

# Evolution Paths to Multimodal Applications

James A. Larson

Larson Technical Services

[jim@larson-tech.com](mailto:jim@larson-tech.com)

# Four Trends

Evolution paths from  
IVR, Text, GUI to Mobil  
Device Multimodal  
Applications

Embed Speech in HTML

Author once, deploy of  
multiple devices

W3C distributed  
modularization for  
mashups

# Three Evolution Paths to Multimodal

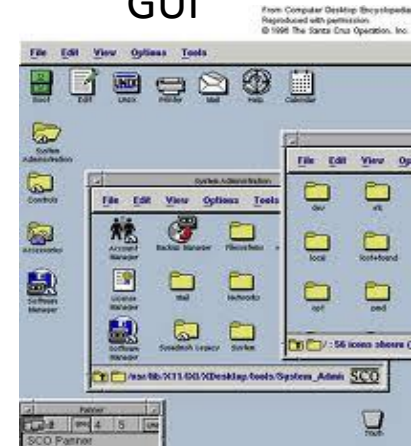
IVR



Text



GUI



# Keyword search



- Key words
  - Blister rust
- Boolean expressions
  - “White pine” and “blister rust”
  - “White pine” or “blister rust”
- Natural Language
  - What causes blister rust in white pines

# Voice User Interfaces



- VoiceXML 2.0/2.1
  - Recorded messages and touch tone
  - Speech recognition and speech synthesis
- VoiceXML 3.0
  - Pictures and Audio
  - Standardization not likely to be completed

# Graphical User Interfaces



- Type via keyboard



# Graphical User Interfaces



- Type via keyboard
- Select via mouse



# Graphical User Interfaces



- Type via keyboard
- Select via mouse
- Write with keyboard replacement





# Graphical User Interfaces



- Type via keyboard
- Select via mouse
- Write with keyboard replacement
- Write, finger gesture, and speak with keyboard replacement

# Write, finger gesture, and speak with keyboard replacement

- Turns existing mobile UIs into multimodal user interfaces today



# Four Trends

Evolution paths from  
IVR, Text, GUI to Mobil  
Device Multimodal  
Applications

- My Advice: Don't waste your time converting your GUI application to multimodal; instead use a keyboard replacement
- Keyboard replacement functions will migrate into mobile device OS

# Four Trends

Evolution paths from  
IVR, Text, GUI to Mobil  
Device Multimodal  
Applications

Embed Speech in HTML

Author once, deploy of  
multiple devices

W3C distributed  
modularization for  
mashups

# Embedding Speech into HTML

- Advantage
  - HTML can be used on any mobile device
- Disadvantage
  - May not run as fast as native code

# Embedding Speech into HTML

## Two W3C candidate standards

- W3C HTML Speech Incubator Group
  - “HTML 5 plus speech”
  - <http://www.w3.org/2005/Incubator/htmlspeech/XGR-htmlspeech-20111206/>
- W3C Speech API Community Group
  - WebSpeech API
  - <https://dvcs.w3.org/hg/speech-api/raw-file/tip/speechapi.html>
- VoiceXML Community Group
  - Use of VoiceXML for mobile applications
- Web Speech Working Group Charter
  - Bring speech to the web
  - <http://www.w3.org/2012/12/speech-charter.html>

# Four Trends

- My Advice: Be careful about lack of voice standards and adherence to HTML 5.0 standards



Embed Speech in HTML

# Four Trends

Evolution paths from  
IVR, Text, GUI to Mobil  
Device Multimodal  
Applications

Embed Speech in HTML

Author once, deploy of  
multiple devices

W3C distributed  
modularization for  
mashups



# Author once deploy on multiple devices

- Open Stream CluMe
- Angel Lexee Speech Assistant

# Four Trends

- My advice: beware of becoming locked into to proprietary platform

Author once, deploy of multiple devices

# Four Trends

Evolution paths from  
IVR, Text, GUI to Mobil  
Device Multimodal  
Applications

