Augmented Reality: Service Construction via a 4D Communication Model

Daping Wang, Alcatel-Lucent Australia
Michael Brenner, Alcatel-Lucent USA

Abstract

Connecting people is the ultimate goal of telecommunication. With advances in computing capability and ubiquitous access to various data, today's connection is focused on the blended information delivered by a converged mechanism, driven by the consumer's thirst for digital social interaction. Leveraging technologies like data mining, service-indication texting, and synthesized imaging, next-generation applications are incorporating more information into our daily communication. More services hence emerge from this kind of data federation. This article explores potential service augmentation by placing communication in a four-dimensional model that combines service delivery and external data augmentation. Using categorized information obtained by data mining techniques, a specially designed service indication text language could bring user interactivity to new levels. Tele-reality converges toward use of synthesized images or videos that are generated by merging and processing information about people, place, time, mode, weather, recent events and more. In addition, this article also illustrates how current standard user profiles and application programming interfaces are leveraged in order to accommodate such services.

Introduction

Seeing is believing — people are more likely to treat things that they can perceive and observe as reality. This is also the starting point of augmented reality (AR). AR systems, as defined in R. T. Azuma's survey paper [1], are systems that have the characteristics of combining interactively in real time the real and the virtual, and registering the result in 3D. If we look at our near-real-time communication methods (e.g., video, audio, or even texting), we may call them "real" since they are all targeted at delivering tele-presence or tele-reality. If we could augment our direct perception of reality by adding information from the virtual world, why couldn't we augment our tele-reality? When technologies such as fiber to the home (FTTH), Long Term Evolution (LTE), or LTE-Advanced get widely deployed, people will face more tele-realities than ever before. At the same time, the advancement of the Internet and sensor networking (e.g., location) greatly boosts the availability and accessibility of various kinds of information, which could be utilized in reality augmentation.

Enhancing existing tele-reality could be the low hanging fruit of AR. First off, it avoids concerns with recording and adaptation problems, which come from the need to align real and virtual worlds. Since what we are seeing or hearing via telecommunication channels is already in the form of digitized information, alignment can easily be carried out in the virtual world. Second, augmenting tele-reality certainly circumvents the need for wearing stereo glasses, when the 3D effect is not necessarily required. Third, instead of relying on multilayer displays to blend real imagery and additional information, augmented tele-realities could have one display with additional information directly fused in. Lastly, if the reality itself has already been virtualized from the beginning, more services could be layered on top.

Our daily used telecommunication services fall into three categories — textual, oral, and pictorial. Voice communication stands for the oral part. Video telephony is simply the combination of pictorial and oral channels. Short messages (SMS) and instant messages (IM) are mainly textual. In other words, all of our telecommunication services are enclosed within this 3D space expanded along the textual, oral, and pictorial axes. The telecommunication industry has never stopped its exploration in the 3D communication space since Alexander Graham Bell invented the telephone. Instead of focusing on how to further extend each communication category, today's challenge consists in expanding this 3D space with more dimensions.

For starters, we first examine the transforma-

0163-6804/12/$25.00 © 2012 IEEE

tions of a telecommunication service along different axes, for example, how to convert an IM session to a voice call by applying text to speech (TTS) technology. With the network acquiring more computing capability and ability to access information, enhancements could go well beyond such kind of communication transformation. The access to "explorable information" via data mining technology enables us to expand our horizon and turn each real-time communication into a full fledged multimedia session.

Starting with the goal of enabling mobile applications, the Open Mobile Alliance (OMA) has invested significant effort in enabler specifications, covered extensively in [2]. For the purpose of this article, the most relevant OMA enabler specifications fall in one of two main categories: communication and content delivery enablers on one hand, and subscriber and contextual data management enablers on the other. In addition, OMA's recent focus on application programming interface (API) specifications is broadening the appeal of its specifications by allowing Web 2.0 developers to easily tap into telcos' exposed capabilities.

This article projects how the traditional communication space could be expanded. We cover how data mining and federation would help us to further explore the communication space. We use an enhanced instant messaging prototype to demonstrate how to construct and augment tele-reality. We discuss recent activities in standards bodies, especially OMA, in terms of communication enhancement.

