

# **Phonology and CALL: A Review of the Research**

Jacqueline Norris-Holt

---

*Studies in Humanities and Cultures*

---

**Vol. 2**

**GRADUATE SCHOOL OF HUMANITIES AND SOCIAL SCIENCES**

NAGOYA CITY UNIVERSITY

NAGOYA JAPAN

JANUARY 2004

# Phonology and CALL: A Review of the Research

Jacqueline Norris-Holt

## Abstract

In recent years an increasing number of EFL/ESL language programs have begun incorporating CALL into their curricula for the purpose of enhancing the pronunciation skills of students in their L2. This paper examines the use of such computer-assisted instruction and its effectiveness in improving the communicative competence of language learners. Speech technology covers a broad range of CALL applications and can be used in systems for the training of segmental and suprasegmental phonology. Segmental phonology evaluates learner competence in terms of phoneme production while suprasegmental features of speech include rhythm, stress and intonation. A number of applications are described and evaluated to determine their value in assisting the learner in the L2 classroom.

*Keywords* : *CALL, speech technology, segmental phonology, suprasegmental phonology*

With advances which have taken place in multimedia technology, and especially in the area of computer-assisted language learning (CALL), a new field of alternative teaching methods have become available. This is not to say that these methods will completely replace the more traditional ways of teaching, but will provide a supplement to instruction and teacher/student interaction that takes place in the classroom. In the last twenty years the importance of developing effective speaking skills in the L2 has received considerable attention from educators involved in the field. EFL/ESL learning environments have begun to focus on improving the communicative competence of students engaged in the study of a second language. This particular direction in language teaching has generated a growing need for instructional materials that provide an opportunity for “controlled interactive speaking practice outside the classroom” (Ehsani & Knodt, 1998, p. 45).

When considering the effectiveness of CALL for use in phonology in the language classroom it is first important to examine the amount of time a student spends in actual contact with the teacher. In an interesting study conducted by Heuston (cited in Donahue, 1999) it was found that individual students spent approximately 10 seconds with the teacher for every hour spent in a classroom. This period of time being represented as one full week over the 13 year period of attendance at school. “By extension, this ten-second-criteria would translate to 450 seconds or about eight minutes of individual treatment for an ESL college level pronunciation course” (Donahue, 1999, p. 1). Donahue also suggests that while the classroom

environment may be enhanced by the supportive use of the language laboratory, time restraints and student numbers restrict lengthy teacher/student interaction. CALL can provide students with the opportunity to actually learn pronunciation and monitor progress at the same time. The classroom however, often focuses on the teaching of pronunciation rather than actual practice. CALL applications move the focus away from teaching to learning, providing students with an educational experience in which they can gain access to a facility and work at a rate they consider suitable.

Eskenazi (1999) points out that while young language learners may be able to produce new sounds in a language with relative ease, adults need to feel motivated and possess self-confidence to produce otherwise unfamiliar sounds not present in their L1. Learners who feel uncomfortable about making such utterances have a higher risk of performing badly or simply abandoning language learning. CALL can provide an ideal environment in which to develop language proficiency.

In recent years an increase in international students in many countries and limitations on existing ESL programs has led to concerns regarding the effective teaching of pronunciation. Although many overseas students are admitted to university with a reasonable understanding of the English language, their ability to make themselves understood is limited. Poor pronunciation of the language has prompted the need to educate such students to be able to communicate effectively in their L2 (Molholt, 1988).

Speech technology covers a wide range of CALL applications and can be used in systems for training segmental and suprasegmental phonology (Ehsani & Knodt, 1998; Eskenazi, 1999; Jo, 1999; Pennington & Esling, 1996). Segmental phonology assesses learner competency in terms of phoneme productions, such as vowels and consonants; while suprasegmental phonology evaluates the features of rhythm, stress and intonation. This paper aims to explore both areas of phonology and the effectiveness of CALL in providing non-native speakers with an opportunity to acquire native-like language proficiency.

### **Segmental Phonology**

Within the area of segmental phonology a number of studies have been conducted to determine the value of using computer technology to effectively teach language learners.

