

# SISTEMA DI INTERAZIONE MULTI-MODALE PER DISPOSITIVI MOBILI

*Daniele Falavigna, Luca Nardelli, Marco Orlandi\**

ITC-IRST, Via Sommarive 18, Povo (TN)  
Waycom S.R.L., Via Borace 54, Albisola Superiore (SV)\*  
([falavi@itc.it](mailto:falavi@itc.it), [lunarde@itc.it](mailto:lunarde@itc.it), [morlandi@waycom.it](mailto:morlandi@waycom.it))

## RIASSUNTO

Gli ultimi anni hanno visto il rapido sviluppo e la grande diffusione di dispositivi mobili quali PDA (Personal Digital Assistant) e smartphone; questi ultimi, in particolare, offrono un buon livello di connettività (GPRS, WiFi, Bluetooth, UMTS), processori potenti (StrongARM, XScale, ecc) e sistemi operativi multi-tasking (Windows Mobile 2003, Symbian O.S., Linux).

Il problema principale nell'utilizzo di tali dispositivi riguarda ancora oggi la difficoltà d'interazione con i servizi, a causa delle ridotte dimensioni del display e dell'impiego di tastiere minuscole e scomode.

Il sistema per il browsing multi-modale (MBS) sviluppato in ITC-irst [1] si propone come possibile alternativa per la realizzazione di applicazioni che rendano semplice e intuitivo l'utilizzo di questi dispositivi, grazie alla possibilità di interagire sia vocalmente, sia attraverso le consuete periferiche di input (tastiera, pennino, ecc).

Il sistema si basa sulla sincronizzazione temporale di documenti HTML, che definiscono il layout bidimensionale (visuale) e di documenti VoiceXML, che definiscono il layout temporale (vocale).

Tale scelta garantisce agli sviluppatori la possibilità di realizzare in modo efficiente un servizio multi-modale rispettando alcune semplici regole di allineamento tra i documenti HTML e VoiceXML; non vengono inoltre limitate in alcun modo le potenzialità offerte dai due linguaggi di mark-up, dato che lo sviluppatore è libero di decidere quali campi allineare. L'architettura è di tipo server-side, nella quale tutte le componenti software (Automatic Speech Recognizer, ASR, Text To Speech, TTS, ecc) risiedono nella parte server, mentre al client non è richiesto alcun software aggiuntivo. Questa soluzione impone però la presenza di connessioni a banda larga, in quanto sia il segnale vocale acquisito sul client, sia i prompt vocali inviati dal server devono essere trasmessi attraverso la rete di comunicazione. L'allineamento dei documenti HTML con quelli VoiceXML viene realizzato attraverso due connessioni socket instaurate tra client e server: un canale di sincronizzazione, sul quale vengono scambiati dati e comandi al fine di realizzare l'aggiornamento dei documenti e un canale vocale, sul quale vengono trasferiti i segnali audio. Quando l'utente fornisce un'informazione a voce il segnale viene inviato al server, il quale identifica la coppia *campo-valore* che sarà in seguito utilizzata per aggiornare il documento VoiceXML; in un secondo momento il sistema MBS richiederà, attraverso il canale di sincronizzazione, l'aggiornamento del corretto campo nella corrente pagina Web. A seguito di un'interazione visuale, ottenuta ad esempio digitando il proprio nome in un campo di testo, il client invia al server MBS le informazioni (*campo-valore*) necessarie per l'aggiornamento del corrispondente documento VoiceXML. L'esempio chiarisce un'altra fondamentale caratteristica del sistema, la trasparenza: l'utente non deve, infatti, notificare al sistema la modalità di interazione scelta.

Il sistema MBS è stato interamente realizzato nel linguaggio di programmazione Java e quindi risulta portabile su differenti piattaforme e permette inoltre di distribuire su diversi PC le principali risorse utilizzate: web server, ASR, TTS, server MBS. Durante la presentazione del lavoro sarà dimostrato il funzionamento del sistema di browsing multi-modale, sviluppato in ITC-irst, attraverso un prototipo che prevede l'interazione tramite uno smartphone Sony-Ericsson P900.

[1] L. Nardelli, D. Falavigna, M. Orlandi: "A Multi-Modal Architecture for Cellular Phones", ICMI 2004, State College, Pennsylvania, USA, October 14 - 15, 2004.