ABSTRACT

Automatic speech recognition is used in a system to offer first-level, computer-based exercises in the Japanese language for beginning high school students. Implementing a preprototype version of the Subarashii system identified strengths and limitations of continuous speech recognition technology in supporting open dialogue practice, in which utterance choices are not written on the screen. Lessons learned from this implementation have led to novel methods in the development of materials for computer-based interactive spoken language education.

KEYWORDS

Speech Recognition, Spoken Dialogue, Japanese, Authoring System, User Testing and Evaluation

INTRODUCTION

The Subarashii system offers beginning students of Japanese the opportunity to solve simple problems through (virtual) spoken interactions with monolingual Japanese natives. Subarashii is an experimental computer-based interactive spoken language education (ISLE) system designed to understand what a student is saying in Japanese and to respond in a meaningful way in spoken Japanese. The computer system poses problems in written English and offers occasional support to the student in the form of written reminders, but problems can be solved only by speaking and un-
Subarashii: Encounters in Japanese

derstanding Japanese. The use of continuous ASR technology enables dialogues that are open in the sense that utterance choices are not displayed on the screen for students to read.

The principal research goal of the Subarashii project has been to extend the range of activities available in ISLE systems and to demonstrate the educational effectiveness of these activities. This paper reports aspects of the early stages of the project, in 1996 and 1997, when the goal was simply to find an engaging intersection between pedagogical design and technological feasibility. Meador, Ehsani, Egan, and Stokowski (1997) report a later implementation of the Subarashii system under AuthorWare and the results of a series of experiments using that system with university and secondary students.

BACKGROUND ASSUMPTIONS

Speech is the core modality of a living language. For English-native learners, the complexity and unfamiliarity of the written form of languages like Japanese make proficiency in the spoken language an especially important part of training and maintenance of language skills. A good way to learn a new language is through individual interaction provided by a skilled tutor with great patience and energy. However, most language learning is not done with individual tutoring. Instead, the teacher has to be shared among many students in a class (see Eskenazi, this issue). A potential alternative is a computer-based ISLE system designed to understand what a student is saying in Japanese (within a constrained context) and to respond in meaningful ways. An ISLE system should relieve the teacher of routine tasks such as engaging beginning students in basic spoken interactions, while reserving the teacher's time and energy for more subtle and creative work. To this end, an ISLE system can use recent advances in speech recognition technology for the purpose of Japanese language instruction.

Many multimedia computer programs are available that promote spoken foreign language acquisition, mostly for English and the more commonly studied languages (Spanish, French, and German) (see review in Wachowicz & Scott, this issue). However, at the time the Subarashii preprototype was built, programs that existed typically offered only highly constrained exercises because of the limited performance of the commercial recognition components that were incorporated in them.

One key to the future of multimedia computer-aided language learning (CALL) will be the ability to understand and judge continuous spoken language with programmable levels of acceptance. That is, an ISLE system (like a real instructor) should have an adjustable threshold of rejection that works with the wide range of errors and distortions that students actually produce when attempting to speak a new language. Furthermore,
the system should be able to simulate essential features of human-human communication. That is, interactions should work without requiring collateral cues from a mouse or keyboard, operate at an appropriate conversational pace, and incorporate verbal strategies for resolving misunderstandings (see Wachowicz & Scott, this issue).

Early development work in Subarashii explores those aspects of speech recognition and user interface technology that are likely to form the basis of advanced ISLE systems for any language. We need to know (a) which interactive formats can (or cannot) be supported by high-performance speech recognizers running on future generations of computers and (b) which of the feasible interactive formats are most enjoyable for users and most effective in producing gains in language proficiency.

What Learners Need

As demand for Japanese instruction continues to increase, so does learners' need for effective methods and materials to teach and maintain skills in spoken Japanese. Unger, Lorish, Noda, and Wada (1993) discuss reasons for emphasizing spoken Japanese, arguments that apply to learning other languages as well. They say, for example, “Speech is primary.... Research on reading shows that subvocalization occurs in the ordinary reading of all languages, including Japanese” (Unger et al., p.12). Indeed, children develop vocal communication before and independently of reading and writing. In most high school language courses speech is relatively neglected—individual students spend little time actually producing speech as compared with reading and writing. Unger et al. also offer principles for speech-focused teaching that underlie our own assumptions for ISLE:

- Learning speech requires practice. As with any skill, acquiring speech understanding and production requires repetition. Repeated exercises that use a small vocabulary and few grammatical constructs arrayed in different ways are exactly the kind of task that a computer can perform well.

