GENERALIZATION OF COMPUTER-ASSISTED PROSODY TRAINING: QUANTITATIVE AND QUALITATIVE FINDINGS[1]

Debra M. Hardison
Michigan State University

ABSTRACT

Two experiments investigated the effectiveness of computer-assisted prosody training, its generalization to novel sentences and segmental accuracy, and the relationship between prosodic and lexical information in long-term memory. Experiment 1, using a pretest-posttest design, provided native English-speaking learners of French with 3 weeks of training focused on prosody using a real-time computerized pitch display. Multiple exemplars produced by native speakers (NSs) of French and stored on hard disk provided training feedback. Learners' recorded pre- and posttest productions were presented to NSs for evaluation in two conditions: filtered (unintelligible segmental information) and unfiltered. Ratings using 7-point scales for the prosody and segmental accuracy of unfiltered samples revealed significant improvement in prosody with generalization to segmental production and novel sentences. Comparison of prosody ratings for filtered and unfiltered samples revealed some segmental influence on the pretest ratings of prosody. In Experiment 2, involving a memory recall task using filtered stimuli of reduced intelligibility, learners identified the exact lexical content of an average of 80% of the training sentences based on prosodic cues consistent with exemplar-based learning models. Questionnaire responses indicated a greater awareness of the various aspects of speech and increased confidence in producing another language.

Many factors may influence a native speaker's judgment of a second or foreign language (FL) learner's accent including suprasegmental features such as stress, rhythm, and intonation (Munro, 1995). These features began to draw the greatest attention from teachers and materials developers with the advent of the discourse-level focus of the communicative approach to language teaching and remained the principal focus throughout the 1980s (e.g., Morley, 1991; Pennington & Richards, 1986). However, in recent years, the field of pronunciation teaching appears to have adopted a more balanced viewpoint with regard to the importance of both the segmental and suprasegmental aspects of language (e.g., Celce-Murcia, Brinton, & Goodwin, 1996; Derwing, Munro, & Wiebe, 1998). Perhaps, then, an ideal training tool is one that can produce a significant improvement in both levels of the spoken language with evidence of generalization to novel stimuli.

The current study investigated such a tool in the form of computer-assisted training with visual feedback. The study was divided into Experiments 1 and 2. Experiment 1 focused on both quantitative and qualitative aspects of the acquisition of French prosody by native speakers (NS) of American English (AE) using computer-assisted training that permits visual display of pitch contours in real time. This experiment was designed to investigate (a) the effects of such training when exemplars from native speakers (NSs) of French were used as feedback rather than models to imitate, (b) the extent to which the training would generalize to improvement in segmental production and to both the prosodic and segmental features of novel sentences, and (c) the potential influence (positive or negative) of segmental quality on the NS ratings of prosody by comparing learners' recorded productions with filtered versions that render the segmental information unintelligible while preserving the prosodic information. In addition to the quantitative findings, observation notes I made during training and the responses to participants' anonymous questionnaires following their training program provided qualitative information

on the learner-technology interaction and the learners' perceived usefulness of computer-assisted speech training.

Experiment 2 was conducted to explore how prosodic and lexical information is stored in memory, specifically whether these components of utterances are stored together in memory traces. A memory recall task was used to investigate (a) whether probing memory with only the prosodic information from familiar training sentences results in recall of the associated lexical information, and (b) what characterizes the prosodic/lexical associations that are easiest for learners to recall. In other words, what constitutes salient input in prosody training?

In this paper, prosody refers collectively to variations in pitch, tempo, and rhythm. Intonation represents the patterns of pitch or melody. This association of intonation with a term such as *melody* that one immediately associates with music will surface again later with a discussion of some of the learners' questionnaire responses. Rhythm may be considered the perceived regularity of prominent units in speech. The rhythmic structure of speech reflects a hierarchical organization of the temporal sequence of speech sounds into syllables and higher level units of prosodic and syntactic structure. Part of learning a spoken language is the acquisition of its systematic rhythmic organization. The current study makes reference to computerized visual displays of pitch contours in prosody training, but because intonation, tempo, and rhythm are inextricably linked, the study is one of the acquisition of prosody, not of intonation per se.

Several reports have advocated the use of speech technology programs providing visual displays of intonation for language learners (Anderson-Hsieh, 1992, 1994; de Bot, 1983; de Bot & Mailfert, 1982; Leather, 1990; Molholt, 1988; Pennington & Esling, 1996; Weltens & de Bot, 1984) though few studies have utilized research methodology to evaluate the effectiveness of this approach. De Bot (1983) investigated visual pitch feedback with Dutch university students learning English. In a pretest-posttest design, auditory-visual (AV) versus auditory-only (A-only) feedback, and the amount of practice (one session of 45 minutes vs. 2 sessions) were investigated. After a sentence was presented auditorily, the pitch contour was displayed and learners imitated the sentence. Their pitch contours then appeared on the display below the target. They were allowed to repeat the process a nonspecified number of times. Ratings of these productions were done by three teachers of English using a 5-point scale. Results indicated that AV feedback was significantly better than A-only, but the amount of practice time was not significant.

The benefits of AV versus A-only training (i.e., two channels of input vs. one) have also been demonstrated in improving the perceptual accuracy of nonnative sounds for learners of English as a second language (ESL) with generalization to novel stimuli and transfer to production improvement (Hardison, 2003). The visual stimuli in this instance were full-sized images of talkers' faces on videotape.

Another study (Weltens & de Bot, 1984) with Dutch learners of English found that the duration of feedback delay, that is, the period of time between the speech signal and pitch contour display (40 ms., 250 ms., or at the end of the signal), did not significantly affect improvement, while the speech material (i.e., the number of unvoiced segments) and the voice characteristics of a speaker were factors in the quality of feedback. These studies with Dutch learners of English involved brief training periods and focused on the learners' imitation of NS sentence models. Questions have arisen regarding the effects of brief training sessions, limitations of earlier speech technology systems, and the learners' ability to interpret displays -- all factors influencing a decision to incorporate such systems into pronunciation teaching (Chun, 1998).

In addition to the above issues involved in prosody training, per se, theoretical and pedagogical questions also arise, specifically regarding how the prosodic and lexical content of speech is stored in memory. Results of an experiment dealing with recognition memory of English sentences by advanced learners (native speakers of Cantonese) showed a high level of lexical memory in recognizing sentences they had

heard before ("old" sentences) but the same learners did not recognize when an old sentence's prosody had changed, even when the change was lexically relevant. This suggests that unanalyzed prosodic patterns had been stored in memory and that learners had failed to generalize the information to new examples (Pennington & Ellis, 2000).