Communication Space

Our world is evolving through patterns of disruption and harmony. When applied to the telecommunication industry, this is reflected in the fact that on one hand there is a boom in voice services on the Internet, which is posing a big revenue threat to telcos; on the other hand, the ever-richer Internet resources greatly attract more users to subscribe to broadband services. To prevent the telcos from being only a bit pipe supplier, they must explore both the wild Internet and themselves [3].

The traditional territory telcos occupy is mainly the 3D communication delivery space defined by textual, oral, and pictorial methods, referred to in this article as the TOP space (Fig. 1).

Seven scenarios (Table 1) can be constructed in the TOP space by combining the three dimensions. It is easy to understand that most of these seven scenarios of our daily communications only happen in one or two ways. To really set up a communication, we need the network's help in connecting two TOP spaces together. Connections along the same axis are straightforward in today's communications. For example, short messages are normally passed between mobile phones (T ↔ T). Bridging different delivery mechanisms together or providing unified communication has become both technically feasible and service attractive after a service provider's endorsement of converged IP-based core networks (e.g., IP multimedia subsystem [IMS] [4]). Cross-domain/channel communication also provides a vehicle to further enhance real-time communication services.

Combination	Communication scenario
T	Short message, instant message, etc.
O	Voice call, voice mail, etc.
P	Video call, video mail, video share, image share, etc.
TO	Instant messaging during a voice call.
TP	Sharing of video or image while in chatting sessions. Most of the time, this is regarded as file sharing in instant messaging.
OP	A typical video/image share scenario in GSMA Rich Communication Suite (RCS).
TOP	Text message exchanges during a video call.

Table 1. *Delivery scenarios.*

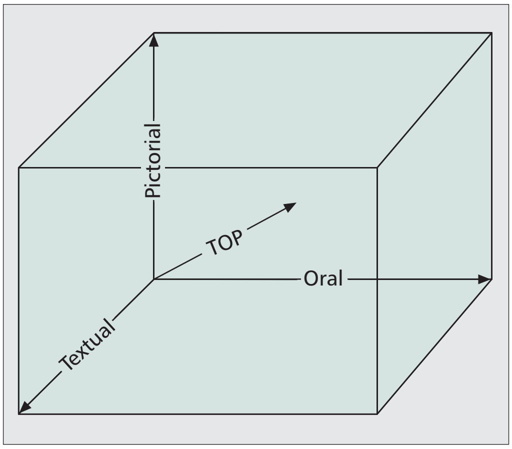

Figure 1. *TOP communication space.*

Deciding how many bridges to build is the first step in realizing cross-domain communication. When applied to peer-to-peer communications, it is really those scenarios that involve different domains that are of most interest. For example, for a voice to voice-plus-messaging communication or T ↔ TO, we only need to cover T ↔ O as T ↔ T or voice to voice is a de facto service. This will leave us with only six bridges to cross. Table 2 summarizes the four major technologies that drive these six kinds of unified communications: text to speech, automatic speech recognition (ASR), descriptive video service (DVS), and text to picture (TTP).

Both TTS and ASR have been very well explored, and it is not difficult to find their uses in various applications. DVS is a service that helps people with visual difficulties to understand what is going on in the display. Such services are provided by some TV stations and cinemas. Preparing DVS for films or shows generally requires people to watch the show, and provide an accurate and vivid description of the scene in a period of time. Such preparation makes real-time rendering almost impossible.

Cross-domain communication	Description
T <=> O	Text to Speech (TTS), Automatic Speech Recognition (ASR)
T <=> P	Text to Picture (TTP), Descriptive Video Service (DVS)
O <=> P	Although a voice call to a video endpoint normally ends up with voice only, this could be enhanced by the combination of ASR+TTP and DVS+TTS.
T <=> PO	TTP, DVS, TTS, ASR
O <=> TP	ASR+TTP, DVS+TTS
P <=> TO	DVS, TTP

Table 2. *Cross-domain communications.*

Compared with DVS, TTP seems more difficult because it uses less information to generate more. As a consequence, the majority of its use lies within the animation industry. Finding ways to circumvent such technical difficulties in applying DVS and TTP in real-time communication will be a determining factor in ascertaining how well our unified communication could be delivered to attain a full-fledged three-dimensional tele-reality.

Unified communication provides the first layer in tele-reality augmentation as it addresses the delivery. The key to the next generation communication lies within the next step, which is to synthesize contents. This is where data federation comes in play.

