

- Part I -: 08 Mai 2006

## Technological Survey of a Multimodal library: CSLU Toolkit

Msc-Students: - Ridha Zarrougui  
- Lawrence Michel  
\*\*\*\*\*

Supervisors: - Denis Lalanne  
- Omar Ab. Khaled  
- Jacques Bapst

---

# Content

---

- What is CSLU ?
- What does it do ?
- For what it can be useful ?
- How to install it ?
- How does it work ?
- What are the CSLU limitations (weakness) ?
- Strong points ?
- Demo ?
- Conclusion: what we have learned ?

# What is CSLU Toolkit ?

---

A comprehensive environment to build, investigate and use interactive language systems.

<http://www.cslu.ogi>

Center for Spoken Language Understanding  
(CSLU)  
OGI Campus,  
Oregon Health & Science University  
USA

Prof. van Santen



- Development started in 1992
- Free for research use; customizable for corporate use

# What does it do ?

---

- Conduct both Basic Speech Science and Technology Development in all areas of Spoken Language Technology.
- Educate a multi-disciplinary workforce with hands-on experience in Spoken Language Systems.
- Help Businesses, Researchers, and Educators build and use Speech Technology.

under constraints of :

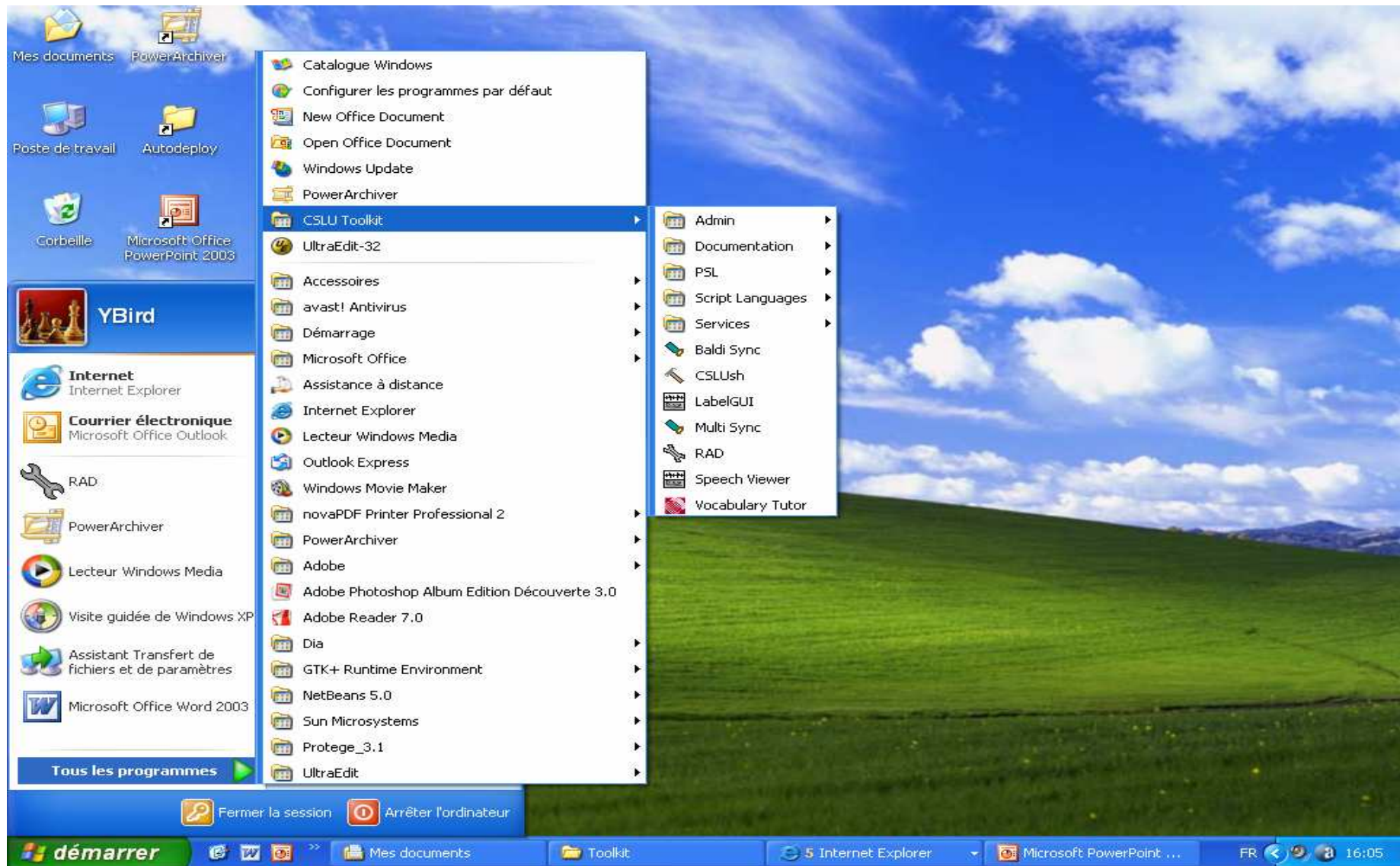
being **comprehensive**, **powerful** and to provide a **flexible** environment for building interactive systems.