In the current study, Experiment 2 used a memory recall task to explore the relationship between the prosodic information and lexical content of the traces of the training stimuli from Experiment 1 stored in long-term memory by using prosody as the lexical access cue through the use of filtered speech. In multiple-trace theory, all attended perceptual details of an event are stored in a trace, so that prosodic patterns as well as lexical information may be stored together in a trace, and a composite of such traces forms the basis of episodic memory (Hardison, 2000[2]; Hintzman, 1986). Recent studies have demonstrated that episodic memory exists both in musically untrained listeners and in trained listeners for "musical prosody," or the acoustic features (termed *performance expression*) that characterize particular music performances (Palmer, Jungers, & Jusczyk, 2001).

Traces stored in long-term memory are said to return an *echo* to primary memory in response to a retrieval cue or probe. The content of the echo is the summed contributions of all the traces reacting in concert according to similarity between their properties and the probe. Differing levels of attention paid to the various aspects of the speech signal result in a hierarchical structure of information in the trace such that the strongest element is the one that received the most attention. This hierarchy, in turn, may be influenced by factors such as the perceived relevance of the content to the performance of a task or the uniqueness of the content. The echo can also enhance a probe's representation by filling in missing details. Echoes for higher-frequency words do not reflect the detailed characteristics of any particular trace, but they do produce a more generic echo as a result of multiple-trace activation; in contrast, the echoes for lower-frequency words more closely resemble the stimulus, as they are the result of activation of specific old traces (Goldinger, 1997).

The acquisition of French by speakers of AE presents an interesting contrast in terms of prosodic features that often pose problems for learners. The traditional syllable-timed/stress-timed distinction used to characterize the difference between French and English has not been supported by experimental findings (e.g., Bertinetto, 1989). Wenk and Wioland (1982) and Fletcher (1991) suggest that the rhythm of French is unlike that of English in that it is *trailer-timed*; that is, the salient element is a right-boundary phrase-final prominence at the end of a word or sense group in contrast to English, which is *leader-timed* (left-boundary foot-initial prominence). In French, the right-boundary stress is generally associated with a perceptible fluctuation in pitch as well as some degree of lengthening.

In his description of the prosodic units of General French,[3] Di Cristo (1998) notes that the language has a single rhythmic stress (primary stress) assigned to the final full syllable of the last lexical item, usually a content word, of a stress group (composed of the "stressable" word and adjacent clitics governed by it). There is also emphatic stress (e.g., focal stress), and non-final (secondary) optional stress described by some as generally assigned to the first full syllable of a phrase-initial content word (e.g., Fónagy, 1979; Hirst & Di Cristo, 1984; Vaissière 1974). This is subject to rhythmic constraints within a stress group (e.g., the number of syllables in the group generally does not exceed three according to Fletcher, 1991 and to Wenk & Wioland, 1982) and across groups (e.g., there is a tendency to avoid adjacent rhythmic stress in the same intonation unit). The stress group has also been referred to as a tonal unit (Hirst and Di Cristo, 1984) because pitch prominence is the main cue in signaling primary and secondary rhythmic stress, but the syllable bearing final or primary stress is also lengthened.

The following basic description of French intonation patterns is guided by the characteristics reported in most studies (for a recent, thorough discussion, see Di Cristo, 1998). The examples below were taken from the stimulus set of the current study. The transcriptions were based on visual displays of the pitch contours of native speakers and were generated by the Real-Time Pitch Program of the Kay Elemetrics

Computerized Speech Lab (CSL) used in this study. The transcription system follows **INTSINT** (**IN**ternational **T**ranscription **S**ystem for **INT**onation) developed by Hirst and Di Cristo (1998) for use specifically with French and English, where ↑ (Higher) and ↓ (Lower) represent pitch points relatively higher or lower than the immediately preceding pitch point; > (Downstep) and < (Upstep) represent a slight downstepping (lowering) or upstepping (raising) of pitch relative to the preceding point (and are used in this paper to indicate smaller pitch changes than those transcribed as Higher or Lower); and (Top) and (Bottom) represent more extreme high and low values with respect to the speaker's vocal range. These symbols are summarized below in Table 1.

Table 1. Pitch-Transcription Symbols

Higher	Lower	Same	Downstep	Upstep	Top	Bottom
↑	↓	→	>	<		

In the examples below, the left square bracket represents the beginning tone level. For a simple declarative sentence, there is a rising pitch movement (from low to high) at the end of each stress group except the last one which is produced with a falling pitch movement as in Example 1.

Example 1. Simple declarative sentence

Je	suis	allée	à l'agence	de	voyage.	"I went to the travel agency."
[	→	↑	> > ↑	>	<	]

Di Cristo (1998) groups interrogative forms into two categories: *total questions* (including syntactically unmarked ones signaled by intonation that are frequently used in contemporary French, those with inversion of a pronoun subject and verb, and those introduced by the expression *Est-ce que*…? "Is it that…?") and *partial* (WH-) *questions*. Total questions seeking information (vs. confirmation of what is already known) are characterized by a rising pitch associated with the last stressed syllable of the utterance as shown in Example 2.

Example 2. Total question

Pardon,	*vous*	*avez*	*l'heure?*	"Excuse/pardon me, do you have the time?"
[	<	> >	]	

Partial (WH-) questions are marked with an interrogative morpheme (e.g., *où* "where"). As shown in Example 3, neutral (vs. echo) partial questions generally have an initial pitch prominence on the stressed syllable of the question word followed by a regular drop in pitch until the final syllable produced with a low pitch in the speaker's vocal range.

Example 3. Partial (WH-) question

Où	*se*	*trouve*	*la*	*gare?*	"Where is the train station?"
[	>	>	>	]	

French speakers can also use focal accents for intensification in which case a particular syllable as in Example 4 (***IN****croyable* "unbelievable") or word as in Example 6 (*assez* "enough") can be highlighted with an extra pitch prominence.

Example 4. Intensification by syllable

Je	*viens*	*de*	*lire*	*un*	*livre*	*incroyable.*	"I've just read an unbelievable book."
[	↑	>	↑	>	↑	> >	]

Recall that the symbol ↑ means the pitch is higher relative to the previous pitch point. In this example, the pitch points for *viens* (part of the expression *venir de*) "have just" and *lire* "read" are at roughly the same

frequency. There is a rise on *livre* "book" (though not as high as the first two) followed by an additional rise for emphasis on the first syllable of ***in****croyable* "unbelievable."