Data Federation

Coming up with a coordinated delivery mechanism, as described in the previous section, is only half the journey in our communication space. The next step relies on how content and delivery can be smoothly combined together by taking into consideration various attributes of both the end user and the network.

Within their TOP domains, telcos have access to and manage two types of important data resources: subscription data and user behavior. Subscription data is relatively static. It covers charging information, user preferences, billing addresses, and supplementary feature settings. User behavior, in comparison, is rather dynamic. Typical user behavior includes location changes, buddy list update, and so on. As user behavior is driven by the subscription data, we could study them together by considering the user behavior as the temporal change of the subscription. A user's initial sign-up to a service is therefore simply an event that happened at time zero ($t = 0$). The subscriber's buddy list or address book setting is actually a very special parameter because it determines the delivery target. Accordingly, in our model we put the subscriber settings into two parameter vectors, A and B:

- A consists of data settings a, c, d, e, l, and m, as in A = (a, b, c, d, e, l, m).
  - –a stands for the subscriber activities. Such information could be set by the user, based on the calendar, or discovered by motion detection on the mobile device.
  - –b stands for the communication type, or bearer. This is related to the user's subscription to voice or video services.
  - –c refers to the conditions of the communication delivery. This is the predefined service logic such as "call forwarding when busy" ($t = 0$) or per call calling line ID (CLI) display.
  - –d stands for device capability. It includes both the communication delivery capability such as WiFi or LTE access and applications installed on top of it.
  - –e is the economic benefit part of the communication. This could be the user's monthly plan ($t = 0$) or cost per call.
  - –l refers to the subscriber's location. This can be the billing address ($t = 0$), or the current location during the communication. Such location information is applicable for all parties, both in the address book and involved in the communication.
  - –Last, m is the mode of the end user, which is either defined by user input or modified via network intelligence.
- B as the second parameter vector simply consists of all the contacts in the address book.

The Explorable Information

There is no doubt that the Internet is the largest knowledge system ever built in the world so far. Significantly, Internet traffic doubles every year [5]. The mere notion of being able to access this gigantic knowledge system provides us with a wealth of possibilities to explore. Beyond the Internet, more adaptations of smart phones in the market have also equipped the service providers with a personal sensor related to the subscriber. With installed Global Positioning System (GPS) and accelerometer, today's smart phone can reveal not only people's geographical location but also their activities. We refer to such information resources as explorable information (EI) to reflect the fact that they are not under telcos' control but carry significant potential.

As our journey in exploring the EI is only beginning, here we only study information that will be helpful in constructing the tele-reality.

Our understanding of the power of the EI starts with the perspective of a person's snapshot. From a typical snapshot of a tourist, we could detect information such as his or her look, dressing style, mood, possible interests, as well as the background scenery. We can also find similar information from video calls. Beyond that, it is up to our imagination and deduction. Recalling the first conversation between Sherlock Homes and Dr. Watson, "… clearly an army doctor … just came from the tropics … left arm has been injured … clearly in Afghanistan," we cannot help but wonder how a normal communication could be augmented by applying and combining a person's knowledge, observation skill, and logical thinking, even if the example given is fictional.

In a certain way, the modern network has acquired such detective abilities to enhance the communication. Consider video calls or video sharing, for example. When your counterpart's image is shown on the screen, the network could carry out various kinds of data mining techniques. Using visual search technology such as Google Goggle, a landmark building could be identified in the background scenery. Hence, more information could be revealed (e.g., geographic location). Starting from the location information, further queries could retrieve local news, weather forecast, and so on. At this point we can categorize all such external information as a third parameter vector C = (g, n, w, q) where g represents the geographic information, n refers to local news for events that happened or are anticipated, and w stands for the current weather. These can all be related to the location parameter A(l) discussed before. Last, q applies to all general online query results, which can also be considered a wildcard parameter. Query inputs are simply based on parameter vectors A and B discussed above. It is impossible to enumerate all possible queries q carries as A (b) parameter covers the communication contents. The q parameter alone can bring in a personal consultancy service that utilizes online search engine to generate comments to the on-going conversation.

