

Ideas for a System Supporting Users' Learning of a Foreign Language

Carol Strohecker

Originally appeared as Note 92-19, Mitsubishi Electric Research Laboratories

Abstract

If you want to learn a foreign language, go to the country in which the language is spoken and live there for a while. That way you'll hear the language constantly, will be inclined to use it, and will do so within a cultural context in which the language has been developing for years and years.

While many might agree with such common-sense wisdom, few can afford to follow it. But a computerized system could provide an alternative by simulating environments in which users do fun things while hearing and speaking a language other than their own.

What follows is a loosely connected set of ideas for such a system. In particular, it is a system for Japanese children who know some English and want to speak it better. The fact that these children need something to talk *about* as they learn to speak better English suggests that the selection of content areas, as well as identification of features, is important in thinking about capabilities of the system.

The development of these ideas is still rough and their presentation is fanciful; you might like to get a cup of coffee, kick off your shoes, and imagine yourself in Atlanta four years from now – yes, you are at the Olympics.

The summer games of the 1996 Olympics will be held in Atlanta, Georgia.

Such a center would be an ideal spot for introducing a set of online environments that supports children's learning of interrelated topics – in this example, sports, language, and culture.

I/O devices are yet to be determined. The metaphor here is that an object activates when "touched."

Using a built-in camera and signal processors, the system isolates key facial features and renders a rough 3D image of the user.

Clocks display the varying times of international locations.

The "new pictures" are associated with a change to the environment in which the boy had been playing.

The system retains the last "chunk" of activity from each user's session.

The images of people and objects available for immediate activity are colorful and mobile; those associated with activity from a previous session appear gray and still.

Mispronunciations, mishearings, and misinterpretations can be part of the fun of speaking a new language and can enhance memorability.

Delia is nine years old and loves gymnastics. Her parents, excited both about their daughter's budding skills and the presence of the international games in their home town, have bought tickets so the three of them can attend the Olympics.

After a couple of days, Delia has seen warm-ups for the equestrian, taekwondo, archery, swimming, diving, baseball, and track and field events. She has also seen her first gymnastics competition, but was more disappointed than thrilled. The athletes seemed very far away, and even though the television monitors helped with close-ups, she could have stayed home to see them on TV. Add to that the crowds and noise, and it is clear that Delia is not having a good time.

Her parents find a children's center where she can join other young people in doing gymnastics, reading, making colorful block constructions, and playing with computers. She stays at the center while her parents go off to see more of the competitions. This is more like it! Delia goes from the rings to the parallel bars to the uneven bars, flipping and swinging, arching her back and poising her chin like the athletes she has seen on TV.

That night she feels better, and as she relives the pleasant experiences of the day, she remembers a boy looking through a kind of window in a corner of the room. He was gesticulating oddly, and speaking, though there wasn't anyone near him. She wants to go back to the center the next day so that she can see what he was doing.

She does. No one is at the window, so she approaches it and peers inside. She sees a specter-like image and suddenly realizes that it is a rough image of her face! It appears in the midst of a wide, dark area bordered by tiny luminous pictures – clocks, familiar symbols of some of the Olympic sports, and faces of different people. Some of them are bright, colorful, and lively, like hers, but some are dim and still. One of these is the boy from yesterday! She touches his picture and hears his voice as new pictures appear around him: "Oh, I thought you said *frying* change!" What was he talking about?

Delia touches his picture again. It moves as he says again: "Oh, I thought you said *frying* change!" What did he mean? She notices some of the other pictures. The Olympic symbol for equestrian events is there, and a TV, and things associated with the events – ramps, log and water jumps, and barriers that remind her of the uneven bars for gymnastics. There are horses, too, and other people – she touches another faded-looking one, a girl. Again, Delia hears a voice, but can't quite make out the words. The girl seems to say, "He's doing a *frying* change." Delia touches the other dim picture, a horse. Suddenly, to her delight, it prances across the screen, as does some text including the words *flying change*.

Certain characters "live" in each environment. Their conversation is modulated both by programmed characteristics and by a computational model of the user. This model updates based on the user's interactions and influences output by activating certain objects and guiding their utterances and movements.

An "environment" is something like a room in which objects can be arranged and manipulated in different ways.

A small "television set" is one of the objects in each environment. With it, users can view recordings of the games, edit their own presentations of events, and control speed and zoom for closer examination of movements.

The best idea may be for the system's characters to be speaking objects or animals rather than people, to ease the requirements for them to have "common sense."

