

white paper

MULTIMODALITY ON THIN CLIENTS

A CLOSER LOOK AT CURRENT MOBILE DEVICES AND
THE MULTIMODAL EXPERIENCE POSSIBLE TODAY

WRITTEN BY: **SUNIL KUMAR** (DIRECTOR, RESEARCH)

V-ENABLE

Preface

This white paper is intended for wireless operators and infrastructure vendors interested in understanding V-Enable's view on how multimodality works on 2G/2.5G devices with thin clients such as WAP, BREW or JAVA.

Multimodality can be introduced in the network in two ways: sequential and simultaneous modality. This white paper will address both categories of multimodality and the user experience on mass market mobile devices, i.e. cellular devices that are widespread in the network.

The purpose of this white paper is to position V-Enable's unique multimodal solution against that of its competitors and showcase solutions in the operator network that work very effectively.

The white paper is based on V-Enable's current experience with operators' trials/deployment processes and our expertise with all major wireless standards, including WAP, SMS, IOTA, MMS etc...

Trademarks and Permissions:

V-Enable and other V-Enable trademarks are the trademarks or registered trademarks of V-Enable, Inc.

All trademarks or registered trademarks are properties of their respective owners.

The contents of this document are subject to revision without notice due to continued progress in methodology and design.

V-Enable shall have no liability for any error or damages of any kind resulting from the use of this document.

© 2003 V-Enable, Inc. All rights reserved.

V-Enable Inc.
4250 Executive Square, #200
La Jolla, California 92037
USA

Executive Summary

Multimodality is the next step in the evolution of data and speech services. The objective of this new enabling technology is to increase the adoption of current data and speech services in the network, and to also leverage the operators' investment in the current infrastructure by increasing the utilization.

With the introduction of V-Enable's Multimodal technology in the network, existing services are enhanced and made more useable by offering the end user the choice of multiple modes such as speech, text, graphics, etc... The huge benefit is that the applications themselves do not have to be modified, they are simply enhanced by dynamic rendering by the multimodal engine.

As the W3C continues standardization efforts towards multimodal interaction, this technology will be ubiquitous in the network and future 3G data services will be built around a more natural, multimodal experience to the end user.

Introduction

Multimodality allows users to access wireless information in the most natural way based on the context of the user and the content type. The user can now input with speech, touch pad, mouse etc... and receive information in synthesized speech, video, text, animation etc...

The type of device used to access multimodal information is vitally important. Based on the device capability the change of modality may introduce delays if the device/network used is not capable of running voice and data session simultaneously. The devices such as mobile phones fall into this category where the user has to switch sessions in order to experience multimodality. The experience could be seamless if the device/network is capable of running simultaneous voice and data session. The devices such as pocket pc, desktop and some of the upcoming 3G devices fall into this category. The device not capable of simultaneous voice and data are only capable of *sequential modality* where the user has to switch modes between voice and data sessions.

The earlier version of mobile devices with browser only capability has forced developers to adapt to sequential modality. Sequential modality is the first step towards multimodality on mobile phones and is quite successful in simple applications such as Driving Directions, Email, and Directory Assistance etc. but has failed to convince the user that it could be used for rich applications such as Airline Reservations, Entertainment etc. The network plays an important role and the users experience depends on the response time of the network.

The advent of thin clients with BREW/JAVA capability has taken multimodality a step ahead. Thin clients provide the necessary

capability to execute sequential modality with better control. Initial sequential definition of multimodality on 2/2.5/3G devices is now evolving towards simultaneous multimodality. Further sections will discuss how thin client devices have transformed the state of multimodality.

Browser Only Client

The most common devices that will benefit from multimodality are mobile phones with a WAP browser on it. Such devices form the mass of wireless users with initial estimates of over 1B such devices. The browser only mobile phones have few limiting factors that are impediments to multimodality and are worth mentioning here before we proceed.

- No software can be installed on the client.
- Using the WAP browser for accessing wireless data and placing a voice call cannot be done at the same time.
- Disconnecting the data browser and then starting a voice call or vice-versa introduces latency.

