SIGNAL ANALYSIS SOFTWARE
FOR TEACHING DISCOURSE INTONATION

Dorothy M. Chun
University of California, Santa Barbara

ABSTRACT

In the last fifteen years, there have been major paradigm shifts in both general and applied linguistics toward acknowledging intonation as an indispensable component of language and communication. In addition, the hardware and software for conducting acoustic phonetic signal analysis have recently become more accessible. The main goal of this paper is thus to integrate the two seemingly disparate subfields of linguistics, acoustic phonetics, and discourse intonation, and to suggest a new framework for facilitating and studying the acquisition of suprasegmental phonology. The purpose of this article is threefold: (1) to review previous research on the acquisition of suprasegmentals by L2 learners and the potential of computer-based instructional materials for improving intonation; (2) to briefly describe and critique some of the software previously available for this purpose; and (3) to suggest criteria for the conceptualization of multimedia software and concomitant research on the teaching of discourse-based phonology and intonation. In looking toward the future, this article will focus on providing learners with discourse-level language input and with specific feedback regarding acoustic features of the intonation patterns they produce. Finally, the article urges that software be designed to include both research tools and tools to facilitate, record, and analyze the intonation produced in real interactions between speakers.

INTRODUCTION

Intonation is said to be among the first aspects of speech that human infants attend to, react to, and produce themselves (Lieberman, 1986). It has also been shown that child L2 learners have little difficulty in acquiring native-like pronunciation and intonation in the L2 (Felix, 1978). Intonation thus seems to be easily, if not automatically, acquired by children in both L1 and L2. However, although it appears easy for adults to maintain and retain intonation in their L1, it is difficult, if not seemingly impossible for them to learn L2 intonation. Due to its inherent complexity and to the ensuing difficulty in learning and mastering it, intonation was ignored for many years in language teaching. However, it is slowly gaining recognition as an integral part of language fluency, competence, and proficiency.

There appear to be several reasons for the generally growing interest in intonation. First, in theoretical linguistics, there have been important new advances in the theory of intonation and its representation, aided by the growing accessibility of acoustic signal analysis. Second, the expansion of the analytical domains of traditional linguistics from sounds (phonetics and phonology), words (morphology), and sentences (syntax) to larger units of inquiry, such as entire texts, discourses, and interactions, has given rise to such subfields as pragmatics, text linguistics, discourse analysis, and conversation analysis (Brazil, 1975; Brown, Currie, & Kenworthy, 1980; Couper-Kuhlen & Selting, 1996). In addition, the realization that language and language acquisition are affected and influenced by other disciplines has strengthened the interdisciplinary focus of linguistics to stress psycholinguistics and sociolinguistics. Third, applied linguistics has grown to emphasize communicative function rather than linguistic form (cf. Canale & Swain, 1980; Kramsch, 1995). With particular regard to the teaching and learning of pronunciation, there has been an appeal to adopt a top-down approach, namely, to focus more on global meaning and

communication (context- and discourse-based instruction), rather than adhere to the traditional bottom-up, phonemic-based approach (based on isolated or contrasted sounds).

In accordance with these major paradigm shifts in both general and applied linguistics toward acknowledging intonation as an indispensable component of language and communication, as well as with the advances in acoustic phonetic technology, this paper has three main goals: (a) to review previous research on the acquisition of suprasegmentals by L2 learners and the potential for computer-based instructional materials to aid in improving intonation; (b) to briefly describe and critique some of the software previously available; and (c) to suggest criteria for the conceptualization and design of multimedia software for the teaching of discourse-based phonology and intonation. It is the third goal that presents a distinct challenge and suggests a new framework for studying the acquisition of suprasegmental phonology.

PERCEPTUAL AND ARTICULATORY TRAINING BEYOND SEGMENTALS: A BRIEF HISTORY OF PREVIOUS RESEARCH

The traditional theoretical linguistic basis for the learning and teaching of pronunciation was a focus on the segmentals, that is, the articulatory phonetics of individual sounds. Intonation, comprised of the so-called suprasegmentals, had not been as extensively researched theoretically or acoustically and was considered a "luxury" in terms of teaching. However, since the early 1980s there have been appeals to reverse the emphasis from segmentals to prosodic patterns (cf., for example, Chun, 1988; Hurley, 1992; Leather, 1983; Luthy, 1983; Morley, 1991; Pennington & Richards, 1986; Wennerstrom, 1994, 1998). There is also a growing literature on the importance of teaching intonation to L2 learners, as well as on the role of prosody in foreign accents (cf. Purcell & Suter, 1980; Scovel, 1988; Van Els & de Bot, 1987). For the purposes of this paper, however, we assume its importance and focus on ways of implementing instruction, based on underlying applied linguistic principles.

In the past, various types of training in L2 pronunciation have been studied, including perceptual, articulatory, or some combination of both. Only those studies dealing with suprasegmentals, in conjunction with speech technology systems or programs, will be summarized here. For example, in perceptual or discrimination experiments, Feldman (1973, 1977) used electronic synthesis to simplify several prosodic features that were presented to hearers for discrimination. After this initial training, natural speech models that contained the same prosodic contrasts were presented, and positive results in discrimination were obtained.

'T Hart and Collier (1975) sought to provide Dutch learners with necessary relevant information about English intonation so that they could learn correct intonation. They found that "four aspects of pitch change are perceptually relevant: (a) direction of pitch change (rise, fall, or level); (b) range of pitch change (difference between high and low levels); (c) speed of pitch change (how abruptly or gradually the change happens); and (d) place of pitch change (in sentence, word, or syllable)" (p. 237). These four relevant aspects of intonation were then explained and demonstrated in a 12-minute tape, using single tones and speech samples generated by a computer. The different pitch changes were also presented on a screen. Though the tape was only concerned with the perception of intonation, it caused a significant improvement in the production of intonation.[1]

As for using visual displays with a focus on improving articulation or production (rather than perceptual discrimination), results of studies have also been varied: Although several studies report positive effects of the use of visual displays of intonation for language learners (de Bot & Mailfert, 1982; de Bot, 1983; Hengestenberg, 1980; James, 1976, 1977, 1979; Lane & Buiten, 1969; Leon & Martin, 1972), other studies did not find these effects (Vardanian, 1964; Wichern & Boves, 1980). De Bot and Mailfert (1982) found that a 45-minute training session in the perception of intonation did result in a statistically significant improvement in the production of English intonation patterns by both Dutch and French

students. Subsequent studies using visual and audio-visual feedback further corroborated the 1982 findings: De Bot (1983) showed that audio-visual feedback is more effective in intonation learning than auditory feedback alone.

However, de Bot (1981) cautioned:

> Methods making use of visualisation as an aid in intonation learning wrongly imply, that visualisation by itself is useful in teaching. Experimental phonetic research has shown that there is no empirical basis for a number of assumptions made in these methods. Especially, the tacit assumption that the pupil is an unbiased perceiver of whatever the teacher is presenting leads to many problems, because the perceptual abilities of the pupils in general have not been developed sufficiently and specifically enough for the perception of intonation. (p. 39)

