



Multimodal Experience from Every Handheld

A Position Paper

Ewald Anderl

Inderpal Singh Mumick

CTO and VP, Kirusa

CEO and Founder, Kirusa

anderl@kirusa.com

mumick@kirusa.com

For the W3C Multimodal Workshop



July 2004
Sophia Antipolis, France

Section I. Introduction

Today's mobile services are limited by user interfaces that allow for voice or visual interaction, rather than an integrated voice and visual interface. Usability problems unique to mobile devices have prevented service and information providers from meeting their consumer wireless data usage and revenue goals/targets. Voice interfaces allow for simplified service input and interaction, but are not practical for conveying graphics (maps) or memory straining information such as long lists (example, lists of messages), complex instructions, or numbers. Conversely, visual interfaces are cumbersome to use but excellent for information output.

By combining both voice and visual interfaces, applications can exploit the strengths of each and minimize the weaknesses of both modes. This integration, known as multimodality, seamlessly combines speech, text and touch input with graphics, text and audio output to deliver enhanced end user mobile applications and services. Multimodal applications such as voice SMS enhance service usability resulting in improved customer satisfaction and usage, and can increase service revenues.

Multimodality has been an active research area. While the research needs to continue, we also need to develop and promote technologies and products that will take multimodality from a research topic, into common use on every one of the 1.5 billion handheld devices around the world.

Kirusa provides production-ready multimodal platform and multimodal application products on a variety of different end-user devices, from common SMS mobile phones to full featured smart devices and PDAs, and everything in between. Installed at several leading mobile carriers and enterprises, Kirusa has products that target next generation GPRS, EDGE, 3G, WiFi, and Flash-OFDM networks, as well as products that deliver multimodality on the widespread 2G networks.

In this position paper, we provide a brief description of how Kirusa is working to build the technology to bring multimodality to every mobile handheld device. We discuss the current solutions, as well as some of the areas of future work needed for widespread adoption of multimodal solutions.

Section 2. Kirusa Solutions

2.1 Overview

Kirusa today builds and markets three products for deployment of multimodal solutions by mobile carriers and enterprises.

- KMS-IPV, Kirusa Multimodal Solution for IP Voice, for delivery of a multimodal experience on smart devices over 3G and 2.5G mobile networks.
- KMS-CV, Kirusa Multimodal Solution for Circuit Voice, for delivery of a multimodal experience on all devices over 3G, 2.5G, and 2G mobile networks.

- KV.SMS, Kirusa Voice SMS Application, a packaged multimodal messaging application for all devices over 3G, 2.5G, and 2G mobile networks.

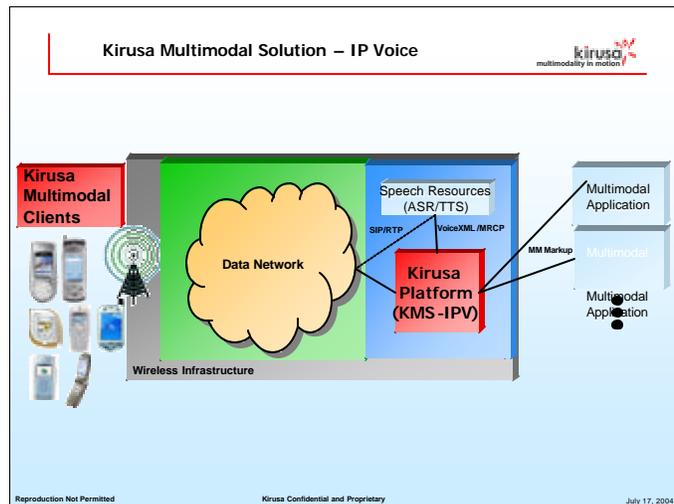
The first two products enable delivery of a multimodal user interface for any application to a mobile handset.

VoiceXML, ASR, and TTS

The KMS-IPV, KMS-CV, and KV.SMS work with a wide variety of VoiceXML 2.0 compliant voice platforms, ASR engines, and text to speech products. SALT can also be supported.

2.2 Kirusa Multimodal Solution for IP Voice

Kirusa’s multimodal solution for IP Voice is a simultaneous multimodal solution that allows any application to be accessed using a combination of voice and visual modes at the same time. The simultaneous multimodal solution consists of a multimodal platform (KMP-IPV) and multimodal clients (KMC-IPV) for mobile handsets, as shown in the adjacent architecture diagram.



The multimodal platform is the “heart” of the multimodal solution. Functioning like a “multimodal gateway,” the platform ties together the activities of the client, application, and the voice browser to realize a multimodal session based on the multimodal markup language. The voice and data connections run simultaneously over an IP connection between the platform and the client. This mechanism enables KMP to utilize a variety of wireless networks including GPRS, Edge, UMTS, CDMA 1X/3X, EV-DO, Flash-OFDM, and 802.11x. The platform is optimized to provide fast response time over low bandwidth wireless networks, such as GPRS.

