Automatic Speech Recognition for CALL: A Task-Specific Application for Training Nurses

Nicholas R. Walker
Henrietta Cedergren
Pavel Trofimovich
Elizabeth Gatbonton

Abstract: In recent years, language researchers and teachers have attempted to put meaningful communication at the centre of learners' classroom interactions. Yet the majority of existing computer-assisted language learning (CALL) applications have relied on largely non-communicative learner-computer interactions. The challenge facing CALL developers, therefore, is to explore new ways of providing learners with communicative practice. This article reviews existing uses of automatic speech recognition in second and foreign language teaching and describes the development of an innovative interactive automatic speech recognition system for developing second language speaking skills. This system uses video clips and the *EduSpeak* speech recognition system to simulate a nurse-patient interview. The system allows learners (for example, health care professionals whose first language is not English) to ask questions to an English-speaking patient and to receive both meaningful responses from the patient and feedback about their own pronunciation accuracy from the speech recognizer.

Keywords: Automatic speech recognition, communicative language teaching, health care, pronunciation, second language, speaking, task-based learning

Résumé : Au cours des dernières années, les chercheurs et les enseignants du domaine des langues ont tenté de placer la communication riche de sens au centre des interactions en classe. Malgré cela, la majorité des logiciels d'apprentissage des langues reposent encore sur des interactions largement non communicationnelles entre l'apprenant et la machine. Pour les développeurs de logiciels, le défi consiste donc à explorer de nouvelles voies permettant aux apprenants de s'exercer aux pratiques communicationnelles. L'article passe en revue les utilisations actuelles de la reconnaissance automatique de la parole dans l'enseignement des langues secondes et étrangères, et présente un système interactif innovateur de reconnaissance de la parole qui permet de développer les habiletés à l'oral en langue seconde. À partir de

© 2011 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 67, 4 (November/novembre), 459–479 doi:10.3138/cmlr.67.4.459

vidéoclips et du système de reconnaissance vocale *EduSpeak*, il simule une entrevue entre une infirmière et un patient. Le système permet aux apprenants (par exemple, des professionnels de la santé dont la langue forte n'est pas l'anglais) de poser des questions à un patient anglophone et de recevoir à la fois des réponses sensées du « patient » et de la rétroaction sur la précision de leur prononciation par le reconnaisseur de parole.

Mots clés : reconnaissance de la parole, enseignement des langues par l'approche communicative, soins de santé, prononciation, langue seconde, oral, apprentissage centré sur les tâches

Many countries, including Canada, share the assumption that their citizens have the right to equitable health care delivery. Thus, Canadian government institutions, both federal and provincial, have created measures intended to contribute to its realization. In compliance with this assumption, Quebec, the only official unilingual French province in Canada, has enacted the *Act Respecting Health and Social Services* (R.S.Q., chapter S-4.2). Underscoring the special status of the English speaking minority in the province, Article 15 of this act states,

> English-speaking persons are entitled to receive health services and social services in the English language, in keeping with the organizational structure and human, material and financial resources of the institutions providing such services and to the extent provided by an access program referred to in section 348. (Gouvernement du Québec, 2011)

The recognition of entitlement to health and social services, however, does not directly translate into implementation. Mediating factors such as demography and material and financial resources also play a role. Although there is a sizable population of English-speaking persons in Quebec – close to 1 million in a province of 7.5 million – they are not evenly distributed throughout the 16 health regions of the province. More than half of the English speakers live in the Montreal metropolitan area, while others are distributed unevenly throughout the different regions of the province (McGill-THRDP, 2006). These demographic factors impinge on the availability and quality of health care for the English speaking minority of the province. As a consequence, many English-speaking patients may not be able to receive service in English depending on where they reside (Bélanger, 2003).

Consequently, health care may be compromised by language availability. Thus language can function as a barrier to health.

'Language is medicine's essential technology,' affirmed Bowen (2001), because encounters between a patient and a caregiver require a shared channel of communication to succeed. Outcomes of health encounters depend on the appropriate exchange of information. Both patients and providers need to be understood by each other. The appropriate use of language both in form and content is a determinant of these encounters.

Two solutions have been used to overcome these language barriers. The first solution involves the presence of a third party – an interpreter – and the transformation of the patient-provider encounter into a triadic encounter. The second is a long-term solution in which the provider must acquire a degree of competence in the patient's language. The success of this solution rests on second language (L2) instruction. This is the relevant context for our article.

This article describes the development of a prototype interactive automatic speech recognition enabled system for developing task-specific L2 speaking skills for francophone nurses, a system that builds upon previously described pedagogical practices. We look at the use of automatic speech recognition (ASR) technology in a system that will allow us to combine meaningful interaction, interactive feedback, simulation, and asynchronous access in L2 training. We will first situate our research by providing an overview of existing uses of ASR in second and foreign language teaching.

The challenge of recognizing human speech

ASR refers to the process of turning speech into text, where *text* refers to a readable representation of the lexical content of an acoustic signal. Speech recognition is generally achieved by using two principal system architectures: knowledge-based systems and pattern-recognition systems (Coleman, 2005). Briefly, a knowledge-based system involves a three-step process. The first step involves deriving acoustic parameters from short intervals of the speech signal and then classifying the interval as a specific phoneme (e.g., /r/ in the word *red*) by comparing the parameters in the speech signal against those stored in the system (i.e., against expert knowledge). The next step involves repeating this classification process for each successive interval within the signal so that, for instance, the string [r], [ɛ], and [d] is identified as /rɛd/. In step 3, the system passes each string through its internal

© 2011 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 67, 4 (November/novembre), 459–479

dictionary. When a match is found, the system is said to have recognized one word (i.e., *red*).