- Instruction should focus on the learner. ISLE systems can control the pace of the material presented as well as the level of pronunciation accepted, based upon the responses from the student.

- Basic exercises must precede more open exercises but must be couched in concrete situations. “[Open dialogue] ... must be supported by mechanical drills—exchanges in which there is only one correct response for the student to give. This is because speaking and listening in social interactions involve the
ability to produce and recognize language promptly... Contextualized exercises help them internalize the memorized forms as well as information about the appropriate situations in which they are used" (Unger et al., p. 59).

These principles are reflected in other current philosophies of language teaching (Omaggio, 1993; Shrum & Gilsan, 1994). If well implemented, ISLE systems using multimedia technology can establish motivation for the production and reception of speech. The computer can create a scenario in which the student is trying to achieve something, for example, asking people for directions to a concert. If successful in this task, the student may be able to attend that virtual concert.

Why Japanese?

An ISLE system can be employed in teaching any foreign language. We choose to focus on beginning Japanese for these reasons:

- The rate of growth of Japanese-language programs in both high schools and colleges exceeds that of any other foreign language taught in the United States (Jorden & Lambert, 1991).
- Japanese-language instruction in the U.S. is a recent phenomenon so has not yet been standardized. New teaching methods and materials can make an impact.
- Japanese is a very difficult language for English speakers to master. According to criteria used by the Foreign Service Institute of the U.S. State Department, Japanese is rated a Category 4 language (the most difficult for native-English speakers).
- Speaking is important to the student of Japanese. Jorden and Lambert (1991) report that in a survey of high school students of Japanese, two thirds who wanted to make changes in their curriculum preferred more emphasis on speaking.
- Reading Japanese is difficult without the ability to speak. Due to multiple readings of Kanji characters based on linguistic context, reading aloud relies on familiarity with spoken Japanese.

OVERVIEW OF SUBARASHII

The Subarashii project has tried to push the art of the possible in CALL by focusing on the most challenging (and to the student, potentially most useful) types of exercises, namely the seemingly open-ended dialogues that
we call “encounters.” The Subarashii system implements a series of explorable encounters. In each encounter the learner has a mission and an interlocutor through whom the mission can be accomplished.

From the vantage of an instructional designer, the first focus is the selection of content for an encounter. Encounters were chosen from typical, everyday situations that a student might be in— for example, meeting someone for the first time; making, accepting, or refusing an invitation; buying something at a store; or visiting a restaurant. Although the primary focus in Subarashii was on high school students, we selected situations that are also relevant to adult students of Japanese. Because of the key role of both gender and age of interlocutors in a Japanese dialogue, such a selection was no simple matter. The sequence of materials in the encounters deployed in Subarashii is somewhat similar to the progression of material in Yookoso: An Invitation to Contemporary Japanese by Tohsaku (1994).

The situations built into Subarashii are not authentic in the sense that a Japanese native might encounter them. Rather, they are authentic in that they present the kinds of situations that students with limited knowledge of Japanese might encounter when visiting Japan. To render the encounters more realistic, we illustrated each with photographed scenes of Japan and realia such as Japanese restaurant menus. Though not an explicit goal of our research, the whole range of multimedia technology could be employed to embellish the encounters.

Subarashii encompasses a set of encounters (with automatic backup coaching) that run with an open microphone and offer opportunities for students to practice partially constrained, goal-oriented spoken interactions. Even in the early versions, the visual and auditory displays were of high quality and provided much of the feel of a conversation with a Japanese speaker.