An example of a program that displays visual pitch curves is a product from Kay Elemetrics called Visi-Pitch that has been available for a number of years for DOS-based personal computers (PCs) (Abberton & Fourcin, 1975; Fischer, 1986; James, 1976, 1979; Molholt, 1988).2 With Visi-Pitch, students are able to see both a native speaker's and their own pitch curve simultaneously. Students first speak a sentence into a microphone; their utterance is then digitized and pitch-tracked, and they can see a display of their pitch curve directly under a native speaker's pitch curve of the same sentence. Figure 1 from Fischer (1986) shows the pitch contours of the French question Qu'est-ce qu'il fait? (What is he doing?) as spoken by a native speaker in the top half, and the same question produced by an American learner in the bottom half.
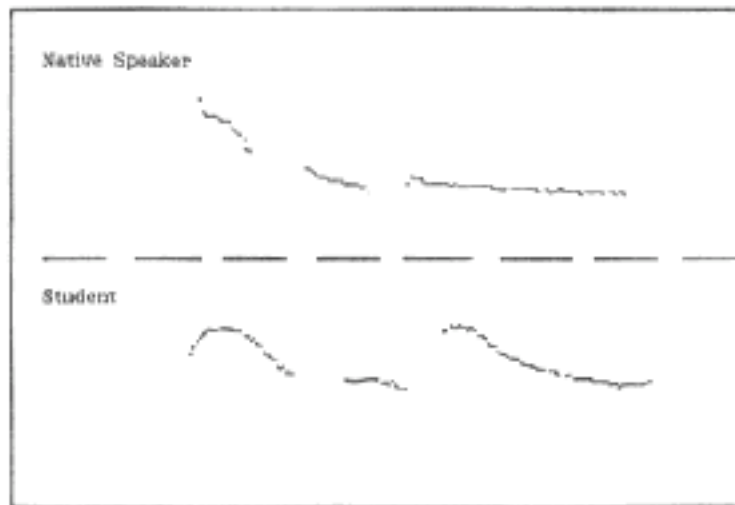

Figure 1. French sentence *Qu'est-ce qu'il fait?* (What is he doing?)

Anderson-Hsieh (1992) reported on using *Visi-Pitch* to teach English suprasegmentals to international teaching assistants and found that

the major benefit of electronic visual feedback for teaching suprasegmentals is that it provides the students with an accurate visual representation of suprasegmentals in real time paired with the normal auditory feedback that occurs during speech. Students can thus more easily replicate native suprasegmental targets using both the target form and the visual feedback from their own speech to guide them. (p. 61)

Although the majority of recent studies have demonstrated the efficacy of audio and visual training to improve learners' perception and production of intonation, the conflicting results of earlier research which will be discussed below can be attributed to several factors. First, the relationship between perception and production is not fully understood, at either the level of segmentals or the level of suprasegmentals. To date, there is no conclusive evidence that a simple relationship exists between learners' perceptual ability

and their productive skills; that is, whether correct perception must precede correct production, whether they can occur simultaneously, or in either order. Nor has it been determined whether these abilities progress in parallel (Leather, 1983). Second, the importance of auditory and/or visual feedback with regard to intonation is difficult to assess because previous programs providing feedback required learners to be able to monitor themselves critically. In other words, in a typical exercise, learners recorded themselves, then replayed their utterance and saw a visual display of their intonation curve. Other than the visual display, no further feedback was provided. This required learners to monitor themselves and to extrapolate their difficulties even though they may well have lacked the phonetic criteria for critical listening. However, it was the objective of the exercise to develop these criteria. Third, questions remain about the long-term effects of any of the brief training sessions reported on in the studies discussed above, and about the effects of more extensive training. Finally, the technical limitations of speech technology systems have also made research in this area significantly more difficult.[3] The limitations of both hardware and software used in earlier studies will be discussed in the following section.

SPEECH TECHNOLOGY SYSTEMS FOR LANGUAGE INSTRUCTION: PROBLEMS AND SHORTCOMINGS

As stated above, research on using speech technology systems for intonation training has been hampered to some degree by the limited capabilities of both the hardware and the software. This section discusses possible reasons why not all previous studies reported on unqualified success of intonation training systems. With regard to hardware, the devices and computers used in previous research were often incapable of dealing with weak speech signals and particularly with voiceless consonants, as lack of voicing means breaks in the fundamental frequency or pitch curves.

Weltens and de Bot (1984) were concerned that the limitations of the hardware and software that caused a slight delay in feedback might impede effectiveness of their display system. However, they showed that "feedback delay is not a critical factor when using a pitch visualizer for intonation teaching, but that the nature of the speech material [voiceless vs. voiced consonants, neutral vs. contrastive intonation]. . .does dramatically affect the quality of the visual feedback" (p. 79). In other words, the fact that feedback was not provided in "real-time" (i.e., with a minimum of delay between the production of a speech signal and its visualization) did not adversely affect the effectiveness of the visualizer. But if sentences contained many unvoiced sounds the quality of the pitch curve was significantly poorer than for sentences with few unvoiced sounds. Another serious difficulty with the early attempts such as the one by Weltens and de Bot was that the devices used were very costly and often "one-of-a-kind." As a result, none of these display systems came into widespread use.

In addition to the technical limitations of the hardware and software, other problems in the above studies concern pedagogical issues related to the underlying assumptions of the software's content and design. For example, subjects in an experiment by Wichern (1979) reported that they had considerable difficulties in relating visual and auditory signals. De Bot (1981) also noted the problem in the application of visual feedback is that the visual feedback does not indicate which parts of the signal are perceptually relevant and which ones are irrelevant from a student's point of view (pp. 38-39). Although only those parts that are perceptually relevant should be paid attention to and imitated as exactly as possible, either students must be able to make these judgments, or the device must be set or programmed to judge student performance as acceptable or unacceptable (cf. Anderson-Hsieh, 1994).

One of the problems with some of the earlier software programs was the lack of feedback processing, for instance, the pitch could be measured and directly fed back to the learner, but interruptions in the intonation contour during unvoiced parts of the utterance, and the inclusion of perceptually irrelevant pitch variations made it difficult for the learner to interpret the feedback. Spaai and Hermes (1993) devised a visual intonation-display system called "Intonation Meter" that presents visual feedback of the intonation as a continuous representation of the pitch contour and contains only the perceptually relevant

aspects of the intonation pattern. It also marks vowel onsets in Dutch, that is, the points at which pitch movement is perceived (see Figures 2 and 3). The Intonation Meter displays pitch contours produced by the foreign language teacher on the upper part of a computer screen, and the student's attempted imitation is shown on the lower part of the screen. Both are presented in stylized form, that is, with a continuous (interpolated) pitch "curve" which eliminates the irrelevant pitch variations. This visual display system is used in a program for training foreign language learners that contains three types of exercises: Auditory discrimination, imitation, and "production on demand" exercises. In the latter, spontaneous production from the learner is elicited by means of visual cues (e.g., sentences, part of stories, or dialogues) (p. 26-27).
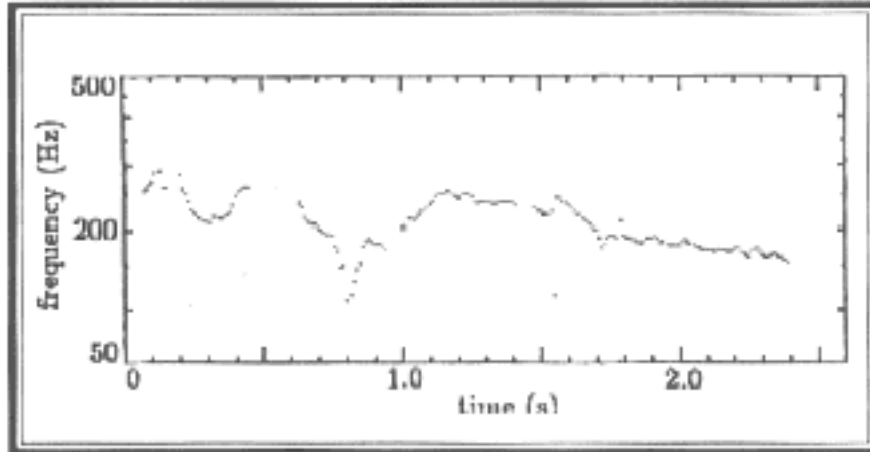