A pattern-recognition system uses a two-step pattern-matching approach. In this approach, the parameters derived from longer stretches of the speech signal are compared to reference patterns stored in an internal database of examples until a match is found. Typically, these longer stretches are words rather than phonemes. For example, the system will first derive the parameters for an entire string (e.g., [rɛd]) and will then compare these parameters to stored reference patterns to find the text version of *red*. The same two-step process can be used for sentence-length speech signals, where instead of assembling a sentence from individual words the entire signal can be compared to stored sentence-length patterns. For instance, the acoustic pattern of *See John run* can be compared with internal dictionary entries like *See John swim*, *Watch John run*, and *See John run* for the best match. Of these two principal system architectures, the pattern-recognition system has been the more commercially successful.

The seemingly straightforward process of getting a machine to recognize speech is made more complicated by variability issues. One issue is that the peaks and troughs of the speech signal and the reference pattern may not line up perfectly when words are said at a faster or slower rate than the rate at which they were said when the reference database was made. Variability among speakers is another problem. Accent, male/female differences in vocal tract length, voice quality, and individual differences in health all affect speech characteristics. Moreover, unlike printed text, spoken language does not have clear-cut boundaries between adjacent words. Therefore, sounds at the beginnings and ends of words could be produced slightly differently depending on their immediate context. The word *let*, for example, will be realized as [lɛm] in *let me* but as [lɛt] in *let Anne*. Coping with all this variability requires a variety of technical solutions, including having the user train the recognizer by reading a script and creating reference patterns for all the possible realizations of every word in the recognizer's lexicon. In a speaker-independent system or a system that does not require the user to train the recognizer (e.g., see Bernstein, Najmi, & Ehsani, 1999), the range of variability for a particular dialect area is built into the system and avoids the need for recognizer training. This promises to be the best choice for language training systems that set a pronunciation standard that must be met.

ASR uses in L2 learning

In spite of lingering technological constraints on recognizing speech in general, and L2 speech in particular, a range of ASR enabled applications is now available to language learners, teachers, and researchers. Converting speech into text allows the learner to put the mouse and keyboard aside in favour of a microphone, turning previously receptive or silent tasks into speaking tasks. Before looking at the most innovative and promising uses of ASR in CALL, it will be instructive to evaluate prior uses of ASR, which will show just how far human-computer interaction has come. Although these earlier uses of ASR seem dated and tend to rely on questionable pedagogy, they nevertheless will be of interest to future material developers, who will want to use them as a source of possible task types to embed within the framework of virtual interaction. As such, the future of ASR enabled human-computer interactions for CALL crucially depends on a critical review and understanding of what has gone before.

ASR for drilling and repeating

Some of the most common applications of ASR in CALL involve drill-and-fill, listen-and-repeat, and read-aloud practice activities (e.g., Coniam, 1998; Mostow & Aist, 1999; Wohlert 1984). In these systems, the learner sits at the computer and listens to an audio prompt played on a cassette player. Then, he or she is presented with a series of questions on the computer screen, each of which is designed to elicit a correct pronunciation of the prompt. Non-recognition triggers a *try again* response from the system until the learner utters something approximating what the system expects to hear. By manipulating the acceptance threshold, the system requires a more exact match for advanced students, as opposed to beginners, between what the learners say and what the system expects.

Listen-and-repeat activities often try to quantify pronunciation problems visually, rather than textually. Just as Léon and Martin's (1972) intonation tutor had learners listen to a recording and repeat it back so that visual representations of the two intonation curves would appear together for comparison, ASR applications can also provide graphical feedback on goodness of fit for pronunciation. Some systems (e.g., Wohlert's [1984] original ASR system) compare a recorded template to the learner's attempt, and an

© 2011 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 67, 4 (November/novembre), 459–479

evaluative score is generated based on how similar the two speech samples are to each other. The result is visually presented in the form of a bar graph. Newer applications use a greater variety of wave forms, graph types, accuracy metres, and game-like visual representations for kids (see Atwell, 1999; Dalby & Kewley-Port, 1999; Wildner, 2002).

ASR for selecting, directing, and sequencing

Using a microphone instead of a computer mouse or keyboard can be helpful or even essential for computer users with limited vision or mobility, but for L2 learners, using one's voice to compose, select, and arrange words and sentences on a computer can transform traditional CALL tasks into speaking tasks. For instance, ASR allows learners to turn writing composition tasks into speaking tasks through the use of dictation software. Dictation allows learners to plan what to say during pauses, receive visual support for utterances, develop monitoring abilities, gain awareness of their articulation, improve their pronunciation, and promote the transfer of oral skills to written production (Myers, 2000).

