Voice Recognition Softwares: Their implications to second language teaching, learning, and research

Chongwon Park*

ABSTRACT

Recently, Computer Assisted Language Learning (CALL) received widely held attention from diverse audiences. However, to the author's knowledge, relatively little attention was paid to the educational implications of voice recognition (VR) softwares in language teaching in general, and teaching and learning pronunciation in particular. This study explores, and extends the applicability of VR softwares toward second language research areas addressing how VR softwares might facilitate interview data entering processes. To aid the readers' understanding in this field, the background of classroom interaction research, and the rationale of why interview data, therefore the role of VR softwares, becomes critical in this realm of inquiry will be discussed. VR softwares' development and a brief report on the features of up-to-date VR softwares will be sketched. Finally, suggestions for future studies investigating the impact of VR softwares on second language learning, teaching, and research will be offered.

Keywords : voice recognition softwares, second language teaching, learning, and research, ethnography, qualitative methods

1. Classroom interaction studies and qualitative approach

In the field of Second Language Acquisition (SLA) research, input, interaction, and output were thought to be critical in one's language acquisition processes and product. These hypotheses have created a popular belief that the one who verbally participates in class is the one who acquires most and fastest. In reality, as a teacher, we seldom see such cases where students voluntarily participate in class without the teachers' request (Park, 1998). The results of studies conducted upon the relationship between students' active participation in the classroom and their overall proficiency are contradictory. Studies by Seliger (1977), Naiman et al. (1978), and Strong (1983;1984) report a positive relationship between various measures of learner participation and proficiency. However, Day (1984), in his replication of Seliger's study, and Ely (1986) found no such

* English Department, Chonan University.

relationships. Allwright (1980) found that the learner who participated the most in the lesson he analyzed was not among those who showed the greatest advances.

The limitations of early correlational studies which emphasize the role of active participation and language acquisition have been pointed out by some SLA researchers. The results of such correlational studies are not easy to interpret. There is no way to determine whether participation causes learning, or proficiency causes participation when a significant relationship is discovered (Chaudron,1993; Ellis, 1988, 1994; Allwright et al. 1991; Johnson 1992). Even if we could determine causality from these studies, they fail to provide explanations for why students behave as they do in the class.

These unsatisfactory findings from classroom interaction research have shifted researchers' methods of inquiry from numbers to learners' voices, perspectives, and attitudes in order to understand what is going on in a language classroom (e.g. Special Issue of Applied Linguistics (1992) and TESOL Quarterly (1995) on Qualitative Research Methods, and Bailey et. al (1996) for empirical evidence from diverse learners and settings).

Not merely due to the failure of choosing the right methods of inquiry, some SLA researchers provide the following reasons as to why these paradigm shifts are inevitable. The first reason for choosing qualitative methodology in general, and ethnography in particular is linked to the nature of classroom research in general. Van Lier (1988) claims that second language classroom studies lead to classification models for the quantitative analysis of variables, and serve as input for comparative or experimental studies. In most cases, he believes, a specific phenomenon, or a prespecified model has been selected beforehand for close analysis, and the tendency is for further work to lead to ever-increasing specification and narrowing down of the phenomena investigated. Both the 'holistic' and the 'emic' principles are therefore absent, and such studies lead away from, rather than towards, a description of classroom interaction processes.

The limitations of the previous studies have led many researchers, if not all, to agree to the fact that we know very little about 'what actually goes on in the classroom (Allwright & Bailey, 1991; Fishman, 1977; Stenhouse, 1975; Stubbs, 1976; Van Lier, 1988).

To understand better what actually happens in the classroom, Van Lier (1988) emphasized the need to look at classroom research as 'a contextually defined setting'. He argues that consideration of context and purposeful interaction are central. Van Lier states that: The classroom as context-based analysis can therefore not have as its primary aim the immediate generalizability of findings. Early generalizing can be detrimental to classroom research; it often prevents depth of analysis (p.2).

For classroom research to be possible, the classroom setting must be 'intact', and not set up for the purpose of research. Such research can not be judged by standards of generalizability and prediction, but rather by its power to create an understanding of how things happen the way they do, and how such research affects individuals making these

things happen. For that reason (Van Lier, 1988), there is no point in judging classroom research by the criteria of internal and external validity. He argues that the basic principles of classroom research which emphasize the importance of 'intersubjectivity' and 'context' share these principles with ethnography. Based upon his claim, Van Lier addresses the need to consider the unique rather than the common, the particular rather than the general. To him, greater understanding can be enhanced by looking at uniqueness through context, because that uniqueness is a part of generality.

