The Development and Feasibility of a Speech Recognition-Enabled Virtual Patient for Training Francophone Nurses to Conduct Medical History Interviews in English

Nicholas Walker

A Thesis

in

The Department

of

Education

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Arts (Applied Linguistics) at
Concordia University
Montreal, Quebec, Canada

March, 2009

Your file Votre référence
ISBN: 978-0-494-63280-2
Our file Notre référence
ISBN: 978-0-494-63280-2

Abstract

The Development and Feasibility of a Speech Recognition-Enabled Virtual Patient for Training Francophone Nurses to Conduct Medical History Interviews in English

Nicholas Walker

Low language proficiency remains a significant barrier to healthcare access for many patients throughout the world. Training healthcare professionals in the language of their minority language patients, therefore, should lead to greater healthcare access, lowered costs, better health outcomes, and improved patient satisfaction (Zambrana, Molnar, Munoz, & Lopez, 2004). One important aspect of language training involves the development of accessible, appropriate, and pedagogically sound language training materials. The first goal of this thesis is to describe the development of the "Virtual Language Patient," a computer-based language training module based on the Virtual Dialogue Method (Harless, Zier, & Duncan, 1999). The prototype system under consideration employs automatic speech recognition (ASR) technology, using video clips of a simulated medical history interview with a minority language patient. The second goal of this thesis is to report the findings of a proof-of-concept feasibility study where the ease of operability and fitness of purpose of this prototype system were explored. Five nursing-students at a French language nursing college in Quebec reported the system to be easy to operate and fit for their anticipated language learning needs in terms of target language, choice of interlocutor, mode of interaction, task type, and corrective feedback. Training effects on participants' pronunciation scores, speech rate, and sense of preparedness for real life medical interviews suggest that the system can be effective in language training for healthcare professionals. Implications for the improvement of this and future virtual dialogue systems are also discussed.

Dedication

This thesis is dedicated to Gregory and Dorothy Walker for all of their encouragement and support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: HEALTHCARE ACCESS FOR LINGUISTIC MINORITIES

Effective communication between healthcare providers and their patients is an important factor in patient satisfaction. Global patient satisfaction with healthcare has been found to be lower among patients who have more provider-patient communication problems (Charles, Goldsmith, Chambers, & Haynes, 1996). The most commonly reported problems involve failures on the part of providers to communicate when communication is expected, such as failing to explain what the provider is intending to do while examining a patient, keeping the patient in the dark about daily routines, and failing to communicate adequately during discharge planning.

Global dissatisfaction with healthcare becomes more acute, however, when the provider and patient cannot effectively communicate in each other's language. In a study of 26 international medical graduates enrolled in an Internal Medicine residency program at Wayne State University, a significant correlation was found between language proficiency and patient satisfaction (Eggly, Musial, & Smulowitz, 1999). In another study, Spanish-speaking patients in San Francisco were also found to be less satisfied with the care they received from non-Spanish speaking physicians (Fernandez et al., 2004), and in the North-eastern United States, a variety of non-English speaking patients reported less satisfaction than their English-speaking counterparts with emergency room care, courtesy and respect, and with discharge instructions (Carrasquillo, Orav, Brennan, & Burstin, 1999). Comparisons made between members of the same linguistic minority group also showed a correlation between language proficiency and satisfaction levels. For example, low-English-proficiency Korean patients over the age of 60 in the U.S. were

less likely to be satisfied with the healthcare service they received than Koreans with higher levels of proficiency (Jang, Kim, & Chiriboga, 2005).

Indeed, not speaking the language of the patient adds to a patient's suffering. One emergency department study found that Spanish-speaking Hispanic patients were half as likely to receive analgesia in the treatment of their long bone fractures as their English-speaking counterparts (Todd, Samaroo, & Hoffman, 1993). Worse still, a failure to anticipate communication problems and accommodate low-language proficiency clientele can turn fatal, as was recently illustrated in a news story of an Albanian immigrant who killed himself, thinking his wife had been diagnosed with AIDS when hospital staff told him his wife's blood type was A-positive (The Canadian Press, 2007).

One obvious solution to increasing healthcare access to linguistic minorities is to use interpreters. Whereas the use of hospital-trained interpreters in pediatric emergency departments was found to increase parents' satisfaction with their physicians and nurses (Garcia, Roy et al., 2006), in primary care medical interviews a reliance upon interpreters is somewhat more problematic. Aranguri, Davidson and Ramirez (2006) observed that during regular doctors' appointments with Hispanic patients about half of the words exchanged between doctor and patient were missing from interpreters' translations. All small talk, known to increase patients' emotional engagement in their treatments and in their doctors' ability to get a comprehensive patient history, was eliminated. Patients' questions, an important indication of patients' engagement with their own care, were also significantly reduced when an interpreter was used.

To reduce the heavy reliance on interpreters in healthcare, Zambrana et al. (2004) recommend hiring more minority, linguistically competent, and culturally competent

healthcare providers in managed care networks. They argue that having healthcare providers that speak the same language as their patients will lead to lowered costs, greater healthcare access, better health outcomes, patient satisfaction, and patient compliance. There is evidence to support this claim. One study investigating patient outcomes found that asthma patients cared for by doctors who spoke their language were more likely to take their medication and less likely to miss office appointments or make resource-intensive emergency room visits than patients with doctors who did not speak their language (Manson , 1988). Another study found that patients whose doctors spoke their language asked more questions and had a better recall of their doctor's recommendations (Seijo, Girmez, & Freidenberg, 1991).

Effective Language Training for Healthcare Providers

Where bilingual healthcare providers are currently in short supply, medical school students and in-service medical professionals will need to receive effective language training to help them become proficient in a second language (Eggly et al., 1999). An investigation of language training preferences at a South Carolina medical school identified listening and speaking skill development as a much higher priority than reading and writing amongst the 165 pre-service health professionals surveyed (Lepetit & Cichocki, 2002). Appreciating the importance of authentic face-to-face encounters, these same respondents also indicated a preference for learning situations that would bring them into direct contact with speakers of the target language.

In contrast, a qualitative case study reporting on the learning needs of in-service healthcare professionals (those who already had considerable contact with Spanish

speakers in the community they serve) identified high priority language learning needs of a more specific nature. They wanted help with pronunciation, a repertoire of commonly asked questions to draw from during routine medical interviews, and a list of phrases to get patients to speak more slowly, explain, or repeat (Lear, 2005).

Three language training programs for healthcare professionals are described in the healthcare communication literature. The first classroom-based program employs an integrated skills reinforcement (ISR) approach to teach basic skills in writing, reading, listening, and speaking to foreign-born nurses with limited English proficiency working in the United States (Guttman, 2004). This program trains learners to develop bilingual vocabulary lists, answer comprehension questions on textbook readings, do library research, and engage in pre-writing and draft-writing activities. Oral skills are developed by getting the students to do oral presentations while a peer listens for and repeats any pronunciation and usage errors that need repair. Listening skills are taught by engaging learners in five-minute note-taking activities.

This classroom based approach to language training may indeed be effective at helping immigrant nurses to improve their general English proficiency, but since all speaking and listening tasks are essentially monologic in nature it is unlikely to prepare learners for the kind of face-to-face encounters that medical students and in-service healthcare professionals both say they want help with.

In addressing the specific language learning needs of healthcare professionals, an Ontario university program where trained actors engage pre-clinical medical students in a series of French language role plays (Drouin & Rivet, 2003) seems more appropriate. Each 10-15 minute simulation involves a medical interview and physical examination in

a controlled laboratory setting on a particular theme such as "a non-compliant patient" or "a depressed elderly patient" (2003, p. 601). The series of role plays are sequenced according to complexity, beginning with rapport-building tasks and continuing with history-taking tasks with actors who simulate being talkative, tearful, anxious, domineering, or presenting some other challenge to communication. Learners receive feedback from their clinical-supervisors through a standard assessment rubric as well as from the actor playing the simulated patient through a post-interview feedback session.

Provider-patient simulations of this sort are likely to be considerably more effective at preparing learners for the type of oral interactions medical professionals encounter on the job than the monologic oral presentation tasks described in Guttman. However, the cost of using trained actors, clinical-supervisors, a medical consultation laboratory, and the one-learner-at-a-time approach to instruction is likely to be too expensive for some medical schools and in-service language training programs. Also, busy healthcare professionals may not be available to study at the times those courses are offered. In such cases, asynchronous learning options will be necessary.

Responding to the need for low-cost, asynchronous second language training solution, Araiza, Klopf, and Kelly (2005) report on a website that was created to teach nursing students simple Spanish medical terms and phrases to use with Spanish-speaking patients in primary care settings in the United States. The website consists of Spanish language scripts of imagined encounters between a nurse and three patients: one adult coming in for a blood-pressure check-up, another adult presenting acute abdominal pain, and a child with a cold. Clicking on each phrase within the script with a computer mouse plays an audio file for the learner to listen to and repeat. Hovering on each phrase with

the mouse cursor reveals the English translation in the status bar at the bottom of the web-browser.

Though limited in scope and interactivity, the authors see this web-based language learning resource as an important step toward building Spanish language support for healthcare providers with Spanish speaking clientele. Indeed, its affordability, 24-hour accessibility, and focus on routine medical interviews should help. Future planned improvements such as database support, interactive quizzes and expanded use of multimedia will be built upon what has already been achieved. Nevertheless, the reliance upon a listen-and-repeat approach to oral skill development it employs is fundamentally unsound. Unless efforts are made to replace this strategy with opportunities for meaningful exchanges, initial enthusiasm will eventually give way to disappointment and finally general abandonment of the system.

In Search of a Sound Pairing of Good Pedagogy and Technology

Meaningless repetition of language forms as a means to acquire a language has been thoroughly discredited (Wong & VanPatten, 2003). Rather, all oral repetitions, it is believed, must occur in a genuinely communicative context where each formulation of a repeated structure is part of a meaningful message conveyed to a receptive interlocutor (Gatbonton & Segalowitz, 2005). Finding meaningful forms of oral interaction are essential for computer-assisted language learning (hereafter CALL) to be able to begin to respond to the language training needs of healthcare professionals in any significant way.

The goal of this research, therefore, will be to look to the literature on CALL for a sound pairing of technology and meaningful oral language practice that can be deployed to help solve the training issues surrounding the problem of healthcare access for

linguistic minorities. If an adequate pairing is available, the second goal of this research will be to look to the literature on healthcare communication for insights into the nature of patient-provider interactions in order to design and build a CALL system that will suit the second language learning needs of healthcare professionals. The third and fourth research goals will be to build a prototype system and test it for fitness of purpose, ease of operability, and effect on members of the target population of language learners.

CHAPTER 2: LITERATURE REVIEW

Technology and Language Learning

In order to achieve the first goal of this research effort, to find a sound pairing of technology and meaningful oral language practice, a search of the Linguistics and Language Behavior Abstracts (LLBA) and Google Scholar research databases using the keywords "voice recognition," or "speech recognition," and "language learning" was performed. A critical review of the literature on automatic speech recognition-enabled CALL followed to establish a core of representative literature on the subject.

Since the purpose of this critical review is to identify opportunities for meaningful human-machine oral interactions in language learning, figurative uses of the words "talk" and "listen" are employed throughout to represent the exchanges between a learner and a machine. It should be understood that the storage and retrieval of voice recordings and the rapid computational analysis of incoming electronic speech signals used to achieve full or partial speech recognition are technological processes employed to create a convincing illusion of a conversation. Machines only talk and listen in a metaphorical sense.

The Talking Machine

The promise of learning to speak another language with the help of a talking machine has been with us for a long time, but not without its disappointments. Almost a century ago, Charles C. Clarke wrote, "The talking machine in teaching foreign languages is by no means new. Many experiments have been made with it in schools and colleges, and the silent verdict brought by its general abandonment is that it is not worth the trouble it involves" (1918, p. 116). The talking machine at that time meant a wind-up

phonograph, which had been in use in the language classroom since about 1900 (Koekkoek, 1959). While Clarke saw its greatest potential in helping language teachers maintain the language skills they had acquired abroad, he lamented that, in the classroom, phonograph recordings offered only a metallic, nasal, and unnatural model of foreign speech for imitation and memorization.

Recording and playback quality have since improved, but getting the talking machine to listen has been much more of a challenge. Language labs of the 1960s and 1970s gave learners a chance to record themselves imitating models of native speech. However, without feedback, learners had to rely upon their own judgment to assess their performance in the target language. It was not until the early 1970s that a machine first started to listen to the language learner. At the University of Toronto, P. R. Léon and P. Martin (Léon & Martin, 1972) developed a computer system that employed pattern-recognition technology to automatically recognize, evaluate and graphically display a learner's intonation contours. The learner could then compare on a TV screen visual representations of his or her own speech contour next to the recorded model. Computer assisted pronunciation training (CAPT) thus introduced the means of recognizing how a learner said something, even if it didn't recognize what was said.

True speech recognition came a few years later with the advent of low-cost "off the shelf" component technologies. Harry Wohlert (1984) reports using an early speech recognition system to drill and test learners with German verb forms. In spite of the fact that the system was at times awkward and frustrating to use, Wohlert reports that students enjoyed speaking to a computer and "obtained a high level of achievement" (p. 84).

A quarter of a century later, "speech recognition is now a reliable and widespread technology" (Coleman, 2005, p. 158), but not necessarily in language learning. Numerous commercial speech recognition-enabled CALL products have recently come onto the market with the promise of helping students of all ages improve their reading, pronunciation, and speaking skills. Still, there is a reticence among language teachers and researchers toward this new technology. Leery of the costs and unsure of what speech recognition might do for students, some ask whether the new-and-improved talking machine is worth the trouble it involves yet.

What is Automatic Speech Recognition?

It should be noted here that although voice recognition has been used to refer to a sub-type of speech recognition in the literature (Coniam, 1999; Myers, 2000), for the purpose of this review voice recognition and speaker recognition are understood to refer to a technology that identifies the person speaking from a known population for security or forensic purposes (Zue & Cole, 2007). Automatic speech recognition (hereafter ASR), on the other hand, refers to the process of turning speech into text (Coleman, 2005; Zue & Cole, 2007), where "text" refers to a non-acoustic representation of the meaning of the speech contained in an acoustic signal. In order to understand just how text is extracted from speech, a short explanation of the technology involved and the problem of variability in speech is likely to be helpful.

Speech recognition is generally achieved using two principal system architectures: knowledge-based systems programmed with expert knowledge of the properties of a language; and pattern-recognition systems that use a pattern matching approach (Coleman, 2005). What follows is a brief summary. A fuller, more detailed

account of these two architectures can be found in John Coleman's book, *Introducing Speech and Language Processing* (2005).

A knowledge-based system follows 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 by comparing its derived parameters to an expert knowledge of parameter ranges for each phoneme in the language. For example, a speaker may pronounce the /r/ in a word like "run" in a variety of ways depending where he or she is from. Whichever way it is realized by the person speaking, the system must be able to recognize it as the equivalent of the phoneme /r/. Step two involves creating a hypothesis about the sequence of phonemes in a word by repeating this classification process for each successive interval within the signal. Keeping with the same example, /r/, /ʌ/, /n/ is hypothesized to be /rʌn/. In step three, the system's hypothesis about what was said is compared against the computer's lexicon to find the best match. Again using the word "run," the system will pass the string /rʌn/ through its internal dictionary. When a match is found, the computer has recognized one word (i.e., "run") in the signal.

A more commercially successful architecture uses a two-step pattern-matching approach. Rather than program expert knowledge of the internal structure of words and phonemes into a system, this approach simply compares the parameters derived from longer stretches of the speech signal to reference patterns stored in an internal database of examples until a match is found. Typically, these longer stretches of the speech signal are words rather than phonemes, although a variety of the pattern matching approaches has been used to treat short sentences as single words (Holland, Kaplan, & A. M. A. Sabol, 1999; Wohlert, 1984). To explain, instead of matching short intervals in the speech signal

to individual phonemes to produce a hypothesized string of phonemes to match to text representations of words, a pattern-matching approach will attempt to match the parameters of a longer word-length interval in the signal to word-length reference pattern entries in the internal dictionary. For example, the parameters of the spoken word /run/ are compared to reference-pattern listings to find the text version of "run." The same two-step process can be used for sentence-length speech signals where instead of assembling a sentence from individual words recognized in the speech signal, the entire signal can be compared to reference patterns of sentences. The pattern in "See Dick run" can be run against internal dictionary entries like "See Dick swim," "Watch Dick run," and "See Dick run" for the best match.

Variability Issues

Variations in the rate of delivery, a change of speaker, and different phonological environments make pattern matching difficult. When words are said at a rate faster or slower than the rate they were said when the reference database was made, the peaks and the troughs of the speech signal and the reference pattern may not line up perfectly. A technique called dynamic time warping compensates for this kind of variability by stretching or compressing the signal so that they do.

Variability among speakers is another problem. Accent, sex differences, vocal tract length, and breathiness, as well as individual differences in health all affect voice quality. For small vocabulary systems, statistical models of the normal range of variability are possible. For larger vocabulary systems, individual variability is dealt with by having the user train the recognizer by reading a script. This creates additional

reference patterns specific to the individual user. This need for extra training is what is meant in the literature by speaker-dependent speech-recognition.

Another problem occurs when speakers don't pause between words. In continuous speech, the onset of a new word will affect the way in which the coda of the preceding syllable is realized. The word 'ten' will be realized as [tɛm] in "ten pence" but [tɛŋ] in "ten cars" (Coleman, 2005, p. 182). Coping with this type of phonological environment effect in connected speech demands either debugging the recognizer with a rule-based system or creating reference patterns for all the possible environmental effects on the realization of every word in the recognizer's lexicon. This will either be very time consuming for the programmer doing the debugging or time consuming for the user who has to train the recognizer by reading long texts with occurrences of every possible phonological environment.

The Challenges of Recognizing Second Language Speech

The challenges to accurate speech recognition are further compounded when users speak in a second language since general purpose recognizers are not designed to handle the specific "errors and disfluencies characteristic of language learners" (LaRocca, Morgan, & Bellinger, 1999, p. 302). For instance, when a general-purpose pattern-matching discrete-speech recognizer (such as the one described in Holland et al., 1999) recognizes speech, it does so one word at a time requiring pauses between words. The only way it can recognize continuous speech is when it is tricked into treating an entire sentence as if it were just a very long word.

Not surprisingly, sentence-as-word recognition strategies will not work when learners use unexpected word orders. ASR-CALL application developers employ two

basic strategies to handle learner-specific errors of this sort. Some systems forestall errors by providing on-screen text-prompts that list only the sentences that the system is programmed to recognize (Eskenazi, 1999; Holland et al., 1999; Harless, Zier, & Duncan, 1999; Rypa & P. Price, 1999). The second strategy involves programming the system's recognition grammar files to include additional sentences that achieve their communicative intent despite containing common learner-specific grammar errors (Bernstein, Najmi, & Ehsani, 1999; Morton & Jack, 2005) .

Pronunciation errors, on the other hand, can be tackled in ASR-CALL by simply lowering the recognizer's acceptance threshold to tolerate less perfect matches between the learner's speech and the expected utterance. Wolhert, the first advises lowering the threshold for beginners but cautions against setting the threshold too low as it leads to misrecognitions, something "students react very negatively to" (1984, p. 83). His students did not complain, however, about the system rejecting their pronunciation too often.

Obviously, another way to improve recognition rates of second language speech is to use a better recognizer. Bernstein, Najmi, and Ehsani claim that Entropic's Hidden Markov Model Toolkit allowed them to develop a system that supports speaker-independent, continuous speech recognition for both native-Japanese and non-native-Japanese speakers. They report that their Subarashii system is not only better at avoiding misrecognitions, but also avoids unwarranted rejections due to insignificant differences in pronunciation.

Nevertheless, Wachowicz and Scott (1999) argue that even human teachers misunderstand students at times, and so developers of ASR-CALL applications must not expect 100% recognition accuracy all of the time. Instead, planning for misunderstanding

through the judicious use of "verification procedures and repair strategies" by getting the application to ask the learner "Did you say X or Y?" (1999, p. 272) or providing visual or textual feedback on what it believes it has heard will make for more robust ASR-CALL systems.

What can ASR do for CALL?

In spite of lingering technological constraints upon recognizing speech in general, and second language speech in particular, a range of ASR enabled applications is now available to language learners, teachers, and researchers. Turning the learner's speech into text through an intermediary step of looking at its sound and/or prosody characteristics makes ASR particularly well-suited for computer assisted pronunciation training (CAPT) as well as form-focused feedback and certain types of oral assessment. Also, converting speech to text allows the learner to put the mouse and keyboard aside in favour of a microphone, turning previously receptive or silent form-focused tasks into speaking tasks. Listening for meaningful text, however, allows for a new type of oral language practice not previously open to language learners before the advent of ASR. By linking recognized speech to audio, video, or animation clips, ASR can create a convincing illusion of communication with a virtual interlocutor. While still preliminary and exploratory in nature, recent research into these new directions of ASR for CALL reveals a promising new source of interaction and learner motivation.

ASR Used in Second Language Learning

Before looking at the most innovative and promising uses of ASR in CALL, it will be instructive to look at the range of uses for ASR that have emerged over the years. These earlier uses of ASR show just how far human-computer interaction has come.

Though in hindsight they seem dated and rely upon questionable pedagogy, earlier ASR-task types will be of interest to future materials developers who will look to them as a source of possible ASR tasks types to embed within the framework of virtual interaction. As such, the future of ASR-enabled human-computer interactions for CALL will depend upon a critical review and understanding of what has gone before.

ASR for Drilling and Repeating

Since all speech-to-text applications go through the intermediary stage of matching properties of the speech signal to either stored patterns or sequences of phonemes, saying things clearly and correctly into the microphone is a precondition for getting the machine to recognize speech with confidence. Non-recognition (other than the kind caused by a problem with the microphone or microphone placement) can indicate problems with the learner's pronunciation, morphology, word order, or word choice. Moreover, the closeness of the match between what was said and what was expected can provide a measure of goodness of pronunciation. Recognition confidence scores are thus used to manage and score drill-and-fill, listen-and-repeat, and read-aloud practice activities in ASR for CALL.

Wohlert's system in the 1980s uses this basic strategy with extensive accuracy-oriented drill-and-fill exercises. The learner sitting at the computer would hear a recorded audio prompt of a German verb form played on a cassette player and then was presented with a series of questions on the computer screen. Each question was designed to elicit a correct pronunciation of the prompt. Non-recognition triggered a "try again" response from the system until the learner said what the system expected to hear. Manipulating the

acceptance threshold, Wohlert required a more exact match between what the learners said and what the system expected for intermediate level students than for beginners.

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 on a television, ASR-enabled applications can also provide graphical feedback on goodness of pronunciation. Wohlert's original ASR system played a recorded prompt and then listened for the learner to repeat the word or sentence. The template of the recording was then compared to the learner's attempt, and an evaluative score was generated based on how similar the two speech samples were to each other. The result was visually presented in the form of a bar graph. Newer applications use a greater variety of wave forms, graph types, accuracy meters (see Wildner, 2002), and game-like visual representations for kids. Dalby and Kewly-Port (1999), for instance, report on one system for young learners that displays the evaluative pronunciation score graphically represented as bowling pins knocked down in a bowling alley. Better pronunciations knock down more pins, worse fewer.

An ASR-enabled dedicated reading tutor described in Mostow and Aist (1999) illustrates how recognizability can be used to help children with their reading fluency. Displaying a sentence on the screen for the young learner to read, the computer application listens for each word to be read aloud. If the learner gets stuck or mispronounces a word, the application interrupts the learner with a cough, or other unobtrusive sound, and by underlining the misread word. If the learner misreads the item again, the application plays an audio prompt.

Dedicated ASR-CALL read-aloud applications might not be necessary for all learners, though. Meyers (2000) reports that, for second language learners, the initial training session for standard speaker dependent dictation software can be used to stabilize pronunciation errors since excessive variability in the pronunciation of a word will trigger a request to the speaker to say the word again. For this reason, second language learners will likely find using this type of software particularly motivating as the implicit goal of the training is to get the software to recognize every word uttered.

David Coniam's exploratory study (1998) of misrecognitions on a read-aloud activity demonstrated implications for computer-based oral testing. Asking ESL learners to read a text aloud to a standard dictation software, he found that the differences between what the software recognized and the script the learners were asked to read, validated against manual transcriptions, revealed that "reading aloud is a good indicator of overall ability in English" (1998, p. 20).

ASR for Selecting, Directing and Sequencing

Using a microphone in the place of a computer mouse or keyboard can be helpful or even essential for computer users with limited vision or mobility, but for second language learners, using one's voice to select, arrange and compose words and sentences on a computer can transform traditional CALL tasks into speaking tasks. LaRocca, Morgan, and Bellinger (1999) use a multiple-choice selection task to focus 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. Wachowicz and Scott (1999) provide another example of a selection task performed in the spoken modality where the learner is asked to choose 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.

An example of a multimodal arrangement task can also be found in Wachowicz and Scott (1999). 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 says in the order said, and sequences the words on the screen to unscramble the sentence. This same task could be done with a mouse or the keyboard, but employing ASR creates a little more oral practice possible for the learner.

Speech recognition furnishes the means to turn writing composition tasks into speaking tasks. Dictation software, little more than an ASR-enabled word processing program, makes this possible. Myers (2000) lists a number of advantages to composing orally with speech-to-text software. Dictation allows learners to plan what to say during pauses, to receive visual support for utterances, to develop monitoring abilities, to gain awareness of their articulation, to improve their pronunciation, and to promote the transfer of oral skills to written production.

