

VideoDIMs as a Framework for Digital Immortality Applications

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

IVA 2003
Irsee, Germany
September 15-17, 2003

Presentation Outline

- Some Goals of Digital Immortality (*only some!*)
- Digital Immortality as an Agent Technology
- Relate VideoDIMs to ChatterBots
- “Believability” and “Credibility”
- My VideoDIM Framework
- Status of Project

Multiple Forms of Immortality are Being Pursued

Biological

Fame

Religious

Uploading *

Genetic

Memetic

Digital Immortality *

Roughly: A Personal Database is made of all of a person's photographs, videos, audio recordings, documents, diaries, interviews, love letters, notes, papers, etc., etc. over a lifetime.

An **interactive agent / avatar** is placed in charge and equated with the collected "memories", events, and assets of the person.

There is a continuum of "types" of Digital Immortality.

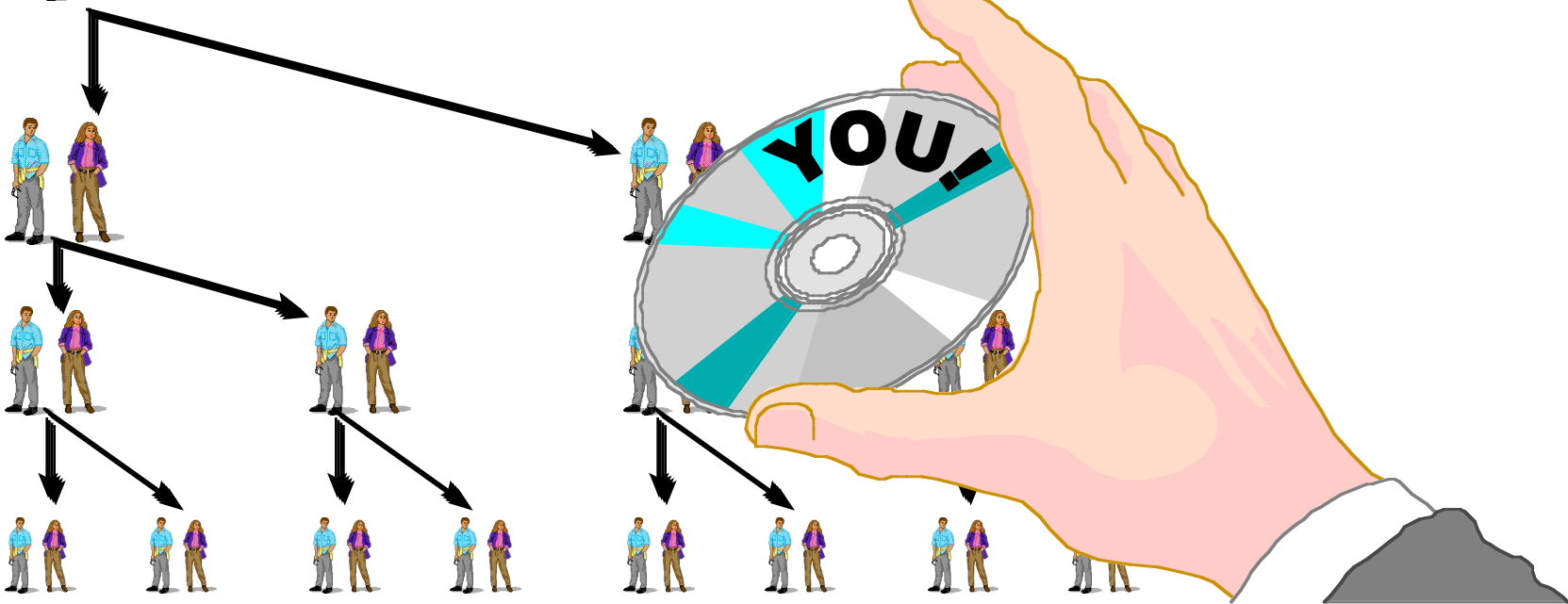
* Digital Immortality and Uploading are complexly interrelated.

Technological Impetuses for Digital Immortality

- **Digital Video & Audio**
 - Digital cameras, videocameras, voice recorders, etc.
- **Digital Photo albums and Movies, Vanity DVDs, etc.**
 - Autobiographical, voice annotations, multiple media.
- **Electronic Writing, Archiving, etc.**
 - Emails, diaries, chat logs, ...
- **Huge increases in storage capacity, at lower costs.**
 - Can (soon) store an entire person's life on disk (cf. Snow Crash).

Everyone Gets One!

Because it's digital, you can make as many copies as you want, each just as perfect as the others. Now you can afford to maintain you entire family's legacy of high-quality family videos. And you can find what you want, who you want, when you want it.

