

Frontiers in Interaction

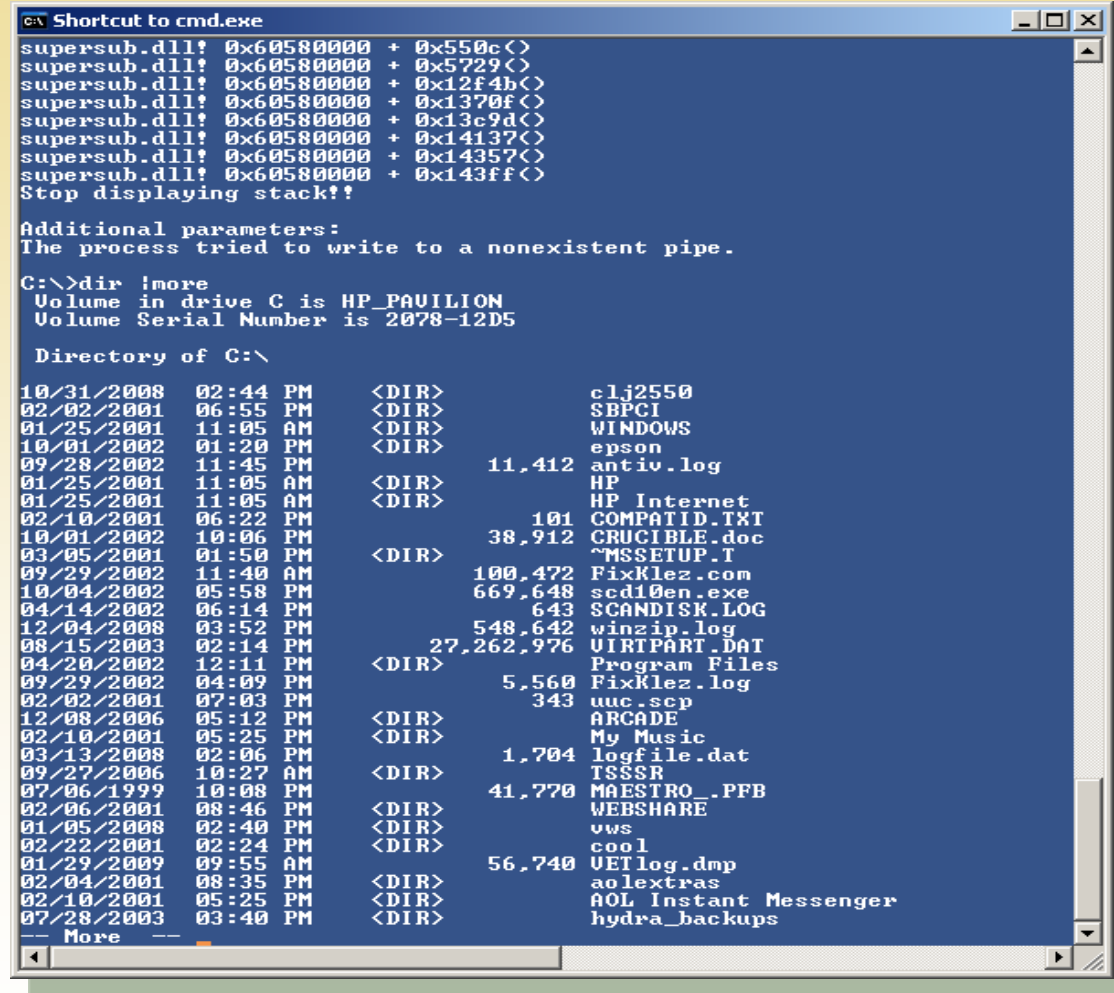
The Power of Multimodal Standards

Deborah Dahl

Principal, Conversational Technologies
Chair, W3C Multimodal Interaction Working Group
VoiceSearch 2009
March 2-4
San Diego, California