Embed Speech in HTML

Author once, deploy of  
multiple devices

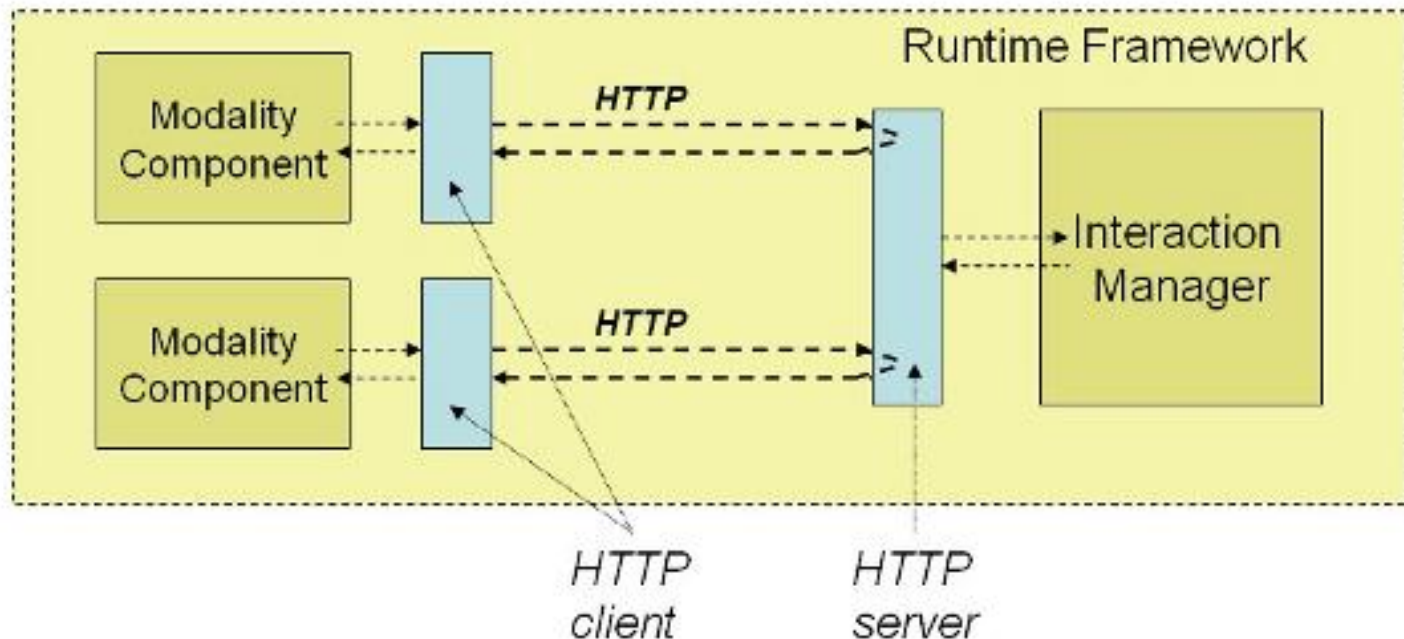
W3C distributed  
modularization for  
mashups

