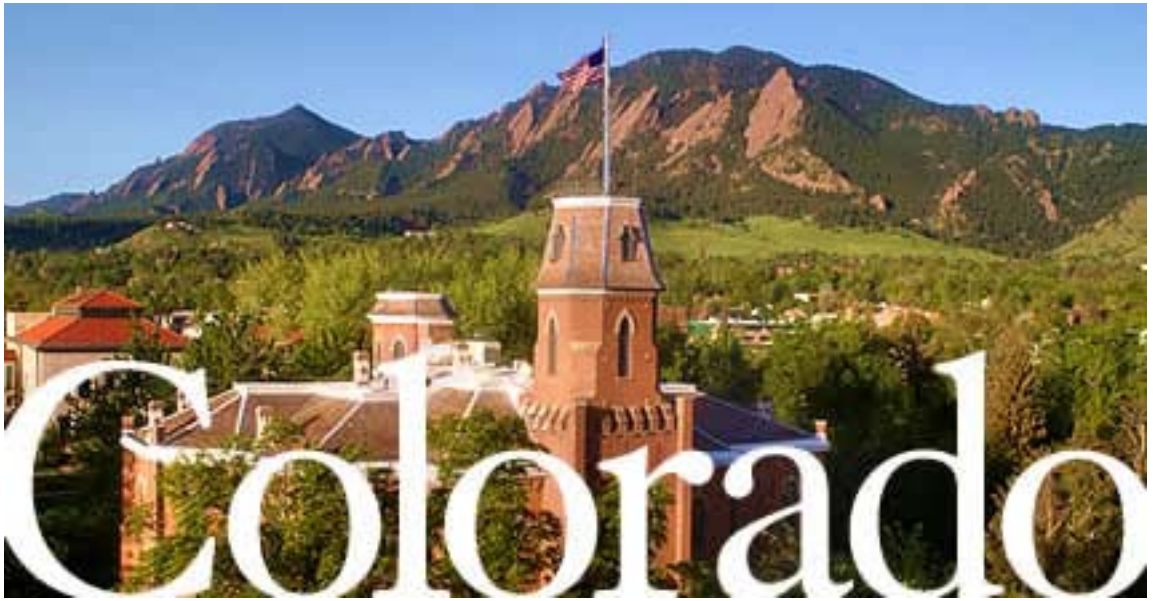
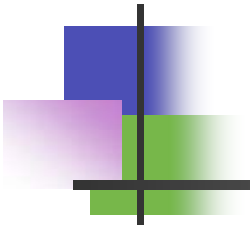
Inventing Virtual Humans
The Next Generation of HCI

Ron Cole
Center for Spoken Language Research

Sept 14, 2004

“The best way to create the future is to invent it” -- Alan Kay





Outline

- The Vision
- Virtual Tutors
- Virtual Therapists
- Grand Challenges

Vision

- **Teach people to read, comprehend text, learn new languages, acquire job skills…**
  - **Including individuals with sensory deficits, cognitive disabilities and neurological disorders**
- **Through conversational interaction with virtual humans**
- **That behave like sensitive and effective teachers or therapists**

Compelling Need

- **One-on-one tutoring provides the most benefit (the earlier, the better)**
- **No national educational or health system has sufficient resources to provide the required level of individualized attention to those who need it**

Solution?

- **Invent virtual tutors**
  - **Perceptive Animated Agents**
- **That behave like sensitive and effective (human) teachers or therapists**
- **To detect problems early, and help individuals learn in immersive, focused learning tasks**

What is a virtual tutor?

A Lifelike 3D Computer Character

- **with personality and attitude**
- **that engages users in natural face-to-face conversation**
- **to produce great learning experiences**

Marge

Creating Virtual Tutors is a Grand Challenge

- Understanding the social dynamics of face-to-face communication
- Inventing the machine perception and generation technologies
  - to model the exchange of signals and cues during face-to-face interaction in real time
- Combining knowledge and technologies in systems that pass the Turing test
  - is it a human “Wizard” behind the screen or a virtual human?

Evolution of Virtual Humans (A personal perspective)

- The remainder of this talk describes virtual humans I have helped develop to:
  - Teach vocabulary to students who are deaf
  - Teach concepts to students with autism
  - Teach students to read
  - Conduct speech therapy with individuals with neurological diseases and aphasia

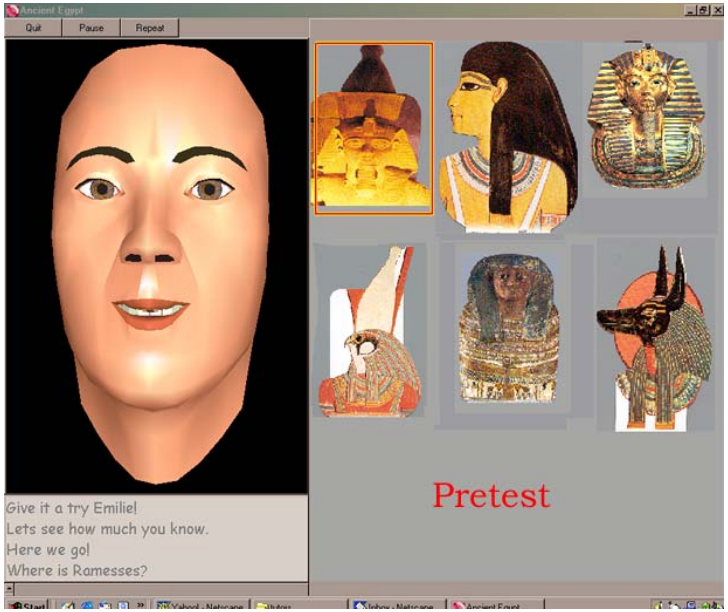
A virtual tutor at the Tucker Maxon Oral School

