EXECUTIVE SUMMARY

This document presents the progress with respect to Work Package 7b.1 and 7b.2 which focuses on the use case “Personalized TV Guide with Adaptive Advertising”. This use case illustrates the design and development of a Personalized Program Guide which can, next to providing personalized content, propose additional services like Multimodal support and personalized advertisements.

This deliverable specifically deals with tasks T7b.3 (Integration of social aspects and multimodal aspects in the PPG), and continues where deliverable D7b.1 and D7b.2 ended. In this deliverable we specifically focus on the technologies to facilitate multimodal interaction and the integration with the Video-On-Demand portal, described in D7b.3.
**DOCUMENT INFORMATION**

<table>
<thead>
<tr>
<th>IST Project Number</th>
<th>Acronym</th>
<th>NoTube</th>
</tr>
</thead>
<tbody>
<tr>
<td>FP7 – 231761</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Full Title</th>
<th>Networks and Ontologies for the Transformation and Unification of Broadcasting and the Internet</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Project URL</th>
<th>Document URL</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://www.notube.tv/">http://www.notube.tv/</a></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>EU Project Officer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leonhard Maqua</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deliverable Number</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>7b.4</td>
<td>Social Interaction Tools &amp; Multimodal Support for PPG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Work Package Number</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>TV-related Use Cases</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date of Delivery</th>
<th>Contractual</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td>M33</td>
<td>31-10-11</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Status</th>
<th>Nature</th>
</tr>
</thead>
<tbody>
<tr>
<td>version 1.0</td>
<td>prototype □ report ☑ dissemination □</td>
</tr>
</tbody>
</table>

| Dissemination Level | |
|---------------------| |
| public ☑ consortium □ |

<table>
<thead>
<tr>
<th>Authors (Partner)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annelies Kaptein, Pieter Bellekens (Stoneroos), Teresa Sanghee Kim, MinJeung Cho (KT)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resp. Author</th>
<th>E-mail</th>
<th>Phone</th>
</tr>
</thead>
<tbody>
<tr>
<td>Annelies Kaptein</td>
<td><a href="mailto:annelies.kaptein@stoneroos.com">annelies.kaptein@stoneroos.com</a></td>
<td>+31 (0) 35 628 47 22</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Abstract (for dissemination)</th>
</tr>
</thead>
<tbody>
<tr>
<td>In this deliverable we report the status and progress made towards the goal of multimodal interaction in the next generation PPG</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Keywords</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multimodal interaction</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Version Log</th>
<th>Issue Date</th>
<th>Rev No.</th>
<th>Author</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>October 27, 2011</td>
<td>1</td>
<td>Pieter Bellekens</td>
<td>Version 1.0</td>
</tr>
<tr>
<td>Participant’s name</td>
<td>Partner</td>
<td>Contact</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Vrije Universiteit Amsterdam | | Guus Schreiber  
Phone: +31 20 598 7739/7718  
Email: schreiber@cs.vu.nl |
| British Broadcasting Corporation | BBC | Libby Miller  
Phone: +44 787 65 65 561  
Email: Libby.Miller@bbc.co.uk |
| Pronetics S.p.A. | (pro-netics) | TBA  
Phone: TBA  
Email: TBA |
| Engin Medya Hizmetleri A.S. | | Ron van der Heiden  
Phone: +31 6 2003 2006  
Email: ron@engin.tv |
| Institut fuer Rundfunktechnik GmbH | IRT | Christoph Dosch  
Phone: +49 89 32399 349  
Email: dosch@irt.de |
| Ontotext AD | ontotext | Atanas Kiryakov  
Phone: +35 928 091 565  
Email: naso@sirma.bg |
| Open University | The Open University | John Domingue  
Phone: +44 1908 655 014  
Email: j.b.domingue@open.ac.uk |
| RAI Radiotelevisione Italiana SPA | Rai | Alberto Morello  
Phone: +39 011 810 31 07  
Email: a.morello@rai.it |
| Semantic Technology Institute International | STi INTERNATIONAL | Lyndon Nixon  
Phone: +43 1 23 64 002  
Email: lyndon.nixon@stti2.org |
| Stoneroos B.V. | Stoneroos | Annelies Kaptein  
Phone: +31 35 628 47 22  
Email: annelies.kaptein@stoneroos |
| Thomson Grass Valley France SA | Grass Valley | Raoul Monnier  
Phone: +33 2 99 27 30 57  
Email: raoul.monnier@thomson-networks.com |
| TXT Polymedia SPA | TXT Polymedia | Sergio Gusmeroli  
Phone: +39 02 2577 1310  
Email: sergio.gusmeroli@txtgroup.com |
| KT Corporation | KT | Myoung-Wan Koo  
Phone: +82 2 526 5090  
Email: mwkoo@kt.com |
TABLE OF CONTENTS