The Student’s View of Subarashii

Students logging on to Subarashii are presented with a graphical user interface that allows the selection of one of several encounters. Each encounter is an interaction between the student and a Japanese interlocutor, generated by the computer, in a situation that the student is likely to encounter in Japan. The student can navigate between encounters using either the mouse-driven graphical user interface or voice commands in Japanese. Although the student is not obliged to follow a particular order, the encounters are arranged in the form of an adventure game where each encounter builds on the preceding one.

Within an encounter, the student is confronted with a Japanese setting and a “mission statement” in English that is the objective of the exercise. For example, in the encounter Are You Busy? the mission statement is:
Yamaguchi is probably going to ask you to do something tonight, but you have to study for tomorrow’s exam. Be tactful, but don’t give in to the urge to go.

The encounter then begins with the student and Yamaguchi, the virtual interlocutor, taking turns speaking (see Figure 1).

Figure 1
A Subarashii encounter: Yamaguchi greets the student visitor

Depending on the objectives of the encounter, the student may initiate turn taking. During the course of the dialogue, the student may receive reminders, hints, and further directions in written English, especially if the student is having difficulty getting through the encounter. The student is never asked to translate into or out of Japanese. All English information is presented textually and addressed to the student. In the encounter the student hears only Japanese and the system responds only to spoken Japanese. Depending on what the student chooses to say, the interaction may go in one of several alternative directions. However, certain conversation paths are not open to the student. For example, if the student’s mission is to refuse an invitation politely, the student is not allowed to accept an offer to go to the movies. If a student attempts to do so, “mission control” reminds the student of the mission and asks for a repeat of the last ex-
change. Figure 2 shows a short reminder that appears in the Glad to Meet You encounter.

Figure 2
A reminder in the Glad to Meet You encounter

The interlocutor has given his name and the student is obliged to respond. This reminder comes when the student's response has not included the name that is expected in the role-play. Upon completion of one encounter, the student is free to select a new one. With high-quality graphics, sound, and voice interaction, the student can get rather close to the feel of a real encounter with a Japanese native. However, the computer-generated interlocutor is tireless, nonjudgmental, and private, thus minimizing that student's performance anxieties.
An Example Encounter

The first implementation of Subarashii had five encounters, listed here in order of complexity and difficulty:

- Glad to Meet You
- How About a Movie
- Are You Busy?
- Got Milk?
- Let’s Go to the Restaurant

After selecting an encounter, the student sees a starting screen, as shown in Figure 3 for the Got Milk? encounter.

Figure 3
Start screen for the Got Milk? encounter

With its opening graphic display and written mission statement, this encounter puts the initiative squarely on the student. At this point, the student is presented with a passive shopkeeper who will respond to a range of queries or statements of the type one might expect at a convenience store in Japan.
A simpler encounter is Are You Busy?, in which an acquaintance, Yamaguchi, starts the dialogue. She approaches as shown in Figure 1 and wishes you a good day. There are many valid responses that the student can make and many invalid responses that the system knows about. If the student is really a novice and does not know the logic or controls of the system, the student can click on the “help” button or say “help” in Japanese. This brings up the help screen shown in Figure 4, which explains the meaning and function of various parts of the display.

Figure 4
Help screen, shown as called from the Are You Busy? encounter

For example, the “replay” button in the early versions of Subarashii replayed the whole interaction—screens, interlocutor voice, and student voice—back to the student at any time. The valid student responses are not available to the student, nor is any representation of the expected discourse structure.

The beginning states and paths through the encounter logic for Are You Busy? are diagrammed in Figure 5.
Subarashii: Encounters in Japanese

Figure 5
Initial version of the first turns of the encounter logic for Are You Busy?

Yamaguchi’s turns are shown in ovals, the student’s anticipated responses are shown in rectangles, and the machine’s advice texts from the Subarashii system are shown in shaded rectangles. Note that only the shaded text is presented in written English; all the ovals are presented in spoken Japanese, and all the clear rectangles are responses that must be produced by the student in spoken Japanese. There is no way to get through the encounter without understanding and speaking Japanese. The ellipses (...) in several of the clear rectangles in Figure 5 indicate that there are many more responses, valid and invalid, that Subarashii expects and can handle than can be shown in the diagram. A fuller representation of the logic of one encounter appears in HyperCard format as an appendix to this paper.