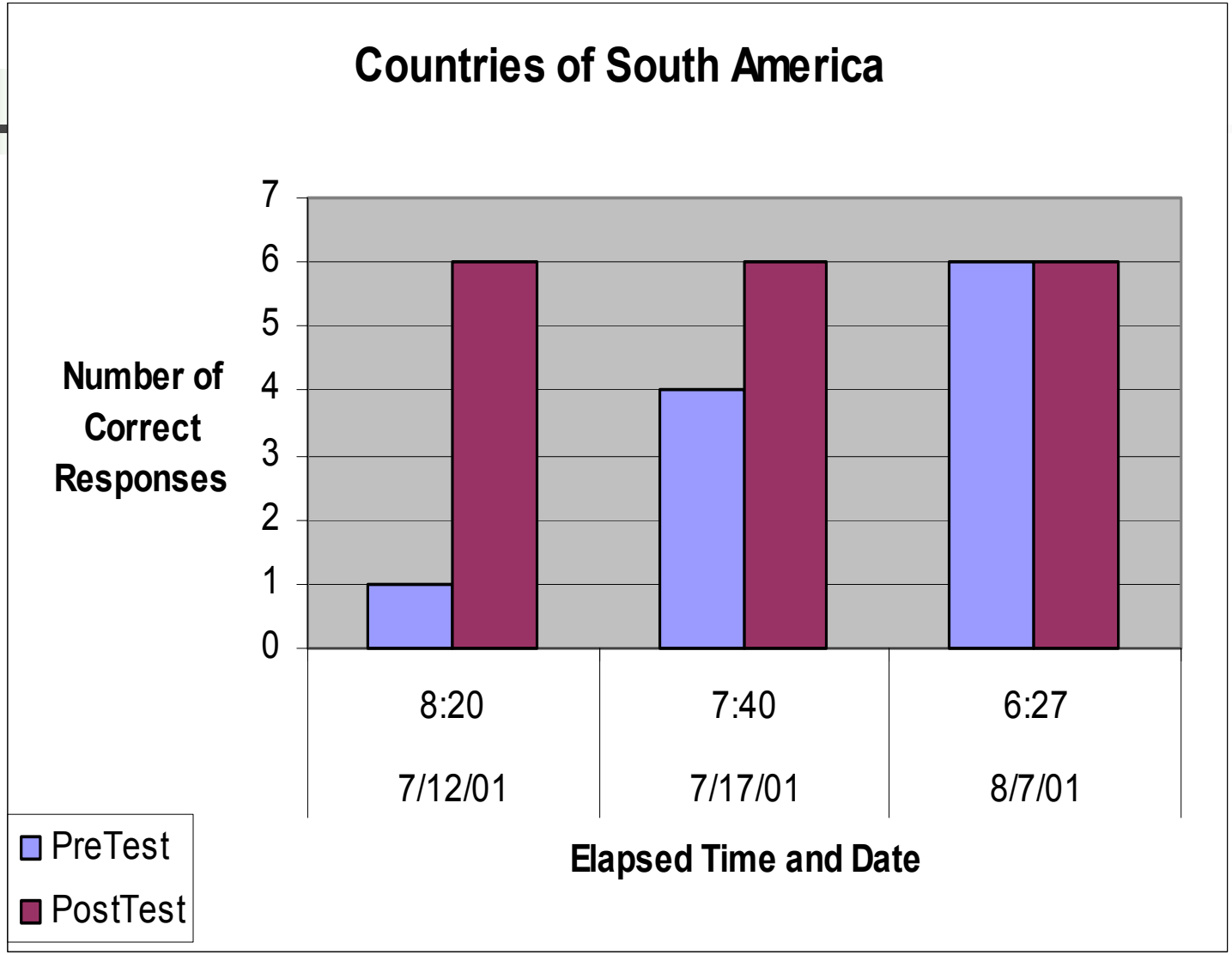
Students with hearing loss

- 1998-2001: Baldi teaches vocabulary to students at Tucker Maxon School
  - Rapid acquisition of vocabulary
  - >50% retention several months later
  - Dramatic improvements in speech production
  - Featured on national TV and NSF Home page
- 2004-2005: Marni teaches reading to students at Tucker Maxon School to read

Teaching Children with Autism (2001)

- **Summer 2001: 200 applications developed with CSLU Toolkit vocabulary tutor**
- **Six children with autism, PDD and Asperger were seen twice a week for 8 weeks**
- **Skills trained included mouse control, geography, telling time, astronomy, etc.**
- **Dramatic improvements in knowledge and skills**




Countries of South America
Number of Correct Responses
7
6
5
4
3
2
1
0
8:20
7/12/01
7:40
7/17/01
6:27
8/7/01
Elapsed Time and Date
PreTest
PostTest

2003: Colorado Literacy Tutor

Colorado literacy tutor

- A comprehensive, scientifically-based reading program
- Designed to teach any child to read
- By interacting with a virtual tutor

Cognitive theory & scientifically-based reading research

Skilled reading is

- Word recognition processes + comprehension processes
- This is called the "Simple Model of Reading" (Gough et al., 1996)

- **Word recognition processes**
  - Alphabet, Phonological awareness, Encoding, Decoding, Sight words
  - Reading in context until fluent & automatic
  - Evidence-based pedagogy: SBRR (Rayner et al, 2001)

- **Comprehension processes**
  - Discourse Processing Model (Kintsch '98, Perfetti & Marron '98)
  - Implemented in Summary Street comprehension program

Sponsors & grants

- **NSF/ITR: REC-0115419 - Kintsch, W., Landauer T., Caccamise, D., Cole, R., "ITR/PE: Latent Semantic Analysis Theory and Technology," $2,400,000, NSF, 09/01/01 - 08/31/06.**
- **NSF/IERI: EIA-0121201 - Kintsch, W., Caccamise, D., Cole, R., Olson, R., Snyder, L., "IERI: Scalable and Sustainable Technologies for Reading Instruction and Assessment," $5,997,404, NSF, 07/01/01 - 06/26/06**
- **NSF/ITR: IIS-0086107 - Cole, R., van Santen, J., Movellan, J., "ITR: Creating the Next Generation of Intelligent Animated Conversational Agents," $4,000,000, NSF, 09/01/00 - 08/31/05.**
- **NICHD/IERI : 1R01HD-44276.01Cole, R., Barker, L., Schwartz S., Snyder, L., Wise, B., "IERI: Scaling up Reading Tutors," $1,000,000.00, NIH. 9/27/02 - 9/30/04.**

Project team

Principal Investigator: Ronald Cole

Head of Research & Development: Sarel van Vuuren

Technology development and Integration: Bryan Pellom, Kadri Hacioglu, Wayne Ward, Javier Movellan, Jie Yan, Jiyong Ma, David Wade-Stein,

Software Development & Programming: Nattawut Ngampatipatpong, Jariya Tuantranont

Educational Research & Participatory Design: Barb Wise, Scott Schwartz, Lecia Barker, Kathy Bunch, Kathy Garvin-Doxis, BVSD educators and kids