Demonstrating how voice alone can be used to select between alternatives displayed on a computer screen, LaRocca, Morgan, and Bellinger (1999) describe a multiple-choice task which focuses the learner's attention on developing fine distinctions in pronunciation. By getting the learner to select between minimal pair items using voice alone, careful pronunciation becomes a matter of fine control. For example, fine control is needed for an L2 speaker to distinguish between, say, *tu* (you) and *toux* (cough) in French or *there are* and *they are* in English. Wachowicz and Scott (1999) provide another example of a selection task. In this task, the learner chooses one of three sentences as the next line of an interactive story. Instead of making the selection with a mouse click, the learner speaks the sentence to indicate the choice. Once the application has recognized the selection, three new related sentences become available to further the story. The learner thus selects and directs the events of a story, providing meaningful opportunities for language use. Wachowicz and Scott also discuss a multimodal arrangement task where the learner is presented with a jumbled sentence to unscramble by speaking the words in the correct order. The application recognizes each word the learner utters and sequences the words on the screen to unscramble the sentence.

Meaning-focused and purpose-oriented ASR

ASR has also made meaning-focused interactive speaking activities possible. Two types of interaction using ASR appear in the literature. The first involves command and control activities in which the learner speaks commands to a virtual agent who obediently does what is asked within a three-dimensional virtual world (Holland, Kaplan, & Sabol, 1999; Kaplan, Sabol, Wisher, & Seidel, 1998; Wachowicz & Scott, 1999). Wachowicz and Scott, for example, describe *TraciTalk*, software that allows the learner to engage in mystery-game problem-solving tasks by telling a virtual agent, Traci, to search for objects in the context of her video-clip world. Wachowicz and Scott note that the deficiencies of the speech recognition engine are cleverly matched to Traci's absent-minded personality so that the technological shortcomings of the software have a less disturbing, human quality to them. This and similar applications (e.g., Holland et al., 1999, Morton & Jack, 2005) are based on several pedagogical principles, which include providing the learner with opportunities for adaptive learning, where the software individualizes learning by responding specifically to the learner's errors through tailored remedial instruction.

The second type of meaning-focused interaction, and perhaps the most exciting opportunity for the use of ASR, is apparent in applications that involve meaningful speaking practice in face-to-face discussions between the learner and a virtual interlocutor. Dubbed the 'virtual dialogue method' by Harless, Zier, and Duncan (1999, p. 318), this use of ASR allows learners to engage in a role play where uttering a question into a headset microphone triggers a video clip (Harless et al., 1999) or animation clip (Ehsani, Bernstein, & Najmi, 2000; Morton & Jack, 2005) with a meaningful response to the question. The result is a convincing illusion of a meaningful two-way dialogue via videophone. For example, Harless et al. (1999) describe a system used by military linguists to practice their Arabic speaking skills. Working with a CD-ROM on a laptop with a microphone and a headset, the learner chooses from a limited set of three question-prompts at the bottom of the screen to interrogate a virtual character (e.g., a prisoner, a pilot). Depending on which question the student utters into the microphone, a different portion of video is played and a new set of relevant questions becomes available.

It should be noted here that while closed-response systems (where the learner's conversational turns are scripted) allow for story branching and therefore some flexibility in achieving conversational goals, more recently, the possibility of freeing the learner from a script by

© 2011 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 67, 4 (November/novembre), 459–479

using *chatbots* – programs simulating an open conversation with a user – as language-practice partners is being explored. It is true that many chatbots still use the keyboard and text-lines to interact with their users, but others 'speak' using speech synthesis and one also 'listens' using ASR (Coniam, 2008a). These advances toward the kind of language practice machine for which Atwell (1999) calls seem promising because of the unscripted and wide-ranging interaction they afford, but Coniam (2008a) reports that the speech quality of the reviewed chatbots is mechanical, artificial, or – at best – staccato. Furthermore, while some could provide corrective feedback on incorrect utterances, the questions from the user had to be well-formed to generate a meaningful response. Even then, the ability of chatbots to maintain a cohesive exchange over multiple conversation turns was limited (Coniam, 2008b). These limitations are non-trivial and cannot be overlooked. However, for practical purposes, the virtual dialogue can avoid the problems of speech synthesis and lack of conversational cohesion and it remains a worthwhile option for CALL application developers for other reasons, some of which are discussed below.

Pedagogical benefits of the virtual dialogue method

The virtual dialogue method described by Harless et al. (1999; see also Harless, Zier, Harless, & Duncan, 2003) represents a significant departure from earlier uses of ASR in CALL in terms of task type and goals. First, in comparison to previous uses of ASR that have followed a non-interactive, listen-and-repeat feedback sequence (as in pronunciation drills or sequencing tasks), the virtual dialogue method attempts to replicate authentic interaction with an interlocutor. Second, in earlier approaches to ASR, the goal of speaking to the machine has been to get explicit, usually graphical feedback on form (e.g., Wohlert, 1984). By way of contrast, the goal of using the virtual dialogue method is to communicate, and feedback is used to prompt repair as a necessary step toward successful communication. Last but not least, previous non-dialogic uses of ASR have typically focused on the form of the learner's utterance (its segmental content or intonation contours), leaving the meaning of a target utterance as an afterthought at best. In contrast, the virtual dialogue method requires the learner to manage cognitive resources more dynamically. By embedding language training within coherent, meaningful communication, the virtual dialogue method requires the learner to repeatedly switch his or her attention from form in the output to meaning in the input, imitating cognitive demands of authentic interaction.