SUBARASHII SYSTEM DEVELOPMENT

System Architecture: A Programmer’s View

Figure 6 shows the modular structure of the program on which the earliest version of Subarashii was based.
All modules except the audio/ASR module were first implemented in the Java programming language. This choice was made for the following reasons: (a) Java is platform independent; (b) Java has a rich set of graphical objects built into the language, facilitating a sophisticated user interface; (c) Java’s support of Unicode enables the presentation of any writing system, thereby making the system extensible to languages with non-Latin scripts, such as Arabic. Ideally, the audio/ASR module would have been implemented in Java. But because Java did not yet provide adequate support for audio input and output, this module was written in C (and would have had to be explicitly ported to different platforms). The preprototype version of Subarashii ran on SGI workstations. The later versions (Meador et al., 1997) were re-implemented in AuthorWare to run under Microsoft Windows.

In the preprototype version, both the presentation environment and the audio/ASR module are independent of the specifics of the Subarashii exercises. They allow for the implementation of a high-level object called a “card.” This object encapsulates subroutine calls to display pictures, play sounds, and get verbal and graphical input from the student. Each exercise is comprised of (a) a series of cards called a “stack” and (b) resources that include sounds, pictures, and recognition grammars. Each card represents a single exchange between a virtual interlocutor and the student. During the exercise, the interaction is implemented by a particular se-
Subarashii: Encounters in Japanese

sequence of card presentations from within the stack, depending on what the student has said. The card object class, while serving the needs of an encounter-type exercise, is flexible enough to allow for any kind of spoken language exercise. The concepts of card and stack have their historical roots in HyperCard simulation. They were found to be useful abstractions in the design of encounters and were therefore retained (although Java does not enforce the card-and-stack metaphor).

Each encounter has a recognition “grammar” for each utterance anticipated from the student, specifying in a compact way how the input should be processed. This grammar helps the ASR module recognize student utterances by limiting its expectations of what will be said. The recognition grammar is specified in romanized Japanese (modified Hepburn notation) and must be parsed and translated into a phoneme-level specification (a recognition network) as required by the ASR module. This is achieved automatically by use of a recognition grammar compiler that was created for the Subarashii project. The recognition module was designed to be speaker independent and can accommodate a variety of nonnative accents.

Prototyping New Exercises: An Author’s Tools

Subarashii is not only a teaching system but also a lesson-authoring system in which new encounters may be prototyped. The author of an encounter must develop the recognition grammar that defines the input expected from the student. An important innovation of Subarashii is to help authors predict this input, before speech is enabled, as follows.

Many dialogue systems are brittle in that they cannot maintain the illusion of a real dialogue if the user deviates from a narrow list of expected utterances. Although a beginning language student has limited proficiency, the student can still produce a variety of potentially valid utterances in any situation, even if some of these are prescriptively incorrect. To plan for every eventuality is almost impossible. Instead, Subarashii lets authors prototype encounters in a traditional HyperCard environment and test them with real students. These prototype encounters use graphics but no speech: The dialogue is strictly text in and text out. Students in a pilot test then try out the prototype encounters and type in their responses. Granted that text responses (right or wrong) will not be identical to spontaneous spoken responses, we assume that they will be similar in many ways. The prototyping environment automatically records all student inputs received for an encounter and organizes them by frequency. Thus, the author can decide at a glance what a student is likely to say at any point and provide for an appropriate course of events or for instructive feedback for common mistakes.

HyperCard provides an efficient means of modifying each encounter as
the result of input from the pilot students. Once the logic of an encounter has been verified to the author’s satisfaction (perhaps after several prototyping iterations), all requisite sounds are recorded, recognition networks generated, and HyperCard scripts translated into Java. The prototyping environment limits the features of HyperTalk (the HyperCard programming language) that authors can use to a simple set of commands that are easily translated into Java. We used HyperCard because of its simplicity and its availability on Macintosh systems at our first high school test site.