Meaning-focused and Purpose-oriented ASR

ASR has also made meaning-focused interactive speaking activities possible. Two types of interaction between the learner and an intelligent virtual agent are apparent in the literature. The first involves command and control activities where the learner speaks commands to a virtual agent who obediently does what is asked within a three

dimensional virtual world (Holland et al., 1999; Kaplan, M. A. Sabol, Wisher, & Seidel, 1998; Morton & Jack, 2005; Wachowicz & Scott, 1999). Using the software called TraciTalk, the learner engages 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 absentminded personality so that the technological shortcomings of the software have a less disturbing, human quality to them.

Holland et al. describe an application created for soldiers learning Arabic that works on the same principle of command and control described above. Three commands are presented at a time for the learner to say to advance the scenario, such as "Walk to the file cabinet," "Open the drawer," and "Turn on the radio" (1999, p. 344). According to the authors, three pedagogical principles informed this design: implicit feedback, which is available to the learner when his or her utterance is understood and the virtual agent does what is intended; over-learning, "repeating training beyond the point of apparent mastery" (1999, p. 341) to promote automaticity and retention during periods of non-use; and 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 authentic listening texts, 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 et al. (1999, p. 318), learners 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.

Two designs are possible (Ehsani & Knodt, 1998). A closed response design lists only the responses that the learner can say on the screen. The learner's lines are completely scripted, and the task of the learner is simply to select from among the available choices and pronounce them correctly for the recognizer. Alternately, the open response design (also programmed to accept a finite number of responses) does not display which responses are available to the student but instead leaves the learner to work out what to say through trial and error. In both cases, the recognizer will be able to recognize only what it has been pre-programmed to recognize. Consequently, novel and ungrammatical utterances will be treated in exactly the same way as nonsense utterances. To limit the number of possible utterances that the learner might produce with an open response design to a manageable number, LaRocca et al. suggest tightly controlling the types of contexts in which the learner can speak to limit the range of possible utterances that could occur.

This type of simulated face-to-face conversational CALL activity was first attempted without speech recognition using videotape and later videodisc systems for language learning in the 1980s. The Autotutor videotape system was, however, embarrassingly slow, requiring the learner to answer open-ended questions using the keyboard and then wait as the videotape spooled forward or backward to the next video clip (Little & Davis, 1986). Videodisc systems described in Schulz (1988) such as VELVET or *Avec Plaisir* were much faster and more sophisticated, allowing instant feedback, story branching, translations, vocabulary glosses, graphics, and access to

electronic reference materials, but were prohibitively expensive, with each disc costing $1,800 and the hardware costing as much as $7,000 per console. Even then, while the learner could recite one of the listed answer options to questions on the screen by speaking into a microphone, the system was not able to recognize the learner's spoken response but waited instead for the corresponding keystroke to advance to the next videoclip.

Virtual Dialogues

So it is the ASR-enabled virtual dialogue system described in Harless, Zier, and Duncan's article *Virtual Dialogues with Native Speakers* (1999) that represents the first meaningful and purposeful two-way spoken conversation with a non-human for language learning. The article details a series of studies performed on a prototype speech activated multimedia system called Conversim™ from Interactive Drama Inc. for instructing U.S. soldiers in Arabic dialects. The system was created to address the problem of proficiency loss in military linguists who, after completing a 63-week Arabic course at the Defense Language Institute (DLI), are required to maintain individual readiness despite a lack of opportunity to practice. In the light of the century of effort to produce a talking machine that is worth the trouble it involves, the development of this ASR-enabled virtual dialogue system represents a long awaited triumph.

In developing the Arabic series, four DLI instructors were recruited to play four Iraqi characters on the virtual dialogue system. Presented as a series of video clips, each character reveals under interrogation that a judgement must be made about him: the prisoner of war is either friend or foe; the defecting air force pilot is either a legitimate

defector or a spy; the civilian émigré looking for asylum could be lying; and the educated Kurdish refugee might be helpful in managing the refugee camp.

Working with a CD-ROM on a laptop with a microphone and headset, the student chooses from a limited set of three question-prompts at the bottom of the screen to interrogate each character. 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. The video footage for each of the characters is extensive enough to allow for more than just military questions. The students can ask about fears, wishes, concerns, and background to discover the character's personality and culture.

In anticipation of each of the four scenarios, a virtual instructor prepares the student with pronunciation help for the phrases and questions he or she will need to conduct the interrogation. During this preparatory stage of the program, the student can record his or her own voice and compare the recording with the instructor's pronunciation, or see a display of the speech recognition software's confidence in the recorded utterance.

Toward a Talking Machine that is Worth the Trouble

Every powerful new technology introduced into language learning since Clarke (1918) early in the last century 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, a disappointment takes hold, leading to general abandonment. Is it any wonder when imitation and memorization drills remain at the heart of the language learning pedagogy? This pattern can be expected to continue to repeat itself with ASR-CALL unless the

exchange of meaningful messages is made the central purpose. Accuracy and fluency have been the earliest targets of ASR-CALL applications, 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 his or her taste. Nevertheless, the most exciting developments in ASR-CALL have targeted meaning-driven and purpose-oriented oral interaction. 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 tasks where the costs and technological limitations remain well below the not-worth-the-trouble threshold for learners and teachers.

While the virtual dialogue system described above is well suited to the needs of military linguists, the task of prisoner interrogation is unlikely to provide healthcare professionals with the kind of language practice they will need for medical communication. Instead, a tailor-made virtual dialogue system will need to be designed and developed around a high-priority, purpose-driven conversational task that healthcare professionals do routinely with linguistic minority patients. With this major goal in mind, the next step, therefore, will be to look to the literature on healthcare communication for insights into the nature of patient-provider interactions in order design and build a CALL system that will suit the second language learning needs of healthcare professionals.

Medical Communication

In order to gain insight into the range and nature of medical communication tasks to inform the design of a virtual dialogue system for healthcare professionals, a search of PubMed and Google Scholar research databases using the keywords "medical interview," "second language," and "nurse-patient communication" was performed. A collection of literature on first and second language communication of a variety of health professionals resulted. The review of this literature is divided into two discrete parts. The first part will explore the general characteristics of the medical interview to establish a global understanding of the structure and nature of medical consultations. The second part will select and describe a specific medical interview task to model a virtual dialogue upon.

Medical Interviews

Three essential communication goals are implicit to every medical interview (Bickley , 2007): the first is to establish a trusting and supportive relationship between healthcare provider and patient, the second is to gather information about the patient, and the third is to share information with the patient. These goals are achieved through the pursuit of seven successive medical communication tasks (Makoul, 2001). They are as follows: (1) open the discussion, (2) build the relationship, (3) gather information, (4) understand the patient's perspective, (5) share information, (6) reach agreement on problems and plans, (7) and provide closure.

"The medical interview is the most common and critical procedure that physicians perform" (Barrier, Li, & Jensen, 2003, p. 214). Its general structure can be viewed as a series of six distinct episodes of communication. The first episode occurs as the doctor or nurse enters the examination room where the patient is waiting. The consultation begins

with the name exchange for first time visits along with pleasantries and small talk. Small talk has been found to be important in the development of rapport between healthcare provider and patient and has been shown to help providers to gather detailed health information (Aranguri et al., 2006). It is here in this portion of the medical interview that the goal of establishing a supportive and trusting relationship is first pursued.

Next comes the solicitation of the reason for the visit. In follow-up visits with a doctor, the doctor may simply ask the patient to confirm the reason for the follow up with a question such as "I asked you back here..." A yes/no question format is preferred in such cases to demonstrate to the patient that the doctor remembers the patient and that it is not necessary to begin *de novo*. Similarly, the doctor may have obtained the reason for the visit from the file or nurse. In such cases, the doctor may simply invite the patient to confirm the reason for the visit by answering in the affirmative to a yes/no question. The danger in such restrictive questioning strategies is that the patient may have another complaint that is the real motivation for the visit. Open-ended questions "How can I help?" are therefore recommended in place of closed-ended questions since the problem presentations they elicit tend to contain more discrete symptoms though more than twice as long (27.1 s versus 11.3 s on average) as those generated by closed-ended questions (Robinson & Heritage, 2006, p. 283). This episode establishes the goal of gathering medical information about the patient and helps to narrow the focus of the information gathering task.

Once the purpose of the visit has been established and the chief complaint has been identified, the physician may begin asking diagnostic questions signalling the third episode of the communication in the medical interview. This will take the form of an

illness history in addition to a comprehensive medical history if there is time or one is warranted. Busy practicing physicians tend not to have the time to take a full medical history, and so may either get a nurse to take it prior to the consultation or ask a more limited set of diagnostic questions as they test a diagnostic hypothesis.

After the history has been taken from the patient, the doctor may perform a physical examination looking to confirm his hypothesis and explaining all the while what he or she is doing. This constitutes the fourth of the six episodes of communication. If the physician feels confident in his or her original hypothesis about the cause of the patient's ailment or has developed a different hypothesis, the physician will be ready to make a diagnosis and issue a prescription, signalling the fifth episode of communication in medical interview. Here the goal of the interview is to share information effectively with the patient. It is important therefore to note that "patients are more likely to take medication effectively if they have been involved in discussions about their treatment options, and understand and support the decision about what is prescribed" (Drew, Chatwin, & Collins, 2001, p. 58).

Leave taking, signalling the end of the final episode, will follow arrangements for a follow-up visit and tests if either is determined to be necessary. The provider may encounter a door handle question at this point, a question from the patient signalling the real purpose of the visit, which will require the doctor to start the consultation from the beginning. In this way, door handle questions have their origins at the beginning of the interview (Baker, O'Connell, & Platt, 2005) when the chief complaint was not properly elicited from the patient at the beginning of the interview, or when the patient was

interrupted before completing his or her illness narrative. A schematic structure of a medical interview is presented in Figure 1.

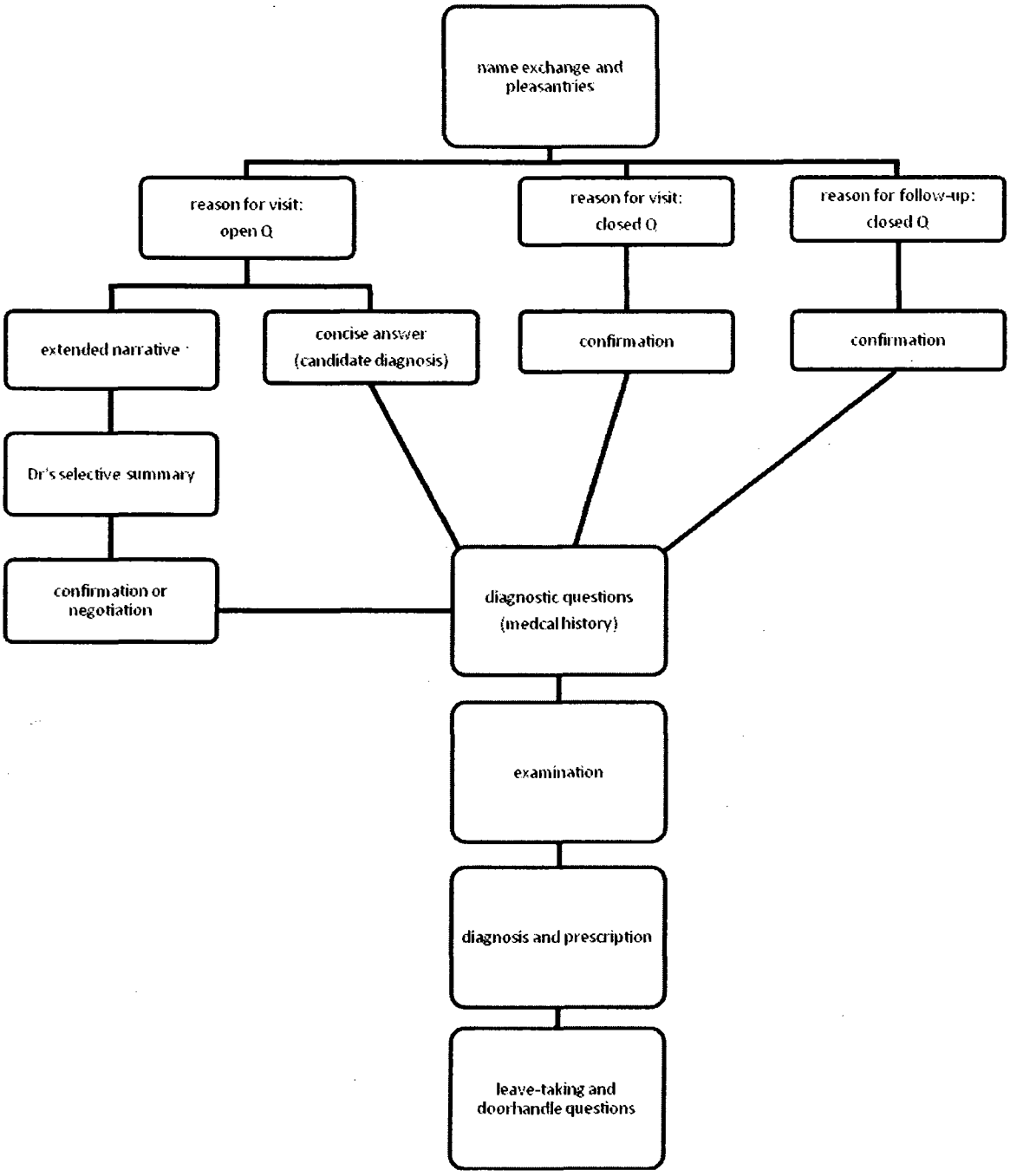


Figure 1. Structure of the primary care interview.

Nurse versus Doctor Patterns of Communication

Nurses, in particular, engage in a variety of communicative exchanges with patients. For example, nurses often serve as intermediary between doctors and patients (Hadlow & Pitts, 1991), as interpreter of medical language for patients (Bourhis, Roth, & MacQueen, 1989), as information provider (Baggens, 2001), as facilitator engaged in listening, supporting, and mobilizing hospital staff and resources for patients (Abbot et al., 2006), and confidant helping terminally ill patients come to terms with death (May, 1995). In fact, according to Price (2004), the six most likely medical interviews that nurses are likely to perform are as follows: (1) taking a health history and understanding a patient situation, (2) explaining diagnostic, investigative or staging measures, (3) dealing with diagnosis, (4) planning care and treatments and understanding options, (5) negotiating rehabilitation and patient education, (6) and planning a follow up.

According to Collins (2005), nurses' communication patterns with patients differ from doctors' communication patterns in that they are oriented toward the patient's personal responsibility and behaviour, whereas doctors' explanations are oriented toward biomedical intervention. As such, their consultations provide different opportunities for patients' involvement. Patients tend to reveal more to nurses than to doctors, finding them more approachable. Nevertheless, doctors tend to treat nurse consultations as supplementary and subordinate to their own rather than as complementary opportunities to promote patient understanding.

Doctors' attitudes toward the value of nurse consultations aside, nurses are being given increasing amounts of clinical work once reserved only for doctors (Charles-Jones, Latimer, & May, 2003). The division of labour between the two groups is becoming more

complex than it used to be. Patients presenting high priority medical problems are given appointments with general practitioner doctors, whereas low priority medical cases are given to nurses or to no one according to a hierarchy of medical expertise, which puts doctors at the top with the most expertise and patients at the bottom with the least. Nurses take up the middle ground between the two extremes. Chest infections are given to doctors because nurses are not generally trained in the use of a stethoscope for diagnosis, but nurses will see patients with sore throats, and small skin rashes. Patients presenting with common colds are sent home without seeing anyone.

Far from being a monolithic class of medical professionals, nursing expertise will place individual nurses higher or lower on this hierarchy according to nursing grade. Apart from de-emphasizing the patient as a person by categorizing him or her as being of minor or major importance according to ailment (Charles-Jones et al., 2003), an unintended consequence of this redistribution of work along biomedical lines is that observed differences in nurse-patient and doctor-patient patterns of communication may be disappearing.

Medical History-Taking

Since this research is primarily a first-effort, prototype-feasibility study exploring what is possible in the development of pedagogically sound second language oral skill training for healthcare professionals, a certain narrowness of purpose is required. Owing to its universality, central importance, and linguistic complexity, the comprehensive medical history interview is the communicative task chosen for the virtual dialogue design. What follows is therefore a detailed review of literature relating to the medical history interview. Being less likely to have the extensive training that doctors have and

being called upon more often to perform a greater number of clinical tasks (Charles-Jones et al., 2003), nurses were selected as the principal target learner population for this materials development effort. It should be noted here that some of the literature discussed below describes how doctors perform a medical history interview, but it is hoped that the findings will be broad enough to generalize to nurses also.

Medical history-taking is the most important diagnostic tool at the disposal of health practitioner. One study (Hampton, Harrison, Mitchell, Pritchard & Seymour, 1975) found that doctors were able to correctly diagnose the patient's medical condition in over 82% of cases from the medical history and referral letter alone. In the remaining 18% of cases where doctors had to revise their post-medical history hypotheses, 9% of revisions were made due to the results of physical examinations of the patients and another 9% were due to laboratory test results.

Time consuming and linguistically demanding in nature, medical histories can be taken by a computer (Bachman, 2003; Dugaw Jr., Civello, Chuinard, & G. N. Jones, 2000), a nurse (B. Price, 2004; Otto, 1999; Cameron & Williams, 1997; Drew et al., 2001; Sherman & Fields, 1988), a physician (Haidet & Paterniti, 2003; Maguire & Rutter, 1976; Paul, Dawson, Lanphear, & Cheema, 1998; Stivers & Heritage, 2001) or a pharmacist (Chaikoolvatana & Goodyer, 2003), medical and nursing students (Sherman & Fields, 1988) and by the patients themselves using a questionnaire (Bachman, 2003).

Patient-centered history-taking skills, while initially easy to acquire for first year medical students, become more of a challenge for fourth year medical students and new doctors. One study of first year medical students practicing their history taking interview skills on volunteer patients found that the students were well received, generating no

negative assessments by any of the seven patients (Thomas, Hafler, & Woo, 1999). However, this initial success at history taking tends not to last. An assessment of 292 University of Connecticut medical students found a decline in medical history-taking skills by the fourth year of medical school. Probable reasons given for this decline were a culture of medicine that de-emphasizes the need for interpersonal skills, residents acting as negative role models in this regard, and students learning to view patients as an annoyance to be avoided (Pfeiffer, Madray, Ardolino, & Willms, 1998). Further deficiencies in providers' history-taking skills identified in the literature include the inability to keep patients to the point and to clarify the real nature of patients' complaints, a reluctance to ask about relevant psychological and social aspects of their histories, and failure to pick up important verbal and non-verbal cues (Maguire & Rutter, 1976). In a study of medical interns' history-taking skills in Bangladesh, it was found that although the interns initially asked patients their names they failed to use them during the history-taking portion of medical consultations 88% of the time. In follow-up interviews, 86% of patients whose doctor did not use their name said that it would have made them feel better about their visit if he did (Rahman, 2000).

Nursing Assessment

Particular to nurses, the initial nursing assessment that professional nurses or nursing students perform begins with a medical history interview (Sherman & Fields, 1988). The medical history will usually begin with an elicitation of the chief complaint, the specific reason the patient has come seeking professional medical attention and a history of the present illness motivating the chief complaint. This will take the form of an open question such as "How may I help you?" (1988, p. 47). However, often a patient

will come in for a routine check-up or pre-employment examination, in which case the chief complaint and present illness will be absent.

Where the patient comes with a complaint, the history of the present illness is taken by establishing the onset of the problem, interval history, current status, and present reason for seeking help. Pain is assessed by asking the patient where it is located, what it feels like and its severity, when it occurs and how long it lasts, what triggers it, and how it affects the patient through associated symptoms.

The next stage in the nursing assessment is to determine the patient's past medical history. Established diagnoses are recorded with dates, severity and complications, along with prior surgeries, injuries, allergies immunizations, and current medications. Sherman and Fields (1988, p. 52) recommend phrasing the elicitation of these details using the prompt, "tell me about all other illnesses and operations you have had".

Following upon the patient's past medical history is the family history. The family history can provide insights into hereditary diseases, reactions to illness and death, and exposure to infectious diseases. The nurse will ask about diseases that run in families such as heart disease, high-blood pressure, diabetes, cancer, allergies, and mental disorders. The construction of a full genogram may be necessary in some cases, which will require as much as an hour to complete (Sherman & Fields, 1988).

Still within the framework of the medical history interview, the next step is to determine what kind of patient has the disease. To find this out will require a personal and sociocultural history of the patient. This part of the medical history is the best place for questions of personal and private nature since success will depend upon the prior development of good rapport.

For adult patients, a review of systems follows. "There is no practical limit to the number of questions that might be asked" (Sherman & Fields, 1988, p. 62) and the questioning proceeds from general to specific and in anatomical order starting at the head and working downward. Systems for review include general systems, skin, head, eyes, nose, mouth, throat, breast, respiratory, cardiovascular, gastronintestinal, genitourinary, gynaecologic, obstetric, musculoskeletal, neuropsychiatric, lymphatic and hematologic, and endocrine systems.

The final part of the nursing medical history interview deals with activities of daily living. Questions will tend to be about occupation and education, recreation, diet, habits, sleep, marital and sexual history, relationships, and self-evaluation.

Modelling the Medical History Interview for a Virtual Dialogue System

The medical history interview is especially well suited for modelling as a virtual dialogue because of its goal-orientation, the non-occurrence of answer-assessments, and the predictable order and routine nature of the questions asked. A necessary feature of interviews with a clear biomedical goal 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 upon 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 healthcare professional using the system. In other words, the learner is unlikely to find 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 the healthcare

provider take control of the conversation in order to achieve the goal of collecting enough pertinent health information to solve the patient's health problem.

Furthermore, since another feature of the biomedical style of interviewing is that providers typically do not react to patients' newsworthy answers, the non-occurrence of answer-assessments mean that awkward silences following a patient's response to a medical question are commonplace in medical interviews (C. M. Jones, 2001). This makes the task of modelling the learner-user's side of a virtual dialogue based on a medical history interview simpler than one based on everyday conversations since expressions of surprise, sympathy, agreement, or affiliation are usually missing from the provider side of provider-patient exchanges. A healthcare provider accustomed to the biomedical style of interviewing patients (as opposed to a biopsychosocial interviewing style) will not find it unnatural to move on to the next question without commenting on the patient's answer to the last question.

Finally, the predictable order and routine nature of the questions asked in a medical history interview not only makes it a relatively simple matter to script the learner-user's lines for the virtual dialogue, but knowing what the learner-user is likely to say beforehand makes the medical history interview particularly suitable for an open-response virtual dialogue design. By concealing the question prompts from the learner-user, it becomes possible to use the virtual medical history dialogue for assessment purposes where the learner-user is tasked with correctly formulating and sequencing medical history questions in English—perhaps after a series of preparatory closed-response dialogues with different patients—in order to take the patient's medical history. This high degree of predictability of the content and form of a medical interviewer's

questions is perhaps unique to a comprehensive medical history interview, making it uniquely well suited for a virtual dialogue.

A Prototype for Feasibility Testing

For the first prototype, a simple virtual dialogue system for medical history interview training in a second language is envisioned, one that is both affordable and feasible. A closed-response design with a linear question-answer sequence—but without small talk, chief complaint or present illness—should be enough to convey to the learner-user the essential range, purpose, and nature of the medical history interview in English. Personal data, medical history, family history, a review of systems and daily living questions represent the core of the medical history interview whether embedded in a primary care doctor's consultation, nursing assessment, or medical and nursing school training activity. As such, these core elements will provide a sound basis for a virtual dialogue prototype-feasibility study of this scope.

The following chapter will outline how such a prototype system will be built and tested. Steps involved in the creation of the system will be described in detail, and the feasibility testing procedure with nursing students will be explained.

CHAPTER 3: METHODOLOGY

Phase 1: Building a Prototype Virtual Dialogue

This two-phase study began with the creation of a virtual dialogue (in phase one), dubbed The Virtual Language Patient (VLP) and followed with its assessment as a worthwhile pedagogical intervention for healthcare professionals learning a second language (in phase two). Phase one, the creation of the medical history interview virtual dialogue prototype, proceeded in four discrete steps. The first step was to create a questionnaire task that would motivate the learner to take a virtual patient's medial history. This was accomplished by performing a Google search for available medical history questionnaires and synthesizing a new, shorter questionnaire from common features. The second step was to video-record a person answering questions derived from items on the synthesized form. The third step involved editing, compressing, and sequencing the video clips to prepare them for use in the creation of an HTML mock-up version of the virtual dialogue, where advancing through the conversation will be achieved by clicking on hyperlinked questions rather than uttering them into a microphone. The fourth and final step was to hire a programmer to build the virtual dialogue using Microsoft Visual Studio and SRI's EduSpeak speech recognition engine.

Creating the VLP Questionnaire

To create a questionnaire task-sheet for the learner to fill-in during the virtual dialogue, a web search was performed using Google. Two questionnaires were selected for synthesis from the search results. The first example is a standard employment medical history questionnaire (Scripps Institution of Oceanography, 2004) used during the application process for a job, and the second is a longer questionnaire used at the

University of Florida Shands Executive Health Center (Flint, 2003) in health assessment. 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 onto a single page (see Appendix A).

Creating the Virtual Dialogue Question Script

Once this VLP questionnaire was ready, questions were created from questionnaire items (using native-speaker intuition) for use during the filming of the virtual dialogue. See Appendix B for a complete list of questions. Two preliminary questions ("Are you here for the medical history interview?" and "Can I ask you a few questions?") were added to the beginning of the interview question script to indicate the purpose and nature of the VLP, making a total of 70 questions. Rapport-building questions and small-talk were not planned for this prototype virtual dialogue.

Following the two introductory questions, the next thirteen questions elicit person data with questions such as "What is your name?"; "How do you spell your last name?"; "How old are you?" and "Are you married?" etc. The next 42 questions elicit information pertaining to a patient's personal and family medical history, and review of systems, for example "Have you ever had a hernia?" "Has anyone in your family ever had heart disease?" and "Do you get frequent headaches?" etc. The next eight relate to daily living questions about tobacco, drug and alcohol use, and type of employment such as "Do you smoke?"; "Do you drink alcohol?" and "How many hours do you work in a week?" etc.

