Speech Technology and Language Learning: Some Examples from VILTS The Voice Interactive Language Training System

Patti Price and Marikka Rypa

Speech Technology and Research Laboratory

SRI International Menlo Park, California

1 Introduction

In this paper we describe the development of the Voice Interactive Language Training System (VILTS) and our experience in exploring the potential of speech technology in service to language learning. We identify ways in which speech technology can support language learning, and we explore possibilities for the future. In particular, we describe the roles speech technology can play in support of language learning (Section 1), and discuss types of activities that can support new learners (Section 2) or sustain and enhance those who have already acquired some language skills (Section 3). We then summarize our main points (Section 4). In the remainder of this introductory section, we outline the role speech technology can play in language pedagogy (1.1), define differences between initial language learning and later sustainment and enhancement (1.2), describe the roles of implicit and explicit support (1.3), sketch the potential for leveraging existing resources (1.4), and provide some background on VILTS (1.5).

1.1 The Role of Speech Technology in Language Pedagogy

Spoken interaction lies at the heart of language learning. When we say that someone knows Swedish, or Swahili, or Mandarin, we generally mean that the individual can speak and understand the language fluently. Although there may be some uses of language skills that do NOT require speaking, most people learn language because of a desire or need to interact with people who speak that language. Further, there is evidence that speaking skills may be necessary to support good listening skills. For example, someone without speaking experience may have great difficulty inferring the articulatory shortcuts that speakers use when speaking casually. Dr. Ray Clifford, Provost at the Defense

Language Institute (DLI) in Monterey, California, describes previous experience with some government language students being trained only for reading and writing skills. In government tests, those students who had additional instruction in speaking skills scored higher in listening comprehension and reading skills than did those who had only listening and reading training. Moreover, transitions toward fluent and appropriate language interactions, such as the transition from passive knowledge to active knowledge, from halting to fluent speech production, and from a controlled instructional setting to a real-life situation, benefit from well-designed integration of speech technology into computer-assisted instruction.

The transition from passive comprehension to active language production is often one of the most challenging parts of language learning. A widely accepted theory of language learning addressing this transition is Krashen's Monitor Model (Krashen, 1981), which stresses the primacy of appropriate, meaningful input as a prelude to production. Krashen's Input Hypothesis claims that acquisition is activated by understanding the target language in a communicative context. This model stresses the importance of input at a level higher than what the student is able to produce. Speech technology can support this pedagogical transition through listening activities with challenging and incremental material from a variety of speakers. Such a prelude to speech activities can support the progression to meaningful linguistic output. A distinct but related speech technology, pronunciation scoring and feedback, can also assist in forming good articulatory habits.

Another transition in language learning is the transition from halting or calculated speech to fluent production. The "output hypothesis" (Swain, 1995) has been proposed as an extension of the input hypothesis as a crucial stage in the acquisition of fluent production. DeBot (1996), for example, argues that output plays a direct role in enhancing communicative competence by turning "declarative knowledge," a set of facts about the skill to be acquired, into "procedural knowledge." Anderson (1982) describes this transition as embodied in applying language skill appropriately and with increasing speed. Speech technology can support this transition by enabling students to practice proper articulation of utterances in isolation and to make progress toward production in a meaningful context.

The third major transition for language learners is the ability to generalize the skills learned and to apply them in the real world. Speech technology can assist this transition by providing examples from several speakers, speaking styles, and dialects of the language. Speech technology can also provide the opportunity to practice learned skills interactively, using robust speech recognition in activities mapping to realistic situations. Such activities can help bridge the gap between a traditional instructional setting and the target environment of language use with real native speakers in the real world.

The core speech technologies we can exploit for various purposes include

- a database of recorded lexical items including basic vocabulary spoken carefully by one or more native speakers and available as part of the pedagogical approach or as a student option

- speech recognition tuned to the recognition of non-native speakers

- speech rejection, so that all or part of an utterance can be rejected as not understood (rather than risk an inappropriate response)

- pronunciation scoring, so that the system can return a score for a sentence, word, or sound that correlates well with human experts; feedback on the score may include visualization components comparing the student production with a native speaker along various dimensions, and may also provide repair mechanisms.