Example 5. Intensification by word

J'en	*ai*	*assez*	*(de)*	*travailler.*	"I've had enough of working."
[	<	→		> > >	]

In this example, *de* appears in parentheses as it was imperceptible in spontaneous speech.

To imply a contrast, a focused item can be characterized by a rising-falling pitch pattern as in Example 6 where the postnominal color adjective *rouge* "red" is in focus.

Example 6. Focus on "red"

Elle	*a*	*choisi*	*la*	*jupe*	*rouge.*		"She chose the red skirt."
[	→	↑	>	<	↑	↓	]

The lexical item in focus is subject to the speaker's interpretation. In the above example, without additional context or instruction, the speaker chose to convey the idea that it was a red skirt and not any other color. Contrast this with the same sentence spoken by a different NS as shown in Example 7 where the focused item is the noun *jupe* "skirt" rather than *rouge* "red" which conveys the idea that it was a red skirt and not any other article of clothing.

Example 7. Focus on "skirt"

Elle	*a*	*choisi*	*la*	*jupe*	*rouge.*	"She chose the red skirt."
[	→	↑	>	↑	↓	]

I have outlined above the primary declarative, interrogative, and focal patterns of rhythmic stress and their pitch characteristics. Although some researchers have noted relatively high onsets and steeper overall slopes as features of pitch in the imperative modality, there is no overall specific imperative pattern (Di Cristo, 1998). Only a few imperatives were included in the training set of the current study.

Experiment 1 was guided by the following objectives and hypotheses:

1. Evaluate the use in FL prosody training of computerized feedback in the form of visual pitch contour displays in real time through a pretest-posttest experimental design and 3 weeks of training. Findings from a previous study, on a more limited basis, had demonstrated that AV input was significantly better than A-only (de Bot, 1983). I hypothesized that visual pitch displays would serve to enhance input and draw learners' attention to the prosodic organization of the language. The decision to use 3 weeks for training in the present experiment was based on previous successful segmental-level training studies for ESL learners (e.g., Hardison, 2003; Lively, Logan, & Pisoni, 1993) and the hypothesis that more exemplars and practice, especially in the absence of explicit instruction, would increase the potential for generalization. Periods longer than 3 weeks were problematic for participants' schedules.
2. Limit the role of NS pitch contour displays to the role of feedback for comparison with learners' attempts. This is in contrast to the use of NS data exclusively as models for imitation (de Bot, 1983). While a direct comparison of these two approaches was not made in the present study, I hypothesized that having learners produce a target sentence before the NS version was presented would provide them with more confidence in their ability to produce FL speech and greater potential for generalization to unscripted speech after the experiment was over.
3. Determine whether improvement following training with feedback on prosody only would generalize (a) to higher ratings of segmental accuracy in the posttest and (b) to the prosodic and segmental features of novel sentences in a generalization task. Two limitations of training in second- or foreign-

language speech have usually been generalization and retention. This was also the rationale behind the longer period of training compared to previous studies. Although retention could not be evaluated for this experiment, a test of generalization involving novel sentences was conducted following the posttest.

4. Compare NS ratings of learners' prosody in filtered versions (in which segmental information is rendered unintelligible while prosody is preserved) and unfiltered versions to investigate the rating procedure. I hypothesized that NSs might not be able to avoid being influenced by segmental quality even when asked to focus their attention only on prosody.
5. Compile experimenter observations and learner comments during training on the use of technology in FL speech learning. I was particularly interested in how interpretable the displays were for the learners (Chun, 1998), how well such displays functioned to draw learners' attention to the prosodic features of French, and how this changed over time.
6. Gather information from anonymous questionnaires completed by the learners on their perceptions of the effectiveness of such training. As researchers and teachers evaluate computer-assisted learning, it is important to consider the interaction between the learner and the technology (Pennington & Esling, 1996).

It is also important to note that this study's goal was not to compare different *types* of training, an issue to which I will return in the General Discussion. The objectives, as stated above, centered around the generalizability of prosody-focused training, the possible influence of segmental quality on ratings of prosody, the preservation of salient pitch contours in memory, the sequence of elements that captured learners' attention, and learners' views on computer-assisted training.

EXPERIMENT 1

Experimental Design

Experiment 1 had quantitative and qualitative components. A pretest-posttest design was used to measure the effects of 3 weeks of training (13 sessions of about 40 minutes each) in French prosody using computerized visual displays of pitch contours as feedback. Following the posttest, participants were asked to produce a set of new sentences to test generalization of training. They were unaware that their productions would also be rated for segmental accuracy as well as for prosody. They were also not informed about Experiment 2 so as not to promote rehearsal or any exceptional attention to the training stimuli. In addition, I kept a record of comments learners made throughout the training period on the aspects of French that had captured their attention. Following the posttest, participants were asked to complete anonymous questionnaires on the value of the training.

Method

Participants. A total of 16 native speakers of General American English volunteered to participate in this study. Data collection required three training periods. All participants were female undergraduate students at a large American university enrolled in the first semester of the second year of French study. Level placement was determined by testing. None were language majors and none had studied or lived abroad at that point. Through preliminary interviews with me (a former college French teacher), I concluded that the participants were representative of the second year college level of proficiency (i.e., high beginner-low intermediate). Classroom instruction was their sole source of input, and instructors had native or near-native proficiency in French. The coordinator of first and second year French study indicated that little class time was generally available for pronunciation practice. All participants were motivated to improve their production of French. Some expressed an interest in studying abroad while others wanted to travel. There was also a control group of 10 students with similar backgrounds. Although equally interested in participating, they could work only the pretest and posttest sessions into their schedules and

could not attend daily training.[4] Those in the experimental group were paid $15.00 for their participation, and all participants were offered the opportunity to obtain feedback on testing performance when data analysis was completed.

Materials. Prosody Training. Selection of sentences for testing and training followed these guidelines: a) familiar vocabulary (determined by examining instructional materials for the first and second years of study and consultation with the supervisor of these courses), b) functional value to college students, c) several exemplars within semantic domains such as food and wine, student life, travel, etc., d) sustained phonation to provide the best possible continuous display of pitch contour, e) relatively short to facilitate production from short-term memory rather than reading, f) structural variety (e.g., declaratives and interrogatives of different types as outlined earlier), g) a range of sounds including those that are often difficult for AE speakers such as nasal vowels and /ʀ/, as well as liaison[5] contexts (see Appendix for a representative sample of testing and training sentences).

