and emulate pedestrian and casual movements that represent task-driven actions, physical limitations and emotional states. Movements of trained physical performers are precisely controlled and repeatable for experimental purposes. Some performed actions are sensor-specific and other actions are unencumbered and naturalistic. IoT devices capture and analyze this data to identify salient features. A performance media network is implemented with MF to respond to data generated by IoT devices. Relevant movement data is applied to control

traversal of the media network. Media is displayed on multiple devices situated in a purpose-built environment. Media modalities include interactive lighting control in the environment, display of thematic images, sounds, and videos, and display of live video streams showing the performers.

Initial results from the media performance project indicate several encouraging lines of further research. Performers are interested in the application of media performance networks and IoT devices in professional creative productions, as they gain additional levels of control and the system provides a kind of prosthetic extension of their actions. Beyond professional theatre applications, there is value in prototyping and assessment with professional performers' abilities to control their physical movements with a high level of reproducibility, and to emulate physical and emotional conditions of different personas. Initial results suggest a model for early testing and prototyping of unencumbered sensor-driven responsive media applications, to support IoT embedded in a variety of domestic and professional settings.

Comparing Ambient Form with Short Form and Long Form applications, we modified our initial assumptions about the differences between these applications. We identify more system requirements in common than first expected. The elements that most distinguish these forms are identified in user experience, specifically in the perceived responsiveness of modality, comparing the active engagement case, such as exploration and discovery, to the ambient case, such as peripheral engagement with environmental cues. In terms of broader applications, the ambient media implementation yields insights toward promising applications in recent trends of urban built environment, including dedicated interior environments such as care homes and similar community spaces [46].

V. INTERFACE PARADIGMS FOR AUTHORING AND DISCOVERY

An authoring interface and a discovery interface are two paradigms that support the process of curating and personalizing a media network [47]. Authoring is a process to formally encode a media network to generate interactive experiences. The encoding configures semantic associations and binds these with procedural dynamics for interactive display of the media resources. Through this encoding, a personalized media network is defined as a set of virtual relationships that are realized dynamically during an interactive display process. Interactive display is exercised during the creation of the network to test semantic relationships, multimodal composition and display dynamics.

A media network is structured by semantic relationships that are independent of particular display devices. However, multimodal display requirements are specified in order to render a media network as an interactive experience. For this reason the process of creating a personalized media network is informed by designating a physical network of display devices and interactive devices. Although a media network may be developed independently from a device configuration, a physical configuration is required for testing the network in terms of interactive performance and user experience.

The process of creating a personalized media network may be understood in terms of bottom-up assembly, beginning with a selection of media resources that are ingested into the MF asset store. A scene is initialized and media resources are selected from the asset store and assigned to the scene. In the scene, semantic tags are asserted for each resource and semantic associations are defined between resources. Also in the scene media display procedures are defined, providing constraints and priorities to generate spatiotemporal order and dynamics for interactive media presentation. As multiple scenes are completed, the media network can be defined with world level associations of scenes, themes, resources and display dynamics. The authoring workflow requires dynamic auditioning of authoring choices, to evaluate the binding of contents, end-user actions, and procedural display. As a media network is developed, interactive interfaces are required that can transmit data to MF.

Bottom-up authoring from media resources to themes and scenes is an intuitive approach, however it does not generate large-scale form. The bottom-up network creation process may be guided by planned formal structures to define constraints on selections made at several levels of authoring. Constraints can be applied to groups of scenes to enforce large-scale relationships among smaller entities such as themes and scenes. In practice it is structurally efficient at the scene level to adopt a selection of tested templates that demonstrate a desired performance of interaction modalities. The interaction design recommendation "Edit, Don't Create" [48] stresses the importance of building media contents by modifying existing templates rather than creating from scratch. Templates also aid authors' understanding of how scenes bind two levels of content creation: the media tagging and the procedural display. We have observed inexperienced authors have a tendency to focus on semantic interpretation of individual resources, and need encouragement to consider the aggregate impact of media resources in an interactive flow. Templates provide a gentle learning curve for beginners to structure media contents with procedural or processing imagination.

Templates at scene level aid beginners' considerations of large-scale formal structure that are realized at world level and score level, where higher-level templates may be applied.

5.1 An Authoring Interface for Media Networks

To focus authoring at the scene level the primary MF authoring interface provides four scene-level windows: 1) browsing existing scenes in code form, 2) viewing scenes as interactive media display, 3) browsing and previewing all media assets ingested in MF available to include in a scene, and 4) a scene editing window that provides developer-level JSON coding tools such as syntax checking for creating semantic and procedural bindings in a scene. These interfaces are currently presented in a web browser window as moveable and scalable frames. Fig. 2 presents a screen capture showing the media icons in the asset browser, and JSON code in the scene editor. Ingested media of every type is given an icon representation with semantic tags displayed below the asset browser window. Media resources of any type can share a tag. To search and sort media resources tags can be

selected and combined in logical expressions. In Fig. 2 the text entry field across top center accepts text strings to be displayed as media resources in a scene, and accepts URLs to ingest cloud-based media assets. Scene viewer and scene browser windows are minimized at the bottom edge of the figure.