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 in the creation of the interview questions for the virtual dialogue script to maximise the learner's exposure to formulaic aspects of English question grammar. Through the elimination of sequential parasitism, non-native healthcare professionals practicing their English with the system will encounter questions containing *do you* 23 times, *are you* and *have you* eight times each, *is there* five times, and *did you* four times. Encounters with *what*, *how*, and *when* questions occur eleven, nine and four 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 question grammar accuracy.

To mark the end of the interview script, a final "Thanks" was added as a leave-taking salutation to complete the series of 71 interviewer utterances.

Creating the Virtual Dialogue

Once both the VLP questionnaire and virtual dialogue question script were ready, arrangements for filming began. A diabetic man in his early forties willing to answer questions about his own medical history was hired, and over the course of five days in June 2007, video recordings of him giving unscripted answers to the prepared medical history plus a variety of clarification requests (*What?; Could you speak up?; What did you say? I didn't get that.*) were made in his home using a Sony DCR-DVD403 Handycam on a tripod. The digital video clips were edited, compressed, sequenced and embedded individually into html pages hyperlinked to each other. In this way, an html mock-up was created to illustrate to the programmer how the ASR-enabled version of the virtual dialogue might look. A female native-English speaker from Toronto, Ontario, was

recorded saying each of the scripted questions using a Sony ICD-P210 digital voice recorder to be used as pronunciation models.

Using SRI's EduSpeak speech recognition engine (Franco et al., 2000) and Microsoft's Visual Studio 2005, a programmer was hired to work on the graphical user interface, defining the grammar template, coding instructions for accessing the recognition engine, managing the video and audio files, making provisions for saving user audio files, logging data on speaking fluency, and coding for the random sequencing of non-recognition responses. In February 2008, the prototype was finished. The result was dubbed the "Virtual Language Patient" (see Walker, Cedergren, Trofimovich, Gatbonton, & Mikhail, 2008).

Features and Content of the VLP

The VLP system runs on a PC computer with Windows XP 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. Anybody using it for the first time will quickly be able to understand how it works.

Upon launching the software, a video image of "Danny," a 40-year-old male patient, appears at the centre of the screen (see Figure 2). Danny 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. Just 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 the "Recognize" button and pronouncing the sentence into the microphone. If the system recognizes the learner's utterance as being similar enough

to the expected sentence, a video plays Danny's response, *Yup*, and the second question appears on the screen.

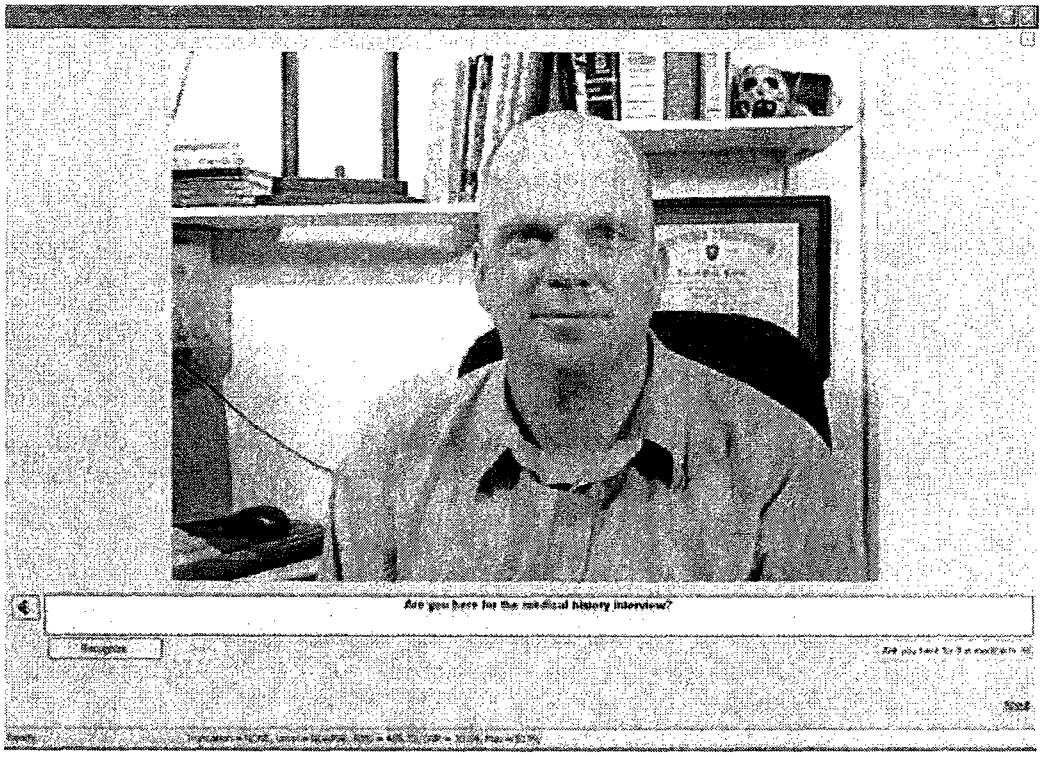


Figure 2. The graphical user interface.

At the same time, a feedback panel is displayed after each successful recognition, providing feedback on confidence ratings associated with each word and the whole utterance (as seen in the top right corner of Figure 3). Ratings for words that fall below a threshold are displayed in red, otherwise in green. The learner can thus get a sense of which words he or she needs to say more clearly. When the entire sentence does not meet the predetermined threshold of what is acceptable due to either poor microphone placement or errors in pronunciation, a video with a clarification request (i.e. *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 a button to the left of the question prompt (identifiable by its small speaker icon) and then listening to the recording through the headphones.

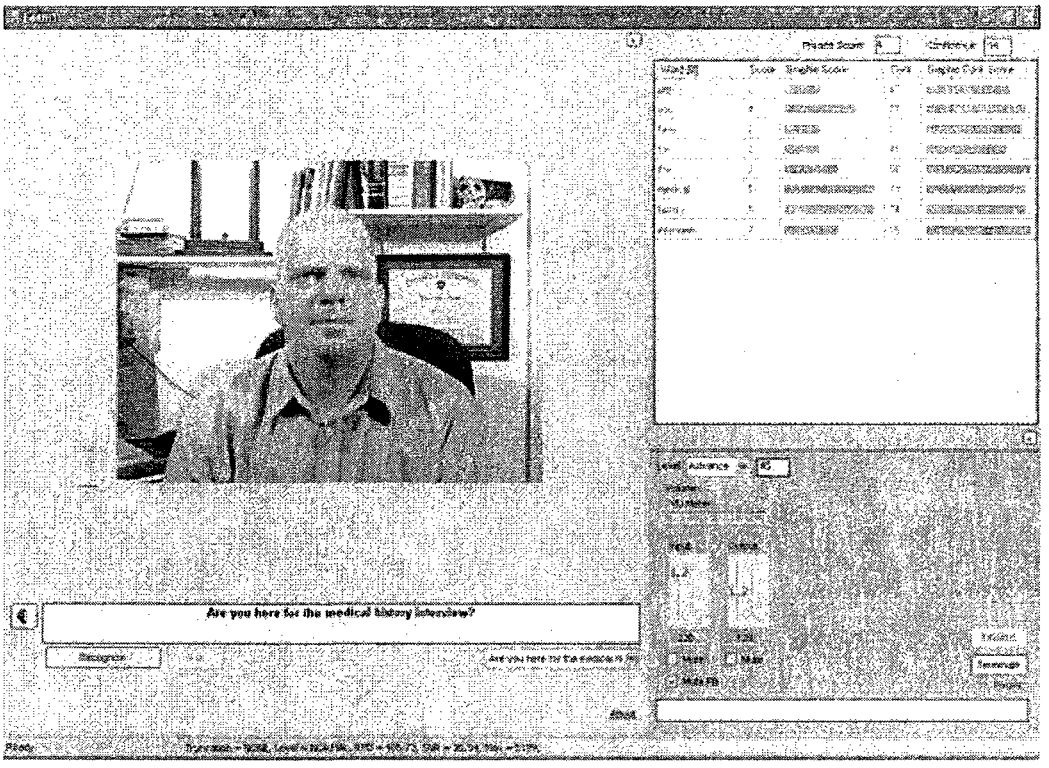


Figure 3. The feedback panel.

Some adjustments to the system's speech recognizer are available to the learner by using a settings panel at the lower right of the screen (again shown in Figure 3). Using the mouse, the learner can change the microphone sensitivity, headset volume, and recognition acceptance threshold. The advantage of being able to set the acceptance threshold to a lower or higher level is that the learner can make Danny more or less forgiving of pronunciation errors and thus make the pronunciation demands of the experience less frustrating or more challenging according to the learner's individual needs. Three acceptance threshold presets are available: beginner (35%), intermediate (45%), and advanced (55%).

This is important because getting Danny to recognize the question is necessary in order 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 (see Appendix A) the learner is prompted to ask each of the 71 virtual dialogue interview questions (Appendix B) in a fixed sequence. Each of the videos contains a video recording of Danny's authentic answer to each question.

Danny is not a professional actor. All of his answers are authentic responses to the prompted questions with the exception of the false address and phone number he gives at the outset. Otherwise, he is talking about his own unscripted medical history (see Appendix C for release permission). His high blood pressure, his insulin dependency, and the history of cancer in his family are subjects that Danny talks about sincerely and in detail without anyone putting words in his mouth. No attempt was used to elicit specific grammar forms or technical jargon. The answers given are unrehearsed and reflect Danny's natural way of speaking English. Furthermore, the medical language he uses to describe himself is the language he has picked up through his own encounters with the local healthcare system. When asked about his eyesight, Danny responds, "My eyesight is relatively good. I have a little bit of retinopathy, though." His reference to retinopathy, a degenerative eye disease caused by the effect of high blood sugar on the small blood vessels in the eye, was unprompted and is entirely consistent with someone with a long history of diabetes.

Adding to the realism, Danny is not always direct at first about his personal habits and so needs to be pressed for an honest answer. When asked, "Do you drink alcohol," he answers, "Occasionally." Following up with the question "Really?" causes him to

reconsider his answer and say, "Actually, I'm lying. Yes, I do drink...frequently." As in real life, this strategy does not always work with Danny. To the question, "When was the last time you got a tetanus shot?" Danny answers, "Hmm. I really can't remember the last time I got a tetanus shot." Pressing him by saying "Try to remember" gets only the answer, "Honestly, I don't know."

Phase 2: Testing the Prototype System

Before the development of any additional virtual medical history interviews can begin, a preliminary proof of concept feasibility study is needed to explore this prototype's suitability for its intended purpose. How suitable, in practical terms, is the VLP likely to be as a pedagogical intervention to train its intended population of learners? To answer this question, two research goals were envisioned. The first goal was to evaluate the ease of operability of the current configuration of the VLP. In short, the ease of operability of the system was determined by assessing whether the system's controls and features are easy to locate and use, and by assessing whether the medical history interview task is easy to complete when standards for pronunciation accuracy are met. Although efforts were made to integrate the speech recognition engine with the multimedia elements of the software into an easy to use graphical user interface (GUI), a reconfiguration of the GUI may be necessary if learners find the system too difficult, too awkward, or too complicated to use without help. Furthermore, it was worthwhile to determine which parts of the medical history interview task present the greatest pronunciation challenges to learners and whether the available models and feedback adequately helped learners to overcome those challenges.

The second research goal was to evaluate the VLP's fitness of purpose. The fitness of purpose of the system was determined by comparing learners' perceived language learning needs with aspects of the VLP task, and by ensuring that any training effects on learners' production are indeed consistent with intended learning outcomes. To these ends, specific questions motivated various elements of the data collection procedures and instruments described in the relevant sections below. References to either Questionnaire 1 (Q1) or Questionnaire 2 (Q2) and other sources of data (described in the Materials subsection below) that were drawn upon to answer the specific questions below are given in parentheses.

Questions

Subjective Measures of Ease of Operability: System Difficulty

1. Do participants report that the VLP is easy to use? (Q2: #12)
2. Do participants report that the VLP takes a long time to learn to use? (Q2: #13)
3. What irritations with the system do participants report? (Q2: #17)
4. Do participants report the video quality as being adequate? (Q2: #18)
5. Do participants report the question prompt to be clear and easy to read? (Q2: #19)
6. Do participants report the aural model to be useful? (Q2: #20)
7. Do participants report the graphic feedback on pronunciation to be useful? (Q2: #21)
8. What improvements do the participants suggest for the VLP? (Q2: #24; Interview Qs 1, 2, 3)

Objective Measures of Ease of Operability: Task Difficulty

9. Which questions produce the lowest confidence scores? (Logfile confidence scores of successful recognitions)
10. Of these low-scoring questions, which words produce the lowest pronunciation scores? (Logfile pronunciation scores of individual words within lowest scoring successful recognitions)
11. Which questions are among the lowest scoring 50% of items for all participants? (Logfile confidence scores of successful recognitions)
12. Which questions produce the highest number of utterance rejections? (Logfile of utterance rejections)
13. Which questions produce utterance rejections for multiple participants? (Logfile of utterance rejections)
14. Which recognizer difficulty level do participants prefer? (Q2: # 16; Interview question #1)
15. What is the average confidence score for each participant?
16. What number of utterance rejections lead participants to prefer a lower confidence rejection threshold setting? (Logfile of utterance rejections)

Fitness of Purpose: Learner Needs

17. Do participants expect English to be useful in their nursing careers? (Q1: #5)
18. Are participants motivated to improve their English interviewing skills? (Q1: #15 Q2: #3)

19. Do participants believe that patients should be able to receive health services in both official languages? (Q1: #11, 12, 13)
20. Do participants see their existing English oracy skills as being adequate to perform their nursing duties in English? (Q1: #14, 20, 21)
21. Are participants more inclined to develop oracy or literacy skills in English? (Q1: #6)
22. What types of situations do participants expect they will need English oracy skills? (Q1:#7, #8)
23. Who do participants expect to converse in English with most frequently within the medical contexts? (Q1: #9)
24. Do learners value the feedback the VLP provides? (Q2 #14, 15, 16, 21)

Fitness of Purpose: Training Effects

25. Do participants report higher or lower levels of confidence in their ability to interact with patients after using the system? (Q1: #10, 20, 21; Q2: #1, 2, 8, 9)
26. Do recognizer confidence scores rise as a function of practice with the system? (Logfile confidence scores of successful recognitions)
27. Do participants' speech rates increase with practice with the system? (The number of words in utterances divided by duration scores for successful recognitions from logfiles)
28. How do repeated utterance rejections affect participants' speech rate? (Logfile data)

Fitness of Purpose: Learners' Perspectives

29. What general assessments of the fitness of purpose of the VLP can participants provide?
30. What changes and improvements can the participants suggest?

Pre and post questionnaires as well as semi-structured post interviews were employed to gather learner preferences and confidence levels as well as to measure any change in question grammar knowledge specific to the medical history interview. Automatic recordings and logging of learner utterance duration were used to derive a measure of speech rate (calculated by dividing the number of words in any given utterance by the total duration of this utterance). A video recording of the software trial was made to assess how participants react to utterance rejections and to record the answers to the post interview questions.

Participants

The five participants in this study were all female, French native-speakers (hereafter Francophones), enrolled in *Technique de Soins Infirmiers*, a three-year, technical nursing program at a French junior college (CÉGEP) in a predominantly French-speaking area of Quebec, Canada (see Table 1). Upon admission to the college, each participant was placed into one of three English levels based on the results of an English proficiency test: false-beginner (learners who remain at the beginner level after years of instruction), low-intermediate, or high intermediate. None of the subjects had taken a pronunciation course before this study, but all had studied some English as their

second language through high school and taken two mandatory 45 hour, 4-skills general grammar courses at the nursing college.

Subject 1, the only high-intermediate speaker of English enrolled in this study, had just completed her second year of the nursing program. Subject 2, Subject 4 (both low-intermediates) and Subject 5 (a false-beginner) had just completed their third year of the program and were about to graduate. Subject 3 (a false-beginner) had the least clinical experience and training of the group, having only completed one full year of the program. Two of the participants had considerable contact with English speakers. Subject 1 (high-intermediate) spent 12 years working in an English-speaking environment in the province of Ontario and continues to have contact with English speaking friends. Subject 4 has an English-speaking boyfriend living in Kansas with whom she has regular contact by phone and visits during holidays. The two false-beginners, Subjects 3 and 5, report having had only limited exposure to English speakers outside of the classroom. Little is known about the language learning experience of Subject 2 as she did not provide any contact information and so could not be contacted to answer the online language experience questionnaire (see Appendix D).

Table 1

Participants' Language Experience

Background characteristics	Participants				
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
Year in program	Second	Third	First	Third	Third
Level of English	High-intermediate	Low-intermediate	False-beginner	Low-intermediate	False-beginner
First language	French	French	French	French	French
Second Language	English	English	English	English	English
Age	45	n/a	41	23	40
What age were you when you began learning English?	12	n/a	8	8	9
Language at home	French	n/a	French	French	French
Years learning English	10	n/a	1 year	15	10
Where did you learn English?	School, work	n/a	School	School, boyfriend	School
Do you speak English with friends?	Yes	n/a	No	Yes	No
Have you ever lived in an English environment? For how long?	Yes, 12 years	n/a	3 months	Yes, several times for 1 month	1 week holiday
Have you ever taken any pronunciation courses?	No	n/a	No	No	No
How often do you speak English?	2%	n/a	5%	20%	1%
How often do you listen to English?	2%	n/a	5%	30%	10%
How often do you read in English?	15%	n/a	2%	50%	3%
How often do you write in English?	5%	n/a	2%	30%	Never
Do you know any other languages?	Spanish	n/a	Arabic	No	No

Materials

Two questionnaires and a semi-structured terminal interview were used to collect the participants' information and attitudes toward the VLP prototype. Both questionnaires were computer-based and were created using WebQuizXP, a quiz development software that can create computer-based surveys for a web browser which save responses locally to a Microsoft Access database on the same machine running the VLP. The three advantages of using browser-based surveys over paper-based surveys are that questions appear one-at-a-time, answers cannot be changed once they have been entered, and the start time and end time are logged automatically.

Questionnaire 1

After the participant had given written consent to participating in the study (see Appendix E), the first questionnaire (see Appendix F), containing 21 questions, was given just and before using the VLP. The first four questions of this quiz were intended to establish the learner's level of nursing knowledge and English ability. The participant was therefore asked to confirm his or her student status in question 1, his or her registration in the nursing program in question 2, his or her year in the program in question 3, and level of college English in question 4.

Question 5 was intended to elicit the learner's expectation of the degree of utility that English has for a working nurse. Answers range from *very useful* to *I won't need it.* Also included were the answers *I don't know* and *I will not work as a nurse* to identify participants who might be likely to have a very low degree of motivation to learn English related to nursing.

Questions 6 to 9 sought to confirm that the participants of this study shared the preference for language training that emphasizes face-to-face oral-aural interaction with patients identified in previous studies (e.g., Lepetit & Cichocki, 2002).

Questions 10 and 14 sought to measure the participants' confidence in their ability to interview a patient in English prior to using the VLP and thereby establish a baseline to compare any increases or decreases in confidence from exposure to the virtual dialogue. The expectation was for an increase in this type of second language confidence to occur.

Since the site of this small-scale feasibility study was in a region of Quebec known for its separatist politics, questions 11, 12 and 13 were intended to identify participants who might have strongly negative attitudes toward Anglophones or Allophones of Quebec asserting their linguistic rights. This allowed for the possibility that a participant might see a need for and utility in learning English but have a strong counterbalancing desire not to provide bilingual services for non-French speaking residents of Quebec to ensure that the language of all business within the borders of Quebec, including health services, remains French.

Question 15 asked the participant to indicate whether or not he or she wanted to improve his or her English ability. The system was, after all, intended for healthcare professionals who were motivated to improve their English speaking skills. Feedback from study participants who were not motivated to learn is unlikely, therefore, to provide any useful information about the system's fitness of purpose.

The next four questions of the first questionnaire, questions 16 to 19, sought to discover the learner's pre-exposure ability to formulate medical history questions in English. They asked, "What question could you ask a patient to find out his or her

name?"; "What question could you ask a patient to find out about any medication he or she is taking?"; "What question could you ask a patient to find if there is a history of heart disease in his or her family?" and "What question could you ask a patient to find if there is a history of cancer in his or her family?" Grammar elicitation test items were used to compare participants' ability to generate well-formed medical history interview questions with their self-reported confidence levels in their knowledge of English grammar.

The last two questions asked participants to indicate their subjective impressions of their own confidence levels in their ability to speak with correct grammar and pronunciation.

Questionnaire 2

The second questionnaire (see Appendix G), containing 24 questions, was administered immediately after using the VLP for the first time. Questions 1, 2, 3, 8, and 9 were each repetitions of questions 10, 14, 15, 20, and 21 respectively of Questionnaire 1. The object of repeating these questions was to determine if using the VLP will have any effect on the participants' self-perceived confidence in their ability to provide health services in English. According to Dörnyei, "language learning in most people's minds is inevitably associated with perceptions of some degree of learning failure" (2001, p. 57). The expectation was that successfully completing a realistic simulation of a medical history interview in English with the VLP would have a positive effect on participants' confidence in their own adequacy to communicate with patients in English.

Questions 4 through 7 were repeats of the question grammar elicitation tasks from the first questionnaire. A comparison between the responses given before and after using

the VLP was expected to reveal any variability in the accuracy and complexity of participants' questions.

Question 10 asked the participants to give their general impression of the VLP. Multiple answers were possible. Question 11 asks for a general impression of Danny, the virtual patient. The possible answers are *excellent, good, bad,* or *terrible.*

Questions 12 and 13 respectively ask how easy or difficult the system was to use and how long it took to learn how to operate it.

Question 14 asked the participants to indicate their levels of frustration at having their pronunciation rejected. It was expected that participants would enjoy the conversational style of feedback on pronunciation that the system provides but could begin to find it frustrating when recognizer confidence acceptance levels were set to "advanced." A global impression was sought. In question 15, participants were asked to choose from a list of adjectives to characterize the system's rejection of poor pronunciation: *irritating, funny, stupid, bad, acceptable*, and *good.* Multiple answers were possible. Question 17 asked the participants to indicate if anything bothered them about the system.

Question 17 asked the participants to indicate their preferred recognizer confidence acceptance threshold: *beginner, intermediate,* or *advanced.* It was expected that the beginner level would not provide a challenge for the learner, and intermediate and advanced would be the preferred level depending on the proficiency level of the learner.

Questions 18 through 21 asked about specific features of the system: the quality of the video, the legibility of the question, the usefulness of the audio pronunciation

model, and the usefulness of the graphical feedback on pronunciation. Question 22 asked if the participants were inclined to use the VLP to interview the virtual patient more than once. It is expected that the participants were unlikely to want to interview the patient a second time once all of his medical information had been collected.

Question 23 asked a participant to suggest features that they would like to see in future dialogues. Participants could select more than one response from among the following choices: different medical conditions, different English accents, different foreign accents, different ages, uncooperative and aggressive patients, or *do not make any additional virtual dialogues because one is enough.* The final question was an open question inviting any suggestions for the improvement of the VLP system.

Semi-Structured Interview Questions

After being given another ten minutes to use the dialogue and manipulate the VLP system settings freely, a terminal semi-structured interview was used (see Appendix H for the questions). The questions were intended to confirm answers given on the second questionnaire and stimulate the participant to share any opinions and ideas for the improvement of the system.

Language Experience Questionnaire

As a follow-up, participants were contacted where possible by email to answer an online language experience questionnaire (see Appendix D). Items in the questionnaire relate to the participants exposure to English and English language instruction.

Procedure

After gaining ethics approval from both Concordia University and the participating nursing college, the study was conducted with one subject at a time in a reserved classroom within the cegep. Subjects entered the testing room one at a time. They were greeted, given the consent form to read and sign (see Appendix E), and asked if there were any questions about it. A video camera was turned on, and the subject was shown how to do the computerized questionnaire and given approximately five minutes to complete it. After the questionnaire was completed, the VLP was initialized, set to the beginner level (confidence threshold 35), and the subject was told to ask the first 20 questions to the virtual patient and then stop. The subject was given 10 minutes to ask all 20 questions. The VLP was then set to the intermediate level setting (confidence threshold 45), and the subject was asked to continue for the next 20 questions (21-40) and given 10 minutes. The VLP was set to advanced (confidence threshold 55), and the subject was asked to continue for the remaining 32 questions (41-72).

Once the interview was completed, the second computerized questionnaire was given. After completing the second questionnaire, the subject was invited to experiment freely with the VLP until the end of the hour. Just before leaving, the subject was asked for any final thoughts on the system (see Appendix H for the terminal interview questions). In addition to the data collection procedures described above, each utterance

produced by the learner for the system to recognize (when the subject clicks the system's "recognize" button) was recorded automatically by the system in .wav format.

CHAPTER 4: EASE OF OPERABILITY RESULTS

Two areas of the ease of operability of the VLP were investigated during the study: participants' subjective impressions of the system difficulty and objective measures of task difficulty. Subjective impressions were collected from participants in the post questionnaire and the post interview where participants were asked to give general impressions about the system in addition to specific evaluations of features of the user interface and materials. The results are given under the heading *System Difficulty*. Objective measures of participants' difficulties with task items were made by identifying the most problematic questions and words for learners in terms of system rejections of participants' utterances, recognizer confidence scores, and system-generated pronunciation scores. The results for these measures are given under the heading *Task Difficulty* below. For readability, the answers to the 16 questions in Chapter 3 (question 1-16) that motivated data collection methods related to the ease of operability of the VLP will not be presented question by question. Rather, results are clustered and reported under the relevant subsections.