The basic skills involved in learning language are listening and speaking. For languages that have a written system (as do most languages being learned), reading and writing are also basic skills. The speech technologies outlined above can support all of these skills:

Listening Speech technology for listening skills has been used widely since recorded materials have been available. Computerized recordings of native speakers and the reduction in the cost of memory have enabled the storing of and random access to a much broader variety of speakers and styles of speech than has been possible in the past. We believe that such material is currently under-used in language learning software, but could have a large impact in helping a student to generalize listening skills toward the goal of interacting with live native speakers in a variety of contexts.

Speaking To be understood easily, one must put words together appropriately and must pronounce them well. Speech recognition can help assess whether words are grammatically and semantically appropriate, and pronunciation scoring can assess how well the student pronounced those words.

Reading Reading aloud can be tracked for errors in fluency as well as in pronunciation. Existing written materials can be leveraged in producing activities that require understanding of such material. It can also give the student experience in seeing the structures of the language in use in a variety of contexts. For many students, the experience of seeing the words in written form is important in learning.

Writing Active creation of materials in written form is required in many contexts. Although spoken language technology is not necessary for teaching or assessing this particular skill, it can be argued that writing ability will be enhanced through experience in the other areas. Writing skills are important in language learning; however, since our focus here is speech technology for language learning, it will not be discussed further.

1.2 Initial Learning versus Sustainment

Perhaps the most important lesson learned in teaching language is that learners differ greatly in what they already know, and in how they learn. An adult has presumably already learned one language and has a sense of how learning has best proceeded in the past. On the other hand, adults have acquired habits about their native language

that seem to be harder to overcome than is the case for children. This paper focuses on adults learning language, but distinguishes two cases: one in which the learner has little or no experience with the language being learned, and the other in which the learner already has some basics in the language and needs to re-learn forgotten knowledge or to build on some existing skills. Of course, in reality there is a continuum between the two, but the basic difference lies in just how much explicit learning needs to be provided; new learners benefit more from explicit support such as explanation, cultural notes, practice and drills, while individuals with higher levels of linguistic competence need to solidify existing knowledge, refresh forgotten abilities, or polish their vocabulary, pronunciation, and fluency. Learners at the sustainment level can take a more active role in their learning, and system design can be more flexible, allowing the learner to choose materials of interest. This is motivating for the learner; it also changes the job of the developer from listing a series of rules to choosing appropriate materials and presenting them in a flexible framework that can support a range of learning styles.

1.3 From Explicit to Implicit Learning

Although Krashen makes a sharp distinction between explicit and implicit knowledge (see Krashen 1982 for a discussion of learning versus acquisition), others argue that explicit knowledge can become implicit through practice (cf. Bialystok, 1978; Kenning and Kenning, 1990). We can describe different kinds of linguistic knowledge that learners acquire as representing a continuum ranging from explicit to implicit knowledge. It is also the case that learners vary greatly in how they make use of these two types of support for learning.

Explicit knowledge is represented by the conscious facts we have about language, those that are usually articulated as "rules" or other information about the linguistic features of a language. Acquisition of this knowledge is supported by components such as linguistic explanations, cultural notes, translations, and categorizations of linguistic structures with accompanying examples. These components provide a cognitive structure for understanding and learning, for clarification, or even for instruction in how to learn (e.g., "Practice saying these items aloud, listen with text and then without text").

Implicit knowledge has been described as that intuitive information upon which learners operate (Gass and Selinker, 1994). The acquisition of implicit knowledge can be supported by exposure to the language, learning the use and meaning of new words in context, exposure to a range of speaking styles and dialects, and so on. The function of such support can be to jog memory, polish fluency, fill in gaps in knowledge, or to raise language abilities to a higher level.

Although speech technologies can be used for explicit and implicit learning, they are perhaps most naturally used in implicit knowledge acquisition. Listening to examples and practicing roles using spoken language interactions can help a learner to infer the more explicit knowledge, and this is, after all, how the first language was learned.