Content: Taylor Struemph, Corby Connolly, Barb Wise, Chandra Bidwell

Comprehension: Donna Caccamise, Walter Kintsch, Eileen Kintsch

Assessment: Lynn Snyder

Components of the Colorado Literacy Tutor

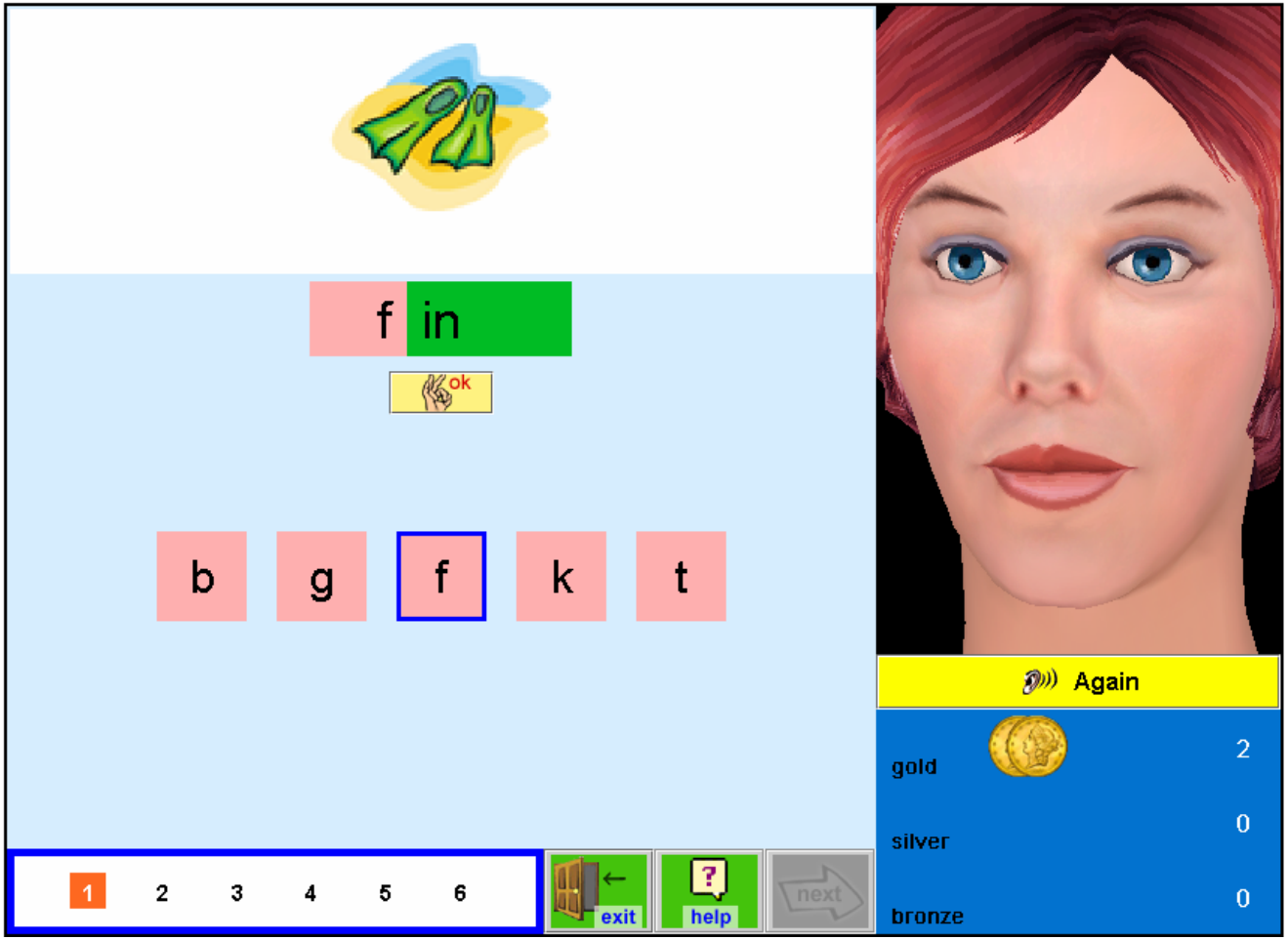
- Animated Learning Tools
  - Foundational Skills Tutors
    - Teach underlying reading skills
  - Interactive Books
    - Teach fluent reading & comprehension
- Managed Learning Environment
  - Tracks student progress
  - Controls Individual study plans

Main activity screen

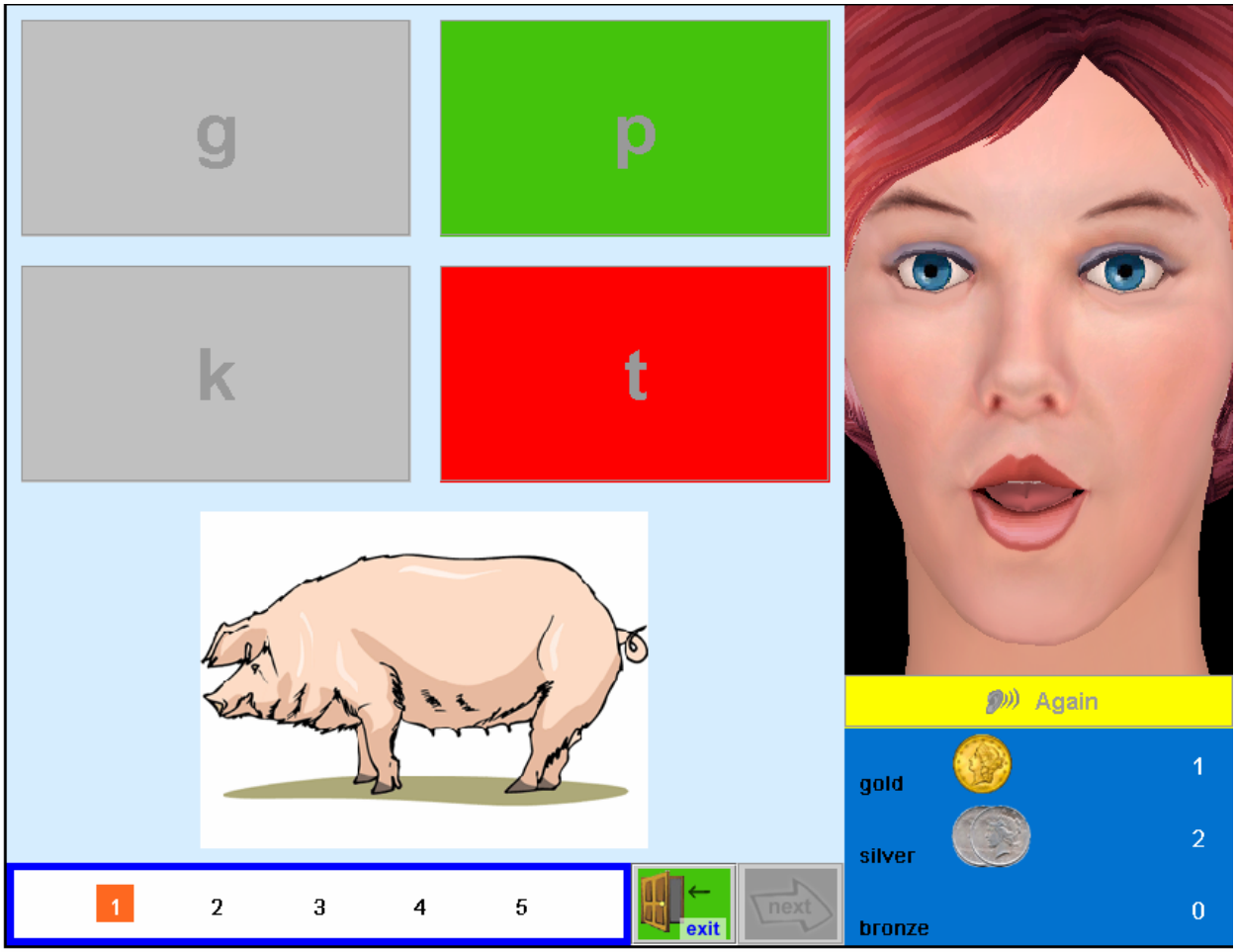
Foundational Reading Skills Exercises

- More than 30 phonological awareness exercises covering domains from alphabet, decoding, encoding, spelling, etc.
- Contextual feedback, reinforcement and individualized instruction by the Virtual Teacher
- All activities are customizable, logged and summarized for reporting in a Managed Learning Environment

