



e-business



Research



Human Language Technologies

Multi-modal Web IBM Position

W3C / WAP Workshop



Mobile Speech Solutions
&
Conversational AdTech

Stéphane H. Maes

smaes@us.ibm.com

TV Raman



Definitions by example: evolution of a stock trading application

Present day

VoiceXML
or
IVRs

What stock or mutual fund please?

Price for XXX is \$104.
Do you want another transaction?




Different Applications - Possible shared back end
Each service is implemented separately for each access channel

WML

Enter Ticker Symbol:

Submit



HTML



Welcome to \$\$Trade:
[Online Banking](#)
[Real Time Stock Quotes](#)
[Trades](#)

Multi-Channel

Welcome to \$\$ Trade
Please select one of the following options:
Online Banking, Real Time Stock Quotes, Trades



Same Application - Multiple Channels - One Mode per Channel
Each service can be exposed via any access channel based on business needs.

Welcome to \$\$Trade:
Select:
• Online Banking
• RT Stock Quotes
• Trades
Submit



Welcome to \$\$Trade:
[Online Banking](#)
[Real Time Stock Quotes](#)
[Trades](#)

Multi-Modal



Welcome to \$\$ Trade
Please select one of the following options:
Online Banking, Real Time Stock Quotes, Trades

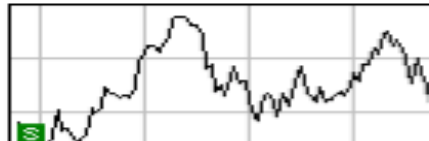
Same Application - New channels combine multiple modes
Application and user can use any available mode for input or output at any time

Welcome to \$\$Trade:
[Online Banking](#)
[Real Time Stock Quotes](#)
[Trades](#)

&

Price for XXX is \$ 104. Do you want another Transaction?

Price for XXX is \$104
[Another Transaction](#)



Trading Application Scenario Evolution

Conversational

This includes
Multi-Modal

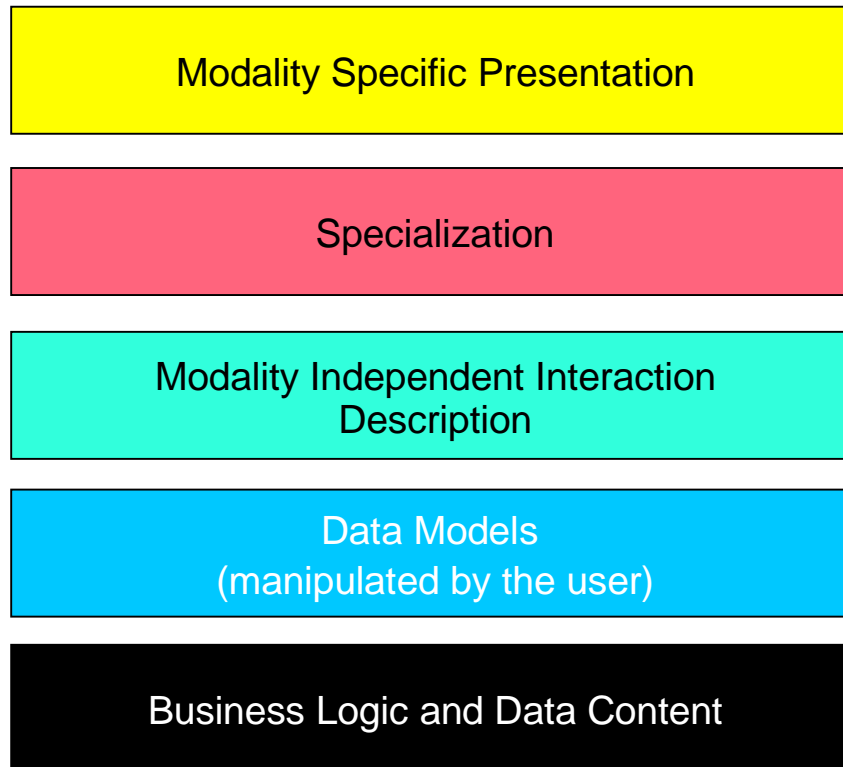
- ▶ Free flow Dialog Application - The dialog is free form /mixed initiative across modalities. Application and user can use any modality at any time, and switch at any time, based on the needs of the application, the situation, and user preferences.
- ▶ User can switch between modalities and applications at any time and share context