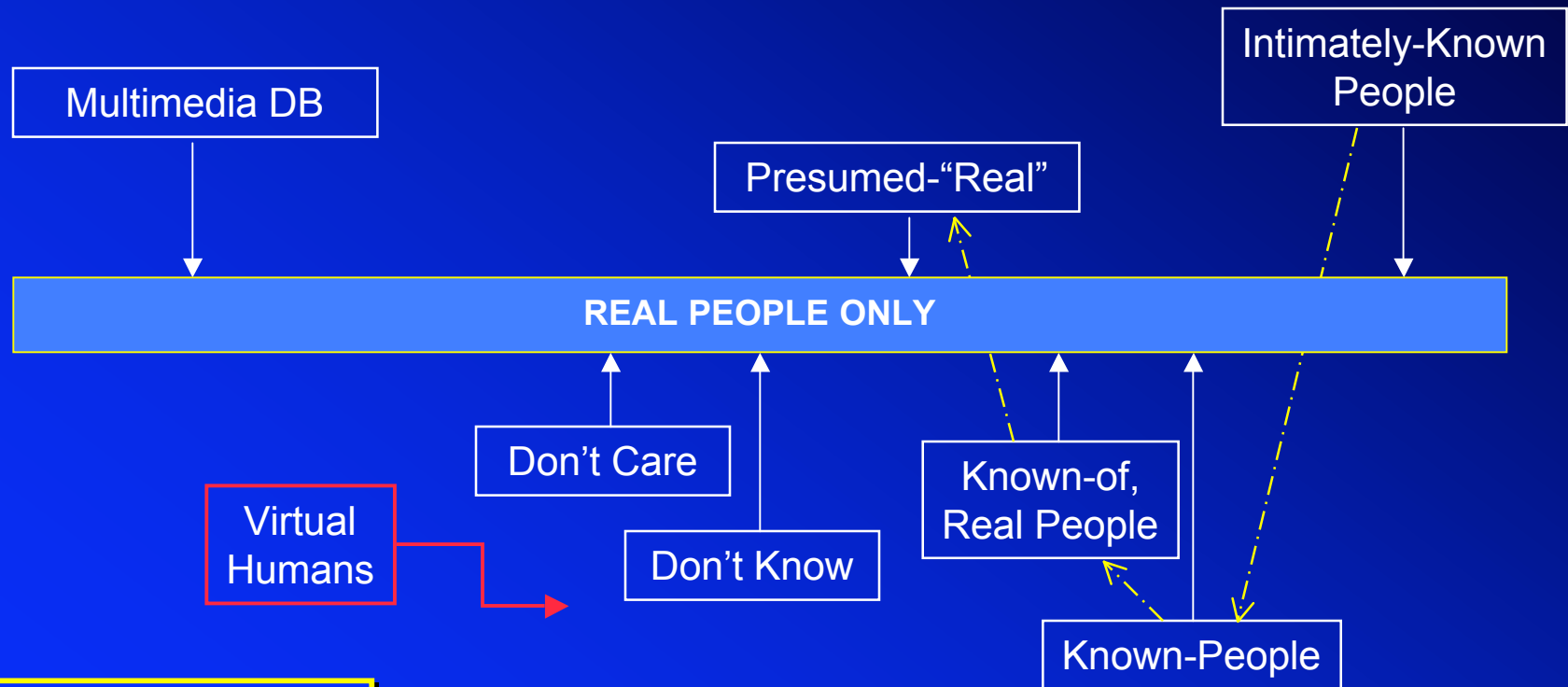


LIFE BITS: A Microsoft Research Labs Project

“ We believe with Ray Kurzweil, Hans Moravec, and others, that it is likely there will be more and more faithful avatars over the next century. By 2040, Moravec predicts that robots will be as smart as humans. Successive generations of **question-answering avatars** will gradually become indistinguishable from the actual persons we know and love in 2001, **enabling that person to appear to ‘live forever’**. ”

Gordon Bell & Jim Gray, MSR

The Digital Immortality Continuum



Things are surely much more complicated than this.

Note that many of these classifications change over multiple lifetimes.

VideoBots compared to VideoDIMs

VideoBots

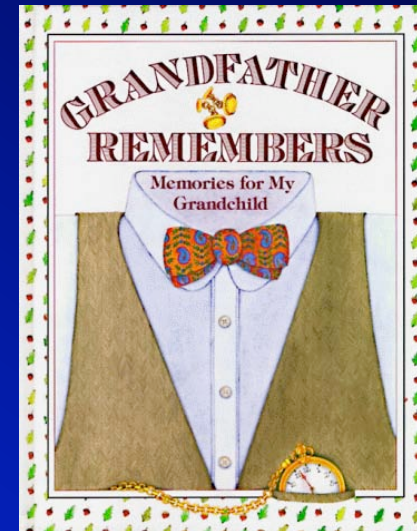
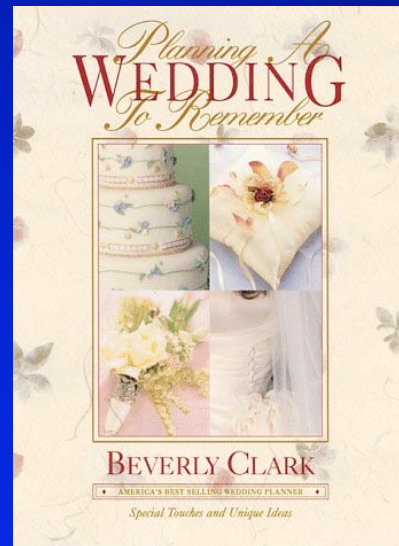
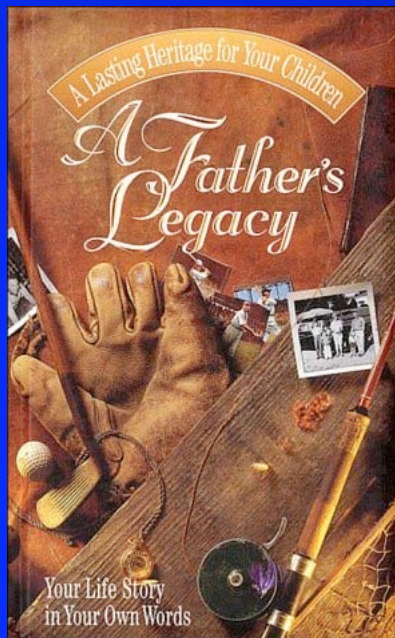
- Video-based visual interface
- Focus is on the “agent” aspects
- Presumably a “virtual” human.
- Presumably, no one knows the “real” person on whom it is based.
- or else it doesn’t matter:
 - Annanova
 - The Elite Models
 - Final Fantasy

VideoDIMs

- Video-based visual interface
- Focus is on the “avatar” aspects
- Presumably a “real” human.
- Presumably, he/she will be known to the user.
- Identity becomes key to believability.
- Can effectively be seen as a VideoBot when the user does not know the “real” person.

- They are used for very different purposes.
- Thus even though they are very similar technologically, their “success criteria” differ significantly.
- The resulting technological implications are extremely significant.

We already have many “kits” for recording our “lives” ...



... and for other people's
use, not (just) our own!

Examples of Interactive VideoBots



Hank and his girlfriend (in *The Sixth Day*).