© 2011 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 67, 4 (November/novembre), 459–479

Creating a virtual dialogue system

Considering the domain specificity of the virtual dialogue system described by Harless et al. (1999) for the particular linguistic needs of military linguists, tailor-made virtual dialogue systems will need to be designed and developed around high-priority, purpose-driven conversational tasks in other professional domains. We therefore set about to create a virtual dialogue system for teaching L2 speaking skills in another context, using task-based pedagogy (Van den Branden, Bygate, & Norris, 2007) and available technology. Our goal was to cater to the language needs of francophone nurses in Quebec, the only official unilingual French province in Canada. Because French is required as the language of work, many health caregivers are French speaking. As revealed in a linguistic needs analysis conducted on francophone nurses (Isaacs, Laurier, Turner, & Segalowitz, 2011), these health care providers often have low levels of spoken English proficiency and cite language as a source of difficulty in carrying out their health caregiving tasks.

The creation of our virtual dialogue system proceeded in several steps, and our goal here is to illustrate these steps for language teachers and researchers interested in the pedagogical uses of ASR technology. The first step was to select a routine, goal-oriented communicative task for learners to use in language training. The next step was to create a questionnaire task that would allow pre-service nursing learners to simulate a specific professional task, namely, taking a patient's medical history. The final steps were to video-record a person being interviewed while answering the medical history questionnaire and to build the virtual dialogue with an ASR system.

Selecting a task to model

Nurses engage in a variety of communicative exchanges with patients. These exchanges include taking a health history, explaining diagnostic measures, dealing with diagnosis, planning care and treatments, negotiating rehabilitation, and planning a follow-up (Price, 2004). Of these, the medical history interview appears to be an especially well-suited exchange on which to model a virtual dialogue. First, the medical interview contains a large set of standard questions that are asked each time a medical history is sought regardless of context and are asked more or less in a predictable order and sequence. This makes it relatively simple to script the learner's part of the virtual dialogue. Second, a necessary feature of medical interviews is that status is distributed

asymmetrically between provider and patient according to biomedical expertise (Gallagher, Gregory, Bianchi, Hartung, & Harkness, 2005). Since virtual dialogues by design depend on the initiative of the learner-user in the selection and maintenance of conversational topics, this asymmetry of control over the conversation will seem natural and appropriate from the perspective of the health care professional using the system. In other words, the learner is unlikely to be surprised that the virtual patient plays too passive a role in the virtual dialogue since it is quite normal and even expected for patients to let health care providers take control of the conversation to achieve the goal of collecting enough pertinent health information to solve the patients' health problems.

Finally, because health care professionals typically do not react to patients' newsworthy answers, the absence of answer-assessments means that the awkward silences following a patient's response to a medical question are commonplace in medical interviews (Jones, 2001). This makes it simpler to model the learner-user's side of a virtual dialogue based on a medical history interview.

Because a speaking task becomes truly communicative only when there is an inherent reason to engage in communication (Gatbonton & Segalowitz, 2005), we created a questionnaire for the learner to complete on paper as part of the virtual interview. The questionnaire was based on two examples of medical surveys. The first example is a standard employment medical history questionnaire (Scripps Institution of Oceanography, 2004) used during the job-application process, and the second is a longer questionnaire used for health assessment at the University of Florida Shands Executive Health Center (Flint, 2003). The task sheet that resulted from a synthesis of the two retains the five core elements of the health questionnaire (personal data, medical history, family history, a review of systems, and daily living questions) condensed into a single page. Questions that are related to administrative tasks, such as asking patients for their medicare numbers (e.g., in Quebec, this is done during the patients' first admission into the hospital), and upon which health care providers may or may not focus during the medical history interview were not investigated in detail.

The questions were not selected on the basis of their grammatical structures but on how well they corresponded to the appropriate dialogue function at a particular point in the interview. That is, they were pragmatically appropriate requests about the medical experiences of the patient. A post-hoc analysis of the structure of some of these questions (see below) revealed that they have common patterns.

During the virtual dialogue, the learner has to complete the questionnaire with information she receives from the virtual patient; thus, she has a real purpose to engage in a conversation with the patient.

Creating the script

Once this questionnaire was ready, questions were created from questionnaire items for use during the filming of the virtual dialogue. Two opening questions ('Are you here for the medical history interview?' and 'Can I ask you a few questions?') were added at the beginning of the interview script to set the purpose and nature of the interview. The next questions elicited personal data (e.g., 'What is your name?' and 'How old are you?') and specific information pertaining to the patient's medical history and review of systems (e.g., 'Have you ever had a hernia?' and 'Do you get frequent headaches?'). Several questions targeted issues of daily living; tobacco, drug, and alcohol use; and type of employment (e.g., 'Do you smoke?' and 'How many hours do you work in a week?'). Finally, a closing statement ended the interview encounter (i.e., 'Thanks'). The medical interview questions are found in Appendix A.

Sequential parasitism – the questioning strategy of adding to a previously asked question using phrasal increments such as 'Do you have any other medical problems? No heart disease? Any lung disease as far as you know? Any diabetes?' (Stivers & Heritage, 2001, p. 152) – was avoided to maximize the learner's exposure to repetitive, formulaic aspects of English question utterances that are typical of an interviewer's role. Thus, questions containing *do you + VERB* occurred in the script 23 times, *are you + NOUN* and *have you + VERB* 8 times each, *is there + NOUN* 5 times, and *did you + VERB* 4 times. Repetitions of *what*, *how*, and *when + VERB* occurred 11, 9, and 4 times, respectively. Such a high number of oral repetitions of question words and phrases deployed in a genuinely communicative exchange is expected, therefore, to have a positive effect on the learner's attainment of appropriate interviewing forms.

