K-buddy:
A Near-spontaneous and an Interactive Program for Learners of Korean as a Foreign Language
(자연대화방식을 이용한 음성인식기반 한국어 학습 프로그램)

Seok-Hoon You*

Abstract

This paper aims to introduce *K*-buddy, a world-first CALL (Computer Aided Language Learning) program with voice recognition and speech synthesis technology implemented. This program has excellent features such as sound user interface, high interactivity,[1] spontaneity of conversation, learner-centeredness, user-friendliness, and user customizability that would help the learners to have his or her own practice sessions for the four major skills of language learning, i.e. reading, writing, speaking, and listening, in comfortable and friendly environment.

Interactivity and spontaneity are two major key concepts of the program, that are borrowed from the field of language pedagogy. In order to guarantee an enhanced level of communicability competence of the language learners, various proposals have been made in the field so far. With the advent of new information and computer technologies in the

* Korea University Department of Linguistics

1) Interactivity is defined as "a dynamic exchange of opinions and ideas for a specific goal or purpose in mutually understood manner and convention". According to You (2000a), there are three types of interactivity relevant to language education: intra-dialogue, inter-dialogue, and open interactivity.

internet era, numerous application programs have been developed so far to facilitate the learning of foreign languages. But none of them were quite successful in its level of interactivity and spontaneity not to mention communication competence. *K*-buddy adopts these concepts as foundations of design to realize more improved and interactive cyber-environment for learners of Korean as a foreign language.

1. Introduction

More and more attentions have been paid to the issue of enhancing communicative competence of the students in language education recently. (Canale and Swain 1980, Canale 1983, Bachman 1990) The trend has been culminated in Brown (2000) in which is proposed that an integration of comprehension and production mechanisms in language learning is required to improve communicative competence of the learners.

Human communication is not complete without active interaction. Active interaction guarantees full comprehension of the content of dialogue between interlocutors. This trend has pushed researchers in the area of second or foreign language education to pursue the goal of developing pedagogical methodologies that would help the learners of a second or foreign language reach the maximum level of communicative competence by guaranteeing full interaction between instructors and learners. Similar approaches have

been made in the field of educational technology. Through joint efforts between language education specialists and the computer software engineers, many computer applications have been developed in various formats to facilitate the language learning process of the foreign learners of a language. Similar effort has been made in the Korean language since mid 1990s. (Kang 1996; Cho, 2000; Cho and Lee, 2000; Choe 1999, 2000, Ji and Lee 1999) However, it has been noted that most applications developed so far are based on a conventional interface between human controller and the learners through mediation of computer and the internet. In other words, human controller is indispensable in the learning process. There are other types of applications in CD-formats to provide more direct interface between human learners and the computer (and possibly Web or the internet). But none of them were successful in providing a spontaneous and learner-centered interface.

K-buddy is a program compatible with Windows 98 and 2000 for Korean, and is designed to facilitate an acquisition process of Korean as a foreign language or as a second language (hereafter KFL or KSL, respectively). This program is developed with a goal to provide a near-spontaneous and an interactive environment for the students of KFL or KSL practicing Korean conversation including speaking and listening on one hand, and reading and writing on the other hand. Furthermore, this program can provide an economic solution both to the learners and the providers of education

by reducing the need of the costly environment of the language education by delivering on-line accessibility through intranet (not internet) in rather economic mode.[2)]

To achieve goal of delivering a near-spontaneous and an interactive conversation practice session to the learners, voice recognition and speech synthesis, top-notch modern technologies, are adopted in *K*-buddy. In order to best utilize the high-tech engineering skills and to expedite the acquisition process, most major interfaces of the program are optimized for the sound-user interface (hereafter SUI). In other words, most graphical buttons and mouse operations that are standard in the current GUI environments such as Microsoft Windows interface are replaced with the SUI-based V-COMS (i.e., voice commands).[3)]

There are two engines incorporated to meet the needs and the requirements during interaction: one for the recognition of human voice and another for the recognition of cognitive-functional situations of the discourse. Use of these engines is to guarantee the maximum level of interaction between computer and learners by providing not only a better interactivity via voice but also more accurate and speedy

2) Due to massive size of database, the current internet facility is not able to provide full functionality in the accuracy and the speed of voice recognition and speech synthesis.