1.4 Leveraging Existing Resources

Because adult learners tend to know how they learn, and to know why they want to learn a specific language, it is particularly important to provide appropriate materials for them. A child may be motivated to learn by simple examples and colorful pictures, in part because such features are suited to a child's cognitive development. A challenge in teaching adults lies in providing a range of interesting and complex material suited to an adult's cognitive development, but also suited to that adult's skills in the language being learned. The more that existing and relevant resources can be leveraged, the better we will be able to provide for the needs and interests of adult learners (and probably of children as well).

Although there is most likely no good replacement for dedicated and skilled language teachers, there are simply not enough to meet all language learning needs. The "shrinking" world means that more people are in contact, and that more corporations are multinational and multilingual. Countries themselves are becoming less homogeneous linguistically, and they are interacting more with each other. Although English has become more of a standard throughout the world, native speakers who are monolingual often find themselves at a severe disadvantage relative to others who can access the wealth of information that is not available in English. Computer-aided instruction can assist in some areas where teachers are not available, but we need more: we need tools that can help people access existing resources, no matter what the language of the source material.

Automatic translation is not capable of meeting this need and, based on recent progress, is not likely to be capable of meeting the need for some time. However, we can envisage a combination of technologies that will allow language learners and teachers to take advantage of existing resources, such as broadcast news and captions, movies and radio plays, and text and video learning materials in the new language. Can we develop technologies that allow students (and/or teachers) to access materials they know are of interest? If so, we can obtain leverage from the additional motivation of the student through use of situated and relevant materials, and we can obtain leverage through a saving in the development of course materials. Teachers and developers will still be needed when one-on-one teaching can be afforded, but their role will increasingly be one of focusing more on pedagogy and on tool development than on teaching a particular set of material to a particular individual. Examples of such technologies include speech transcription, on-line dictionaries, role playing (karaoke style) of video or radio material, automatic translation within the language to a targeted complexity level, and database access to the same word or concept in many different contexts. This may still be a futuristic vision, but keeping this goal in mind can help us make effective design and resource decisions now.

1.5 Introduction to VILTS

A major goal of the VILTS project was to build a demonstration system that combines listening comprehension, reading, and speaking in a rich learning environment. Based on a core of natural, unscripted dialogues recorded at various skill levels, each VILTS

lesson comprises five activities for each of the three language skills explicitly taught: listening comprehension, spoken conversation, and reading. Such a system can provide affordable, available, convenient, private, patient practice and feedback in support of language learning and sustainment.

The challenge of the project was to use speech technology (speech recognition and pronunciation scoring) to support the acquisition of listening, speaking, and reading skills. A concomitant challenge lay in bridging the gap between the many disciplines involved in the project. Helping the team of pedagogical experts to understand the possibilities and the limits of the technology was a major task, as was helping the speech engineers to understand the pedagogical goals. The activities developed support a communicative approach, and center around authentic, unscripted dialogues and related newspaper texts.

In the next section, we discuss the potential roles of various activities to support listening, speaking, and reading at initial learning and at sustainment levels. More details on the VILTS project can be found in Rypa (1996); the rest of this paper focuses on organizing what we have learned in the VILTS project and related projects in terms of how to support language learning and sustainment. In Sections 2 and 3, we outline how we believe speech technology can support user needs, give a few examples we have developed, and explore future directions.

2 Types of Activities Appropriate for Initial Learning

As argued above, initial rapid progress (relative to sustainment) in learning a language requires more explicit learning, and more support, especially grammatical and lexical. Translation of some key concepts beyond words may also be useful. The rest of this section focuses on activities that support skills in listening (2.1), reading (2.2), and speaking (2.3).

2.1 Learning to Listen

Listening is half of the communicative model in oral interactivity. As argued in Section 1, a major goal of listening exercises for a language learner is to build confidence in language production. We believe that it is important for learners to hear more than one speech style and voice to be able to generalize from specific native speech instances they have heard to the new voices and styles they will encounter in the real world. Furthermore, the pedagogical literature (see Section 1) suggests that acquainting students with oral materials just beyond their abilities is useful, and it can help them to develop and use word-spotting skills in contexts where complete understanding is not yet possible. Different levels and kinds of support must also be provided as the language is acquired (some students or teachers may want to turn off all text support on the screen, and/or all direct translation, while others may find this useful, or useful at different stages of learning). Learners need feedback to assess their own comprehension, and they need to be able to feel they are building on what has been learned to access more difficult

language tasks.