According to Goh (1993) there are a number of advantages in using computer software to teach pronunciation. Students are provided with a method of practicing, with some guidance, the pronunciation of words or phonemes without teacher intervention or supervision. They may also keep a record of their progress by recording verbalizations on disk to compare with utterances made at some later stage in L2 development. Jo (1999) suggests that CALL provides students with a powerful self-access facility where

they can make their own decisions about when they wish to study and how long they want to work. Hammerly (cited in Jo, 1999) found that adult language learners often preferred the option of using self-instructional programs, as they were allowed to work independently of the teacher, not having to constantly follow instructions. This being a common feature of the classroom environment.

Software used in the area of pronunciation training can also be tailored to meet the needs of various ethnic groups and the particular problems they have acquiring a second language. The programs utilized by Goh (1993) targeted Asian students learning English as a second language, with lists of words containing specific phonemes that Asians had predominantly more problems with, being a feature of the software. Goh (1993) concluded that although the initial responses of the users to the program were favorable, on working with the program a number of problems were identified. Some concern was expressed regarding the accuracy of the speech teaching algorithms. Specifically, in word teaching the algorithm was sensitive to the neighboring phonemes and did not monitor the phoneme being taught. As the software was a first attempt by Goh to address the teaching of pronunciation by aid of a computer, he suggested that many features of the program could be improved for future use with L2 learners.

In a study conducted by Hiller, Rooney, Vaughan, Eckert, Laver and Jack (1994) SPELL (Interactive System for Spoken European Language Training) is examined as a useful tool for the automated evaluation and improvement of foreign language pronunciation. The aim of the SPELL project was to incorporate a set of speech processing algorithms with a demonstrator pronunciation teaching system. The demonstration system was composed of units for teaching information about consonants, vowel quality and other features of speech to students learning English, French and Italian. The system also aimed to provide students with immediate corrective feedback (Ehsani & Knodt, 1998).

The software designed for use with SPELL pronunciation consisted of four phases. The initial part of the program involved the use of a computer to provide the student with audible demonstrations of utterances in the target language. Following this the students completed a small number of tests in order to determine their ability to perceive the particular sounds presented. Once students were able to make such judgments they were then required to familiarize themselves with features of the target language by practicing pronunciation. In this part of the program the students were provided with quantitative feedback and further directions for modifying utterances. Finally, the students were formally assessed to determine their ability to make verbalizations in the L2.

The SPELL system pays particular attention to features of the target language that will give students the greatest learning difficulty or those which will provide the greatest benefits for establishing effective

communication. This is also the focus of methods employed to assist East Asian language speakers, and in particular speakers of Chinese to acquire reasonable levels of English pronunciation to communicate (Molholt, 1988).

In some of the modules developed to teach consonants common errors in each language pair have been identified (Hiller et al., 1994). Using minimal pair examples to emphasize the contrast between the L1 and the L2, the language learner is made more aware of sound differences. Hiller et al. (1994) provide the example of a native Italian speaker learning the English sound /th/ in isolated words. Using a Microsoft Windows ® graphics environment, the teacher window displays the word which will be pronounced along with an appropriate picture.

The teacher window PLAY button allows the students to listen to the teacher's model. The student pushes the SAY button to record and analyze her/his own attempt. The result of the analysis is displayed in the student's window; if the student has pronounced the word correctly then the teacher's picture appears, otherwise the picture appropriate to the error made is displayed. Pushing the NEXT button lets the student choose another word for practice (Hiller et al., 1994, p. 55).

The SPELL vowel-teaching program provides the student with monosyllabic words to practice vowel sounds in the target language. Appropriate vowel sounds are derived from a set of vowel tokens produced by a group of native speakers. The student's speech sample is then compared against the programs to determine if the vocalizations made occur within the range or target vowel space allocated. Feedback is provided in the form of a graphic display enabling the student to see a visual representation of the target vowel and what they have produced verbally.

Although the system appears to be a very effective means of developing language proficiency, Ehsani and Knodt (1998), cite some problems with the program. One of the main criticisms is that although found to be effective in correcting known problems of L1 interference, it was less effective in detecting more idiosyncratic pronunciation errors. The program assumes that the phonetic system of the L2 is similar to the L1. There are few problems when the two languages are phonetically similar, however it does not work well for languages, which have dissimilar sounds.