# For what it can be useful ?

---

- Goal: A few years from now, *conversing with our computers* will be as common as making phone calls from our car, sending pictures over the telephone, or browsing the Internet is today.
- To reach this Goal: CSLU contributes in speech technologies to improve helping people with visual, speech, or learning problems, dictating letters directly into a computer, accessing the internet over the phone, using voice commands to control devices, and education.

# How to install it ?



# ***What Tools are in the Toolkit?***

---

- **Rapid Application Developer (RAD):**
  - + Graphical tool for creating structured dialog between the user and the computer.
- **Baldi/-Sync**
  - + synchronizes arbitrary speech with animated faces
  - + mouth movements
- **PSL Tools**
  - + facilitates human perceptual experiments within RAD

# *RAD* 1/2

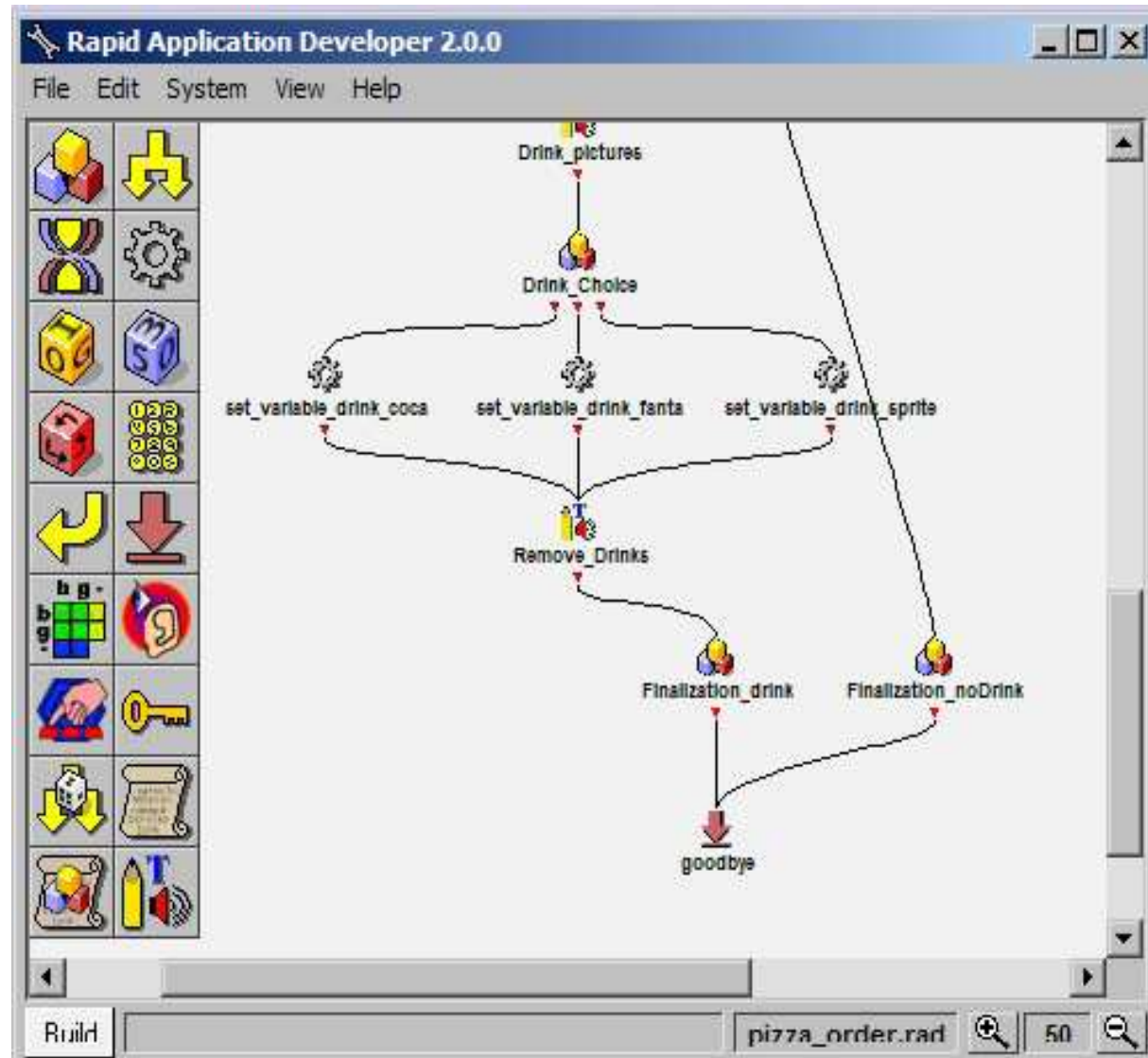
---

- ◆ Graphical tool for creating structured dialog between the user and the computer.
  - build applications by arranging basic objects together
  - requires knowledge of TCL/TK to enable enhanced application usage
  - User front-end and high level modules have been implemented in TCL/TK for ease of use, while back-end modules, such as live speech recognition (Festival), have been written in C for speedness purposes.



# RAD 2/2

- Screenshot:

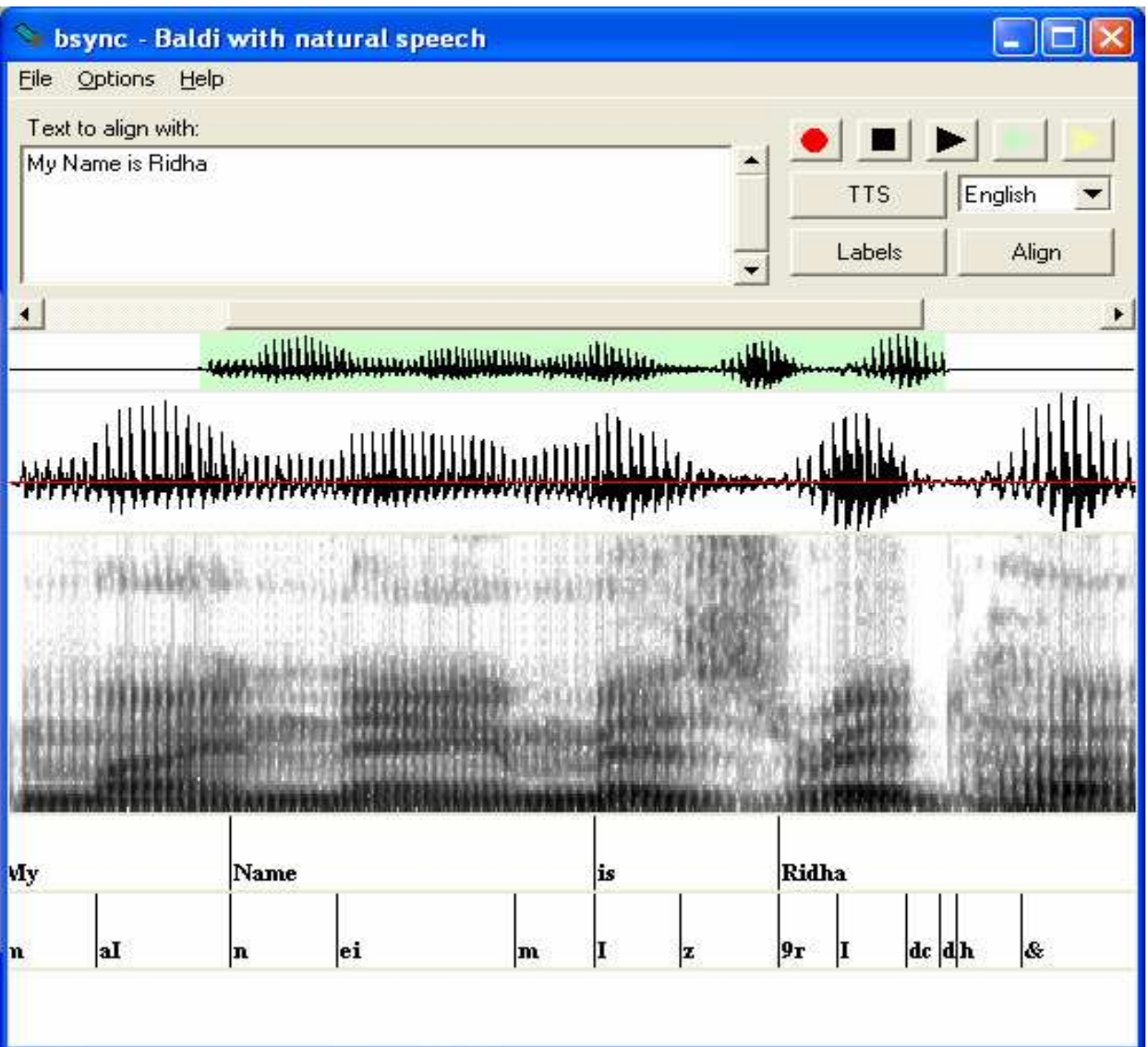



# *Baldi/-Sync* <sup>1/2</sup>

- ❖ Tool which you can use to view and create facial animation that is aligned with recorded audio.
- ❖ Can read and write wave files
- ❖ Animates a 3D face according to what is being spoken (or read)
- ❖ Requires speaker alignment support for correct phonem-face-expression synchronization

# Baldi/-Sync 2/2

- Screenshot:



The screenshot shows the **bsync - Baldi with natural speech** application interface. The window title is "bsync - Baldi with natural speech". The menu bar includes "File", "Options", and "Help".

The main interface features a text input field labeled "Text to align with:" containing the text "My Name is Ridha". To the right of this field are several control buttons: a red circle (stop), a black square (pause), a black triangle (play), a green square (record), and a yellow square (stop). Below these are buttons for "TTS" (with a dropdown menu set to "English"), "Labels", and "Align".

The central area displays three audio waveforms. The top waveform is highlighted in green. Below the waveforms is a spectrogram showing the frequency spectrum of the audio.