2.1.1 Examples

The VILTS architecture supports incremental progression to more complex structures through its architecture organized around beginning, intermediate, and advanced conversational dialogues. The conversations at the core of these activities were collected from 60 native speakers by a male interviewer and a female interviewer. Each interviewee was recruited based on demographics and ability to converse on one of 10 topics selected by our collaborating French teachers. The beginning conversations were based on simple questions that could be answered by "yes" or "no" (though they rarely were answered that way), or a one-line response. The intermediate conversations were based on questions that were not quite so simple, but could not be answered by one word, and the advanced conversations were associated with questions designed to elicit monologues from the interviewees on topics that engaged them.

The 60 speakers talking in each of the three modes described above accounted for 180 conversations on 10 separate topics, with several speakers for each topic. These resources allow students to practice with different speaking styles on one topic without the cognitive load of also switching topics entirely. Support for adapting to different speech styles is provided by versions of the same speakers reading a transcript of the original spontaneous conversation. Both the more careful, read speech style and the more casual, spontaneous style are available for listening in a summary exercise that bridges the listening and the speaking activities.

For some lesson materials, it may not be possible to find spontaneous conversations that have all the vocabulary and constructions that need to be taught. In a project that built on our VILTS work, we have collected speech, mapped to specific task requirements, in a variety of styles by having scripted materials produced in three different modes: (1) rapid reading, (2) carefully articulated and slow reading, and (3) non-read speech in which the participants carried on a similar dialogue with only keywords to guide their conversation.

In learning spoken skills, it is often desirable to focus on one standard dialect, but a student may need to have the ability to understand a greater variety of styles and dialects. Therefore, we have also explored the development of listening activities using new and very different dialects such as Haitian Creole as a complement to teaching standard French.

In addition to providing examples of various speech styles and dialects, we designed materials to develop listening skills that help the student to feel comfortable with word-spotting or incomplete understanding. Examples we have developed include

- "Qu'avez-vous entendu" (What did you Hear), in which students can listen to words in isolation from the conversation They are asked to click on phrases when they hear the same instance of these phrases in the actual conversation. The focus of such an activity can be on phrases with difficult sounds for the learner, or on target vocabulary.

- "Bingo," in which the difficulty of the word-spotting task is increased in that the

student must match a different speaker's utterance in a different style with the same word from the conversation. The difficulty is also increased because single words, as opposed to phrases, must be spotted.

- "Qu'avez-vouz entendu" has also been generalized to a word-spotting task in a Haitian Creole in the teaching of French.

In these listening activities, support is provided to the student through access to key words. At the student's request, a bilingual lexicon of key words relevant to the lesson appears. The student can hear words in isolation and in as many different contexts as available in the lesson, or in the on-line database.

2.1.2 Future Directions for Learning to Listen

As argued in Section 1, we would like to leverage existing resources of spontaneous speech to teach listening skills. Broadcast news with closed captions might be a good resource for such use. However, broadcast news sources may not expose the student to the variety of dialects and casual speech styles that can assist in generalization. Such resources may come from interviews, movies and radio plays, and other sources.

Since learners differ greatly in how they learn, and since language learning theorists differ greatly as well, we would like to explore architectures that support more student-guided use of resources, such as optional use of transcripts and use of on-line dictionaries. If sufficient feedback and assessment are also available, it is possible that students will quickly find the support and feedback mechanisms that foster their own progress.

In addition, we would like to explore the generalization of the use of key word assistance as a richer teaching tool with branches to examples not only of the specific instance, but of a structure or pedagogical component that is being targeted for the lesson (verb types, forms, and so on). Another important area to explore is more specific or targeted feedback; rather than just responding that an answer is wrong, the system should present a hypothesis as to why it is wrong and suggest or impose repair mechanisms.

2.2 Learning to Read

Reading, like listening, is generally a more passive skill. It can be important in support of listening skills for many learners by helping them to get the feel for words in a new language. Many learners rely heavily on visual input to aid comprehension.