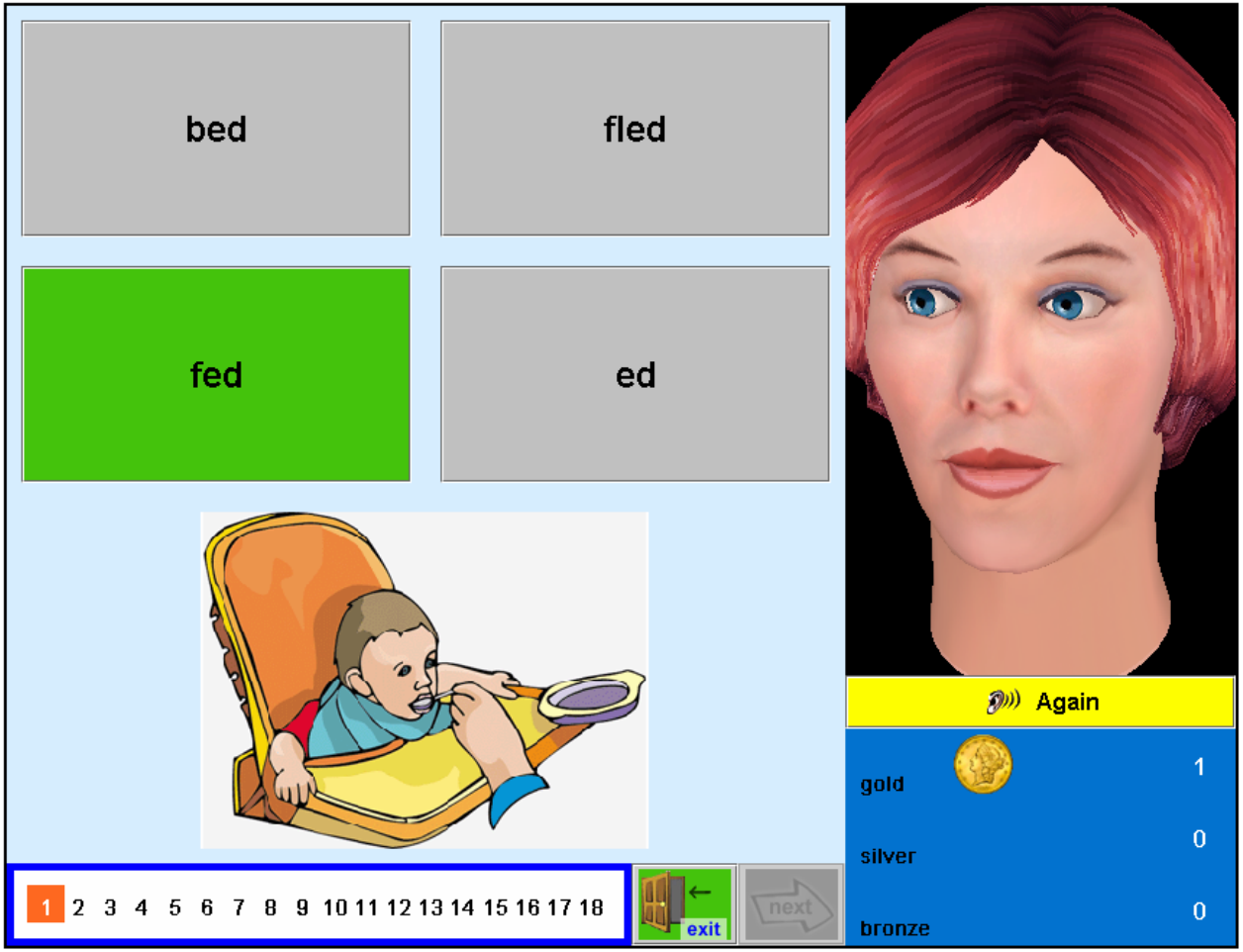
Beginning sounds



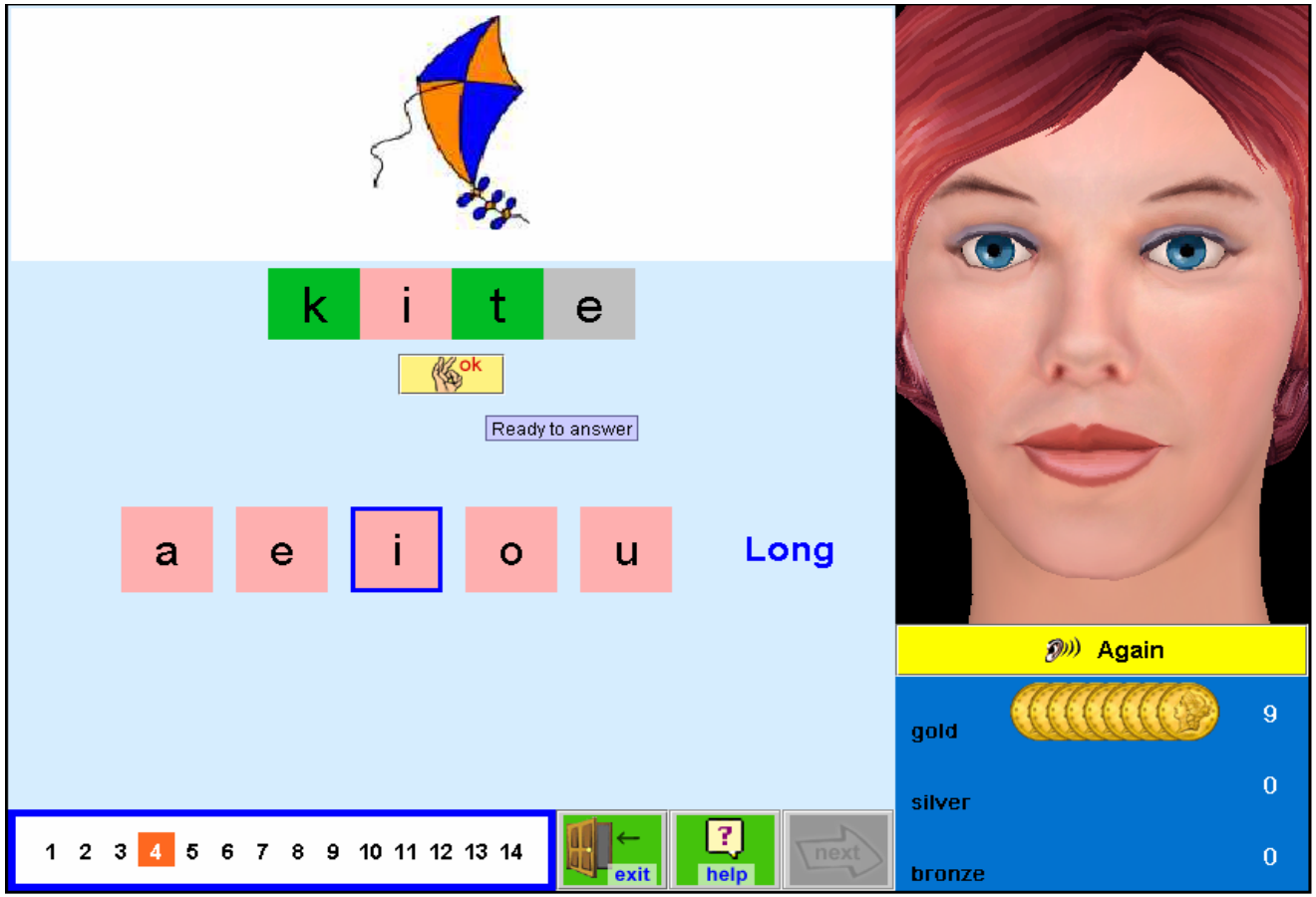
Letter sounds



Word reading



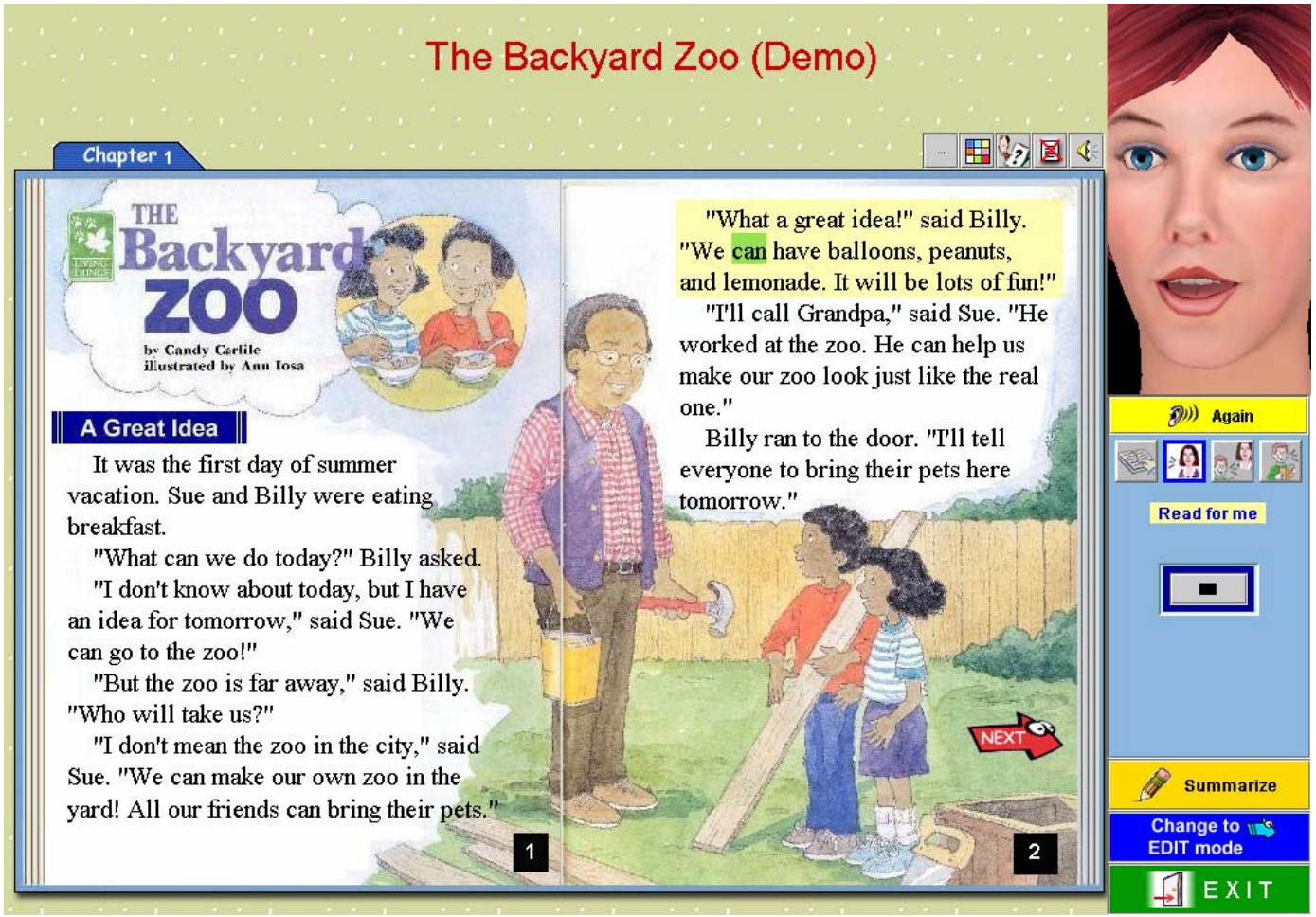
Silent /e/


k i t e
ok
Ready to answer
a e i o u
Long
Again
gold
9
silver
0
bronze
0
1 2 3 4 5 6 7 8 9 10 11 12 13 14
exit
help
next

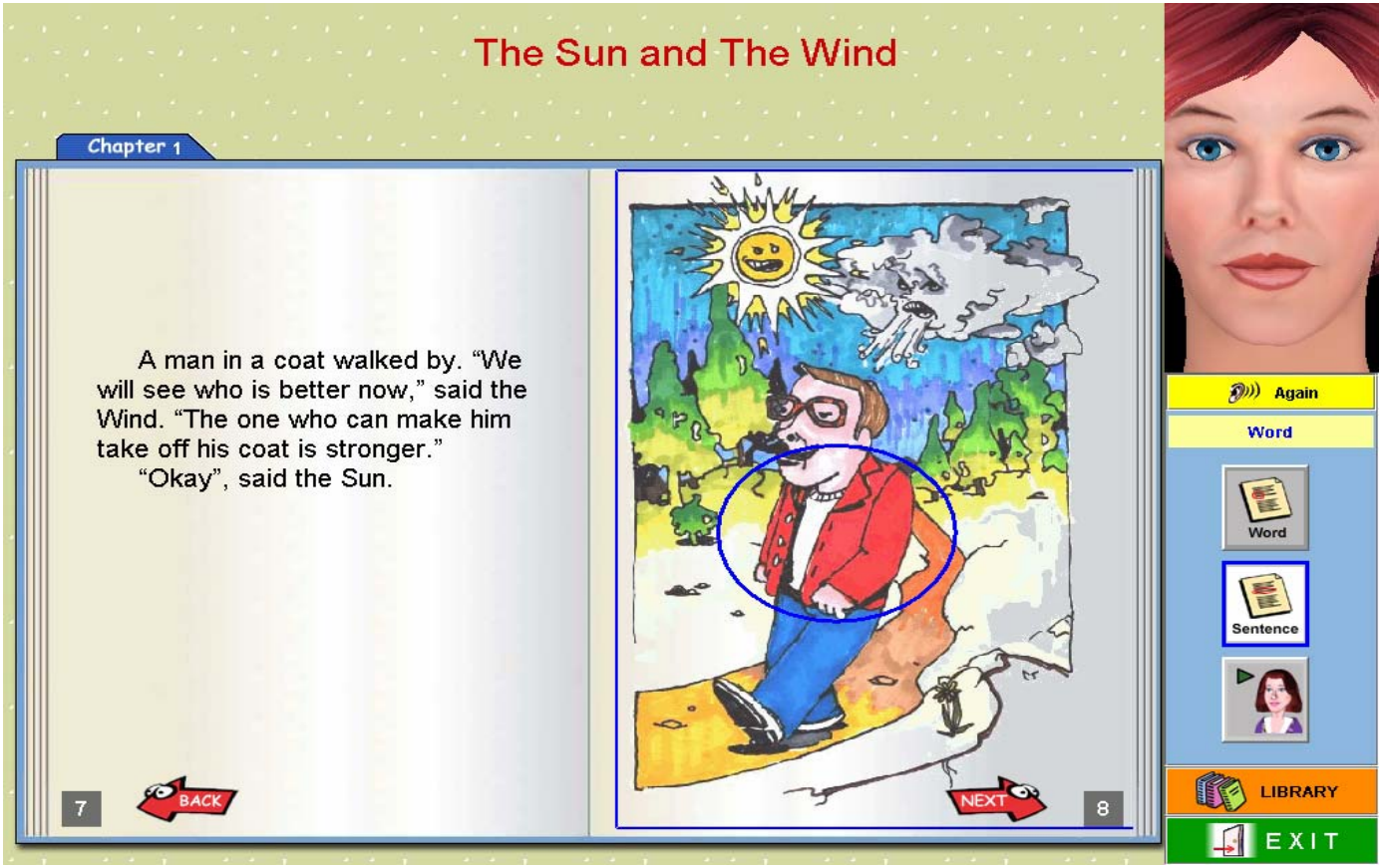
Interactive Books

- Teach word recognition, fluent reading and comprehension
  - Teacher reads to student
  - Student reads aloud while computer listens
  - Multimodal interaction
    - Assess comprehension by asking questions with student responding by clicking on images, answering multiple choice questions or summarizing the story

