VIEWER'S CHOICE: RETHINKING MEDIA PERSONALIZATION IN A MULTI-SCREEN WORLD

Venu Vasudevan

ARRIS Group Inc, USA

ABSTRACT

TV as a medium is undergoing two notable trends - ***dispersion*** and ***atomization***. *Dispersion* is the evolution of TV from single screen at-a-time viewing into a rich, distributed viewing experience across multiple screens. *Atomization* is the transition from linear to non-linear storytelling with substantial user control of the media consumption process. Without a technology solution, these two trends have the potential to adversely affect the economics of TV in increased content costs, increased user experience (UX) complexity, and therefore decreased user participation.

This paper proposes a media services architecture that enables the delivery of richer multi-screen media experiences, while still maintaining the coherence of the experience and the cost of service delivery. We share a practical experience in running such a system over a large, globally diverse media corpus. The working system supports a Social Electronic Program Guide (EPG) user experience leveraging both current and emerging (Wearables and IoT) device platforms.

MOTIVATION AND PROBLEM

The 'Viewer's Choice' Challenge

TV as a medium is undergoing two notable changes - *dispersion* and *atomization*. ***Dispersion*** is the evolution of TV into a distributed experience across multiple screens, and ***atomization*** the componentization of parts of a TV experience into modular media segments that can be reassembled into derivative media. These trends change the nature of media viewing and impact what it means to personalize the viewing experience.

As the Google multi-screen study points out [1], ***dispersion*** is more than a simplistic use of different screens for one-at-a-time video streaming. The sophisticated use of multiple screens sequentially and simultaneously introduces two new UX elements – *divided attention* and new *lean-in levels*. With multiple screens involved, there is both a movement of user attention across screens, as well as the potential for a user to be *simultaneously* paying *partial* attention to multiple screens. In addition to *divided attention* across screens, the user experience on any single screen has become more intricate as TV UX platforms include support for tiling, transparent overlays and other advanced user interfaces. For instance, the humble EPG has gone from a rectangular grid on TV, to a tablet interface with multiple levels of click-thru screens, each with a rich, interactive interface. This means that the minimalistic two-level attention model of *lean-back* vs. *lean-forward* needs to incorporate multiple lean-in levels that lie in between those extremes. As an example of lean-in models, an EPG user could be in navigating to a known program, open to passive discovery, binge recording using the EPG as a tool, exploring the Twitter conversation for a single show, or looking for applications (apps) associated with the show as an overlay experience to TV watching.

Atomization is the transition from linear to non-linear storytelling (and non-linear viewing), where the viewer customizes the *what* (entry point, content granularity and perspective), in addition to the *how* (devices, duration, interactivity) of the viewing experience. As an example of entry points, the choice of user experience could be a highlight reel based on a particular player or character as a seed. Alternatively, the user could pivot from a car chase scene to all movies with scenes of similar mood, movement and genre. With an increasing number of content publishers also publishing companion apps [2,3], atomization also means the viewer mixes modalities – interacting with flexible mixes of video (e.g. an NFL game) and associated apps (e.g. a Fantasy App).

Personalization in a Viewer's Choice World

A user experience with this many choices can be overwhelming without *personalization*. That said, traditional personalization (e.g. collaborative filtering) works for linear viewing but is too heavy-handed for a complex, layered viewing model such as the above. In a viewer's choice world, personalization needs to transition from prescriptive recommendations of entire media items to lightweight *conversations* and exploratory *hints* that the user chooses to opt-in to (or not). The conversation itself may pivot from content titles, to actors, moods, modalities (raw content vs. apps), pricing (free vs. paid), timing (immediate view vs. cached for future viewing), and a growing number of new dimensions and entry points.

To support frequent and lightweight forms of discovery, two services used as building blocks for personalization are proposed. The first uses *media analytics* to enable content providers to create dynamic metadata driven experiences. This enables new assemblies of derivative content to be presented to end users as substantially new experiences, without substantially higher editorial cost. The second is a *social fusion* service that manages multi-device content delivery and 'artful user interruption'. This service maps the available user attention to content delivery on screens with varying capabilities and affordances. These building blocks will be described in detail in a following section on Frameworks. However, the overall design philosophy of a personalized UX has to follow the principles listed below to be effective with this many degrees of user freedom.