Similarly, acknowledging the potential value of occasional mispronunciations, mishearings, etc., can ease requirements for "accuracy" on the part of both speech recognizer and student:

If the system makes a misinterpretation, the associated object or character might just seem momentarily silly. And if the user says something incorrect or unintelligible, the system can respond to an approximation of what was said – often, no doubt, with humorous results.

Excited, she touches one of the brightly colored horses. It seems to come to life, and says in an odd voice, "I wish I could get rid of this bridle." Delia laughs, and says, "What?" The horse says, "I said, I wish I could get rid of this bridle."

Suddenly another girl nudges Delia aside and peers into the window, shouting, "I love horses!" She quickly touches every horse on the screen, and the girls witness a dazzling show of walks, trots, and canters.

"Let's see if anyone's on!" the girl says. She reaches over and touches the "diving" picture, explaining, "Hiro likes this." "There he is! But he's not on. It's too early."

Delia barely has time to wonder what she means. The girl has touched Hiro's dimly lit picture and he is saying, "I'll show you my dive tomorrow." The girls play with the television and, to their delight, discover some stroboscopic images of high dives.

Having learned from the girl that Japanese children are sometimes on the system around suppertime, Delia arranges with her parents to stay at the center longer the next day. At 6:00, she goes to the window and touches the symbol for "swimming" events. Yes! Someone is there – a girl's brightly lit face moves as she talks with the pool caretaker, whose speech seems oddly similar to yesterday's horse. But then, the girl seems to speak a bit oddly too – Delia realizes that she is Japanese!

"Hi!" says Delia.

"Hello," says the girl. "My name is Yoshiko Narakawa." "My name is Delia Green."

"We did not meet before."

"This is my third time here."

"I did not see you in swimming."

"No, I haven't been in swimming before. I was in gymnastics and equestrian."

"E-kes-te-an?"

"Equestrian. Horses."

"I go to school in one hour."

Delia notices that the Japan clock says 7 A.M. "I don't have school in the summer," she replies.

They continue to talk, and Delia learns that Yoshiko loves to swim and knows a lot about the strokes used in Olympic competitions. Yoshiko touches a picture of a swimmer doing the crawl, and the environment seems to become watery. The girls feel that they are underwater with a swimmer, whose arms create turbulent pathways and churn patterns of bubbles through viscous resistance.

An object's speech output should sometimes be easier and sometimes more difficult than the user's ascertained level of skill – and sometimes, of course, the two may "match." The idea is to provide a challenge without becoming frustrating.

Having some music on the system might be a good idea. Tunes can help in remembering words. In fact, composing one's own song in English could be an excellent language-learning activity.

This computer system may as well have a name. I suggest *mediator*: It is a system that extends "multimedia" as we now think of it. It is also a system that mediates between people and objects, and between people and each other.

A high-bandwidth network capable of supporting digital video is slated for installation throughout Japan by 1996.

Currently, Atlanta has no sister city in Japan.

The southern accent of certain Atlantan users may provide an interesting challenge for both the system and its Japanese users.

Speech recognizers need to distinguish the voices of multiple users.

They play until Yoshiko has to leave for school, and then Delia plays alone in the gymnastics environment. By working with pictures of an athlete wielding a hoop and a ribbon, she constructs a floor routine that she can show to a friend later. And by changing the length of a gymnast's legs, she simulates an unusual set of jumps.

Yoshiko is in the fourth grade and has been studying English for three years. By now she knows some basic expressions, the days of the week, and can count to 100. But she has trouble putting the words she knows into sentences, and often finds it difficult to understand her Japanese teacher when he is pronouncing unfamiliar English words.

Like many of her friends, Yoshiko collects things that are from the US: she has T-shirts printed with English words, CD's of Madonna and the Beatles, comic books, and baseball cards. She knows the words to "Yellow Submarine" and "Michelle" (in both English and French). A pitcher for her softball team, she loves wearing her Chicago Cubs baseball cap.

Yoshiko is friends with Hiro Tomishima, who is about the same age and whose mother gives after-school English lessons to young people in the neighborhood. Tomishima-san is leasing a new computer system, which she likes because it keeps the students' attention; allows one, two, or three students to use it at a time; and seems to enhance their English speaking. It comes as no surprise to Hiro's mom that her students like this system – it is highly graphical and interactive, and it offers interesting things for them to talk about.

Tomishima-san is participating in the trial of a particular capability of this system: during the summer Olympics, young people in Atlanta can access a similar system and play with students in her town, which has just been named a sister city of Atlanta. This is a chance for Japanese youngsters to converse with real English speakers! The only problem is the time difference – Tomishima-san is starting her day very early in order to let the students use the system for a couple of hours before school.