Molholt (1988), in an article focusing on the needs of Chinese students learning English points out the benefits of using CALL applications, specifically a Speech Spectrographic Display (SSD) 8800, to successfully overcome pronunciation problems in the L2. The article looks at three levels of pronunciation, these being phoneme, word and sentence. From research which has been conducted in the area of

phonology it has generally been established that consonants occurring in Chinese have a higher frequency than those which occur in American English. Speakers of Chinese are also unfamiliar with the duration of some phonemes in English, as they do not occur in their L1.

Since Chinese has no voiced stops and only one voiced fricative, the language in general has a higher frequency range than English. Therefore, it is important at the beginning of pronunciation lessons for Chinese students to start building more sensitivity to sounds in the low-frequency range. One way to do this is to help them learn how to control frequency. For example, by pronouncing /s/ with the tongue very near the teeth, we have a very high-frequency sound. As the tongue is gradually moved back along the alveolar ridge and onto the palate, the frequency lowers (Molholt, 1988, p. 95).

By the use of a visual display students are able to learn how to produce the appropriate sound by associating the display with the actual feeling of making the sound. For many students the learning step of being able to see a visual display assists them to reach the final stages of acquisition much faster. The visual display provides students with an objective measure whereby they can focus attention on the exact features of the sound, which need to be altered. Although it may be difficult for students to recognize sound differences in the initial stages of language acquisition, they can feel and see the difference on the computer screen. Another important feature of the learning process is that as students are provided with visual information they do not need to firstly learn the linguistic vocabulary associated with a phonological analysis (Molholt, 1988). This is a common problem faced by many second language learners.

A recent CALL application has been the development of a system for teaching the pronunciation of long vowels in Japanese, at the University of Tokyo (Kawai & Hirose, cited in Ehsani & Knodt, 1998; McBride, 1990). With this particular program students are able to practice phonemic differences in Japanese, which often present a problem for L2 learners. Students are given minimal pairs to work through, with both long and short vowels, of which they are then given immediate feedback regarding segment duration. Research has shown that learners are quick to master the relevant duration cues and that the time spent on acquiring the particular pronunciation skills is within the limits of the Japanese L2 curricula. However, what must be noted is that no data is yet available regarding the long-term effects of using the system

There are a number of other techniques that have been developed for automatic recognition and the evaluation of non-native speech. One particular system developed by Rochet (cited in Pennington & Esling, 1996) makes use of a Hypercard on the Macintosh to enable students to recognize and be able to produce a vowel distinction that occurs with English and French. "His system is designed to train first the most central representatives, or "exemplars" of the /ü/ phoneme, then the more peripheral targets, only later

adding “distractors” representing typical errors made by English speakers learning French” (Pennington & Esling, 1996, p. 166). The program also allows learners to listen to sounds in the target language and assist them to discriminate between /ü/ and /u/. The program has a further facility producing target sounds and errors, allowing students to distinguish those phonemes, which do not form part of the repertoire of sounds of the language. This is a useful feature of the program as it enables students to focus on sounds which are perhaps a familiar component of the L1 but do not form part of the sound system of the L2.

### **Suprasegmental Phonology**

CALL applications in the area of suprasegmental phonology can be used to assist the language learner with pitch, intensity, duration and the location of pauses (Jo, 1999; Lian & Lian, 1997; McMeniman & Evans, 1998; Pennington & Esling, 1996). Much of the literature points out that suprasegmentals are an important component of utterances, not only directing the learner’s attention to important information in discourse, but also helping to bring about cultural understanding between a speaker and a listener (Anderson-Hsieh, 1992; De Bot & Mailfert, 1982; Hermes, 1998). It is also suggested that many ESL students spend very little time focused on this area of language acquisition, being unable to hear these features of speech, or for that matter reproduce them in the L2. It is for this reason the use of electronic visual feedback can be useful in assisting to facilitate the perception and production of suprasegmentals in the second language learner (Spaai & Hermes, 1993).