Kinds of Dialogues: System Versus Student Initiated

With the support of the prototyping environment, we were able to generate new encounters and other exercises fairly quickly. Consonant with our goal of pushing the limits of current technology, we experimented with different kinds of encounters aimed at beginning students at different levels of Japanese proficiency. We also varied the amount of prompting in English within a given encounter. Like other authors of CALL, we have found it easier to create system-initiated encounters, which require the student to respond to questions from the virtual interlocutor, because that clearly limits the range of what the student can legitimately say. However, such dialogues tend to fix the student in a passive role. To restrict Subarashii in this way would give the student only half of what it means to communicate.

Therefore, we provided ample opportunity in most encounters for students to take the initiative with questions and comments of their own that might occur logically in the course of the dialogue. We also created encounters that require the student to take the initiative. In Got Milk?, for example, the student must buy something at a grocery store where the character on the screen will only respond to questions or directions from the student, just like polite shopkeepers in Japan. An example of a brief encounter implemented in the preprototype Subarashii is shown in the appendix (in standard HyperCard format).

Speech Recognition Issues

To recognize what the student says, the ASR module uses the encounter-specific recognition grammar together with a set of hidden Markov models (HMMs) of the sounds in the Japanese language. These HMMs were produced for Subarashii by training the recognizer on a corpus of spoken Japanese. Speech produced by nonnative speakers of a language is fundamentally different in certain important respects from that of native
speakers. This is because the speakers typically carry over artifacts (phonological to pragmatic) from their native language when speaking a new language (see in this issue Dalby & Kewley-Port; Eskenazi). This difference between native and nonnative speech is a primary obstacle in the employment of speech recognition technology for CALL and a primary focus of our scientific inquiry.

The recent commercial CALL products based on speaker-dependent recognition (see Wachowicz & Scott, this issue) have typically been trained on a small sample of native speakers. Such products have no internal representation of the difference between, for example, a faulty microphone and a faulty pronunciation. In Subarashii we used HTK (Entropic’s HMM Tool Kit), an ASR development environment that supports the construction of speaker-independent, continuous speech recognition systems. With customizable ASR technology such as HTK, we were able to build a system that recognizes both native-Japanese and English-Japanese and so does better at avoiding both erroneous acceptance of unwanted forms and erroneous rejection of insignificant differences.

Our approach has been not to reject utterances merely on the basis of deviation from a single “gold-standard” model of the correct response. Subarashii compares each utterance both with a set of models of correct responses and a set of models of likely incorrect responses. The model (correct or incorrect) that most closely matches the utterance to be recognized is what the computer understands the speaker to have said. In the event that none of the models compares well with the utterance given, the computer rejects the utterance. This strategy can incorporate prior comparative analysis of differences between English and Japanese to predict likely student errors, which are not random but follow structures well known to a skilled language teacher. However, in designing the recognition networks in the earliest version of Subarashii, we relied on empirical identification of actual errors collected from students’ typed input in the prototyping phase.

PRELIMINARY EVALUATION

After the preprototype system was built, the next challenge was to conduct experiments to reveal how students use the system and to show any positive effect of this use. As a first step, we put Subarashii in Silver Creek High School in San Jose, California, and informally observed its use by 10th and 11th grade students. We also talked to the students after their encounters with the system. We found that the ASR-based system worked in a school setting, that most students got through the encounters to solve the problem, and that almost all said they enjoyed the experience.

We also found that we could evaluate basic design questions through the data collection utilities in the Subarashii software. Utilities in the
preprototype system included provision of logs that yield an exhaustive record of the machine’s use—what the machine did and what someone said to it, with time stamps on every event. We can derive an arbitrarily fine analysis of interaction patterns from these logs. The logging function also allowed students, teachers, or researchers to play back whole student sessions, seeing and hearing everything the machine said and hearing everything the student said. With these facilities we were able to evaluate the following features: the flow control strategies in the encounters, the settings of recognizer parameters (e.g., time-out lengths, rejection thresholds, alternative sets), and the design features of the recognition itself (e.g., the use of native-Japanese models or English-Japanese models, or both). The later versions of Subarashii had automatic sign-in and sign-out procedures that let users react and analyze their experience with Japanese on the machine and then compare it to other methods, such as book study and class time.