The second reason for using qualitative approaches is linked to the their goal (Locke et al., 1993; Maxwell, 1996; Nunan, 1992; Seidman, 1991; Van Lier, 1988; LeCompte and Geetz, 1982; Yin, 1989). The goal of conducting qualitative study is making or constructing 'meaning making processes' from the realities which people have taken for granted (Spradley 1980). Selected Intensive English Program (IEP) Korean students from Park's study (1998) stated that they came to the U.S. to improve their oral proficiency. In a contradictory fashion, they usually socialized only with fellow Koreans; but at the same time, they feared they were losing their chances at practicing English. If they were aware of the disadvantage of this ethnocentric socialization, what made it difficult for them to break into mainstream American culture and society? If there was a reason or reasons for socializing with fellow Koreans, which is quite opposite to their goals of coming to the United States, adult IEP teachers as well as selected Korean students needed to know what were the causes of the conflicts between their goals and realities. To practice speaking English, it is sensible to practice speaking English with native speakers in social situations, and restrict the number of contacts with fellow Koreans. However, behind the question of their ethnocentric nature, there may be many reasons that might be critical hindrances to the selected Korean respondents of this study.

By explicitly describing the conflict, one might be able to find some solutions by which to resolve it. The purpose of Park's (1998) study by using an ethnographic approach is to assist language teachers and administrators in understanding the nature of Korean adult English as a Second Language (ESL) learners, who are implicitly governed by their cultural norms and values, and who therefore suffer from the conflict between their adjustment into American society and the day-to-day events in their lives.

In his Acculturation Model, Schumann (1986) observed that students' acculturation to the target-language group will control the degree to which they acquire the second language. What makes each student's acculturation level different is his or her social and psychological distance. Social distance involves the extent to which individual students become members of the target-language group, and as a result thereof have an opportunity to interact with that group. Although Schumman's model offers many observations which are relevant to Park's study in terms of language learning processes, we must be cautious in applying his insights to the situation because his model is

designed to account for SLA under conditions of immigration. The students in an IEP are sojourners, not immigrants and are consequently less subject to some of the factors Schumann discusses.

Mainstream research has focused on the factors that affect students within the classroom. Most of the research has adopted quantitative methodology. However, using naturalistic inquiry can reveal not only factors within the classroom, but also those outside the classroom which ultimately affect a student's overall proficiency (Loughrin-Sacco, 1990). It is only by examining the group dynamics among peers and with teachers, as well as the teaching methods employed, that an overall picture of the language learning process can be developed (Ellis 1994).

The third reason for employing an ethnographic approach is that its 'emic' and the 'holistic' principles have considerable applicability to the classroom interaction studies. The 'emic' principle assumes the importance of understanding participants' perspectives (Bodgan & Biklen, 1982; Lincoln & Guba, 1985; Magoon, 1977; Patton, 1990: Rist, 1977; Wilson, 1977: Jacob, 1988). The 'holistic' principle emphasizes the fact that an aspect of culture, or behavior must be described, and explained in relation to the whole system of which it is a part (Watson-Gegeo, 1988). An alternative way of expressing this principle is that an adequate ethnographic analysis must account for both the behavior and the context in which the behavior occurred. Context and behavior are seen as mutually causal. Here, "context includes not only the immediate circumstances in which an activity or interaction occurred (micro context), but also relevant socio-cultural relationships and institutions (macro context)." (Watson-Gegeo & Welchman-Gegeo, 1995).

1.1 Interviewing and transcribing as a main source of qualitative data input

According to Jacob (1988), there are six qualitative traditions from the disciplines of psychology, anthropology, and sociology. These traditions are human ethology, ecological psychology, holistic ethnography, cognitive anthropology, ethnography of communication, and symbolic interactionism.

Holistic ethnographers assume that certain aspects of human culture are central to understanding human life in all societies. Malinowski (1922/1961) identified three basic characteristics of holistic ethnography. First, holistic ethnographers directly gather empirical evidence themselves through fieldwork in the culture they are studying which usually involves participant observation and informal interviews. Second, holistic ethnographers endeavor to document the participants' points of view, preferably through verbatim statements. Third, holistic ethnographers collect a wide range of data using a wide range of methods.

It is important, however, for the reader to understand that holistic ethnographers do not focus on the goals and emotions of individuals, but instead attempt to understand

human society through the concept of culture, which is seen as involving subjective component-shared patterns for behavior. Many holistic ethnographers are interested in objectively documenting patterns of behavior in a culture. Ethnographers of communication, by means of their use of audio tape and videotaped data are also concerned with collecting objective, retrievable records of social interaction (Mehan, 1979). The holistic ethnographers are seen as data collection instruments themselves and their subjective" knowing" of the culture provides important information (Malinowski, 1922/1961).

1.2 The urgent need for using voice recognition software as a way of entering the interview data

Perhaps one of the most difficult and time consuming processes in conducting qualitative research is the transcription of large amounts of interview data from audiotape. No matter how much effort and time it takes, it is necessary for the researcher to transcribe all the interview data into written texts. Considering the nature of ethnographic study which emphasizes 'emic' perspectives, preselecting parts of the tapes to transcribe and omitting others would tend to lead to premature judgments about what is important and what is not (Seidman, 1991). Seidman warns: Once the decision is made not to transcribe a portion of the tape, that portion of the interview is usually lost to the researcher. So although labor is saved in preselecting the interview data, the cost may be high (p.88).