Cranen, Weltens, De Bot and Van Rossum (1984) suggest that it is very difficult to both teach and explain the area of intonation, and that even experts in the field of linguistics find evaluating this feature of language acquisition quite challenging. They propose that language learners therefore have considerable trouble in both perceiving intonation in a L2 as well as making the appropriate utterances when learning the new language. For this reason the use of an appropriate aid can benefit the L2 learner when attempting to develop language competence. In one research experiment, which was carried out to assess the effectiveness of visualization intonation contours, one Dutch group studying English and another control group were investigated to determine what effects the use of visualization had on language development (Cranen et al., 1984). Upon examination of the results it appeared that the subjects who had practiced visualization had significantly improved. The same students were also found to be eager to continue with the use of the equipment at the conclusion of the experiment. In a further study (Cranen et al., 1984) the efficiency of pitch visualization was examined on a group of Turkish subjects, having been selected due to their apparent difficulty in acquiring Dutch intonation. Turkish subjects were also selected, as their language is quite different from Dutch and that of the English-Dutch combination of the first research group. “The results of this experiment clearly showed that nearly all subjects improved their imitations of

Dutch intonation contours, irrespective of their general proficiency in Dutch” (Cranen et al., 1984, p. 28).

Although much of the research indicates a positive link between the use of visual feedback and the development of language proficiency there are a number of problems, which have become evident with the system. One of these being, that not all utterances are suitable for plotting on a screen for visualization. If, for example a considerable number of voiceless segments occur, an intermittent contour will result, making it difficult to interpret the information presented. The microphone is also susceptible to background noise, which distorts the pitch contour plotted on the screen. Weltens and De Bot (1984) point out that the feedback provided to students must be clear and interpretable if language competency is to develop and improvements are to be made in the area of CALL and intonation.

In a study conducted by Anderson-Hsieh (1992) a group of Chinese and Korean students studying English were instructed for a 6-8 hour period, focusing on stress, rhythm, linking and intonation with the use of electronic visual feedback. The results of the study indicate that the students did benefit from the use of a CALL application and that they were assisted to develop more native-like production. The students were found to correct their mistakes after a few attempts, with only an occasional student reporting limited assistance from the feedback. It is however, suggested that having made the appropriate corrections to speech, does not mean the student will no longer make the same errors at some future stage. Such patterns of self-correction and experimentation are common among learners and seem to assist them in understanding their own pronunciation and at the same time the overall learning process involved. Anderson-Hsieh concludes that electronic visual feedback is useful to teach features of the suprasegmental system, but should be used in conjunction with other learning activities. “Thus, electronic visual feedback is used as a tool to raise the student’s awareness of suprasegmentals so that they can more easily practice them outside of class and acquire them through communicative use of the language” (Anderson-Hsieh, 1992, p. 61).

Other major advantages of electronic visual feedback are that it provides the learner with an accurate visual representation of suprasegmentals in real time. Students are able to more easily attempt to reproduce the target language by visual stimulus and the comparisons they make with their own speech production. Students have also been found to become less self-conscious about pronunciation in the target language as they focus on the visual display. While the system provides numerous benefits for the learner, it is by no means the answer to successful language acquisition. Students must practice what they have gained from using the system and they must constantly monitor their speech production. They must also be able to make the transition from such language exercises to the communicative use of the language.

In another study conducted by Hermes (1998) the audible differences between two pitch contours and the visual differences between displayed pitch contours were examined. Although the experiment was carried out with experienced participants, a number of important findings emerged. The subjects found visual comparisons much easier to perform than auditory comparisons. This was determined from the amount of time taken to perform the visual task, as compared with the auditory one. The participants were also observed using the repeat button more when the auditory difference between the two utterances was relatively small. In contrast to this, the visual difference between utterances was established quickly. As the subjects involved in the study were experienced in the area of speech and intonation it would seem to indicate that a L2 learner would have difficulty detecting differences in intonation between a L1 and a L2. This is perhaps one important reason why visual feedback may be more effective than auditory feedback in intonation teaching. Although the study did not measure the cognitive load of the tasks, the results suggest that comparing visual contours is less demanding than making auditory comparisons (Hermes, 1998).