Fig. 2 illustrates how coproduction is deeply structured in the MF interface and system components in a workflow. A semantic vocabulary becomes a code-based production vocabulary. Tags are not merely descriptive annotations; tags are a data type required to code the connectivity of a media network. Team discourse is rigorously exercised through the formalized activity of resource tagging and theme construction where some common understanding is agreed and used across the production. In the GDC production the tagging process developed as intercultural dialogue; students had to work with team members from other nations, and exercised articulation, abstraction, multiple perspective orientation, and most importantly negotiation and reconciliation in order to arrive at a consensus of tags and themes for structuring associations. The result is an associative network of the material senses, which are the expressions of media resources, and texts; in

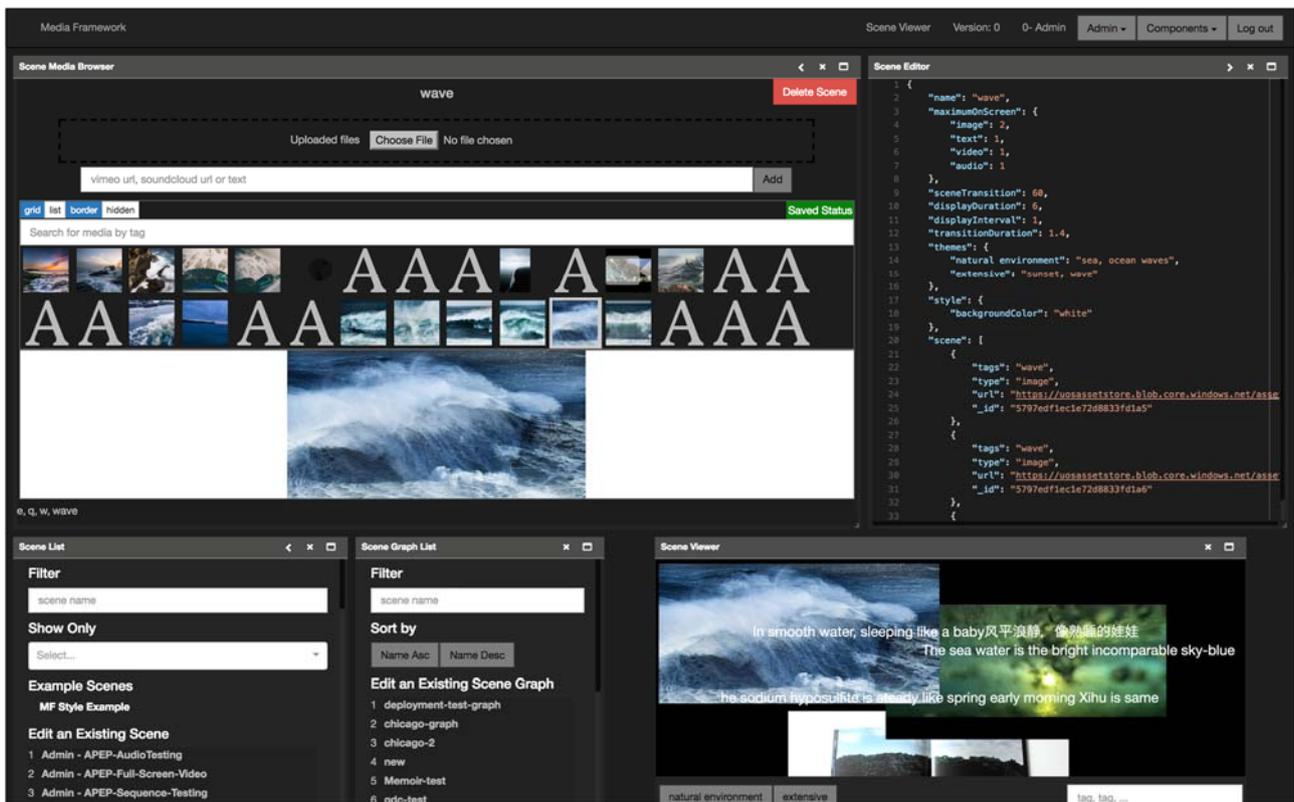


Fig. 2: Media Framework multi-frame Authoring Interface. Frame functions clockwise from Top Left: Media Resource Browser, Top Right: Scene Editor, Lower Right: Scene Viewer, Lower Left Scene Graph List with Search Functions, Scene List with Search Functions. Icons in the Resource Browser indicate video, sound, and text assets; thumbnails are still images. Below the icon tray a preview window displays a selected image, video or sound. The Scene Viewer responds to a Scene Authoring interpreter as a function of Authoring run-time and generates layered media display on the fly; this example includes English and Chinese texts.

turn the network is an expression of a structure of signifying rules for the media resources. In short, the GDC production process cultivates young people with media literacy authentic to their time, and with coding skills through collaborative multimedia authoring methodology for conducting intercultural discourse.

5.2 Discovery Interface for Semantic Networks

A direct semantic connection between two media resources is asserted by assigning a common tag to both resources. A common tag represents a similarity—the two resources depict similar content. An indirect semantic connection may be identified between two media resources that do not share tags in common but each shares a tag with a third resource. The connecting resource depicts content similar to each of the other resources that do not share a direct connection. Figure 3 illustrates these relationships. A set of resources that share both direct and indirect connections can be represented as a graph with resources as nodes and shared tags as edges. Each resource in this graph will have multiple direct connections and indirect connections to other resources through degrees of separation. A semantic graph could be dense where resources have multiple tags. For a given set of resources, the selection of tags and how they are applied will determine whether the semantic graph is dense or sparse, and will reflect the refinement of “similarity” in terms of level of detail, which is difficult to maintain at a uniform level across all tags. Grouping resources meaningfully requires control over the level of detail of tags to maintain relevance of their use. Too few or too many connections reduce relevance of tags.