There were 20 pretest/posttest sentences and 20 novel sentences for the test of generalization. For feedback purposes, the training sentences were recorded by NSs of General French direct to computer hard disk using Kay Elemetrics Computerized Speech Lab (CSL) with a Shure unidirectional microphone and JBL studio speaker. Each of three female NSs recorded a different set of 30 sentences for a total of 90 used in training. Each NS was instructed to look at a sentence printed on a card, look up, and then produce it at a conversational rate of speech into the microphone. Sentences were played back to check intelligibility and naturalness of expression. Some of the testing sentences were also recorded by NSs and stored on hard disk for later comparison purposes.

Questionnaire. Participants were asked to complete an anonymous questionnaire consisting of the following 5 open-ended questions: 1) What were the most difficult elements of French pronunciation for you **before** this program? 2) What do you feel you focused your attention on **during** the 3 weeks of practice? 3) What elements of the speech of native French speakers did you notice **after** this program that you had not noticed before? 4) What have you noticed about your own pronunciation in French as a result of this program? 5) What do you feel you've accomplished in terms of your pronunciation in French?

Procedure. Prosody Training. Participants were tested and trained individually. For the pretest, posttest, and test of generalization, they were shown sentences printed on a card. They were allowed to practice the sentence aloud before recording (primarily to deal with any anxiety especially during the pretest). After looking at each sentence, they were instructed to look up and produce the sentence at a conversational rate into the microphone. The sentences were stored as separate files on hard disk. During testing, there was no visual or auditory feedback. Participants could not see the monitor screen. Following the posttest, participants were given a test of generalization involving 20 novel sentences with no feedback.

For the training sessions, participants were shown each sentence printed on a card. As with testing, they were then instructed to look away from the card, pause and produce the sentence into the microphone to avoid 'read' speech. Each training session focused on one set of 30 sentences. The pitch contour of a learner's utterance was displayed in real time in View Screen B on the bottom of the monitor screen and played out through the speaker. The utterance was then replayed and redrawn. In View Screen A on the top of the screen, the NS's version was displayed providing auditory and visual feedback. It was then overlaid on the learner's in a contrasting color in View Screen B. The screens were then cleared and the sentence was practiced again.

Questionnaire. At the conclusion of the study, participants were given the questionnaire to complete anonymously and return to me through campus mail or to the department office.

Results and Discussion

Prosody Training. Participants' recorded productions were evaluated on a 7-point scale by a total of three NSs of French. Rating sessions were done individually. Each learner's pretest, posttest, and generalization sentences were randomized. Raters were not told which productions had preceded or followed training. The productions were blocked by learner for presentation to raters in filtered and then unfiltered versions. These two versions were used in order to determine whether raters' judgments of prosody were being influenced by segmental quality (either in a positive or negative direction). In addition, in the filtered version, some NS samples were included to ensure that raters were able to rate native-like prosody appropriately as none had had any experience listening to filtered speech. These samples received ratings of 7 ("definitely native-like").

The digital filter (low pass, Blackman, 100th order) was created using the CSL to render the segmental content of speech unintelligible, leaving only the prosodic information. The cutoff was set at 300 Hz appropriate for adult female voices. To determine this value, a spectrogram with overlay pitch extraction was generated for each speech sample to establish the point above which there were few, if any, F_0 components. Each filtered sample was saved as a file on hard disk and then evaluated by NSs of French to ensure that words could not be identified.[6]

For each version, raters were given the sentence on a response sheet along with a 7-point scale ranging from "1" (definitely not native) to "7" (definitely native-like) for the prosody rating. In the unfiltered version, there were two scales for each sentence: one for a rating of prosody and one for segmental accuracy. Providing the sentence was necessary for the rating of filtered speech so this was also done for the unfiltered samples. All speech samples were played directly from the computer through the studio speaker for ratings. Ratings for the filtered ones were obtained first. To familiarize raters with filtered speech, they were presented with filtered versions of several sentences I had recorded earlier that were not part of the study.

Mean ratings were calculated separately for filtered and unfiltered prosody and for segmental accuracy (present only in unfiltered speech). Interrater reliability was assessed using a method suggested by Hatch and Lazaraton (1991) involving the calculation of mean interrater correlations with correction for use with ordinal data by Fisher Z transformation. A Pearson (*r*) value of .83 was considered satisfactory reliability for three raters especially in view of the absence of prespecified criteria for the evaluation of prosody. The control group did not show any significant improvement. The mean prosody rating for that group declined from 4.28 in the pretest to 4.18 in the posttest; the mean rating for segmental accuracy was 4.12 in the pretest and 4.16 in the posttest. This is not surprising given that little attention can be paid in their classes to features of the spoken language and the students were not engaged in any pronunciation practice outside of class. As such, statistical analyses were conducted only with data from the experimental group.[7]

Results for the experimental group are shown in Figure 1. The first two sets of bars represent the mean ratings for prosody in the pretest, posttest, and test of generalization. The first of these sets is from filtered speech and the second from unfiltered. The third set of bars provides data for segmental accuracy in the three tests (recall that segmental information is present only in unfiltered speech). A two-factor ANOVA involving Time (pretest, posttest) and Feature (prosody, segments) revealed a significant effect of time [$F(1,30) = 8.76$, $p < .01$] indicating an improvement in both prosody and segmental accuracy as a result of prosody training. Recall that learners had not been told that segmental accuracy would be assessed. The Time x Feature interaction was not significant [$F(1,30) = .719$, *n.s.*].