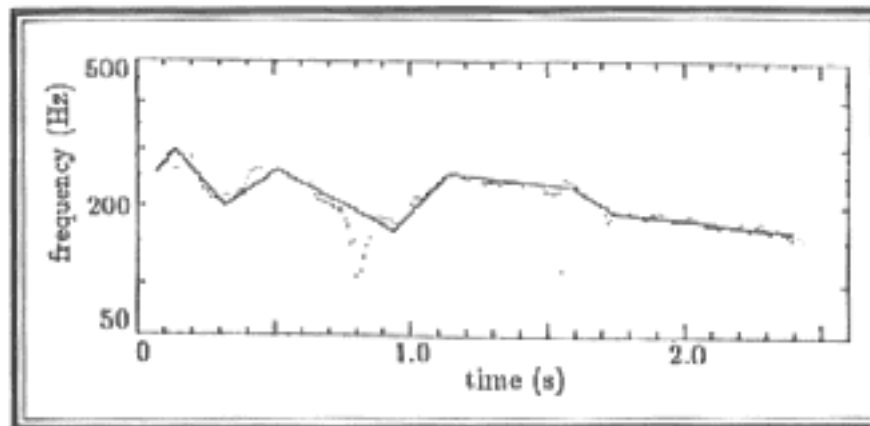Figure 2. "Unprocessed" fundamental frequency measurements

Figure 3. "Unprocessed" pitch contour (dots) and stylized pitch contour (straight line)

Anderson-Hsieh (1994) discussed some of the technical problems with existing software and sought to provide learners with information on interpreting visual feedback graphs of intonation contours produced by computer programs, in particular, those generated by Visi-Pitch. In addressing the problem noted by Spaai and Hermes (1993), she compared pitch contours for sentences with mainly voiced sounds (more or less continuous pitch curves) with sentences containing many voiceless sounds (broken or truncated pitch curves). In Figure 4 below, the final syllable of both sentences is accented, but in the sentence with the word nine, containing vowels and nasals (all voiced), the final accent is shown as a long fall, while in the sentence with the word six, the two voiceless consonants cause the final accent to be cut off and to look like a truncated fall. This could be misleading to a learner. Moreover, deep and truncated falls convey very different meanings in English; Bolinger (1986), for example, attributed finality or assertiveness to deep falls and a sense of offhandedness or tentativeness to truncated falls. The solution to these inherent

problems of voiceless consonants being incapable of carrying pitch is, at least initially, to use model sentences in training that contain mainly voiced sounds, particularly at the end of utterances, in accented syllables, and wherever it is imperative for learners to see the movement of pitch.
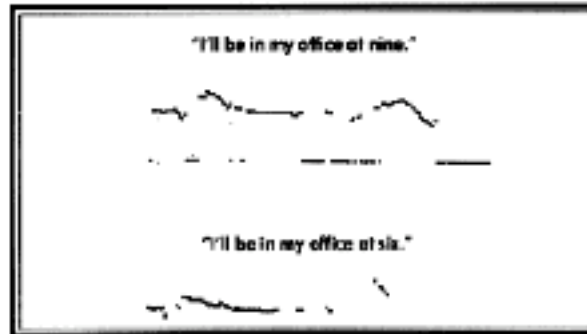

Figure 4. Sentences with voiced vs. unvoiced final words

To summarize, the following main research issues remain unanswered:

- What is the relationship between perception and production ability?
- What is the effectiveness of computer training materials, particularly what are the longer-term effects?

The main shortcomings of hardware and software used previously for training and research can be summed up as follows:

Accessibility and Cost:

- Inaccessibility (often "one-of-a-kind" devices)
- Cost of the equipment required (high-end machines required for the most accurate processing)

Technical Problems:

- Weak speech signals
- No extrapolation for voiceless sounds

Pedagogical Shortcomings:

- Intonation contours were derived from synthesized speech (e.g., simple tones generated by a computer)
- Pedagogical applications were often not user-friendly (i.e., learners were not told how to interpret displays)
- Examples and exercises consisted of sentence-level intonation

In all of the studies discussed above, the units of analysis were sentence-level intonation contours. The next section will discuss the recent calls for teaching discourse-level intonation and ways of integrating this level of intonation teaching with the increasingly accessible signal analysis software tools. New principles for the conceptualization and design of software as well as accompanying research on the effectiveness of using such software for acquisition of L2 intonation will be explored.

LOOKING TOWARD THE FUTURE: BASIC PRINCIPLES FOR CONCEPTUALIZATION AND DESIGN OF SOFTWARE FOR TEACHING INTONATION

Although computers have been utilized for providing learners with visualizations of their intonational patterns since the 1970s, it is only recently that hardware and software for acoustic signal analysis with microcomputers have become increasingly more accessible. This new technology comes in the form of speech digitizers, pitch trackers to produce displays of intonation curves, and a small number of

pronunciation tutors with audio and graphic components. At the same time, in the subfield of applied linguistics, the pedagogical goals with regard to pronunciation and intonation teaching have been focusing increasingly on discourse-level phenomena. It is therefore an appropriate time to integrate two areas of linguistics that have often been placed on opposite ends of the continuum in terms of the scope of their respective domains: Acoustic phonetic analyses, on the one hand, have traditionally dealt with individual vowels and consonants, whereas discourse-level linguistic analyses, on the other hand, have been concerned with entire texts and discourses.

The application of discourse-analytic principles to using computer-based tools for teaching pronunciation leads to the following discussion regarding the design of signal analysis software and the research questions that can be addressed using such software. Pennington (1989a) pointed out that computers can both provide training in the production and perception of speech and create environments that facilitate interaction. Citing the work of Chaudron (1985) and Richards (1986), Pennington stressed the need for language learning software to move to skill-based and task-based learning activities that not only offer users practice in listening comprehension but also elicit and encourage practice of specific types of interactions, language forms, sound contrasts, or nuances of meaning signaled by intonation (cf. also Piper, 1986; Young, 1988). In other words, she extended the scope of intonation practice to include context and transactions, not only sentences.

Some research has already been conducted to expand the scope of intonation study beyond the sentence level and to identify contrasting acoustic intonational features between languages. For example, Hurley (1992) showed how differences in intonation can cause sociocultural misunderstanding. He found that while drops in loudness and pitch are turn-relinquishing signals in English, Arabic speakers of English often use non-native like loudness instead. This could be misinterpreted by English speakers as an effort to hold the floor (pp. 272-273). Similarly, in a study of politeness with Japanese and English speakers, Loveday (1981) found more sharply defined differences in both absolute pitch and within-utterance pitch variation between males and females in Japanese than between English males and females in English politeness formulas. In addition, the Japanese subjects transferred their lower native language pitch ranges when uttering the English formulas. Low intonation contours are judged by native speakers of English to indicate boredom and detachment, and if male Japanese speakers transfer their low contours from Japanese to English when trying to be polite, this could result in misunderstandings by native English speakers. Van Bezooijen (1995) corroborated earlier studies and found Japanese women to have higher pitches than Dutch women for reasons not only of physical and psychological powerlessness, but also due to a stronger differentiation between the ideal woman and man in Japan than in the Netherlands.

As evidence for culture-specificity with regard to the encoding and perception of affective states in intonation contours, Luthy (1983) reported that although a set of "nonlexical intonation signals" (p. 19) (associated with expressions like *uh-oh* or *mm-hm* in English) were interpreted consistently by a control group of English native speakers, non-native speakers of varied L1 backgrounds tended to misinterpret them more often. He concluded that many foreign students appear to have difficulty understanding the intended meanings of some intonation signals in English because these nuances are not being explicitly taught.