Similarity as a classification is well organized in scientific practice, applied through expert consensus as a foundation for empirical observation. Similarity is often applied as a classification in media practice to organize assets for storage and for marketing – see for example the naming conventions used by commercial music licensing services at BeatPick [49]. The names are far removed from musical terminology. Consumer entertainment recommendation systems are based upon similarity classifications that are generated by human interpretation supplemented by computational analysis. As introduced in Section 1 above, the brute force implementation of most large-scale similarity analysis systems indicates the shortcomings in automated analysis. We hypothesize these shortcomings are due in part to the use of classification by similarity. MF is designed to investigate alternative semantic metadata beyond similarity. Another approach is described below to create a structural perspective for authoring.

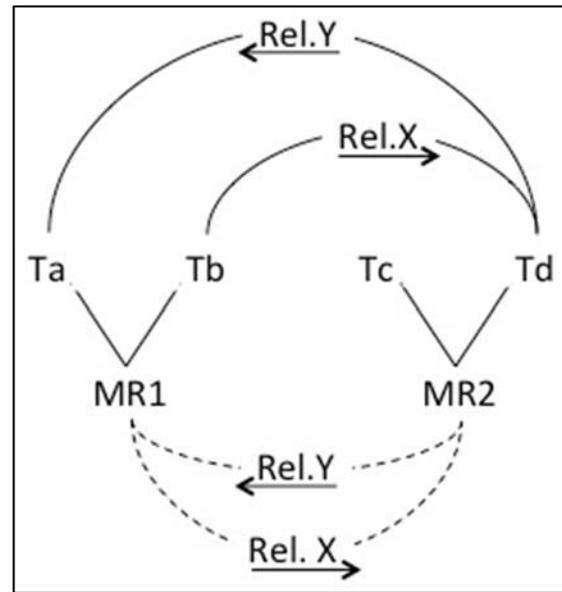


Figure 4. Media Resources MR1 and MR2 do not share tags in common, but they are connected by relationships between their tags. Relationship Rel.X asserts an association from tag Tb to Td, and Rel.Y asserts an association from tag Td to Ta. Therefore MR1 and MR2 are valid members of relationships Rel.X and Rel.Y by association through tags Ta, Tb and Td.

5.2.1. Predicate Associations

Similarity is not the only type of relationship relevant for generating an interactive media experience, as it does not

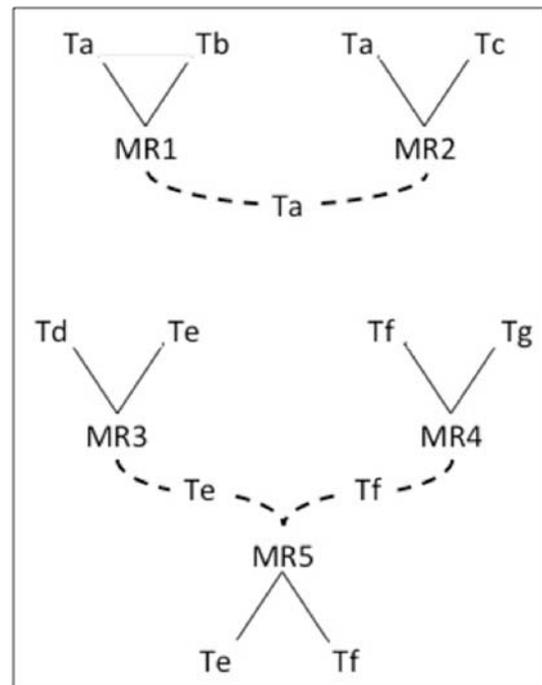


Fig. 3. Sharing tag Ta directly connects Media Resources MR1 and MR2 by similarity. MR3 and MR4 are not directly connected, as they share no tags in common. They are indirectly connected as each share a tag with MR5.

account for many interesting relationships that may exist or may be desired. Media communications use structures that exhibit semantic complementarity, associations that support contexts and logics. Contexts include place, character, social group, and action; logics include causality and rationale. Semantic metadata can reflect contexts and logics by associations other than similarity; two tags that describe different content may be connected to each other by a concept that describes a relationship. The connection forms a predicate: noun-verb-noun, represented in RDF data format [50]: tag *A* exhibits relationship *R* to tag *B*. Applied to media, the relationship between two tags is conveyed to the media resources associated with each tag. Fig. 4 illustrates these relationships. This approach can generate structure for presenting media based on contexts and logics other than classification.

A set of tags with direct connections by semantic association can be represented as a graph with tags as nodes and predicate relationships as edges. If tags are associated only by similarity the graph will be a tree. However if tags have predicate associations the graph can be non-hierarchical and acyclic. Predicate structure is more diverse than the similarity graph of media resources described in section V.5.2 above. A graph of tags and predicate edges can represent similarity relationships and at the same time represent contexts beyond similarity to bring sets of resources into multiple associations.

5.2.2. Graph Visualization as Discovery interface

In the MF media network structure, themes and tropes support predicate relationships. Tagging tends to focus on individual media resources, whereas authoring requires attention to relationships between resources and the larger semantic structures that generate these. MF provides tools to build semantic graphs of tags, themes and tropes. MF visualizes the resulting structure as a graph that functions as an interface for exploring associations among media resources. We refer to the interactive graph as a Discovery interface: a user may select a node on the graph to display its associated media resources. When a node is selected, temporal dynamics of media display are determined by procedures declared in scenes.