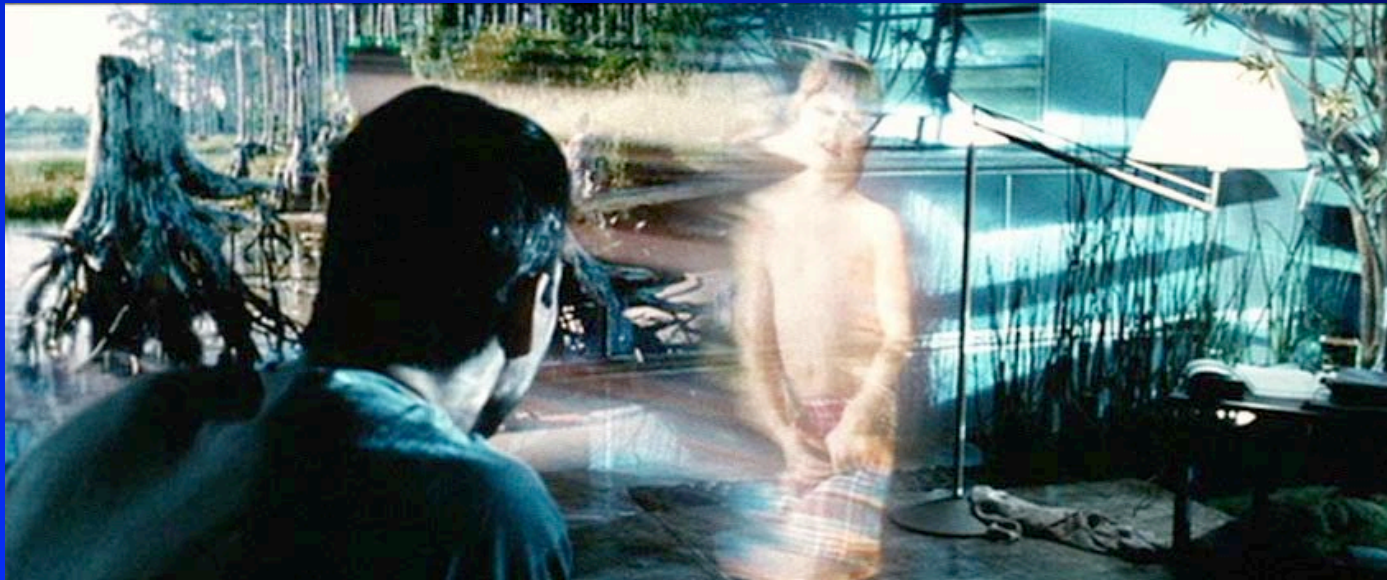


The Fifth Avenue Public Library's informational kiosk: a VOX – registration NY-114.

VirtuaGirls: tiny VideoBots on your screen.



Visiting the Past with Digital Immortality



John visits and talks to his dead son, Sean.

Minority Report, Philip K. Dick

Sarah Bellum, our proto-TV Agent VideoBot.

- Sarah “lives” in your TV.
 - (actually, in your TV “system.”)
 - (well, actually, in your “House”)
 - (Actually, we’re not sure where she (should) live(s).)
- She is smart, so you don’t have to be.
- Like a Butler, you tell her only once (not every time).
- She learns your behaviors and preferences.
- The longer she works for you,
 - the less you have to interact with her
 - the less you will probably see her.
- She knows who you are and where you are.
 - Voice ID, RFID tags, etc.
- So she can find you to remind you of your schedule, to notify you of email, phone calls, caller IDs, alarms, etc.
- Animated, speech recognizing persona.



Some Central Research Issues:

Within the domain of

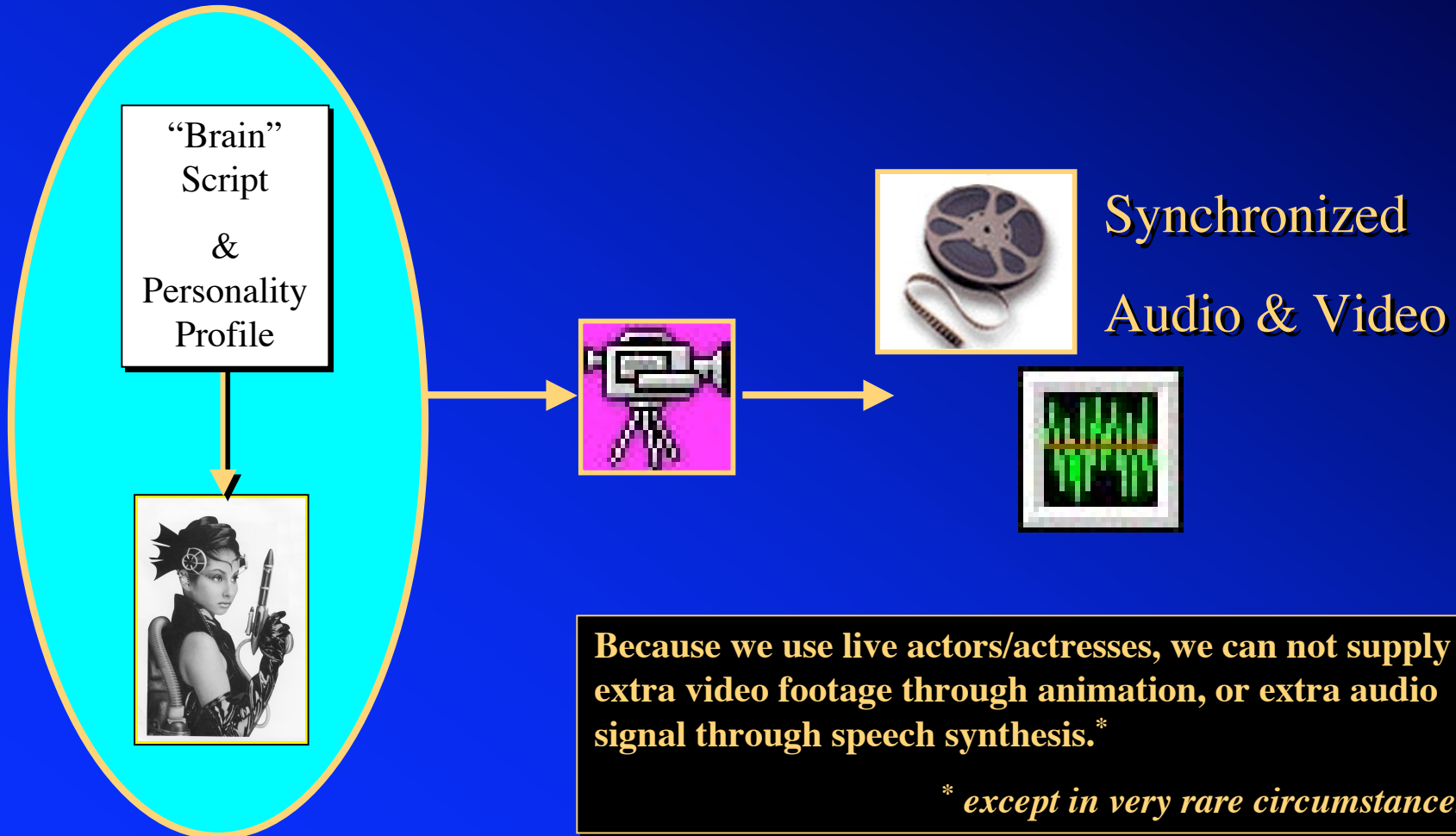
Visual/Animated & Conversational Agents,
much emphasis is placed on