System Difficulty

Apart from the laptop computer hardware and headset, the VLP system has two principle components that participants were asked to assess: the VLP software which includes the GUI with its audio and video components, and a pen-and-paper medical history questionnaire form (see Appendix A). Data from the post-questionnaire and the video-recorded post-interviews are presented for each of these components in turn.

In general, all five participants reported that they found the VLP software either very easy to use (n = 4) or easy to use (n = 1). In terms of the time it took to understand how to operate the system, participants indicated that it took either no time (n = 3) or little time (n = 2). All participants found the question prompts to be "very easy" (n = 4) or "easy" (n = 1) to read and the quality of the compressed videos of the virtual patient to be excellent (n = 3), good (n = 1) or adequate (n = 1).

However, the native speaker audio pronunciation models caused two of the participants some difficulty. While Subject 1 indicated that she did not use the pronunciation models at all, Subjects 2, 3 and 5 reported the audio models to be very useful. However, when asked to give her impression of the worst aspect of the software, Subject 3 reported the following as her assessment of the pronunciation models.

> Juste au niveau de quand elle parle, quand on fait l'écoute, elle parle tellement vite. Mois c'est que je trouve le pire. Même que je le commence 5 fois, je vais entendre la même chose. [Translation: Just in terms of when she speaks, when I try to listen, she speaks very fast. For me, that's what I find the most difficult. Even if I activate it five times, I will hear the same thing.]

This reference to hearing the same thing five times seems to reflect a frustration with not being able to perceive the segmental details of the audio model even after listening to it repeatedly.

Subject 4 also reports some difficulty with the audio models. She was the only participant to indicate that they were only somewhat useful on the post-questionnaire.

She explained later during the post interview that she similarly found the native-speaker pronunciation model too fast. She said this:

> The worst thing? I think really it is the girl [the female speaker who recorded the audio models]. Maybe she could do different speeds. We could choose different speeds. Like, that one is fast, and sometimes I would take it slower and try to say it and then take it faster, then try to see the difference.

For both Subject 3 and 4, the natural speech rate of the native speaker audio models made it difficult to hear the precise pronunciation of individual words and their segments. Adding additional recordings of the script item at different speech rates therefore would allow learners the option of attending to segmental features of words before attending to prosodic features of the sentence. One or more additional recordings could be made of each script item, and additional buttons could be provided on the interface to activate the slower pronunciation models. Alternately, each word in the text prompt could be linked to an audio or video recording of the word spoken in isolation from the sentence. The merits of the various options will need to be considered.

Two participants indicated that not being able to get the patient to repeat his answers had the effect of making the task of recording medical history information on the medical history questionnaire more difficult. Subject 3 said in the post interview that the system needs a repeat button on the GUI to get the patient to repeat what he said. She added that she found the combination of listening to the patient, finding her place on the questionnaire, and writing down his information too difficult without a repeat feature. Subject 4 had a similar comment.

> There is no rewind. There is answers I would like to ask him again, or part of the answer he gave me that I need and I didn't write and I just didn't know how to get it again. It would be really great to be able to say "Could you say that again?" to really get all the information I need. For me, he talks fast. If I could ask him to say the answer again, it would be easier to get the information.

In fact, it is possible with the present prototype to get the patient to repeat. Using the mouse, the learner can open the drop-down menu, scroll up to the previous question, and ask it again. However, both of the participants' suggested solutions are, from the learner's standpoint, easier to achieve and would therefore enhance the VLP's ease of operability. Of the two suggestions, however, the addition of phrases to recognizer's grammar file to cause the virtual patient to speak more slowly, explain, or repeat his last answer would be preferable for the additional oral practice it would afford the learner and for the opportunity to learn the kind of phrases Lear (2005) identified as being needed by clinicians in the field.

Difficulties with the pen-and-paper component of the VLP system were also identified in the post interview. Subjects 1, 2, and 4 all mentioned that there was not enough space on the medical history questionnaire (Appendix A) to record all of the information the virtual patient provided. Subject 1 resorted to using the back of the page. Subjects 2 and 4 both indicated that they would have written more if there had been space to do so. Subject 4 also suggested that the number of columns on the form be reduced from three to two.

To summarize, the participants found the VLP system easy to use with only minor frustrations with the audio, video, and questionnaire. The audio models were found to be too fast, so slower audio models should be added. Knowing how to get the virtual patient to repeat his answer was not evident to participants, making it more difficult to record the virtual patient's information on the form, therefore alternate methods for getting the virtual patient to repeat his answer should be explored. Finally, the medical history questionnaire was found to be too crowded. More blank space is required for the task.

Task Difficulty

In terms of product design, it is important that the participants found the system generally easy and intuitive to use. With a few exceptions, the VLP appears to be well-designed in this respect. However, in terms of task pedagogy it is important that the difficulty of the VLP task, as much as possible, come from the challenges of pronouncing the second language in a communicative exchange and not from extraneous, preventable sources.

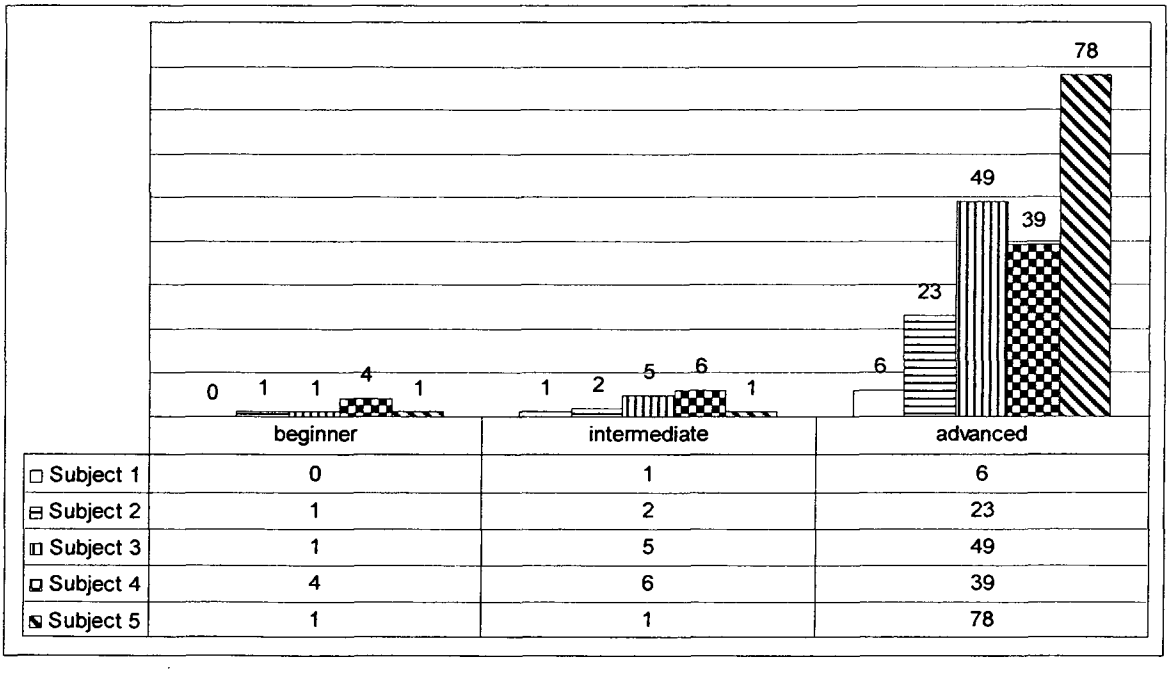
As explained above, the degree of pronunciation challenge for learners is determined by each of the three recognizer confidence rejection threshold settings. They are labelled for the learner as *beginner*, *intermediate*, and *advanced.* From the learner's point of view, these settings represent three standards of pronunciation accuracy that each utterance must meet in order for the system to pass to the next question. Not meeting the threshold prompts the learner to try again. In this way, utterance rejections provide the learner with useful opportunities to modify his or her pronunciation but will become frustrating if utterances repeatedly fall short of the threshold and the dialogue does not advance to the next question.

The number of rejections over the course of the virtual dialogue for participants by proficiency level is as follows. The false-beginners received the most rejections, with Subject 3 and Subject 5 receiving 55 and 80 rejections respectively, while the low-intermediate learners, Subject 2 and Subject 4, received 26 and 49 utterance rejections respectively. Subject 1, the only high-intermediate participant, received only 7 rejections (see Figure 4). This shows that lower initial proficiency levels correlate with higher utterance rejection rates. As intended, the ease of operability of the system in terms of task duration and ultimate success can be managed by manipulating the recognizer confidence threshold levels.

Participants all indicated positive attitudes on the post questionnaire toward the VLP's rejection of utterances that scored below recognizer confidence-rejection thresholds. Since none of the subjects indicated a preference for the beginner rejection threshold, it suggests that learners value the recognizer rejections and may not seek to avoid them by setting the difficulty level to its lowest level. Looking more closely at the participants' utterance rejection rate for each of the three difficulty settings (beginner, intermediate, advanced), rejections rates rise sharply for all participants once the system is set to the advanced threshold level (see Figure 4). When asked which threshold each participant preferred, the two participants with the highest number of rejections (S3, S5) both expressed a preference for the intermediate setting, suggesting that the number of utterance rejections they received at the advanced level (49 and 78, respectively) caused excessive frustration. The other three participants each expressed a preference for the advanced setting, indicating that they appreciated the challenge of the highest setting. It

also suggests that an even higher difficulty level might be valued by learners seeking a greater challenge.

Since subjects 1, 2, and 4 received 6, 23, and 39 utterance rejections respectively at the advanced setting, it can be concluded that the optimal rejection rate for motivated learners may lie somewhere between 39 rejections (the highest rate for a participant who prefers the advanced setting) and 49 rejections (the lowest rate for a participant who prefers the intermediate rate) per 34 question-prompts. Expressed another way, the optimal balance of task challenge and task frustration in future virtual dialogues will likely be found between 1.15 and 1.44 rejections per question prompt.



	beginner	intermediate	advanced
Subject 1	0	1	6
Subject 2	1	2	23
Subject 3	1	5	49
Subject 4	4	6	39
Subject 5	1	1	78

Figure 4. Rejections by threshold setting.

Script Item Difficulty

Another dimension of the system's ease of operability is the relative difficulty of each of the script items in the medical history VLP interview dialogue task (see Appendix B). Four measures of item difficulty were explored: measure 1—the average combined confidence score for each script item; measure 2—the frequency with which an item was among the lowest scoring 50% of items across participants; measure 3—the total number of combined utterance rejections generated for each script item; and measure 4—the number of participants that received utterance rejections for each script item. These four measures roughly correspond to the questions 9 to 16 listed in Chapter 3. Since script items at the beginner and intermediate settings were not held to the same standard of pronunciation accuracy triggering fewer clarification requests and accepting lower confidence scores for each script item, only the relative difficulty of script items 40 to 71 at the advanced difficulty setting was investigated.

Measure 1

Measure 1 was the average combined confidence score for each script item. This measure was determined by averaging confidence scores for each script item across the five participants. The script items at the advanced difficulty setting ($n = 31$) that received the ten lowest average confidence scores (see Figure 5) in ascending order (from lowest to highest) were as follows: #66, *Do you drink alcohol?*; #45, *Do you get frequent headaches?*; #59 and #60, *Have you noticed any penile discharge?*; #61, *Have you noticed any testicular lumps?*; #51, *Do you have asthma?*; #53, *Do you suffer from varicose veins?*; #63, *Do you smoke?*; #56, *A long time ago?*; #67, *Really?*; #43, *Have you ever had a mental breakdown?*.

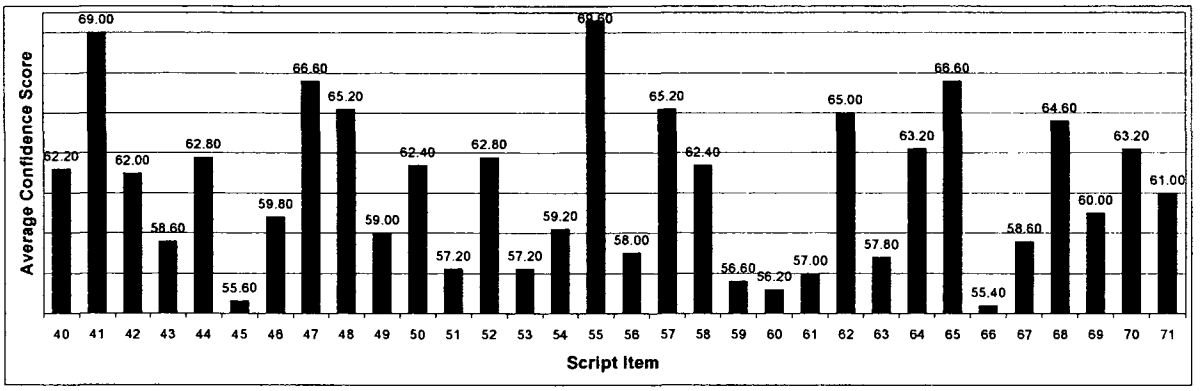


Figure 5. Average confidence score for each script item at the advanced difficulty setting.

Measure 2

Measure 2 was the frequency with which an item was among the lowest scoring 50% of items across participants. This measure of difficulty was determined by the frequency with which an item was among the lowest scoring 50% of items for four or more participants. Five lists, one for each participant, were prepared of confidence scores produced for script items at the advanced difficulty setting ($n = 31$) from the data retrieved from the system log files. By comparing the five lists for overlap, script items that were among the lowest scoring 50% of items for four or more participants were identified. They are as follows: #66, *Do you drink alcohol?* (all 5/5); #49, *Do you do drugs?* (4/5, except S3); #51, *Has there been any tuberculosis in the family?* (4/5, except S4); #53, *Do you suffer from varicose veins?* (4/5, except S4); #59, *Have you noticed any penile discharge?* (4/5, except S1); #63, *Do you smoke?* (4/5 except S1); and #67, *Really?* (4/5, except S1).

Measure 3

Measure 3 was the total number of combined utterance rejections generated for each script item. This measure of script item difficulty was calculated by counting the number of utterances that the system rejected, for the five participants as a group, as

being below the threshold for pronunciation accuracy for each script item at the advanced difficulty setting (see Figure 6). As explained above, when the learner reads the dialogue script item aloud from the text prompt on the screen into the headset microphone, the recorded utterance is sent to the system's speech recognition engine and a confidence score for the entire utterance is compared to the confidence rejection threshold. If the confidence score is below the threshold, the system returns an utterance rejection result and the learner is prompted to try again. The 10 items that received the highest total number of rejections for the five participants as a group were, in descending order, as follows: #45, *Do you get frequent headaches?* (n = 27); #66, *Do you drink alcohol?* (n = 19); #40, *Do you ever get seizures?* (n = 15); #67, *Really?* (n = 13); #51, *Has there been any tuberculosis in the family?* (n = 51); #61, *Have you noticed any testicular lumps?* (n = 11); and #59, *Have you noticed any penile discharge?* (n = 10); #46, *Do you suffer from dizziness?* (n = 8); #42, *Do you have hearing difficulties?* (n = 8); #41, *Do you ever get rashes or skin troubles?* (n = 6).

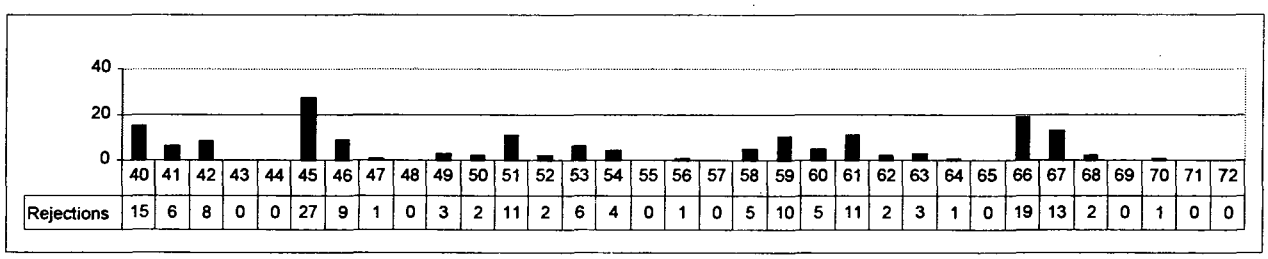


Figure 6. Total number of utterance rejections per script item at the advanced difficulty setting.

Measure 4

Measure 4 of script item difficulty was a count of the number of participants who received one or more utterance rejections for a particular script item. Put simply,

utterance rejections occur when the system judges the learner's utterance to be below the standard of accuracy determined by the recognizer confidence rejection threshold (set at 55) represented by the advanced difficulty setting. A list of the script items that generated confidence scores below the rejection threshold of 55 (the threshold value at the advanced difficulty setting) was prepared for each participant. The resulting five lists were compared for overlap, and the 10 script items that generated utterance rejections by at least three of the five participants were identified. By this measure, the 10 most difficult script items to pronounce are as follows: #53, *Do you suffer from varicose veins?* (n = 5); #45, *Do you get frequent headaches?* (n = 4); #61, *Have you noticed any testicular lumps?* (n = 4); #66 *Do you drink alcohol?* (n = 4); #67, *Really?* (n = 3); #63, *Do you smoke?* (n = 3); #59, *Have you noticed any penile discharge?* (n = 3); #51, *Has there been any tuberculosis in the family?* (n = 3); #42, *Do you have hearing difficulties?* (n = 3); #40, *Do you ever get seizures?* (n = 3).

Multiple Measures

Each of the above measures produces a list of 10 challenging script items that is slightly different from the other three. Combined, there are 16 sentences flagged for difficulty by one or more of these measures (see Table 2), but only 10 items were flagged as difficult by two or more of the above measures. The 10 sentences that were flagged twice or more (see column "Total" in Table 2) are as follows: #40, #42, #45, #51, #53, #61, #63, #59-60, #66, #67. It is curious that this multiple-measures list of difficult script items is identical to the list generated by Measure 4. Considering the small number of participants in this study, it is difficult to determine without further testing whether this

consistency between measures is simply a coincidence or significant evidence of reliability for Measure 4 in future script item difficulty assessments.

Table 2

Comparison of Script Item Difficult Measures

Script item #	Measures of difficulty				
	Measure 1	Measure 2	Measure 3	Measure 4	Total
40			✓	✓	2
41			✓		1
42			✓	✓	2
43	✓				1
45	✓		✓	✓	3
46			✓		1
49		✓			1
51		✓			1
51	✓		✓	✓	3
53	✓	✓		✓	3
56	✓				1
59-60	✓	✓	✓	✓	4
61	✓		✓	✓	3
63	✓	✓		✓	3
66	✓	✓	✓	✓	4
67	✓	✓	✓	✓	4

Script Item Difficulty Conclusion

In sum, this exploration of the relative pronunciation difficulty of dialogue script items has yielded a list of items that presented the greatest challenge to the participants of this study. In light of the difficulties with the audio pronunciation models noted above, these 10 sentences (#40, #42, #45, #51, #53, #61, #63, #59-60, #66, and #67) constitute

the highest priority for new pronunciation model recordings. However, what it is about these 10 items that makes them challenging for learners to pronounce accurately at the advanced threshold setting is not clear. How word difficulty contributes to script item difficulty will therefore be taken up in the next section.

Word Difficulty

It is not clear without further investigation what it is about the 10 script items identified above that made them so challenging to the participants. Were the content words or the function words the primary cause of the participants' difficulty? Which words received the most repair after feedback? Moreover, how did the learners overcome the difficulty posed by these items? To answer these questions, the relative difficulty of individual words in this ten-most-challenging list of script items needed to be explored. To this end, recognizer confidence scores and system generated pronunciation scores retrieved from the system log file were analyzed. The results of the analysis should suggest possible system improvements and implications for future development.

Looking at the word confidence scores for each word in the 10 sentences identified, average final confidence scores across participants were calculated for each word. All of the words that fell below the confidence rejection threshold of 55 (the threshold of the advanced setting) were identified. They are underlined in the list below. The words *do you* appear in six of the 10 sentences. In each instance, they appear at the sentence-initial position, and in all instances the word *do* fell below the recognizer confidence rejection threshold of 55. The word *you* fell below the threshold in four of the six instances. In five of the 10 sentences, content words fell below the threshold. *Have* fell below the threshold twice, but *has* achieves a score well above the threshold at 64.4.

In all but two of these sentences, words in the sentence-initial position fell below the recognizer confidence threshold. The frequency of below-threshold scores for words in this sentence-initial position could suggest a problem with the way these words are scored, and this issue requires further analysis.

1. #40 Do you ever get seizures?
2. #42 Do you have hearing difficulties?
3. #45 Do you get frequent headaches?
4. #51 Has there been any tuberculosis in the family?
5. #53 Do you suffer from varicose veins?
6. #59 Have you noticed any penile discharge?
7. #61 Have you noticed any testicular lumps?
8. #63 Do you smoke?
9. #66 Do you drink alcohol?
10. #67 Really?

First, however, it is apparent in that there is a problem with judging word difficulty from log file word confidence scores. This is most clearly illustrated with script item #67. It is the only utterance in which none of the words in the question prompt falls below the confidence threshold. It is also the only script item within the above ten-most-challenging list that contains only one word. These two distinctions are noteworthy as they illustrate the limitation inherent in using word confidence scores from successful attempts to identify problematic words in the virtual dialogue script items: they are generated after feedback and repair. Since the current configuration of the VLP does not log confidence scores for words from utterances that fall below the recognizer confidence

rejection threshold, comparisons of scores from pre-feedback attempts and post-feedback attempts are impossible using recognizer confidence scores alone.

However, the system does log system-generated 6-point pronunciation scores for both accepted and rejected utterances. For the purpose of assessing of task difficulty in terms of script items, difficult to pronounce words can be readily identified from their low (1/6 or 2/6) pronunciation scores. As such, repair was quantified as the difference between the pronunciation scores from the earlier rejected attempt and the later successful attempt. Where average differences between pre and post feedback are small, at least some conclusions about the adequacy pronunciation models and feedback and scoring can be made.

Script item #53 (*Do you suffer from varicose veins?*) furnishes a ready example of the operability and pronunciation challenges facing learners in that it is the one sentence for which all five subjects received a rejection (see Table 3). On average, the lowest pronunciation scores for words in attempts prior to feedback and repair are both content words: *varicose* and *veins* (see Table 3). A comparison of pronunciation scores before feedback and repair with pronunciation scores after feedback and repair indicate that *veins* gets more repair after feedback (+1.6) than *varicose* (+0.2).

One explanation for more repair might be that the sentence-final position that *veins* occupies makes it more salient to the participants when listening to the native speaker audio model. In contrast, *varicose* gets very little repair after feedback with a change in the average pronunciation score for all five participants of only +0.2. This result suggests that an exploration of alternatives to sentence-length pronunciation models—the VLP's present approach to modelling pronunciation—could further enhance

ease of operability in terms of task difficulty by making the correct pronunciation of words in the non-final position of the sentence more salient to learners. For example, in the questions *Do you get frequent headaches?* and *Do you suffer from varicose veins?* the word *frequent* and the word *varicose* are sentence-medial and so less salient to learners. Sentence-medial words could be modelled on their own and outside of the environment of the questions in which they appear to make them easier to perceive. Better models might then lead to fewer utterance rejections, and learners might find the virtual dialogue less frustrating as a consequence.

Table 3

Pronunciation Scores for Script Item #53

	Pronunciation scores before and after repair						
	Do	*you*	*suffer*	*from*	*varicose*	*veins?*	average
S1 before	3.00	2.00	4.00	5.00	1.00	1.00	2.67
S1 after	3.00	3.00	5.00	6.00	1.00	3.00	3.50
S2 before	1.00	1.00	2.00	1.00	1.00	1.00	1.17
S2 after	1.00	2.00	4.00	4.00	1.00	2.00	2.33
S3 before	1.00	1.00	1.00	2.00	1.00	1.00	1.17
S3 after	1.00	2.00	2.00	2.00	1.00	3.00	1.83
S4 before	1.00	1.00	3.00	3.00	1.00	1.00	1.67
S4 after	1.00	1.00	4.00	5.00	2.00	4.00	2.83
S5 before	3.00	2.00	1.00	1.00	1.00	1.00	1.50
S5 after	3.00	3.00	2.00	5.00	1.00	1.00	2.50
average	1.80	1.80	2.80	3.40	1.10	1.80	2.11
before	1.80	1.40	2.20	2.40	10	1.00	1.63
after	1.80	2.20	3.40	4.40	1.20	2.60	2.6
difference	0.00	0.80	1.20	2.00	0.20	1.60	0.96

Function words also appear to have been largely overlooked by participants in item #53. *You* gets only moderate repair after feedback by four of the five participants, and *do* gets no repair at all from any of the five participants. Although, this lack of an increase in the pronunciation score for *do* suggests a similar lack of attention given to it in the audio model by the participants, this may not be reason for concern as the following example from script item #45 seems to illustrate.

Table 4

Pronunciation Scores for Script Item #45

	Pronunciation scores before and after repair					
	Do	*you*	*get*	*frequent*	*headaches?*	average
S2 before	1.00	2.00	4.00	1.00	1.00	1.80
S2 after	1.00	1.00	5.00	3.00	3.00	2.60
S3 before	1.00	1.00	1.00	1.00	1.00	1.00
S3 after	3.00	1.00	4.00	1.00	4.00	2.60
S4 before	1.00	1.00	2.00	1.00	1.00	1.20
S4 after	3.00	4.00	4.00	3.00	3.00	3.40
S5 before	1.00	1.00	1.00	1.00	1.00	1.00
S5 after	1.00	1.00	1.00	1.00	1.00	1.00
average	1.50	1.50	2.75	1.50	1.88	1.83
before	1.00	1.25	2.00	1.00	1.00	1.25
after	2.00	1.75	3.50	2.00	2.75	2.40
difference	1.00	0.50	1.50	1.00	1.75	1.15

In script item #45, the three lowest scoring words prior to feedback and repair are *do, frequent,* and *headaches*. Each receives the lowest possible score of 1/6 (see Table 4). Again the same pattern emerges where the content word *headaches* in the sentence final position gets considerably more repair (+1.75) than the word *frequent* in the second to

last position. This is not surprising as sentence-final content words receive tonic (major sentence) stress in an utterance making them more salient to the learner.