User Experience vs. Complexity 1

The command prompt – technically *simple* but a *bad* user experience



```
Shortcut to cmd.exe
supersub.dll! 0x60580000 + 0x550c(<)
supersub.dll! 0x60580000 + 0x5729(<)
supersub.dll! 0x60580000 + 0x12f4b(<)
supersub.dll! 0x60580000 + 0x1370f(<)
supersub.dll! 0x60580000 + 0x13c9d(<)
supersub.dll! 0x60580000 + 0x14137(<)
supersub.dll! 0x60580000 + 0x14357(<)
supersub.dll! 0x60580000 + 0x143ff(<)
Stop displaying stack!!

Additional parameters:
The process tried to write to a nonexistent pipe.

C:\>dir /more
Volume in drive C is HP_PAUILLION
Volume Serial Number is 2078-12D5

Directory of C:\

10/31/2008 02:44 PM <DIR> c:\j2550
02/02/2001 06:55 PM <DIR> SBPCI
01/25/2001 11:05 AM <DIR> WINDOWS
10/01/2002 01:20 PM <DIR> epson
09/28/2002 11:45 PM <DIR> 11,412 antiv.log
01/25/2001 11:05 AM <DIR> HP
01/25/2001 11:05 AM <DIR> HP Internet
02/10/2001 06:22 PM <DIR> 101 COMPATID.TXT
10/01/2002 10:06 PM <DIR> 38,912 CRUCIBLE.doc
03/05/2001 01:50 PM <DIR> ~MSSETUP.T
09/29/2002 11:40 AM <DIR> 100,472 FixKlez.com
10/04/2002 05:58 PM <DIR> 669,648 scd10en.exe
04/14/2002 06:14 PM <DIR> 643 SCANDISK.LOG
12/04/2008 03:52 PM <DIR> 548,642 winzip.log
08/15/2003 02:14 PM <DIR> 27,262,976 UIRTPART.DAT
04/20/2002 12:11 PM <DIR> Program Files
09/29/2002 04:09 PM <DIR> 5,560 FixKlez.log
02/02/2001 07:03 PM <DIR> 343 uuc.scp
12/08/2006 05:12 PM <DIR> ARCADE
02/10/2001 05:25 PM <DIR> My Music
03/13/2008 02:06 PM <DIR> 1,704 logfile.dat
09/27/2006 10:27 AM <DIR> TSSSR
07/06/1999 10:08 PM <DIR> 41,770 MAESTRO_.PFB
02/06/2001 08:46 PM <DIR> WEBSHARE
01/05/2008 02:40 PM <DIR> vws
02/22/2001 02:24 PM <DIR> cool
01/29/2009 09:55 AM <DIR> 56,740 UETlog.dmp
02/04/2001 08:35 PM <DIR> aolextras
02/10/2001 05:25 PM <DIR> AOL Instant Messenger
07/28/2003 03:40 PM <DIR> hydra_backups

More --
```

User Experience vs. Complexity 2

The iPhone Google Earth application – technically *complex* but a *good* user experience



As the user experience becomes more natural and intuitive, the more complex the supporting technology becomes

Multimodality offers an opportunity for a more natural user experience, but ...

There are a lot of technologies
involved in multimodal applications!