Figure 5 is the Discovery interface for GDC. In the center are three central tropes: City, Movement, and People. Surrounding these a circle of Scene Groups nodes represent the 10 cities. Smaller nodes represent themes of individual scenes and tropes that connect cross-city themes. A user can random-access any of the nodes and when a node is selected the interface highlights its connections to nearest neighbors, a feature intended to enable exploration of graph neighborhoods.

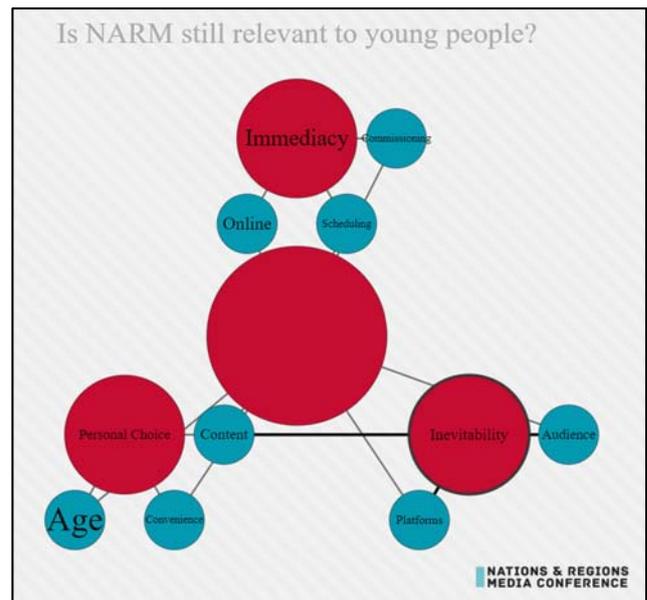


Fig. 6. Short Form Discovery Interface for NARM project

5.2.3. Assessing the Discovery Interface

The extent and complexity of the graph reflects the Long Form of GDC as interactive media. As discussed above the Long Form as visualized in the Discovery interface does not obviate the exploration of subsections such as providing paths from start node to end node. These conditions have been assessed with informal user tests of the Discovery interface during the GDC workshop. 39 subjects who had been working only with the MF authoring interface were introduced to several versions of the Discovery interface design, and were presented with 16 timed tasks for interface use. Each task required users to identify node names and relationships, after which users were asked to rank their preference between two versions of graph design. Five design versions of the Discovery interface were compared, each with varying levels of interactive visual feedback. Interaction variables that were assessed separately include: providing visual feedback for mouse-over, single click, or double-click functions. Visual

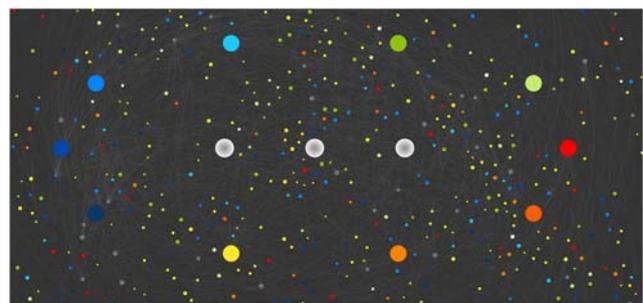


Fig. 5. Long Form Discovery Interface for GDC project: a global map by data visualization from user-generated semantic networks. Personalized semantic networks about individual cities were integrated by users' negotiation of shared semantic data.

cues that were assessed separately include: highlighting a node only when selected, highlighting nodes that connect to a selected node, displaying text strings with names of nodes, and animating the graph to change its node layout by arranging first-order connections in a circle around the selected node. The graph animation option presents an alternative priority to the user's orientation, by disrupting the global layout of the graph in favor of simplifying the local layout around a selection. Assessment responses indicate a strong preference for visual cues that represent semantic context supported by local connectivity, and a preference for connectivity to be visualized dynamically for regions local to an interactive exploration, rather than constantly displayed. Global orientation was considered less relevant and the density of global information was not considered a high priority as an interface display function.

Strong preference for local context encouraged further work on the development of Short Form media. This is reflected in the Discovery interface for the NARM project, illustrated in Figure 6. The Short Form structure of the project is reflected in the clarity of the interface. A root node connects to eight Scene Groups: Online, Commissioning, Scheduling, Content, Convenience, Age, Platforms and Audience. Themes from these scenes are gathered into three tropes: Immediacy, Personal Choice and Inevitability. The tropes provide point of view orientation for exploration and discussion. The GUI layout is based on graph connectivity of direct and indirect semantic associations. Similarity hierarchy is not used.

VI. SUMMARY AND DISCUSSION

Media Framework's architectural design addresses integrated functionalities in one workflow while aiming to facilitate the seven dimensions of system responsiveness described in Section 3.1. In addition, MF's iterative system development is driven by seven meta-directives:

- **Structured Tagging:** MF employs structured tagging as a major collaborative production requirement for sharing, reuse, and repurposing media resources. Tagging facilitates all related interface functions.
- **Multimedia Sketchpad:** MF facilitates end-to-end prototyping for multimedia production in ways to function as a *multimedia sketchpad*. Iterative refinement applies alongside iterative prototyping so that no prototypes are throwaway. This includes lightweight asset management and MAM strategies.
- **Multimodal Interaction:** For both authoring and auditioning, MF facilitates multimodal interaction. The system does not differentiate between author's auditioning and audience viewing in terms of the interactive experience other than differences of

read/write access control.