Kelm (1987), acknowledging that "correct intonation is a vital part of being understood" (p. 627), focused on the different ways of expressing contrastive emphasis in Spanish and English. He investigated acoustically whether the range of pitch of non-native Spanish speakers differed from that of native Spanish speakers. Previous research by Bowen (1975) had found that improper intonation in moments of high emotion might cause a non-native speaker of Spanish to sound angry or disgusted. Kelm found that the native Spanish-speaking group clearly varied in pitch less than the two American groups; that is, native English speakers used pitch and intensity to contrast words in their native language and transferred this intonation when speaking Spanish. Although the results showed a difference between native and non-

native Spanish intonation in contrasts, they did not show the degree to which those differences affect or interfere with communication.

The studies described above provide us with preliminary criteria for the content of software to teach discourse intonation. Examples of what might be incorporated in such software using a very accessible signal analysis program are discussed below. In addition, it should be borne in mind that experimentation and research to determine the effectiveness of the various methods is also needed (cf. Pennington, 1989b). One of the greatest advantages of using computer-assisted pronunciation and intonation tutors, for example, is that the computer serves both as a medium of instruction and as a tool for research; that is, a software program, while teaching pronunciation, can simultaneously keep detailed and thorough records of student performance and progress. Consequently, the recent studies on the effectiveness of visual feedback in teaching intonation and on the use of technology to develop new measures of speaking proficiency represent the type of ongoing research that must be done in addition to developing better and more efficient computer software. Pennington stated, "Perhaps the most exciting possibilities combining language training, assessment and research involve two-person interactions which are both facilitated and analyzed by a computer" (p. 119).

Morley (1991) echoed Pennington in noting the need for "more definitive evaluative measures and methods to quantify changes and improvements in the learner's intelligibility and communicability" and for "controlled studies of changes in learner pronunciation patterns as the result of specific instructional procedures" (p. 511). Both of these needs can be addressed, first, by continuing to study the acoustic parameters that can be used to quantify intonational features, and second, by building data collection tools into signal analysis software.

In summary, four specific areas in which technology can be integrated into intonation instruction and research can be suggested: (1) providing learners with visualizations of their intonational patterns and with specific feedback to help them perceive the meaningful intonational contrasts between L1 and L2 so that they can improve their speech production; (2) providing learners with authentic and extensive speech and cultural input to represent the diversity of speech sounds and the great variation that exists within a language, and then in turn to hone learners' perceptual abilities; (3) using computer software to facilitate, record, and analyze interactions between speakers; (4) using computers for research purposes by recording students' performance, progress, and steps toward self correction.

Based on the foregoing needs assessment, some specific examples for each of the four areas described above are presented below regarding the conceptualization and design of multimedia software for teaching discourse intonation:[4]

1. Provide learners with visualization of their intonation patterns and specific contrastive feedback.

In intonation teaching, one focus has traditionally been contrasting the typical patterns of different sentence types. Pitch-tracking software can certainly be used to teach these basic intonation contours, but for the future, in accordance with the current emphasis on *communicative* and *sociocultural* competence, more attention should be paid to discourse-level communication and to cross-cultural differences in pitch patterns. Selected examples are presented below and illustrated in Figures 5 to 9. Software programs must have the capability to:

- Distinguish the meaningful intonational features with regard to four aspects of pitch change: (a) direction of pitch change (rise, fall, or level), (2) range of pitch change (difference between high and low levels), (3) speed of pitch change (how abruptly or gradually the change happens), and (4) place of pitch change (which syllable(s) in an utterance)
- Go beyond the sentence level and address the multiple levels of communicative competence: *grammatical*, *attitudinal*, *discourse*, and *sociolinguistic*

- Start with utterances containing a minimum of voiceless sounds in order to ensure decipherable pitch curves, particularly for the initial exercises, where learners must familiarize themselves with the visual representations of the pitch curves.

Grammatical Functions of Intonation. Figure 5 contains three so-called "neutral" utterances typically used to illustrate the grammatical function of intonation: the question Where are you going? and the replies I'm going home and Home. It is interesting to note that all three utterances have similar intonation patterns, typical of English wh-questions and statements. They show pitch peaks on stressed syllables, with the greatest peak on the syllable with sentence stress. At the end of the utterance, the pitch falls to its lowest level. This illustrates the fact that there is not a one-to-one correspondence between intonation pattern and sentence type. In this case, the same pattern is used for different syntactic types.

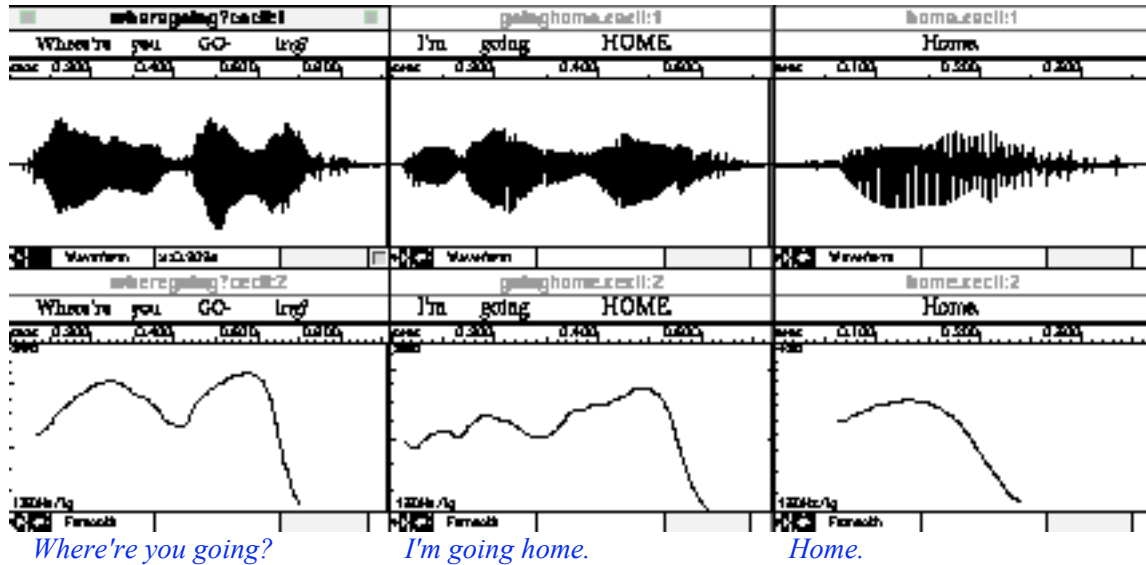*Where're you going?*   *I'm going home.*   *Home.*

Figure 5. "Neutral" question and two statements (replies)

Figure 6 shows the yes-no question *Are you going?* with rising intonation and the reply *Yeah, I'm going* with a falling intonation pattern similar to those shown in Figure 5.