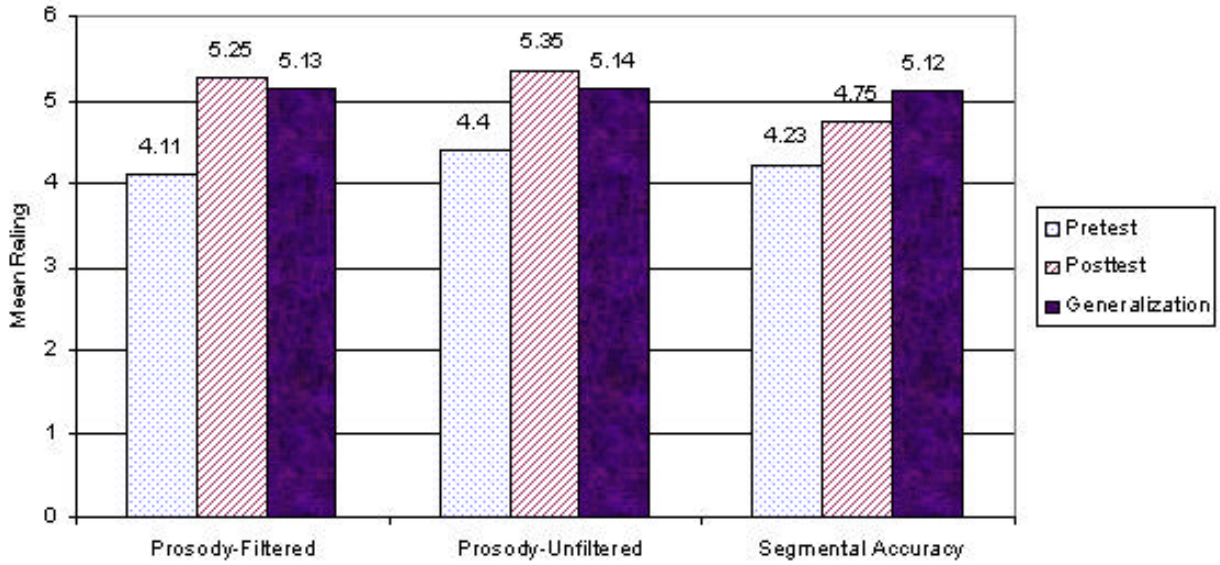

Figure 1. Effects of prosody training: Mean accuracy ratings for prosody (filtered and unfiltered versions) and segmental accuracy in pretest, posttest, and test of generalization.

To determine whether improvement in both prosody and segmental accuracy had generalized to novel sentences, ratings of the unfiltered productions in the test of generalization were compared to those of the pretest. A two-factor ANOVA involving Test (pretest, test of generalization) and Feature (prosody, segments) revealed a significant effect of test [$F(1,30) = 11.43$, $p = .001$] but no significant interaction indicating comparable generalization of improvement in both prosody and segmental accuracy to novel sentences.

As shown in Figure 1, mean ratings in the pretest were significantly lower for filtered prosody (4.11) compared to unfiltered (4.40) [$df = 15$, $t = 4.14$, $p < .001$]; in unfiltered speech, segmental information may have had a positive influence on NS assessment. Mean prosody ratings in the posttest for filtered and unfiltered speech were not significantly different (means of 5.25 and 5.35 respectively) [$t = 2.00$, *n.s.*]. Following training, prosody had improved significantly, and ratings of its native-like quality were not influenced by segmental accuracy as in the pretest.

Note that the three ratings for the generalization sentences shown in Figure 1 were quite similar (5.13 filtered prosody, 5.14 unfiltered prosody, 5.12 segmental accuracy). Segmental accuracy was actually rated higher for these sentences than for the posttest ones. I would discount a mere practice effect here as the prosody ratings (the focus of the training) were not similarly influenced. The mean prosody ratings for filtered and unfiltered generalization sentences, while good and significantly higher than those for the pretest, were a bit below those for the posttest. The high segmental accuracy could be the influence of the content of the sentences. Although the testing sentences were made as equivalent as possible in terms of length, familiarity of vocabulary, type, and number of syntactic structures, no attempt was made to count up the number of /ʀ/ sounds, nasal vowels, and so forth; as described below, learners did begin to notice these segmental features in the later stages of training.

To demonstrate the type of visual feedback provided by this training program and the improvement in pitch contour, Screen A on the top of Figure 2 shows a sample pretest version of the sentence *Elles adorent la couleur rouge* (They love the color red) with the learner's posttest utterance in Screen B at the bottom. One of the raters commented that the pretest was a good example of "English marching rhythm" with no pitch prominence evident; however, in the posttest, the appropriate contour appears with the highest pitch level on the second syllable of the verb *adorent.* In Figure 3, the NS version of this sentence

appears alone in Screen A on the top, and on the bottom in Screen B it is overlaid on this learner's posttest production. On the computer screen, these are shown in different colors.