- **Cloud Based Workflow:** MF facilitates online authoring capacity by connecting and utilizing networked services, data centers, and accessibility from multiple sites, which enables collaborative, distributed, and cloud based production.
- **Semantic Vertical Integration:** MF facilitates a vertical integration of semantic metadata from auditioning individuals' UGC to collective UGC for collaborative authoring when a project requires.
- **Cloud Based Connectivity:** MF facilitates future applications on IoT enabled remote performances on and with ICT devices.
- **Procedural Dissemination:** Ultimately UGC from MF is disseminated and dynamically generated through networked services, and experienced on distributed devices, while it can be locally contained and tested.

The first five directives are well under development in the current progress in MF while the last two directives are partially tested. The meta-directives above are designed to accelerate gap reduction between authoring practice and information systems so that the two practices can share more sustainable workflows.

To evaluate design hypotheses and implementations of architecture, MF was deployed into project based learning schemes that enabled ethnographic observations with iterative assessment. User tests of the MF interfaces have enabled elicitation of constraints and requirements for multimodal authoring practices, aimed to support performance of multimedia information systems that are both interactive and distributed. To facilitate users' perception of immediacy and information relevance, two classes of system responsiveness are required: action responsiveness and information responsiveness. Accordingly, to enable and author quality user experiences, MF presently supports two classes of signal processing: spatiotemporal organization of media display, and semantic organization of media content.

The MF capacity to use distributed network services for information processing enables authored media push onto multiple devices, and also facilitates second and third screen experiences across distributed devices. The architecture adopts several network models: transactions with networks of distributed services, networks of devices for coordinated multimodal display, and semantic networks of media resources for generating multimodal content. The term "media network" may refer to any of these models, as the architecture adopts extensible context for constructing media networks.

The four factors discussed in Section II can be abstracted as a "four factors theory," a hypothesis that four factors

together determine sustainability of multimedia authoring practice. They also indicate the need to tighten the relationship between an authoring system and its underpinning multimedia information system. Information or data as well as interface capacity will impact the capacity for authoring multimodal experiences as well as ensuring users' experiences. The factors are restated as follows:

1. **Platform as Technology Configuration** defines emerging technology practice in a way that can help predict emerging media genres.
2. **New use cases** create concept shift in system function definitions, and new practices lead to users' further expectations of multimodal responsiveness.
3. **Media genres** emerge as an exchange value that leverages and promotes capacity for multimodal responsiveness. Genres no longer appear as immutable; they are emergent.
4. **Transactional structure** of network services anticipates a platform's capacity for hosting new media genres, and hybrid transaction models are a source of prototyping new genres. Transactional structure always shifts within the limited combinations of models and types as shown in Table 1, but its connectivity shifts infinitely in an evolving network landscape.

Cognizance of these four factors helps identify and envision long-term goals for MF. These four factors inform MF's meta-directives for planning strategic implementations and iterative development.

Platforms are virtual and accessed by endpoint devices. Current consumer devices support users' participation in entry-level multimedia authoring, mostly in the form of UGC for sharing through social media. To an increasing capacity, social media is where users adopt the dual roles of content consumer and content provider. Users' alternating roles contribute to development of new media genres. Media Framework anticipates emerging platforms that redefine users' role by enabling them to access integrated functionalities for media content production, distribution and display. We note the challenges to tackle interactivity. Both authoring as an interactive experience and authoring interactivity to be experienced by media consumers further accelerate functional convergence between authoring and information processing.

Deep structure of media content has historically been identified through content analysis using methods such as semiotics [51], structuralism [52] and deconstruction [53]. Whereas, media signals and signal processing are viewed as encoding that is neutral of media content, based on the perspectives originated from Shannon's information theory [54]. The Media Framework architecture brings

into view a shift in the relationship between media signal processing and encoding for transmission, and media content production and consumer experience. Structuring content for procedural delivery mirrors instruction sets for multimedia information processing. Consider content as having deep structure of meaning. As we represent media content using semantic data for information processing, we open up representation of content structured in the forms that may be aligned with signal pathways for multimedia information processing. Using semantic data to organize media for interactive display, drawing upon distributed media services and information processing, indicates a direction for a more agile application of AI models to produce interactive multimodal experiences. Semantic networks provide a capacity for media resource modeling that can combine content requirements with engineering requirements in workflows that are structurally integrated, physically distributed, and temporally concurrent.

The Global Digital City project demonstrates an early realization of this capacity. The GDC creative process produced layers of semantic structure connected to interactive display functions for multimodal authoring. The authoring workflow begins with low-level tags applied to individual resources, and continues to create scenes that bind semantic associations and spatiotemporal display dynamics. Each scene's semantics are personalized by an author's creative perspective. Nested semantic scope enables vertical integration of scenes and associative integration of themes to create a global system network, a world graph. In GDC this integration involved 60 students co-creating semantic associations to connect their personalized scenes by practicing peer-to-peer negotiation. This process demonstrates a social media community model of production workflow. The Discovery Interface visualizes the structure of the resulting world graph that objectifies the collective media network. Interactive graph traversal enables navigation through the negotiated and objectified semantics to dynamically generate media in display streams.

Collectively the GDC promotes a cycle of personalized and socialized production generating media structure in encapsulated layers with vertical semantic integration. The semantic authoring process and production of a collaborative global graph and Discovery interface are bellwethers of future media genres and future platforms. The GDC global graph is a prototype representation of deep content structure that can also be interpreted as information structure for signal processing pathways and social structure of personalization and content co-production. Introducing semantic data implies the future use of semantic computing embedded in media production

workflows, such as the application of RDF data representations, structured vocabulary such as ontology [55], and computational inference such as semantic reasoning [56]. Industry production tools show indications of this direction for managing assets [57, 58] and for assisting video signal processing [59].