In Park's study, in order to transcribe all the interview texts, he utilized a computer based voice typed dictation software called IBM Voice Type 3.0 (1994), Simply Speaking Gold (1996), and Via Voice (1997). After each interview was conducted, he listened to the tapes and verbally translated what the tape said directly to the computer. When translating and transcribing the interview data without using voice type, it usually took him 12 hours to finish a 90 minute tape. With the voice typed system, it took about three hours to finish one tape, although there were individual differences from tape to tape and session to session.

2. Four leading Voice Recognition softwares

2.1 The background

The idea of simply speaking to the computer has been investigated by International Business Machines (IBM) since the 1950's. However, it is only in the last few decades that voice recognition softwares have developed to the point where it is more a reality than science fiction (Coniam, 1999). Right before the time when IBM's Simply Speaking Gold version came on to the street, purchasing and utilizing voice recognition softwares

was thought to be an expensive as well as a frustrating experience. In terms of frustration, one of IBM's Simply Speaking Gold (1996) users expressed her extreme anger toward this program and admonished not to use this program: Simply Speaking Gold might not recognize your speech, but it will occasionally transform it into interesting—even inspired—lines: "Up fruitful ground!" The green in screen called mainly on applying." And our favorite: "It's working vine." (Branson, 1997).

The author of this study started the VR program with great hope, believing that the program would eventually terminate his almost uncountable numbers of interview texts with less effort and time consumption. It turned out to be that there was as much, or even more frustration than Branson expressed. There seemed to be no option for the researcher except using the program, and he observed three possible reasons for the inaccurate production of voice recognition softwares. These are, 1) inappropriate microphone use·(noise reduction one is recommended), 2) inappropriate computer working environment (satisfying system requirements is a prerequisite before using a specific program), and 3) a vague expectation toward computer software programs. As far as the voice recognition program is concerned, there is no program whose recognition rate is 100 % accurate right after installation. In a sense, it requires consistent effort and patience. Think about the time when you first used your own word processor, whether it was Microsoft Word, Word Perfect, or whatever. To be a competent user of a specific program, definite time investment is highly advisable. As time passed by, as the vendor had insisted, the recognition rate increased up to the point of 80 percent which seemed still not ready for my purpose of utilizing this program.

2.2 The development

Behind the idea of simply talking to the computer, three conditions must be met, that is, 1) anyone can speak anytime 2) with natural speed to the computer, and the computer is expected to recognize the voice 3) accurately. These three conditions, however, were not fulfilled in the previous stages of voice recognition. Coniam (1999) argues; "…A distinction first needs to be drawn between voice recognition and speech recognition. Voice recognition requires machine training and is speaker-dependent; speech recognition, in contrast, is speaker-independent."[1] This concept is critical in understanding the development of voice recognition softwares, and will be discussed in turn.

Compared to the previous versions of VR softwares, three things are developed and these developments are deeply related with VR softwares' implications to second

1) Unfortunately, all of the currently existing voice recognition softwares, Dragon Naturally Speaking Preferred 4.0 (2,000), Philip's Free Speech 2000 (2,000), L & H Voice Xpress Professional 4 (2,000), and Via Voice Pro Millennium Edition (2,000) are voice recognition softwares.

language learning, teaching, and research. These developments are, 1) shortened enrollment time, 2) from discrete to continuous speech module, and 3) enhanced accuracy rate.

2.2.1 shortened enrollment time

As pointed out earlier, the limitation of current VR is its speaker-dependent function. Therefore, after the program installation, users are expected to train the computer with given speech samples. Figure 1 illustrates the first enrollment dialogue box of Via Voice (1997).

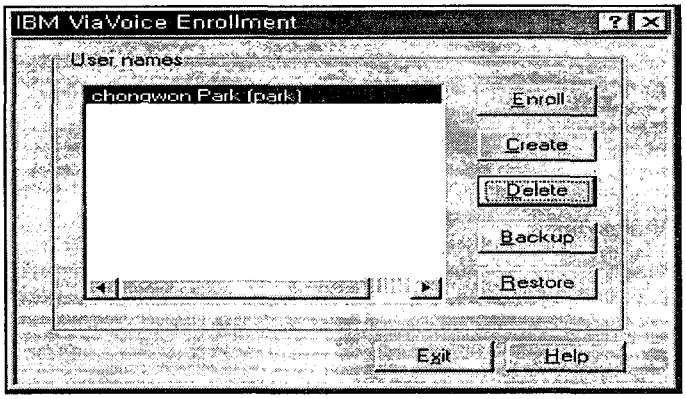

Figure 1. IBM Via Voice enrollment dialogue box

Not everyone needs to go through the enrollment process, because Via Voice interprets speech patterns accurately most of the time. If you don't have a foreign accent, you only need to read and train the first 50 sentences of the enrollment script. (Via Voice, 1996: help file) However, for ESL users, it is highly recommended to go through the whole enrollment process. Creating as many users as one wishes and deleting unwanted users anytime increases the possibility of applying this program to teaching, and researching a second language. The function of backing up and restoring your speech file to transfer speech information from one computer to another will make it possible for the teachers and researchers to have easy access and utilization whenever needed. However, in the case of the previous VR softwares (IBM's Simply Speaking Gold (1996), Via Voice (1997), and Dragon Naturally Speaking (1997)), 60 to 90 minutes are expected to finish the whole enrollment procedure. However, the innovative millenium versions of VR promise less than 10 to 15 minutes of reading enrollment text (Coniam (1999); Alwang (1999); Jecker (1999); Newman 1999). This relatively short and simple enrollment procedure offers even more opportunities of applying VR to the field of second language acquisition research, teaching, and learning. Coniam (1999) explores the possibility of developing an assessment tool, such as a reading aloud text via voice recognition