1. *Power tools provide choice.* The personalized experience can be viewed as an agile experience that can be easily steered in directions that users choose, not as a destination pre-determined by an algorithm. This leads to the next point of the personalization UX being easy to opt-in to, and use.

2. *Personalization should/must be 'native' to the UX*. A key to simplifying the personalization UX is to make it 'native' to the application UX. By putting ambient personalization at top level interfaces, and more active 'lean in' personalization at nested interfaces, we match the personalization interface to the expected level of user engagement when a user is at that particular level of the application interface.

3. *Screens are personalities, not rectangles.* Studies [6,1] have pointed out that it is useful in a multi-screen experience to think of screens as people with personalities, rather than inter-changeable rectangles. Users may use a particular screen out of

convenience, but the mapping of user task to screen is a complex function of demographic, and level of available user attention (e.g. lean-back to lean-forward).

4. *Entertaining is not just utilitarian*. Personalization adds to the information presented to the user, but may or may not be something the user always opts-in to. Any interface is likely to vary greatly in whether the information presented is an exact match for the user need. By making the journey entertaining, user engagement can be maintained while refining content relevancy via iterative conversation.

The next few sections describe a services architecture and framework for multi-screen personalization based on operating a lab system continuously for over 3 years and 200K pieces of TV content. We describe our experience in using this Framework to augment an Electronic Program Guide (EPG) product with Social TV capabilities. The paper concludes by outlining future work driven by lessons learned, as well as extensions to existing standards to multi-screen storytelling.

FRAMEWORK

The framework resulting from our experience building several multi-screen UXes is divided into two inter-operating web services that provide *descriptive* and *responsive* metadata pertaining to media respectively. The *Media Analysis Framework* (**MAF**) ingests media and extracts a feature vector of *descriptive metadata*, i.e. metadata that describes what is happening **within** a video content item. The *Fusion* Service (**FS**) ingests social network feeds (e.g. Twitter) and relevant web information services (e.g. Yahoo Fantasy) to extract *responsive metadata*. This type of metadata is either created *in response* to the content (e.g. Tweets) or as a responsive overlay on content (e.g. Fantasy scores, relevant TV apps).

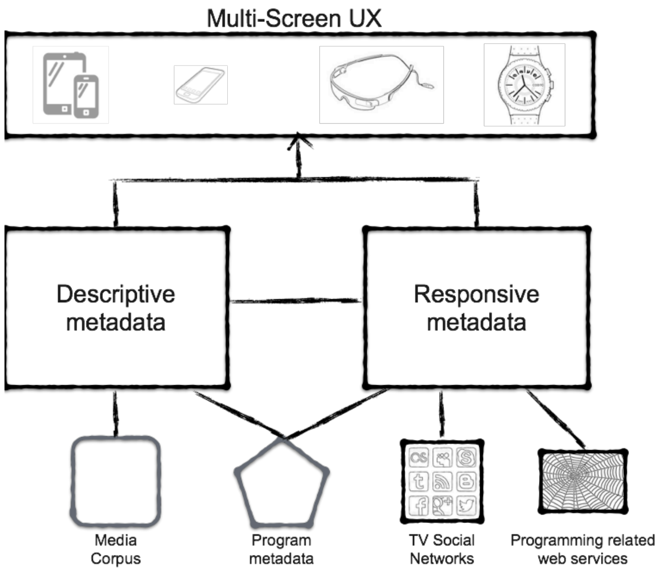

Figure 1 – Metadata Services Architecture

The two metadata services feed into a *multi-screen UX controller* system, which determines the timing and payload of the content displayed on different screens. This subsystem is aware of the capabilities and affordances of different screens, and transcodes the content to match the target device(s). These three components are described in more detail below.

