SpeechWeb & Adobe Captivate

towards a revolution in education

Richard Frost
School of Computer Science
University of Windsor

ITS 2012 Windsor

Questions

1. **Given that speech is a fundamental method of communication:**

   - Why are there so few web-based “speech applications”.
   - Why are there so few natural-language “English” interfaces to web applications and data?
   - Why are there hardly any “speech games” on the web?

2. **Given that YouTube is so easy to use:**

   - Why do we not have more “college and university lessons” available on YouTube?

Possible Answers

1. **Speech and natural-language applications**
   - Speech technology is immature.
   - NL theories cannot be computerized.
   - There is no market for such applications
   - Few people are interested in creating speech & NL apps.
   - Speech and NL technologies are extremely difficult.

2. **YouTube lessons**
   - Instructors are not interested in creating on-line lessons.
   - Video capture technology is difficult to use.

A different perspective

- **Speech technology is very mature (e,g, Google speech apps, iPhone 4S)**
- **Compositional theories of natural language are available.**
- **The market for NL speech applications is huge, as is on-line learning.**
- **Many people are interested in these technologies BUT think that they are very difficult.**

My Thesis

- **Technology, interest and NOTATION is now available for non-experts to create natural-language speech applications and deploy them on the web.**
- **Video capture technology is available that allows non-experts to build computer based lessons and deploy them on YouTube and elsewhere.**
- **In the next few years we will see a massive increase in NL speech interfaces to knowledge and access to on-line lessons which will revolutionize education.**

We begin with an analogy

An old tune goes global

- **Pachelbel composed the “Canon” (late 1600s)**

  http://www.youtube.com/watch?v=8Af372EQLck

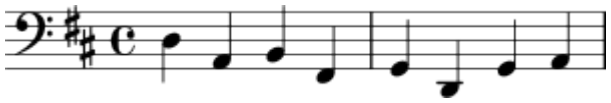
- **Jerry C (Chang) re-arranged for electric guitar around “Canon Rock” (2005)**

  http://www.youtube.com/watch?v=by8oyJztzwo

- **A youtube user, Impeto, spliced together 39 excerpts of musicians playing and called it the “Ultimate Canon Rock” (2007)**

  http://www.youtube.com/watch?v=dMWl_5NujBw

What helped Jerry C teach a wide range of people to play the Canon and participate in the "Ultimate Canon Rock"



- Electric guitar (1930's)
- The Web (Tim Berners-Lee 1990's)
- YouTube
- Guitar TAB reborn in 40's, widely used now

```
----------------------|----------------------|----------------------
--19~--------15~------|--16~--------13~------|--14~--------11~------
----------------------|----------------------|----------------------
----------------------|----------------------|----------------------
----------------------|----------------------|----------------------
----------------------|----------------------|----------------------
```

And now for something completely different

A video demonstration of SpeechWeb created using Adobe Captivate Software.

www.youtube.com (and type in “speechweb”)