3) The GUI interface is improved to meet the need of new environment. Icons and graphics take major role in GUI environment. Mouse is an indispensible device in this already well-attested interaction environment. Please refer to 4.1 for improved GUI interface of the program.

exchange between the two based on the cognitive and discourse knowledge-base covering both sociolinguistic and cultural aspects of whole conversation.

2. Basic Design Concept of *K*-buddy

The basic design concept of *K*-buddy is to deliver quite a natural and a near-spontaneous interactivity in the learning process of Korean language as a foreign language. To guarantee this, four major factors are incorporated in the design: sound user interface (hereafter, SUI), studentcenteredness, user-customizability, and interactivity. Following is a brief description of each of them.

2.1. Sound User Interface (SUI)

K-buddy is based on so-called Sound User Interface, which is expected to be incorporated in the next versions of major operating systems in the computer industry.[4] To comprise the environment, the program is equipped with a top-notch voice recognition engine and cognitive functional grammar engine that is developed by Lernout & Hauspie Korea and released in June 2000 and the Team Lingua in July 2000, respectively. The voice recognition engine

4) In some operating systems such as Macintosh OS, voice natigation has been provided as an option.

provides TTS (Text-To-Speech) and STT (Speech-To-Text) features including voice recognition and speech synthesis while the latter provides more efficient and correct speech recognition and production of the program.

These two are kernel of the program through which a near-spontaneous conversation is realized. Once the program is executed, all sorts of commands including standard hierarchical menu system, icons, and dialogue boxes that comprise current standard user interface become invisible. Instead, V-COMS take over the role.

2.2. Student-centeredness

The fundamental design concept of the program is to provide KFL students a learning environment in which they can stand in the center of learning activities. In this environment, students take full control of the learning interaction provided that they are equipped with decent intranet facilities and the computers with matching specification. Since the program is developed targeting the intranet environment, it is guaranteed that students have free access to computers connected to each other via intranet. Also, it is required that the computers are appropriately equipped with the following minimal specification:

Parts	Minimum specification	Etc.
Pentium Processor	400MHz	
RAM	64Mb	
Video RAM	4Mb	
Storage	6Gb	
Sound	Sound Blaster or compatible	
Microphone	HiFi	Direction Detection
Speaker	Standard	

Table 1. minimum hardware specification

With the given environment, students are allowed to interact with the computer via voice on the lessons and the topics of their interest. Again, students have full control with regard to the selection of target lesson, length of interaction, choice of conversation partner, speech level, leadership of conversation, and so on. Most of all, students are free of all kinds of personal barrier or interferences such as losing face, embarrassment, time limitation, peer competition, unwanted place of instruction, etc.

2.3. User-customizability

One excellent feature among others is that users, i.e., students, can choose and even change the environment of practice session by giving his or her own preference list of options for conversation partner within the program.[5] Fur-

thermore, the program, not students, automatically adjusts to levels and expectations of the students to keep the conversation practice session more interactive and adaptable.

2.4. Intra-dialogue Interactivity[6]

This is a prime feature of the program. *K*-buddy is developed to emulate natural conversation situation. Thus, it allows near-spontaneous interaction between computer and learners. Other programs developed so far were unidirectional in the sense that interactivity was extremely limited. To be more precise, once the programs started, they are to wait the input or any kind of response of the students. Otherwise, they are not taking any active action. However, *K*-buddy does not wait too long until the students respond in pre-determined manner. Rather, it takes over the leadership of the conversation and starts or resumes the interaction. Furthermore, once students want to claim back the leadership, they can take it back by interrupting the dialogue uttered by *K*-buddy. It even allows skipping irrelevant turns within the dialogue as long as the action abiding by the rules of discourse (Grice, 1975). Consider the following.[7] Circled numbers are added to indicate the turns:

---

5) This feature is to be realized and included in the full version in the later stage. To construct this environment, various related technologies such as 3D-rendering, morphing, and lip synchronization are required.

6) Refer to You (2000) for further discussion on the interactivity.