Creating the virtual dialogue

Our virtual patient was a diabetic man (native speaker of Quebec English) in his early forties who was willing to answer questions about his own medical history. This patient's unscripted answers to the prepared medical history schedule were video-recorded using a Sony DCR-DVD403 Handycam. The patient also videotaped a

variety of clarification requests (e.g., 'What?' and 'Could you speak up?') for their subsequent use as system prompts for the learner to try again. The patient was not a professional actor. All his answers were authentic responses to the prompted questions. No attempt was used to elicit specific grammar forms or technical jargon. To provide users of the virtual dialogue with pronunciation models of interview questions, the voice of a female native speaker of standard Canadian English was added, producing each of the scripted questions using a Sony ICD-P210 digital voice recorder. A female speaker was considered appropriate for this female-dominated profession. Dialogic interaction between the user and the virtual patient was enabled with the use of SRI International's *EduSpeak* speaker-independent speech recognition engine (Franco et al., 2000), and dialogue management interface grammar was developed with Microsoft's Visual Studio 2005. The result was named the *Virtual Language Patient* (VLP; for technical specifications, see Walker, Cedergren, Trofimovich, Gatbonton, & Mikhail [2008]).

Features and content of the VLP

The VLP system runs on a PC equipped with a noise-cancelling microphone with headphones and a mouse. The intuitive graphical-user interface of the VLP is set up to be simple to use without the need for extensive training, instructions, or demonstration videos. Upon launching the software, a video image of the patient appears at the centre of the screen (Figure 1). He does not say anything at first but just looks around, waiting quietly for the learner to begin the virtual dialogue by asking the first question. Underneath the video image is a box with the first question to be asked: 'Are you here for the medical history interview?' The learner initiates the virtual dialogue by clicking on the 'Recognize' button and pronouncing the question into the microphone. If the system recognizes the learner's utterance as the expected sentence, a video plays the patient's response, 'Yup,' and the second question appears on the screen. In this way, the VLP employs a closed-response design where the user is prompted with one script item at a time to maintain a focus on pronunciation accuracy.

The system displays a feedback panel after each successful recognition, providing feedback on the confidence ratings associated with each word and the whole utterance (as seen in Figure 2). On the screen, ratings for words that fall below a threshold of acceptability are displayed in red, while ratings for all other words are displayed

FIGURE 1
The graphical user interface of the VLP

in green. The learner can thus get a sense of which words he or she needs to say more clearly. If the system does not recognize the question asked or if the entire sentence does not meet the predetermined threshold of what is acceptable as a result of either poor microphone placement or errors in pronunciation, a video with a clarification request (e.g., 'Could you say that again, please?') plays and an opportunity to try again is made available. The learner may at this point wish to hear an audio recording of a native speaker pronouncing the sentence. This is possible at any time by clicking on a button to the left of the question-prompt and then listening to the recording through the headset.

Some adjustments to the system's speech recognizer are available to the learner who can change the microphone sensitivity, headset volume, and recognition acceptance threshold using the mouse. The advantage of being able to set the acceptance threshold to a lower or higher level is that the learner can make the patient more or less forgiving of pronunciation errors and thus make the pronunciation demands of the experience less frustrating or more challenging according to his or her needs. This is important because getting the patient to recognize the question is necessary to advance through the dialogue and complete the task of taking his medical history. Provided with a pen and a paper copy of the medical history questionnaire, the learner is prompted to ask each of the 71 virtual dialogue interview questions