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

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

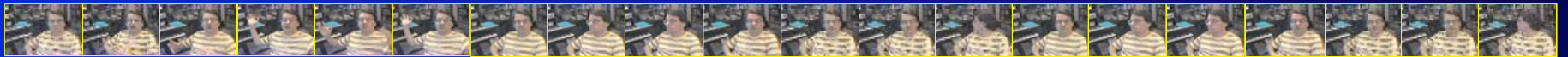
- **do we start out with “believability”?**
- **is it easier to achieve & maintain “suspension of disbelief”?**
- **do we still have to trick the user into believing the machine is human, or do we have to prevent the loss of this assumption?**

The Production Process limits us to actual, filmed sequences.

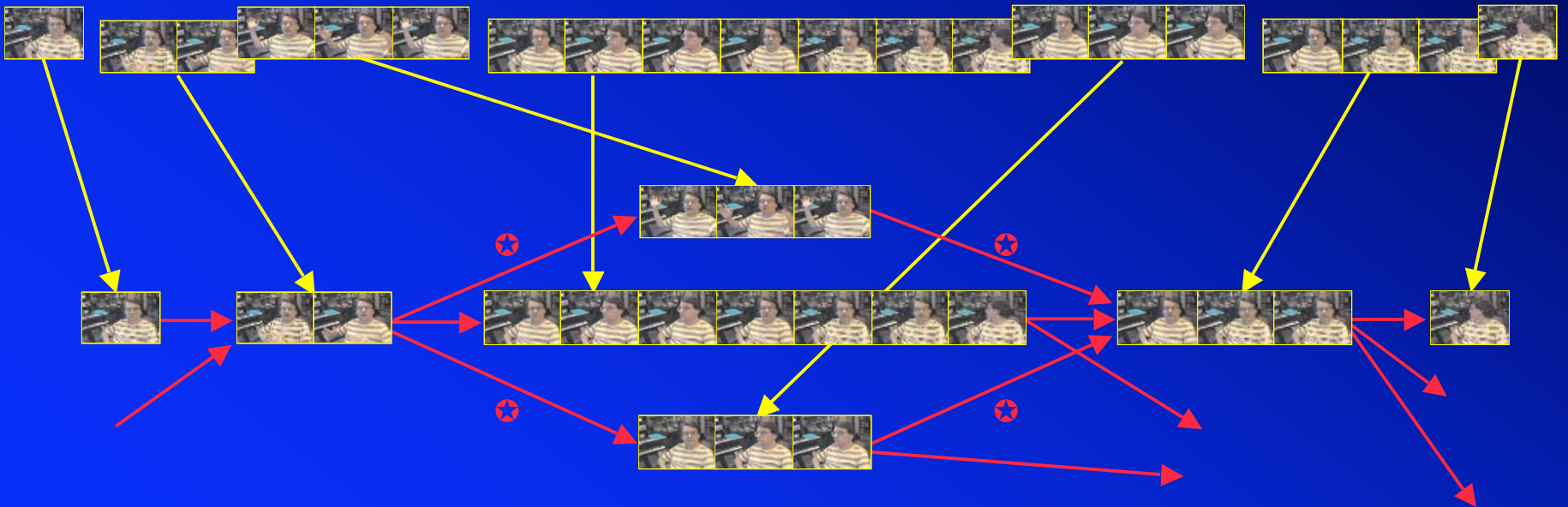


Linear to Random Video

1. The video is captured normally – in scripted, linear sequences



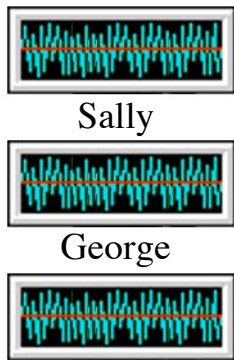
2. Break points are chosen / identified.



