

# Maintaining Credible Dialogs in a VideoBot System: Special Audio Techniques

Doug DeGroot  
Leiden Institute of Advanced Computer Science  
Leiden University  
The Netherlands

ICTAI 2001

# Presentation Outline

- Describe our SmartTV/SmartHouse Domain
- Describe our VideoBot Approach
- Relate VideoBots to ChatterBots
- Discuss “Believability” and “Credibility”
- Present Several Audio “Tricks”

# Our Central Research Issues:

Within the domain of

**Visual/Animated & Conversational Agents,**  
much emphasis is placed on

- **“believability”**
- **the Turing & Loebner tests**

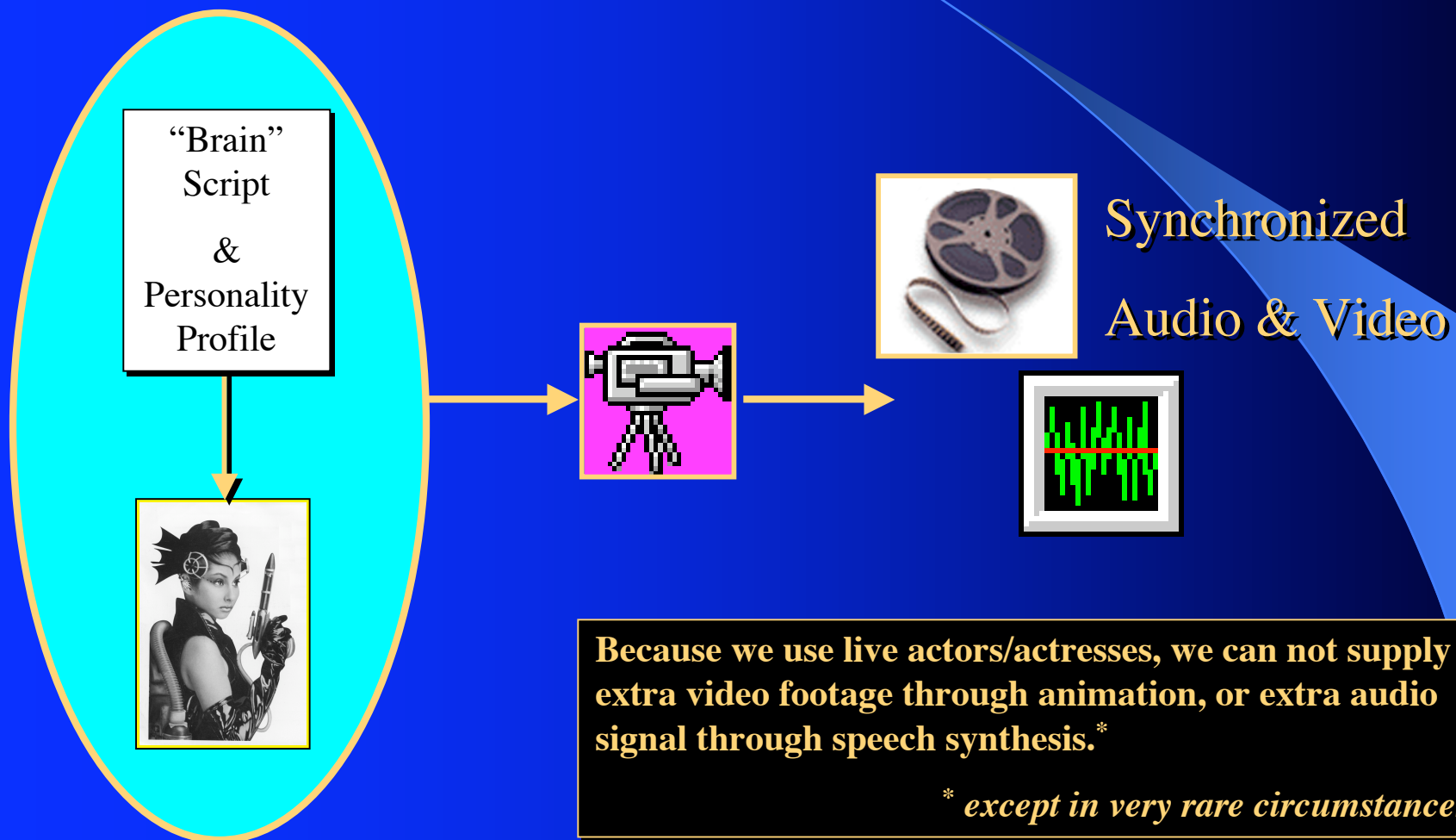
If we start out with “real” humans, ...

