

# Multimodal Application Using XHTML+Voice

---

Anthony Sheng

## Abstract

As Web technology advances, we have seen lots of browser-based visual applications, such as web mail, online auctions and online movie. Most of these applications only accept mouse, keyboard, or keypad inputs. These interfaces are not so efficient for people who have some handicap like vision impairments. Even for normal people, extra input method could be useful in situations when they are in hands busy environment. During the same time period, growth has occurred in the number of handheld devices, including cell phones, PDAs and other access interface running web browsers. As the wireless networks improve to support new devices, which become much smaller, we face the challenge that traditional input methods of old web applications are not always efficient enough to be used by end users. So other means of input in addition to keyboard, mouse or pen input are becoming necessary. This has led to a new application technology called multimodal, which provides multiple methods of communication between the user and a device.

This report analyzes approach for developing multimodal applications by using XHTML+Voice and illustrates the structure and contents of an XHTML+Voice multimodal application. It investigates the possibility of applying speech to common desktop and mobile applications.

Three sample multimodal applications including voice-enabled album, voice-enabled clock and voice-enabled TV control are developed using XHTML+Voice to support the analysis and evaluation. Detail of the applications is described and source code is provided.