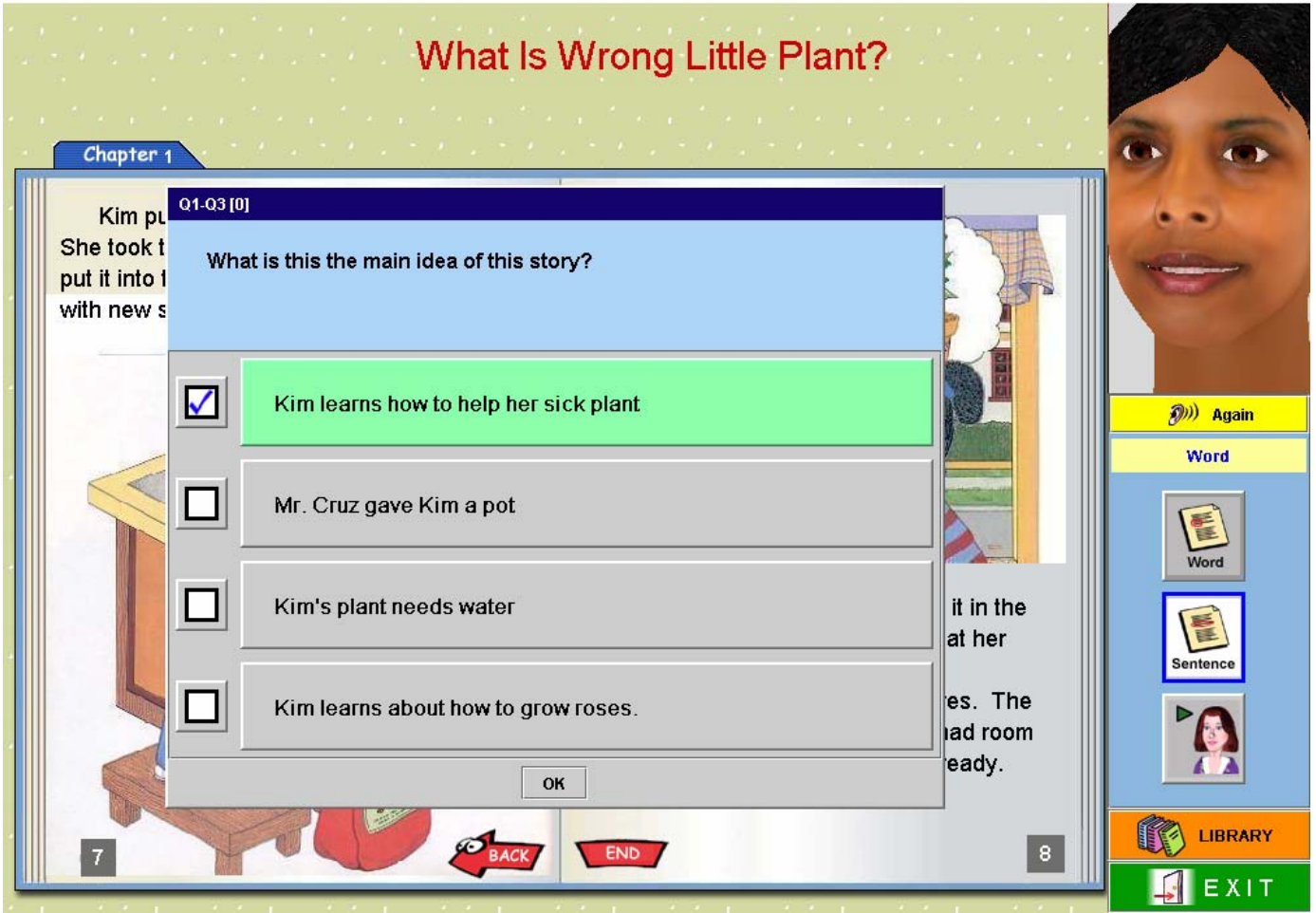
Read to me and read aloud

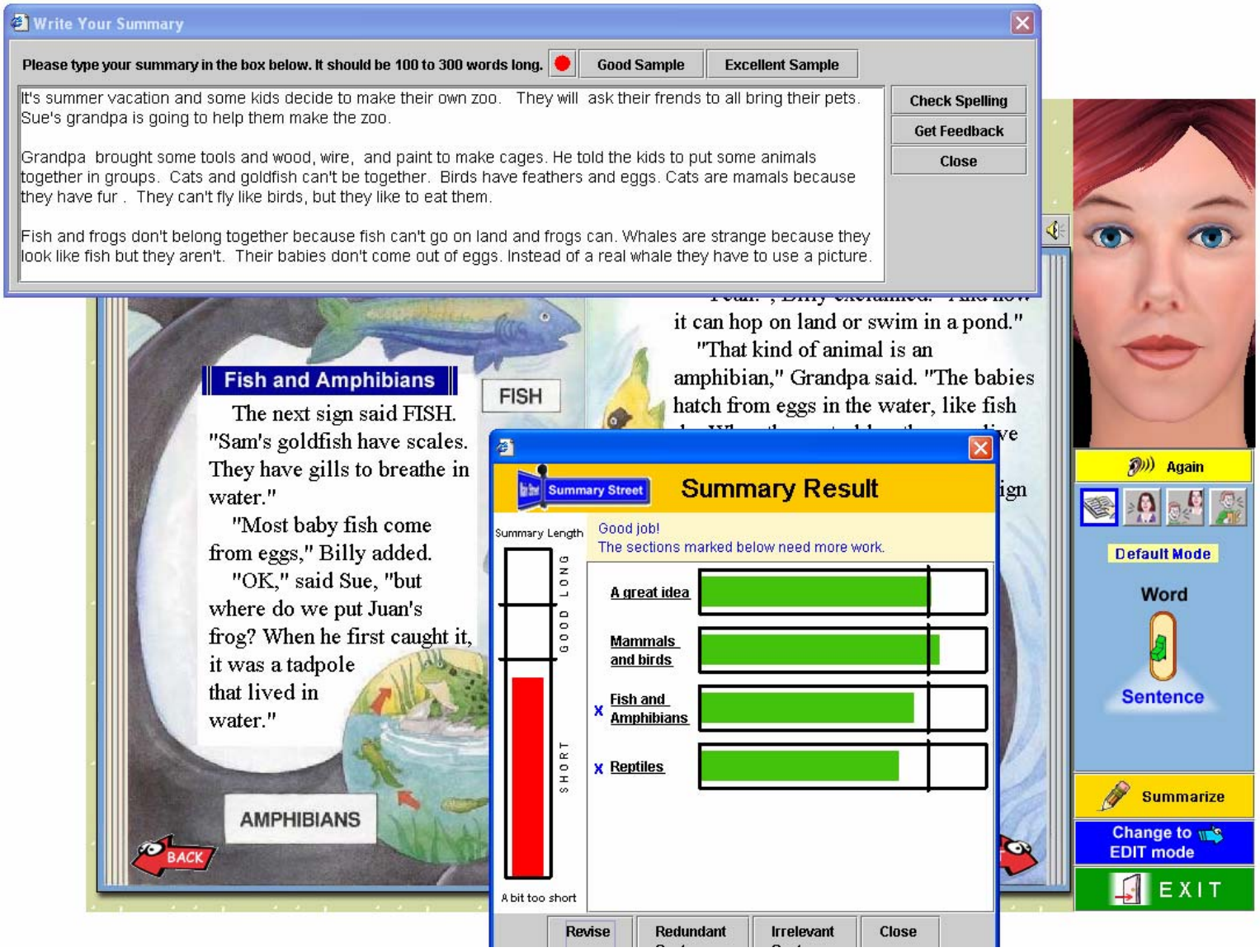


Click on image question interaction



Multiple choice question interaction

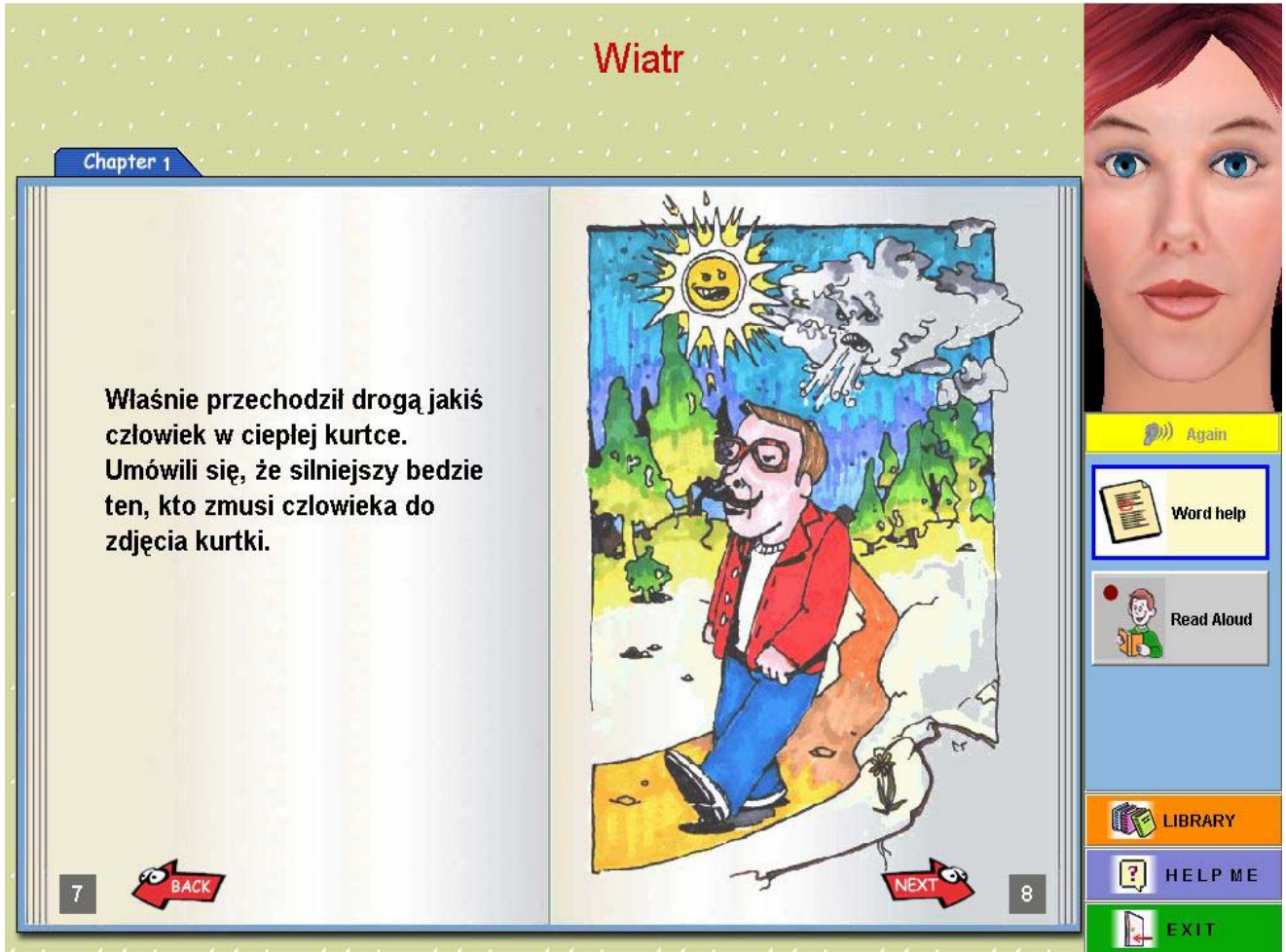



Write Your Summary
Please type your summary in the box below. It should be 100 to 300 words long.
Good Sample
Excellent Sample
It's summer vacation and some kids decide to make their own zoo. They will ask their frends to all bring their pets. Sue's grandpa is going to help them make the zoo.
Grandpa brought some tools and wood, wire, and paint to make cages. He told the kids to put some animals together in groups. Cats and goldfish can't be together. Birds have feathers and eggs. Cats are mamals because they have fur . They can't fly like birds, but they like to eat them.
Fish and frogs don't belong together because fish can't go on land and frogs can. Whales are strange because they look like fish but they aren't. Their babies don't come out of eggs. Instead of a real whale they have to use a picture.
Check Spelling
Get Feedback
Close
Fish and Amphibians
FISH
The next sign said FISH. "Sam's goldfish have scales. They have gills to breathe in water."
"Most baby fish come from eggs," Billy added.