technology and determining a score through an analysis of the output. In the study, Coniam utilized Dragon system's 'Dragon Naturally Speaking' (DNS) since it is one of the first programs which claim to use continuous voice recognition technology. Ten non-native speaker participants read aloud an excerpt from Arthur C. Clark's novel '3001' (1997), consisting of a total of 3850 word tokens in 230 sentences. Participants' productions were analyzed in terms of T-units, clausal units, sub-clausal units, and single words. Coniam (1999) compared the results of the study with the previous one (1998) and concludes; "In terms of accuracy, the second language speakers' output on each category of analysis was significantly lower than that achieved by the native speakers."

The implication of Coniam's studies is that commercial VR softwares merely neither take what the non-native speaker said, nor guess and mechanically dictate. Rather, VR softwares have their own criteria, or norms of what native speakers' production should be. This characteristic of VR makes Coniam's proposition plausible. However, the problem still remaining was the enrollment time which took 60 to 90 minutes. As Coniam accepted, with 10 subjects, having 60 to 90 minutes for an enrollment session might not be a big issue, but, for a very large study where the number of subjects is more than 100, Coniam's propositions might not work.

Right after Coniam's (1999) article was published, as far as this enrollment procedure is concerned, more research and application friendly VRs were released to the street, tested, and verified by experts (Jecker, 1999). According to Jecker (1999), all four leading VR softwares mentioned above require 10 to 15 minutes of enrollment time. This recent development in enrollment time will surely enhance their applicability to a diverse area, especially in the realm of second language teaching, learning, and research. In addition to its wide applicability, the short and simple enrollment processes guarantee higher accuracy rate which will be the next topic of the discussion.

2.2.2 From discrete to continuous speech module

When it comes to the input methods of VR softwares, remarkable development has been made. Previous generations of VR products (for example, Kurzuweil's 'Voice' (1997) or IBM's 'Simply Speaking Gold' (1996)) have discrete speech as the form of input; "…As you know by now you must pause briefly between words in order for the dictation system to work. This is called speaking in isolated words. Computers can recognize speech better when people pause between words rather than speaking continuously. This makes it easier to tell the beginning and end of each word. It also decreases the variability in the way words sound on different occasions which makes recognition easier (Simply Speaking Gold, 1996: enrollment).

Example: b*egin dictation* **Speech** *pause* **recognition** *pause* **offers** *pause*
several *pause* **advantages** *pause* **over** *pause* **traditional** *pause*
dictation *pause* **and** *pause* **transcription** *pause* **period** *stop dictation*

Talking to the computer with a discrete manner might cause more problems than it solves. Imagine when you have to understand a person who speaks word-for-word. Contrary to the discrete manner, in continuous speech, sounds across words run together. As we listen to continuous speech, we understand the message from the prominences, which generally fall on the main lexical words (Coniam, 1999). The new generation of VR, which is now available (as of July of 2000) is that of continuous voice recognition where pauses between words are not necessary. This holds out more promise for man-machine interaction than does discrete-word speech (e.g. Kempaineu, 1997, cited in Coniam's study (1999)). Figure 2 represents the second dialogue box of Via Voice. showing input choices, either continuous or isolated.