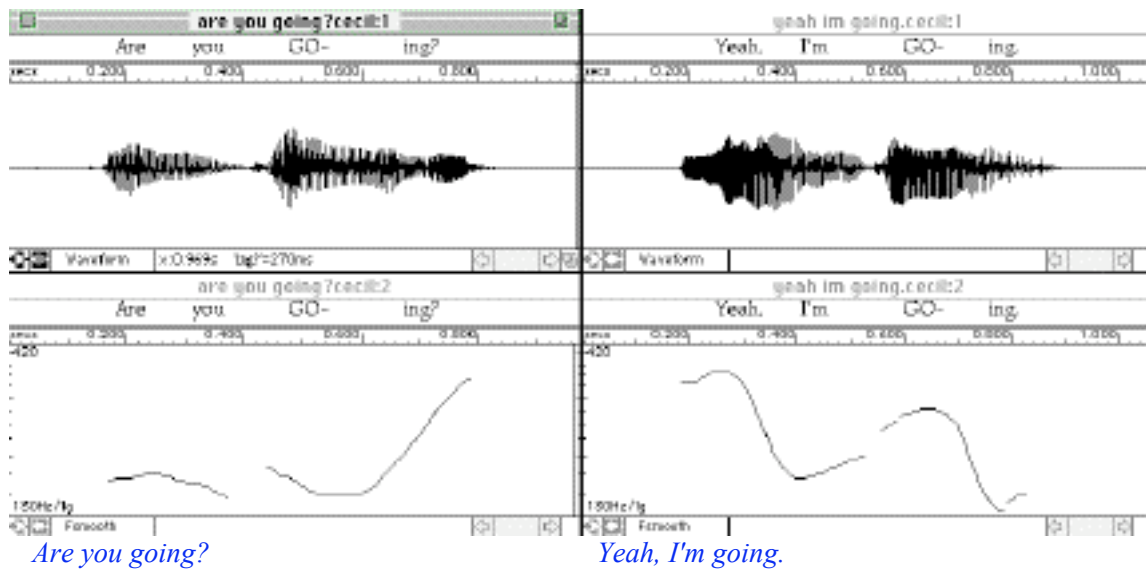*Are you going?*   *Yeah, I'm going.*

Figure 6. Yes-no question and reply

In accordance with Anderson-Hsieh's (1994) recommendation, these examples contain primarily voiced sounds, and one notes that the pitch curves are relatively "continuous." Initially, for practicing basic intonation patterns, constructed examples and exercises should strive to maximize words with a minimum of voiceless sounds, though this will certainly be more problematic with natural discourse.

Attitudinal Functions of Intonation. *Figure 7 presents examples of so-called attitudinal* functions of intonation, here, the surprised or incredulous question *You're going WHERE?* and the exasperated answer *I said I'm going HOME!* Of note here are the wider pitch ranges and the steeper rising and falling curves in both utterances. These are the acoustic features that can be brought to the learner's attention. Software should thus present both a native speaker's pitch curve and that of the learner. In the case of speed or slope of pitch change, for example, the software can also (a) instruct learners to compare the steepness of their falling or rising pitch to that of the native speaker, and/or (b) provide a quantitative measurement of the actual pitch slopes of both the native speaker and the learner.
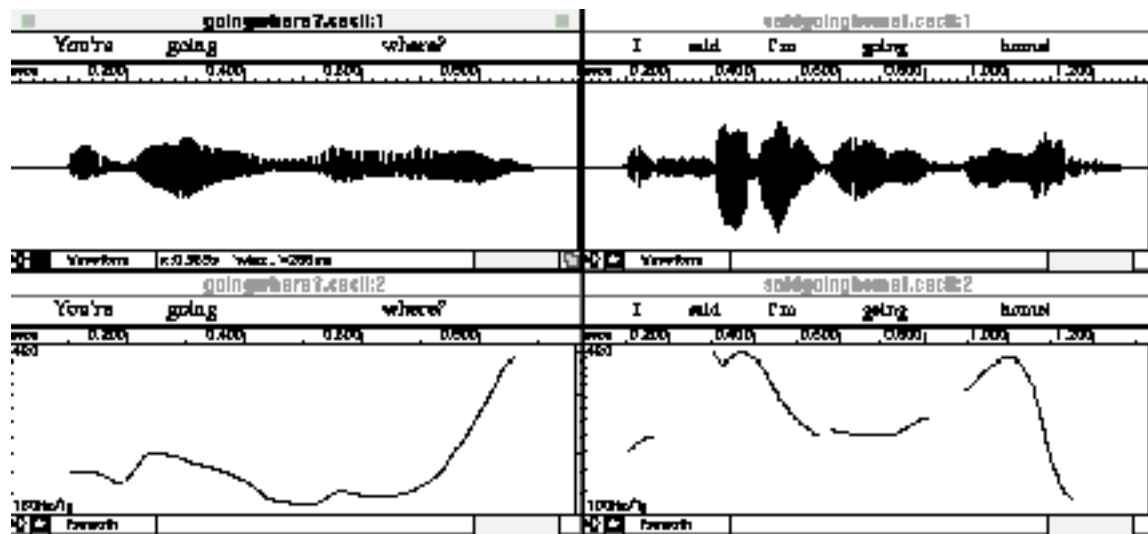
You're going WHERE?   *I said I'm going HOME!*

Figure 7. "Marked" question and reply

Figure 8 shows three renditions of the utterance *Let's go*, first said as a relatively neutral imperative, then as an impatient command, and finally as an attempt to cajole or persuade. The differences in the intonation curves are apparent, and learners can first be asked to practice producing these different patterns. They can then be presented with different situations or scenarios and asked how they would respond. For instance, they could do role-plays of (a) saying *Let's go* to a friend, (b) saying *Let's go* as a sports coach would say to a team, or (c) saying *Let's go* as a polite request to a superior. The next step for language instructors would be to do a cross-linguistic comparison in order to determine whether the target language uses the same means for expressing such nuances of meaning. In the examples below, the impatient command has a higher peak on *go* and a steeper falling pitch curve. The cajoling utterance shows a more sustained, level pitch and does not fall at the end.

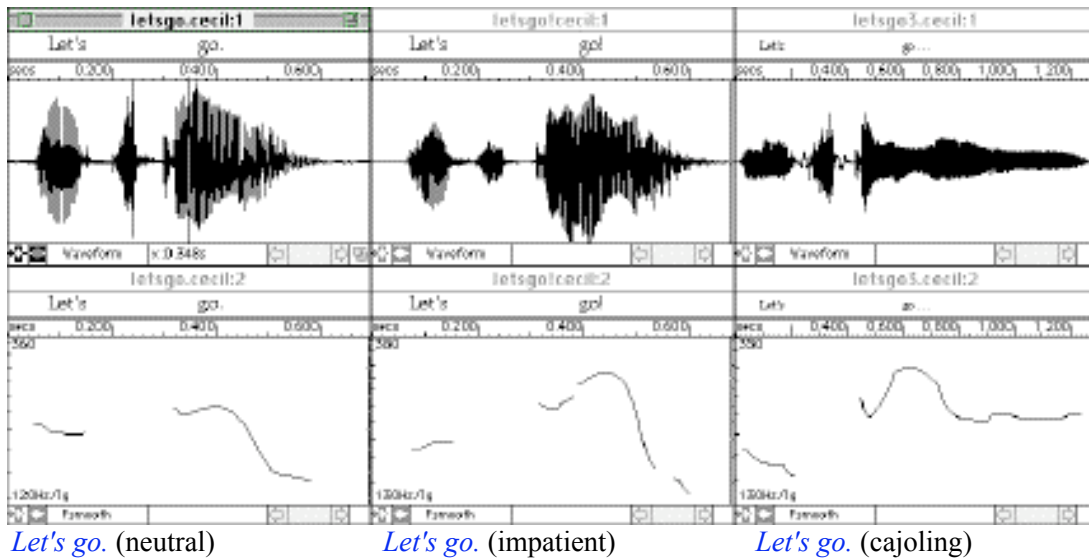*Let's go.* (neutral)   *Let's go.* (impatient)   *Let's go.* (cajoling)

Figure 8. Three renditions of *Let's go*

While "surprise," "exasperation," and "impatience" can be thought of as attitudes, expressing the illocutionary function of "cajoling" might be classified by some as a discourse function rather than an attitudinal function of intonation. Both attitudinal and discourse functions can be considered "pragmatic" as opposed to "grammatical" functions, although the distinctions are far from clear (Couper-Kuhlen & Selting, 1996; Lambrecht, 1994). The important idea is that intonation is used as a signaling device in discourse and must be analyzed and determined on the basis of the surrounding context. Speakers make real-time assessments of what words need to be stressed, of whether to mimic their interlocutor's pitch range, or of what attitude or intention they wish to convey by means of intonation.