Properly applied, augmented tele-reality can protect the end user's privacy instead of jeopardizing it. Operators offering an "opt-in" with user consent is a necessity, but augmented tele-reality itself can be used to enhance both the visibility and the importance of protecting the end user's vulnerabilities. Protecting the granularity and depth of user information is usually more important than protecting the entire breadth of user information. Most people will not mind having their location shown in Paris, but may be concerned with it being pinpointed to a particular hotel room. For such a case, when making a video call from a Paris hotel room, we might want the network to replace the real background with a picture of the Louvre, for example. Augmented tele-reality opt-in is what end users really would appreciate: providing information, while privacy is protected.

DATA MINING

Any services in the TOP space can be treated as a function with A, B, and C as the key variables or parameters. One can also call these parameters application enablers. Take parameter A(c), for example: it could be treated as the ringback tone service. When combined with the news parameter C(n), a service called "local news ringback tone" can hence be constructed.

Even without enumerating all the parameter interactions, a new space is expanding in front of us. Adding subscriber settings, Internet-based information, and individual targets to our existing TOP space gives us a new 4D communication space (Fig. 2). These parameters not only provide the possibility of creating new applications, but also smoothly utilize the unified communication service delivery. Take one-to-one chat, for example. It obviously occurs in the one-dimensional space, along the T axis. Once we add a text analyzer to help dynamically change the avatar images of both parties involved in the chat, the service will belong in the 2D TP space.

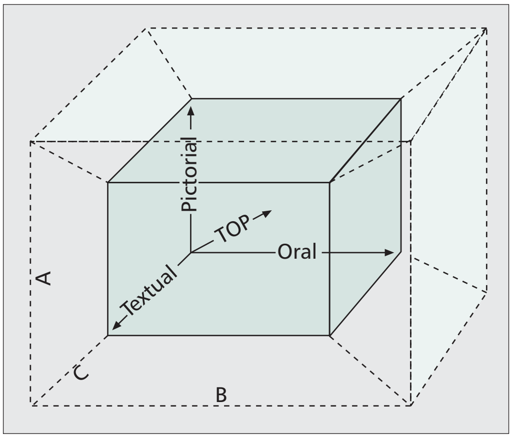

Figure 2. *4D communication space.*

Furthermore, we could add location information A(l) to obtain the time zone information (via C (q)). As a result, a chatting session could change the avatar background to highlight the time difference (if any).

Service providers and equipment vendors alike have made significant achievements in the exploration of subscription and user behavior management. With so many parameters to consider, a critical test will consist of how well they can integrate and leverage all the information to deliver a higher-value service to the end user. OMA's Converged Address Book (CAB) work and Adress Book API are targeted at exploration via parameter B, while several other OMA enablers and OMA RESTful network APIs are more related to exploration via parameters A and C. Through the blending of traditional telecommunication and Internet domains, and supported by relevant standards, fantastic unified service delivery becomes a realistic goal.

Service providers and equipment vendors have made significant achievements in the exploration of subscription and user behavior management. With so many parameters to consider, a critical test will consist of how well they can integrate and leverage all the information to deliver a higher-value service to the end user.

REALITY CONSTRUCTION

AT&T first introduced picture-phone services back in the 1960s. After half a century, video telephony only managed to find a niche in enterprise videoconferencing and family chat applications. Even after it was ported to mobile devices, customers still express a reserved attitude toward it [6]. Afraid of being overexposed in a video call is one privacy issue that makes most people shy away from it. One strategy to address the privacy concern is to use animated images.

Recall the example of a snapshot of a tourist we used in the previous section: we could actually construct a semi-realistic image of a person by using additional information plus a cartoon character. First, based on the location information (g), a background scene can be constructed. Second, we can render the background by using the weather information (w). Additional stuff can be added in by queries of local news and social events. Last, with A = (a, b, c, d, e, l, m) available, and in particular using the mood of the person (m), the animated character in the picture could also change in a dynamic way. Hence, a virtual tele-reality is constructed.