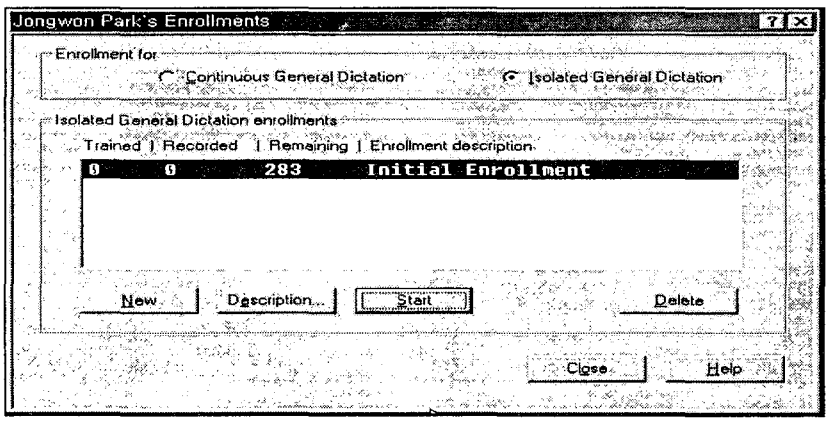

Figure 2. The second dialogue box of Via Voice

Compared to 'Simply Speaking Gold' (1996), command enrollment was incorporated into dictation enrollment. Via Voice (1997) users are required to read 283 sample sentences for the enrollment. Figure 3 illustrates the enrollment dialogue box introducing the first of 283 sentences. After enrollment, the accuracy rate is expected to increase the starting point to more than 90 percent. Although the Via Voice (1997) can be directly used in all areas of Windows programs, the highest accuracy can be achieved from Voice Pad which is the main work station of the program. Figure 4 depicts Voice Pad (for Figures 3 and 4, please refer to the appendix section of this study). As one can notice from Figure 4, VR programs have almost the same formation with other word processing programs except their dictation, reading aloud, and correcting errors. If a user is already familiar with Windows 95 or the above system, it will not take much time to understand

and run this program.

2.2.3 accuracy rate

Alwang (1999) conducted the benchmark test of four leading VR softwares.[2] After enrollment, two men and women dictated documents into the program to measure the initial accuracy. They corrected these documents to enable the program to build its vocabulary. Then, they dictated the same documents a second time to determine the final accuracy scores, which showed how much the program was able to learn from the process of dictation and correction (Alwang, 1999). Table 1 illustrates the results of the tests.

Table 1. The results of four leading VR softwares' benchmark test

Accuracy	Tester 1		Tester 2		Tester 3		Tester 4		Average	
Dragon Naturally Speaking	96	96	93	95	96	96	96	95	95	96
Free Speech 2000	91	95	97	92	92	94	95	92	91	93
L&H Voice Xpress Professional	91	93	93	93	92	96	95	95	93	94
Via Voice Pro Millenium Edition	94	98	96	98	93	99	96	97	95	**98**

note. The left column represents initial tests, and the right column represents final ones. adopted from http://www.zdnet.com/products/stories/reviews/0,4161,2388296,00.html

Alwang concludes that the accuracy rate is improving, with even less training time. In 1998's roundup, the tests achieved an average accuracy of 80 to 90 percent (in the case of Simply Speaking Gold(1996) or Via Voice (1997)). But for the 1999's story, initial accuracy ranged from 91 to 95 percent across the board. As table 1 indicates, in the case of Via Voice Millenium version, the accuracy rate went up to 98 percent as the frequency of the program use increased. For those who are interested in the features of the four major VR softwares, please refer to table 2 of this study which appears in the appendix section.

2) Participating VR softwares were Dragon Naturally Speaking Preferred 4.0 (2,000), Philip's Free Speech 2000 (2,000), L & H Voice Xpress Professional 4 (2,000), and Via Voice Pro Millennium Edition (2,000).

3. What VR means to the field of Second Language learning and teaching

Perhaps one of the most influential factors determining Korean college or university students' learning success would be class size. As the world is getting smaller and smaller, the drive for learning English has exploded. In the case of the university where the author teaches, numbers of student enrollment for Practical English courses tend to be more than 40 to 60. In the case of TOEIC (Test of English as International Communication), the numbers of students in one class used to be 60 to 100. In these situations, what can teachers do about students' speech production whenever correction or giving feedback is needed? If one would agree that the real purpose of teaching and learning English is to foster students' communicative abilities in and out of school environment, it is obvious that teaching pronunciation should be considered as a prerequisite. That prerequisite might be neglected because of the institutional constraints. How can a teacher provide appropriate feedback and evaluate learners' speech production development where learners' levels are so diversified?

Emphasizing computer programs' role as an aid, not a substitute for teachers, Warschauer and Healey (1998) pointed out seven benefits of adding a computer component to language instruction; 1) multimodal practice with feedback, 2) individualization in a large class, 3) pair and small group work on projects, either collaboratively or competitively, 4) the fun factor, 5) variety in the resources available and learning styles used, 6) exploratory learning with large amounts of language data, and 7) real-life skill-building in computer use. According to Warschauer and Healey (1998), utilizing educational softwares and making use of them seems to be one way of solving our educational dilemma.

Several programs for language teaching now incorporate speech recognition, not *voice recognition*, for example, The Learning Company's Learn to Speak Series; Triple Play Plus from Syracuse Language Systems; Course ware Publishing International's See It, Hear It, Say It, English Vocabulary; and Traci Talk; and Dynamic English from DynEd. One of the limitations of these programs compared to the VR softwares is that these programs can not guarantee free speech, but rather recognize a multiple choice answer (Warschauer and Healey, 1998). With the rapid development of VR softwares described so far, multimedia programs incorporating voice and speech recognition software can immerse students into rich environments for language practice. However, to obtain best accuracy rate, VR softwares require an isolated space with specified requirements.[3] The

3) A 300-MHZ Pentium II with 128 MB of RAM, running windows 98 and with a creative Lab sound Blaster Live! sound card and a 550-MHZ Pentium III with 128 MB of RAM and a creative Labs Sound Blaster Live! sound card running windows 98 were utilized for the benchmark test.

practical issue of how can we apply VR in an institutional level still needs to be solved. Providing a private room in a university computer lab and assigning each student according to the schedule might be one way of enhancing the applicability of VR in reality.