"OK," said Sue, "but where do we put Juan's frog? When he first caught it, it was a tadpole that lived in water."
AMPHIBIANS
BACK
it can hop on land or swim in a pond."
"That kind of animal is an amphibian," Grandpa said. "The babies hatch from eggs in the water, like fish
Summary Street
Summary Result
Summary Length
GOOD LONG
SHORT
A bit too short
Good job!
The sections marked below need more work.
A great idea
Mammals and birds
Fish and Amphibians
Reptiles
Revise
Redundant
Irrelevant
Close
Again
Default Mode
Word
Sentence
Summarize
Change to EDIT mode
EXIT

Multilingual Literacy Systems

- Interactive Books in Spanish, Italian, German, French, Polish
- Includes
  - Speech recognition in each language--reading out loud with feedback
  - Animated speech production with accurate movement of the lips in each language

Polish Interactive Books



Spanish Interactive Books

Demos

- I will show you
  - Foundation reading skills activities
  - Interactive Books

Parkinson's disease

1.5 Million individuals US alone
Over 6 million worldwide

89% have a speech or voice problem
(Logemann et al.,1978)

4% receive traditional speech therapy
(Hartelius & Swenson, 1994; Oxtoby, 1982)

1990 Consensus: Speech treatment does not work
(Sarno, 1968; Allan, 1970; Green,1980; Aronson, 1990; Weiner & Singer, 1989)