3. Morphing “joins” (★) allow segments to be replayed in alternate sequences.

Personalizing Audio Segments

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

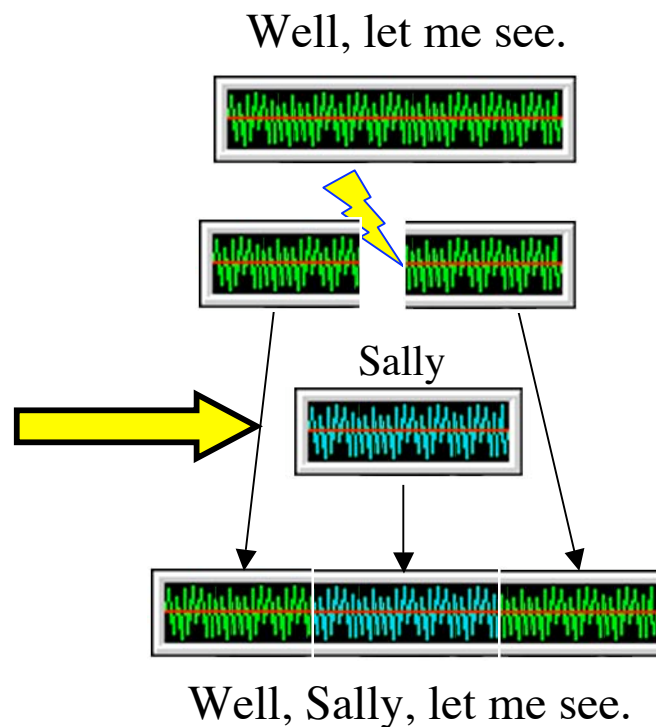


Sally

George

Good Buddy

or, *Voice Cloning*



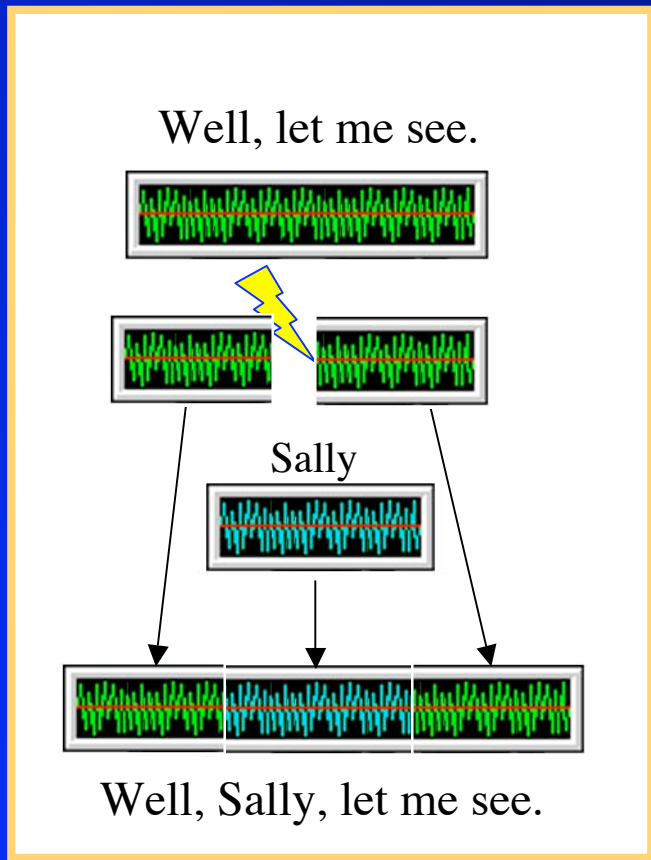
But of course, this completely goofs up the video frames.

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 <X> ?

- **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.



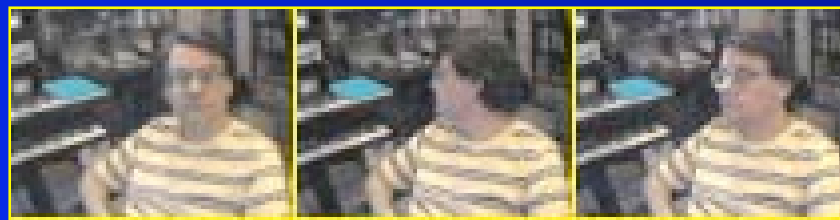
- Emotion can be driven by
 - the nature of the dialog
 - the nature of the “state” or “content”
 - etc.

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.