Not only for teachers and students, but for researchers in any field, especially ethnographers suffering from huge numbers of interview data, VR softwares can offer time-saving and cost effective solutions (Park, 1998). For example, Park (forth-coming) introduces three computer aided programs for data entering (IBM Via Voice), analyzing (QSR NUD.IST), and final writing up (QSR NUD.IST and Endnote Plus) and how these programs affected and facilitated his research processes.

4. Future research

So far, the researcher has introduced recent developments of VR softwares, and explored the possibility of their application to second language learning, teaching, and research. Paucity of related studies indicates that jumping to any conclusion is obviously premature at this time. More empirically based studies with rigid study design will validate whether VR softwares can be beneficial or not. As peoples' interests toward CALL has increased, numerous choices are available for language teachers. Most vendors of CALL software promote the effectiveness of their products and guarantee the ultimate achievement in language learning. However, what they lacked was the empirical evidence. If they argue that their products will offer fun for the learners, and eventually increase language learners' motivation, then, one could think of the reverse question. Is using computer software the only way to get students motivated? If not, why should the choice always be CALL? It is not a matter of superiority, that is, which software is better than the others. It should be a matter of context whether a specific software chosen would meet learners' current levels and needs, and will contribute to the second language learners' route and speed of acquisition. For future study, therefore, we should take into account a) the number of respondents, b) combining process and product (Spada, 1987; Allwright, 1991), and c) long term effects of learning through VR softwares. A study considering these factors will provide answers as to how, why, and for whom a specific CALL software is needed.

REFERENCES

[1] Allwright, D. and K. M. Bailey. 1991. *Focus on the language classroom: an introduction to classroom research for language teachers*. New York: Cambridge University Press.

[2] Allwright, R. 1980. Turns, topics and tasks: Patterns of participation in language

learning and teaching. D. Larsen-Freeman. MA: Newbury House.
[3] Alwang, G. 1999. Speech Recognition. http://www.zdnet.com/pcmag/stories/reviews/0,6755,2385295,00.html.
[4] Bailey, K. M. and D. Nunan. 1996. *Voices from the language classroom: qualitative research in second language education.* New York : Cambridge University Press.
[5] Bodgan, R. and S. Biklen. 1982. "Foundations of Qualitative research in Education: An Introduction." *Qualitative Research for Education.* Boston: Allym & Bacon, 1-53.
[6] Branson, J. 1997. VoiceType Simply Speaking Gold: don't speak! this voice recognition software isn'treadyyet. http://coverage.cnet.com/content/Reviews/Justin/items/1,118,199.00.html.
[7] Chaudron, C. 1993. *Second Language Classroom.* Cambridge: Cambridge university press.
[8] Coniam, D. 1999. "Voice recognition software accuracy with second language speakers of English." *System,* 27, 49-64.
[9] Day, R. 1984. "Student participation in the ESL classroom, or some imperfections of practice." *Language Learning,* 34, 69-102.
[10] Ellis, R. 1988. "The role of practice in classroom language learning." *AILA Review,* 5, 20-39.
[11] Ellis, R. 1994. *The Study of Second Language Acquisition.* Oxford: Oxford University Press.
[12] Ely, C. 1986. "An analysis of discomfort, risktaking, sociability, and motivation in the L2 classroom." *Language Learning,* 36, 1-25.
[13] Fishman, J. 1977. "The social science perspective." *Bilingual education: current perspectives 1(Social science),* 1-49.
[14] IBM 1993. IBM Voice Type Dictation. Thornwood: NY, IBM.
[15] IBM 1997. Via Voice. http: //www.software.ibm.com/is/voicetype
[16] IBM 1996. Simply Speaking Gold. Thornwood: NY, IBM.
[17] Jecker, D. 1999. Benchmark Tests: Speech Recognition. http://www.zdnet.com/products/stories/reviews/0,4161,2385302,00.html.
[18] Johnson, D. M. 1991. *Approaches to research in second language learning.* New York: Longman.
[19] Kempainen, S. 1997. "Automatic voice recognition lets machines listen and comprehend." *EDN,* 42, 5, 73-80.
[20] LeCompte, M. D. and J. Geetz. 1982. "Problems of reliability and validity in ethnographic research." *Review of Educational Research,* 52, 1.
[21] Lincoln, Y. S. and E. G. Guba. 1985. *Naturalistic inquiry.* California: Sage Publications.
[22] Locke, L. F., W. W. Spirduso, et al. 1993. *Proposals that work : a guide for planning dissertations and grant proposals.* Newbury Park: Sage Publications.
[23] Loughrin-Sacco, S. J. 1990. "Inside the 'black box' revisited: Toward the intergration of naturalistic inquiry in classroom resaerch on foreign language learning." *Polilingua: a college journal of foreign language 1,* 1, 22-26.
[24] Magoon, A. J. 1977. "Constructivist approaches in educational research." *Review of educational research,* 47, 651-693.
[25] Malinowski, B. 1961. *Argonauts of the western pacific.* New York: E.P.Dutton (original work published in 1922).