As argued in Section 1, especially for adults, motivation is greatly enhanced by access to a choice of materials. In the VILTS architecture, after selecting the level (beginning, intermediate, advanced), the student can select the topic of the lesson. Translation of key words can be useful, as in the listening activities. Just as for word-spotting in listening, giving students assistance in gisting can greatly leverage their existing skills. Finally, as with listening activities, presenting materials just a bit beyond abilities will help learners to rely on context and to learn for themselves (as they learned their first language) what a new word might mean, based on its use.

2.2.1 Examples

In these example activities, a student is not provided a complete translation and may need to guess at some meanings:

- "Quel est le Titre" (What is the Title) forces the student to demonstrate comprehension of a passage by selecting the best title for a text. Feedback on wrong answers focuses on what might have been misunderstood ("It happened on Tuesday? Are you sure?").
- "Remplir les Blancs" (Fill-in-the-Blanks) forces the student to demonstrate comprehension by selecting words that are appropriate for missing words in a text. Such exercises can also teach the student about language structure by showing which words occur in similar environments.

2.2.2 Future Directions for Learning to Read

Given the state of the art of automatic translation, we would not rely on cross-language translation for language learning. However, the current state of the technology may be appropriate for translation within a language. For example, a translation system in the near future may be able to replace rare words with words or phrases the student has encountered in the lessons so far, and may be able to break up complex sentences into shorter, simpler sentences. Allowing the student to see both versions may be helpful. Other directions for future work include tool development for teachers in the design of such exercises, and tools for incorporating current texts chosen by the teacher or the student.

2.3 Learning to Speak

Speaking is a primary active language skill, and is often a primary motivation for learning a language. As argued in Section 1, speaking skills enhance listening skills as well. Both are needed for language's primary purpose: interactive communication.

New language learners have little experience in producing utterances in the target language, and need significant interactive practice in a conversational setting where they hear a native speaker, respond appropriately, and practice intelligibility, fluency, and pronunciation of the new language. Ultimately, the goal of practice is to be able to reuse learned pieces of a language to create utterances in navigating a new situation.

In our work we separate two aspects of speech: what was said, and how it was said. Our work in speech recognition aims to, as accurately as possible, determine what the student was trying to say, no matter how badly it might have been said. Pronunciation scoring, on the other hand, aims to score how well the utterance, or sounds comprising it, were produced. A major goal of the pronunciation scoring work is to provide a score that correlates well with human expert ratings (see Neumeyer et al., 1996 for details).

2.3.1 Examples

In VILTS, conversational activities progress from the simple to the more complex. Beginning activities requiring spoken output require the student to read one of three sentences on the screen that best answers an oral question. Later activities involve participation in a dialogue by reading a turn, and subsequently participating in a branching activity where students determine the direction of the conversation by choosing from a selection of responses. These activities have two types of feedback. First, the system indicates whether or not the response is understood, and then whether or not the response is correct. If the response is not articulated properly, either because the user is not trying to articulate one of the possible choices, or because the level of pronunciation is unintelligible, then the system prompts for a repeat. Specific examples are

- "Dites-Moi" (Tell Me). This is the first activity in the program where the user must produce speech, so it is deliberately simple. The topic is the same as the core dialogue, where the user hears questions similar to but not direct repetitions of questions that appeared in the dialogue. The user sees a graphic clue to the desired appropriate response. In keeping within the communicative framework, none of the responses is grammatically incorrect, although only one is appropriate in the given context.

- "Le Reporter" (The Reporter). In this activity, the learner assumes and reads one role in the lesson dialogue. This activity simulates conversation in that the dialogue will not continue if the learner does not articulate the turn properly. The user is prompted with a response such as "Pardon?" or "Je ne comprends pas" (I don't understand) if the turn is not understood. However, to avoid a situation in which the user cannot articulate properly and becomes trapped and frustrated, the dialogue continues after two tries that are not understood. The technological support here is relatively simple, but the opportunity for learners to assume a role and hear themselves in conversation seems to greatly aid prosodic control and fluency.

- "L'Interview" (The Interview, which is a branching activity). After completing the series of one-sentence multiple-choice activities and the role-playing activity described above, the learner proceeds to a branching activity where all answers are possible and appropriate; the student's selection guides the outcome of the conversation.