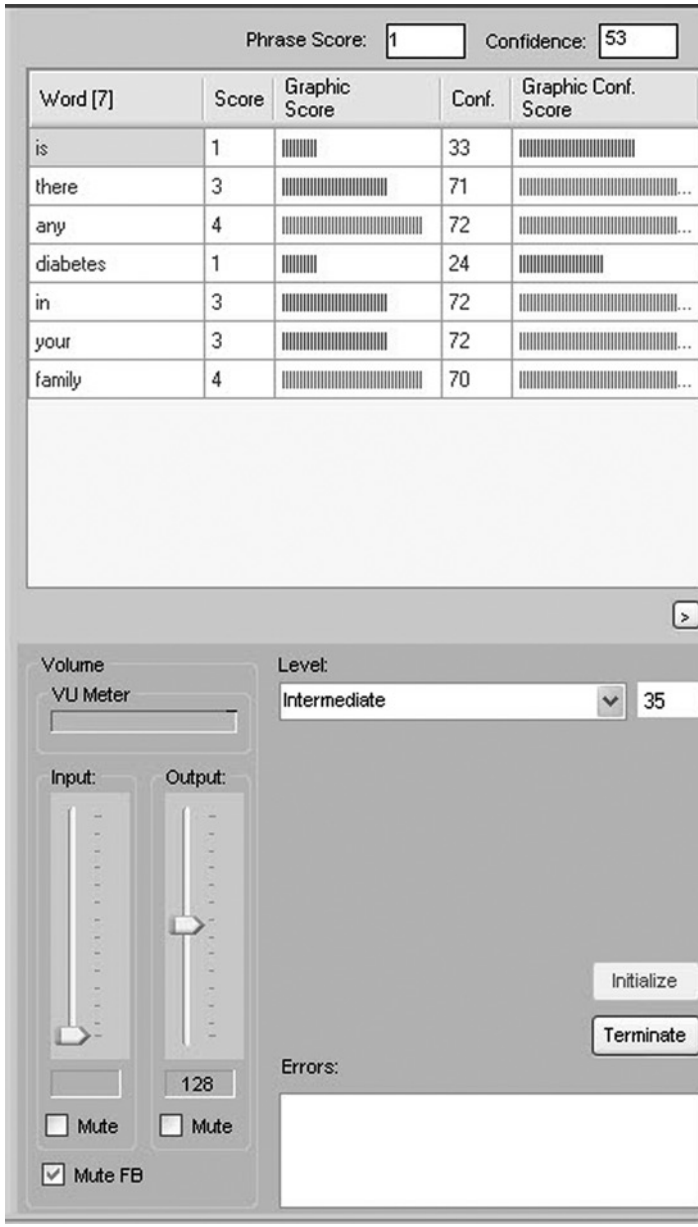

FIGURE 2
The feedback panel

in a fixed sequence and to view a video recording of the patient's unrehearsed answer to each question.

The suitable pairing of pedagogy and technology

Our principal goal for the VLP was to find a suitable pairing of pedagogy and technology, which could inform the design of an instructional tool that might prove useful for nurses motivated to improve health care for linguistic minorities through language training. We intended to avoid decontextualized drilling of language forms since the reliance upon mechanical practice without opportunities for communicative practice has been largely discredited as an approach to effective language teaching (Wong & VanPatten, 2003). We hoped, instead, to ensure that all oral repetitions would occur in a genuinely communicative context in which each formulation of a repeated structure would be part of a meaningful message conveyed to a receptive interlocutor (see Gatbonton & Segalowitz, 2005). The speech recognition system described here – at least from a purely design standpoint – appears to fit the bill of being a talking machine that might be worth the effort it involves to create.

An additional value of ASR enabled virtual dialogues is in their versatility to address a variety of learning issues. In addition to the use of virtual dialogues for speaking practice (as we have shown here), other aspects of communication, various comprehension strategies, and goal-oriented social talk – not to mention aspects of politeness, grammar, and word choice – all seem promising areas for research and development in the future of the virtual dialogue. Walker, Trofimovich, Cedergren, and Gatbonton (2011) report a study that critically evaluates the effectiveness of the VLP. In brief, the five francophone nurses who used the VLP improved their pronunciation, among other things, and the improvement is indicative of the feasibility of using the VLP for this purpose. Along the way, if conversation forms of feedback can be shown to be easier for learners to interpret and more effective at drawing the learner's attention to formal aspects of speech, it could be that graphical feedback will disappear. Learner-adaptive difficulty settings could become a type of video-game-like score-keeping feature of virtual dialogues that learners use to gauge their progress. In this way, language learning may begin to be seen for what it has always been – an exciting opportunity to meet new people (virtual or otherwise) and a sophisticated form of play.

© 2011 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 67, 4 (November/novembre), 459–479

Limitations

The system we have described has three design limitations that deserve mention here. The first is the lack of information given with an utterance rejection. While an utterance that produces a recognizer score that surpasses a particular acceptance threshold will display the high and low scoring words in red and green, an utterance that does not meet the threshold provides the learner with a general clarification request only. No other information about the learner's performance is given. One improvement could be to program the recognizer to 'listen' for specific content words and trigger video clips of clarification requests that could help the learner localize the source of the problem. For example, if the learner pronounces every word in the prompt 'Is there any diabetes in your family?' correctly except the word 'diabetes,' the virtual patient could respond, 'Is there any WHAT in your family?' The second limitation of the system is its closed-dialogue design. The lines are scripted to aid nurses in the development of a repertoire of accurately pronounced medical history questions. While useful, given the routine and formulaic nature of the majority of medical history interviews that nurses might encounter, the system does not provide opportunities for self-expression or opportunities to test nurses' own formulations of medical history questions in a communicative exchange. The third limitation of the system is that learners are currently exposed to only one variety of Quebec English. Our virtual patient is a native Montrealer with no marked ethnic identity. Future development of the system could include virtual patients that represent the social and regional diversity of Quebec English (Boberg, 2004, 2010).

Conclusion

Every powerful new piece of technology introduced into language learning since Clarke (1918) seems to follow the same pattern. Each is first greeted with hope and some degree of enthusiasm as learner motivation increases briefly in response to the novelty of the technology, but as limits become apparent disappointment takes hold, leading to general abandonment. This pattern could repeat itself with the uses of ASR in CALL unless the central purpose becomes the exchange of meaningful messages. Accuracy and fluency were among the early targets of CALL applications using ASR, with listen-and-repeat type focus-on-form activities. Opportunities for

creative language use have also been revealed by exploiting the hands-free multimodal functionality of ASR for the learner to compose, select, direct, and sequence language according to his or her taste. Nevertheless, the most exciting developments for the use of ASR in CALL have targeted meaning-driven and purpose-oriented oral interaction (Harless et al., 1999, 2003; Holland et al., 1999; Morton & Jack, 2005). For the ASR enabled talking machine to avoid the fate of the phonograph, teachers and learners will want to see the full integration of language accuracy and fluency practice within opportunities for creative language use in meaning-driven and purpose-oriented speaking tasks where the costs, usability, and technological limitations remain well below the not-worth-the-trouble threshold for learners and teachers.