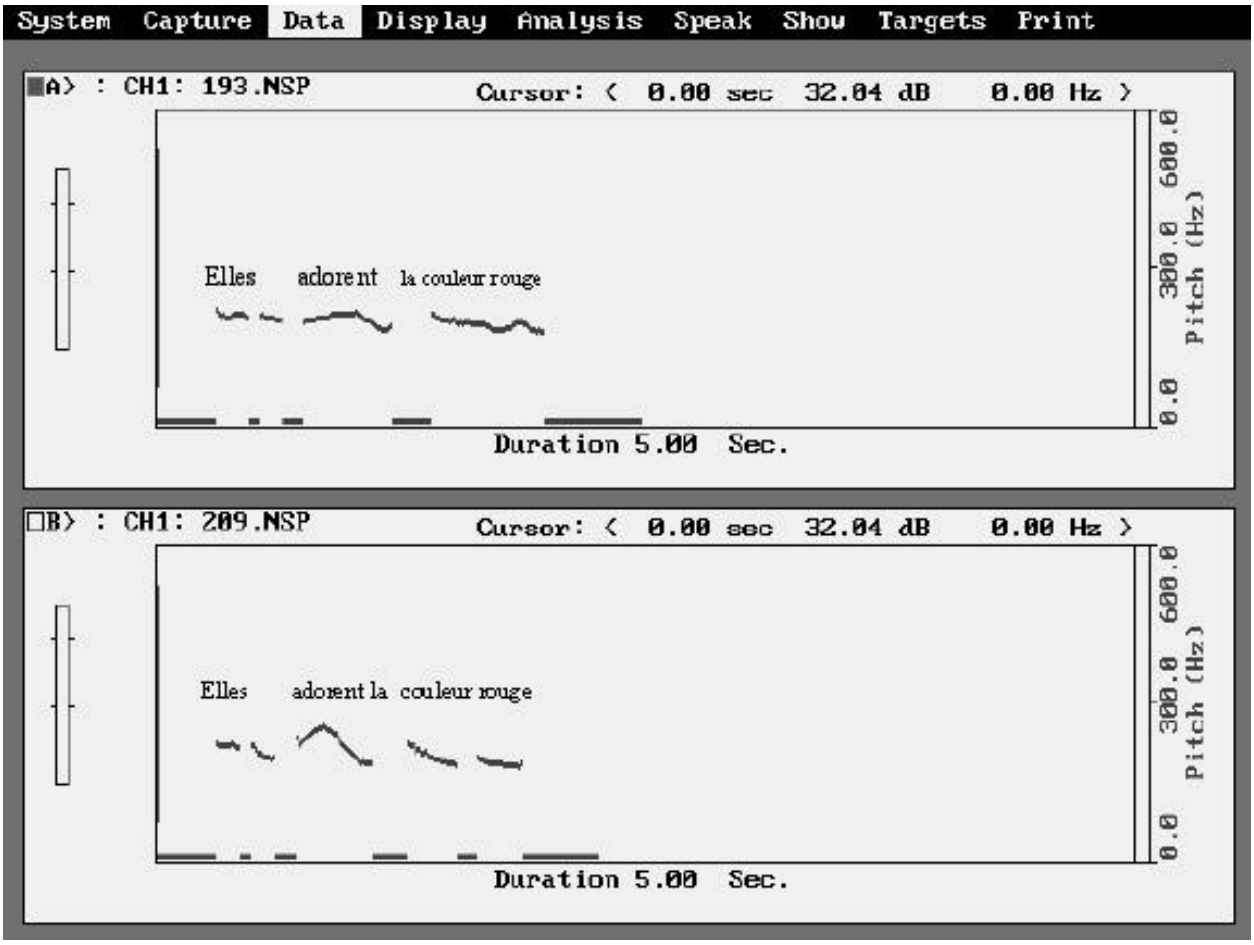

Figure 2. Comparison of learner's pitch contour in pretest (top, Screen A) and posttest (bottom, Screen B). *Elles adorent la couleur rouge* (They love the color red). The text follows the corresponding areas of the pitch contour.

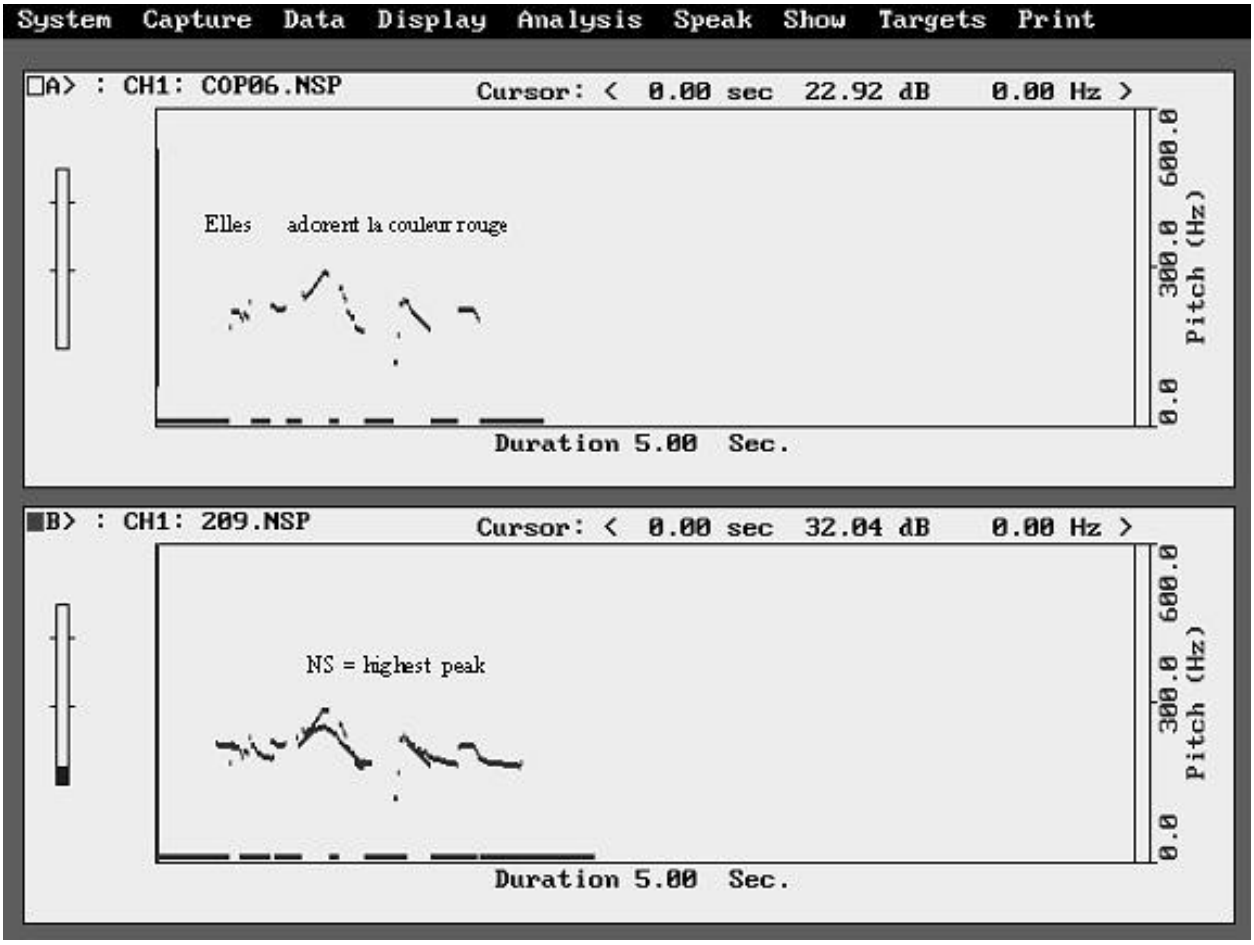

Figure 3. NS pitch contour (top, Screen A). *Elles adorent la couleur rouge* (They love the color red). The text follows the corresponding areas of the contour. Learners' posttest production (Screen B) shows NS pitch overlay (the higher peak on the second syllable of the verb *adorent*).

Training Observations

Throughout the training program, I wrote down the comments that participants made at various points with particular emphasis on the features of the language that drew their attention. The information presented in this section is limited to what they actually verbalized and may not represent all the elements of speech to which they attended.

Initially, all learners focused their attention on the "flatness" of their pitch contours which sometimes appeared as just "dots on the screen" (also subject to the sustained phonation quality of the sentence) in contrast to what many referred to as the "flowing" speech or "peaks and valleys" evident in the contours produced by the NSs and displayed as feedback during training. Even though learners were not encouraged to focus on rate of speech, the native-nonnative difference did capture their attention and it appears on the screen as the duration of pitch tracking. At first, no one commented on liaison, vowels, or other segmental features.