&

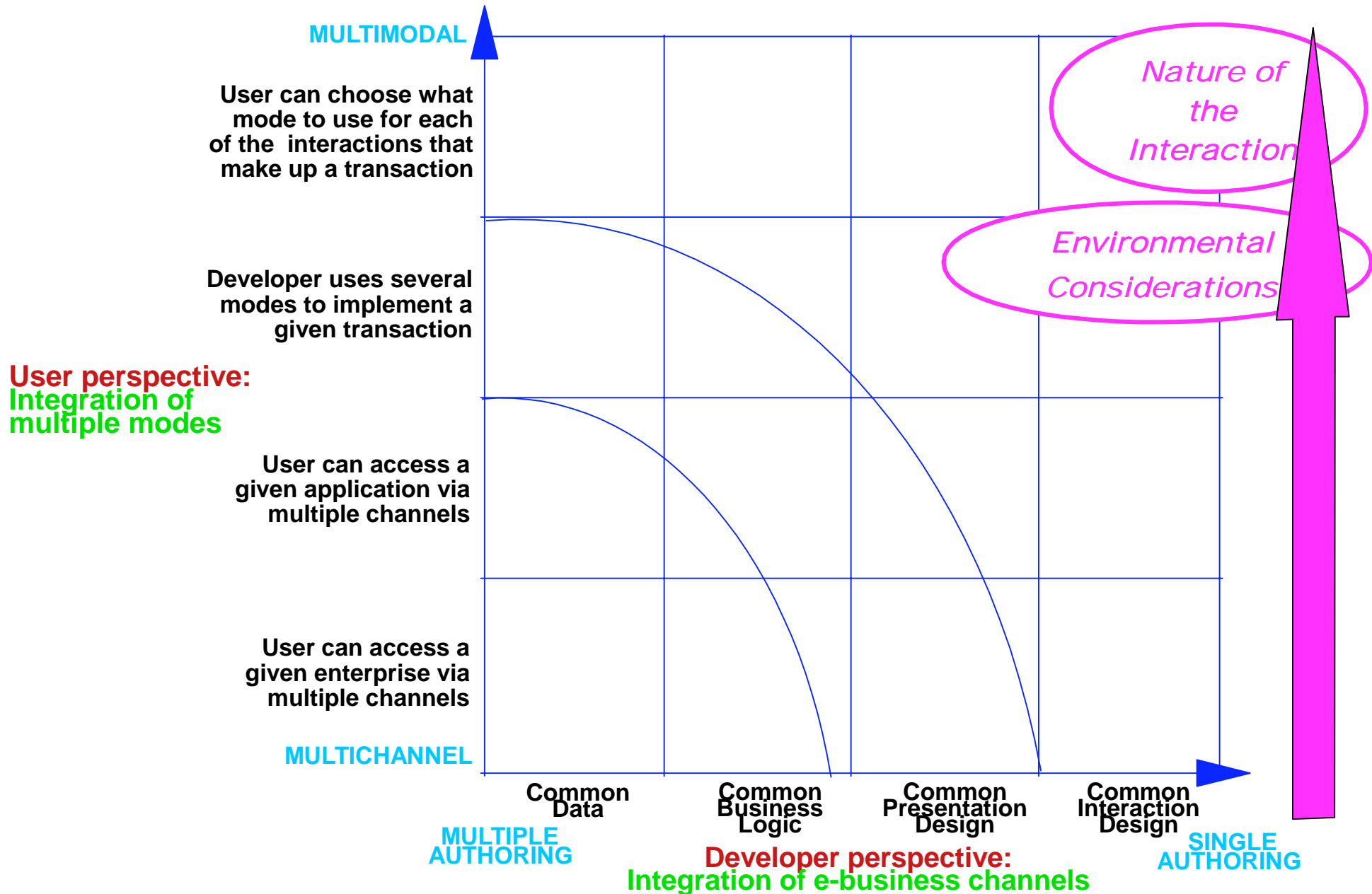
Single Authoring for the Multi-modal web

- ▶ Issues with multiple authoring(Manual or automated through transcoding)
 - ▶ M pages for N channels requires M x N stylesheets
 - ▶ How to support new modalities
- + Authoring of the synchronization for multi-modal applications: Merged languages or synchronization tags!
- ▶ Single authoring:
 - ▶ Channel independent representation of the applications
 - ▶ Channel independent interaction description layer and data model (manipulated by user) between content and presentation - finer factorization
 - ▶ Re-usable stylesheets / transformations
 - ▶ In-line with current evolution of programming model:
 - ▶ Data model representation (XFORMS, Xschema)
 - ▶ Specialization / versioning step for target modalities
 - ▶ Separation of concerns: content & business logic vs. application and design
 - ▶ Support for multi-channel and multi-modal applications
 - ▶ Re-use existing final form languages and renderers