7) We appreciate KLEAR for allowing us to use the text.

Dialogue 1 (Lesson 4, Integrated Korean Vol. 2)

(Steve runs into Jenny on campus on his way to Korean class.)

① 제니 스티브 씨, 어디 가세요?

Jenny Hi, Steve. Where are you going?

② 스티브 한국어 수업에 가요.

Steve I am going to Korean language class.

③ 제니 한국어 수업은 어디서 해요?

Jenny Where do you have the class?

④ 스티브 이스트 홀에서 해요.

Steve At East Hall.

⑤ 제니 교실이 몇 층이에요?

Jenny (On) what floor

⑥ 스티브 2층이에요.

Steve Second.

⑦ 제니 학생이 많아요?

Jenny Are there many students in the class?

⑧ 스티브 좀 많아요. 스무 명이에요.

Steve Quite a lot. Twenty.

⑨ 스티브 *참, 지금 몇 시예요?*

Steve Oh, my god, what time is it now?

⑩ 제니 열두 시 사십오 분이에요.

Jenny It s 12:45.

⑪ 스티브 *아이구, 오늘 한 시에 한국어 시험이 있어요.*

Steve My goodness, I have a Korean exam at 1 p.m.

⑫ 제니 그럼, 빨리 가세요. (Steve leaves.)

Jenny You'd better hurry.

⑬ 제니 안녕히 가세요.

Jenny Good bye

⑭ 스티브 네, 안녕히 가세요.

Steve Good bye.

In Dialogue 1, expressions in italics (⑨ and ⑪), especially exclamation markers such as *my god! or oh my goodness!*, are so-called communication breakers that function as breakers in natural conversation. The speaker can utter the expressions in urgent or emergency situations as can be seen in the dialogue. Once the expressions are uttered, a spontaneous flow of conversation is interrupted or broken, and new topic or closure of the conversation is expected to follow in natural conversation. *K*-buddy can detect this kind of breakers and respond in appropriate manner to make the conversation more spontaneous and interactive.

3. Components of *K*-buddy

K-buddy is composed of following five components: convertsation, dictation, read aloud, listening comprehension, and review practices. Following are description of each component:

3.1. Conversation

In this component, a near-spontaneous and interactive conversation practice is possible. Students are expected to say "대화(conversation)" to launch the function. In most CD-ROM titles developed so far, the sequence of conversation is limited to the order of text predefined in the scenario.

But, in *K*-buddy, learners are allowed to control the flow of text in a dialogue. In other words, they can start the dialogue even from the turn that they want to practice and the computer will response in appropriate turn as a reply to keep the conversation meaningful.

3.2. Dictation

In this component, a dictation drill is possible. *K*-buddy reads text and learners are expected to dictate what they hear using keyboard and dialogue box displayed on the screen and *K*-buddy check whether given answer is correct or not. If the input provided by students is not correct, *K*-buddy notifies students that the given answer is wrong and prompts to retry to type correct answer, or to proceed to next practice or repeat same one for reinforcement purpose.

3.3. Pronunciation/Read-aloud

In this component, *K*-buddy presents text in a small dialogue box so that students can read it aloud. *K*-buddy detects the reading and converts voice input into text to check whether the input is correct in its content and pronunciation. If the input pronunciation provided by students is not correct, *K*-buddy prompts students to try to read it aloud again. If students choose not to try the sentence again

for some reason and continue on with next sentences, *K*-buddy allows it to proceed to next sentences.

3.4. Listening Comprehension

In this component, *K*-buddy reads certain amount of text for students. Students listen to the reading and take a brief quiz or test to check whether they understand the text or not. *K*-buddy tracks the interactions between computer and students, provides helps when needs arise, and even corrects wrong answers. It also provides students chances to review the text as a reinforcement.