Semantic support in the interface is a necessary step for supporting diversity of individual media authoring with personalized semantics while also enabling higher-level coherence of collective media transactions and negotiated vocabulary. The MF authoring workflow is designed to capture the semantics of naming in the interface with the intent to embed individual perspectives as persistent properties of media networks. The GDC long form project represents this approach applied to preserve cultural diversity among an international community. Access to personalized metadata for context-aware discovery of media provides an alternative to recommendation systems that tend to homogenize and simplify content classification around similarity rather than diversity.

We foresee two paths forward toward adoption of semantic authoring workflows. One path facilitates personalized authoring of shareable media networks with greater ease of use for acquiring and applying metadata and generating higher-level associative structure, including social media collaboration. This path requires more flexible interfaces supported with personalized context-sensitive AI. Another path applies automation for initial media tagging based upon established vocabularies and templates of higher-level media network structures. This path depends upon stakeholders developing templates that encourage community breadth by simplifying the initial level of personalization. The degree of simplification to increase usability may determine the limit of expressive vocabulary. Scalable industry templates may be inversely proportional to media network personalization, indicating a practical limit to commercial support for level of detail. A number of social media platforms encourage participants to form subgroups based on similarity of profiles and interests. These platforms could also provide instruments for sharing or co-creating semantic metadata, aspiring to levels of detail beyond Facebook's "like" tag. Formations of social media groups indicate that templates are practical for scalability and ease of use; scalability of templates may also represent a slippery slope to homogeneity of semantic expression.

Addressing two questions posed in section I.1.3: 1) *Is it feasible to implement an authoring system to transform creative practice to be more collaborative, subjected to and relevant to information processing without compromising individuals' creativity?* Current MF project results show that collaborative authoring systems can be

implemented in ways that enhance personalization and do not compromise individual creativity. Further development requires computational support for shared control of media content generation in a collaborative workflow. 2) *Is it feasible to harvest semantics using a process that is native to the multimedia information and authoring system in one framework?* Current MF project results show that semantic data facilitates collaborative authoring, and enables efficient workflows once the data is entered in the authoring system. The process to develop semantic data is not streamlined. Future integration of semantic capture with WYSIWYG interface paradigms will enable more intuitive capture for scalable application of semantic data in media authoring.

As a concluding statement, the outcome of integrating a multimedia authoring and information system in MF is essentially a paradigmatic shift of the concept of UGC as personalized media network, which is dynamical and evolvable.

Acknowledgement

The Authors wish to thank Allan Walker for international partnership initiatives; Jeff Meyers, Allen Fairchild, Aaron Phillips, and Angel Petrov for technology support.

REFERENCES

- [1] E. England and A. Finney, *Managing Multimedia: Project Management for Interactive Media with CD-ROM*. Boston, MA: Addison-Wesley Longman, Inc., 1999.
- [2] A. Lenhart, "Teens, Smart Phones, and Texting." Pew Research Center, Mar. 19 2012. <http://www.pewinternet.org/2012/03/19/teens-smartphones-texting/>.
- [3] R. Belk, "Sharing." *Journal of Consumer Research* 36(5): pp.715-734, 2010.
- [4] J. Hamari, M. Sjöklint, M., & A. Ukkonen, "The sharing economy: Why people participate in collaborative consumption," *Journal of the Association for Information Science and Technology*, doi: 10.1002/asi.23552, 2015.
- [5] K. Dervojeda, D. Verzijl, F. Nagtegaal, and M. Lengton. "The Sharing Economy. Accessibility Based Business Models for Peer-to-Peer Markets," Brussels: European Commission, Directorate-General for Enterprise and Industry, 2013.
- [6] H. Heinrichs, "Sharing economy: A potential new pathway to sustainability," *GAIA* 22(4), 228-231, 2013.