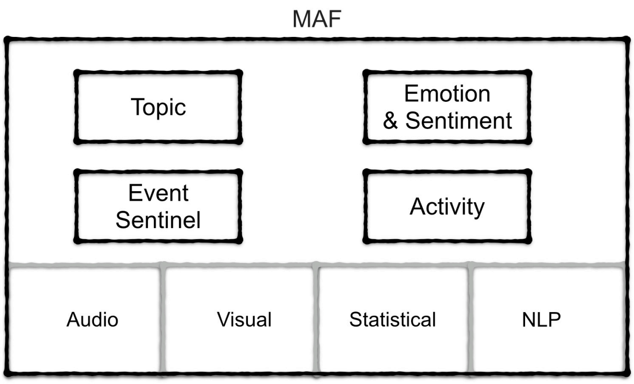

Figure 2 – MAF: Descriptive Metadata

Descriptive Metadata and MAF

Descriptive metadata can be extracted directly from the media item, and provides the capability to flexibly segment a piece of media into new experiences. The descriptors can further be divided into *structural* and *semantic* descriptors. *Structural descriptors* detected by ***MAF*** include the delineation of black frames, audio silences, colour variance indicators of the level of activity on a scene, and video hard-cuts that are structural properties of the media. Such structural metadata can then be used to identify UX level segments such as replays in sports programs (via hard-cuts), and program-ad boundaries (via the presence of black frames and silence).

Semantic descriptors extract 'media chunks with meaning'. Semantic descriptors in MAF include: sentiment/emotion extractors, topic detectors, speaker identifiers, and named entity extractors to name a few. Use cases of market interest around semantic metadata include *semantic search* based on People/Places/Brands, *smart previews* of content based on semantic summarization, and *intelligent notification*.

The MAF system automatically ingests content and exports feature vectors containing its descriptive metadata to a cloud service, to enhance scaling and reachability. The ingest architecture enables the extraction of descriptive metadata in both stored content (e.g. on demand content stored in a content management system) or in-stream (content ingested from a real-time feed) modes. In terms of the estimated scaling and feature vector requirements, the current MAF service has been continuously ingesting 8-10 cable channels over 3 years, and the associated feature vector metadata for ~150K ingested media elements is about 100GB across SQL and NoSQL metadata stores.

Responsive Metadata and Fusion Server

The **Fusion** Server collects, extracts, and aggregates two types of responsive metadata – *explicit* and *associative*. *Explicit* responsive denotes user-generated content that is explicitly created by users who are watching the content. Examples of this include tweets that co-occur with the broadcast of an NBA basketball game, or Facebook posts about the stars on the catwalk during an Oscar Award ceremony broadcast. *Associative* responsive metadata is web content that deepens and enriches the media experience, but *was not* explicitly created for the show or in response to watching it. For example, Yummly recipes can enrich any Food Network airing, and the same holds for Wikipedia content that is relevant to the topic of a news story. With the increasing trend of packaging responsive metadata experiences into marketplace apps, TV apps (created by both the studio and third-parties) are a significant component of the responsive metadata category.