- **do we start out with “believability”?**
- **is it easier to achieve & maintain “suspension of disbelief”?**
- **do we still have to trick the user into believing the machine is human, or do we have to prevent the loss of this assumption?**

# We assume some minimal SmartHome components in the system architecture.

- X-10 as Command & Control
  - utilizes the electricity distribution system
  - low-cost, low-power, consumer & professional
- Multiple TVs/Video-Displays
- Integrated TV/PC/Internet communication
- Next-Generation Remote Controls<sub>(programmable)</sub>
- Voice & IR control, RFID tags
- Motion Detectors & Video Cameras

# The Production Process limits us to actual, filmed sequences.



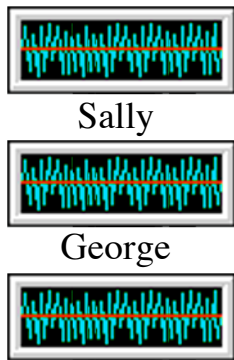
# Personalizing Audio Segments

Library of pre-recorded names (e.g.).

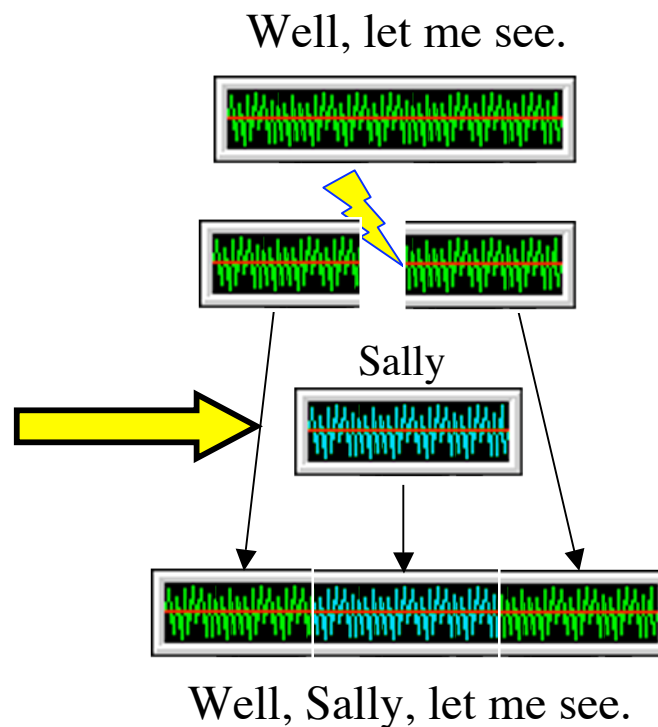
Sally

George

Good Buddy



or, *Voice Cloning*



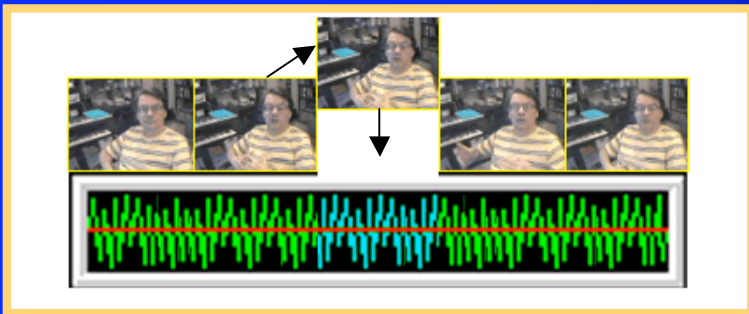
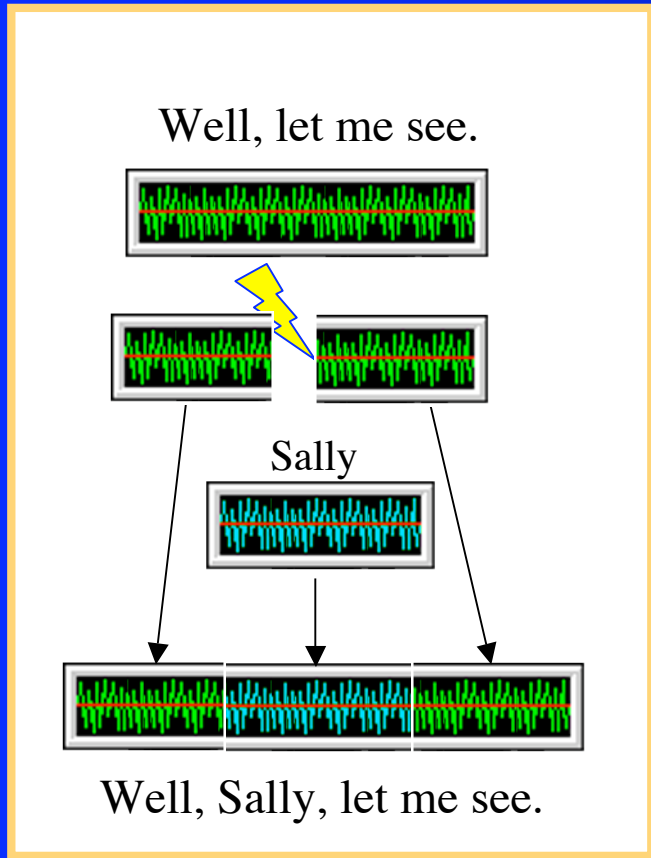
*But of course, this completely goofs up the video frames.*