Correspondence should be addressed to **Pavel Trofimovich**, Concordia University, Department of Education LB-579, 1455 de Maisonneuve Blvd. W., Montreal, Quebec, H3G 1M8. E-mail: pavel.trofimovich@concordia.ca.

References

Atwell, E. (1999). *The language machine*. London, UK: The British Council.

Bélanger, M. (2003). *L'accès aux soins de santé pour les communautés minoritaires de langue officielle*. Ottawa: Chambres des Communes.

Bernstein, J., Najmi, A., & Ehsani, F. (1999). Subarashii: Encounters in Japanese spoken language education. *CALICO Journal*, *16*(3), 361–384.

Boberg, C. (2004). Ethnic patterns in the phonetics of Montreal English. *Journal of Sociolinguistics*, *8*(4), 538–568. doi:10.1111/j.1467-9841.2004.00273.x

Boberg, C. (2010). *The English language in Canada: Status, history and comparative analysis*. Cambridge, UK: Cambridge University Press.

Bowen, S. (2001). *Language barriers in access to health care*. Ottawa: Health Canada.

Clarke, C. (1918). The phonograph in modern language teaching. *Modern Language Journal*, *3*(3), 116–122. doi:10.2307/313729

Coleman, J. (2005). *Introducing speech and language processing*. Cambridge, UK: Cambridge University Press.

Coniam, D. (1998). The use of speech recognition software as an English language oral assessment instrument: An exploratory study. *CALICO Journal*, *15*, 7–23.

Coniam, D. (2008a). An evaluation of chatbots as software aids to learning English as a second language. *ReCALL*, *13*. Retrieved from: http://www.eurocall-languages.org/review/13/#coniam

© 2011 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 67, 4 (November/novembre), 459–479

Coniam, D. (2008b). Evaluating the language resources of chatbots for their potential in English as a second language. *ReCALL*, *20*(1), 98–116. doi:10.1017/S0958344008000815

Dalby, J., & Kewley-Port, D. (1999). Explicit pronunciation training using automatic speech recognition technology. *CALICO Journal*, *16*, 425–445.

Ehsani, F., Bernstein, J., & Najmi, A. (2000). An interactive dialog system for learning Japanese. *Speech Communication*, *30*(2–3), 167–177. doi:10.1016/S0167-6393(99)00042-4

Flint, N. (2003). Medical history questionnaire. University of Florida Shands Executive Health Center. Retrieved from: http://www.exechealth.ufl.edu/MHQ_questionnaire/EHC-MHQ.pdf

Franco, H., Abrash, V., Precoda, K., Bratt, H., Rao, R., Butzberger, J., … Cesari, F. (2000). The SRI EduSpeak system: Recognition and pronunciation scoring for language learning. *Proceedings of InSTILL (Integrating Speech Technology in Language Learning)*. Dundee, Scotland: University of Abertay.

Gallagher, T.J., Gregory, J., Bianchi, A.J., Hartung, P.J., & Harkness, S. (2005). Examining medical interview asymmetry using the expectation states approach. *Social Psychology Quarterly*, *68*(3), 187–203. doi:10.1177/019027250506800301

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: A focus on access to fluency. *Canadian Modern Language Review, 61*(3), 325–353. doi:10.3138/cmlr.61.3.325

Gouvernement du Québec. (2011). *Loi sur les services de santé et les services sociaux L.R.Q., c. S-4.2.* Retrieved from: http://www.publicationsduquebec.gouv.qc.ca/fre/products/978-2-551-24567-3

Harless, W.G., Zier, M.A., & Duncan, R.C. (1999). Virtual dialogues with native speakers: The evaluation of an interactive multimedia method. *CALICO Journal*, *16*, 313–337.

Harless, W.G., Zier, M.A., Harless, M.G., & Duncan, R.C. (2003). Virtual conversations: An interface to knowledge. *IEEE Computer Graphics and Applications*, *23*(5), 46–52. doi:10.1109/MCG.2003.1231177

Holland, V.M., Kaplan, J.D., & Sabol, M.A. (1999). Preliminary tests of language learning in a speech-interactive graphics microworld. *CALICO Journal*, *16*, 339–359.

Isaacs, T., Laurier, M.D., Turner, C.E., & Segalowitz, N. (2011). Identifying second language speech tasks and ability levels for successful nurse oral interaction with patients in a linguistic minority setting: An instrument development project. *Health Communication*, *26*(6), 560–570. doi:10.1080/10410236.2011.558336. Medline:21512923

Jones, C.M. (2001). Missing assessments: Lay and professional orientations in medical interviews. *Text*, *21*(1,2), 113–150. doi:10.1515/text.1.21.1-2.113

Kaplan, J.D., Sabol, M.A., Wisher, R.A., & Seidel, R.J. (1998). The Military Language Tutor (MILT) program: An advanced authoring system. *Computer Assisted Language Learning*, *11*(3), 265–287. doi:10.1076/call.11.3.265.5679

LaRocca, S.A., Morgan, J.J., & Bellinger, S.M. (1999). On the path to 2X learning: Exploring the possibilities of advanced speech recognition. *CALICO Journal*, *16*, 295–310.