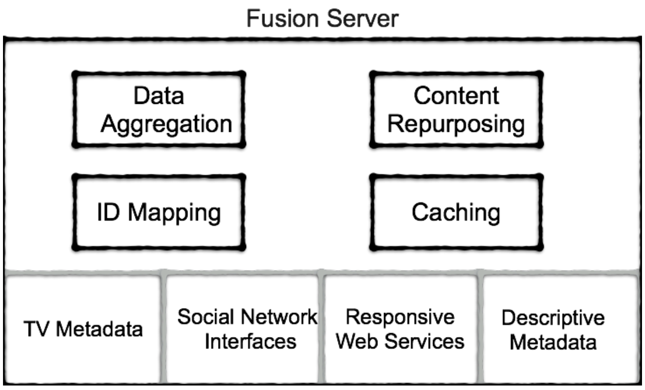

Figure 3 – Fusion Server

Beyond the diversity of web data source interfaces, Fusion Server provides capabilities that simplify the creation and delivery of responsive metadata. The *ID Mapper* enables data aggregation across TV metadata with mismatching ID standards (Twitter TV Program Ids vs. TMS vs. Rovi). Intelligent *result caching* enables high-performance aggregation across information sources where data changes at different rates (from few seconds to once every 30 minutes). The Data Aggregation module enables the aggregation of metadata from internal (e.g. MAF) and external (e.g. Twitter) sources into a single, differentiated result for the content provider and operator. *Content repurposing* enables the creation of rich interactivity for the more capable devices (e.g. Tablets) and specialized content bundles for wearables and IoT. The latter enables the experience on Wearables and IoT devices that is both device appropriate and optimal in bandwidth and battery usage.

In the *explicit* category, the Fusion Server implements filtered and curated interfaces to Twitter and Thuuz. In the *associative* category, it implements genre specific interfaces to Yahoo Fantasy (Sports) and Yummly (Food). It also interfaces to ProductHunt (an app and content curator), and uses Bayesian Learning to categorize trending apps on ProductHunt into their associated content categories. With additional interfaces to NFL and MLB player information, both content and app overlays to linear TV line-ups in Sports are supported. Also supported are app overlays in Food and News categories with a database of over 250 apps, and filtered tweets and Twitter trend data ranging over 6-8 months. There is automated support for app discovery, download and invocation for Android and Web apps. Fusion Services also support ID translation between Twitter, TMS, and Rovi Ids.

Multi-Screen UX Controller

As wearables and IoT proliferate, and there's a need to extend device-appropriate content feeds to new device classes without expensive content authoring costs. These devices have additional constraints on bandwidth and power. The multi-screen UX controller addresses these by supporting both selective content targeting and content repurposing services. The former enables the programmer to determine the timing and scope (device classes) that any push notification or interactive content unit will be targeted. Quick glance-able interactions (e.g. favourite player scoring touchdown) are targeted at wearables, where longer interactions (e.g. Q and A, or an app download promotion) are scoped down to tablet or smartphone devices only. Notification scheduling can additionally leverage user context (in front of a physical TV set vs. watching on a mobile device or smaller screen) or device affordance (e.g. whether device supports video streaming). The current Multi-Screen UX controller supports Android Wear devices and Google Glass. Extending support to new devices and iOS is anticipated in the near future.

SYSTEM AND EXPERIENCE

TV interactivity has been both lucrative and risky as a market proposition. There have been a number of failed attempts at interactive TV over the last couple of decades. Yet much of the success of many social networks (Twitter and Tumblr in particular) comes from being a destination for responsive user generated content. To understand the dividing line between desirable and cumbersome, we conducted several studies on user (and advertiser) attitudes to video interactivity, and UX preferences for advanced TV interactivity. Our findings were:

- Users desire light interactivity, but have a strong negative reaction to interface complexity (e.g. complex nested structures with no 'Return Home' button)
- Users think of search, discovery and personalization as an organic composite, and see the current discovery UX as both unoriginal (same old recommendations), 'gamed' (recommendations are thinly veiled promos) and at the wrong level of granularity (entire media items, not key scenes, actors)

- Users have had a poor experience with complexity in TV interfaces and equate it to poor performance (back to the point about lightweight interfaces)
- Advanced users find multi-user interactivity natural. But concerned about inappropriate interruption.

In the rest of this section, we illustrate how Fusion Server and MAF enable an enhanced EPG that aligns to the principles stated above. The goal here isn't an exhaustive coverage of Social EPG features, as much as illustrating the architectural patterns and leverage created by MAF and Fusion services.

Top-Level Ambient Interfaces

For light interactivity with a low learning curve, the guide was augmented with *contextualized ambient interfaces*. Instead of showing visually overwhelming tweets, Twitter trends were incorporated into the top-level interface as an ambient indicator (see Fig 4). Using the *Twitter Trends* API, social discovery was integrated into the core program descriptor based discovery EPG's support. Fusion Server capabilities for time zone and threshold based trend filtering also enable operator customization of Social Discovery.