LIST OF FIGURES 6

1 INTRODUCTION 7

2 INTEGRATION OF MULTIMODAL ASPECTS IN PPG 8
   2.1 A Keyword/Semantic VOD Search 8
   2.2 AT&T Speech Mashup (SM) Web Portal 9
   2.3 Demo Scenario 11
      2.3.1 Voice recognition engine implementation 11
      2.3.2 Video-On-Demand Web portal implementation 12

3 CONCLUSIONS 20

REFERENCES 20
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Speech Mashup Architecture</td>
<td>9</td>
</tr>
<tr>
<td>2.2</td>
<td>Example of the grammar file in XML</td>
<td>10</td>
</tr>
<tr>
<td>2.3</td>
<td>An architecture of an extended OllehTVNow</td>
<td>11</td>
</tr>
<tr>
<td>2.4</td>
<td>The OllehTVNow interface recording the user’s speech</td>
<td>14</td>
</tr>
<tr>
<td>2.5</td>
<td>Searching after the user’s input</td>
<td>15</td>
</tr>
<tr>
<td>2.6</td>
<td>Showing the auto-completion search results of the search</td>
<td>16</td>
</tr>
<tr>
<td>2.7</td>
<td>The search results of the query</td>
<td>17</td>
</tr>
<tr>
<td>2.8</td>
<td>The VOD portal link in the OllehTVNow app</td>
<td>18</td>
</tr>
<tr>
<td>2.9</td>
<td>Korean content in the VOD portal</td>
<td>19</td>
</tr>
<tr>
<td>2.10</td>
<td>An incoming request from the OllehTVNow app</td>
<td>19</td>
</tr>
</tbody>
</table>
1. Introduction

Previously, in deliverable D7b.1 and D7b.2, we extensively discussed the technologies behind personalized advertisements, multimodal interaction as well as the insertion of ads in video content. Therefore, when in need of more information we would like to refer to this document. In this deliverable, we report on the continuation with respect to the multimodal technology. While in deliverable D7b.3 we report on the progress of the multimodal technology with respect to the technological improvements in the field of multi-linguality, in this deliverable we focus on the demonstrator built to showcase this technology as well as the speech engine to facilitate the multimodal input. The demonstrator presented here and developed by kt is called the OllehTVNow app, which among others runs on the iPad. At the end of the document we will zoom in on the integration of the OllehTVNow app with the iFanzy VOD (Video On Demand) portal\(^1\), which is already extensively described in deliverable D7b.3.

The structure of this document is as follows: In Chapter 2 we discuss the practical application of the multimodal technology. Within this chapter we have Section 2.1 which concisely describes the search technology of the OllehTVNow app, Section 2.2 explaining the Speech Mashup service which interprets spoken language, and Section 2.3 which elucidates the actual demo by means of screenshots. Within this last section, we also take a closer look at the integration between the OllehTVNow app and the VOD portal. We conclude this deliverable with some conclusions in Chapter 3.

For more background on the NoTube context and how this use case fits within that context, we would like to refer the reader to previous deliverables.

\(^1\)http://adsdemo.stoneroos.nl/VOD
2. Integration of Multimodal Aspects in PPG

To demonstrate the multimodal interaction in PPGs, kt has released an iPad-version IPTV mobile service (OllehTVNow) in 2011 which contains over 5000 VODs and 30 live channels. It is currently freely available. This app is kt’s NScreen application for an IPTV service, developed as a commercial product.

2.1 A Keyword/Semantic VOD Search

Previously and outside NoTube, kt has developed a keyword-based and/or semantic-based search engine for VOD content. A keyword search retrieves documents where terms in a user query occur. It does not take into account the meanings or context of the terms in documents or queries. Terms in the documents are the main means of indexing and retrieving related documents. On the other hand, semantic search tries to improve search accuracy by understanding the searcher’s intent and contextual meanings of query terms when retrieving related documents. It uses domain knowledge called an ontology that is predefined by human experts and allows reasoning in contexts. An ontology is a formal and explicit specification of a shared conceptualization, and models a domain in terms of objects, properties and relations. kt uses the Owlim framework of Saltlux company as an underlying reasoning engine [1].

Further, kt uses a keyword search for retrieving content based on titles and a semantic search for retrieving other information such as actors, directors, plot keywords, and genres. Search results are presented with detailed information such as Korean/English titles, directors, actors, subtitles, synopses, series information, production companies, production year, producing countries, viewer rates, runtimes, images and release dates.

Semantic searches are mainly deployed for auto-completion purposes. For example, when a user submits a query of “Julia Roberts”, the auto-completion results are “Julia Roberts Movie” and “Julia Roberts Documentary”. In this case, a user can select one of these results to find corresponding search results. In this way, a user is presented with a better interface where he/she can understand what kinds of content about “Julia Roberts” could be found in our IPTV service and can quickly find the information they want.