Parts of a Multimodal Application

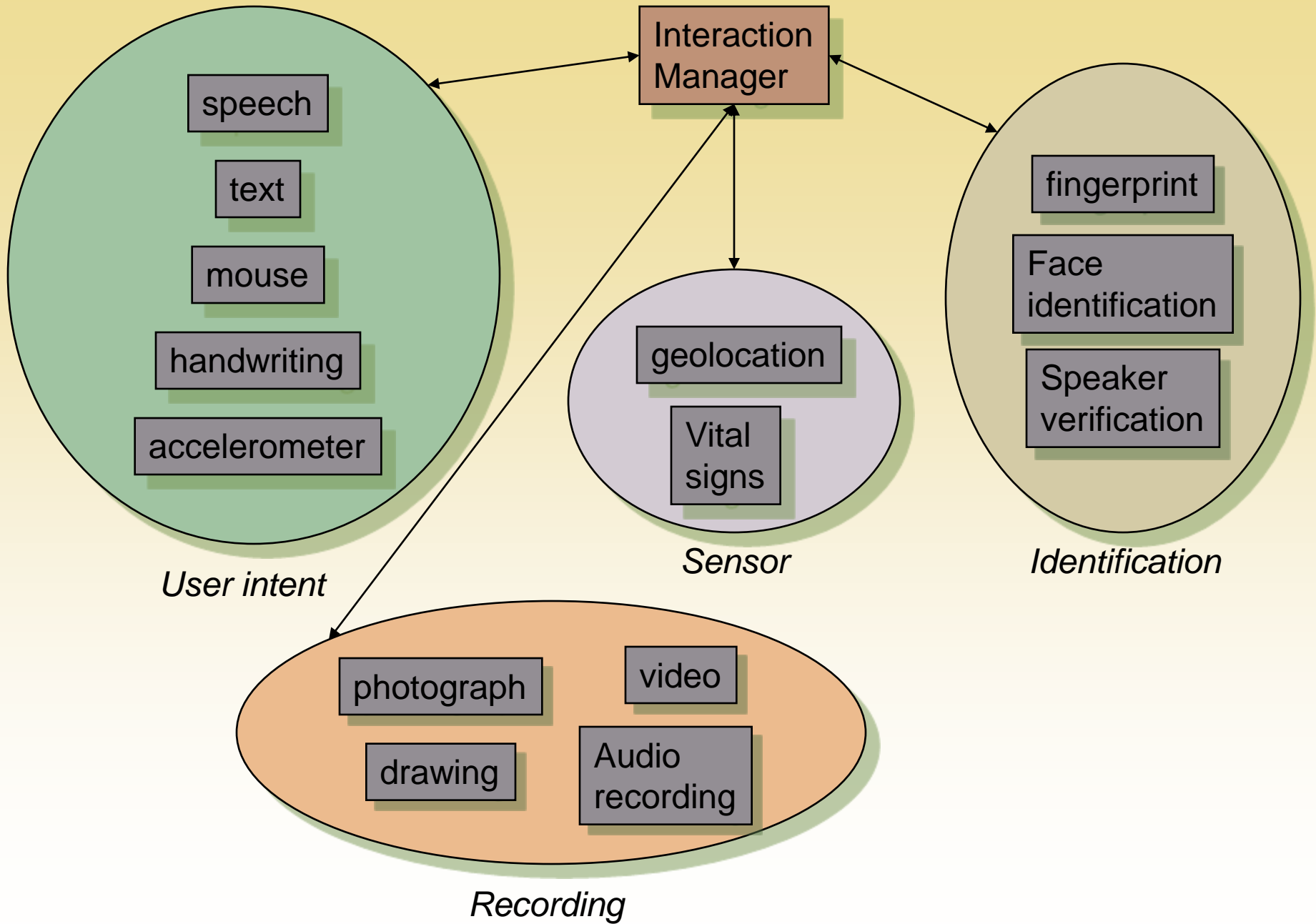
- User experience
- Technology
 - Technologies for managing the interaction
 - Technologies for capturing and interpreting user input
 - Technologies for capturing sensor data (geolocation)
 - Technologies for presenting output
- Communication protocols
- Devices
- Servers
- Networks

Multimodal applications require coordinating many forms of user and sensor input...

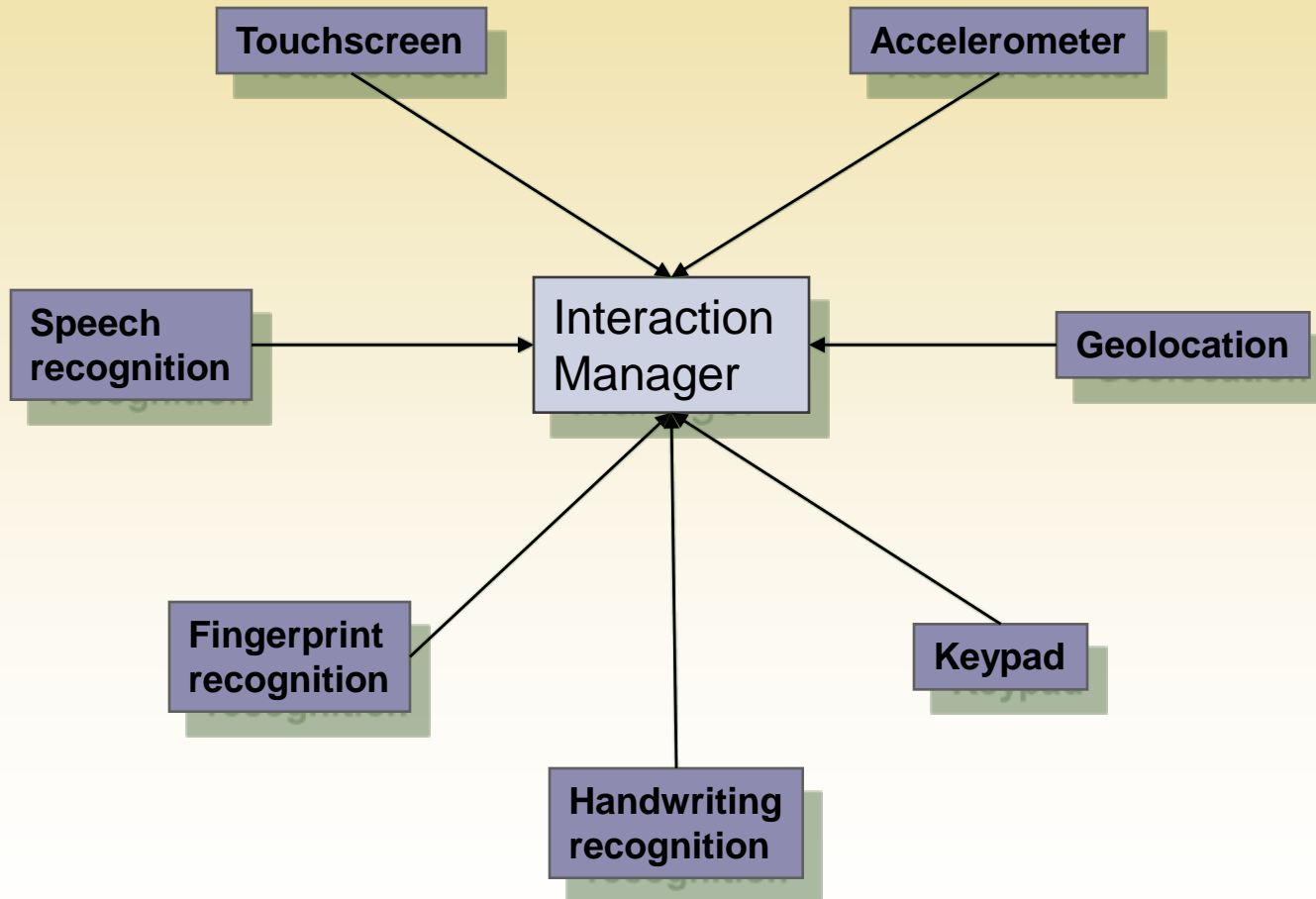
- Speech
- Audio
- Stylus
- Touch
- Accelerometer
- Keyboard/keypad
- Mouse/touchpad
- Camera
- Geolocation
- Handwriting recognition
- Speaker verification
- Signature verification
- Fingerprint identification
-