Hiro, of course, spends a lot of time on the system. By using it, he has met an American boy who shares his interest in the diving competitions. They each pretend to be Kent Ferguson and experiment with changes to his forward somersault by altering the springiness of the diving board (which occasionally offers suggestions about such manipulations).

One morning Yoshiko joins Hiro, and with his American friend they compare times of somersaults executed with varying degrees of tucking the body. They help each other to construct new dives by combining maneuvers.

"I know," says the American boy. "Let's look at Ferguson's vertebrae in the 2 1/2 twist."

Hiro says, "What is ve-te-bae?"

"Backbone."

"Oh, like spine," says Yoshiko.

"Yeah, the spinal chord goes with the vertebrae."

They study the articulation of these bones in various views of the twist movement, until the American boy realizes that he has to leave.

"Do you mind if we stop now?"

"Yes. No. I don't know what do you mean?" Hiro asks.

"I have to stop now. I will come back tomorrow."

"OK. I'll show you my dive tomorrow."

Hiro and Yoshiko stop playing and gather their things so they can get to school on time. As they leave, Yoshiko says, "Tomorrow I will try swimming." Tomishima-san hears them chattering in English as they walk down the street.

Notes:

Obviously, there are lots of "holes" in these scenarios. Many of the ideas will not be possible within four years, and some may already be cliché – but the scenario may help to provide a frame for discussion of directions and possibilities. *A key premise is that learning happens best through immersion in the subject and through its interrelationship with supporting ideas.*

The "computational model of the user" may result from information gleaned by a collection of input monitors – possibly of speech, facial expressions, and/or data related to some handheld device/s (gauging frequency of use, coordinates of selections, nature of manipulations, etc.).

Say, for example, that two users are developing a new dive and one says, "I'm bending the right arm now" while performing that act. Speech recognizers might notice the form of the verb, the adjective and noun, and some aspect of content. This information might then influence what happens next in the interaction: one of the objects might make a point of using another present participle, the system may provide opportunities to use directions like "right" and "left," and so on. Similarly, the computational module that achieves the tension and animation associated with "bending" may have connections to other bendable objects or instances of bending, and effect the inclusion of such instances within the experience.

The same technologies could underlie many environments and scenarios: restaurants, taxis, hotels, groceries, factories, etc. However, the Olympics is a subject realm that offers a rich set of possibilities for animations and simulations pertaining to human physiology and, more broadly, for the study of motion. These areas are interesting for the challenges they present in technical implementation and as domains for epistemological inquiry.

At about \$10K-\$20K, systems such as the "mediator" could appear in American centers for children, communities, and language learning, as well as outside-of-school learning centers in Japan. Obviously, the "buy-in" of an existing infrastructure for distribution and use would be ideal. Given the widespread interest among Japanese in learning English, looking for such an infrastructure in Japan may make the best sense.

Acknowledgments:

Some of these ideas stem from my own musings; some, from conversations with other staff members at MERL; and some, from things that are "in the air" among the larger "computer media" community. In particular, I would like to thank Joe Bates, Pavel Curtis, Andy Golding, Hugh Lauer, Tohei Nitta, Carl Pixley, Marc Raibert, Chuck Rich, Dick Waters, and members of the Epistemology & Learning and Narrative Intelligence groups at the MIT Media Lab.

See also:

- Jenkins, Henry. 1991. "'x Logic': Repositioning Nintendo in Children's Lives." Book review of *Playing with Power* (Kinder, Univ. of California Press, Berkeley, 1991) and *Video Kids* (Provenzo, Harvard Univ. Press, Cambridge, 1991). Available through Media Studies Department, MIT.
- Morningstar, Chip, and F. Randall Farmer. 1991. "The Lessons of Lucasfilms' *Habitat*," *Cyberspace: First Steps*. ed. Michael Benedikt. Cambridge, MIT Press.
- Papert, Seymour. 1980. *Mindstorms: Children, Computers, and Powerful Ideas*. New York: Basic Books.
- Papert, Seymour. 1990. "A Unified Computer Environment for Schools: A Cultural/Constructionist Approach." Proposal to the National Science Foundation. Cambridge, MA: Epistemology & Learning Group, MIT.
- Papert, Seymour. 1991. "A Teachers' Learning Collaborative: A Model for the Appropriation of Science by Elementary Teachers." Proposal to the National Science Foundation. MIT Media Lab in cooperation with the Science and Whole Learning Teachers Collaborative.