The three exercises described above can be useful in placing learners in an interaction with a native where they hear native speech and must respond appropriately. For these interactions to proceed, user input must be intelligible. Users can practice iteratively until they can smoothly step through the question-and-answer exercises or interact fluently through the entire interview or branching dialogue. Building on the VILTS work, we have also explored ways in which language learners can interact more freely with the system without reading materials that are on the screen.

Fluency, or the timely, proper, and smooth articulation of a response, is encouraged in the speech-based activities in two ways: first, each expected utterance from the learner is

timed by end-pointing so that too much hesitation or latency is cause for non-acceptance of the response. In these cases, the user can try again. Native speech examples of expected input by the user are always available as a model. In preliminary user testing of this type of activity, users almost universally indicated that they liked being "pushed" to formulate the input utterance well enough to be understood as well as quickly enough to map to native speech.

Although the VILTS project was designed to include pronunciation scoring, its scope did not allow for integration of this technology into the prototype. The project did support research and development in pronunciation evaluation. Our pronunciation scoring algorithms can return a score for sentences or for words, as well as for individual sounds. The scoring algorithms are designed to correlate well with human expert raters (see Neumeyer et al. 1996).

2.3.2 Future Directions for Learning to Speak

A major area for further work lies in feedback. Users who are interested in improvement invariably ask why a response was wrong, or how they can improve their pronunciation. This type of feedback must be present for language learners so that they can see and understand the results of their efforts. This feedback is present in some of the activities we have developed, but could be much more widely used. Our position is that any feedback should be validated to ensure that it is consistent and accurate. Inconsistent or inaccurate feedback can confuse and mislead a language learner; sadly, this is the case in many products currently on the market.

A generalization of our work in providing a variety of styles could lead to prompts and native models systematically available in more than one style. Of course, having two examples of each doubles the development load and speech storage requirements; therefore, the ability to adjust the speed of the output speech may be an appropriate technology to include. However, it is known that when people speed up or slow down their speech they do not simply scale the durations of all the sounds: some sounds are reduced or lengthened more than others. Therefore, research would be required to determine whether or not this is helpful to language learners.

Another area for expansion of speech exercises, one that pertains to other activities as well, is the inclusion of hints on how to use the activities, for example, suggestions on how to learn. Input from the experience of teachers could be incorporated, such as moving from slower and more careful pronunciation to more natural speech, or trying the system out at a higher level and moving back as necessary, depending on preferences. Other methodological suggestions might include initially using text to help with understanding the prompts, but then moving to a more realistic setting (native speakers that students will encounter in the street will probably not have subtitles at their feet).

Since the VILTS project was originally designed to support pronunciation assessment research, all activities were designed to prompt the student with whole text utterances along with one fill-in-the-blank exercise where the range of possible completions was limited to eight words or phrases. We felt that in order to give high-quality feedback on the pronunciation, the system would need to know exactly what the student had said. We still believe that this is important for pronunciation scoring, but we now believe that

it is valuable to separate the goal of learning what to say from the goal of learning how to say it. There are times to concentrate on one, or the other, and times to concentrate on both. We are developing new dialogue activities where the student does not rely on reading text in order to produce speech that a speech recognizer can reliably and robustly recognize.

3 Types of Activities Appropriate for Sustainment

As learners progress, their need for explicit support decreases as their level of language becomes more sophisticated. In the following sections, we review VILTS activities as sustainment support as opposed to material for new learners, though we realize that many activities are appropriate for both. Within each skill area, we motivate and discuss current examples and describe future directions.

3.1 Listening Sustainment

At more advanced levels where the learner needs to refresh or maintain existing language levels, more challenging, real-life materials form the most useful set of listening activities. Linguistic structures have presumably been learned and can be reinforced by hearing natural, challenging speech. In addition, more domain-specific vocabulary is often needed, as is vocabulary appropriate for more abstract subjects that are not as easily handled at lower levels (e.g., politics, technology, or philosophy). These conversations often include less commonly used vocabulary, and greater frequency of idioms or slang expressions (perhaps because the speaker may be more emotionally involved and concentrating more on content and less on form for these topics).