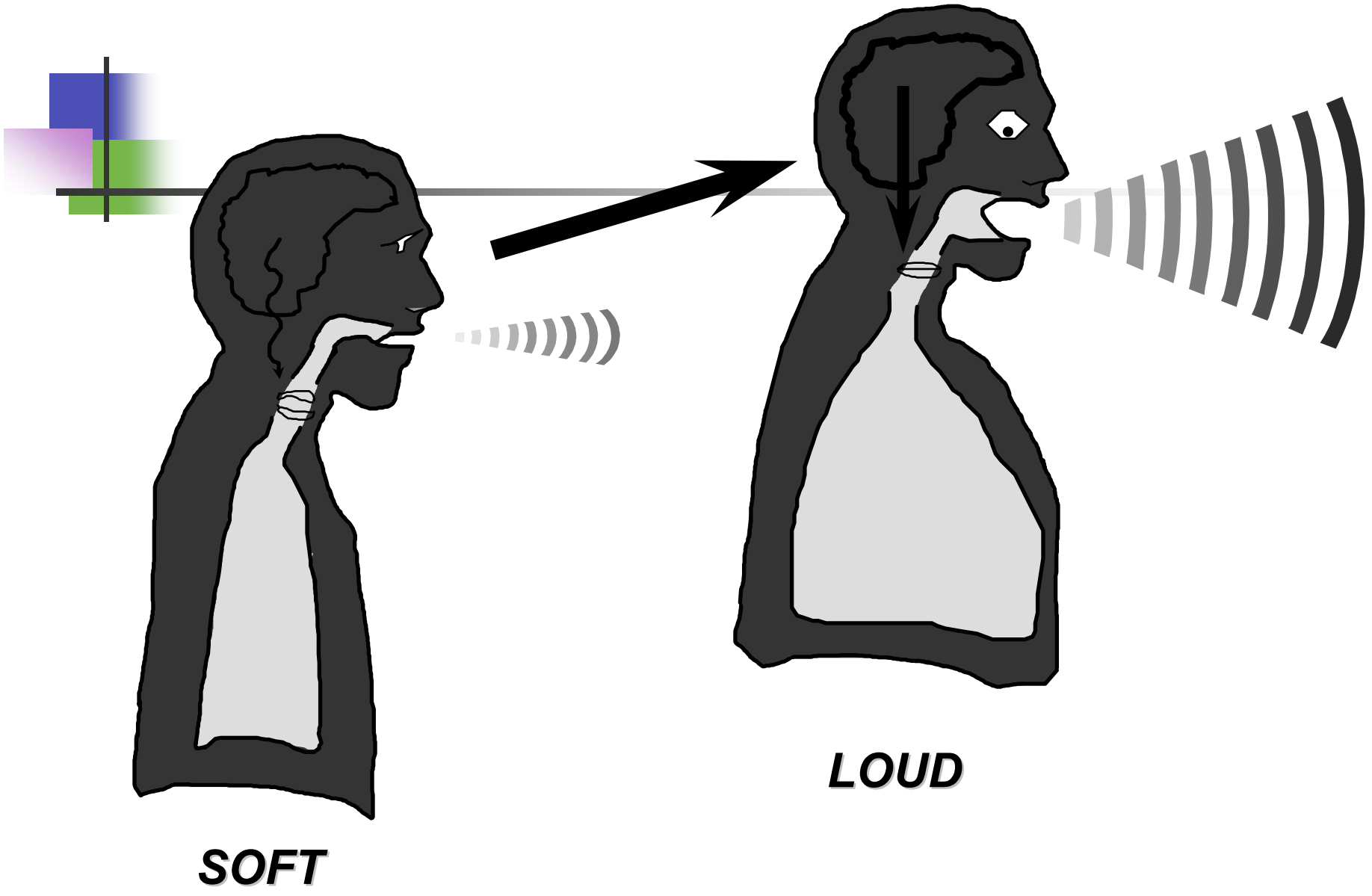
Perceptual Characteristics of Speech

Reduced loudness
Hoarse voice quality
Monotone
Imprecise articulation
Vocal tremor

Some patients report volume, hoarse voice or monotone as the first PD symptom
(Aronson, 1990)

“If you don’t talk loud enough, people stop listening”

-Individual with Parkinson Disease
Boston, May 1996


LOUD
SOFT

To a patient……major life impact

“My voice is alive again”

“I can talk to my grandchildren!”

“I feel like my old self”

“I am confident I can communicate!”

Lori Loudmouth Lori Ramig

Animated Voice Therapist

Computerized Training of Conversational Scripts to Facilitate Integration into the Community & Workforce

Leora R. Cherney, PhD
Anita S. Halper, MA

Funded by:
National Institute on Disability and Rehabilitation Research, Grant # H133B031127

Specific Aims

- Develop conversational scripts that are personally relevant to the individual with aphasia
- Implement the computerized intervention to facilitate script production through cued mass practice
- Evaluate the efficacy of the intervention when provided by a speech-language pathologist
- Evaluate the effectiveness of the intervention when used by vocational rehabilitation specialists

What is a Script?

- A sequence of sentences that a person typically speaks in routine communication situations
- Examples
  - Ordering pizza over the phone
  - Making a doctor's appointment
  - Job interview

What is Cued Mass Practice?

- Provides maximum support to facilitate accurate production; support is gradually decreased
- Intensive repetitive practice
- Accomplishes automatization of script production
- Improves communication during participation in specific everyday activities

C-CoSTA: Computerized Conversational Script Training for Aphasia

WHOLE CONVERSATION

Pizza Man: Hello can I help you
You: I want to order a large pizza.

Pizza Man: What do you want on it?
You: Pepperoni and extra cheese.

Pizza Man: What's your address?
You: 4945 East Commissary Court.

Pizza Man: Cross streets?
You: Swan and Fort Dowell

Pizza Man: What's your phone number?
You: 243-2926

You: How much will it cost?
Pizza Man: Fourteen dollars and thirty two cents.

You: How long will it take?
Pizza Man: It'll be there in half an hour.

You: Thank you

Scaling up new technologies

for more powerful, immersive learning experiences

	Auditory	Visual
Recognition / Understanding	**Speech recognition**	**Interpret visible speech, emotions, gestures**
Generation / Synthesis	**Language generation & synthesis**	**Animated characters: visible speech, emotions & gestures**

Animation Challenges

- Predict and generate head movements and facial expressions from text and/or speech
- Have the agent provide appropriate head movements and facial expressions when the student or patient is talking
- Implement “shared attention” in learning tasks
  - Tutor looks where student is looking
  - Follows the mouse, etc.
- Natural “Idle” animation & daydreaming

Motion capture and synthesis

CSLR characters

Ms. ReadWrite

Marni

Julie

Singo

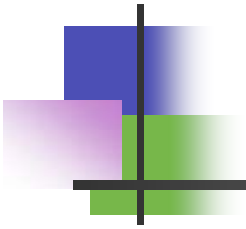
Speech Recognition

- Reading out loud with feedback
- Spoken responses to focused comprehension questions
- Transcription of spoken summaries for automatic grading
- 8 languages
  - Arabic, English, French, German, Italian, Japanese, Portuguese, Spanish

SONIC is the world's most accurate kids' speech recognizer

Computer vision

- Orient to the user (using face tracking algorithm, developed at UCSD)
- Social resonance (smiles when student smiles, using emotion recognition system developed at UCSD)
- Eye tracking (UCSD)



Discussion