V-Enable has developed its multimodal technology such that any device with data and voice capability can experience Multimodality. The browser only client can experience a limited form of multimodality known as sequential modality. *“The sequential modality is one where a user can experience multiple modes, but only one mode at any particular instant.”* The user of such a modality will have to switch between modes in order to access a multimodal

application such as Driving Directions; where the user inputs in voice mode and uses visual (data) mode to see the output. The latency between switching of modes may hinder the usage of complicated applications that needs alternate switching of modes. Hence it is essential for the application developer to write applications that don't switch often and thus leverage the resources efficiently. Badly designed applications and unnecessary switching of mode will not take sequential modality anywhere. One of the examples of a good design would be the Driving Directions application where user enters source and destination and gets the direction to the place on mobile screen. The Driving Direction application presents a good case for supporting sequential multimodality. The driving direction application is incomplete if it runs in unimode (either data mode only or voice mode only). The data only mode is cumbersome while entering source/destination information because of the mini keypad used in the mobile phones. The voice only mode solves the problem of input but presents problems of comprehension if the directions are too big and you cannot listen to a picture (MAP). A good design of sequential multimodal driving direction application would be to accept the input (source/destination) in voice mode but present the output (directions) in data mode. This application would only require one switch from voice to data but would increase the usage of the application many folds. Numerous applications

can be redesigned to make use of sequential multimodality.

The important factor in designing an application is the switching latency that user has to face. The latency is directly proportional to the performance of the carriers network. V-Enable has done extensive work with reducing the latency in the network. The numbers below are based on various carrier networks that include Sprint, Verizon, AT&T, Nextel, T-Mobile, Vodafone, Orange, SKT, NTTDOCOMO etc...

	Worst	Best	Average
Voice to Data	>10 sec	1 sec	4 sec
Data to Voice	>10sec	4 sec	5 sec

The latency numbers in above table indicate fairly acceptable switching times on an average basis. The user while using a driving direction application can wait for 3-4 seconds to receive the directions on mobile screen in data mode after providing the input in voice mode.

Everyday applications such as Driving Directions, Email, Directory Assistance etc. could benefit from simplicity of usage with sequential multimodality.

Intelligent Thin Clients

The next generation of devices has intelligence where the user can download software and does not have to depend on pre-installed manufacturers' browser to present information visually. Carriers have

recently launched these 2.5/3G devices in abundance and they are rapidly penetrating the subscriber base. However most of these devices belong to CLASS B category, which are not capable of running voice and data sessions at the same time. These devices pose few problems that earlier browser only devices also had, but provide higher control to the application developer while writing applications.

Advantage that thin devices bring with them:

1. The application can reside locally on the phone thus reducing the latency involved in fetching the application from the server.
2. Data session can be suspended when a voice session starts, so the data session is resumed not initiated again once the voice session has ended. This feature reduces the time required to restart the data session unlike in browser only client where the server has to send a message to the mobile phone in order to start the data session or user has to manually start the data session by starting the browser.
3. Further the developer can write applications for the phone that can initiate voice session using standard TAPI. Earlier, the voice session was started using features available within the browser.
4. Higher network bandwidth for data operations.

- 5. The developer can use the VoIP solution for making voice calls and still experience simultaneous modality.

The multimodal solution on thin clients can be presented in two different ways. The first way is the traditional sequential approach that uses existing browsers on the device. This scenario is similar to the 2G devices except that 2.5/3.0 will bring more data bandwidth for the user.

Another way to achieve multimodality will be to define a multimode client software residing on the mobile phone. The client software is responsible for making all the communication with the server that is earlier done by the browser. The client is also responsible for presenting the data on the mobile screen. Thus this solution bypasses the browser and uses phone API (e.g. JAVA/BREW) to present information on the phone.

The capability to download small clients on thin devices has allowed developers to provide better sequential modality experience to the users. The earlier browser only clients allowed the design of multimodal solutions based on server side sequential modality. The part of the server only solution can now be implemented on the client,

which helps in better control over the interaction and reduces the network latency.

The thin clients should not be confused with the power of a palm device, iPAQ etc... The processing power provided by thin clients is limited and has to be used sparingly. So it is important to carefully design the distributed architecture of a multimodal solution so as to put

limited CPU cycles on the client such as browsers and speech encoding whereas the CPU intensive software still remains on the server side such as speech recognition, VoiceXML browsers. All the events generated visually with the help of touch pad clicks, soft key clicks are locally processed at the client side however the voice initiated events are processed remotely by more powerful machines at the server end.

An intelligent thin client with JAVA/BREW capability enhances the multimodal experience with reduced latency and better interaction.