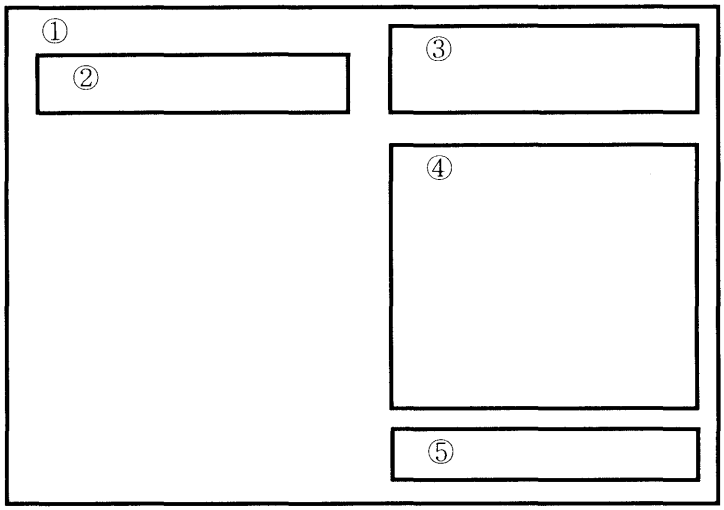
3.5. Review Practices

Review practices on vocabulary, phrases, grammatical tips, idiomatic expressions, and cultural facts are available upon request of students.

4. *K*-buddy: Opening Screen

The opening screen of *K*-buddy is as follows:

The program starts with a brief (about 5 seconds long) opening animation and a voice introduction to the program. Also, a short introductory text gives a brief introduction on the use of the program. The base screen is cluttered with



Picture 1. Opening screen of *K*-buddy

four smaller panes to display different types of menus and information. Each pane functions as follows:

4.1. Video Panel

In this pane, a selected conversation partner appears according to the preferences given by learners. 3-D(imen-sional) rendering and lip synchronization technologies[8] will

8) 3-D rendering is a computer emulation of generating an object in 3-dimensional mode on the screen. This has an effect of rendering the figure looking more vivid than in 2-D mode. Lip synchronization is a technology to render the lips and the muscles around the lips moving in more natural manner by mimicking the movement of lips and the muscles when we utter the target expressions. Both require quite an advanced level of computer technology to be fully developed yet.

make the image on this panel looking more vivid and friendly to render activities more productive and fun.

4.2. Function Buttons

This is a menu item that displays five functions of the program, i.e., conversation, dictation, read-aloud, listening comprehension, and review practices. Students may select one function from these by saying one of the V-COMS (i.e., saying the name of the function or command in voice).

4.3. Visual Panel

In this pane, voice input patterns from both learners and computer are displayed in wave format. The pane is divided into two: one for wave pattern produced by the computer and another one for the learner's wave pattern. The reason for this is to give chances for students to practice pronunciation of the input by referring to model wave pattern displayed.

4.4. Text Interface Panel

This pane is for text interface between students and *K*-buddy. And, this is a sole part in which keyboard and mouse interface, the de facto standard in most current computer operating systems including previous versions of

MS Windows, Macintosh OS, and UNIX OS environments so far, is reserved to serve the conventional goal. Otherwise, SUI in V-COMS takes full control of whole program.

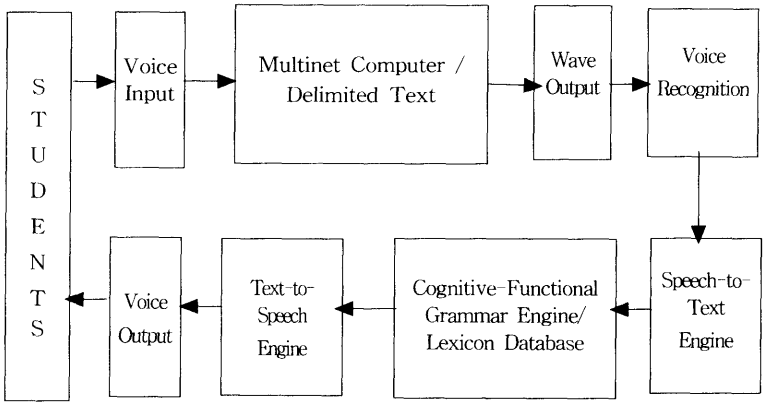
4.5. Multi-Function Panel[9)]

In this pane, various wizards for miscellaneous functions of *K*-buddy, i.e. buddy selection, dictionary checking, idiom primer, grammar book, cultural tips, and general help are provided. Students can activate any of these functions by saying corresponding V-COM. For example, buddy selection can be activated by providing a list of preferences on gender, age, race, physical characteristics, background language, etc., of desired partner. Once the list is completed, assorted top-notch computer and internet technologies mentioned in 4.1 above compose a cyber avatar that is to act as a buddy in the interaction through the video panel.