[26] Maxwell, J. A. (1996). *Qualitative research design*. California: Sage.
[27] Mehan, H. (1979). *Learning lessons: Social organization in the classroom*. MA: Harvard university press.
[28] Naiman, N., M. Frohlich, et al. 1978. *The Good language learner*. Toronto: The Ontario Institute for Studies in education.
[29] Newman, D. 1999. Via Voice Pro Millennium Edition. http://www.zdnet.com/filters/printerfriendly/0,6061,2385299-3,00.html.
[30] Nunan, D. 1992. *Research methods in language learning*. NY: Cambridge University Press.
[31] Park, C. 1998. *Why not speak English?: A study of language use among Korean students in an intensive English program in the United States*. Learning and Instruction. Buffalo: State Universityof New York at Buffalo.
[32] Patton, M. Q. 1990. *Qualitative evaluation and research methods*. Newbury Park:Sage.
[33] Rist, R. 1977. "On the relationships among educational research paradigms: From disdain to detente." *Anthropology and Education Quarterly*,8, 42-49.
[34] Schumann, J. H. 1986. "Research on the acculturation model for second language acquisition." *Journal of multilingual and multicultural development*, 7, 5, 379-392.
[35] Seidman, I. 1991. *Interviewing as qualitative research : a guide for researchers in education and the social sciences*. New York: Teachers College Press.
[36] Seliger, H. 1977. "Does practice make perfect? A study of the interaction patterns and L2 competence." *Language Learning*, 27, 263-78.
[37] Spada, N. 1987. "Relationships between instructional differences and learning outcomes: a process-product study of communicative language teaching." *Applied Linguistics*, 8, 137-61.
[38] Spradley, J. P. 1980. *Participant observation*. New York: Rinehart, & Winston.
[39] Stenhouse, L. 1975. *An introduction to curriculum research and development*. London: Heinemann.
[40] Strong, M. 1983. "Social styles and second language acquisition of Spanish -speaking kindergartners." *TESOL Quarterly*, 17, 241-58.
[41] Strong, M. 1984. "Integrative motivation: cause or result of successful second language acquisition?" *Language Learning*, 34, 1-14.
[42] Stubbs, M. 1976. *Language, schools and classrooms*. London: Methuen.
[43] Van Lier, L. 1988. *The classroom and the language learner*. New York: Longman.
[44] Warschauer, M. and D. Healey. 1998. "Computers and language learning: An overview." *Language Teaching*, 31, 53-71.
[45] Watson-Gegeo, K. A. 1988. "Ethnography in ESL: Defining the essentials." *TESOL Quarterly*, 22, 575-592.
[46] Watson-Gegeo, K. A. and Welchman-Gegeo. 1995. "Understanding language and power in the Solomon Islands: Methodological lessons for educational intervention." *Power and Inequality in language education*. M. H. L. a. J. C. Richards. Cambridge: Cambridge university press, 59-72.
[47] Wilson, S. 1977. "The use of ethnographic techniques in educational research." *Review of Educational Research*, 47, 245-265.
[48] Yin, R. K. 1989. *Case study research : design and methods*. California: Sage Publications.