- [7] O. Lobel, "The Law of the Platform," *Minnesota Law Review*, San Diego Legal Studies Paper No. 16-212, March 4, 2016.
- [8] H. Park and C. Armstrong, "Collaborative apparel consumption in the digital sharing economy: An agenda for academic inquiry," *International Journal of Consumer Studies*, Vol. 41, Issue 5, pp. 465–474, Sept. 2017.
- [9] Y. Benkler, *The wealth of networks: how social production transforms markets and freedom*. New Haven, CT: Yale University Press, 2006.
- [10] Y. Benkler, *The penguin and the Leviathan: the triumph of cooperation over self-interest* (1st ed.). New York, NY: Crown Business, 2011.
- [11] Y. Benkler, A. Shaw, and B. M. Hill. "Peer Production: A Form of Collective Intelligence." In *Hand-book of Collective Intelligence*, T. Malone and M. Bernstein, Eds., Cambridge, MA: MIT Press, 2015.
- [12] V. Bush, "As We May Think," *Atlantic Monthly* 176, pp.101-108, July 1945.
- [13] T. Nelson, "Complex information processing: a file structure for the complex, the changing and the indeterminate," In *Proceedings of the 1965 20th National Conference*, pp. 84-100, 1965.
- [14] T. Nelson, "The Xanadu paradigm" [a poster], San Antonio, TX: Theodor H. Nelson, 1987.
- [15] T. Nelson, "Xanalogical structure, needed now more than ever: parallel documents, deep links to content, deep versioning, and deep re-use," *ACM Computing Surveys (CSUR): Volume 31 Issue 4es*, 1999.
- [16] F. Lehmann and E. Y. Rodin, Eds. "Semantic networks in artificial intelligence," *International series in modern applied mathematics and computer science*. 24. New York, NY: Pergamon Press, p.6, 1992.
- [17] R. F. Simmons, "Synthetic language behaviour," *Data Processing Management* 5 (12): pp.11–18, 1963.
- [18] R. Quillian, "A notation for representing conceptual information: An application to semantics and mechanical English paraphrasing." SP-1395, System Development Corporation, Santa Monica, 1963.
- [19] T. Berners-Lee and M. Fischetti, *Weaving the Web: The Original Design and Ultimate Destiny of the World Wide Web by its inventor*, Britain: Orion Business, 1999.
- [20] T. Berners-Lee, W. Hall, J. Hendler, N. Shadbolt and D. Weitzner, (2006). "Computer Science: Enhanced: Creating a Science of the Web," *Science*, 313 (5788): pp.769–771, 2006.
- [21] T. Berners-Lee, J. Hendler and O. Lassila, "The Semantic Web," *Scientific American Magazine*, May 17, 2001.
- [22] Apple Inc., "Organize and Find Your Photos," <https://support.apple.com/en-us/HT207368>, accessed 3 Sept. 2017.
- [23] M. Nunez, "Want to Know What Facebook Really Thinks of Journalists? Here's What Happened When It Hired Some," *Gizmodo*, <http://gizmodo.com/want-to-know-what-facebook-really-thinks-of-journalists-1773916117>, 5 Mar. 2016.
- [24] A. Chen, "The Laborers Who Keep Dick Pics and Beheadings Out of Your Facebook Feed," *Wired*, <https://www.wired.com/2014/10/content-moderation/>, 23 Oct. 2104.
- [25] R. Graff, "How the Washington Post used data and natural language processing to get people to read more news," <https://knightlab.northwestern.edu/2015/06/03/how-the-washington-posts-clavis-tool-helps-to-make-news-personal/>, 3 June 2015.
- [26] I. Choi, "Distributed and Collaborative Multimedia Authoring Architecture and its Application to Learning Platform," in *Proceedings of the 13th Annual Conference on Multimedia Technology and Applications*, Cyberjaya, Malaysia, pp.87-90, July 2017.
- [27] G. Shao, "Understanding the appeal of user-generated media: a uses and gratification perspective", *Internet Research*, Vol. 19 Iss: 1, pp.7-25, 2009.
- [28] Wikipedia, "YouTube Channels Launched in 2005," https://en.wikipedia.org/wiki/Category:YouTube_channels_launched_in_2005, accessed 5 Sept. 2017.
- [29] D. J. O'Keefe, *Persuasion: Theory and Research*. Beverly Hills, CA: Sage, 2002.
- [30] M. Metzger and A. Flanagin, "Psychological Approaches to Credibility Assessment Online," in S. S. Sundar, Ed., *The Handbook of the Psychology of Communication Technology*, West Sussex, UK: John Wiley & Sons, 2015.
- [31] S. Sundar, H. Jia, T. F. Waddell, and Y. Huang, "Toward a Theory of Interactive Media Effects (TIME): Four Models for Explaining How Interface Features Affect User Psychology," in S. S. Sundar, Ed., *The Handbook of the Psychology of Communication Technology*, West Sussex, UK: John Wiley & Sons, 2015.
- [32] S. Kalyanaraman and B. W. Wojdyski, "Affording Control: How Customization, Interactivity, and Navigability Affect Psychological Responses to Technology," in S. S. Sundar, Ed., *The Handbook of the Psychology of Communication Technology*, West Sussex, UK: John Wiley & Sons, 2015.
- [33] N. B. Ellison and J. Vitak, "Social Network Site Affordances and Their Relationship to Social Capital Processes," in S. S. Sundar, Ed., *The Handbook of the Psychology of Communication Technology*, West Sussex, UK: John Wiley & Sons, 2015.
- [34] M. Bruhn, V. Schoenmueller and D. B. Schäfer, "Are social media replacing traditional media in terms of brand equity creation?," *Management Research Review*, Vol. 35 Issue: 9, pp.770-790, <https://doi.org/10.1108/01409171211255948>, 2012.
- [35] K. Narhstedt, H. Chu and S. Narayan, "QoS-aware resource management for distributed multimedia applications," *J. High Speed Networks* 7, pp. 229–257, 1998.