5. Basic Flowchart of *K*-buddy

K-buddy has the following flowchart for the processing of inputs and outputs:

9) Yet to be realized.



Picture 2. Flowchart of *K*-buddy

Students' voice responses including V-COMS and text input to the prompts made by *K*-buddy are perceived and processed by multinet-ready computer equipped with relevant hardware specifications given in table 1 and related voice recognition-production capabilities within delimited text.[10] The computer, in turn, produces wave output and send it to *K*-buddy for further processing. *K*-buddy, being equipped with devices such as voice recognition-production and cognitive-functional grammar(CFG) engines, and a lexicon database, processes the input in the following sequence:

1) speech-to-text conversion
2) checking the lexicon database for matching lexical items

10) Refer to section 7 for more description on text delimitedness.

3) checking the CFG engine with for grammaticality and acceptability of the input expressions within delimited text
4) text-to-speech conversion

In the first step, *K*-buddy converts input wave form into matching string of text based on the morphemerecognition. The resulting strings are forwarded to cognitivefunctional grammar engine to be checked their grammaticality and cognitive-functional acceptability within given delimited text. If it is turned out that the input strings are not found in the given delimited text or not matching any of morphemes in the text, the program issues warnings to notify students that the responses are not correct. It then gives students options to choose between either to continue or to retry in voice.[11] On the other hand, once the responses pass the checking, they are forwarded to next step to be checked their grammaticality and acceptability in the cognitive-functional grammar engine.

In the whole interaction, use of mouse and keyboard is minimized. Instead, V-COMS take their place to satisfy the SUI specification.

---

11) If the responses to not match the database, the program activates its wisdom to issue a warning in wave format by saying "Sorry you are wrong. Please say it again or just say 'next' to continue the interaction". Database is fully loaded with all sorts of possible responses. If there are more than one possible responses in the database, the engine automatically choose the first one in the database in which items are listed in the order of probability.

6. Two Engines: Voice Recognition/Production[12] and Cognitive-Functional Grammar

In the center of *K*-buddy are two engines, one for voice recognition-production and another one for cognitive-functional grammar. For voice recognition and production, *K*-buddy adopts an engine developed by Lernout & Hauspie's that was released in June 2000. This engine is well known for its high accuracy rate and processing speed, and naturalness of voice in the speech synthesis. For the latter, an engine developed by Team Lingua, a small venture research group in Korea university, is adopted. Adoption of the engines accelerates processing speed of voice inputs and speech synthesis on one hand and improves the accuracy of the response, which, by voice recognition alone, might have been much more difficult to achieve.

7. Technical Limitations and Wish List

K-buddy is the world-first prototype with fully-functional voice recognition and speech synthesis features implemented. Also, it is the first among its kinds with such an advanced level of spontaneity and interactivity of the program. However, it still has much more to be desired. Following is

12) Also known as speech-to-text (STT) or speech perception and text-to-speech (TTS) or speech synthesis, respectively.

the list of limitations:

7.1. Text-delimitedness

K-buddy is delimited to the specific text limited in many aspects. This means that the spontaneity of conversation is achieved only within context of the text selected. This is to guarantee the highest accuracy and fastest interaction between computer and learners in given intranet environment. It is expected that a modular and sub-language approach would render *K*-buddy free of this limitation in the future, though.

7.2. Noise filtering

Noise is a most critical interfering factor in *K*-buddy. There have been numerous efforts in the field of noise filtering/reduction technology to reduce a negative effect of ambient noise in the perception of human voice. Despite significant success in the industry, noise still causes tremendous trouble in the recognition process. One practical method of reducing the noise interference is to use *K*-buddy in ideally quiet place such as language lab or one's own room with less ambient noise. Another option is to use the program with a high quality microphone with direction detection feature.