# Distributed modularization

- Multimodal architecture
- EMMA
- Discovery

# W3C Multimodal Architecture

- Loosely coupled architecture for multimodal user interfaces



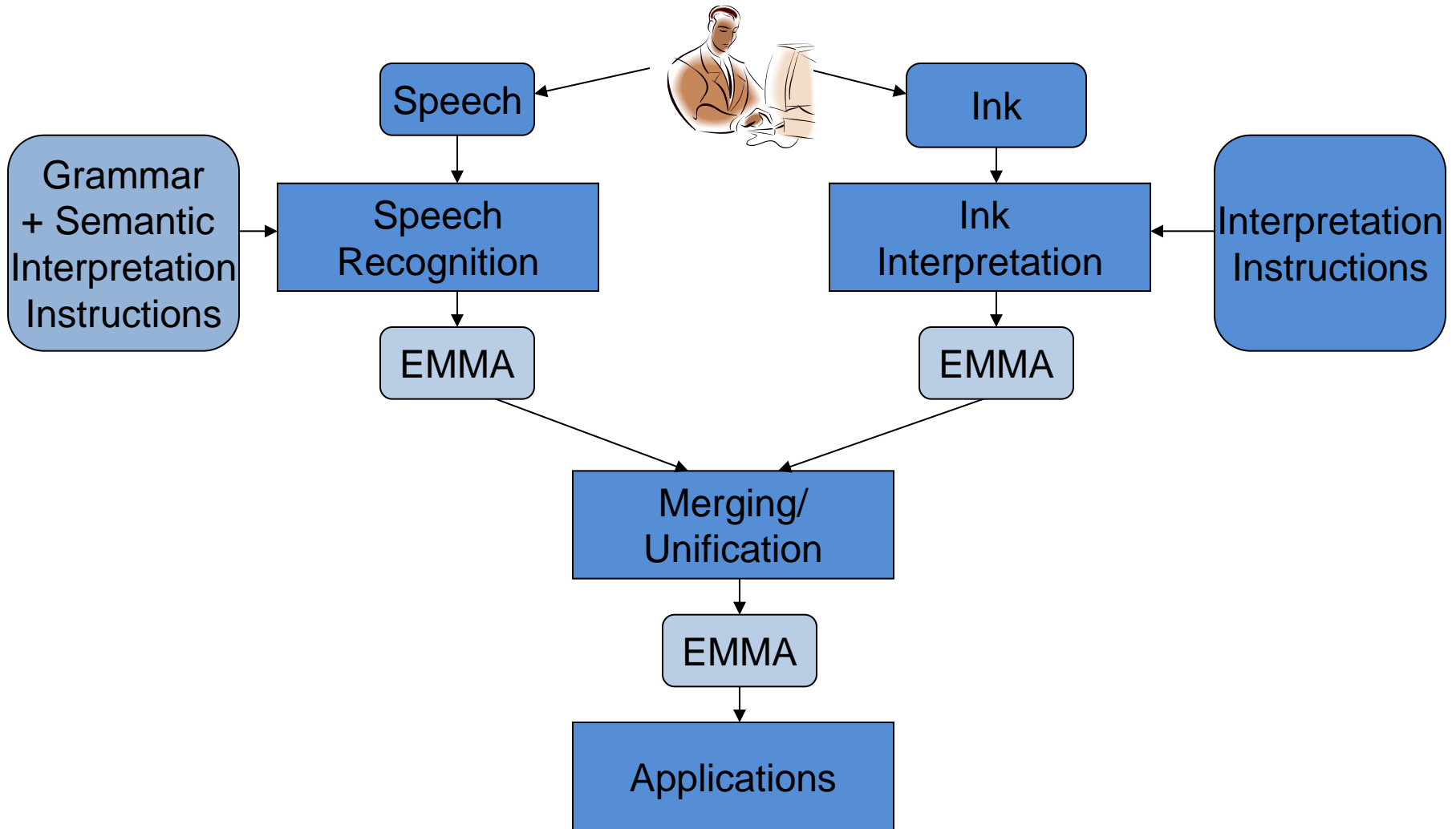
# Possible W3C Modality Components

- Speech Recognition
- Speech Synthesis
- Pen
- GUI
- Voice Biometrics
- Machine Translation
- NL processing
- GPS
- Logic Engine
- ...
- Relational Database
- Sales Force Data
- Map
- Yellow Pages
- Weather
- Financial data
- E-mail
- Social network
- Audio and Video
- ...