The visually salient convex shape of some final intonation patterns in French drew the attention of most learners as did the steep slope of rising intonation to signal questions with declarative structure and no question word (e.g., *Pardon, vous avez l'heure?* "Excuse me, do you have the time?"). This training sentence example is shown in Figure 4 where Screen A is the NS version and Screen B is a learner's production with NS overlay. These prosodic contours are so close that only the sections showing the rise

on *l'heure* (time) are distinguishable because of the slight difference in the duration of their productions although the slopes are comparable.

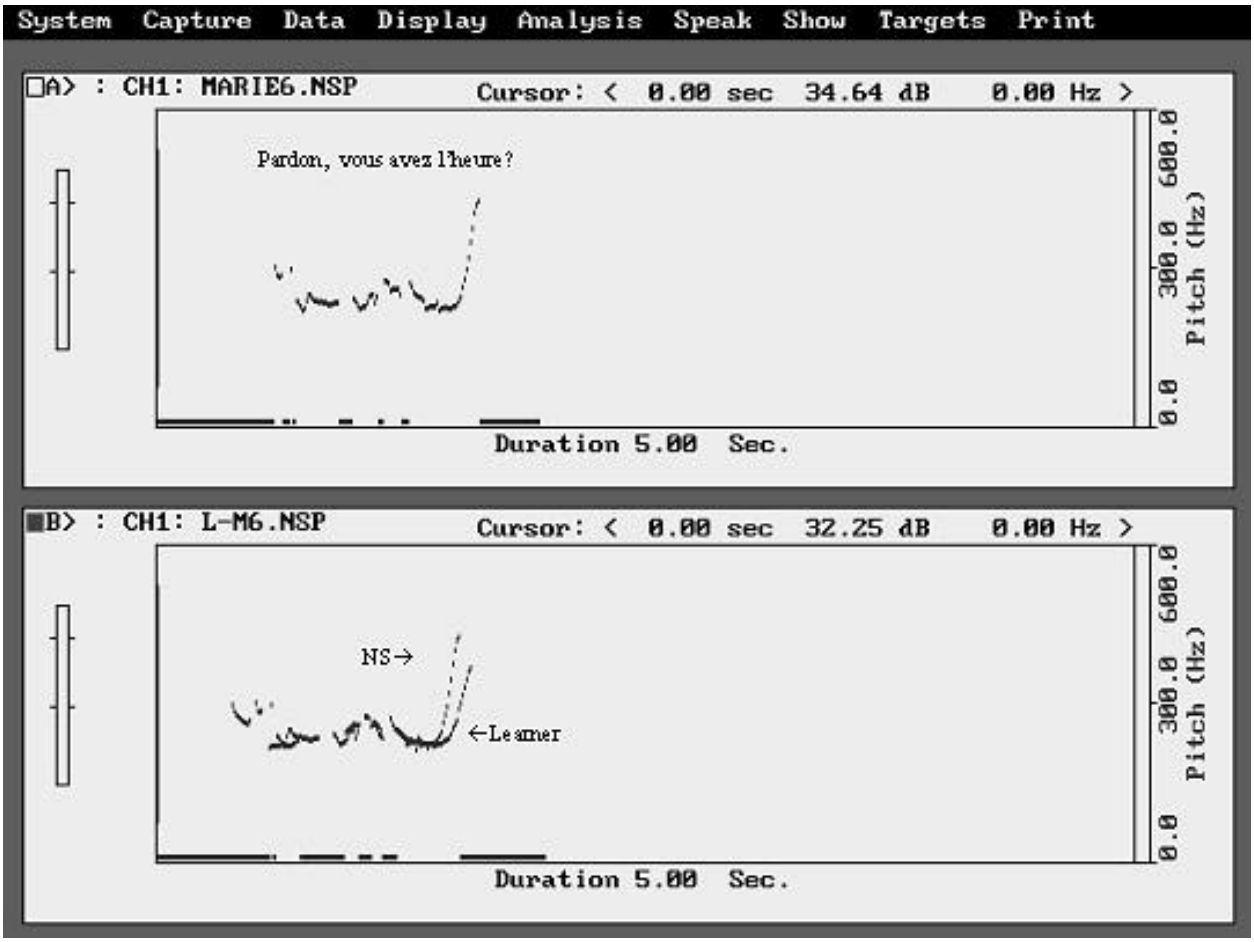

Figure 4. NS pitch contour (top, Screen A). Pardon, vous avez l'heure? (Excuse me, do you have the time?) The text follows the corresponding areas of the pitch contour. The final steep rise is on the last word *l'heure.* Learner's training production (Screen B) shows NS pitch overlay. The final rise on *l'heure* "time" is the only clearly distinguishable difference; the leftmost pitch track belongs to the native speaker.

By the middle of Week 2, other elements were then mentioned. For example, one learner noticed she said [dɐ] instead of [d(ə)][8] for the preposition *de* ("of" in this case although it has various meanings), failed to make an obligatory liaison, and mispronounced some content words (e.g., *maison* "house"). Another learner said she realized she sometimes mispronounced the interrogative pronoun *que* "what" [k(ə)] as [kɐ], which she attributed to having studied Spanish.

At the end of week 2 and the beginning of week 3, participants commented on the following features: the contribution of liaison environments to fluent speech (e.g., *en‿hiver* "in winter" where the nasal consonant [n] of *en* is pronounced before the noun *hiver* which begins with a vowel sound)[9]; words with "dropped" sounds (e.g., pauv(re) garçon "poor boy"); barely perceptible words (e.g, *Je vais (lui) téléphoner* "I'm going to call him/her"); and the pronunciation of elided forms of articles (e.g., *l'université* "the university"), some main content words, and finally individual sounds, especially [ʀ].

In sum, over the course of the 3 weeks of training, there appeared to be a hierarchy of what the learners noticed (i.e., what they explicitly mentioned), beginning with more global elements such as the pitch

contour -- the obvious focus of training and visual feedback -- and moving towards an awareness of more local elements such as individual sounds.

Questionnaire Responses

As noted earlier, at the conclusion of each training program, participants were given a questionnaire to complete and return to me anonymously in order to assess their perceived value of this type of speech technology in foreign language instruction. Of the 16 participants in the experimental group, 13 returned the questionnaires. The responses shown below are listed according to frequency of occurrence on the returned questionnaires. Several (here, in italics) emphasize the relationship between prosody and music or make specific reference to voice characteristics that are often components of voice-training techniques commonly used in the fields of music and theatre.