7.3. Morpheme as a unit of voice recognition

Whole capability of *K*-buddy is bound to features of the STT-TTS engines. In other words, unit of voice processing of the L&H voice recognition engine is morpheme. Thus, it crucially reduces *K*-buddy's sensitivity to supra-segmental features of natural human language such as tone, accent, intonation, length, and so on not to mention various phonetic features.

7.4. Sensitivity to Individual Speech Style

Due to the reasons mentioned in 7.3 above, *K*-buddy is critically sensitive to the speech style of an individual's. Especially, differences in gender, voice color, and dialects may negatively affect result of the voice processing seriously.

7.5. Intranet connectivity

Due to massive size of database for voice processing,[13] *K*-buddy is currently available only for intranet environment. However, the continuing effort in the information and technology industry will improve the quality of the internet connection dramatically very soon. Once this happens, *K*-buddy can be easily transported to an internet-ready

13) Combined technology of STT and TTS reduces the load dramatically. But it is still quite a burden through internet.

format so that it can be used in the internet environment.

8. Concluding Remarks

K-buddy is a prototype developed as a CALL or CAI (Computer-Aided Instruction) device for the purpose of assisting learners of Korean as a foreign language who are in need of more practices with native speakers. It has new features such as voice recognition, speech synthesis, and near-spontaneous interactivity. Despite all these top-notch features, it has some crucial limitations, too. However, *K*-buddy has shown the future of CALL in connection with new-born technologies in the fields such as the internet and computer technologies. Most of all, highly improved interactivity of the program has been made possible thanks to a fine cognitive-functional grammar engine in conjunction with an advanced voice processing technology.

Considering all mentioned above, *K*-buddy presents future of CALL in many significant aspects despite its many-folded limitations.

References

Bachman, L. F. (1990), *Fundamental Considerations in Language Test- ing*, Oxford: Oxford University Press.

Brown, H. D. (2000), *Principles of Language Learning and Teaching* (4th ed.) Englewood Cliffs, NJ: Prentice-Hall.

Canale, Michael and Swain, Merrill (1980), Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1:1-47.

Canale, Michael (1983), From communicative competence to communicative language pedagogy. In Richards and Schmidt 1983.

Cho, H. and Jeongun Lee (2000), Joint project of developing on-line Korean language education material between Yonsei university (Center for Korean Language Education) and Joongang Daily (Joins Learning). *Proceedings of a Conference on Korean language education and Information Science*. Seoul: Center for Language and Informatics, Yonsei Univ..

Choe, J.S. (1999), The Internet (Web) as the new instructional Medium. *Journal of Korean Language Education*. 10.2. Seoul: The International Association For Korean Language Education.

Choe, J.S. (2000), Progress report on the development of Web-based Korean language education program at Sogang University. *Proceedings of a Conference on Korean language education and Information Science*. Seoul: Center for Language and Informatics, Yonsei Univ.

Grice, P. (1975), Logic and conversation. In P. Cole and J. Morgan (eds.) *Syntax and Semantics*, Vol. 3. Speech Acts (pp. 41-58). New York: Academic Press.

Ji, Hyunsook and Hyojeong Lee (1999), Korean instruction through the internet: 'Novice-Net Korean I' at Sogang University. *Journal of*

Korean Language Education. 10.2. Seoul: The International Association For Korean Language Education.

Kang, Inae (1996), Computer Network-ey uyhan swuepkwa kwusengcwuui: kyoywukceok hwalyongkwa uymi (Class Instruction in Computer Network and Constructivism: It's educational application and the significance), *Cengpokwakhakhoyci* (Journal of Information Science). 14(12). 15-29. Seoul: Cengpokwakhakhoy.

Richards, Jack C. and Schmidt, Richard. eds. (1983), Language and *Communication*. London: Longman Group Ltd.

You, S. (2000), Interactive Conversation Practice. Special Lecture on the language education through Internet. The 6th AATK Workshop. Los Angeles.

유석훈(Seok-Hoon You)
고려대학교 언어과학과 교수
주소 : (136-701) 서울시 성북구 안암동 5가 1번지
고려대학교 문과대학 언어과학과
Tel : 02-3290-2176
E-mail : syou@korea.ac.kr