

Speech: A Privileged Modality

Luc JULIA Adam CHEYER



SRI International

julia@speech.sri.com

cheyer@ai.sri.com

Model of a Multimodal Dialog

VO*V*: Verb Object[s] Variable[s]

- V: word[s] expressing an **ACTION**
- O*: **OBJECT**[s] for which the Verb applies
- V*: **ATTRIBUTE**[s] to complete the command

Input modalities fill slots in any order (**synergistic**)

Little need for time stamping, **based on completion**

Multimedia prompting mechanism to assist the user

Chosen Modalities

- **Speech**

- Nuance speech recognizer



- **Handwriting**

- CIC handwriting recognizer



- **Gesture**

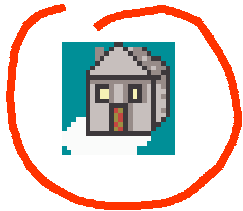
- Proprietary gesture recognizer



Speech leads the way!

A **gesture** does **NOT** affect the fundamental **meaning of a spoken query**, it can complete it (**complementary**) or confirm it (**redundancy**)

**Delete this
house**

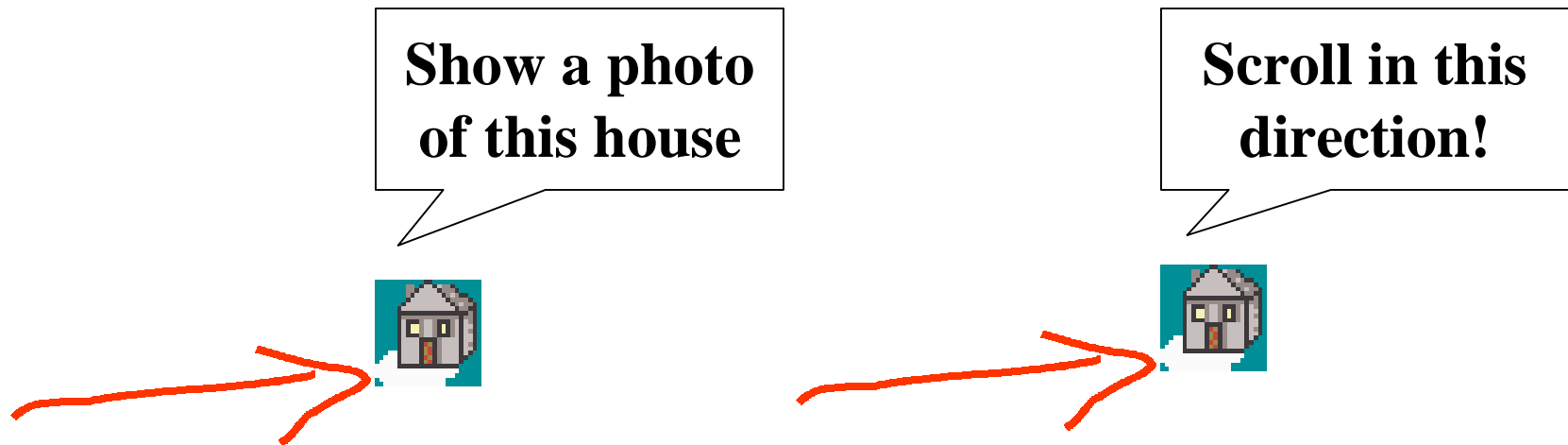


**Delete this
house**



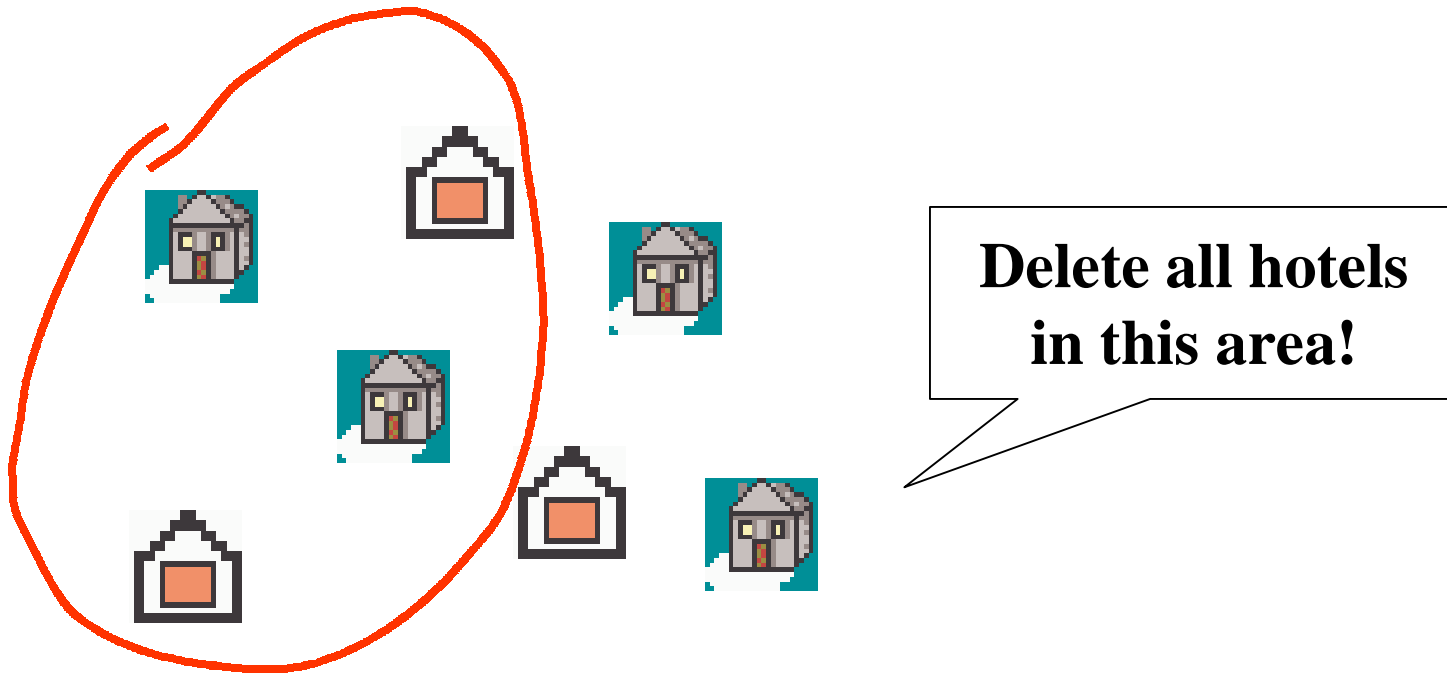
Speech leads the way (2)!

If the user speaks while gesturing, the **gesture's interpretation IS AFFECTED** by the speech



Speech leads the way (3)!

Spoken queries are even more powerful
thanks to their **filtering abilities**



Exception!