Conversely, function words, which are normally reduced in everyday speech, are among the least salient elements of the sentence. Nevertheless, in the sentence initial position, *do* gets more repair (+1.0) than *you,* which gains only half a point (+0.5). Listening to the audio file of Subject 3's utterances makes it clear that the jump in pronunciation score for *do* from 1/6 to 3/6 is likely due to a slowed speech rate (from an initial 2.54 seconds to 4.01 seconds) and the replacement of schwa with [u]. Subject 4 also successfully employs this strategy of speech rate reduction (from 2.20 seconds to 3.59 seconds) and over-articulation of function words to surpass the rejection threshold. It should be noted that the audio model is 1.15 seconds long with *do you* reduced as in normal speech to [dəyʊ]. Ostensibly, Subjects 3 and 4 have developed this strategy on their own in order to produce an utterance that meets the recognizer confidence rejection threshold.

Referring to the recognition engine's dictionary file, indeed *do* is defined as the combination of /d/ and /uw/ and *you* as /y/ and /uw/, their non-reduced forms. No alternate definitions are given, and so it seems that the recognizer was not expecting the vowel reduction modelled by the native speaker in the pronunciation model. This kind of mismatch between audio pronunciation models and recognizer scoring models is bound to make the negative assessments (i.e. utterance rejections and low pronunciation scores on the graphical display) frustrating to the learner since greater fidelity to the model after feedback will produce lower scores and more negative assessments. Indeed, Subject 5 received sixteen utterance rejections for script item #45 alone. An alternate system

dictionary entry with vowels reduced for the question formula *do + you* would therefore help to enhance the system's ease of operability for this and similar script items, and obviate the need to resort to over-articulation of function words. If this ease of operability issue is not addressed, pronunciation errors in content words may persist, as the next example illustrates.

While Subjects 3 and 4's hypercorrection strategy (employed with item #45) was indeed effective at improving system recognition scores, it could be counterproductive to the pedagogical purpose of the system if the mispronunciation of key words goes unchallenged. This appears to be what happened when Subject 3 first pronounces *frequent* as [frænɪkli] and *headache* as [hɛt] in a rejected attempt, but then more recognizably as [fri:kʊɛn] and [hɛdetʃ] in combination with the over-articulation of *do you* to raise the system's confidence score over the threshold for the entire utterance. There is considerable repair to these two content words making the entire sentence more comprehensible, but sentence-level confidence scoring could make it possible for learners to use over-articulation to get over the threshold. When this strategy is employed, the opportunity for further repair of key content words is missed. This result suggests that an exploration of alternate methods of utterance scoring may also be worthwhile. For example, in addition to a minimum confidence threshold of 55 for the sentence as a whole, the added condition of minimum thresholds for key content words could also be made to apply. To illustrate, upon parsing and scoring the question "*Do you get frequent headaches?*" if the sentence achieves the threshold of 55 but the word *frequent* does not, the system could trigger a specific video clip of the virtual patient making the clarification request, *What kind of headaches?* Alternately, an ironic clip of the virtual

patient asking, "Did you say 'freaky' headaches?" or "What is a 'freaky' headache?" could be played.

Ease of Operability Conclusion

To sum up, testing the operability of the system with learners from a range of proficiency levels at the three difficulty settings showed that rejection rates were consistent with learners' starting proficiency. As expected, higher proficiency learners received fewer rejections than lower proficiency learners. However, difficulty-level recognizer rejection thresholds and the increments between rejection thresholds will need to be adjusted to ensure that all learners encounter the appropriate balance of challenge and success. Ostensibly, the optimal utterance rejection rate appears to be about one (1.15) rejection per dialogue script item. This means that in a virtual dialogue with 100 dialogue script items, the learner should be willing to pronounce without getting frustrated all 100 items plus 115 corrections after feedback (i.e., repairs) , making a total of 215 utterances within a communicative exchange. In other words, the maximal number of utterances a learner is willing to produce within the communicative exchange of a virtual dialogue before frustration sets in is equal to the number of dialogue script items plus the maximum tolerable number of retrials [Utterances = Items + (Items x 1.15)]. To achieve this ratio consistently for learners of all proficiency levels, a learner-adaptive approach to setting recognizer confidence thresholds may need to be considered.

In general, the participants found the system easy to operate. They found the GUI to be well-arranged, easy to understand how to use, and were largely satisfied with the multimedia aspects of the system. However, ease of operability issues surrounding the pronunciation models, feedback and scoring features were revealed through a detailed

analysis of quantitative measures of script item difficulty and contributing word difficulty. Two issues with the pronunciation models became apparent. Firstly, content words in the penultimate position of an audio pronunciation model recording seem to have posed perceptual problems for learners, evidenced by the minimal repair they received after feedback. Secondly, and perhaps more importantly, an oversight was discovered in that the VLP's internal dictionary does not contain reduced forms of common question formula function words. Consequently, it appears that participants' fidelity to the pronunciation models triggered inappropriate and excessive rejections of their utterances, prompting them to resort to reduced speech rates and over-articulation of function words instead of focussing on making additional repair to their pronunciation of content words.

Recommendations

The addition of reduced forms to a user-defined dictionary that would override specific existing dictionary entries of the current VLP is recommended. Also, audio models with slower speech rates or individual audio models of content words are suggested for future prototypes. These measures would help to ensure that learners focus on improving their content-word pronunciation accuracy rather than resort to hypercorrection of function words.

Additionally, the development and testing of alternate scoring procedures and feedback that target specific content word pronunciation errors may also help in this regard. Penultimate content words in script items were identified as a worthwhile starting point for such development efforts due to perceptual difficulties of words in this sentence position. Novel scoring routines that make utterance acceptance conditional upon key

word pronunciation accuracy could be developed for just this purpose. Making it clear to the learner why an utterance with a key word error was rejected, video clips that provide focussed feedback on a specific part of the sentence or provide a kind of ironic feedback where the virtual patient appears to have misheard the patient seems appropriate in this respect.

CHAPTER 5: FITNESS OF PURPOSE RESULTS

In addition to ease of operability, three key areas of the VLP's fitness for its intended purpose were evaluated: learner needs, training effects, and learners' perspectives. The first area of fitness was evaluated by determining whether or not the VLP correctly anticipates the learner-specific language learning priorities that members of the target population of learners have for themselves. In other words, is the present design of the VLP capable of responding to learners' perceived second language needs? Using the answers to closed-response questionnaire items in the pre-questionnaire, learners' reported language learning needs were compared with specific aspects of the VLP task. Participants were asked before seeing the VLP questions about how useful English was expected to be in their future nursing careers, who they believed they might need to communicate in English with, in what circumstances, how ready they felt themselves to be, and whether they were motivated to improve their skills for those interlocutors and circumstances. Responses to these items are arranged below under the subheadings *Perceived Utility of English, Expected Modes of Interaction, Expected Interlocutors, Sense of Preparedness* and *Reported Motivation to Improve English Interviewing Skills*, respectively. Again for the sake of readability, answers to the 14 questions (questions 17-30) in Chapter 3 that motivated data collection methods related to the fitness of purpose of the VLP will not be presented question by question. Rather, results are clustered and reported under the relevant subsections.

An evaluation of the system's ability to respond to general language learning needs was also made by investigating specific training effects of the system. Three training effects were explored. First, changes in participants' sense of preparedness were

observed by comparing responses given before and after exposure to the VLP to determine what effect the VLP might have on task-related confidence. Second, changes in participants' speech rate derived from data collected in the VLP's system log files were identified by calculating speech rate averages at regular intervals throughout the dialogue. Trends in average speech rate were identified in this way to determine what effect the VLP might have on this aspect of oral fluency. Third, changes in system-generated recognizer confidence scores were observed by calculating confidence score averages at regular intervals throughout the dialogue to determine whether changes in pronunciation accuracy might occur from using the VLP. These three training effects are explored in turn under the subheadings *Training Effect on Sense of Preparedness*, *Training Effect on Speech Rate,* and *Training Effect on Pronunciation Accuracy.*

Finally, participants were asked to suggest changes and improvements to the system that would make it more suitable for their needs. Relevant comments from the post questionnaire and semi-structured interview answers are given under the heading *Suggestions for Improvement from the Learner.*

Learners' Needs

Perceived Utility of English

When asked whether they expected English to be useful or not in their nursing careers, four of the five (Subjects 1-4) believed it would be "very useful," with one participant (Subject 5) expecting English to be only "somewhat useful" (see Table 5). Later in the post interview, this same participant revealed that in the rural area where she worked there were more monolingual Spanish speakers (i.e. seasonal agricultural

workers) than monolingual English speakers. As such, she felt that there was a more pressing need for Spanish at her hospital.

Table 5

Aspects of Participants' Motivation to Learn and Use English

Questions	Participants				
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
Would you like to improve your ability to interview patients in English?	Yes	Yes	Yes	Yes	yes
Do you think English will be useful to you in your career as a nurse?	very useful	very useful	very useful	very useful	somewhat useful
If necessary, could you interview a patient in English?	certainly, yes	certainly yes	certainly yes	certainly yes	I don't know
How confident would you be about communicating with a patient entirely in English?	very confident	somewhat confident	very confident	very confident	I would not be able to speak
How confident would you be about your ability to use correct English grammar?	Somewhat confident	somewhat confident	somewhat nervous	somewhat confident	very nervous
How confident would you be about your ability to use correct English pronunciation?	Somewhat confident	somewhat confident	somewhat nervous	somewhat confident	very nervous

Expected Modes of Interaction in English

When asked to indicate whether English literacy or oracy skills were a higher priority in their future nursing careers, all participants answered that their primary need for English would most likely involve face-to-face oral exchanges rather than phone conversations or having to give monologic presentations in English, confirming the findings of prior research (Lear, 2005; Lepetit & Cichocki, 2002).

Expected Interlocutors

In describing who they expect to be their conversational partners in English, all participants identified patients—rather than doctors, other nurses, administrators, or other individuals—as their most likely interlocutors.

Sense of Preparedness

All except one of the participants reported being sufficiently proficient in English to interview a patient (see Table 5). Only Subject 5, a false beginner, reported that she did not think she would be able to conduct the entire interview in English. Subject 2, a low intermediate learner, reported that she was only somewhat confident in this regard. Subjects 1, 3, and 4 were confident that they would be able to establish their communicative intent but expect errors in their pronunciation and grammar.

In the post-interview, Subject 4 explained that speaking English was for her a greater challenge than listening and reading in English because of her tendency for /h/ epenthesis and /h/ deletion, plus her difficulty predicting the correct pronunciation of English words. She says:

> Speaking is the most hard because of the pronunciation. With the 'h' for example, I will often say like 'heat' and 'eat.' I will switch by accident. Also, wind (gesturing a breeze and pronouncing it [waɪnd]) and wind (gesturing a turning motion and pronouncing it [wɪnd]), I always mess it up. Also, L-O-N-G and L-U-N-G (spelling aloud), I cannot say the difference. If I want to talk to the patient and he don't understand what I say in English, then he will not be able to answer

to my question and I will not be able to do that evaluation (holding up Appendix A).

Participants' doubts about the accuracy of their grammar seem only partly justified in light of evidence of question formation errors in their typed responses to open-response grammar test items on the pre-questionnaire and post-questionnaire (see Table 6). For example, participants were asked, "What question could you ask a patient to find out his or her name?" and "What question could you ask a patient to find out about any medication he or she is taking?" (See Appendices 4 and 5 for all four grammar test items). Subject 5 wrote, "What your name?" on the pre-questionnaire and "What is your name?" on the post questionnaire. While it is not expected that significant systemic changes in a learner's interlanguage would appear after a short exposure to any pedagogical intervention—including the VLP—differences here in questionnaire responses between the pre and post are likely due to reactivation of prior knowledge after a period of non-use. In contrast, Subject 3 wrote "What is your name?" on the pre-questionnaire and "What is name?" on the post. Equally, this is more likely a sign of fatigue than evidence of language attrition. Nevertheless, grammatical variability in Subjects 2, 3, and 5's question formation suggests that subjects are not merely being modest in their concern about the accuracy of their grammar. Further instruction and feedback on errors in some cases may indeed be warranted.

Table 6

Evidence of Grammar Errors

	Elicited questions forms			
		Pre-Questionnaire		Post-Questionnaire
Subj. 1	#16	What is your name?	#4	What is your name?
	#17	Can you give me the list of the medications you are taking currently?	#5	Are you on any medication now?
	#18	Do you know if in your family there is history of heart disease?	#6	Is there any history of heart disease in your family?
	#19	Is there any cancer history in your family?	#7	Any cancer history in your family?
Subj. 2	#16	Your name is?	#4	What's your name?
	#17	Could you tell me what is the medication you use at home?	#5	Do you take any medication?
	#18	Have you ever had a heart problem?	#6	Do you have heart problem in the family?
	#19	Do you have cancer in your family?	#7	Do you have any form of cancer in the family?
Subj. 3	#16	What is your name?	#4	What is name?
	#17	What is you taking?	#5	Do you taking a medication?
	#18	How is the problem health for the family?	#6	What is heart history for family?
	#19	What is the cancer for family?	#7	Is the cancer for family?
Subj. 4	#16	What is your full name?	#4	What is your name?
	#17	What type of medicine are you taking at home?	#5	Do you take any medication?
	#18	Do you have any heart disease in your family history?	#6	Is there any heart disease in your family?
	#19	Do you have any cancer history inside your family?	#7	Is there any cancer in your family?
Subj. 5	#16	What your name?	#4	What is your name?
	#17	What do you take medication in your home?	#5	What medication you take usually?
	#18	What do you history of heart and your family?	#6	What your history of heart in your family?
	#19	Have you cancer in his family?	#7	Have you cancer in your family?

However, Subject 1 does not seem to make any errors in her question formation. Her response to item 18, "Do you know if in your family there is history of heart disease?" may seem a bit awkward with the placement of "in your family" in the sentence-medial position, but it is not strictly an error. Subject 4's responses also seem largely accurate, with the exception of the word *inside* where *in* or *within* might be more appropriate. However, as she explained in the post interview, despite her considerable proficiency in English, without explicit correction, she is never sure what is correct and what is not. Subject 4 explained it this way.

> Often, I will say grammar problems, and I don't know that it is not right. Instead to say "it's over there," I said "it's by there." I didn't know it was wrong, so I keeps to saying it. In French it's "by there" [par là] but in English it doesn't work and I didn't know it, and my boyfriend was like, "There's something wrong in what you say." Finally, he just said, "Oh, it's not *by there*. You have to say, *It's over there*." Now I say, "over there." In class, you learn that this is wrong and that is right, but you don't use it so you don't remember it.

Reported Motivation to Improve English Interviewing Skills

Two aspects of language learning motivation were explored in the first questionnaire. The first aspect involved asking each participant directly whether she wanted to improve her ability to interview patients in English. All five participants indicated that, yes, they wanted to improve.

The second aspect of their motivation relates to whether they each believed that Quebec society had an obligation to provide health services to Anglophones and Allophones in English. Only one participant, Subject 2, strongly agreed that Anglophones

and Allophones should be able to receive health services in English and that Francophones should be able to receive health services in French in the rest of Canada (see Table 7). Subjects 3, 4, and 5 all answered more tentatively that they only "somewhat" agreed that Anglophones and Allophones should receive English language access to health services in Quebec and that Francophones deserved the same level of access in French outside of Quebec. This constitutes weak support for minority language access among four of the five participants of this study.

Table 7

Support for Minority Language Access to Healthcare

	Participants				
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
French entitlement to HC access outside of QC	I somewhat agree	I strongly agree	I somewhat agree	I somewhat agree	I somewhat agree
English entitlement to HC access in English inside Qc.	I strongly agree	I strongly agree	I somewhat agree	I somewhat agree	I somewhat agree
Allophone entitlement to HC access in English inside Qc.	I somewhat agree	I strongly agree	I somewhat agree	I somewhat agree	I somewhat agree

In the post interview, Subject 4 explained her reason for answering that she only somewhat agreed that patients should be able to get health services in either French or English.

> I don't want to make it obligatory so that [every] nurse must learn English to work because I think it is something personal, but I think it is important that the hospital have people to translate. I had only two courses of English here, and it is not enough.

The concern expressed here seems to be that making bilingualism a prerequisite for nurses would create an excessive burden on nursing students, who may not have the motivation or opportunity to improve their English ability. Nurses already struggle with a demanding course load, and the current amount of instruction provided by the college is insufficient for bringing about major changes in learners' general proficiency. The degree of English proficiency needed to conduct medical interviews with Anglophone patients, in her view, is a matter of personal responsibility outside of the professional commitment to become a nurse in Francophone Quebec.

The asymmetry in Subject 1's answers about healthcare access for linguistic minorities across Canada is somewhat more perplexing. She indicated strong support for English language access to healthcare for Anglophones in Quebec but less support for English access to healthcare inside Quebec for Allophones and less support for French language access to healthcare for Francophones outside of Quebec. It may be that she confers greater support for English access to health services for Anglophones because of an experience that touched her personally. In the post interview, she related the following story.

> I'm working right now at the hospital as a préposée [nursing assistant], and I think it's pathetic because some people speaks English and they don't even speak English to them. And there was a lady once. They thought that she couldn't speak, that she was mute. She wasn't mute! She was English! When I read her name, it was something like Caruthers or something, and I started to speak English with her, and she was the most happy patient in the world. Oh my God! But here we don't have very much English. But for my part, to become a nurse you should have at least two languages that you can currently speak, or be willing to do so, you know? Maybe not learn English, but learn the medical vocabulary. Because when people are sick, they need confidence, they need to be reassured. I thought it was sad.

Whereas Subject 4 sees minority language access to healthcare as the hospital's responsibility, Subject 1 sees it as a nurse's professional responsibility, at least initially. At first, Subject 1 argues for general bilingualism as a minimum requirement for nurses, but then quickly tempers her position by suggesting instead that nurses should have a basic knowledge of medical English. Clearly, she also recognizes that the level of bilingualism that she has attained after twelve years of living and working in an English environment is unattainable for the general population of nursing students considering the current level of language instruction available to them.

Though minority language access to healthcare as a universal principle received only weak support from four of the five nursing students, this anecdote is helpful in that it points to both a present lack of fitness of purpose in the VLP and a way to improve future prototypes. As the anecdote illustrates, Subject 1 recognized the value of using

English with an Anglophone patient when she witnessed the relief and happiness that her patient felt at being addressed in English. In terms of fitness of purpose, the implication is that this prototype lacks fitness to the degree that it fails to contextualize the conversation with the virtual minority language patient within a healthcare system that is reluctant or unable to provide health services in English. Danny, the virtual patient, does not indicate in any way that a lack of minority language access has silenced him or decreased his satisfaction with his healthcare. In this regard, the fitness of purpose of the VLP will be enhanced in future prototypes by simulating the relief and happiness that patients feel when addressed in a language they understand. For the current prototype, this could be achieved simply by substituting "Oh good! You speak English" for Danny's current response "Yup" when asked, "Are you here for the medical history interview?" He could also add, "I was really worried that you would only speak to me in French, like the last nurse who was in here." This might provide learners with insight into one source of dissatisfaction patients have with their healthcare.

Training Effects

Training Effect on Sense of Preparedness

To gauge the effect of the VLP on participants' confidence in their perceived ability to perform a real-life medical interview after using the VLP, differences between pre and post questionnaire responses were compared (see Table 8). Subject 1, the most highly proficient of the group in English, appears to have lost some confidence in her own ability to conduct a medical interview entirely in English by answering "somewhat confident" to the question "How confident would you be about communicating with a

patient entirely in English?" on the post after initially answering "very confident" on the pre questionnaire. Nevertheless, she remained "somewhat confident" that she could perform the interview with correct pronunciation and grammar.

Subject 2, the low-intermediate learner, also seems to be less certain about her interview skills in English after the simulation, having changed her answer to the question, "If necessary, could you interview a patient in English?" from "certainly, yes" to "probably, yes."

Subject 4, the other low-intermediate of the group, shows the greatest loss of confidence in her English interviewing skills. She also changed her answer from "certainly, yes" to "probably, yes" when asked if she could interview an English patient. More significantly, she changed her answer from "very confident" to "somewhat confident" when asked about conducting the entire interview in English and changed her answer of "somewhat confident" to "somewhat nervous" when asked about her ability to use correct pronunciation.

In contrast, subject 3, a false-beginner, seems to have gained slightly in confidence from her interaction with the VLP. On the pre questionnaire, she indicated that she would be "somewhat nervous" about her grammar and pronunciation. On the post, she indicated that she would be "very confident" in both of these areas.

Subject 5 remained almost equally uncertain about her English interviewing ability before and after using the system, indicating "I don't know" when asked if she thought she could conduct an interview with an English-speaking patient and "very nervous" about her grammar and pronunciation accuracy. The only difference between her pre and post questionnaire answers with respect to second language confidence was

where she changed her answer from initially indicating that she "would not be able to speak" in an interview conducted entirely in English to afterward indicating that she might be able to speak but would be "very nervous," a small gain in self-confidence.

Table 8

Post Questionnaire Confidence Results Showing Gains and Losses

	Participants				
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
# of rejections	7	26	55	49	56
Descriptor given for rejections	good	acceptable	good	Great	Good
Difficulty preference	advanced	advanced	intermediate	advanced	Intermediate
If necessary, could you interview a patient in English?	certainly yes	**probably yes (-)**	certainly yes	**probably yes (-)**	I don't know
How confident would you be about communicating with a patient entirely in English?	**somewhat confident (-)**	somewhat confident	very confident	**somewhat confident (-)**	**very nervous (+)**
How confident would you be about your ability to use correct English grammar?	somewhat confident	somewhat confident	**very confident (+)**	somewhat confident	very nervous
How confident would you be about your ability to use correct English pronunciation?	somewhat confident	somewhat confident	**very confident (+)**	**somewhat nervous (-)**	very nervous

These effects on participants' sense of their own preparedness to interact with English speaking patients suggest that the VLP provides a useful simulation of a medical history interview. Subject 1 remarked that the simulation was "really, really the real thing" though without "the same impact [on the patient] because it is like TV. [...] He is not going to get angry if you make a mistake." This fidelity to a real life encounter achieves a twofold effect. It causes more proficient learners to re-evaluate their perhaps slightly exaggerated sense of preparedness and less proficient learners (for whom talking to English speaking patients seemed an impossibility) to see the task as being slightly more achievable.

Training Effect on Speech Rate

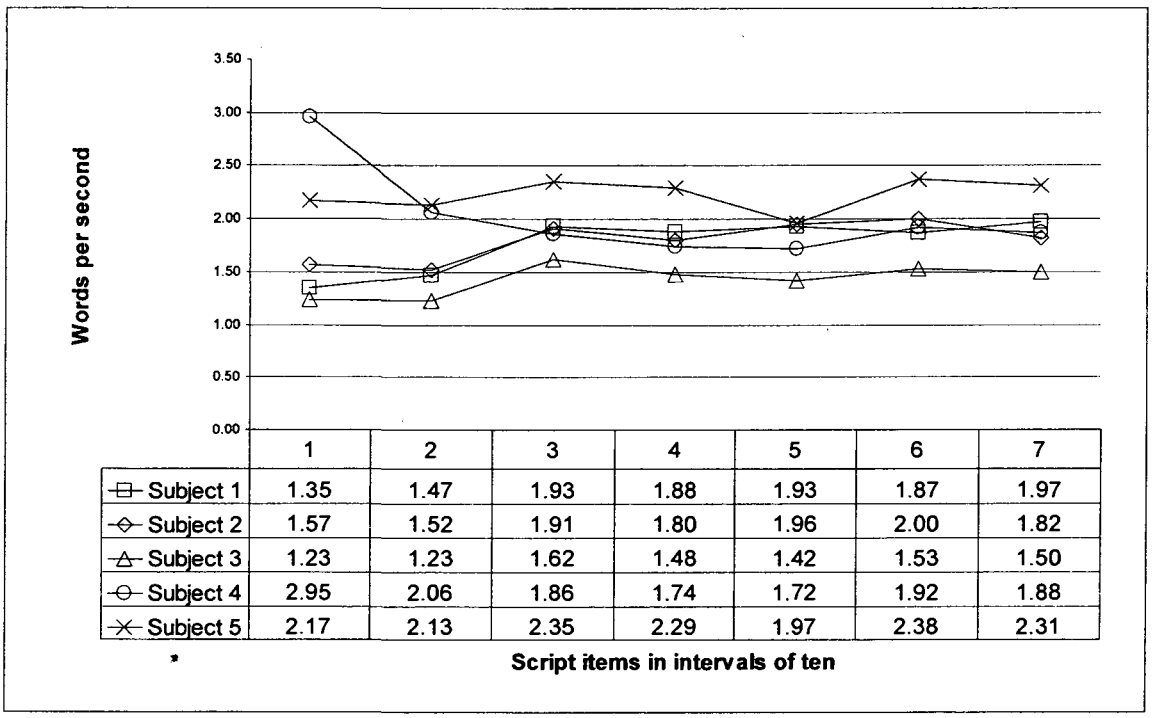
One of the intended purposes of the VLP is to provide learners with an opportunity to increase their English fluency while performing a medical history interview. Although it would be unrealistic to expect global changes in second language fluency after an intervention of only 40 minutes, subtle increases in speech rate, one indicator of fluency, would provide evidence that the VLP task-type is fit for fluency training.