At the bottom, there is a text alignment table with the following content:

My	Name	is	Ridha								
n	aI	n	ei	m	I	z	9r	I	dc	dh	&

# What are the CSLU limitations (weakness) ?

---

- Speech recognition conditions
- We can develop only small applications
- With long text will take more time to pass to the next step...
- Emotion management (gesture versus Voice)

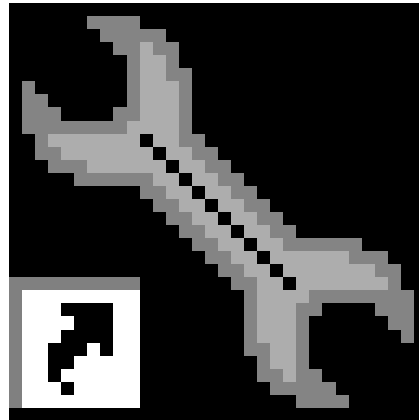
# Strong points ?

---

- Easy to use
- Powerful for research and development use
- Systems that work in real world
- Incorporates research advances
- Modular System (Baldi/-Sync, Festival)
- Eventual integration within other applications (VR, client Service voice based, ...)

# Demo 1

---



Rad.Ink

# Demo 2

---

- **BALDI-Sync**

# Conclusion: what we have learned ?

---

- Introduction over CSLU including what is it, how to install it, ...
- What are the added values of the CSLU to the existing technologies
- How it works..
- The weakness and the strong points within CSLU



*Thank you for your attention*