As further steps in the evaluation, we are conducting more controlled and monitored field placement studies of the later implementations of Subarashii at Silver Creek High School and at Stanford University. Meador et al. (1998) and Ehsani, Bernstein, and Najmi (forthcoming) present emerging data from these studies. Questions of interest include:

- What kinds of ISLE dialogues and games do students enjoy most?
- Which paths through which encounters do students actually use?
- Which ISLE dialogues and games (if any) produce measurable increments in skill?

We also plan to consider user variables as they affect usage and attitudes, including demographics (e.g., age, gender, education level, grade-point average) and language-learning variables (e.g., number of languages spoken, achievement level in Japanese, Japanese skill profile). Note that precise techniques for estimating the rate of learning require very significant blocks of student and teacher time and are not planned for a high school test. Another important practical question that cannot be addressed in a school placement is how a set of tools can be packaged to enable Japanese teachers to produce content. These questions must be addressed in other settings. (Some difficulties of testing teaching methods in public school settings are discussed in Logan and Dobb [1985].)

**LESSONS LEARNED, IMPROVEMENTS, AND EXTENSIONS**

The preprototype Subarashii system broke new ground with its seemingly open encounters. Our early observations in a high school provided
preliminary evidence that meaningful conversational practice can be authored and implemented, that the system with its ASR works in a school setting, and that high school students seem to enjoy it. The preprototype system has been extended in two ways. First, it was re-implemented under AuthorWare for operation in the Microsoft Windows environment so that larger scale and more reliable evaluations could be performed on an affordable hardware platform. Second, instructional exercises were built to support student performance in the open encounters. The second extension was motivated by lessons learned in the preliminary evaluation. We found that, although Subarashii allowed students to practice social interaction at a polite pace, sometimes the encounters were too difficult and students did not know what to say. They needed opportunity to find or practice an appropriate response to a machine query before engaging in the encounters. We also found that students could profitably access a dictionary, video examples of suitable exchanges, or other resources provided with the system. We reasoned that these could help students prepare to participate in dialogue and so included them in later versions of Subarashii. Moreover, simple drills of the patterns that comprise the Subarashii encounters might provide gradual instructional support and bring students up to dialogue readiness. We conceived drills in three basic modes of oral interaction, ordered as follows:

- Pronunciation training
- Closed response queries
- Constrained grammar exercises

Pronunciation training focuses on errors or distortions made by the student in articulating words and phrases in the target language. The student may be asked to read words, phrases, or entire sentences displayed on the screen. This technique may be most beneficial for languages that do not use the Roman alphabet, especially for nonphonetic scripts such as Japanese (Kanji characters). Experiments have demonstrated that systems based on speech recognition can reliably score several aspects of spoken language production, including segmental pronunciation, rhythm, and intonation (Bernstein et al., 1990; Dalby & Kewley-Port, this issue; Neumeyer et al., 1996; Shirotsuka et al., 1992). Aspects of the underlying techniques used in some of these experiments are presented in Bernstein and Franco (1996).

In closed response queries, the student selects one response from a number of options presented by the computer. The student may be prompted for speech by written, graphic, auditory, or verbal cues. For Japanese we might also include selection of the appropriate degree of politeness based on the situation and the interlocutor. This question and answer format

In constrained exercises the student is not directly presented with a list of response choices, even though responses are still limited to a few correct ones. The goal of a constrained exercise is to train the student for encounters. Constrained exercises may address sentence negation, the numbering system, telling time, identifying colors, and so forth. In the course of constrained exercises, the student is likely to make a number of errors, grammatical and otherwise, that are commonly made by English learners of Japanese. The computer can track these errors and, wherever possible, provide the student with an explanation of errors they make. Simple drills of the patterns that comprise the encounters may also be helpful to beginning students.

The student can be presented a series of exercises in these modes, grouped into lessons. Each mode would require that the student achieve some level of proficiency in the preceding modes. During a lesson, students would be given immediate feedback on their performance whenever possible. Upon completion of an exercise, students would be given a performance evaluation if appropriate. If appropriate resources were available, more advanced interactions might be offered that rely on specific area vocabularies.