Data from the system log file provided automatically generated utterance duration times for each participant's utterance, calculated automatically when the system detects the beginning and end of the user's utterance. The number of words in each script item was then divided by utterance duration times to produce a speech rate score, expressed as words per second. Speech rate trends for each participant were then calculated by averaging speech rate scores at intervals of ten.

Four of the five subjects' speech rates increased with practice (see Figure 7), suggesting a slight gain in one measure of oral fluency over the course of the dialogue. Secondly, as can be seen in Figure 7, the initial variability in each of the subjects' speech rate at the beginning of the dialogue tends to stabilize and flatten out by the end of the dialogue. While encouraging, this general increase in speech rate is not meaningful unless there is also an increase in pronunciation accuracy. Faster, less accurate speech is hardly a desired outcome. Also, the decrease in Subject 4's speech rate over the course of the dialogue needs to be explained in the context of the general increase in the rates of the other four.

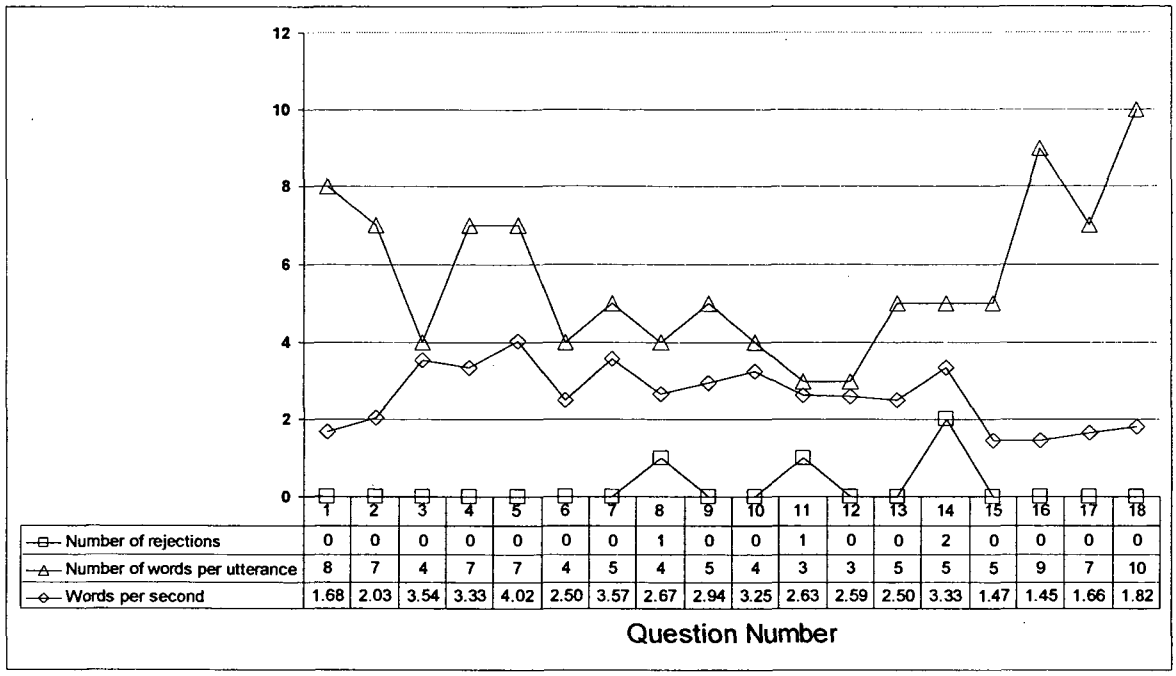
Subject 4 stands out from the other four participants in two important respects. First, while the other participants' speech rates increase, her speech rate drops with practice. Second, unlike the other participants who have little trouble at the beginner difficulty setting, she has a rejection rate of 20% (see Figure 4), a rate four times higher than the next highest rejection rate of 5%. A closer look at her initial speech rate and recognizer rejections (see Figure 8) provides clues as to why this may be the case. From the outset, her speech rate is significantly higher than the other participants' but then drops sharply after encountering two rejections for the same interview question. Although, she produced her third attempt of question 14 at a rate of 3.33 words per second, a rate higher than the average score of the preceding fourteen questions (2.87 w/s), she reduces her average speech rate on the next four questions to half the speed (1.60 w/s) .

One reason for Subject 4 pronouncing subsequent script items more slowly could be attributed to an intervening factor such as utterance length. This can be ruled out since script items #15 to 18 increase in length, and it has been shown that speech rate generally increases when utterance length increases (Yuan, Liberman, & Cieri, 2006). In the context of a general increase in speech rate by the other four participants over the course of the medical history interview, Subject 4's decrease in speech rate is most likely the result of a compensatory strategy to improve the system's recognition rate after a series of early rejections. The effect of the VLP's feedback on this learner is slower, more careful speech.



	1	2	3	4	5	6	7
Subject 1	1.35	1.47	1.93	1.88	1.93	1.87	1.97
Subject 2	1.57	1.52	1.91	1.80	1.96	2.00	1.82
Subject 3	1.23	1.23	1.62	1.48	1.42	1.53	1.50
Subject 4	2.95	2.06	1.86	1.74	1.72	1.92	1.88
Subject 5	2.17	2.13	2.35	2.29	1.97	2.38	2.31

Figure 7. Speech rate trends by subject averaged every ten script items.



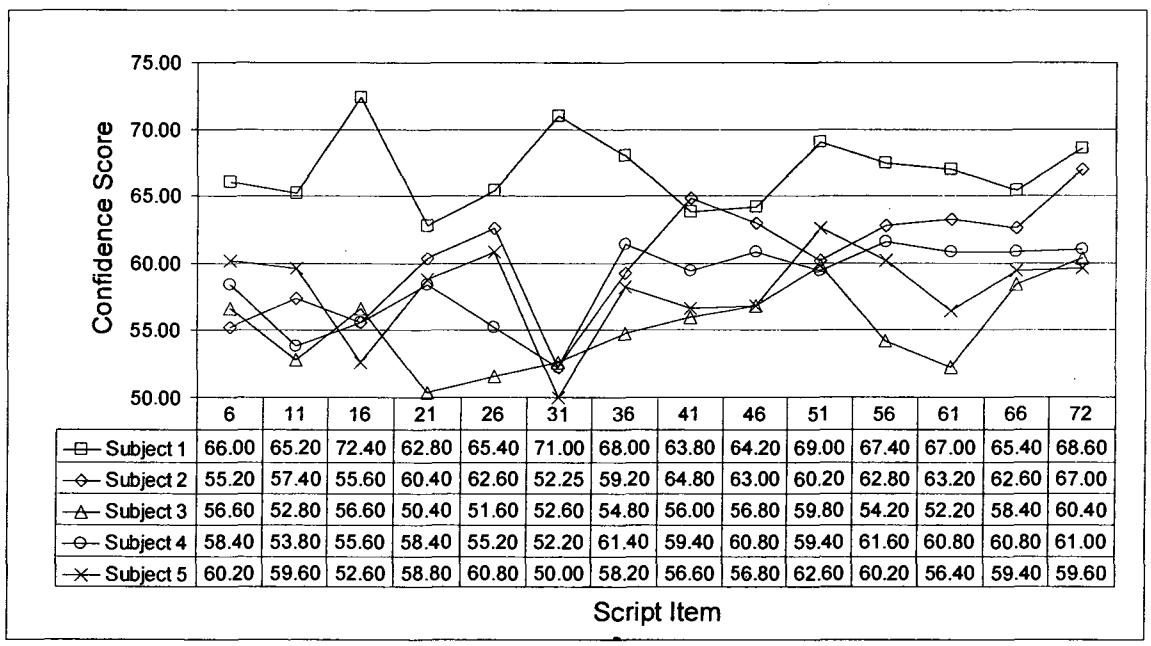
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Number of rejections	0	0	0	0	0	0	0	1	0	0	1	0	0	2	0	0	0	0
Number of words per utterance	8	7	4	7	7	4	5	4	5	4	3	3	5	5	5	9	7	10
Words per second	1.68	2.03	3.54	3.33	4.02	2.50	3.57	2.67	2.94	3.25	2.63	2.59	2.50	3.33	1.47	1.45	1.66	1.82

Figure 8. Speech rate, utterance length, and number of rejections for Subject 4

In this way, the VLP appears to have a standardizing effect on the group, narrowing the range of variability between speakers. The slowest of the group at the beginning of the virtual dialogue was Subject 3 with a speech rate of 1.23 words per second. The fastest was Subject 4 with 2.95 words per second, a difference of 1.72 words per second. By the end of the dialogue, the slowest is still Subject 3 with an increased speech rate of 1.50 words per second, but the fastest of the group is now Subject 5 with 2.31 words per second, a difference of 0.81 words per second. What this all means is that the VLP could reduce the variability between speakers, making fast talkers more careful and slow talkers more fluent.

Training Effect on Accuracy

Apart from fluency increases, another goal for the VLP is that it should improve the accuracy of learners' speech while they are performing a medical history interview. To investigate a training effect, changes in pronunciation accuracy were quantified in terms of the degree and number of improvements. First, the degree of improvement for each participant was determined by averaging recognizer confidence scores at increments over the course of the virtual dialogue. Next, the number of specific improvements to pronunciation made over the course of the dialogue was determined by identifying the number of successful repairs from the log file data.



	6	11	16	21	26	31	36	41	46	51	56	61	66	72
Subject 1	66.00	65.20	72.40	62.80	65.40	71.00	68.00	63.80	64.20	69.00	67.40	67.00	65.40	68.60
Subject 2	55.20	57.40	55.60	60.40	62.60	52.25	59.20	64.80	63.00	60.20	62.80	63.20	62.60	67.00
Subject 3	56.60	52.80	56.60	50.40	51.60	52.60	54.80	56.00	56.80	59.80	54.20	52.20	58.40	60.40
Subject 4	58.40	53.80	55.60	58.40	55.20	52.20	61.40	59.40	60.80	59.40	61.60	60.80	60.80	61.00
Subject 5	60.20	59.60	52.60	58.80	60.80	50.00	58.20	56.60	56.80	62.60	60.20	56.40	59.40	59.60

Figure 9. Recognizer confidence score trends averaged every five items.

On average, recognizer confidence scores increased over the course of the virtual dialogue from 59.28 to 63.32, an increase of 4.04 points or 6.8% (see Figure 9).

Individually, confidence scores increased for four of the five participants. The largest gain was seen in Subject 2's score with a 21.37% increase. Subject 5 saw no increase in recognizer confidence scores by the end of the dialogue, holding steady with 59.6 after starting with 60.2.

The gains can be attributed in part to the effect of incremental increases of the confidence rejection threshold through the researchers manipulation of the difficulty settings. To explain, at the beginning of the dialogue (script items from #1 to 19), only utterances with scores of $\leq$34 were rejected; in the middle of the dialogue (script items from #20 to 39), only utterances with scores of $\leq$44 were rejected; and by the end only utterances with scores of $\leq$54 were rejected. As the rejection threshold increased, participants were required to meet higher standards of accuracy. With repair, participants succeeded in meeting the higher thresholds, and so confidence scores from successful attempts tend to show an increase. However, rising recognizer confidence score averages do not tell the whole story.

Increases in pronunciation accuracy can also be quantified in terms of the number of improvements in pronunciation accuracy made by each participant over the course of the virtual dialogue. A successful repair, indicated by an utterance acceptance after one or more rejections, represents a specific instance of increase in pronunciation accuracy. As such, all of the participants improved the accuracy of their pronunciation of medical history interview questions to some extent (see Figure 10). Subject 4 made the greatest number of improvements to her pronunciation with 22 successful repairs, while Subjects 5, 3, and 2 made 19, 17 and 15 successful repairs respectively. Subject 1, by this measure, made the fewest gains in pronunciation accuracy ($n = 4$). A caveat is offered here since

this measure does not indicate the amount of the increase in accuracy with each successful repair. Where the difference between the utterance rejection and the utterance acceptance scores was small, the increase in pronunciation accuracy may be negligible.

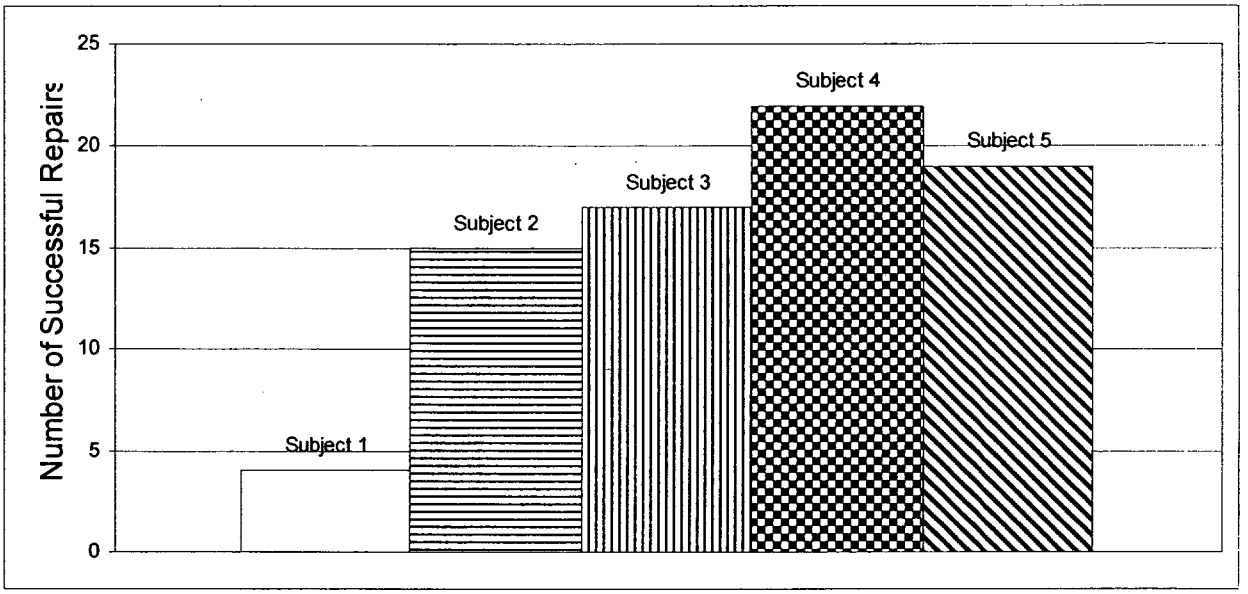


Figure 10. The number of successful repairs made over the course of the virtual dialogue for each participant.

Summary of Training Effects on Accuracy

In summary, there appears to be some evidence for a positive training effect from practicing orally with the VLP on participants' pronunciation accuracy. By the end of the dialogue, participants saw either an increase in their pronunciation accuracy in terms of average recognizer confidence scores, or an increase in the number of medical history interview questions each could pronounce correctly, or an increase in both.

Learners' Perspectives

In the two sections above on learner needs and training effects, evidence was given for the fitness of purpose of the present prototype VLP system. In the following section, suggestions made by the study's participants for enhancing the system's fitness of purpose will be given. These comments were collected from responses provided on the post questionnaire as well as suggestions made during the post interview.

Some of the suggestions for improvement from the post interview have already been mentioned. For instance, Subject 4 suggested that audio pronunciation models with a slower speech rate should be added. Subjects 1, 2 and 4 each mentioned that the medical history interview questionnaire needed more space to write down the virtual patient's answers. Subject 4 suggested that there needs to be a way to "rewind" the virtual patient's answers with an oral command. Subject 3 suggested a repeat button.

Other suggestions for improvements made by participants are as follows. Subject 4 suggested that the graphical feedback panel could be larger. She said: "Maybe put that bigger (pointing). I think it is really important, but I didn't use it as much as I should." While changes to the GUI of this sort do not sound too difficult to achieve, this raises the question whether graphical feedback is really all that useful to learners. Subject 2 suggests that it is, but perhaps learners would value conversational types of feedback more. Verification requests could be filmed where the learner is asked, "Did you say X or Y?" Otherwise, more focussed clarification requests than "Huh?" could be triggered such as "I didn't hear the second to last word in your sentence." This issue will be taken up further in the following chapter.

Subject 2 suggested that the system should increase the pronunciation challenge of the dialogue as the learner progresses through it. She said: "Changes? No, just that the more that people go, the more harder the questions should be." The learner can already achieve this manually to some degree by selecting either the intermediate or advanced difficulty setting for him or herself. However, what Subject 4 is probably getting at is for the system to increase the difficulty setting automatically through a learner-adaptive method of automatically setting the recognizer confidence threshold. For example, the dialogue could begin with a default confidence threshold of 35, and every successful utterance recognition thereafter would raise the confidence threshold by increments of two (i.e. 35, 37, 39, 41, 43, 45, etc.). This could go on until the learner produces an utterance that scores below the threshold and gets rejected. At which point, the system could automatically reduce the threshold by three (i.e. 45, 42, 39, etc.) for each successive utterance rejection until the learner produces an utterance that meets or surpasses the recognizer confidence rejection threshold again. Such a method of automatically adjusting the difficulty setting up or down would ensure that the system always poses a challenge for any learner who uses it.

Subject 2 also said that script items that allow the nurse to ask the virtual patient for extra details should be added to the dialogue. She said, "I would have asked him 'How do you spell your medication Cozar?' because I didn't know how it is, how it was spelled and maybe also the frequency that he takes it, like everyday, 2 mg per day." Follow-up questions of this nature can be added to future modifications of the VLP and with future virtual dialogues.

On the post questionnaire (see Appendix G), participants were asked to select from a list of virtual patient characteristics they would like to see in future prototype systems. All participants responded that they thought that future versions of the VLP should include patients of different ages with a variety of medical problems. Three of the five participants also indicated an interest in virtual dialogues with patients with a variety of English and foreign accents (see Table 9).

Table 9

Suggested Characteristics for Future Virtual Patients

	Participants				
	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5
Different medical problems	✓	✓	✓	✓	✓
Different English Accents	✓		✓	✓	
Different foreign accents	✓		✓	✓	
Different ages	✓	✓	✓	✓	✓
Include aggressive and uncooperative patients	✓		✓	✓	✓

Subject 5 made two interesting suggestions for alternate virtual dialogues to the present prototype. First, she said there is considerable need for Spanish language health services in her area, something already mentioned above. Secondly, Subject 5 said that the present virtual dialogue is of only limited usefulness in that the VLP provides practice asking questions that are restricted to the short admission process. Of more vital importance is the need for help conducting a pain assessment interview. She explained her reasons in the following way. A translation follows.

> Là, c'est sure que c'est faite pour quand on va faire une admission pour quelqu'un que je rencontre la première fois. Ça c'est très bref et ce n'est pas là qu'on rentre le plus en contact avec la personne anglophone. On remplit son app., et c'est fini là. Après ça, on va aller plus au niveau de soins, l'évaluation, la douleur. C'est-ce qui est le plus important chez la personne pour voir comment bien évaluer la douleur, pour pouvoir bien soulager la personne. Il y en a beaucoup de problème respiratoire à l'hôpital puis des problèmes cardiaques. Après ça, il y en a des fractures et des problèmes circulatoires. C'est la douleur qui serait vraiment pertinent parce que c'est quelque chose ou on va avoir les difficultés. [So, this is clearly made for when we do an admission for someone I meet for the first time. That's very brief and it is not where we have the most contact with Anglophones. We fill in their admission papers, and it ends there. After that, we begin care, their health assessment, their pain. That's the most important for the person, to see how to assess their pain, to give proper relief. There are a lot of respiratory problems at the hospital and heart problems. After

that, it's fractures and circulatory problems. It's pain that is really important because it is something we have difficulties with.]

As the above quotation illustrates, the fitness of purpose of the VLP will depend largely upon the language learning and professional priorities of the learner. For Subject 5, this prototype prepares the learner for an administrative task which she feels is of trifling importance in comparison to the more pressing need for help using English to administer care and to relieve the patient's suffering. Her point is well made. It will be worthwhile therefore to continue to include learners in the selection and design of language learning materials that reflect their priorities to support them in their efforts to provide healthcare access to linguistic minorities.

Summary of Fitness of Purpose Results

In summary, the VLP appears to have adequately anticipated some of the English language training needs of these learners by targeting face-to-face oral interaction with an English speaking patient with a focus on pronunciation training. However, the VLP's fitness of purpose could be further enhanced by dramatizing the negative consequences for patients of not receiving healthcare services in their preferred language. The VLP gave learners a more realistic view of the demands of the medical history task, causing participants to re-evaluate their sense of their own preparedness. Learners were quicker, more accurate and more careful in their pronunciation of medical history interview questions at the end of the virtual dialogue. Overall, these training effects suggest a degree of fitness of purpose that is encouraging. However, further enhancements to the VLP will be best achieved by involving the learner in the development of language learning materials that reflect their language learning and professional priorities.

CHAPTER 6: DISCUSSION AND CONCLUSIONS

Summary

The purpose of this research effort was to design and assess a prototype computer assisted language learning tool for medical professionals learning spoken English. At the outset, the preliminary goal was to find a suitable pairing of pedagogy and technology that could serve the cause of improving access to healthcare for linguistic minorities through language training. A critical review of available literature concerning applications of ASR for second language instruction identified the virtual dialogue (as described in Harless et al., 1999) as being one pairing that could provide learners with opportunities for meaningful oral repetitions with a receptive interlocutor. Looking at healthcare communication literature for suitable task types to model, the medical history interview was selected for its ubiquity, length, lexical range, and importance as a language-dependent diagnostic tool. The prototype VLP that emerged was the result of the choice to develop a virtual medical history interview with a focus on English pronunciation.

The VLP system has a straightforward design. Learners of medical English are charged with the familiar task of taking a patient's medical history. The learner sits in front of a laptop computer wearing a headset and is given a paper-based medical history questionnaire. In the center of the computer screen, there is a window playing a video of a male patient in his forties looking around. Under the video, there is a medical history question in English next to a button labelled *recognize*. To activate the system, the learner simply clicks the *recognize* button, and reads the question. If the learner does not

pronounce the question accurately, a video clip is triggered with the patient asking the learner to try again. Help is provided to the learner in the form of another clickable button that activates an audio recording of a native speaker reading the question at a natural rate. If the learner subsequently pronounces the question accurately, a video clip plays with the patient providing the requested information and the next in the series of medical history questions appears. The learner can then use the information to fill in the paper-based questionnaire and also view visual representations of his or her accuracy scores for each word in the last question asked. This simple procedure repeats 72 times, until all of the questions have been successfully asked and the virtual patient's information has been recorded on the paper-based questionnaire.

The practical goal of the pilot study was to provide a preliminary assessment of the prototype's suitability as a learning tool for medical professionals. Five Francophone nursing students were asked to try out the system at three difficulty settings and provide feedback. The qualitative and quantitative data collected during these trials were then analyzed in terms of the VLP's ease of operability and its fitness of purpose in order to make a recommendation to either abandon this line of development, or pursue it further.

Results of the study indicate that participants found the system to be generally easy to use and well-designed, although certain modifications could enhance its ease of operability. Two issues came up with the native speaker's pronunciation models. Participants had trouble perceiving the exact pronunciation of individual words within the recordings due to the natural speech rate used by the native speaker. To resolve this perceptual difficulty, additional pronunciation models of individual content words (i.e., words such as *frequent* + *headaches* spoken outside of the phonological environment of

the questions in which they appear) or a model of the entire sentence pronounced more slowly and clearly could be recorded and made available as an additional resource for learners to use. If constraints upon resources prevent rerecording every script item in this way, then rerecording only those script items that trigger the most rejections seems a reasonable alternative.

The second problem discovered with the recorded pronunciation models involved a mismatch between the pronunciation of "Do you?" in the audio model and the recognizer's dictionary entry for this question formula. The native speaker had reduced her vowels as is commonly done in everyday speech, but the system's internal dictionary did not contain reduced forms for these words. Consequently, the more closely the participants emulated the recorded model, the lower their accuracy score. This scoring error caused added frustration and proved counterproductive since the negative feedback on the first two words of the sentence detracted attention from correction of pronunciation errors in key content words at the end of the question. While it is encouraging to see that learners use the graphical feedback to locate pronunciation errors in the sentence, a careful review of the function words in the internal dictionary would help to eliminate future scoring errors of this nature. A choice will have to be made either to add reduced forms of function words in the script to a user defined dictionary as alternative acceptable pronunciations (i.e., /duw/ or /də/ + /juw/ or /jə/) or replace the non-reduced forms of individual function words with phrase-as-word dictionary entries (i.e., /dəjə/) to ensure that reduced forms are consistently scored higher than unreduced forms.

Other factors contributing unnecessarily to task difficulty were uncovered. Participants complained that there was not enough space on the paper-based medical history questionnaire to fill in all of the information the virtual patient provided. This would seem a small matter as fewer columns and wider margins would settle it, but the problem of deciding where to put the information made it more difficult for participants to remember what to put, highlighting another issue. There was no apparent way to get the virtual patient to repeat his answer. The suggestion to add buttons or voice commands (i.e., additional script prompts to get the patient to repeat, explain, or slow down) therefore came up. From a language practice point of view the inclusion of additional script prompts seems a better choice than buttons. From a design point of view, adding script prompts raises the question of what other kinds of prompts could be added and to what effect.

Before getting to that, a few points should be made about the system's difficulty settings. First, there was a clear correlation between starting proficiency and the total number of utterance rejections each participant ultimately received. As expected, lower proficiency learners received more utterance rejections than higher proficiency learners. Second, the three settings (beginner, intermediate and advanced levels) proved to be effective at managing task difficulty in terms of triggering more utterance rejections and feedback on pronunciation accuracy. Third, results point to an optimal ratio of utterance rejections to script items. By choosing a preferred difficulty setting for themselves, participants indirectly indicated a preference for about one to about one and a half (1.15-1.44) utterance rejections per medical history question. Presumably, fewer would make the task too easy, and more would make it too frustrating. A learner adaptive method of

regulating the system's difficulty setting may prove to be the best way to maintain this balance of task challenge to task frustration.