Léon, P.R., & Martin, P. (1972). Applied linguistics and the teaching of intonation. *Modern Language Journal*, *56*, 139–144. doi:10.2307/324034

McGill-THRDP. (2006). Training and Human Resources Development Program, funded by Santé Canada.

Morton, H., & Jack, M.A. (2005). Scenario-based spoken interaction with virtual agents. *Computer Assisted Language Learning*, *18*(3), 171–191. doi:10.1080/09588220500173344

Mostow, J., & Aist, G. (1999). Giving help and praise in a reading tutor with imperfect listening – Because automated speech recognition means never being able to say you're certain. *CALICO Journal*, *16*, 407–424.

Myers, M.J. (2000). Voice recognition software and a hand-held translation machine for second language learning. *Computer Assisted Language Learning*, *13*(1), 29–41. doi:10.1076/0958-8221(200002)13:1;1-K;FT029

Price, B. (2004). Conducting sensitive patient interviews. *Nursing Standard*, *18*(38), 45–55. Medline:15199710

Scripps Institution of Oceanography. (2004). Medical history questionnaire. Retrieved from: http://shipsked.ucsd.edu/General_Info/Forms/mhq-1.pdf

Stivers, T., & Heritage, J. (2001). Breaking the sequential mold: Answering 'more than the question' during comprehensive history taking. *Text*, *21*(1–2), 151–185. doi:10.1515/text.1.21.1-2.151

Van den Branden, K., Bygate, M., & Norris, J. (Eds.). (2007). *Task-based language teaching: A reader*. Amsterdam: John Benjamins.

Wachowicz, K.A., & Scott, B.L. (1999). Software that listens: It's not a question of whether, it's a question of how. *CALICO Journal*, *16*, 253–276.

Walker, N., Cedergren, H., Trofimovich, P., Gatbonton, E., & Mikhail, E. (2008). Someone to talk to: A virtual patient for medical history interview training in a second language. In L. Gómez Chova, D. Martí Belenguer & I. Candel Torres (Eds.), *INTED 2008 Proceedings* (pp. 1–9). Valencia, Spain: International Association of Technology, Education and Development.

Walker, N., Trofimovich, P., Cedergren, H., & Gatbonton, E. (2011). Using ASR technology in language training for specific purposes: A perspective from Quebec, Canada. *CALICO Journal*, *28*, 721–743.

Wildner, S. (2002). 'Learn German Now! Version 8' software review. *CALICO Journal*, *20*, 161–174.

Wohlert, H.S. (1984). Voice input/output speech technologies for German language learning. Die Unterrichtspraxis. *Teaching German*, *17*, 76–84.
Wong, W., & Van Patten, B. (2003). The evidence is in: Drills are out. *Foreign Language Annals*, *36*(3), 403–423. doi:10.1111/j.1944-9720.2003.tb02123.x

Appendix A

1. Are you here for the medical history interview?
2. Can I ask you a few questions?
3. What is your name?
4. How do you spell your first name?
5. How do you spell your last name?
6. What is your address?
7. What is your phone number?
8. When is your birthday?
9. What year were you born?
10. How old are you?
11. Are you married?
12. Are you divorced?
13. What is your sexual preference?
14. Do you have any allergies?
15. What is your blood type?
16. Has anyone in your family ever had heart disease?
17. Is there any epilepsy in your family?
18. Is there a history of mental illness in your family?
19. Is there any cancer in your family?
20. Did your aunts smoke?
21. Is there any high blood pressure in your family?
22. Do you have high blood pressure?
23. Are you taking any medication for it?
24. What are you taking?
25. How often do you take it?
26. Medication should be taken regularly.
27. Are there any side-effects to the medication?
28. Is there any diabetes in your family?
29. What can you tell me about your diabetes?
30. How did you discover that you were a diabetic?
31. Is your diabetes under control?
32. How is your blood glucose?
33. How is your eyesight?

34. Do you wear glasses or contact lenses?
35. Have you ever had a hernia?
36. Do you have a chronic cough?
37. Do you get chest pains?
38. Do you have arthritis?
39. Have you had any kidney trouble?
40. Do you ever get seizures?
41. Do you ever get rashes or skin troubles?
42. Do you have hearing difficulties?
43. Have you ever had a mental breakdown?
44. What happened?
45. Do you get frequent headaches?
46. Do you suffer from dizziness?
47. What causes it?
48. Have you ever had a sexually transmitted disease?
49. Do you do drugs?
50. Do you have asthma?
51. Has there been any tuberculosis in the family?
52. Do you suffer from back problems?
53. Do you suffer from varicose veins?
54. When was the last time you got a tetanus shot?
55. Try to remember.
56. A long time ago?
57. When did you last see a dentist?
58. Have you had any problems with your prostate?
59. Have you noticed any penile discharge?
60. Have you noticed any testicular lumps?
61. Are you seeing a doctor for anything at the moment?
62. Do you smoke?
63. How many cigarettes do you smoke in a day?
64. When did you start?
65. Do you drink alcohol?
66. Really?
67. Do you take any drugs?
68. Are you taking any medication?
69. What is your job?
70. How many hours do you work in a week?
71. Thanks.

© 2011 *The Canadian Modern Language Review/La Revue canadienne des langues vivantes*, 67, 4 (November/novembre), 459–479