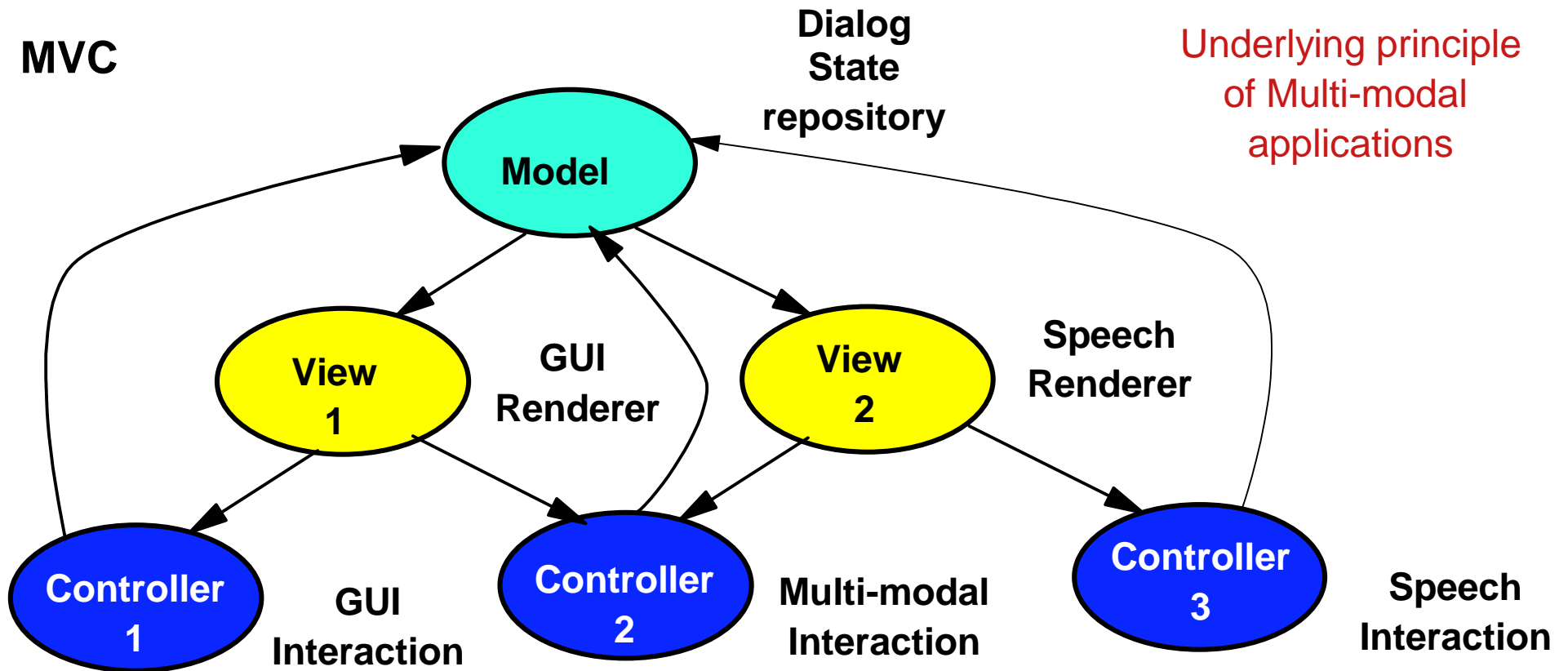
Programming Model for the Multi-modal web



Multi-modal Space Description



Model View Controller



- ▶ The different views of the application are guaranteed to always be in the same state
- ▶ User can change modality at any time
- ▶ Distributable (VoiceTIMES Consortium) & Multi-device

Conclusions: High Level Recommendations

- ▶ We recommend creation of a Multi-channel / Multi-modal Web Authoring Working Group within a standard organization
- ▶ We consider that single authoring is the key requirement of the multi-modal web. Therefore, single authoring of multi-channel and multi-modal applications should be an explicit requirement of the working group charter
- ▶ We need to study this new programming model.

Conclusions: MM Web Authoring Requirements

- ▶ XML compliant
 - ▶ Vendor neutral
 - ▶ Any tool developer can target it or use it as an input representation
 - ▶ It can be used not only to express data within an application, but also to pass it to a network services provider, portal, or directly to an end-user device
 - ▶ A single language should handle both multi-channel applications and multi-modal applications
 - ▶ Can be mapped using style sheets to an open-ended set of device specific markups including VoiceXML, WML, CHTML, HTML and others. Minimal / no change the languages
 - ▶ Can accommodate channel- or device-specific specialization either in-line, as annotations, or using style sheets
 - ▶ Supports a developer-definable hierarchy of channels and devices
 - ▶ Supports specification of data models in Xforms / Xschema to model the data that can be manipulated by the end user
 - ▶ Enables fine-grain synchronization of multi-modal interaction
 - ▶ Can accommodate both synchronous and asynchronous data exchange, and connected as well as disconnected operation

 - ▶ In addition, for multi-modal rendering, we recommend to leverage the forthcoming DOM level 2 specifications to enable the implementation of the MVC with legacy browsers. It implies that the supported channel specific languages must have a DOM level 2 standardized specification.
-