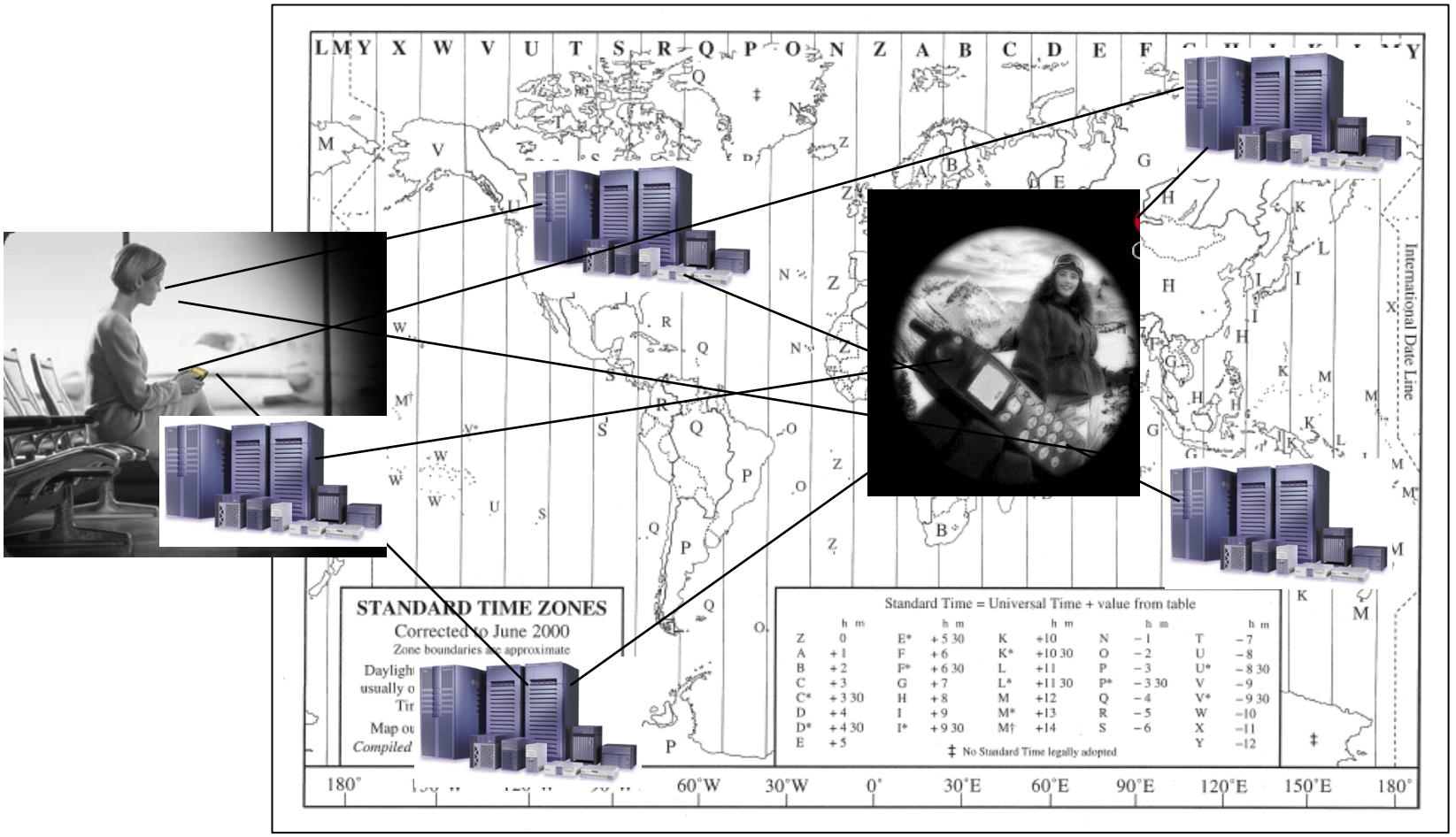
or go directly:

http://www.youtube.com/watch?v=Axa-n4etdZE

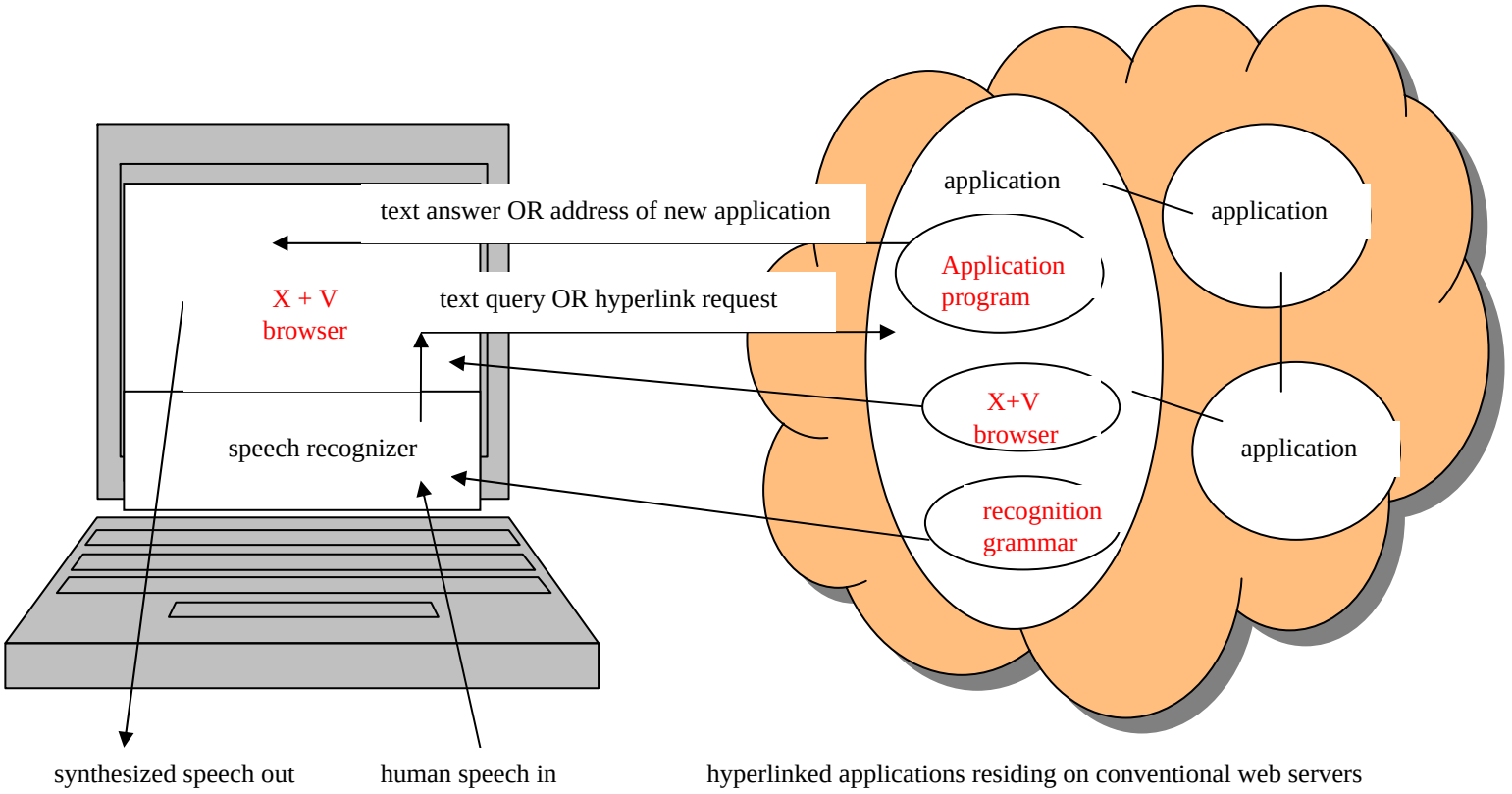
A Brief Overview of SpeechWeb Technology

- The SpeechWeb architecture
- The speech browser interface
- How to create a SpeechWeb application and deploy on the web.
- The mathematical basis of natural language processing.
- A summary of the notation which has made it possible.

Local Recognition Remote Processing (LRRP) Architecture



Applications in the cloud



To Create a SpeechWeb Application

- **Copy three files into a web directory**
  1. The X+V browser
  2. A sample grammar
  3. A sample program

- **Modify four lines in the X+V browser**

- **Change the grammar for your application's input language.**

- **Modify the sample program or replace with a program, written in any language to process the input.**

ALL SIMPLE NOTATION?