The VLP fared well in terms of fitness of purpose. The system largely anticipated the learners' needs in terms of task type, mode of interaction, type of interlocutor, and focus on pronunciation. Furthermore, three training effects on the participants were observed. The first observable training effect was that the experience of taking the virtual patient's medical history seems to have made the more proficient learners of the group view the task of interviewing English-speaking patients as being a little more demanding than they had previously anticipated, whereas the less proficient learners came to view such interviews as being somewhat more achievable. This speaks well of the system's level of realism.

The second observable training effect on participants was that four of the five learners were more fluent in their pronunciation of medical history questions by the end of the virtual dialogue, at least in terms of speech rate. For one of the participants, a decrease in the rate of her speech was observed after she received a number of utterance rejections early in the virtual dialogue. She subsequently pronounced the remaining medical history questions more slowly and more accurately to accommodate the communication needs of her virtual interlocutor, demonstrating a kind of careful-speech effect.

The third training effect observed in the study was an increase in pronunciation accuracy. Participants corrected their pronunciation of between 6% and 30% of the medical history questions in the script through post-feedback repair. This effect is attributable to the negative feedback the system provided each time it judged an utterance

to be mispronounced. Participants were told (sometimes repeatedly) to try again until the system's standard of pronunciation accuracy was met. They did, with success.

An increase in average pronunciation accuracy scores was also seen in four of the five participants. In short, the average system-generated accuracy scores by the end of the virtual dialogue increased for participants by up to 21%. However, this effect should probably be attributed to the researcher's manipulation of the system's difficulty settings during trials. By raising the difficulty settings twice over the course of the dialogue, learners were in effect being held to increasingly higher standards of pronunciation accuracy. This is not therefore strictly an effect of the design of the VLP system but an effect of the design of the study since learners using the system on their own without the supervision of the researcher are in no way required to increase the difficulty setting during the virtual dialogue. Nevertheless, this is a fortuitous finding since it suggests that programming the system to automatically increase the pronunciation accuracy standards could achieve a similar effect.

Looking for a Suitable Pairing of Pedagogy and Technology

These preliminary results suggest that this incarnation of the talking machine might indeed be worth the trouble involved—to borrow a phrase from Chapter Two. Indeed, it appears that with a few minor modifications this VLP prototype could make a suitable pronunciation activity to occupy French-speaking nursing students for a few hours. Spending much more time with this particular virtual patient is not recommended since the pedagogy that informs the VLP calls for task repetition with novel interlocutors and not strict task duplication with the same interlocutor. Whatever becomes of this

particular prototype, however, the approach to pronunciation training that the VLP employs has general advantages over earlier forms of computer assisted pronunciation training that make it likely to endure.

In no small way, the virtual dialogue approach to pronunciation training represents a significant departure from earlier, non-dialogic pairings of language pedagogy and ASR. Up until recently, the usual approach to computer-assisted-pronunciation-training (CAPT) has been a non-dialogic pairing of technology and pedagogy that follows a listen-repeat-feedback (hereafter LRF) sequence of human-computer exchanges. The machine initiates the interaction by playing an audio model for the learner to listen to, repeat, and then receive feedback on. This sequence of moves, though well-intentioned, has been sending a subtle, unspoken message to "shut-up and listen," silencing the learner, dismissing prior pronunciation knowledge of the target language and emphasizing learner receptivity and passivity. The machine speaks first, it assumes no prior knowledge of how to pronounce a target sentence, and the learner must listen attentively and follow the machine's lead. In contrast, the virtual dialogue's hypothesize-feedback-communicate approach (hereafter HFC) begins by inviting the learner to test a pre-existing hypothesis about the pronunciation of a sentence. The learner speaks first, and more appropriately, the system remains passive and receptive.

Another difference between the two approaches is their respective goals for the interaction. In the traditional LRF approach to pronunciation training, the goal of speaking to the machine was to get explicit, usually graphical feedback on form. In contrast, with the virtual dialogue HFC approach, the goal is to communicate, and feedback is provided as a means to that end. Implicit negative and positive feedback is

provided in the form of communicative success or failure: the virtual interlocutor either "understands" or asks for clarification. In contrast to LRF, with HFC, negative feedback prompts repair as a necessary step toward successful communication. Despite considerable efforts by materials developers from Léon and Martin (1972) to the present, receiving feedback in the form of intonation curves, waveforms and accuracy meters may not be such an appropriate goal for learners after all. As noted in Chapter Two, learners sometimes face difficulties interpreting graphical feedback (Wildner, 2002), and in this present study we see that at least two of the five learners (Subjects 1 and 4) were content to ignore the accuracy meters altogether, relying instead on conversational cues from their virtual interlocutor for feedback.

Perhaps, graphical feedback can be done away with entirely. If, for example, a medical simulation actor were to be employed in the development of the next in the series of medical history interviews where his or her lines were carefully scripted, all feedback on pronunciation accuracy could potentially be provided conversationally. Instead of an accuracy-meter showing which words need the most repair, video clips containing conversational verification phrases could be employed in their place. When, for instance, the question "Do you ever get seizures?" falls below the confidence rejection threshold, the virtual patient could ask "Did you say seizures or scissors?" or even more simply "Did you say seizures?" A verification procedure of this type provides the added benefits of identifying the word that contained the pronunciation error, prioritizing its repair over other words in the question, and modelling its correct pronunciation. Instead of graphical feedback, clarification requests could also be linked to specific ranges of accuracy so that very low-scoring attempts could trigger phrases like, "I have no idea what you just said,"

and slightly below threshold attempts could trigger, "I almost understood you there, so please try again." Explicit, detailed feedback that indicates a low recognizer confidence score, or identifies the location of the error, or communicates the severity of the error in a conversational manner would help learners stay within the bounds of the oral interaction, un-distracted by colourful lights and bar graphs that might otherwise vie for their attention. It seems unlikely that learners would continue to claim an appetite for graphical feedback once conversational feedback on form from a virtual agent has been more fully developed.

As an aside, though verification procedures have been used before in ASR for CALL (see Wachowicz & Scott, (1999), nowhere, it seems, has anyone ever tried irony as an alternative form of feedback with ASR applications. Readers of Greek philosophy will remember how Socrates pretended not to know in order to help his students discover the errors in their reasoning (now called Socratic irony). Likewise, language teachers sometimes pretend not to understand in order to help learners discover errors in their production. Continuing with the example above, if the system detected an error in the learner's pronunciation of the word *seizures*, the virtual patient could reply "I don't need scissors. I have a sharp pair at home. Why do you ask?" Similarly, scripted misinterpretations of utterances with missing morphology in the pronunciation of past forms (*want/wanted*) or plural forms (*lens/lenses*) could trigger ironic conversational forms of feedback that would help to emphasize the sound-meaning connection that is so often overlooked in pronunciation teaching. Conversational irony could thereby encourage the fine control that multiple-choice minimal-pair activities have targeted in earlier ASR-CALL applications (i.e., LaRocca et al., 1999).

The third important difference between the LRF and the HFC approaches to pronunciation training relates to the demands each places on the learner's attention. The LRF approach usually focuses entirely upon form, leaving the meaning of a target utterance as an afterthought at best. Ostensibly, the reason for this was to help learners to concentrate on meeting pronunciation challenges without the challenges of comprehending the utterance competing for their attention. The HFC approach, in contrast, requires the learner to manage cognitive resources more dynamically. By embedding pronunciation training within meaningful communication, the HFC approach requires the learner to repeatedly switch his or her attention from form in the output to meaning in the input. This is key. When the learner endeavours to correctly pronounce a question in the furtherance of a communicative task with the expectation of receiving a meaningful and appropriate answer to the question, the learner gains practice in selectively and appropriately attending to aspects of form and meaning in response to the demands of situation. This kind of attention-switching practice is completely neglected by language teaching materials that require the learner simply to listen and repeat.

This is why current thinking on good pedagogy requires that all oral repetitions must occur in a genuinely communicative context and where each formulation of a repeated structure is part of a meaningful message conveyed to a receptive interlocutor (Gatbonton & Segalowitz, 2005). For a context to be genuinely communicative, the message must be meaningful. In a medical interview, real or virtual, the message will be meaningful whenever it works to establish a trusting and supportive relationship between healthcare provider and patient, it gathers needed health information about the patient, or it entails sharing new health information with the patient.

Good pedagogy will not be served, therefore, by repeating the virtual dialogue with Danny (our virtual patient) past the point where the messages exchanged in the virtual dialogue stop being meaningful. During the study, the five participants were asked to say how many times they would like to practice the medical interview with the VLP before they would want to stop. Subject 1 said once, subjects 2 and 3 said twice, and subjects 4 and 5 said five times would be enough. Perhaps, the low-intermediate learners and the false-beginners expect to understand more by listening to Danny's answers again, but ideally learners would benefit more from the opportunity to repeat the medical interview task with other virtual patients.

While strict task duplication might increase the comprehensibility of a virtual patient's answers for lower-proficiency learners, real-life patients are less likely to be so generous with their time. Rather, the addition of question prompts to the GUI to get the virtual patient to slow down, repeat or explain could have far reaching benefits. Apart from simply adding to the range of questions available to the learner for pronunciation practice, these clarification requests are transferable to real-life encounters with the potential to improve the comprehensibility of the available linguistic input. Questions such as "What does ____ mean?" and "Could you give me an example?" would provide a level of conversational control over the comprehensibility of aural input that has been absent from earlier language learning materials.

Adding clarification requests to the available prompts of the VLP represents an opportunity for the learner to engage in only minor digressions in an otherwise linear conversation. In contrast, the Conversim™ system described in Harless et al. (1999) used multiple prompts in a different way. At decision points within their virtual dialogues,

choosing to ask one question instead of the others allowed for greater conversational digressions and led to story branching. This innovation meant that revisiting a virtual dialogue and choosing to ask a different question at that decision point might uncover new information from a virtual prisoner about his personal history. Since learners were tasked with making a decision to trust the prisoner or not, returning to the virtual dialogue to explore aspects of the conversation not covered previously helped the learner to make the right decision. For medical interviews, the burden of trust seems to be the other way around. It is the provider who needs to win the patient's trust in order to gather the intimate details of the patient's health history and gain a commitment of compliance/adherence to treatments.

Providing additional prompts that allow digressions into small talk in a virtual medical history interview has the potential of helping learners develop their skill in pragmatic aspects of second language medical communication that goes beyond comprehension and pronunciation training. It is known that social talk is often missing from second language medical encounters even though it serves the important purpose of building rapport with patients, increasing trust, and consequently leading to a more comprehensive patient history containing more detailed answers from patients (Aranguri et al., 2006). Future virtual medical interviews could provide the learner with a variety of prompt choices that include small-talk questions to invite the virtual patient to develop a more trusting relationship. Choosing to ask those small-talk questions could lead to a branch of the virtual dialogue containing health information omitted from exchanges not prefaced with social talk, thereby helping the learner to make the correct diagnosis and develop important second language communication skills in the process.

Other aspects of pronunciation, 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. Along the way, it could be that graphical feedback will disappear if conversational 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. 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.

While good language pedagogy requires meaningful repetitions of target structures in a communicative context, better language pedagogy may require an approach to language instruction with a greater emphasis on learner motivation. In this study, we saw how nursing students were careful not to suggest that all nurses be fully bilingual, cognizant perhaps of the motivational intensity and long hours required to master a second language in adulthood. Nursing students, it seems, just do not have the time to devote to language learning, or do they? Luis von Ahn, a researcher of human computation at Carnegie Mellon University, approximates that in 2003 alone 6 billion human hours of computerized solitaire were played for fun (von Ahn, 2006). He notes that in comparison, it took only 7 million human hours to build the Empire State Building. While von Ahn's research is directed toward capturing some of these hours of game playing to solve large-scale computational problems, the motivation to play coupled with the good pedagogy of virtual dialogues may be just what is needed to solve

large-scale problems in second language healthcare communication. If some of the time that healthcare professionals freely devote to computer game playing could be redirected toward highly social, absorbing, goal-oriented, competitive, skill-building, ego-gratifying language learning fun, solving the problem of healthcare access for linguistic minorities might be less of a challenge than it appears today.

The potential of the virtual dialogue in this regard is enormous, and so it is not difficult to get carried away. Nevertheless, it remains difficult to see today what future reviewers will make of these preliminary and exploratory efforts to match good pedagogy with speech recognition technology. At minimum the virtual dialogue method appears ready today to provide an interesting contrast to earlier approaches to CAPT and worthy of further research.

REFERENCES

Araiza, R., Klopf, M. I., & Kelley, F. J. (2005). Forces of change. Nurse practitioner education and second language learning: using the Internet as a tool. *Clinical Excellence for Nurse Practitioners, 9*(1), 41-47.

Aranguri, C., Davidson, B., & Ramirez, R. (2006). Patterns of Communication through Interpreters: A Detailed Sociolinguistic Analysis. *Journal of General Internal Medicine, 21*(6), 623-629.

Bachman, J. W. (2003). The Patient-Computer Interview: A Neglected Tool That Can Aid the Clinician. *Mayo Clinic Proceedings, 78*(1), 67. doi: Article.

Baker, L. H., O'Connell, D., & Platt, F. W. (2005). "What Else?" Setting the Agenda for the Clinical Interview. *Annals of Internal Medicine, 143*(10), 766-770. doi: Article.

Barrier, P. A., Li, J. T. C., & Jensen, N. M. (2003). Two Words to Improve Physician-Patient Communication: What Else? *Mayo Clinic Proceedings, 78*(2), 211. doi: Article.

Bernstein, J., Najmi, A., & Ehsani, F. (1999). Subarashii: Encounters in Japanese Spoken Language Education. *CALICO Journal, 16*(3), 361-384.

Bickley , L. S. (2007). *Bates' guide to physical examination and history taking.* (pp. xxv, 902 p. : ill. + 1 CD-ROM (4 3/4 in.).). Philadelphia: Lippincott Williams & Wilkins, c2007. Retrieved from http://catalogue.mcgill.ca/F/JTN1FDQAHKMNAE9R82J85V6GPRX5SK5GUQCUQRMD4YFLB1PKEY-72547?func=item-global&doc_library=MGU01&doc_number=002266670&year=&volume=&sub_library=HLTH.

Cameron, R., & Williams, J. (1997). Sentence to ten cents: a case study of relevance and communicative success in nonnative-native speaker interactions in a medical setting. *Applied Linguistics, 18*(4), 415-445.

Carrasquillo, O., Orav, E. J., Brennan, T. A., & Burstin, H. R. (1999). Impact of language barriers on patient satisfaction in an emergency department. *Journal of General Internal Medicine, 14*(2), 82-87.

Chaikoolvatana, A., & Goodyer, L. (2003). Evaluation of a multimedia case-history simulation program for pharmacy students. *American Journal of Pharmaceutical Education, 67*(1), 108-115.

Charles, C., Goldsmith, L. J., Chambers, L., & Haynes, R. B. G. (1996). Provider--Patient Communication Among Elderly and Nonelderly Patients in Canadian Hospitals: A National Survey. *Health Communication, 8*(3), 281-302.

Charles-Jones, H., Latimer, J., & May, C. (2003). Transforming general practice: the redistribution of medical work in primary care. *Sociology of Health & Illness, 25*(1), 71-92.

Clarke, C. (1918). The phonograph in modern language teaching. *The Modern Language Journal, 3*(3), 116-122.

Coleman, J. (2005). *Introducing Speech and Language Processing*. Cambridge University Press.

Collins, S. (2005). Explanations in consultations: the combined effectiveness of doctors' and nurses' communication with patients. *Medical Education, 39*(8), 785-796.

Coniam, D. (1998). The Use of Speech Recognition Software as an English Language Oral Assessment Instrument: An Exploratory Study. *CALICO Journal, 15*(4), 7-23.

Dalby, J., & Kewley-Port, D. (1999). Explicit Pronunciation Training Using Automatic Speech Recognition Technology. *CALICO Journal, 16*(3). Retrieved from http://project5.nurseenglish.com/files/update12.zip.

Dörnyei, Z. (2001). *Teaching and researching motivation* (p. 295). Harlow: Longman.

Drew, P., Chatwin, J., & Collins, S. (2001). Conversation analysis: a method for research into interactions between patients and health-care professionals. *Health Expectations*, *4*(1), 58-70.

Drouin, J., & Rivet, C. (2003). Training medical students to communicate with a linguistic minority group. *Acad Med*, *78*(6), 599-604.

Dugaw Jr., J. E., Civello, K., Chuinard, C., & Jones, G. N. (2000). Will Patients Use a Computer to Give a Medical History? *Journal of Family Practice*, *49*(10), 921. doi: Article.

Eggly, S., Musial, J., & Smulowitz, J. (1999). Research and Discussion Note The Relationship between English Language Proficiency and Success as a Medical Resident. *English for Specific Purposes*, *18*(2), 201-208.

Ehsani, F., Bernstein, J., & Najmi, A. (2000). An interactive dialog system for learning Japanese. *Speech Communication*, *30*(2-3), 167-177.

Ehsani, F., & Knodt, E. (1998). Speech technology in computer-aided language learning: strengths and weaknesses of a new call paradigm. *Language Learning & Technology* , *2*(1), 54-73.

Eskenazi, M. (1999). Using a Computer in Foreign Language Pronunciation Training: What Advantages? *CALICO Journal*, *16*(3), 447-469.

Fernandez, A., Schillinger, D., Grumbach, K., Rosenthal, A., Stewart, A. L., Wang, F., et al. (2004). Physician language ability and cultural competence an exploratory study of communication with Spanish-speaking patients. *Journal of General Internal Medicine*, *19*(2), 167-174.

Flint, N. (2003). Medical history questionnaire. University of Florida Shands Executive Health Center. Retrieved May 23, 2008, from http://www.exechealth.ufl.edu/MHQ_questionnaire/EHC-MHQ.pdf.

Franco, H., Abrash, V., Precoda, K., Bratt, H., Ramana, R., Butzberger, J., et al. (2000). The SRI Eduspeak system: recognition and pronunciation scoring for language learning (pp. 123-128). Scotland. Retrieved from http://citeseer.ist.psu.edu/franco00sri.html .

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: Article.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking Communicative Language Teaching: A Focus on Access to Fluency. *The Canadian Modern Language Review, 61*(3), 325-353.

Guttman, M. S. (2004). Increasing the linguistic competence of the nurse with limited English proficiency. *J Contin Educ Nurs, 35*(6), 264-9.

Hadlow, J., & Pitts, M. (1991). The understanding of common health terms by doctors, nurses and patients. *Social Science & Medicine, 32*(2), 193-196.

Haidet, P., & Paterniti, D. A. (2003). "Building" a History Rather Than "Taking" One: A Perspective on Information Sharing During the Medical Interview. *Archives of Internal Medicine, 163*(10), 1134. doi: Article.

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*(3), 313-337.

Holland, V. M., Kaplan, J. D., & Sabol, A. M. A. (1999). Preliminary Tests of Language Learning in a Speech-Interactive Graphics Microworld. *CALICO Journal, 16*(3), 339-359.

Jang, Y., Kim, G., & Chiriboga, D. A. (2005). Health, healthcare utilization, and satisfaction with service: barriers and facilitators for older Korean Americans. *J Am Geriatr Soc, 53*(9), 1613-7.

Jones, C. M. (2001). Missing assessments: Lay and professional orientations in medical interviews. *Text, 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.*

Koekkoek, B. J. (1959). The Advent of the Language Laboratory. *The Modern Language Journal, 43*(1), 4-5.

LaRocca, S. A., Morgan, J. J., & Bellinger, S. M. (1999). On the Path to 2X Learning: Exploring the Possiblilities of Advanced Speech Recognition. *CALICO Journal, 16*(3).

Lear, D. W. (2005). Spanish for Working Medical Professionals: Linguistic Needs. *Foreign Language Annals, 38*(2), 223-35.

Léon, P. R., & Martin, P. (1972). Applied Linguistics and the Teaching of Intonation. *The Modern Language Journal, 56*(3), 139-144.

Lepetit, D., & Cichocki, W. (2002). Teaching Languages to Future Health Professionals: A Needs Assessment Study. *The Modern Language Journal, 86*(3), 384-396.

Little, D., & Davis, E. (1986). Interactive video for language learning: the Autotutor project. *System, 14*(1), 29-34.

Maguire, G. P., & Rutter, D. R. (1976). History-taking for medical students, I—deficiencies in performance. *The Lancet, 308*(7985), 556-558.

Manson , A. (1988). Language Concordance as a Determinant of Patient Compliance and Emergency Room Use in Patients with Asthma. *Medical Care, 26*(12), 1119-1128.

Morton, H., & Jack, M. A. (2005). Scenario-Based Spoken Interaction with Virtual Agents. *Computer Assisted Language Learning, 18*(3), 171-191.

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*(3), 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.

Otto, B. J. (1999). The interview. *Nursing, 29*(8), 77. doi: Article.

Paul, S., Dawson, K. P., Lanphear, J. H., & Cheema, M. Y. (1998). Video recording feedback: a feasible and effective approach to teaching history-taking and physical examination skills in undergraduate paediatric medicine. *Medical Education, 32*(3), 332-336.

Pfeiffer, C., Madray, H., Ardolino, A., & Willms, J. (1998). The rise and fall of students' skill in obtaining a medical history. *Medical Education, 32*(3), 283-288. doi: Article.

Price, B. (2004). Conducting sensitive patient interviews. *Nurs Stand, 18*(38), 45-52; quiz 54-5.

Rahman, A. (2000). Initial assessment of communication skills of intern doctors in history-taking. *Medical Teacher, 22*(2), 184-188. doi: Article.

Robinson, J. D., & Heritage, J. (2006). Physicians' opening questions and patients' satisfaction. *Patient Education and Counseling , 60*(3), 279-285.

Rypa, M. E., & Price, P. (1999). VILTS: A Tale of Two Technologies. *CALICO Journal, 16*(3), 385-404.

Schulz, R. (1988). The interactive videodisc and its potential for second language acquisition. *Die Unterrichtspraxis / Teaching German, 21*(1), 55-59.

Scripps Institution of Oceanography. (2004). Medical history questionnaire. *Medical History Questionnaire*. Retrieved May 23, 2008, from http://shipsked.ucsd.edu/general/forms/mhq.pdf.

Seijo, R., Girmez, H., & Freidenberg, J. (1991). Language as a communication barrier in medical care for Latino patients. *Hisp J Behav Sci, 13*(363).

Sherman, J. L., & Fields, S. K. (1988). *Guide to patient evaluation: history taking, physical examination, and the nursing process* (5th ed., p. 649). New York: Medical Examination Pub. Co.

Stivers, T., & Heritage, J. (2001). Breaking the sequential mold: Answering `more than the question' during comprehensive history taking*. *Text, 21*(1/2), 151.

The Canadian Press. (2007, December 11). Caregivers must be open to cultural differences, commission told. *CBC news*. Retrieved January 2, 2008, from http://www.cbc.ca/canada/montreal/story/2007/12/11/qc-bouchardtaylor.html.

Thomas, E. J., Hafler, J. P., & Woo, B. (1999). The patient's experience of being interviewed by first-year medical students. *Medical Teacher, 21*(3), 311-314. doi: Article.

Todd, K. H., Samaroo, N., & Hoffman, J. R. (1993). Ethnicity as a risk factor for inadequate emergency department analgesia. *JAMA, 269*(12), 1537-1539.

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*(3), 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 Gómez Chova, L., Martí Belenguer, D, & Candel Torres, I. (Eds.), *INTED 2008 Proceedings* (pp. 1-9). Valencia, Spain: International Association of Technology, Education and Development.

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*(1), 76-84.

Wong, W., & VanPatten, B. (2003). The evidence is in: drills are out. *Foreign Language Annals, 36*(3), 403-423.

Yuan, J., Liberman, M., & Cieri, C. (2006). Towards an integrated understanding of speaking rate in conversation. In *Interspeech 2006*. Pittsburgh.

Zambrana, R. E., Molnar, C., Munoz, H. B., & Lopez, D. S. (2004). Cultural competency as it intersects with racial/ethnic, linguistic, and class disparities in managed healthcare organizations. *Am J Manag Care, 10 Spec No,* SP37-44.

Zue, V., & Cole, R. (2007, February 9). Survey of the state of the art in human language technology. Retrieved from http://www.coli.uni-saarland.de/publikationen/softcopies/Cole:1997:SSA.pdf.

APPENDIX A: MEDICAL HISTORY QUESTIONNAIRE

H-CALM Virtual Dialogue Medical History Questionnaire

1

NAME_____ PHONE _____

ADDRESS_____

BIRTHDATE____/____/____ MARITAL STATUS _____

SEXUAL ORIENTATION ○ Heterosexual ○ Homosexual

ALLERGIES:_____

BLOOD TYPE: A B AB O RH FACTOR: POS NEG

FAMILY HISTORY: Has anyone in your family ever had (check box if yes):

○ Heart Disease	○ Mental Illness	○ High Blood Pressure
○ Epilepsy/seizures	○ Cancer	○ Diabetes

Check the box if you have had or received medical treatment for:

○ Heart disease	○ Arthritis	○ Sexually Transmitted Disease
○ Cancer	○ Kidney Trouble	
○ High Blood Pressure	○ Epilepsy/seizures	○ Drug Problems
○ Diabetes	○ Hearing Problems	○ Asthma
○ Vision Problems	○ Mental breakdown	○ Tuberculosis
○ Hernia	○ Frequent Headaches	○ Back Problems
○ Chronic Cough	○ Dizziness	○ Varicose Veins
○ Chest Pain		

FEMALES ONLY

○ Pregnancy	○ Menstrual Problems	○ Breast Lumps

MALES ONLY

○ Prostate Problems	○ Penile Discharge	○ Testicular Lumps

HABITS:

Do you smoke? _____ How long? _____ Cigarettes per day? ___

Do you drink alcohol? _____ How much/often? _____

Do you use or take any drugs? _____ What kinds? _____

List prescription and over the counter medications you take regularly: _____

_____

What is your usual occupation? _____ How many hours a week do you work? _____

APPENDIX B: VLP SCRIPT ITEMS

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 sugar?
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 penile discharge?
61. Have you noticed any testicular lumps?
62. Are you seeing a doctor for anything at the moment?
63. Do you smoke?
64. How many cigarettes do you smoke in a day?
65. When did you start?
66. Do you drink alcohol?
67. Really?
68. Do you take any drugs?
69. Are you taking any medication?