Not only does the use of network constructed images tackle some of the privacy concerns, but it

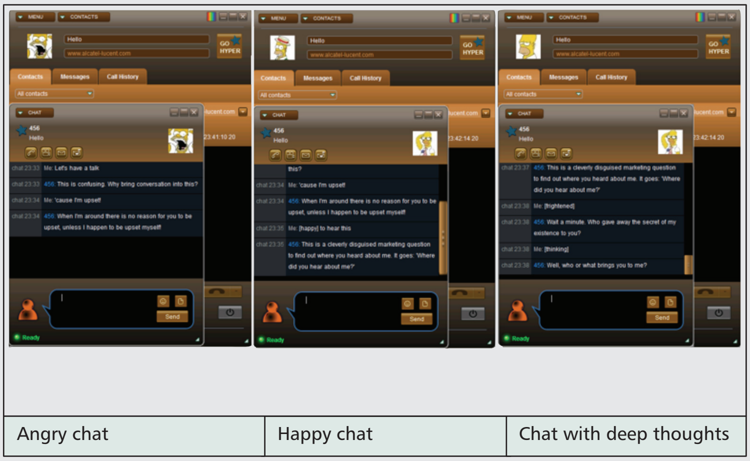

Figure 3. *Enhanced chat.*

also provides transition possibilities in the unified TOP communication space. Take T <=> TP, for example. This is a typical case where a mobile user uses SMS to communicate with an IM user. As SMS users do not have an icon feature, the network will help construct the SMS user's icon based on the network parameter. The IM user will then have the same user experience in viewing the far end user's image changes. At the same time, the network could deliver the IM user's virtual image information (e.g. mood) directly to the SMS user in text. This will smoothly accomplish the augmented tele-reality delivery between two domains, T and TP. This is also in line with OMA's Converged IP Messaging (CPM) work, which we briefly discussed in the last section.

As a proof of concept, we have set up an enhanced IM prototype by using existing GSMA Rich Communication Suite (RCS) infrastructure, which is built on top of OMA's Presence model. A text analyzer is running as an application server in order to detect the user's mood. During a chat session, the status icon will change from time to time, depending on the content of the exchanged messages. The application server is also connected to some Chatbots that can be called in when any of the involved parties want to add some fun to the conversation. The icon delivery uses the standard SIP-NOTIFY-HTTP-GET approach. The screenshot in Fig. 3 shows a series of chat sessions between subscribers 456 and 457.

Standards Support

This section's purpose is to emphasize how OMA standards are evolving to support the development of augmented reality applications.

Let us start with an enabler in the category of enhanced communications and content delivery enablers: converged IP messaging (CPM) [7], an enabler with the explicit goal of providing convergence of multimedia communications services, while leveraging standardized service functionalities from existing communication enablers like instant messaging, short messaging, or multimedia messaging. This enabler is paving the way for telcos to manage their communication space by supporting any of the communication scenarios discussed earlier. CPM's main features include immediate and/or deferred messaging, multimedia contents file transfer, adding/removing media streams on the fly, adding/removing users on the fly, support for oral, pictorial, and textual messaging, support for group communications, support for multiple devices per end user, and interactions with the Presence enabler. Finally, CPM also supports interworking with non-CPM communication services, which would allow an end user to enjoy some of the side effects of the features for which his/her wiser friend on the other end of the communication channel signed up. Other OMA enablers in this category include OMA SIP Push [8] and OMA Dynamic Content Delivery (DCD) [9].

Another OMA enabler of interest, this time from the subscriber and contextual data management category, is the Converged Address Book (CAB) [10]. The CAB Enabler allows a CAB user to add, delete, and modify his/her Address Book data (e.g., contact information, authorization rules, buddies, and contact status). It also allows a CAB user to search for other CAB users' contact information or request to subscribe to other CAB users' contact information updates. Through the use of CAB, an end user could see that his/her buddy is present and available, and therefore that he/she can contact him/her. The end user is eager to share with this friend some recent events of common interest and at least his/her own mood regarding those events — and since a CAB user can share contact information with his/her buddies, regardless of whether they are CAB users or not, the end user could share with his/her buddy contact information about some other long forgotten common friend, even though the end user's buddy is not a CAB user.

Other OMA enablers in this category include OMA Presence [11], OMA Device Profiles Evolution (DPE) [12], and Service User Profile Management (SUPM) [13] — the latter supporting data federation.