3.1.1 Examples

The 10 topics covered in the approximately 180 conversations of the VILTS corpus tended to cluster more into low- or high-level conversations by topic. Although interviewers were instructed to build each conversation from beginning, through intermediate, to advanced language use, this did not always happen. There were more interviews at the beginning level in such topics as health, travel, and leisure, for example, and more high-level conversations on topics such as technology, politics, and the environment. The VILTS prototype includes an example of an advanced conversation that centers on politics in France. Examples of advanced comprehension activity types are

- Dialogue. The conversation about French politics is 10 minutes as opposed to a 3-minute beginning dialogue, individual responses and sentences are longer, and there is a much wider range of vocabulary. Since the interview questions at the high level were often more abstract and designed to elicit longer responses (for example, "How did you develop your political convictions?"), the responses are often an exposition of an opinion and justification for that opinion, as opposed to the one-line responses in the beginning-level dialogues.

- Word-spotting. At the advanced level in the word-spotting exercises, the focus is on less-common expressions, used in a figurative expression, and on finer acoustic distinctions in vocabulary. In rapid speech, these distinctions are difficult to detect, but the knowledge that a high-level user has can help.

- Dialect practice. In work building on the VILTS architecture, the same conversation was recorded in standard French and in Haitian Creole, and both were used as a core dialogue for listening activities. This offers the student the opportunity to compare the two forms and to practice comprehension of Creole French. A subsequent activity modeled after "Qu'avez-vous entendu" (see above) allows the user to practice phrase-spotting. The pedagogical model here is to first build up confidence by selecting Creole phrases to be spotted that are similar to their French counterparts, as in cognates, and then to move on to less similar yet crucial vocabulary.

3.1.2 Future Directions for Listening Sustainment

Further support could be provided for sustainment use with VILTS by including several speakers responding to the same questions or interacting on the same topic in a series of dialogues. Several speakers reading the same text could provide exposure to different voices, accents, and speech styles. Lessons could also be updated with current materials, such as CNN broadcasts now available on the World Wide Web, or televised broadcasts in several languages. As a complement to new, incoming materials, tools such as on-line domain dictionaries could help users through unfamiliar texts. One could envision the use of speech recognition in combination with on-line dictionaries and domain-related reference materials to enable a user to listen and understand. Another area to explore is explicitly teaching students to become comfortable listening to speech in realistic noise environments, including competing speech from other talkers and/or over the telephone.

3.2 Reading Sustainment

At the lower levels of language learning, leveraging existing resources may be more difficult than at the sustainment level, since available resources, such as on-line newspaper text, may be too complex for the low language levels. At the sustainment level, with appropriate lexical tools, no rewriting should be necessary for most standard newspaper materials.

3.2.1 Examples

In the VILTS prototype we took advantage of available articles from LeMonde. Articles were selected to coordinate with the lessons in topic and in language level. Some software tools were used to search for appropriate sets of sentences and stories of appropriate length. Comprehension activities based on these materials required high-level skills in comprehension of the texts.

3.2.2 Future Directions for Reading Sustainment

As for the broadcast news audio resources, other on-line resources could assist in providing current and relevant text materials in support of reading sustainment and enhancement. There is a vast set of resources in the form of on-line newspapers and the Web that could be targeted to particular language learning needs, provided the right support mechanisms are in place to enable learning.

It is exciting to imagine the accessibility of up-to-date and relevant materials in service to language learning. Tools are needed for selecting these materials (on the basis of both topic relevance and language-level appropriateness). As mentioned earlier, within-language translation or paraphrasing might help advanced as well as beginning language learners.

3.3 Speaking Sustainment

An advanced learner needs to focus on "native-like" skills, for example, the ability to put words together and to pronounce them correctly without hesitation, and the learner needs to be able to do both at the same time without significantly affecting either. Furthermore, the learner needs to be challenged to participate in increasingly complex conversations, with many overlapping characteristics of high-level reading materials, for example, more domain-specific vocabulary, slang, and less commonly used expressions.

3.3.1 Examples

A dialogue on politics forms the core of lesson activities for the advanced lesson developed as a part of the VILTS prototype. This dialogue differs from lower-level dialogues in the program in several ways, most notably the length of the dialogue and the level of vocabulary.