# EMMA

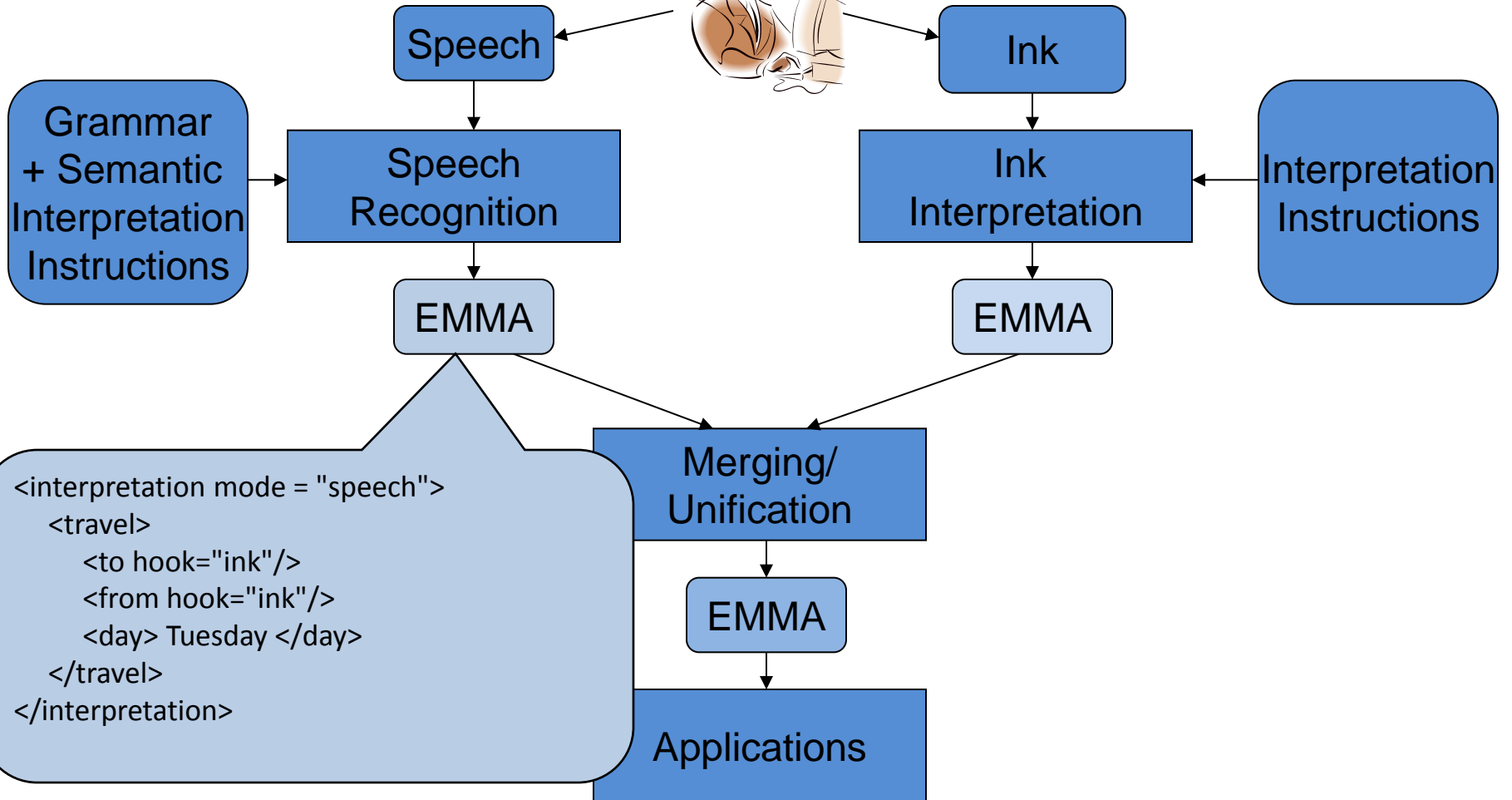
- Extensible Multimodal Annotation markup language
- Canonical structure containing information from modality components
- Used to build “mashup” applications
  - Combines data, presentation and functionality from two or more sources to create new services

