

# Speech-Enabled Video Service using VoiceXML

---

Anthony Lam

## Abstract

As technology advances, handheld devices can now support a much wider range of functionalities. Phones, PDAs, video cameras and mp3 players – all of these have been separate devices, but they are now converging to the new generation of multi-purpose devices. At the same time, network consolidation is taking place. Mobile, telephone, and video services are migrating to the IP infrastructure. All of these changes are re-shaping how services will be provided and are creating new opportunities for new application developments.

This report analyzes various approaches for developing next generation video applications. In particular, it investigates the possibility of applying speech to video applications. It proposes how next generation video services can take advantage of the IP infrastructure and VoiceXML, and be provided in a more interactive and user-friendly fashion.

The VoiceXML approach bases upon IVR (Interactive Voice Response) technologies and is compared against the HTML-based web model for developing and delivering video applications. The key evaluation factors include usability, technological adoptability and feasibility, maintainability and ease of development, scalability, and support for video delivery. Various technologies and protocols including HTTP, SIP, RTP and IMS are also analyzed in details for understanding how they fit in the big picture for providing next generation video applications and solutions.

A sample video application, Movies Finder, is developed using VoiceXML to support the analysis and evaluation. It is concluded that the VoiceXML approach enhances the users' experience by enabling speech support, while the development cost is comparable to that of traditional web applications. Various extensions such as Visual TTS, CCXML, RTSP and multimodality are also proposed to improve the capability of the system.