APPENDICES

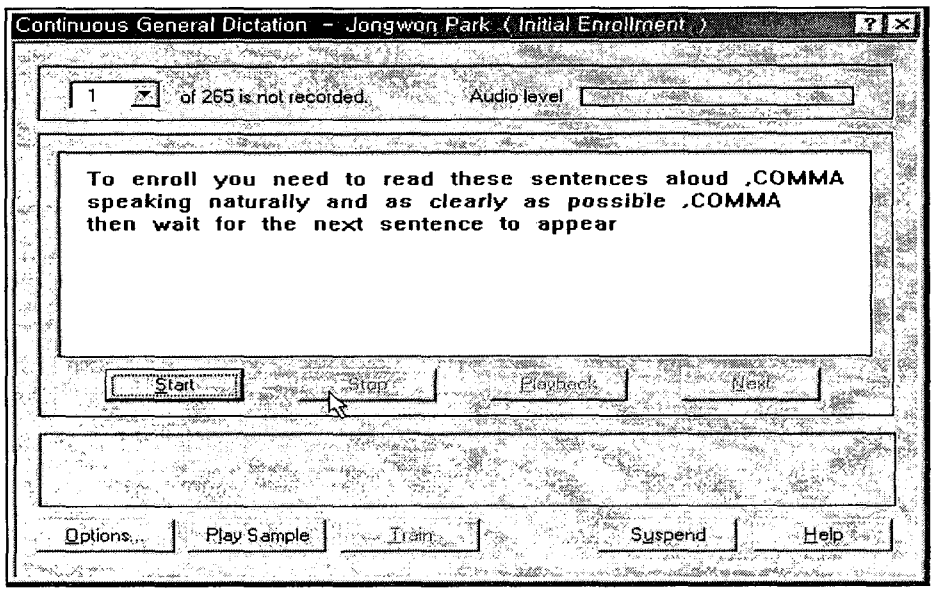

Figure 3. Continuous general dictation dialogue box

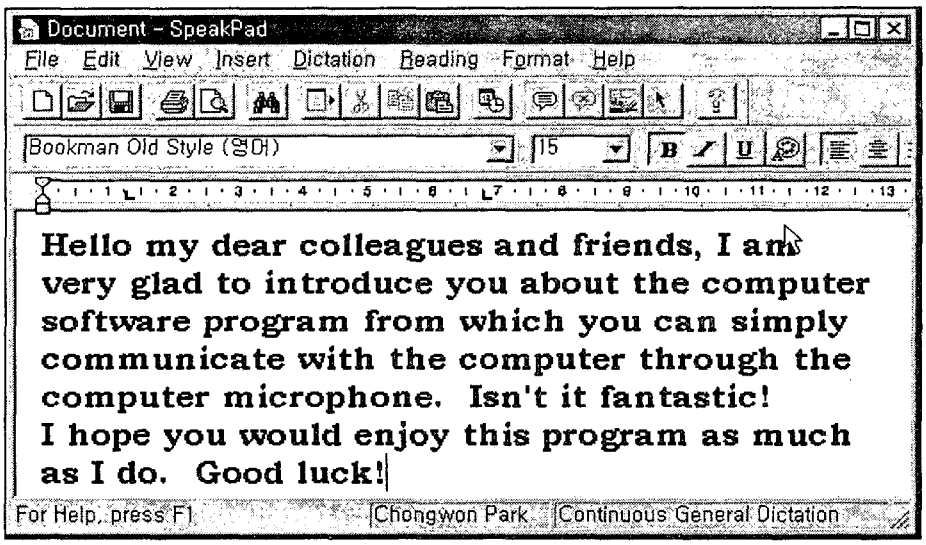

Figure 4. A sample of voice typed text from Speak Pad

Table 2. Summary of features of voice recognition software

Products	Dragon Naturally Speaking Preferred 4.0	Philips Free Speech 2000	L&H Voice Xpress Professional 4	Via Voice Pro Millennium Edition
Included Microphone Headset	VXI Parrot Translator	Plantronics SR 1	TelexUSBHeadwor n Microphone	Andrea NC-61
Base active vocabulary / Maximum active vocabulary	160,000 / 250,000	60,000 / 670,000	34,000 / 64,000	64,000 / 2,000,000
Supports multiple users / dictionaries	No / Yes	No / Yes	Yes / Yes	Yes / Yes
Modeless operation	Yes	No	Yes	Yes
Can play back dictation for proofreading	Yes	No	No	No
Can convert text to speech?	No	No	Yes	No
Support USB microphones	Yes	Yes	Yes	Yes
Transcribes from hand-held	No	No	No	No
Users can dictate into windows application	Yes	No	Yes	Yes
Command and control / natural-language command of				
Windows Desktop	Yes / No	Yes / Yes	Yes / Yes	Yes / Yes
Corel WordPerfect	Yes / Yes	Yes / Yes	Yes/ No	Yes / No
Microsoft Excel 97 and 2000	No / No	Yes / No	Yes / Yes	Yes / Yes
Lotus 1-2-3	No / No	No / No	Yes / No	Yes / No
Microsoft Outlook	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes
Microsoft IE 4.0 and 5.0	Yes / Yes	Yes / Yes	Yes / Yes	Yes / Yes
Netscape 4.0	Yes / No	No / No	Yes / No	Yes (with free down load) / No
AOL	Yes / No	No / No	Yes / No	Yes / Yes
Microsoft exchange	No / No	Yes / Yes	Yes / No	Yes / Yes
Lotus Notes	Yes / No	No / No	Yes / No	Yes / No
Microsoft PowerPoint	Yes / No	Yes / Yes	Yes / Yes	Yes / No

Products	Dragon Naturally Speaking Preferred 4.0	Philips Free Speech 2000	L&H Voice Xpress Professional 4	Via Voice Pro Millennium Edition
Microsoft Project	Yes / No	Yes / Yes	Yes / No	Yes / No
Microsoft Money	No / No	Yes / No	Yes / No	Yes / No
Quicken	No / No	Yes / Yes	Yes / No	Yes / No
Supports command macros	Yes	Yes	Yes	Yes
Specialized dictionaries included	None	French, Italian	Business/Finance, in the News, Leisure, Technology	Business/Finance, Chatter's Jargon, Computer, Cuisine
Dictionaries supported within other versions	Medical, Legal	Radiology, Auto Claims Insurance, Property Claims Insurance	Medical, legal, police	medical, legal
Can import user speech files?	Yes	Yes	Yes	Yes
Can export user speech files?	Yes	Yes	Yes	Yes

Received : August 5, 2000.
Accepted : September 1, 2000.

▲ Chongwon Park
English Department, Chonan University
115 Anseo Dong Chonan, Chungnam, 330-180, Korea
Tel: +82-41-550-9183, FAX : +82-41-575-1826
H/P: 016-220-0769
e-mail: chongwon@mail.chonan.ac.kr