An another example is for finding content based on semantic terms such as ‘accident’. The semantic searches then find content that are tagged as ‘accident’ and using the semantic hierarchy definition, it also finds content which are tagged with the sub-concepts of ‘accident’ such as “car accident”, “plane accident”, “crash”, etc.

These semantic searches are different from keyword searches in that instead of mainly matching a query term against content terms, we extract most the important (or representative) terms from the content even through the terms are not explicitly mentioned in the titles or synopses of content items. That is, the semantic searches mine the representative terms from titles and synopsis of the content and also from external information (e.g. user reviews).
2.2 AT&T Speech Mashup (SM) Web Portal

Within NoTube and as continuation on the search engine, kt has implemented a speech engine which allows users to search for content by talking to the device. The technology which was used to facilitate this feature consists of the Speech Mashup service library built by AT&T. This Speech Mashup service (SM) by AT&T is relatively new in comparison to speech recognizers by Google and Nuance [3]. Nonetheless, we have decided to use SM with the following reasons: 1) it allows a developer to customize a speech grammar; 2) the grammar can be defined in an XML format (i.e. which is easy to use); 3) it is a web-based; and 4) it is open source and there is no limitation for a query request. In order to use this service, a developer has to create a login-account and once login, a new grammar file can be uploaded or an existing file can be modified. More details can be found at the AT&T website\(^1\). Figure 2.1 shows the architecture of the SM.

![Speech Mashup Architecture](image)

Figure 2.1: Speech Mashup Architecture.

As shown in Figure 2.1, SM application consists of three main components that enable clients on mobile device to connect to AT&T speech servers:

- SM Manager opens and manages direct connections to the appropriate AT&T speech servers on behalf of the client, including resolving device dependency issues, and performing authentication and general accounting.

- A speech mash-up client that relays audio (using HTTP) to the WATSON servers and accepts the recognition result.

- AT&T speech servers, including WATSON servers configured for ASR, and Natural Voices servers, which converts text to speech and returns streaming audio back to the web application.

\(^1\)https://service.research.att.com/smm/login.jsp
kt has used AT&T’s ASR Service for finding content by titles and semantics and in doing so, a rule-based grammar file needs to be defined. In Figure 2.2, an example of the grammar file in XML is shown.

```xml
<grammar version="1.0"
    root="object"
    tag-format="semantics/1.0">
    <rule id="iptvVODSearch">
        <ruleref uri="#preamble"/>
        <ruleref uri="#object"/>
        <tag>
            out = rules.object;
        </tag>
    </rule>

    <rule id="preamble">
        <item repeat="0-1">
            <one-of>
                <item>
                    find me
                </item>
                <item>
                    I want to
                </item>
                <item>
                    give me
                </item>
            </one-of>
        </item>
    </rule>

    <rule id="object">
        <one-of>
            <item>Dong yi</item>
            <tag>out="DongYi"/></tag>
        </item>
        <item>Dongyi</item>
        <tag> out="DongYi" </tag>
    </one-of>
</rule>
</grammar>
```

Figure 2.2: Example of the grammar file in XML.

Using a rule-based grammar format, it is possible to include variations for a given item by adding the string `tag-format="semantics/1.0"` in the first `<grammar>` line, i.e. `<grammar tag-format="semantics/1.0" root='object'>`. The `<tag>` elements are used for this. For example, the same string ‘DongYi’ is returned depending on whether the speaker says ‘Dong yi’ or ‘Dongyi’. Currently, the grammar file has 87 titles and 12 semantic words. Figure 2.3 shows an overview of the extended OllehTVNow application.
2.3 Demo Scenario

This scenario provides details about the collaboration between Stoneroos and kt, in particular for a multimodal integration interface. As previously described, the OllehTVNow app provides different ways to find content. E.g. he or she starts the OllehTVNow iPad app and searches for a video using its title, either using a text or speech interface. Whenever new content is selected many details and extra information (e.g. rating) can be asked for. To illustrate, when a user utters “Winter Sonata” (i.e. a famous love story) as shown in Figure 2.4, in Figure 2.5 the system is busy searching, the auto-completion results are shown in Figure 2.6 and its search results are shown in Figure 2.7. The details include running time, the number of viewers, and ratings.