...while presenting the user with a
natural and usable interface

A multimodal application can include many kinds of input, used for many purposes



It can include many technologies



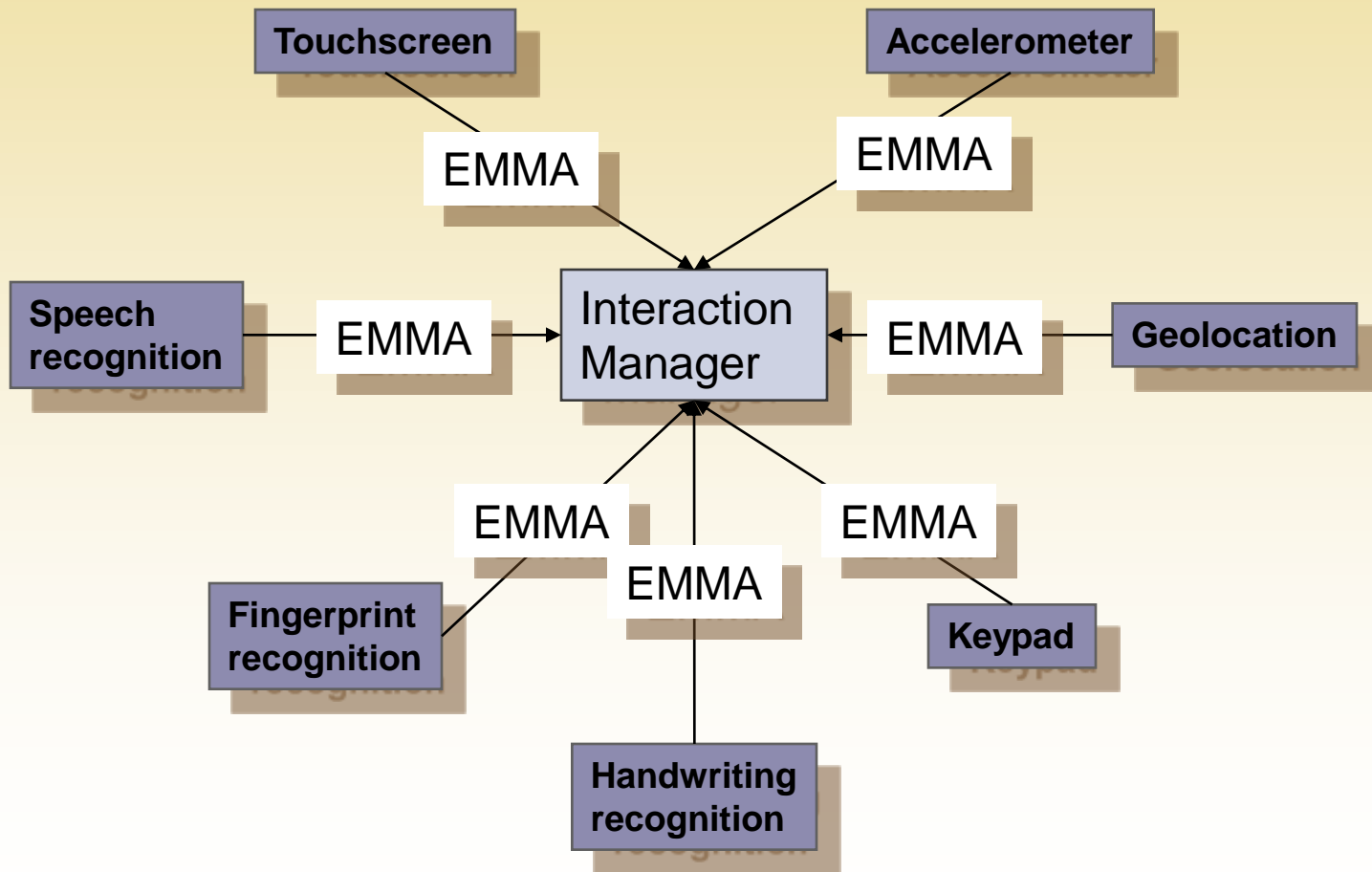
Getting everything to work together is complicated.