Chatterbots can use a set of well-known “tricks” to maintain the illusion of their “humanness.”

- Randomly change the subject.
- Use Questions to get the user to talk about him/herself.
- Use the user’s statements to formulate replies.
- Admit ignorance.
- Rigidly continue the previous topic.
- Use humor, controversy, insults, etc.
- Excerpt USENET news stories.
- Simulate human typing errors.
- Utilize “stories” to maintain continuity of dialog.

**Q: Can these or similar tricks work effectively in VideoBots?**

# Personalizing Audio Segments



- **To fix the Video:**

- can morph the two end frames
- can simply duplicate the end frame of the first segment (simply “stretching” the video)

- **This provides the needed, extra video frames, but...**

- the video is no longer lip-synched
- stop-motion may become observable

- **However, ...**

- this whole process can be automated
- all statements can be cloned into “personalized” versions, in addition to the original, unpersonalized audio/video
- supports multiple personalizations, too

# Generic & Non-committal Responses

- **Eliza:**

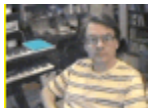
- Why do you believe that  $\langle X \rangle$  ?

- **Hex:**

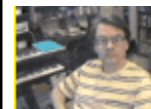
- OK, let's change the subject, shall we?
- or, generic insults, etc.

- **VideoBots**

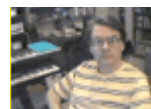
- Can't employ *Eliza-type* tricks.
- But can easily exploit a set of “generic” responses, segues, questions, noises, insults, etc.



OK, I'm totally confused. Sorry, but can we start all over on that?



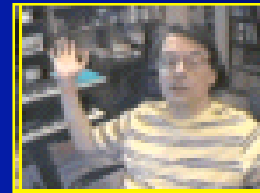
That's really outside my area of expertise.



Hey, Bub, if you want me to pay attention, you'll have to speak Greek.

# Full Exploitation of Emotions

- All, most, or some statements can be made under different emotional states:
  - boredom
  - excitement
  - politeness
  - etc.
- Emotion can be driven by
  - the nature of the dialog
  - the nature of the “state” or “content”
  - etc.



*Much easier to express & exploit emotion than in text-based chatterbots.*

# Diversionsary Video Clips

To incorporate non-lip-synched audio clips,

- the bot can “turn away”  
*(from the camera/viewer)*
- can cough and cover his/her mouth
- play with his/her mouth
- chew a pencil
- etc.



Umm, let me  
see...

It's on at  
7:00 tonight.

I'll record it for  
you on VCR2.

# Many Other Audio/Video Tricks Possible

- Repeat statements but use different intonations.
- Repeat statements but use different video clips.
- Use simulated distance between the VideoBot and the User, e.g.,
  - Jail-house window-panes and phones
  - space-ship command & control room
  - simple video-phone setup
  - artificial environment
- multiple “bail points” within the A/V
- visual clues to indicate “non-listening” mode

# Summary

- VideoBots have been shown to be more “instantly” believable as “Intelligent”.
- Suspension of Disbelief should be easier to maintain, as well as easier to achieve.
- Many ChatterBot tricks can be exploited by VideoBots, but some are problematic.
- Production costs & difficulties are stronger.