The latest version of Subarashii implements both closed response queries and constrained exercises but not pronunciation training. Ehsani and Egan (1998) present aspects of these system features as well as an interactive practice function that uses real-time video.
APPENDIX

The Are You Busy? encounter stack after iterations (in text form) with students.

theme: polite refusal

** Initial Information

Your name is Smith. You are an exchange student in Japan. Tomorrow you have your first big test and you really need to study. You are on your way into the library when you see an acquaintance, Kobayashi, coming your way.

** card S1
* information
She walks towards you.
* Japanese
Konnichi wa! (Hello!)
* grammar
basic = [Kobayashi san ] konnichi wa, konnichi wa Kobayashi san;
genki = ([Kobayashi san ] konnichi wa | konnichi wa Kobayashi san )
genki [desu ka];
longtime = (hisashiburi |shibaraku ) desu nee;
wrongtime = [Kobayashi san ] (ohayoo [gozaimasu] | konban wa);
haji = [Kobayashi san ] hajimemashite [Kobayashi san];

on userInput label
   if label = “basic” then
      go to card “S2”
   else if label = “genki” then
      outputToUser “Okagesama de, genki desu. Tokoro de,”
      enableContinueButton “S2”
   else if label = “longtime” then
      outputToUser “Soo desu nee. Tokoro de,”
      enableContinueButton “S2”
   else if label = “wrongtime” then
      outputComment “Think about the time of day. She said ‘konnichi wa’.”
      enableRepeatButton
   else if label = “haji” then
      outputComment “This isn’t the first time you’ve met. Try again.”
      enableRepeatButton
   else
      outputToUser “Chotto wakarimasen”
Jared Bernstein, Amir Najmi, and Farzad Ehsani

enableRepeatButton
end if
end userInput

** card S2
* information
Don't forget about the test you have tomorrow!
* Japanese
Sumisu san wa, kon'ya, isogashii desu ka? (Smith-san, are you busy tonight?)
* grammar
basic = [(zannen desu ga |hai |ee )] chotto [isogashii [n ] desu],(hai |ee ) isogashii [desu], (hai |ee ) soo desu;
short = (hai|ee);
anata = (hai |ee ) (soo |isogashii ) desu anata wa [isogashii desu ka];
anataben = [(hai |ee )] (shiken ga arimasu |benkyoo desu ) anata wa [isogashii desu ka];
iie = [iie ] [amari ] isogashiku (arimasen)nai desu), [iie ] hima desu, iie;
study = [(hai |ee ]chotto [[chotto ] isogashii [desu ]]) [ashita ] shiken ga arimasu [kara];
on userInput label
  if label = “basic” then
    go to card “S2B”
  else if label = “short” then
    outputComment “That could simply mean you are listening. Make it a little longer.”
    enableRepeatButton
  else if label = “anata” then
    outputToUser “Amari isogashiku nai desu yo. Sumisu san wa,”
    enableContinueButton “S3”
  else if label = “anataben” then
    outputToUser “Amari isogashiku nai desu yo. Sore ja,”
    enableContinueButton “S4”
  else if label = “iie” then
    outputComment “Sure you are! That test is important! Try again.”
    enableRepeatButton
  else if label = “study” then
    outputToUser “Soo desu ka?”
    enableContinueButton “S4”
  else
    outputToUser “Chotto wakarimansen”
    enableRepeatButton
end if
end userInput

Volume 16 Number 3 379
She looks at the books you are carrying and asks,

**Japanese**
Hontoo desu ka? (Really?)

**grammar**

basic = [(zannen desu ga | hai | ee )] chotto [isogashii [n ] desu], (hai | ee )
isogashii [desu], (hai | ee ) soo desu, hai;
anata = (hai | ee ) (soo | isogashii ) desu anata wa [isogashii desu ka];
anataben = [(hai | ee )] (shiken ga arimasu | benkyoo desu ) anata wa
[isogashii desu ka];
iie = [iie ] [amari ] isogashiku (arimasen | hai desu), [iie ] hima desu, iie;
study = [(hai | ee | chotto | [chotto] isogashii [desu ])] [ashita] shiken ga
arimasu [kara], (hai | ee ) benkyoo (desu | shimasu);

She cocks her head to one side with a questioning look.