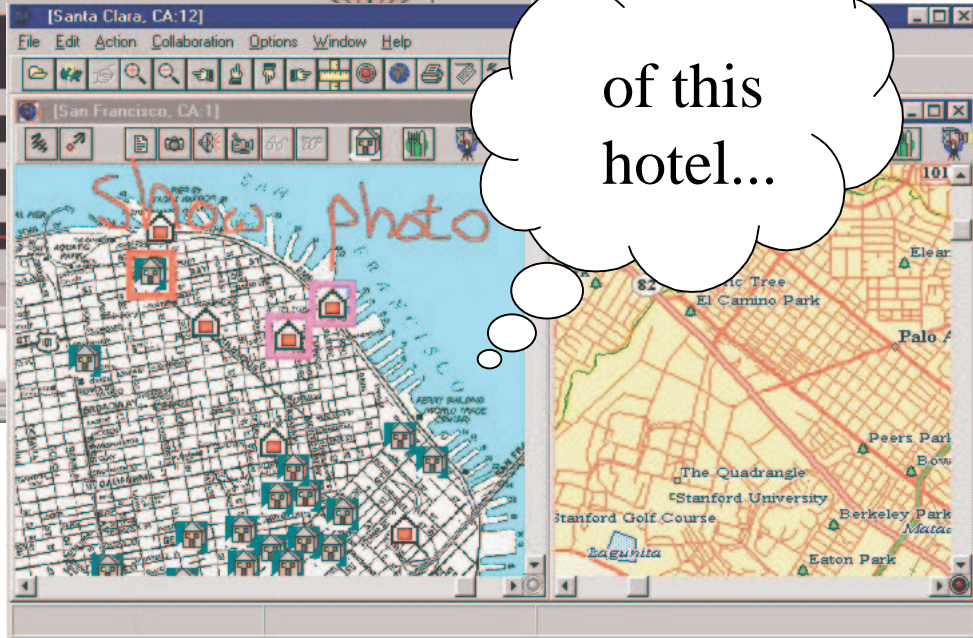
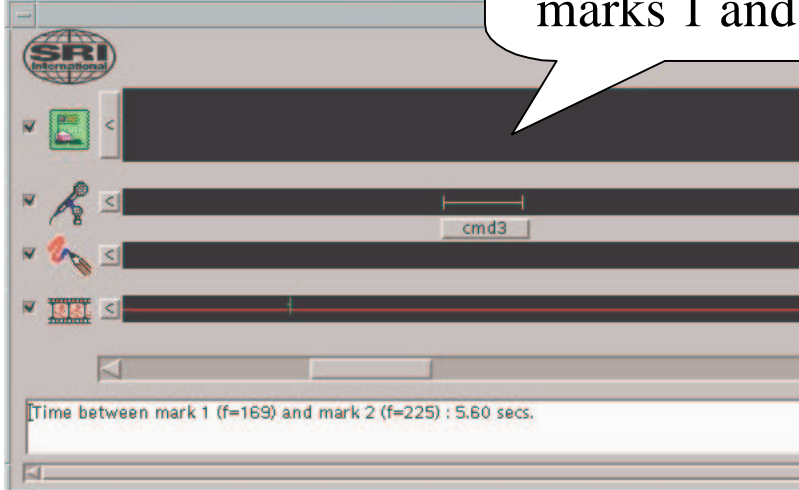
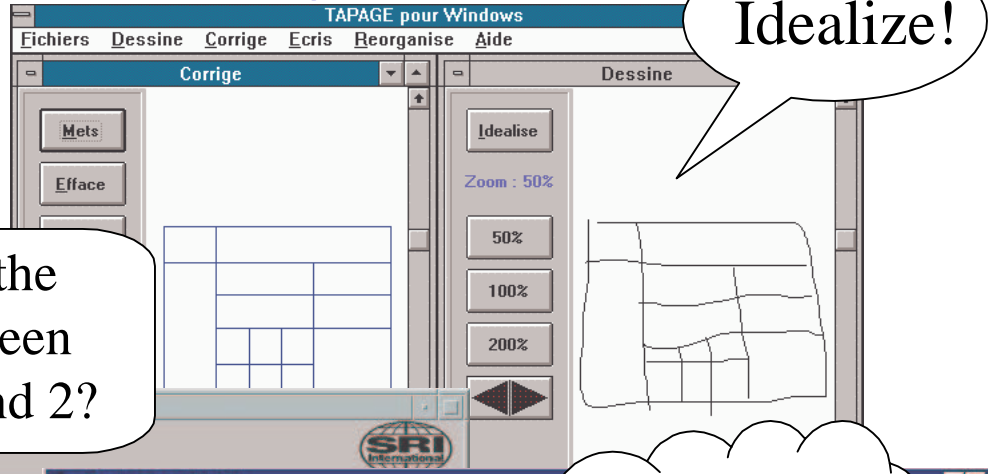


Users make **errors for directional**
commands when they **speak** them!
The **gesture interpretation wins**

Implemented Systems



What is the time between marks 1 and 2?



Allow **complex** multimodal interactions
Resolution of **ambiguities**, predictions
Real subjects observations and studies

Open Agent Architecture (OAA) based

Conclusion

WHAT IS SAID

is

WHAT IS MEANT

References

- **Cheyser:** <http://www.ai.sri.com/~cheyer>
- **Julia:** <http://www.speech.sri.com/people/julia>
- **CIC:** <http://www.cic.com>
- **Nuance:** <http://www.nuancecom.com>
- **OAA:** <http://www.ai.sri.com/~oaa>