OMA enabler specifications have been defined over a number of years, and the interfaces they expose are usually bound to a particular protocol, which supports the communication pattern desired. However, this sometimes results in a less than perfect match between the OMA enabler bindings and the propensity of the Web 2.0 development community to a particular architectural style and interaction patterns with exposed capabilities. In order to better serve the Web 2.0 development community, OMA has engaged in a comprehensive program of developing API specifications, in particular those using the REST architectural style. API specifications in support of applications that could target augmented reality now provide developers with Web 2.0 style access to telco resources such as terminal location, SMS, MMS, payment, device capabilities, presence, address book, terminal status, call control, audio call, and call notification — all included in OMA ParlayREST [14]. The Messaging API, for example, supports a unified style for sending/receiving SMS, IM, MMS, and WAP messages; the Terminal Location API allows an application to obtain a user's location and/or distance from a location; the Presence API allows a user to manage his/her presence information, content related to it, and conditions in which they can be shared with other users. A newer set of RESTful network APIs are in OMA's pipeline to support chat, file transfer, image share, video share, and address book, while the Notification Channel API is being developed to extend the accessibility to event notifications from a server-to-server environment, to device-to-server and browser-to-server environments. A Customer Profile API and an Anonymous Customer Reference API may soon follow, among others.

References

[1] Ronald T. Azuma, "A Survey of Augmented Reality," Presence: Teleoperators and Virtual Environments 6, 4 (Aug. 1997), pp. 355–85.

[2] M. Brenner and M. Unmehopa, *The Open Mobile Alliance — Delivering Service Enabler for Next-Generation Applications*, Wiley, 2008.

[3] M. Brenner and D. Wang, "Mining the Bit Pipes: Discovering and Leveraging Users' Behavior," *Proc. ICIN '09*.

[4] G. Camarillo and M. A. Garcia-Martin, *The 3G IP Multimedia Subsystem*, Wiley, 2004.

[5] K. G. Coffman and A. M. Odlyzko, "Internet Growth: Is There A "Moore Law" for Data Traffic," *Handbook of Massive Data Sets*, 4 June 2001.

[6] L. Makinen, "Mobile Videophone," *Proc. Research Seminar on Telecommun. Business*, Helsinki Univ. of Technology, Spring 2007.

[7] OMA Converged IP Messaging, v. 1.0.

[8] OMA SIP Push, v. 1.0.

[9] OMA Dynamic Content Delivery, v. 1.0.

[10] OMA Converged Address Book, v. 1.0.

[11] OMA Presence SIMPLE, v. 2.0.

[12] OMA Device Profiles Evolution, v. 1.0.

[13] OMA Services User Profile Management, v. 1.0.

[14] OMA ParlayREST v. 1.0 and v. 2.0.

Biographies

DAPING WANG (daniel.d.wang@alcatel-lucent.com) is an IMS solution architect in Alcatel-Lucent's Asia Pacific Regional Competence Center, Melbourne, Australia. His major responsibility is to provide sound IMS solutions for both fixed and mobile operators in the APAC region. He received his B.S. and M.S. degrees in applied mathematics from Jilin University, China. His previous work in Alcatel-Lucent included being a technical manager in R&D for IMS systems engineering, Lucent Softswitch, and 5ESS development. His interests span the areas of IMS application development, software metrics, and data mining technologies.

MICHAEL BRENNER (Michael.Brenner@alcatel-lucent.com) is chief architect, team leader, and a consulting member of technical staff in the Applications Standards Department of the CTO of Alcatel-Lucent's Applications Group. His major responsibilities are to lead the Alcatel-Lucent OMA and OASIS standards delegations, and function as chief architect and primary technical contact for these delegations internally and externally. In the last years, his main focus was on creating and implementing an API standards strategy consistent with Alcatel-Lucent's Application Enablement strategy, currently materializing in OMA ParlayREST. In 2007 he was the recipient of the prestigious OMA Contributor Award for Excellence, and in 2008 he co-authored the book *The Open Mobile Alliance: Delivering Service Enablers for Next-Generation Applications* (Wiley). Prior to joining the standards track, he achieved extensive software development and management expertise, first as an individual contributor and later as a director, in particular in the areas of service and network management, policy management, artificial intelligence, real-time data acquisition, and picture archiving and communication systems. He earned his combined B.S./M.S. degree in computer science from the Polytechnic University Bucharest, Romania.

In order to better serve the Web 2.0 development community, OMA has engaged into a comprehensive program of developing API specifications, in particular thoseusing the Representational State Transfer architectural style.