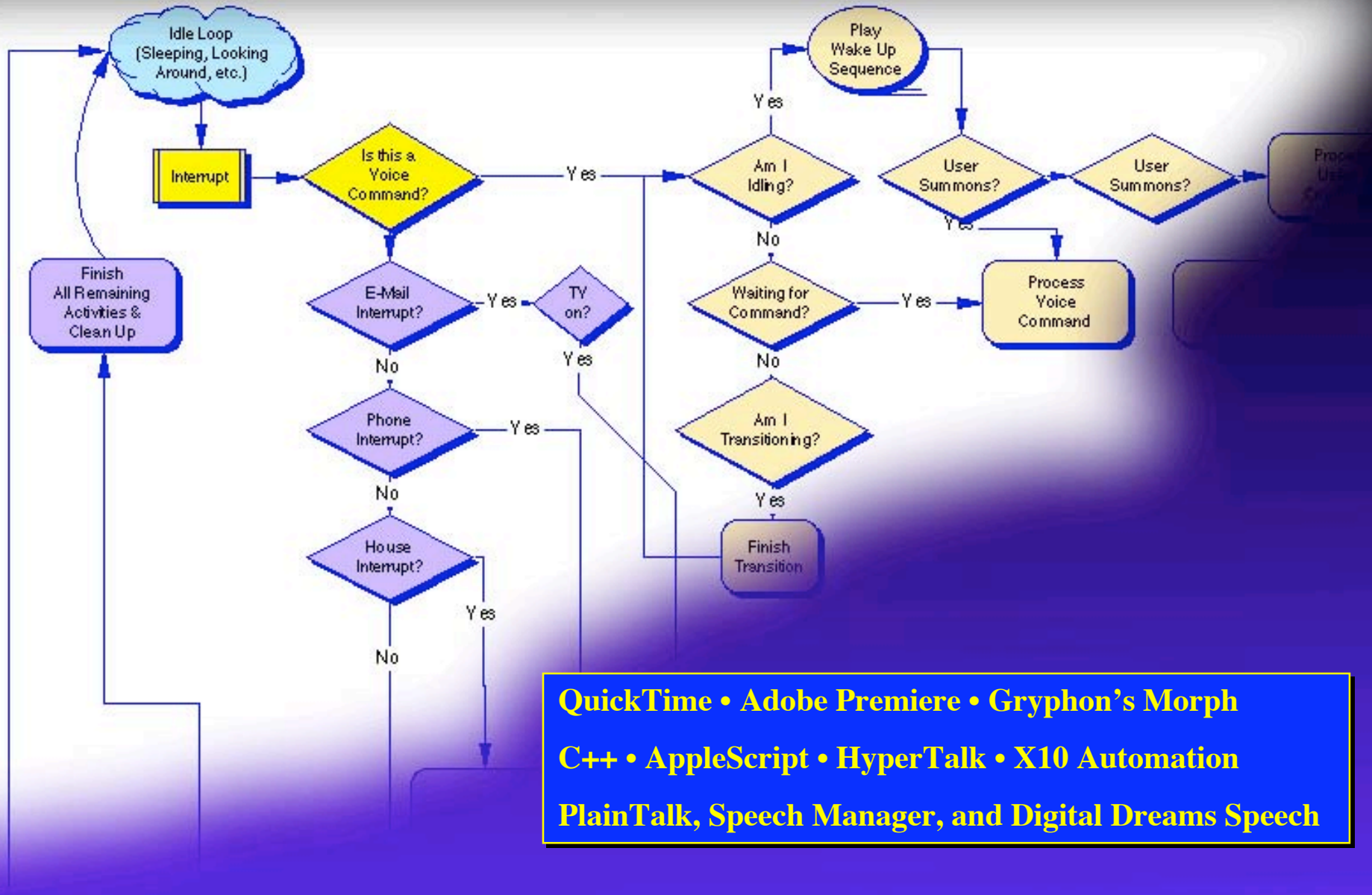
Many Other Audio/Video Tricks Possible

- 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.

At the heart of our VideoDIM is an interrupt-driven OS kernel.



**QuickTime • Adobe Premiere • Gryphon's Morph
C++ • AppleScript • HyperTalk • X10 Automation
PlainTalk, Speech Manager, and Digital Dreams Speech**

An Example Interaction with Sarah

The Human Says...

Sarah, come here please.

What is Said

Here I am, Jane.

Video Clip(s) Shown

(Sarah comes up on-screen.)
Statement video clip
Idle, awaiting command (Group!)
(We can show any number of idle clips while we wait for Jane to speak again.)

Are there any movies on now?

Sure, and it's a love story!

idle, listening (while Jane speaks)
Sarah thinks for a moment.
idle, busy
Statement video clip
idle, awaiting response

OK, let's watch it.

OK, Jane, let's!

idle, listening
Statement video clip
(maybe no audio here)
Sarah punches a few buttons or something, and the TV channel changes to the appropriate channel.

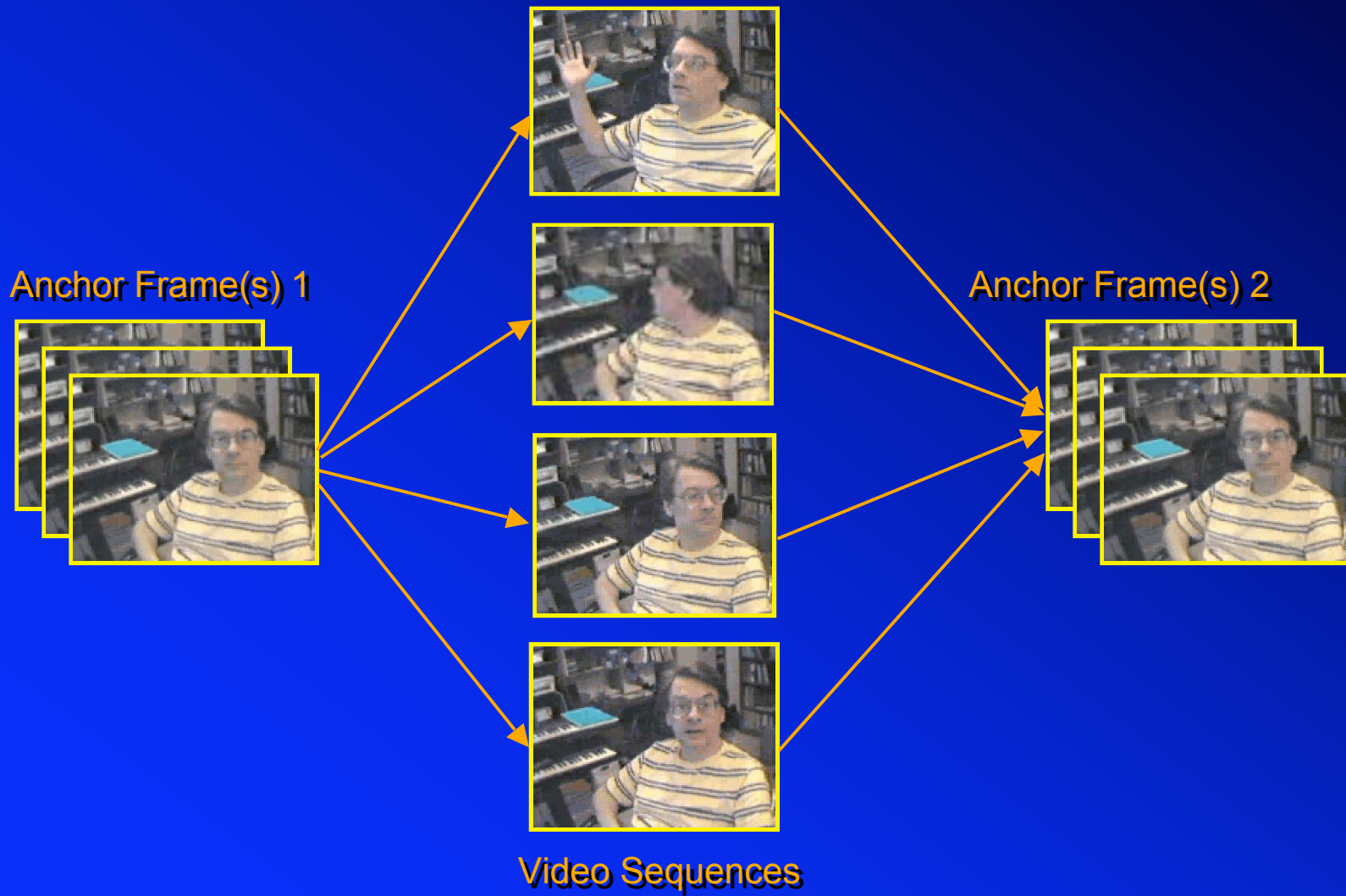
Sarah becomes "idle"; fade out.