- "Dites-moi" (Tell Me). In contrast to the beginning-level exercise, where the questions and responses are straightforward and and a few words in length, the interactions at the advanced level require greater subtlety and more complex responses.
- "Le Reporter" (The Reporter). The role-playing task here is the same as at the low-level, but the responses required of the learner are longer, reflecting the longer sentences and more complex structure of the core dialogue. Pronunciation and fluency become more of a challenge as the language level increases.
- "L'interview" (The Interview). The task and focus of this activity differs from the same activity at the low-level in that the learner must now pose a question, listen to the response, and select one of three choices for the most appropriate follow-up question.

3.3.2 Future Directions for Speaking Sustainment

At the sustainment level of language use, it is desirable to turn over more control to the user. For example, users might be allowed to set the thresholds for pauses allowed, so

that they are forced to formulate a response more quickly and to utter it more fluently once formulated. It would also be desirable to integrate the pronunciation scoring with conversational practice so that students must focus at the same time on what they are saying and how they are saying it in some exercises. This latter activity reflects what they will need to do in actual conversations.

Further work is also needed in devising and assessing the impact of different kinds of feedback on pronunciation problems. This might mean specific targeted exercises, various types of graphic and audio feedback, and comparisons with native speakers.

Finally, the real challenge of using speech technology in support of sustaining speaking skills is to simulate real conversations, in which the student is not reading from a set of sentences on the screen, but is creating language as in the real world. Because speech recognition accuracy is still far from that of humans, the best results depend on careful design of lessons and good communication between technologists and pedagogical experts.

4 Summary

Speech technology can serve pedagogical goals in many ways. It is only perhaps because of cultural and technical differences between speech algorithm developers and language teachers that more work has not begun in this area. Speech technology can be used in both initial language learning and language sustainment, although the use of the technology needs to take into account the differences between these two applications. In particular, initial language learners tend to need more explicit support, whereas learners at the sustainment level can be greatly empowered by tools to access current and relevant material of their own choice, in which language learning can be implicitly furthered. We need additional work in development of tools and architectures in order to achieve this vision of access to current resources such as broadcast news, television, and radio plays. The development of such tools could vastly increase our ability to both teach language and to give access to materials in new languages to more people. We do not see technology ever replacing humans. We do see it as a way to increase our human ability to communicate in the face of linguistic differences.

5 Acknowledgements

VILTS development was carried out at SRI International with the collaboration of government language teachers and software developers. We also acknowledge the cooperation of the Defense Language Institute (DLI) at Monterey, California. We thank Leo Neumeyer for leading the algorithm development in VILTS, for George Chen for system integration, Harry Bratt for further developing materials for targeted instruction, Laurence Devillers and Kate Hunicke-Smith for data collection, and Didier Disenhaus for pedagogical consulting.

References

[1] Anderson, J.R., 1982. "Acquisition of Cognitive Skills," Psychological Review 89, 69-406.

[2] Bialystok, E., 1978. "A Theoretical Model of Second Language Learning," Language Learning, 28:1, pp. 69-83.

[3] deBot, K., 1996. "The Psycholinguistics of the Output Hypothesis," Language Learning, 46:3, pp. 528-555.

[4] Gass, S., and Selinker, L., 1994. Second Language Acquisition, New Jersey: Lawrence Erlbaum Associates.

[5] Kenning, M., and Kenning, M., 1990. Computers and Language Learning, New York: E. Horwood.

[6] Krashen, S., 1981. Second Language Acquisition and Second Language Learning. 1st ed. Oxford: Pergamon Press.

[7] Krashen, S., 1982. Principles and Practice in Second Language Acquisition, New York: Pergamon Press.

[8] Neumeyer, L., Franco, H., Weintraub, M., and Price, P., 1996. "Automatic Text-independent Pronunciation Scoring of Foreign Language Student Speech," Proc. ICSLP 96, pp. 1457-1460, Philadelphia.

[9] Rypa, M., 1996. "VILTS: A Voice Interactive Language Training System," Proc. Computer Assisted Language Instruction Consortium, Albuquerque, New Mexico.

[10] Swain, M., 1995. "Three Functions of Output in Second Language Learning." In G. Cook and B. Seidlhofer eds., Principle and Practice in Applied Linguistics: Studies in honour of H.G. Widdowson, Oxford: Oxford University Press.