

Giving voice command to lifeless applications

By Sufia Tippu/TNN

Bangalore: A neurosurgeon looking at the CAT scan of a patient does not click for different views on the screen. He vocally asks the machine to give him either the anterior or the posterior view of the brain to be scanned.

A stockmarket trader doesn't click on his PDA to place a buy or a sell order. He speaks to his PDA screen and the transaction is executed without him going through multiple clicks or interactive voice recording (IVR).

Called multimodal access, this is about including the voice component in what is essentially data-oriented applications. They are becoming critical day by day with businesses and customers lapping it up in every form possible.

Multimodal technologies combine multiple modes or channels in the same interaction or session. The methods of input include speech recognition, keyboard, touch screen, and stylus. Depending on the situation and the device, a combination of input modes will make it easier to use a small device.

According to a survey by leading service operators in the US, every time you click and wait for a transaction to be executed, you lose 50% of your clients. This is seen to be fuelling the excitement in companies right from IBM to independent providers of mobile Internet infrastructure platforms like Openstream, a New Jersey-based product company whose mobile messaging platform has been embedded with IBM's WebSphere. Openstream has a development centre in Bangalore which has been working on these technologies for some time.

Says Raj Tumhuri, CEO, Openstream Technologies, "Human interaction is basically a multimodal form of interaction. When you want to say something, you use gestures and speech. Until now, technologically, this was not possible but now we are seeing significant action in this space."

Recently, Motorola, Opera Software ASA and IBM submitted a proposal to the W3C for a multimodal markup language standard called XHTML+Voice (X+V for short) that provides a way to create multimodal Web applications.

Brian Garr, program director and segment manager for contact centre solutions at IBM Software Group, was recently quoted as saying: "If you look at the way speech technologies have evolved over the last 10 years, you will see that is something like the wild wild west.... Everybody had their own speech, their own proprietary methods, their own proprietary connectors, and then standards started to kick in and the world changed." So, speech, the first step marking the rise of the human civilisation, seems to be the last frontier for the world wide web.