Providing Pseudo-Realistic Responses:



- There must be multiple responses to every query or instruction.
- These must be replayed randomly.
- All video sequences must begin & end “identically”.
- Morphing tools are required to make the frame transitions seamless.

Constructing Alternate Sequences



Some Sample Views of “Buster” at Work.



“Buster, please find me a Horror Movie to watch.”
“Well, let me see if we have any right now. Yes, ...”



“Hey Buster, do NOT let me forget to watch Seinfeld this Friday.”
“OK, right. That’ll be easy.”



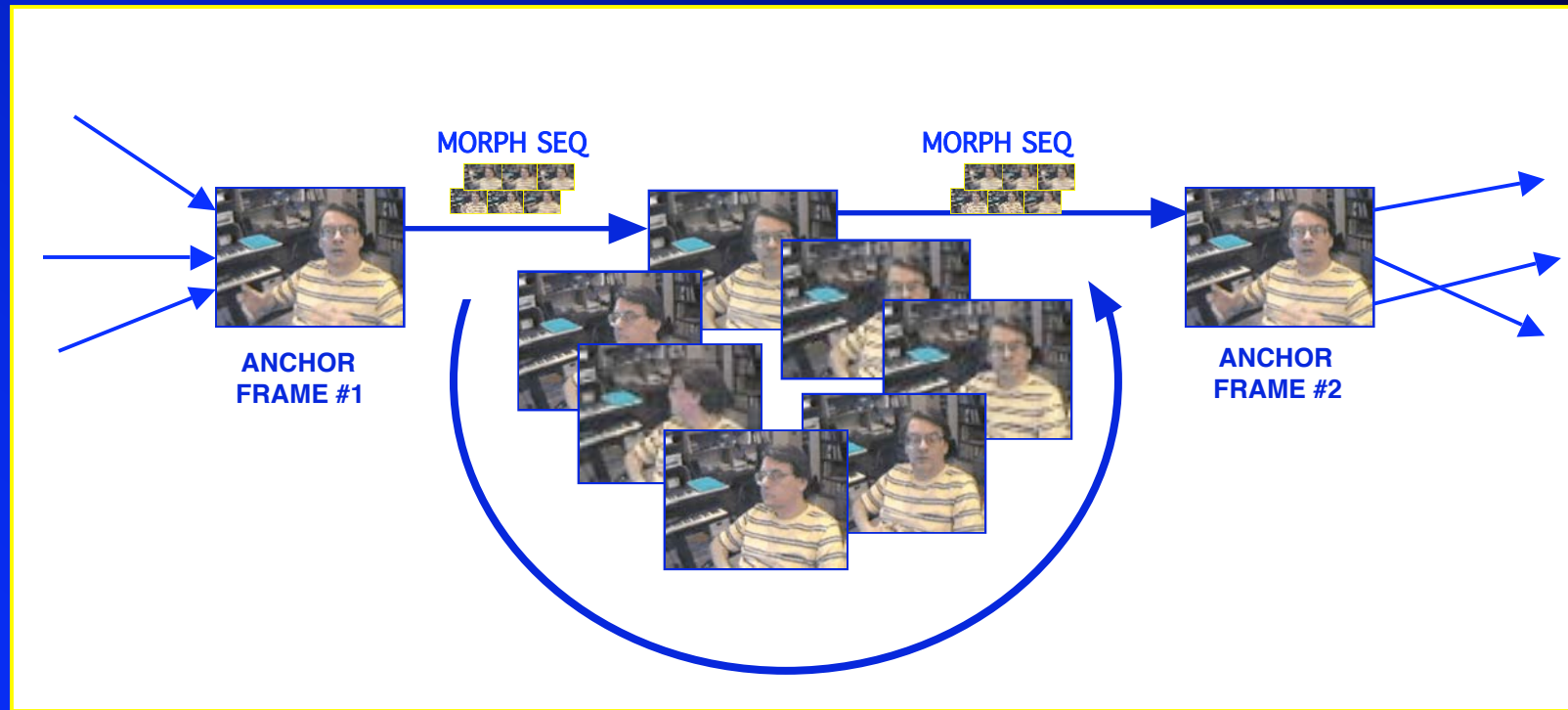
“Buster, please find “Oprah.”
“Oh my God, we can’t do that.”



“Oh, Yoo-hoo, Buster; are the X-Files on tonight?”
“Yep, but it’s a re-run; you’ve already seen it. But next week is an all new 2-hour special. I have presumed you will want to watch it.”