2.3.1 Voice recognition engine implementation

We have described an iPad application that enables both text-based and voice-activated searches and lets a user watch a chosen VOD on a larger TV screen with other people. The main objective is to demonstrate the integration of a multimodal interface into an existing VOD portal. A speech interface is known to be more useful on mobile devices as it is often difficult to type search words using a tiny keyboard on mobile devices [2]. Kt has developed a speech recognition engine for Korean, but have not
worked for an English speech recognizer. For our purpose, we need an English speech recognizer which can be embedded into an iOS platform and configured specially for an IPTV domain rather than a general-purpose recognizer. In Section 2.2, we have mentioned some reasons for choosing AT&T speech recognizer over others. We informally evaluated the performance of the AT&T Speech Mashup engine with that of Google’s voice-activated search. Where that Google recognizes normal English nouns (e.g. Policeman) better than that of AT&T, for uncommon English words Google is not very suitable. For example, it is often the case that titles of Korean contents are literally translated into English without taking into account the meanings. As a general-purpose engine, Google is not good at recognizing such words and therefore we needed a recognizer which could be customized. As explained in Section 2.2, the AT&T recognizer allowed a developer to submit his/her own grammar file which enables the recognizer to be domain specific.

2.3.2 Video-On-Demand Web portal implementation

Further within work package 7b, we have implemented a Video-On-Demand (VOD) Web portal running on top of the iFanzy architecture. This Web portal is just one of many iFanzy interfaces running on top of the back-end server (next to the regular iFanzy Web interface, the iPhone app and the set-top box applications). The start page of this portal shows a set of favourite movies (when no user is logged in, we see the movies which are popular among all users) through which a user can browse. Every item has some accompanying metadata fields such as duration, description, director, actors and a rating. In Figure 2.9, we see this portal filled with Korean VOD data. The VOD repository is connected to the VOD database which is also available in the OllehTVNow application.

This integration between the OllehTVNow app and the Stoneroos VOD Web portal was established to accommodate users with a larger screen display. So whenever a user uses the OllehTVNow app to search interesting content and selects a specific item, a link is provided to the VOD portal. In Figure 2.8 we see the detail page for the program “Winter Sonata”. Underneath the descriptive text we see three action buttons. The first is the play button to play the content on the iPad, the second is to add a user’s rating and the third button (labeled as ‘Stoneroos’) sends a play request to the Stoneroos’s VOD portal server. When this request arrives at the server and the user is logged-in, it displays a message on the screen asking the user whether he wants to play the content he just selected on the iPad app. In Figure 2.10 we see this message popping-up to play the VOD content called “High Kick” during the play-out of another piece of content (“Big Buck Bunny” in this case). When the user confirms, the current video play-out is stopped and the newly selected content starts playing (after paying if applicable).

Technically, the integration was facilitated by means of the HTTP long polling technique, a type of push technology. Long polling is a variation of the traditional polling technique and allows emulation of an information push from a server to a client. With long polling, the client requests information from the server in a similar way to a normal poll. However, if the server does not have any information available for the client, instead of sending an empty response, the server holds the request and waits for some information to be available. Once the information becomes available

\[http://adsdemo.stoneroos.nl/VOD\]
(or after a suitable timeout), a complete response is sent to the client. The client will normally then immediately re-request information from the server, so that the server will almost always have an available waiting request that it can use to deliver data in response to an event\(^3\).

In this demo the VOD portal Web application sends a request to the server, asking whether it needs to play a specific video. The server answers negatively, but keeps the connection open. At some point when the kt iPad app sends a play request to our VOD server, the server forwards the request to the VOD portal Web application. The Web app receives this request and starts playing the correct video after confirmation by the user (like shown in Figure 2.10).

\(^3\)http://en.wikipedia.org/wiki/Push_technology
Figure 2.4: The OllehTVNow interface recording the user’s speech.
Figure 2.5: Searching after the user’s input.
Figure 2.6: Showing the auto-completion search results of the search.
Figure 2.7: The search results of the query.
Figure 2.8: The VOD portal link in the OllehTVNow app.
Figure 2.9: Korean content in the VOD portal.

Figure 2.10: An incoming request from the OllehTVNow app.
3. Conclusions

In this deliverable we presented the progress in Work Package 7 which focuses on the use case “Personalized TV Guide with Adaptive Advertising”. This particular deliverable deals with tasks T7b.3 (Integration of social aspects and multimodal aspects in PPG). Although this task also mentions the social aspect, this deliverable focuses for the largest extent on the multimodal aspect.

With respect to the evolution towards a multimodal PPG environment, we have presented a nice iPad demonstrator (which is actually almost market ready). In OllehTVNow, originally a Korean demonstrator, VODs can be searched either by keyword-based or semantic-based queries, by understanding the searcher’s intent and contextual meanings of query terms when retrieving documents. Further, this document reports on the use of the Speech Mashup service (SM) by AT&T which is a relatively new (but free!) speech recognizer.

Lastly, this document shows a nice overview of the iPad OllehTVNow app as well as the integration of it with the iFanzy VOD portal. The need, or at least the possibility, for a big screen play-out service will continue to be necessary for the future generations of TV consumers.
REFERENCES