# EMMA

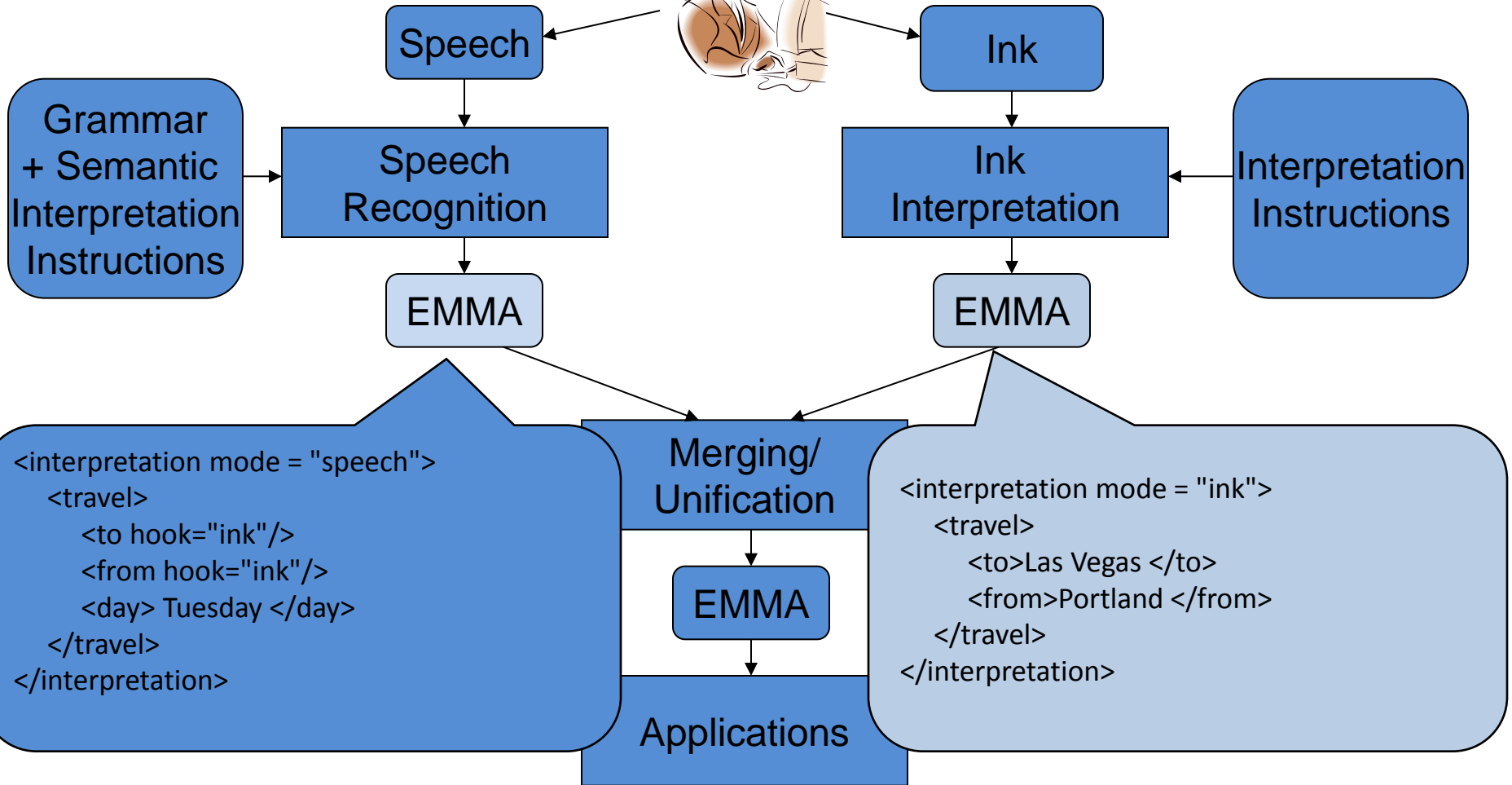




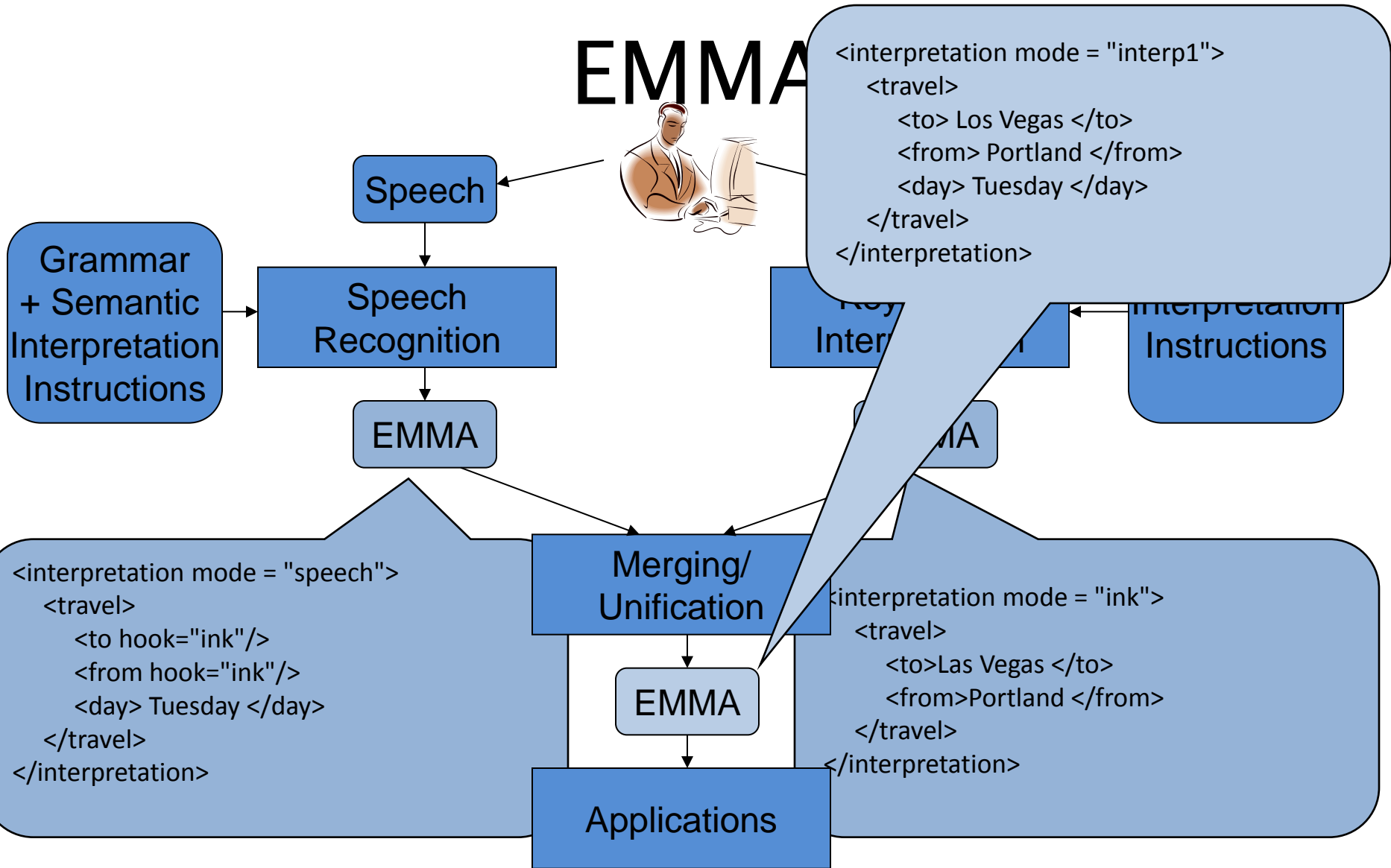
# EMMA



# EMMA



# EMMA



# Modality Component Discovery

- Describe
- Publish
- Discover
- Register
- Subscribe

# Four Trends

- My Advice: Make your data and/or service a “modality” so it can be used in future mashups

W3C distributed  
modularization for  
mashups

# Four Trends

Evolution paths from  
IVR, Text, GUI to Mobil  
Device Multimodal  
Applications

Embed Speech in HTML

Author once, deploy of  
multiple devices

W3C distributed  
modularization for  
mashups

# Situations where voice technologies are useful

- Hands busy
- Eyes busy
- Automated assistants/guides/instructors
- Impaired users
- Traditional input devices not available
- Traditional input devices are not easily useable

# Potential Applications of Voice

- Voice prompts (alerts, error messages, wizards)
- Presentation controller (Juke box, slide show, TV controller)
- Self-help (recipe, do-it-yourself)
  - Assemble, train, diagnose, repair
- Intelligent agents
  - Initially voice-only
  - Text
  - Sequential Multimodal (Put that there)
- Parallel multimodal
  - Drawing apps (change width of line while drawing)
- Voice (ESL, learn to sing, translation, )
- Gaming (third hand)





# My advice

- Don't waste your time converting your GUI application to multimodal; instead use a keyboard replacement
- IVR apps will be difficult to convert to mobile device multimodal applications.
  - IVR are system driven
  - multimodal apps are user driven.
- Concerns about lack of voice standards and adherence to HTML 5.0 standards
- Make your data and/or service a “modality” so it can be used in future mashups

# Questions