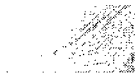
70. What is your job?

71. How many hours do you work in a week?

72. Thanks.

APPENDIX D: LANGUAGE EXPERIENCE QUESTIONNAIRE

Questionnaire d'experience en anglais



Toute l'information recueillie restera strictement confidentielle.

Si vous avez des questions concernant vos droits en tant que participants à l'étude, S.V.P. contactez Adela Reid, Agente d'éthique en recherche/conformité, Université Concordia, au 514-848-2424 poste 7481 ou par courriel au adela.reid@concordia.ca

First name: **Last name:**

1 Est tu un homme ou une femme?

- un homme
- une femme

2 Quelle est votre date de naissance?

3 Votre ouïe est-elle normale, au meilleur de votre connaissance?

- Oui
- Non

4 Quelle est votre langue maternelle ?

5 Avez-vous été exposé à cette langue depuis la naissance?

- Oui
- Non

6 Quelle est votre langue seconde?

7 Parlez-vous d'autres langues que l'anglais ou le français?

8 **À quel âge avez-vous commencé à apprendre votre langue seconde?**

9 **Quelle langue parlez-vous à la maison?**

10 **Dans quelle langue avez-vous étudié à l'école primaire? Choisissez la langue appropriée.**

- Français seulement
- Anglais seulement
- Autre

11 **Dans quelle langue avez-vous étudié à l'école secondaire? Choisissez la langue appropriée.**

- Français seulement
- Anglais seulement
- Autre

12 **Dans quelle langue avez-vous étudié au Cégep? Choisissez la langue appropriée.**

- Français seulement
- Anglais seulement
- Autre

13 **Décrivez votre apprentissage de l'anglais en termes de nombre d'années.**

14 **Décrivez votre apprentissage de l'anglais en termes de nombre d'heures par semaine.**

15 **Décrivez votre apprentissage de l'anglais en termes de environnement (école, maison, travail).**

16 **Parler en français:**
Veuillez évaluer votre habileté à parler en français. Utilisez les chiffres de 1 à 5 dans les cases ici-bas. Veuillez noter que 1 = très faible et 5 = excellente.

- 1 = très faible
- 2
- 3
- 4
- 5 = excellente

17 **Écouter en français:**
Veuillez évaluer votre habileté à écouter en français. Utilisez les chiffres de 1 à 5 dans les cases ici-bas. Veuillez noter que 1 = très faible et 5 = excellente.

- 1 = très faible
- 2
- 3
- 4
- 5 = excellente

18 **Lire en français:**
Veuillez évaluer votre habileté à lire en français. Utilisez les chiffres de 1 à 5 dans les cases ici-bas. Veuillez noter que 1 = très faible et 5 = excellente.

- 1 = très faible
- 2
- 3
- 4
- 5 = excellente

19 **Écrire en français:**
Veuillez évaluer votre habileté à écrire en français. Utilisez les chiffres de 1 à 5 dans les cases ici-bas. Veuillez noter que 1 = très faible et 9 = excellente.

- 1 = très faible
- 2
- 3
- 4
- 9 = excellente

20 **Parler en anglais:**
Veuillez évaluer votre habileté à parler en anglais. Utilisez les chiffres de 1 à 5 dans les cases ici-bas. Veuillez noter que 1 = très faible et 5 = excellente.

- 1 = très faible
- 2
- 3
- 4
- 5 = excellente

21 **Écouter en anglais:**
Veuillez évaluer votre habileté à écouter en anglais. Utilisez les chiffres de 1 à 5 dans les cases ici-bas. Veuillez noter que 1 = très faible et 5 = excellente.

- 1 = très faible
- 2
- 3
- 4
- 5 = excellente

22 **Lire en anglais:**
Veuillez évaluer votre habileté à lire en anglais. Utilisez les chiffres de 1 à 5 dans les cases ici-bas. Veuillez noter que 1 = très faible et 5 = excellente.

- 1 = très faible
- 2
- 3
- 4
- 5 = excellente

23 **Écrire en anglais:**
Veuillez évaluer votre habileté à écrire en anglais. Utilisez les chiffres de 1 à 5 dans les cases ici-bas. Veuillez noter que 1 = très faible et 9 = excellente.

- 1 = très faible
- 2
- 3
- 4
- 9 = excellente

24 **Avez-vous déjà visité/habité un milieu anglophone ? Combien de temps ? (Si non, tapez *non*.)**

25 **Parlez-vous anglais à la maison ? Avec des amis ?**

- Oui
- Non

26 **Avez-vous déjà participé à un programme de langue (échange étudiant)? Si oui, où et pendant combien de temps ? (Si non, tapez *non*.)**

27 **Avez-vous déjà travaillé dans un endroit où l'anglais était utilisé ? Si oui, où et pendant combien de temps ? (Si non, tapez *non*.)**

28 **Avez-vous déjà suivi un cours de prononciation en anglais ? Si oui, où et pendant combien de temps ? (Si non, tapez *non*.)**

29 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous parlez en français.**

30 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous écoutez en français.**

31 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous lisez en français.**

32 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous écrivez en français.**

33 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous où vous parlez en anglais.**

34 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous écoutez en anglais.**

35 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous lisez en anglais.**

36 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous écrivez en anglais.**

37 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous où vous parlez dans une autre langue.**

38 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous écoutez dans une autre langue.**

39 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous lisez dans une autre langue.**

40 **S.V.P. indiquez le pourcentage (%) de temps approximatif où vous écrivez dans une autre langue.**

Toute l'information recueillie restera strictement confidentielle.

Si vous avez des questions concernant vos droits en tant que participants à l'étude, S.V.P. contactez Adela Reid, Agente d'éthique en recherche/conformité, Université Concordia, au 514-848-2424 poste 7481 ou par courriel au adela.reid@concordia.ca

Submit Reset

APPENDIX E: CONSENT FORM

FORMULAIRE DE CONSENTEMENT

Par la présente, je consens à participer au programme de recherche mené par Monsieur Nicholas Walker (514-690-2953 ou walknick@hotmail.com), sous la direction du Professeur Pavel Trofimovich (EN-413, 1455, boul. de Maisonneuve Ouest, Montréal, H3G 1M8; téléphone : (514) 848-2424 poste 2448; courriel : pavel.trofimovich@concordia.ca) du Département des sciences de l'éducation/Centre de didactique de l'anglais langue seconde de l'Université Concordia, et sous la direction du Professeure Henrietta Jonas-Cedergren (DS-5529, 320 rue Sainte-Catherine Est, Montréal, H2X 1L7; (514) 987-3000 poste 2673; courriel : cedergren.henrietta@uqam.ca) Département de linguistique et de didactiques des langues/Institut des sciences cognitives de l'Université du Québec à Montréal.

A. BUT DE LA RECHERCHE

On m'a informé(e) du but de la recherche, qui est de vérifier la fonctionnalité d'un logiciel multimédia d'apprentissage d'anglais par une évaluation de ma prononciation grâce à un module de reconnaissance vocale, et de collecter des idées de comment l'améliorer.

B. PROCÉDURES

Le temps requis pour compléter cette expérience est d'une heure. L'expérience se déroulera individuellement à l'aide d'un ordinateur, un microphone, et un Caméoscope dans une salle tranquille en présence du chercher. Le premier volet de la procédure consistera à remplir un bref questionnaire. Ensuite, vous serez appelés de faire une entrevue avec un patient virtuel en utilisant un module de reconnaissance vocale sur un ordinateur. Après, il y aura un deuxième questionnaire. À la fin, vous pourrez jouer librement avec le logiciel durant les dernières 10 minutes de l'heure avant de partir.

C. CONDITIONS DE PARTICIPATION

- Je comprends que je peux refuser de participer à cette recherche sans que cela m'occasionne des préjudices.
- Je comprends que je suis libre de retirer mon consentement et de mettre fin à ma participation en tout temps sans que cela m'occasionne des préjudices.
- Je comprends que ma participation à cette étude demeurera confidentielle (C'est-à-dire, le chercheur connaîtra mais ne divulguera pas mon identité).
- Je comprends que les données recueillies dans cette étude peuvent être publiées ou présentées lors de congrès scientifiques; les données seront disséminées de façon à protéger l'identité de chaque participant à cette étude.
- Je comprends que je peux recevoir une copie du rapport de recherche si j'en fais la demande, ce que je peux faire auprès de Nicholas Walker ou du professeur Trofimovich ou Professeure Henrietta Jonas-Cedergren pendant cet entretien ou plus tard par écrit.
- Je comprends que ces activités ne sont, en rien, liées à mon cours d'anglais.

JE DÉCLARE AVOIR LU CETTE ENTENTE ET AVOIR COMPRIS LES INFORMATIONS QU'ELLE RENFERME. JE CONSENS À PARTICIPER DE PLEIN GRÉ À CETTE ÉTUDE.

NOM, PRÉNOM : _____
(lettres d'imprimerie S.V.P.)

SIGNATURE : _____

SIGNATURE DU CHERCHEUR : _____

DATE : _____ 2008.

Pour toute question relative à vos droits à titre de participant à ce projet de recherche, veuillez vous adresser à Adela Reid, Agente, Étique et normes de recherche, l'Université Concordia, au (514) 848-2424 poste 7481 ou par courriel à areid@alcor.concordia.ca.

APPENDIX F: VLP QUESTIONNAIRE 1

VLP Questionnaire 1

First name: **Last name:**

Email **Score:** 0

Date: 3/14/2009 3:22:12 PM **IP:** 127.0.0.1

1 Are you a student or teacher?

- ○ Student
- ○ Teacher

2 What is your program of study?

- ○ Soins Infirmiers (Nursing)
- ○ Techniques de diététique (Nutrition)
- ○ Techniques d'hygiène dentaire (Dental Hygiene)
- ○ autre programme (other program)
- ○ I am not a student

3 Is this your first, second, or third year of study in your program?

- ○ I am a first year student
- ○ I am a second year student
- ○ I am a third year student
- ○ I have not started the program yet.
- ○ I am not a student.

4 What is your level of English at the cegep?

- ○ 100 Anglais de Base
- ○ 101 Anglais et Communication
- ○ 102 Anglais et Culture
- ○ autre (other)
- ○ Je ne sais pas (I don't know)

5 Do you think English will be useful to you in your career as a nurse?

- Yes, it will be very useful.
- Yes, it will be somewhat useful.
- No, it don't think I will need much English.
- No, I won't need any English.
- I don't know.
- I will not work as a nurse.

6 As a nurse, which English skills will be more important for you to learn?

- English speaking and listening
- English reading and writing

7 As a nurse, which situation do you think is most important for you to develop your English speaking skills?

- Speaking English in face-to-face conversations
- Speaking English in phone conversations
- Giving oral presentations to a group in English
- other

8 As a nurse, which situation do you think is most important for you to develop your English listening skills?

- listening for face-to-face conversations
- listening for phone conversations
- listening to oral presentations
- listening to movies, TV, music, or radio

9 As a nurse, who do you think you will speak English to most frequently?

- doctors
- nurses
- administrators
- patients
- other

10 If necessary, could you interview a patient in English?

- Certainly yes.
- Probably yes.
- I don't know.
- Probably no.
- Certainly no.

11 Do you agree or disagree? Outside Quebec, Francophone patients should be able to receive health services in French.

- I strongly agree.
- I somewhat agree.
- I somewhat disagree.
- I strongly disagree.

12 Do you agree or disagree? Inside Quebec, Anglophone patients should be able receive health services in English.

- I strongly agree.
- I somewhat agree.
- I somewhat disagree.
- I strongly disagree.

13 Do you agree or disagree? Inside Quebec, Allophone patients (not Francophones or Anglophones) should be able to receive health services in English if they want.

- I strongly agree.
- I somewhat agree.

- I somewhat disagree.
- I strongly disagree.

14 If an Anglophone needed medical help, how confident would you be about communicating with him or her entirely in English?

- Very confident
- Somewhat confident
- Somewhat nervous
- Very nervous
- I would not be able to speak

15 Would you like to improve your ability to interview patients in English?

- Yes
- No
- I don't care

16 What question could you ask a patient to find out his or her name?

17 What question could you ask a patient to find out about any medication he or she is taking?

18 What question could you ask a patient to find if there is a history of heart disease in his or her family?

19 What question could you ask a patient to find if there is a history of cancer in his or her family?

20 If you had to interview an English speaking patient, how confident would you feel about your ability to use correct English grammar?

- Very confident
- Somewhat confident
- Somewhat nervous
- Very nervous

21 If you had to interview an English speaking patient, how confident would you feel about your ability to use correct English pronunciation?

- Very confident
- Somewhat confident
- Somewhat nervous
- Very nervous

Print

APPENDIX G: VLP QUESTIONNAIRE 2

Grammar IQ | VLP Questionnaire 2

First name: **Last name:**

Date: 5/31/2008 3:31:12 PM **IP:** 127.0.0.1

Score: 0

1 **If necessary, could you interview a patient in English?**

- Certainly yes.
- Probably yes.
- I don't know.
- Probably no.
- Certainly no.

2 **If an Anglophone needed medical help, how confident would you be about communicating with him or her entirely in English?**

- Very confident
- Somewhat confident
- Somewhat nervous
- Very nervous
- I would not be able to speak

3 **Would you like to improve your ability to interview patients in English?**

- Yes
- No
- I don't care

4 **What question could you ask a patient to find out his or her name?**

5 **What question could you ask a patient to find out about any medication he or she is taking?**

6 **What question could you ask a patient to find if there is a history of heart disease in his or her family?**

7 **What question could you ask a patient to find if there is a history of cancer in his or her family?**

8 **Grammar: If you had to interview an English speaking patient, how confident would you feel about your ability to use correct English grammar?**

- Very confident
- Somewhat confident
- Somewhat nervous
- Very nervous

9 **Pronunciation: If you had to interview an English speaking patient, how confident would you feel about your ability to use correct English pronunciation?**

- Very confident
- Somewhat confident
- Somewhat nervous
- Very nervous

10 **What did you think of this Virtual Patient system in general?**

(Multiple answer: choose one, more than one, or none from the list)

- ☐ I liked it.
- ☐ It is interesting.
- ☐ It is realistic.
- ☐ It is easy to use.
- ☐ It is useful.
- ☐ I would not use it again.

11 **What did you think of the patient?**

- ○ Excellent
- ○ Good
- ○ Bad
- ○ Terrible

12 **How easy is it to use this Virtual Patient software?**

- ○ Very easy to use
- ○ Easy to use.
- ○ Difficult to use.
- ○ Very difficult to use.

13 **How much time does it take to learn how to use this system?**

- ○ No time--I understood how to use it immediately.
- ○ Little time--I could understand how to use it quickly.
- ○ Some time--I could understand how to use it after some practice.
- ○ A lot of time--It took a lot of practice to learn how to use it.
- ○ Too much time---It takes too long to understand how to make it work.
- ○ Je ne comprends pas cette question.

14 **It was _____ when the system rejected my pronunciation.**

- ○ very frustrating
- ○ frustrating
- ○ a little frustrating
- ○ acceptable
- ○ good
- ○ great

15 **It was _____ when the system accepted bad pronunciation.**

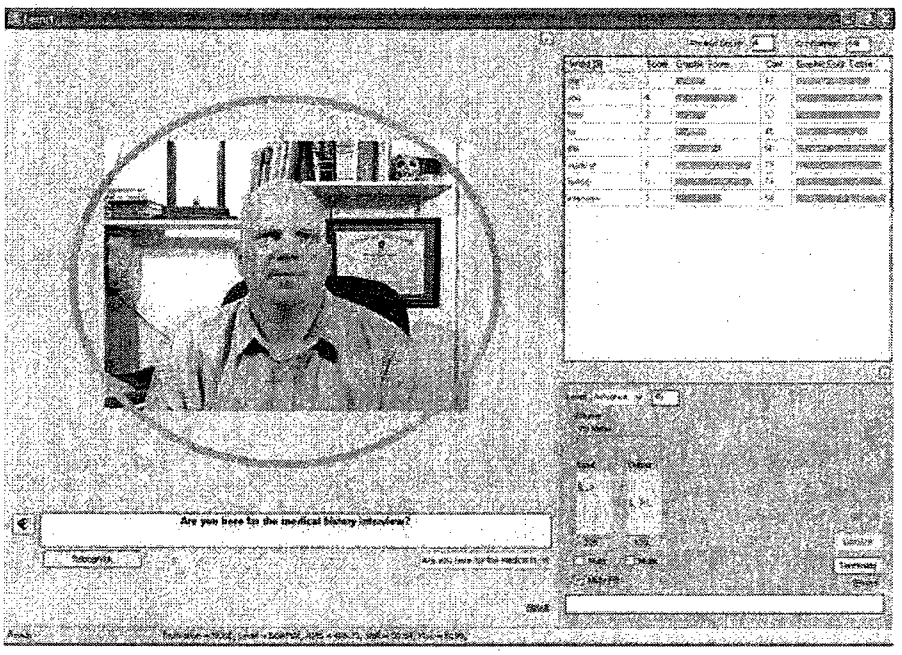
(Multiple answer: choose one, more than one, or none from the list)

- ☐ Irritating
- ☐ Funny
- ☐ Stupid
- ☐ Bad
- ☐ Acceptable
- ☐ Good

16 **I prefer the system when it is set to the _____ level.**

- ○ beginner
- ○ intermediate
- ○ advance

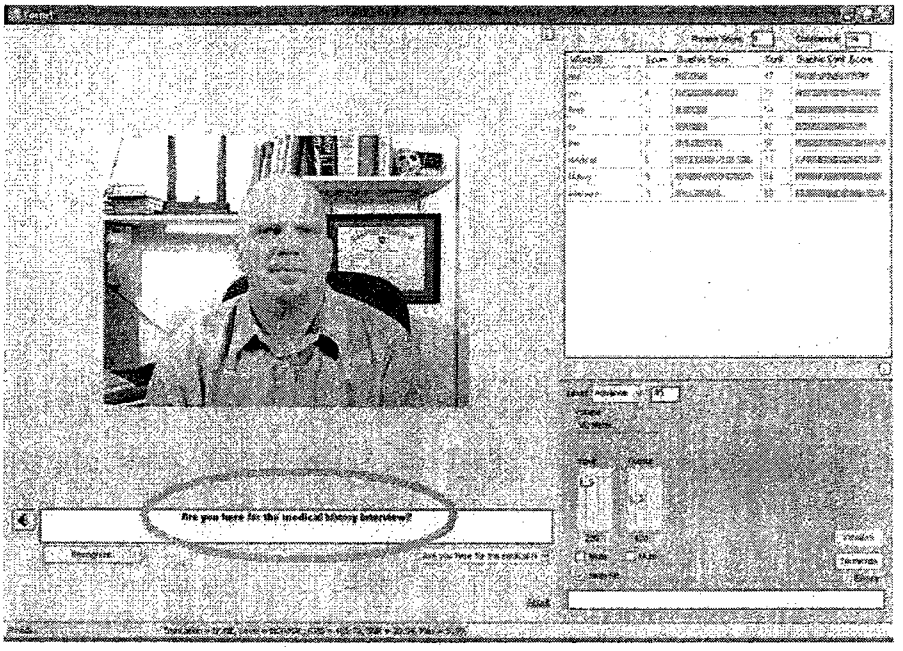
17 **Did anything bother you about the system?**

18 

What did you think of the video quality?

- Excellent
- Good
- Adequate
- Poor quality

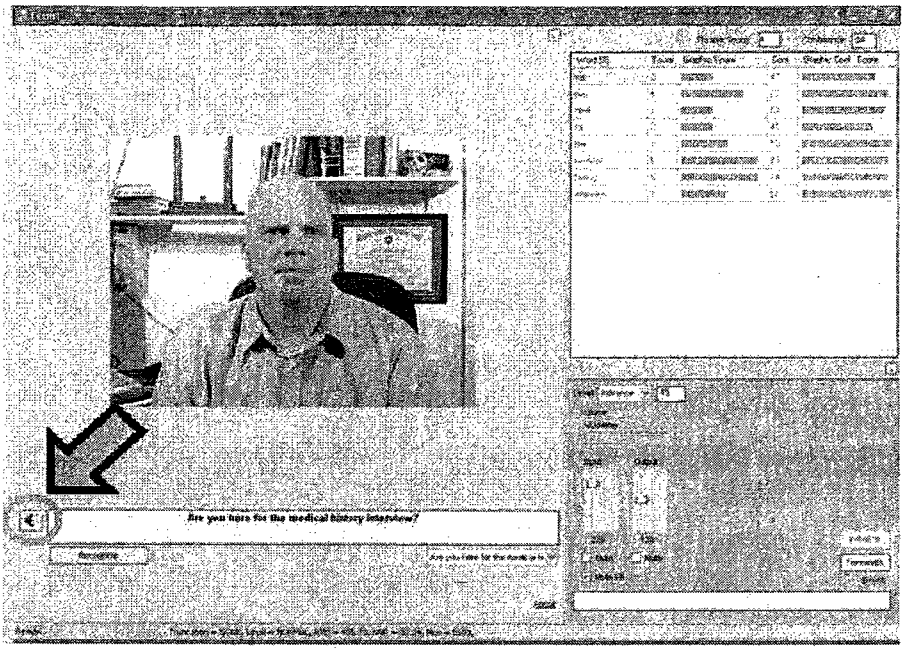
19



Was the size and position of the question easy for you to read?

- Very easy
- Easy
- Difficult
- Very difficult

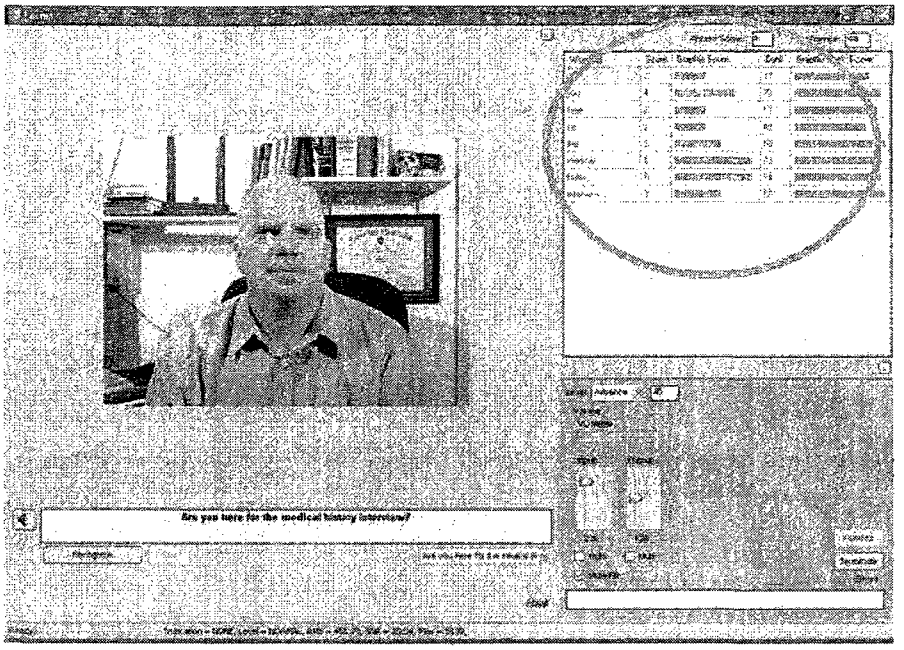
20



Was listening to the native speaker pronounce the question useful for you?

- Very useful
- Somewhat useful
- Not useful
- I did not use it

21



Was the graphic pronunciation feedback useful for you?

- Very useful
- Somewhat useful
- Not useful
- I did not use it

22 **How many times would you like to practice with this patient?**

- Once is enough.
- Twice is enough.
- Three times is enough.
- Four times is enough.
- Five times is enough.

23

Daniel Bonin is our first virtual patient. If we make more virtual patients, what should we do differently?

(Multiple answer: choose one or more from the list)

- ☐ One is enough. Don't make any more.
- ☐ Different medical problems (cancer, high blood pressure, asthma).
- ☐ Different English accents (US accent, British accent, Jamaican accent).
- ☐ Different foreign accents (Chinese accent, Russian accent).
- ☐ Different ages (children, old people)
- ☐ Include aggressive and uncooperative patients.

24 **How should we improve this system?**

Print

APPENDIX H: SEMI-STRUCTURED INTERVIEW QUESTIONS

1. Which difficulty setting do you prefer, advanced, intermediate, or beginner?
    a. Quel niveau préférez-vous, avancé, intermédiaire ou débutant?
2. Is the software still useful the second time using it?
    a. Es'ce que le logiciel est toujours utile la deuxième fois l'utiliser?
3. What do you think is the best and worst thing about this software?
    a. Selon vous, quelle est la meilleure et la pire des choses à propos de ce logiciel?
4. Can you suggest any changes we should make to the software?
    a. Pouvez-vous proposer des modifications, nous devrions apporter au logiciel?