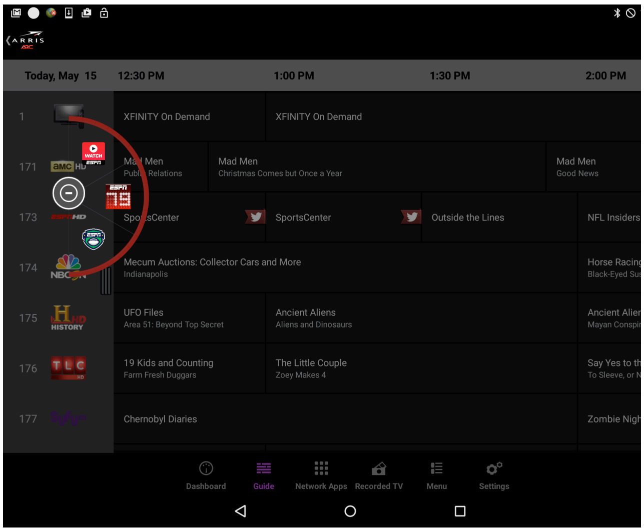

Figure 4 – Ambient, & Low Friction UX

Another common desire for TV viewers is an ability to *pivot* their search from shows to particular actors or show guests. At the same time, there is a desire for operators with largely un-viewed line-ups (given the average user watches 7-11 channels in his line-up) to create exposure for the un-viewed majority. Emerging UX features such as the Radial Menu [5] along with nuanced touch (short vs. long press) gestures are used to organically integrate a *pivot ecosystem* into the core EPG experience. Traditional pivoting on actor and guest are enhanced to filter results within a relevant space and time window. This enables operators to promote less visited channels in the process of providing user value. It also caters to *binge recording* and the batch creation of DVR recordings.

In addition to pivoting on information, pivoting from *TV content to TV apps* (and vice-versa) via the enhanced guide [7] is supported. The Fusion Server maintains a curated list of (over 300) channel or program related apps, and exposes these via the radial menu. Fusion is aware of the app state (preloaded or not), and accordingly downloads and/or invokes the selected app. Combination of ambient information, short click and long click in different contexts enables rich information sets without UX overload.

Personalized Social Streams

Most EPGs extend the linear 'grid' with an on-demand carousel style interface (a la Netflix) to support access to on-demand content. From a UX perspective, we view a carousel as a metaphor to support personalized social streams of *both full content items and content* *segments.* In this way, new forms of segmented content are shown in ways that are organic extensions of existing metaphors.

Several different kinds of social streams are supported, of which the dominant ones are *localized*, *faceted*, and *searched*. An example of a *localized stream* is one that uses location-based DVR analytics to render a 'Viewed History' from the viewer's immediate neighbourhood. MAF filters by content facet (e.g. music performances) are used to enable faceted streams (e.g. only the music performances in past David Letterman shows). Fusion and MAF enable the creation of streams based on a variety of facets beyond music. *Searched* streams are created in response to an explicit (voice) search for a keyword in the watched program – such as all of Lionel Messi's goals in the World Cup. Searched streams could also be thought of as a user initiated faceted stream.

Smart Notifications: Multi-Screen UX

In addition to user-initiated functions, *push notifications* play an important role in creating personalized experiences. Viewers want to know when their favourite show is about to start (especially where Twitter feeds can play 'spoilers'). Sports fans want to be notified of significant player events while away from the TV. Foodies want to be able to opt-in to inline promotions of a chef's cookbooks while the show is active.

The multi-screen UX module deals with both the 'when' and the 'where' of push notifications. Based on a user *attention model* (e.g. based on both ***content sentiment***), push notifications are scheduled at natural pauses in the story, and inappropriate interruptions are avoided. These content signals are integrated with social feeds (e.g. Twitter signals that provide a user engagement amplitude) to further refine the timing of push notifications, and to *transcode* right-sized content items to secondary screens.

WEARABLES AND IOT : CHALLENGES AND OPPORTUNITIES

While much of the preceding discussion regarding lightweight interfaces and timely interruption applies equally to all classes of screens, there are some unique aspects to very small (or no) screen devices – namely *Wearables* and *IoT*. The following discussion provides an Android data point, but is equally relevant to the iOS platform.