The Kirusa Multimodal Client is device-resident software installed via a quick and simple download process. The multimodal client interfaces with the native visual and voice resources of the smart handheld device. Multimodal clients are available or under development for several smart devices, including Symbian, MS Smartphone, J2ME, Brew, and PalmOS, MS PocketPC, and MS PocketPC Phone Edition.

2.3 Kirusa Multimodal Solution for Circuit Voice

Kirusa’s multimodal solution for Circuit Voice maintains two separate connections with the mobile device, a circuit connection for voice, and a second connection for data. The solution consists of a multimodal platform (KMP-CV), and supports SMS, sequential, and simultaneous multimodality, depending on the capabilities of the device and network. No clients are required for this solution. The architecture is as shown in the adjacent figure.

2.4 Kirusa Voice SMS Application

The Kirusa Voice SMS (KV.SMS) application is a multimodal application that integrates voice messaging capability with text based SMS and multimedia based MMS. Based on patent pending technology, the application is designed for use over existing 2G and GPRS networks and handsets, as well as over the next generation EDGE and 3G networks and handsets. The application enables operators to launch easy to use, next generation messaging services quickly and cost effectively without needing to upgrade or replace their current infrastructure. Operators can choose a variety of advanced messaging services, tailored to the needs and interests of the youth, enterprise and mass-market segments.

The Kirusa Voice SMS application allows several options for users:

- The user can initiate sending an SMS message using a voice message.
- The user can send a RingFreeSM voice message, without ringing the phone.
- The user can send a message to one person, or to a group of people.
- Using patent pending Click2ListenSM technology, the user can click on an SMS message to hear the voice message.
- The user can respond by voice, or by text to an SMS message or to a voice messages.
- On MMS handsets, the service integrates with MMS.

Section 3. Future Work

It is our position that within 3-5 years, all applications used from mobile handsets will be multimodal. Needless to say, realization of this vision requires advances in networks, handsets, multimodal technologies, and emergence of appropriate business models. Initial successes are needed to show the value to all participants in the multimodal ecosystem.

We list some of some of the areas where we see focus in the future:

1. **Development of high speed wireless networks.** UMTS and EV-DO networks are beginning to be deployed across the world, providing bandwidths in excess of 150 kbs to mobile users for the first time. Flash-OFDM networks further increase the bandwidth to 1.5 Mbs, approaching broadband speeds. 802.11x networks are becoming commonplace, and the first WiFi handsets are being introduced this year. All these networks provide data speeds at which it is possible to deliver an extremely compelling simultaneous multimodal experience to the user. Multimodal technologies and applications must be developed to take advantages of the full range of networks and devices soon to be available.
2. **Development of smart devices.** Smart devices based on SymbianTM, MS SmartphoneTM, PalmOSTM, BrewTM, and JavaTM are being introduced at an increasing rate by device manufacturers and carriers around the world. Several of these are targeted at enterprise customers. Clients need to be developed for the various smart device platforms to take advantage of their capabilities for multimodality. At the same time, the phone platforms need to build support for multimodality from the ground up.
3. **Support of multimodality on non-smart devices.** While it is tempting to focus on only the newer smart devices when developing multimodal solutions it is projected that less

than 20 million smart devices will be sold in 2004, and, in 2008, only 30-40% of the devices sold will be smart devices. In order to bring a multimodal experience to every handheld, it is equally important to develop multimodal technologies that supports non-smart devices.

4. **More than vocal and visual modalities.** While the early work has been on combining the vocal and visual modalities, multimodality incorporates other modes of input, such as ink and gestures.
5. **Advances in speech recognition and synthesis.** Speech recognition technologies needs to be more robust, and more tolerant of noise from mobile devices. Distributing speech recognition and speech synthesis between the server and devices should be explored with the goals of improving the user experience.
6. **Standardization of mark-up languages.** The effort to standardize the multimodal mark-up language at the W3C is a critical step for the widespread adoption of multimodality. The availability of standards will also facilitate the availability of multimodal authoring tools.
7. **Applications, and more applications.** As an industry we need to develop applications that highlight the value of multimodality to a broad range of users, while providing a clear revenue model or cost savings for the carriers or enterprises.
8. **Leadership role by wireless carriers.** As the channel to 1.5 billion customers around the globe, it is hard to imagine a pervasive multimodal ecosystem without the wireless carriers playing a leadership role in delivering multimodal applications to their subscribers.

Section 4. Conclusions

Kirusa has developed core multimodal technologies and applications that provide a multimodal experience optimized to the full range of mobile handsets, from SMS enabled handsets on 2G networks, to smart devices on 3G networks, to Flash-OFDM enabled devices. As early adopters of multimodality, several leading wireless carriers and enterprises, including France Telecom and Vodafone have embraced Kirusa's products.

Further advances in multimodal technologies will continue. We have identified several areas of work we believe are important for widespread availability and adoption of multimodality. Kirusa continues to invest and be at the forefront of multimodal technology in partnership with carriers, enterprises, and technology providers.