One way to make things easier is to represent the same information from different modalities in the same format.

We need a common language for representing the same information from different modalities.

EMMA: (Extensible MultiModal Annotation)

A uniform representation for multimodal information



EMMA can represent many types of information because it includes both *general* annotations as well as *application-specific* semantics

General Annotations

- Confidence
- Timestamps
- Alternative interpretations
- Language
- Medium (visual, acoustic, tactile)
- Modality (voice, keys, photograph)
- Function (dialog, recording, verification...)

Representing user intent

Example: Travel Application

“I want to go from
Boston to Denver on
March 11”



The screenshot shows a web browser window titled "Flight Selector - Mozilla Firefox". The browser's menu bar includes "File", "Edit", "View", "History", "Bookmarks", "Tools", and "Help". The main content area is titled "Flight Selector" and contains the following elements:

- A "Hold to Talk" button.
- Three input fields labeled "Departure", "Arrival", and "Date".
- A "Submit" button.
- A calendar widget for "Mar 2009" showing days from 1 to 31. The date "11" is highlighted in blue.

The status bar at the bottom of the browser window displays "Done".

Application Semantics (modality-independent)

```
<emma:interpretation id="int1">  
  <origin>Boston</origin>  
  <destination>Denver</destination>  
  <date>03112008</date>  
</emma:interpretation>
```

This could have originated from speech, keyboard, pull-down menus, handwriting...

But it looks (almost) the same in EMMA because it has the same meaning

The *same meaning* with speech and mouse input

```
<emma:interpretation medium="acoustic" mode="voice"  
id="int1">
```

```
<origin>Boston</origin>
```

```
<destination>Denver</destination>
```

```
<date>03112008</date>
```

```
</emma:interpretation>
```

Speech

```
<emma:interpretation medium="tactile" mode="gui"  
id="int1">
```

```
<origin>Boston</origin>
```

```
<destination>Denver</destination>
```

```
<date>03112008</date>
```

```
</emma:interpretation>
```

Mouse

Handwriting, too

```
<emma:interpretation medium="tactile" mode="ink" id="int1">  
  <origin>Boston</origin>  
  <destination>Denver</destination>  
  <date>03112008</date>  
</emma:interpretation>
```



Identifying users -- biometrics

The *same identification* with face recognition or speaker identification

```
<emma:emma version="1.0">  
  <emma:interpretation  
    id="int1  
    emma:confidence=".75"  
    emma:medium="visual"  
    emma:mode="photograph"  
    emma:verbal="false"  
    emma:function="identification">  
    <person>12345</person>  
    <name>Mary Smith</name>  
  </emma:interpretation>  
</emma:emma>
```



```
<emma:emma version="1.0">  
  <emma:interpretation  
    id="int1  
    emma:confidence=".80"  
    emma:medium="acoustic"  
    emma:mode="voice"  
    emma:verbal="false"  
    emma:function="identification">  
    <person>12345</person>  
    <name>Mary Smith</name>  
  </emma:interpretation>  
</emma:emma>
```



Other Modalities

- Geolocation – latitude, longitude, speed
- Text – plain old typing
- Accelerometer – position of device
- ...

More Information

- The specification: <http://www.w3.org/TR/emma/>
- The Multimodal Interaction Working Group:
<http://www.w3.org/2002/mmi/>
- An open source implementation:
<http://sourceforge.net/projects/nlworkbench/>