Other software, which has become available and is possibly more appropriate for use with younger learners, makes use of a different type of display. This may take the shape of an appealing graphic, which is used to indicate either the success or failure of a learner's utterance. Rather than relying on a pitch contour display, a picture of an animal may be utilized. The example provided in Pennington and Esling (1996) is that of a giraffe, in which the neck becomes longer or shorter in accordance with the accuracy of the student's verbalization in the L2. As the giraffe's neck grows longer it is able to reach a number of objects positioned in different places around the screen. The aim of the exercise being, to collect all the objects in the picture.

It has been suggested that the quality and quantity of 'input' is a key factor in the acquisition of a foreign or second language. In an ideal situation this would amount to considerable teacher/student interaction within the confines of the classroom. However, the reality is that class sizes do not allow for this extended direct contact with the teacher. In such a situation CALL can be utilized to enhance the learning environment of the student and provide them with the additional input they require to develop speaking skills in the L2 (McBride, 1990). Students are given an opportunity to work with a computer display where they can compare their own pronunciation with that of a native speaker's model and attempt to match it. They are provided with objective corrective feedback and do not need to rely on their own perceptions.

To date one of the main criticisms of CALL is the lack of a unified theoretical framework for both designing and evaluating software programs (Ehsani & Knodt, 1998). Many researchers note that while there is a positive reaction to the use of CALL applications, the majority of language teachers are unable to exploit the potential of electronic devices to assist L2 learners (De Bot & Mailfert, 1982). While considerable research still needs to be done in the field teachers need to make a conscious effort to

incorporate CALL in the language classroom.

## References

- Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System*, 20(1), 51-62.
- Cranen, B., Weltens, B., De Bot, K. & Van Rossum, N. (1984). An aid in language teaching: The visualization of pitch. *System*, 12(1), 25-29.
- De Bot, B. & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, 16, 71-77.
- Donahue, S. 1999, Teaching Intonation Online [Online]. Available: <http://www.broward.edu/~sdonahue/teaching-online.html>  
[Accessed 31 July. 2001].
- Ehsani, F. & Knodt, E. (1998). Speech technology in computer-aided language learning: Strengths and limitations of a new CALL paradigm. *Language Learning & Technology*, 2(1), 45-60.
- Eskenazi, M. (1999). Using automatic speech processing for foreign language pronunciation tutoring: some issues and a prototype. *Language Learning & Technology*, 2(2), 62-76.
- Goh, I. (1993). A low-cost speech teaching aid for teaching English to speakers of other languages. *System*, 21(3), 349-357.
- Hermes, D.J. (1998). Auditory and visual similarity of pitch contours. *Journal of Speech, language and Hearing Research*, 41(1), 63-73.
- Hiller, S., Rooney, E., Vaughan, R., Eckert, M., Laver, J. & Jack, M. (1994). An automated system for computer-aided pronunciation learning. *Computer Assisted Language Learning*, 7(1), 51-63.
- Jo, C. -H. 1999, Studies on Computer-Assisted Pronunciation: Learning Systems for Non-Native Learners based on Speech Recognition Techniques [Online]. Available: <http://winnie.kuis.kyoto-u.ac.jp/members/chjo/main/node7.html>  
[Accessed 31 July. 2001].
- Lian, A. & Lian, A. (1997). The secret of the shao-lin monk: Contribution to an intellectual framework for language learning. *ON-CALL*, 11(2), 2-18.
- McBride, J. (1990). Computer assisted instruction in the teaching of Japanese as a second language. Commonwealth of Australia. Print Team.
- McMeniman, M. & Evans, R. (1998). CALL through the eyes of teachers and learners of Asian languages: Panacea or business as usual? *ON-CALL*, 12(1), 2-9.
- Molholt, G. (1988). Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly*, 22(1), 91-111.
- Pennington, M. & Esling, J. (1996). Computer-assisted development of spoken language skills. In M. C. Pennington, *The power of CALL* (153-189). Houston, TX: Athelstan.
- Spaai, G. & Hermes, D. (1993). A visual display for the teaching of intonation. *The CALICO Journal*, 10(3), 19-30.
- Weltens, B. & De Bot, K. (1984). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech*, 27(1), 79-88.