Wearables: Promising, and in Flux

When evaluating Wearables, it is important to consider both a user experience (therefore opportunity) and an engineering (therefore cost) perspective. The positives with both *Android Wear* (watch platform) and *Google Glass* pertain to the quality of user experience. Android Wear provided an effective experience for responsive push notifications (e.g. Fantasy Football player alerts) and is a surprisingly good voice input experience. Even in the presence of other voice-enabled devices (e.g. Tablets and Smart Phones), it is quicker and easier to use a watch as the voice input device than reach for a mobile computing device. Google Glass is surprisingly good at rendering short streaming videos both in quality and latency (e.g. Sports Replays or alternate angles), and this capability is only likely to be better in future releases. Our preliminary experience with Apple Watch indicates it to be a crossover device with a mix of Android Wear and Google Glass capabilities.

The weaknesses of both wearable categories lie in the user effort and interoperability issues in creating a seamless cross-device experience. Watch and Glass, while both Wearables, have fairly different App deployment models. Watch Apps have to be bundled into mobile apps, and do not exist independently in the App marketplace as Glass Apps do. Watch Apps on Android require the mobile for most critical functions, where Watch Apps on Apple Watch and Google Glass can operate. As the market stands, mixed platform Wearables (e.g. Apple Watch and Google Glass) are somewhere between cumbersome and infeasible depending on the particular combination.

IoT: Where is the opportunity in the near term?

From a Media personalization perspective there are two intersection points between Media and IoT. The first where an IoT 'thing' is a participant in the end user *IoT extended media experience*, and the second where there is the creation of IoT *Triggered* Media based on monitored IoT activity. While past industry efforts in ambient devices (e.g. Nabaztag and Ambient Orb) point to use cases for an IoT Extended Media Experience, that ambient device ecosystem has not yet re-emerged to a commercially meaningful level. The DIY nature of IoT device purchases, and the struggles that end-users have in getting IoT devices to work together, suggests a more promising opportunity in IoT Triggered Media. On-going efforts in this area focus on answering the following questions – a) techniques (e.g. audio fingerprinting) to import 'legacy' IoT (e.g. my old Espresso Machine) into the IoT ecosystem, b) challenges in interfacing IoT protocols to IP protocols and c) information retrieval challenges in converting learnt IoT activity to 'found' media in ways that help the end user better use and maintain their collection of IoT devices.

CONCLUSIONS AND FUTURE WORK

We presented the multi-screen personalization problem, and the cost and complexity concerns in effective UX personalization in the emerging video delivery environment. We then showed a framework that enabled cost effective and personalized media experiences. Social EPGs extended with organic content discovery provided one proof point of the viability of this framework, one that can benefit from the evolution of both media and device standards. Future work on the media dimension focuses on protocols for more efficient media handling and handoff in resource constrained devices via enhancements to existing standards (e.g. MPEG multi-view). Data aggregation work focuses on automated ways to normalize data sets with varying timelines and provenance.

REFERENCES

[1] Google Multi-Screen World Study - https://goo.gl/AeRsqQ

[2] Vasudevan, Venu, and Jehan Wickramasuriya. Social & Interactive TV: An outside-in approach, Future TV, EUROITV (2011).

[3] A. Nandakumar and J. Murray. 2014. Companion apps for long arc TV series: supporting new viewers in complex storyworlds. In Proc of the 2014 ACM Intl Conference on Interactive experiences for TV (TVX '14). ACM, New York, NY, USA, 3-10.

[4] Vasudevan, Venu, et al. Can context tame web TV? Reconciling infinite choice and zero effort, 2013 IEEE International Conference on Pervasive Computing. IEEE, 2013.

[5] Google's seriously cool new Android radial menu design revealed, http://goo.gl/olPwB2

[6] Meet the Screens, Improve your multi-screen marketing by understanding how consumers relate differently to each screen, http://goo.gl/FWL7RM

[7] The TV App Opportunity, http://goo.gl/BYWU1q