- [36] I. Choi, "Media Authoring with Ontological Reasoning: Use Case for Multimedia Information Extraction, in M. Maybury, Ed., *Multimedia Information Extraction: Advances in Video, Audio and Imagery Analysis for Search, Data Mining, Surveillance, and Authoring*. Hoboken: Wiley & Sons, 385-400, 2012.
- [37] World Wide Web Consortium Multimodal Interaction Working Group, "Multimodal Interaction Working Group Charter," <http://www.w3.org/2013/10/mmi-charter>, 31 Dec. 2016.
- [38] World Wide Web Consortium Multimodal Interaction Working Group, "Multimodal Architecture and Interfaces" <http://www.w3.org/TR/2012/REC-mmi-arch-20121025/#d3e386>, 25 Oct. 2012.
- [39] The 23rd Nations and Regions Media Conference, <http://salfordmediafestival.co.uk/events/nations-regions-media-conference/>, accessed 3 Sept. 2017.
- [40] J. Stevens, "Backpack Journalism Is Here to Stay," *Online Journalism Review*, <http://www.ojr.org/ojr/workplace/1017771575.php>, Feb. 2002.
- [41] A. Lugmayr, "Connecting the real world with the digital overlay with smart ambient media—applying Peirce's categories in the context of ambient media". *Multimedia Tools and Applications*. **58** (2): 385–398, 2010.
- [42] I. Choi, "Gestural Primitives and the context for computational processing in an interactive performance system," in M. Battier and M. Wanderly, Eds., *Trends in Gestural Control of Music*, Paris: IRCAM, pp.139-172, 2000.
- [43] I. Choi, "A Manifold Interface for Kinesthetic Notation in High-Dimensional Systems," in M. Battier and M. Wanderly, Eds., *Trends in Gestural Control of Music*, Paris: IRCAM, 2000.
- [44] I. Choi, "From motion to emotion: Synthesis of interactivity with gestural primitives," *AAAI 1998 Fall Symposium*, Technical Report FS-98-03. Published as: *Emotional and Intelligent: The Tangled Knot of Cognition*, AAAI Press, pp.61-67, 1998.
- [45] I. Choi and C. Ricci, "Foot-mounted gesture detection and its application in a virtual environment," *Proceedings of the 1997 IEEE International Conference on Systems, Man and Cybernetics*, v.5, pp.4248 – 4253, Oct. 12-15, 1997.
- [46] D. Offenhuber, "The Invisible Display—Design Strategies for Ambient Media in the Urban Context," *Proceedings of the 10th Annual Conference on Ubiquitous Computing, Ubicomp 2008 Workshop Program*, Seoul, Korea Sept. 2008.
- [47] I. Choi, "From Tradition to Emerging Practice: A Hybrid Computational Production Model for Interactive Documentary," *Entertainment Computing*, Vol. 1 Issue 3-4. Amsterdam: Elsevier, pp. 105-117, Dec. 2010.
- [48] J. Nielsen, *Designing Web Usability*, Berkeley, CA: New Riders, 1999.
- [49] Beatpick Music Licensing Online Catalogue, <http://www.beatpick.com/player/index>, accessed 5 Sept 2017.
- [50] World Wide Web Consortium RDF Working Group, "Resource Description Framework (RDF)," <https://www.w3.org/RDF/>, 25 Feb. 2014.
- [51] R. Barthes, *Camera Lucida*, R. Howard, Trans., New York, NY: Hill and Wang, 1981.
- [52] T. Hawkes, *Structuralism and Semiotics*, Berkeley, CA: University of California Press, 1977.
- [53] T. W. Adorno, *The Jargon of Authenticity*, K. Tarnowski and F. Will, Trans, Evanston, IL, Northwestern University Press, 1973.
- [54] C. Shannon and W. Weaver, *The Mathematical Theory of Communication*, Urbana IL: University of Illinois Press, 1949.
- [55] World Wide Web Consortium OWL Working Group, "Web Ontology Language (OWL)," <https://www.w3.org/OWL/>, 11 Dec. 2012.
- [56] I. Horrocks, U. Sattler, and S. Tobies, "Practical Reasoning for Expressive Description Logics," in H. Ganzinger, D. McAllester, and A. Voronkov, Eds., *Proceedings of the 6th International Conference on Logic for Programming and Automated Reasoning (LPAR'99)*, number 1705 in Lecture Notes in Artificial Intelligence, pages 161–180. Springer-Verlag, 1999.
- [57] Adobe Inc., "Adobe Prelude," <http://www.adobe.com/uk/products/prelude.html>, accessed 5 Sept. 2017.
- [58] BBC News Lab, "The Network of Content," 2012. <http://bbcnewslabs.co.uk/projects/network-of-content/>, accessed 5 Sept. 2017.
- [59] Onframe, Inc. "Software Defined Media Intelligence," <https://onframe.io/>, accessed 5 Sept. 2017.

Authors



Insook Choi is Professor of Interdisciplinary Arts and Technology and holds a Chair in Creative Media and Technology at MediaCityUK, University of Salford in Greater Manchester, UK. Her previous posts include Associate Provost at Columbia College Chicago; founding director of the Emerging Media Technologies Program in the College of Technology, City University of New York; and

member of Doctoral Faculty of Computer Science at the CUNY Graduate Center. She received the MA degree in music at Yonsei University in Korea and the Doctor of Musical Arts (1995) from the University of Illinois at Urbana-Champaign (UIUC) with the thesis topic on interactive chaotic circuit with multidimensional interface. She held a research post in Beckman Institute of Advanced Science and Technology and National Center for Supercomputing Applications. Her research is widely published in international journals and proceedings across engineering and computer science communities. She participates on international editorial boards and program committees in the areas of creative technology, human-computer interaction, interactive media, and artificial intelligence. Her research interests focus on both tangible and intangible immersive experiences, interaction design with novel interfaces, interactive narrative structure and multimedia authoring, and computational architecture where new use cases drive technology innovation with interdisciplinary research agenda.



Robin Bargar is Professor in the School of Arts & Media and in the School of Computing Sciences & Engineering, University of Salford. He is founding Executive Director of the Studio for International Media & Technology at MediaCityUK, and Director of U. Salford's Digital and Creative Industry Collaboration Zone.

He previously served as Dean of the School of Media Arts and Professor of Interactive Arts and Media, Columbia College Chicago, and Dean of the School of Technology and Design and Professor of Entertainment Technology at the City University of New York. Robin received a DMA (2010) from the University of Illinois at Urbana-Champaign and has a creative and research background in computer music and sound synthesis, interactive media and virtual reality, including software development and systems integration for cross-media applications. He is recognized for creative works in music, media and technology.

