Position Paper for W3C Workshop on Rich Multimodal Application Development

Authors: Wei-Yun Yau (Institute for Infocomm Research, Singapore), Sheau Ng (NBC Universal)

Title: Multimodal Interaction and TV

TV content source has expanded rapidly providing viewers with large options such as from the cable, free to air broadcast, over-the-top and internet content. On top of that, users are also engaged in their social network sites. In addition, using a second screen, they could generate new content in the form of new updates, comments, reviews etc. For example, there is news coverage on the Boston bombing incidence being broadcasted on a cable network. Similar news but from different angle of coverage is also being shown at an internet video site. At the same time, viewers at Boston area also provide live comments on the latest happening on their social network sites while some who caught live actions on video camera could post the clip on Youtube. The question is then how all the above can be easily presented to the viewer such that the user don't have to keep changing channels and sources or having to google to find out the latest happenings based on the user's own preference or inclination such as to know what is happening near to the home of a friend or relative, what happen to the marathon participants or to follow a group and provide the user's comments or concerns of interest.

The multimodal interaction framework could be harnessed to address the above needs. Some of the requirements are:

1. How to provide effective personalised content recommendation that takes into account popular or "hot" external event happening and social recommendation in addition to the traditional collaborative filtering approach based on the number of people who have consumed the content and their usage history associated to it?
2. How to aggregate the various input sources of audio, video and text in a coherent and synchronized manner in order that the content can be further curated in a more automated fashion and personalised to suit the preference of a particular user?
3. How the user's preference can be better modelled through tracking the interaction and consumption pattern on the various platform such that a more holistic model of the user can be obtained?
4. How the content can be better analysed automatically by harnessing the text, speech and video analytics in such a way that accurate information can be extracted to represent the content at fine level instead of the entire multimedia content (eg. entire video)? This would then allow the content to be better related to other similar content, removing the duplicates and providing summary of different views?
5. How is the privacy concerns be mitigated while new business case evolve to harness the above?

We have investigated some of the above questions. Primarily, the approach includes:

1. Video content analysis combining text, speech recogniton and video analysis to extract metadata from the content and correlating with text content to improve the accuracy of information such as names of places and people and to find similar content.
2. Personalised recommendation using weighted preference model from 5 sources, namely external event, usage log analysis, social recommendation, user preference and editing to the recommended content.

This position paper seeks to provide input and solicit discussion on the relevance of the MMI framework for the new consumption model of the video content in general and TV in particular.