**Japanese**
Benkyoo desu ka? (Do you have to study?)

**grammar**

basic = (hai | ee | [(hai | ee )] soo desu ) [[ashita] shiken ga arimasu
[kara]], [(hai | ee )] benkyoo [[desu | desu ne]]
anata = (hai | ee ) (soo |isogashii ) desu anata wa [isogashii desu ka];
on userInput label
  if label = “basic” then
    outputToUser “Soo desu ka.”
    enableContinueButton “S4”
  else if label = “anata” then
    outputToUser “A mari isogashiku nai desu yo.”
    enableContinueButton “S4”
  else
    outputToUser “I did not understand.”
    enableRepeatButton
  end if
end userInput

** card S4
* information
She looks a bit disappointed, but smiles.
* Japanese
Ja, ganbatte kudasai! (Well, do your best!)

* grammar
basic = (hai | ee ) [ganbarimasu], doomo [arigato [gozaimasu]];
mata = [hai ganbarimasu ] (sayonara) ja mata [ashita]);
on userInput label
  if label = “basic” then
    go to card “S5”
  else if label = “mata” then
    outputToUser “Mata, ne!”
    enableContinueButton “done”
  else
    outputToUser “Chotto wakarimasen”
    enableRepeatButton
  end if
end userInput

** card S5
* information
Half-way turned around, she gives you a friendly wave.

* Japanese
Ja, mata ashita! (See you tomorrow!)
Subarashii: Encounters in Japanese

* grammar
basic = [ja ] mata [ashita], [ja ] sayonara;

on userInput label
    if label = “basic” then
        go to card “done”
    else
        outputToUser “Chotto wakarimasen”
        enableRepeatButton
    end if
end userInput

** final card

Make sure to give Kobayashi a call tomorrow. Now, go get ready for that test!
Yoku dekimashita nee! (You did well!)

REFERENCES


Jared Bernstein, Amir Najmi, and Farzad Ehsani


ACKNOWLEDGMENTS

The Subarashii project was conducted by the Language Systems Group of Entropic Research Laboratory, Inc., in Menlo Park, California. The work was funded by the U.S. Department of Education (Grant P017A50019) and by Entropic Research Laboratory, Inc. We thank Rushton Hurley and the students at Silver Creek High School, San Jose, CA, for their help in developing the encounters. We thank the Federal Language Training Laboratory of the U.S. Government for support in graphical design and content, and we thank the president, faculty, staff, and students of Chubu University, Aichi, Japan, for their gracious help and spirit of cooperation.

AUTHORS’ BIODATA

Jared Bernstein is president of Ordinate Corporation and consulting associate professor of Linguistics at Stanford University. He received his Ph.D. in psycholinguistics from the University of Michigan and has been active in speech recognition technology for over 20 years, focusing on applications in language education and in systems for disabled people. He previously conducted research at SRI International and at Entropic, Inc.
Amir Najmi is a Ph.D. student of Electrical Engineering at Stanford University. His research explores the application of information theory to statistical pattern recognition and has addressed the application of speech recognition to language learning systems. He has worked in the field of machine learning for over 10 years.

Farzad Ehsani is one of the founders of Sehda, Inc., a small start-up focusing on human-machine dialogue technology. Previously head of the Language Systems Lab at Entropic, Inc., he conducted research in the areas of speech recognition and dialogue systems for language education. He has also conducted research and product development at Motorola, NEC, and MIT and holds three patents in the areas of speech and handwriting recognition. He has an M.S. and Engineer's Degree in Electrical Engineering from MIT.

AUTHORS' ADDRESSES

Jared Bernstein
ORDINATE Corporation
1040 Noel Drive, Suite 102
Menlo Park, California 94025
Phone: 650/327 4449
Fax: 650/328 8866
E-Mail: jared@ordinate.com
WWW: www.ordinate.com

Amir Najmi
P.O. Box 9069,
Stanford, CA 94309
E-Mail: zaalim@leland.stanford.edu

Farzad Ehsani
Sehda Inc.
1040 Noel Drive, Suite 100,
Menlo Park, CA 94025
E-Mail: farzad@sehda.com