Discourse Functions of Intonation. Similar to the examples in Figure 7, the curves in Figure 9 also show steep rises and falls, in this case with the question and answer containing stressed pronouns *Are YOU going? I'M going.* Since pronouns are generally not stressed, these examples illustrate a *discourse* function of intonation, that is, how contrast and emphasis can be signaled by intonation and the importance of the surrounding context of discourse as well as the expectations of the hearer and the speaker. In this case, the speaker may have singled out the addressee in order to find out if s/he is going, and possibly expected that the other person would be surprised to find out the speaker is going.

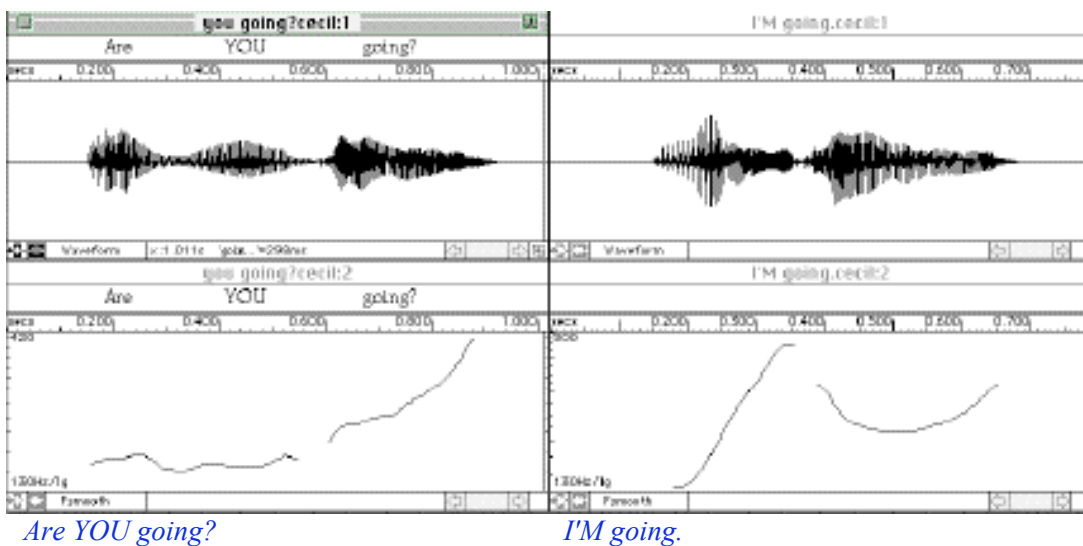*Are YOU going?*   *I'M going.*

Figure 9. Stressed pronouns in a question-answer sequence

The utterances in Figure 9, similar to those in Figure 7, employ a wider pitch range than the more "neutral" utterances in Figure 5. The stressed pronoun *I* in Figure 8 has a fundamental frequency (pitch) of nearly 500 Hz, and the statement does not end on a low pitch as would be more typical of an unemotional or non-contrastive statement.

The final three figures (Figures 10 to 12) show pitch curves for the dialogue below:

> Speaker A: I'm really worried about my GRADE.
>
> Speaker B: You SHOULDn't be. You're NOT gonna fail.
>
> Speaker A: I AM gonna fail.
>
> Speaker B: I think you'll do oKAY.
>
> Speaker A: I HOPE I'll do okay.

The features of note in this dialogue are the varying positions of the sentence stresses, which is representative of conversations. The flaw in many earlier discussions of intonation, both within linguistics proper and in applied linguistics, was that sentence-level intonation was analyzed and taught in terms of where sentence stress typically occurs in a particular language, so stresses were indicated and learners were given isolated sentences to practice, *You're not gonna FAIL* or *I hope I'll do oKAY*. But as is being shown in conversation analysis, natural discourse exhibits anything but "default" intonational patterns, and L2 learners must be made aware early on of how stress, emphasis, contrast, and illocutionary speech are expressed in the L2. As Kelm (1987) found, native speakers of Spanish do not use pitch and intensity to contrast words to the same degree that native English speakers do.
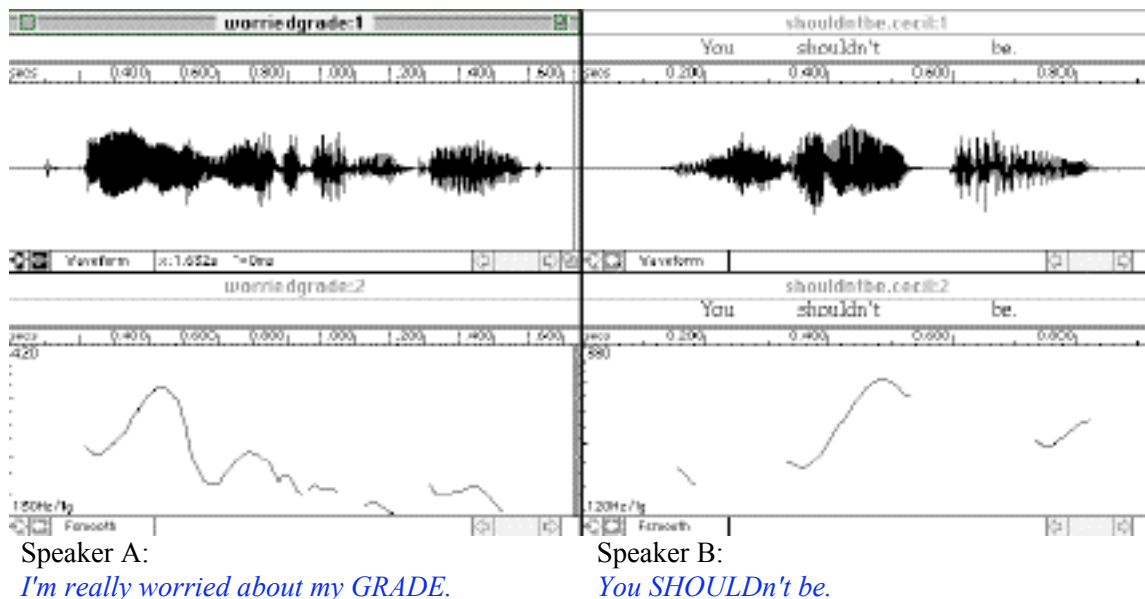

Speaker A:
I'm really worried about my GRADE.

Speaker B:
You SHOULDn't be.

Figure 10.

Speaker A: Speaker B:
You're NOT gonna fail.   *I AM gonna fail.*

Figure 11.