The X+V Browser

```
•<html xmlns="http://www.w3.org/1999/xhtml" xmlns:vxml……
•  <head>
•    <title id="title" />


•    <!-- the name of the speechweb application and its opening statement are specified here -->
•    <script type="text/javascript">
•      var appName = “Monty";
•      var appFullName = “speechweb.cs.uwindsor.ca/applications/Monty";
•      var greeting = “Hello. My name is Monty. I know a joke.";
•    </script>


•    <!-- main vxml form for handling the user/application dialogue -->
•    <vxml:form id="vxml_main">
•      <vxml:field name="vxml_field" modal="true">
•              <vxml:grammar type="application/x-jsgf" src=“Monty.jsgf" />

•              <vxml:prompt cond="greeting.length > 0">
•                <vxml:value expr="showMessage('greeting', greeting)" />
•                <vxml:value expr="greeting" />
•                <vxml:value expr="greeting = "" />
•              </vxml:prompt>
•etc …….
```

Recognition Grammars Guide Search

```
<question>      = what is your name
                | where do you live
                | what do you know
                | tell me a joke
                | can I talk to <person>
                | etc ;

<person>        = judy | solarman | pete …
```

The Programs can be as simple as you want

```
interpret "what is your name" = "My name is Monty.“

interpret "where do you live"
                  = "I hang out in one of Frosties computers. “

interpret "what do you know"
                  = "I got a joke or two. Not much else.“

interpret “tell me a joke"
                  = “Did you hear about the two professors……."
```

The Basis of the Natural language Technology

Variation of Montague's NL semantics (1970's) developed in the λ-calculus (Church 1930's), and implemented in set-theory.

$[[\text{Mars}]] = \lambda s \quad e_{mars} \in s$

$[[\text{spin}]] = \{e_{earth}, e_{mars}, e_{luna}, \ldots\}$

$[[\text{moon}]] = \{e_{luna}, e_{phobos}, \ldots..\}$

$[[\text{Mars}]]\ [[\text{spins}]] \Rightarrow (\lambda s \ e_{mars} \in s)\ \{e_{earth}, e_{mars}, \ldots\}$
$\Rightarrow e_{mars} \in \{e_{earth}, e_{mars}, \ldots\}$
$\Rightarrow$ True

$[[\text{every}]] = \lambda p\ \lambda q\ \ p \text{ subset } q$

The result is a fully compositional semantics

- **The composition rule is always simple function application, e.g.**

  (hall or kuiper) (discovered (every moon))

- **The semantics covers a large sub-set of classical first-order English.**

  `does every moon and every planet spin`

  `how many moons that orbit a red planet were discovered by the person who discovered Nereid`

  `which planet is orbited by no moon`

- **The meaning of words can be defined in terms of other words.**

$$[\![\texttt{discoverer}]\!] = [\![\texttt{person who discovered a thing}]\!]$$

The notation which simplifies creation and deployment of NL speech applications

- VXML (X+V) to configure/interface to the speech recognizer
- BNF notation for recognizer grammars
- Declarative/equational programming languages
- λ calculus and set theory for NL

Adobe Captivate

- **Captures all screen activity and voice over (and sounds from a computer session).**
- **Clever capture minimizes resulting video.**
- **Publish as .pdf, .mp4 etc and directly to YouTube.**
- **Can edit video and sound.**
- **Learning curve similar to PowerPoint.**
- **Can be used with tablets to create “Khan-style” online lessons: http://www.khanacademy.org/**

Use of speech and captivate technology in Education

- **Non experts can add speech interfaces to their web applications.**
- **Non experts can create lessons about anything and deploy them on the web.**
- **In the future we will be able to create interactive on-line lessons with spoken natural-language interfaces.**

Multi-Modal Online Education

Using speech games to create cognitive profiles

- Video games are being used to develop cognitive profiles of users. Can help identify learning strengths and weaknesses in children.
- Speech games can add another “dimension” to the cognitive profiles.
- We are currently designing speech-only games for children aged 6 and above.

Acknowledgements

Graduate Students: Sanjay Chitte, William Ma, Fadi Hanna, Jack Su, Shahriar Chandon, Nabil Abdullah, (Sunny) Yue Shi, and Rahmatullah Hafiz.

Undergraduate students: Ali Karaki, David Dufour, Josh Greig, S. Daichendt, Justin Barolak, Randy Fortier, Bryan St Amour, Jon Donais, Paul Meyer and Matthew Clifford.

The research is funded by NSERC discovery grants, NSERC USRAs, and U. of Windsor Outstanding Scholar awards.