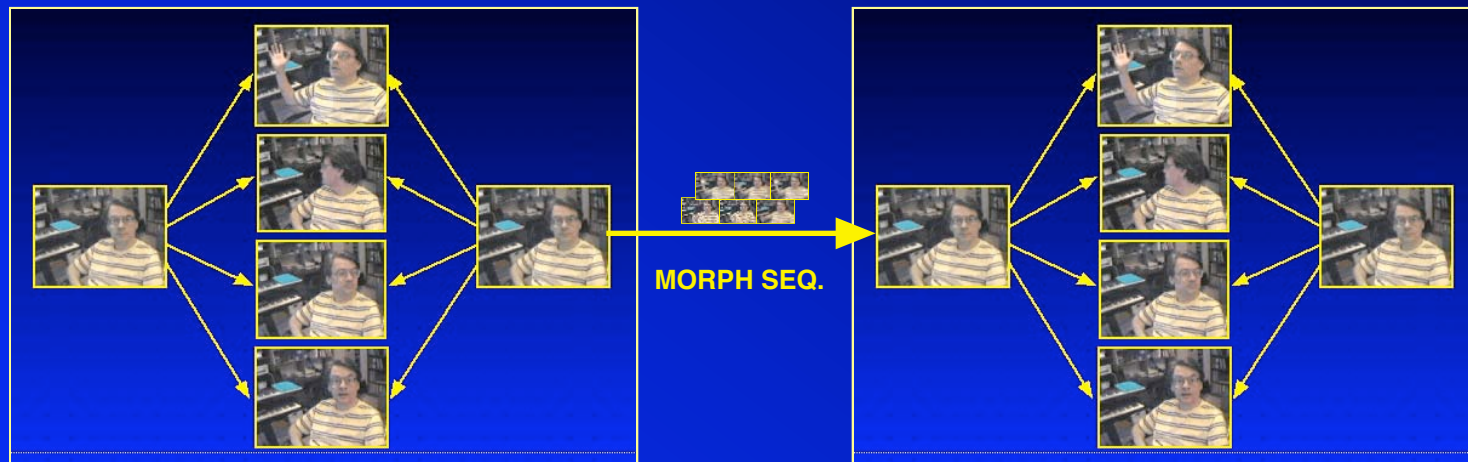
In the prototype, Movements and Emotions are purposely exaggerated.

Constructing Idle Loops

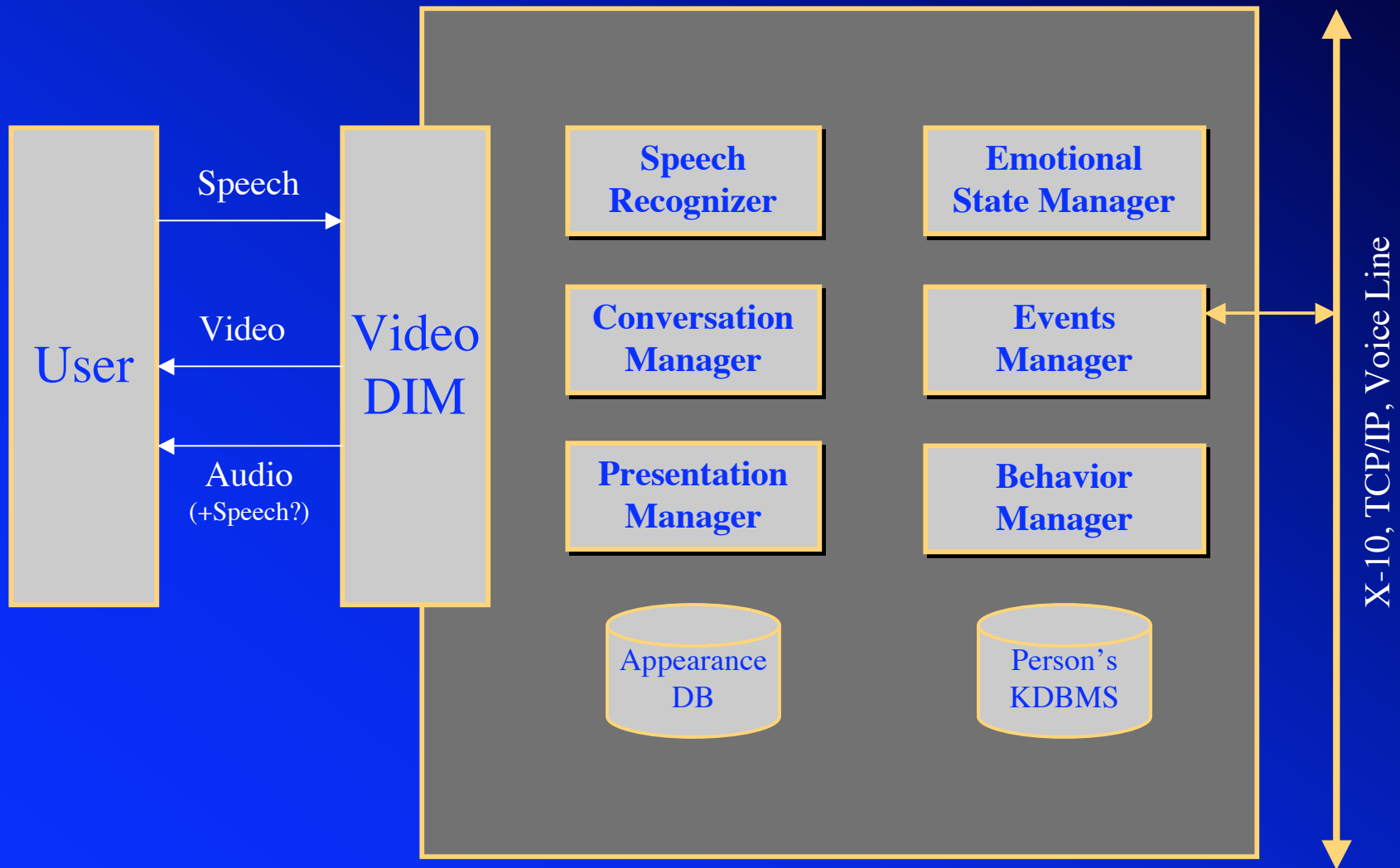


- Idle loops can be constructed from any number of “human”, “emotive” behaviors.
- (wasting time, drumming fingers, making faces, etc.)
- There must actually be multiple ways of entering each loop.
- Thus extensive sequence morphing is often required.

Constructing Complex Sequences of Behavior



Our VideoDIM Framework



Conclusions & Summary

- **Status:**
 - It works (complete prototype system); needs upgrading to OSX.
- **Problems:**
 - Severely limited A/V repertoire, though (very manually intensive).
 - No barge-in capability at present.
- **Experiences Gained:**
 - Scripting and Shooting requirements and styles.
 - Asset creation and management techniques.
 - Real-Time Conversation management techniques.
 - Diversionary & Trick techniques developed.
- **Focus:**
 - Major techniques for maintaining “suspension of disbelief”.
 - Understanding more about key differences between DIMs and other VHs.
 - Preparing for 2.0.