Speaker A: Speaker B:
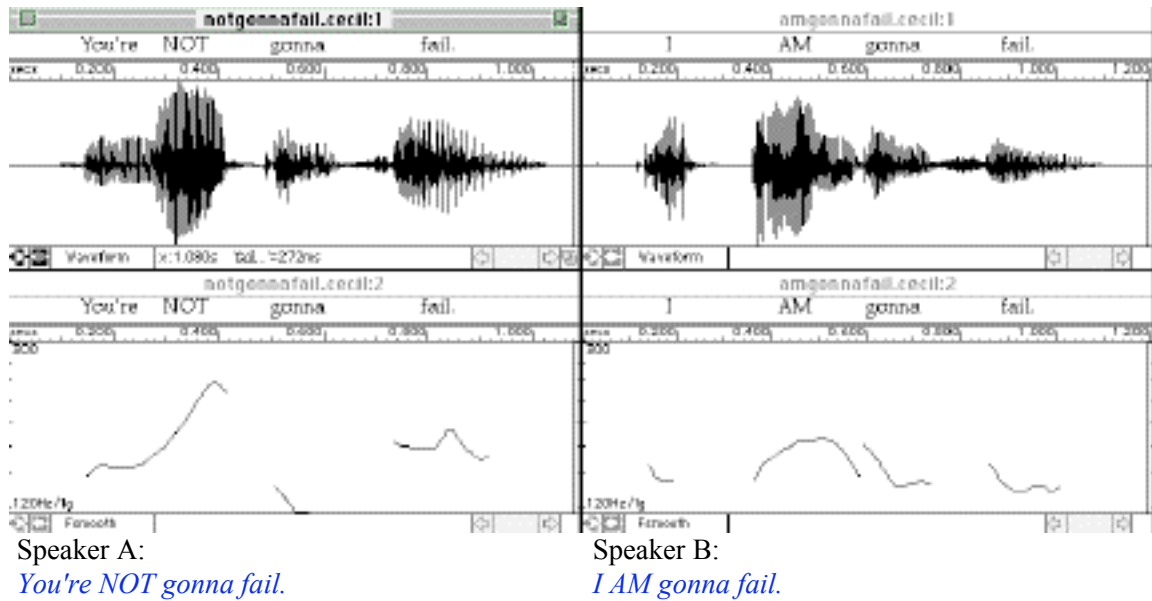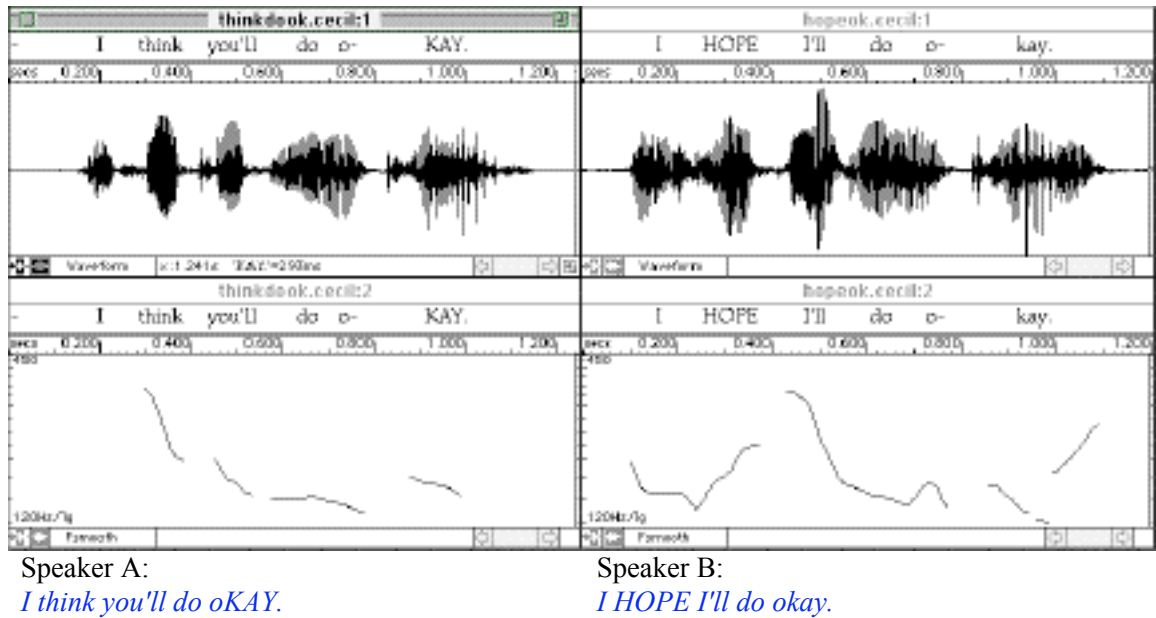*I think you'll do oKAY.*   *I HOPE I'll do okay.*

Figure 12.

The examples above, though not extrapolated from authentic speech, demonstrate that sentence stress does not occur in a relatively fixed position towards the end of the utterance in English. Rather, based on the surrounding context, the speaker makes decisions about what word to stress and what attitude or intention to express. In addition to the discourse-level decisions, learners must also be made aware of sociolinguistic and sociocultural differences in intonation, so that they can practice using appropriate pitch ranges and pitch patterns in the L2. This leads to the next consideration in the development of discourse intonation software, namely, providing learners with ample authentic input so that they can develop their perceptual along with their productive abilities.

2. Provide learners with authentic and extensive speech and cultural input to hone learners' perceptual abilities.

Although the examples presented above were taken from constructed dialogues, authentic speech should be used when possible in later instructional units. For example, short conversations between native speakers can be taken from corpora such as the Corpus of Spoken English (CSE); they can then be pitch tracked and presented to learners. To as great an extent as feasible, conversations with as few voiceless segmentals as possible should be selected. Learners would first do a very close listening and analysis, that is, listen to the conversation or selected parts thereof multiple times as well as view the intonation curves. They would be told which acoustic intonational features to focus on (e.g., male-female differences in pitch patterns or pitch range). Alternatively, they would be presented with utterances and asked to determine the nuance or attitude being expressed. Following the perception training activities, they would be asked to practice these utterances by recording themselves and comparing their pitch tracks to those of the native models. Finally, they would be asked to role-play a conversation or an interaction with a partner on a similar topic. They would record themselves, pitch track their utterances and note the direction of pitch changes, on which syllable the pitch changes occurred, the steepness of the pitch falls and rises, as well as the overall pitch range used. If their natural speech contained a great deal of voiceless sounds, the pitch curves might contain many breaks, unless a system such as the one described by Spaai and Hermes (1993) were being used. If they did not have access to such a system, they would need to have their instructor listen to their conversation and provide feedback and guidance. Although this points to a definite limitation of current speech analysis systems, natural discourse should still be encouraged, and instructors will have to provide the definitive feedback.

3. Design software to facilitate, record, and analyze interactions between speakers.

As suggested in the section above, due to the importance of having natural discourse as the target input and output, software should present authentic speech samples within their cultural contexts and call learners' attention to important intonational features. The second component of the software should contain suggested activities for learners to interact and converse in pairs. Thirdly, the software should provide tools for the learners to record their utterances, pitch track them, and view visual representations of their pitch curves. These features are admittedly quite extensive and require sophisticated implementation, but they address the deficiencies in previous software and also support all of the recent attention in our field on the discourse-level linguistic and cultural aspects of language learning.

4. Build research tools into the software to record students' performance, progress, and attempts at self correction.

In addition to providing audio and visual feedback to learners of their intonation patterns and to having them engage in dialogue with each other, computer software can simultaneously serve as a data collector. A database for each speaker--assuming adequate hard drive space or storage capacity on peripheral devices--can be compiled. Utterances produced by learners can be ordered chronologically and compared for progress over time. Research studies can be designed to address a number of questions, for example, (a) to compare the effects of providing only visual feedback, only audio feedback, only descriptive contrastive feedback, or different combinations of the above three types of feedback; (b) to test the relationship between perception and production; and (c) to assess the long-term effects of intonation training using different types of feedback and using both perception and production exercises.

CONCLUSION

The main goals of this paper have been, first, to review the literature on previous studies of intonation training with computers, and second, to offer reasons for the mixed results of the effectiveness of such training. The limitations of earlier software and technical difficulties as well as certain pedagogical deficiencies were pointed out. Third, it was suggested that the critical next step pedagogically is to

integrate signal analysis capabilities with current linguistic and applied linguistic principles of focusing on discourse-level language use. Four critical components to be implemented in future pronunciation teaching tools were proposed, but this is not a simple task. Although the technology and emerging software are continually and rapidly improving, some aspects are still limited when it comes to analyzing natural speech. In addition, intonation in authentic, spontaneous conversation is also not fully understood. However, using multimedia software for the purposes of perceiving and practicing discourse intonation places one on the cutting edge of an important aspect of second language acquisition, both from a pedagogical as well as from a research perspective. One definite advantage of using technology for intonation training is that concurrent research can be conducted into how prosody in L2 is acquired (through extensive data collection and analysis), as well as into how signal analysis software can facilitate this aspect of phonological acquisition (through testing the effectiveness of visualization and different types of feedback).

ABOUT THE AUTHOR

Dorothy M. Chun, Ph.D. University of California, Berkeley, is Associate Professor of German at the University of California, Santa Barbara. Her main research areas are discourse intonation, second language acquisition, and computer-assisted language learning. She is completing a book on discourse intonation and is co-author of two multimedia readers, CyberBuch and Ciberteca.

E-mail: dmchun@humanitas.ucsb.edu

REFERENCES

Abberton, E., & Fourcin, A. J. (1975). Visual feedback and the acquisition of intonation. In E. H. Lenneberg, & E. Lenneberg (Eds.), *Foundations of Language Development* (pp. 157-165). New York: Academic Press.

Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal, 11*(4), 5-22.

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System, 20*, 51-62.

Bolinger, D. (1986). *Intonation and its parts*. Palo Alto: Stanford University Press.

Bowen, J. D. (1975). *Patterns of English pronunciation*. New York: Newbury House.

Brazil, D. (1975). *Discourse intonation I.* Birmingham: English Language Research Monographs.

Broselow, E., Hurtig, R. R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup, & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp 350-364). Cambridge, MA: Newbury House.

Brown, G., Currie, K. L., & Kenworthy, J. (1980). *Questions of intonation*. London: Helm.

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

Chaudron, C. (1985). Intake: On models and methods for discovering learners' processing of input. *Studies in Second Language Acquisition*, *7*, 1-14.

Chun, D. M. (1988). The neglected role of intonation in communicative competence and proficiency. *Modern Language Journal, 72*, 295-303.

Chun, D. M. (1991). The state of the art in teaching pronunciation. In J. E. Alatis (Ed.), *Georgetown University Round Table on Languages and Linguistics 1991 - Linguistics and language pedagogy: The state of the art* (pp. 179-193). Washington, D.C.: Georgetown University Press.

Couper-Kuhlen, E., & Selting, M. (Eds.). (1996). *Prosody in conversation*. Cambridge: Cambridge University Press.

de Bot, K. (1981). Intonation teaching and pitch control. *ITL Review of Applied Linguistics, 52*, 31-42.

de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech, 26*(4), 331-350.

de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly, 16*, 71-77.

Esling, J. H. (1992). Speech technology systems in applied linguistics instruction. In M. C. Pennington, & V. Stevens (Eds.), *Computers in applied linguistics: An international perspective* (pp. 244-272). Clevedon: Multilingual Matters.

Feldman, D. M. (1973). Measuring auditory discrimination of suprasegmental features in Spanish. *IRAL, 11*, 195-209.

Feldman, D. M. (1977). Instrumental methods for the teaching of English pronunciation. In H. P. Kelz (Ed.), *Phonetische Grundlagen der Ausspracheschulung*, Hamburg: Buske.

Felix, S. W. (1978). *Linguistische Untersuchungen zum natürlichen Zweitsprachenerwerb*. München: Wilhelm Fink.

Fidelman, C., & Keller, E. (1994). *SpeechLab* [Computer Program]. Charlestown, MA: Agora Language Marketplace.

Fischer, L. B. (1986). The use of audio/visual aids in the teaching and learning of French. Pine Brook, NJ: Kay Elemetrics Corporation.

Hengstenberg, P. (1980). *Suprasegmentalia und Aspekte ihrer Vermittlung in sprachlichen Lehr- und Lernprozessen*. Tübingen.

Hurley, D. S. (1992). Issues in teaching pragmatics, prosody, and non-verbal communication. *Applied Linguistics, 13*(3), 259-281.

James, E. (1976). The acquisition of prosodic features of speech using a speech visualizer. *IRAL, 14*(3), 227-243.

James, E. (1977). The acquisition of second language intonation using a visualizer. *Canadian Modern Language Review, 33*(4), 503-506.

James, E. (1979). Intonation through visualization. In H. Hollien, & P. Hollien (Eds.), *Current Issues in the Phonetic Sciences* Vol. IV, Amsterdam Studies in the Theory and History of Linguistic Science (pp. 295-301). Amsterdam: John Benjamins.

Keller, E. (1988). *Signalyze* [Computer Program]. Lausanne: InfoSignal, Inc.

Kelm, O. R. (1987). An acoustic study on the differences of contrastive emphasis between native and non-native Spanish speakers. *Hispania, 70*, 627-633.

Kramsch, C. (Ed.). (1995). *Redefining the boundaries of language study*. Boston: Heinle and Heinle.

Lambrecht, K. (1994). *Information structure and sentence form*. Cambridge: Cambridge University Press.

Lane, H., & Buiten, R. (1969). A self-instructional device for conditioning accurate prosody. In A. Valdman (Ed.), *Trends in language teaching* (pp. 159-174). New York.

Leather, J. (1983). Second-language pronunciation learning and teaching. *Language Teaching, 16*, 198-219.

Leon, P., & Martin, P. (1972). Applied linguistics and the teaching of intonation. *Modern Language Journal, 56*, 139-44.

Lieberman, P. (1986). The acquisition of intonation by infants: Physiology and neural control. In C. Johns-Lewis (Ed.), *Intonation in discourse* (pp. 239-257). London: Croom Helm.

Loveday, L. (1981). Pitch, politeness and sexual role: An exploratory investigation into the pitch correlates of English and Japanese politeness formulae. *Language and Speech, 24*, 71-89.

Luthy, M. J. (1983). Nonnative speakers' perceptions of English "nonlexical" intonation signals. *Language Learning, 33*(1), 19-36.

Molholt, G. (1988). Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly, 22*(1), 91-111.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly, 25*(3), 481-520.

Pennington, M. C. (1989a). *Teaching languages with computers: The state of the art.*. La Jolla: Athelstan.

Pennington, M. C. (1989b). Teaching pronunciation from the top down. *RELC Journal, 20*, 20-38.

Pennington, M. C., & Richards, J. C. (1986). Pronunciation revisited. *TESOL Quarterly, 20*(2), 207-225.

Piper, A. (1986). Conversation and the computer: A study of conversational spin-off generated among learners of English as a foreign language working in groups. *System, 14*(2), 187-98.

Purcell, E., & Suter, R. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning, 30*, 271-287.

Richards, J. C. (1986). Focus on the learner. *University of Hawai'i Working Papers in ESL, 5*, 61-86.

Scovel, T. (1988). *A time to speak: A psycholinguistic inquiry into the critical period for human speech*. Cambridge, MA: Newbury House.

Spaai, G. W. G., & Hermes, D. J. (1993). A visual display for the teaching of intonation. *CALICO Journal, 10*(3), 19-30.

'T Hart, J., & Collier, R. (1975). Integrating different levels of intonation analysis. *Journal of Phonetics, 3*, 235-55.

van Bezooijen, R. (1995). Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech, 38*(3), 253-265.

Van Els, T., & de Bot, K. (1987). The role of intonation in foreign accent. *Modern Language Journal, 71*, 147-155.

Vardanian, R. M. (1964). Teaching English through oscilloscope displays. *Language Learning, 3/4*, 109-118.

Weltens, B., & de Bot, K. (1984). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech, 27*(1), 79-88.

Wennerstrom, A. (1994). Intonational meaning in English discourse: A study of non-native speakers. *Applied Linguistics, 15*(4), 399-420.

Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition 20,* 1-25.

Wichern, P. (1979). Visuele terugmelding als hulpmiddel bij het aanleren van onbeende intonatieverlopen. Nijmegen: Intern Rapport Instituut Fonetiek.

Wichern, P. U. M., & Boves, L. (1980). Visual feedback of Fo curves as an aid in learning intonation-contours. *Proceedings Institute of Phonetics Nijmegen, 4*, 53-63.

Young, R. (1988). Computer-assisted language learning conversations: Negotiating an outcome. *CALICO Journal, 5*(3), 65-83.