1) What were the most difficult elements of French pronunciation for you **before** this program?
   - avoiding monotone speech, intonation, "r" sound, liaison
2) What do you feel you focused your attention on **during** the 3 weeks of practice?
   - intonation, pronunciation, speed, liaison
3) What elements of the speech of native French speakers did you notice **after** this program that you had not noticed before?
   - rise and fall pattern of intonation, more range of tones, pronunciation of "r, "declining mountain peaks of pitch, pitch contour at the end of sentences, fluid speech like "*humming a tune*"
4) What have you noticed about your own pronunciation in French as a result of this program?
   - intonation, more movement in my speech, comfortable with more words, "r," speed, vowels, liaison
5) What do you feel you've accomplished in terms of your pronunciation in French?
   - "I try to *sing phrases* more."
   - "My sentences became more *flowing*, instead of just dots on the screen."
   - "I really had never thought about the intonation affecting the clarity and fluency of a language. It is something I keep in mind now as I learn the language."
   - "I feel that a native French speaker will have an easier time understanding my speech as well as the *smoother rhythm* of my sentences."
   - "I'm much better at *pitch variation* and less afraid of going to high pitches."
   - "I noticed that I *could use my voice range more* and that I could connect certain words for a more authentic speaking rhythm."
   - "I feel more confident with my pronunciation for sure."
   - "I gained a better understanding of the language as well as developed my tone variation and I had a lot of fun. Thank you."

EXPERIMENT 2

Experiment 2 explored the relationship between prosody and the lexical content of sentences in long-term memory. In a training study with learners' attention focused on prosody, the question arises as to whether the prosodic patterns of the training sentences, each of which was presented and practiced several times over the training period, had become key components of memory traces so that the prosody itself (i.e.,

through the use of filtered speech) could facilitate retrieval of the lexical content of a sentence from memory. Based on Goldinger's (1997) findings, I hypothesized that the lexical information corresponding to the prosodic patterns that drew the most learner attention would be the easiest to recall.

Method

Participants. All 16 participants from the experimental group and 10 from the control group in Experiment 1 completed Experiment 2. Recall that they had not been told of this task in advance.

Materials. From the set of sentences used in training, 20 were selected to represent a range of structural types and all training talkers. This subset was filtered using the CSL filtering program; however, as there is not a unique correspondence between the suprasegmental and segmental features of sentences, the objective for this experiment was to reduce substantially, but not eliminate, the intelligibility of the lexical content of the sentences so that prosodic information was the principal lexical access cue. NS judgments were used to determine the appropriate level of reduction.

Procedure. Experiment 2 was conducted following the test of generalization from Experiment 1. The filtered sentences were played through the studio speaker to the participants who were asked to try to recall as many of the words of each sentence as they could. Each sentence was played three times. Their responses were made orally.

Results and Discussion

Results revealed that the training group from Experiment 1 was able to recall the exact lexical content of an average of 80% of the filtered sentences, and the content words only of an additional 10% of the sentences. Performance was quite consistent across participants. The recall of sentences *J'en ai marre!* (I'm fed up!) and *Comme c'est bizarre!* (How strange!) by all participants may be attributed to their more expressive semantic nature and corresponding prosodic contours. Other sentences were also recalled by all participants such as *Pardon, vous avez l'heure?* (Excuse me, do you have the time?) and *Vraiment? Vous avez mangé tout ça?* (Really? You ate all that?). These were associated with the auditorily and visually salient pitch rise marking the utterance as a question that learners frequently commented on when seeing the steep slope of the contour (see Figure 4).

The lexical content of sentences such as *Elle a une soeur et deux frères* (She has one sister and two brothers) and *Voilà Louise. Elle est française* (There's Louise. She's French), both with simple declarative intonation patterns, was recalled completely by about half of the participants. The others said the prosody sounded "familiar" but could not recall the words. The lower rate of recall for simple declaratives is compatible with Goldinger's (1997) findings for generic echoes from multiple-trace activation although there were some apparent exceptions. For example, the sentence *Elle a choisi la jupe rouge* (She chose the red skirt) may have been recalled because most learners had commented on the speaker's ability to link the final consonant of *jupe* with the following word *rouge.*

Therefore, prosodic and lexical information does appear to be stored together in memory traces. The best recall results were obtained for training sentences that attracted the attention of learners because of visually and/or auditorily salient prosodic contours (e.g., those showing a substantial pitch range and/or steep slope), expressive lexical content, and features they had identified as particularly difficult for them to produce (e.g., the linking between *jupe* and *rouge* in the above example). Participants from the control group of Experiment 1 who had not been exposed to the training sentences at all were unable to identify any words.

GENERAL DISCUSSION

Results of Experiment 1 revealed significant effects of computer-assisted training in the acquisition of L2 prosody and, importantly, generalization to segmental accuracy and novel sentences.

The question might arise as to whether other non computer-based training approaches such as the traditional teacher-led instruction would be as or more effective, and one might be inclined to attempt to compare these approaches experimentally. However, I would suggest that such a comparison is inherently flawed. As there are numerous elements that make up a training approach, all but the specific one under investigation would need to be the same in both approaches to avoid a confound. Such a degree of control appears, at best, challenging. Simply using the same materials for the same period of time would not provide a basis of comparison.

Consider the following elements of the present computer-assisted training program that would need to be duplicated in an instructor-led approach. Feedback involved 30 sentences spoken by each of three talkers. We know that talker and stimulus variability contribute significantly to successful L2 speech training (Lively et al., 1993); therefore, in a non computer-assisted approach, three different instructors would be needed to provide feedback throughout data collection that, in the present study, required blocks of several hours set aside each day throughout the week for 3 weeks, and in total, spanned several months. Moreover, what feedback (also a significant factor in successful training) would be given by the instructors? Recall that segmental accuracy in this study was not a focus of training for the participants but part of the investigation of generalization; therefore, an instructor would have to restrict feedback to prosody only. Several more questions then arise. Could all instructors do **exactly** the same thing? If not, another variable enters the picture. In addition, how would feedback be provided by instructors on prosody only? The computer program provides a visual display in real-time and the opportunity to overlay the NS version on the learner's -- a salient form of feedback that drew many positive comments from participants. There is also the issue of the effects of learning styles and preferences. For some learners, technology holds greater interest, which influences motivation. Some enjoy a greater comfort level in working with a computer program than in face-to-face interaction where other personality factors are involved. While not an exhaustive list, the above points serve to emphasize that comparison of approaches, in general, is highly problematic. Note, however, that these comments are not intended as a claim that a particular type of training is best for all learners nor that instructor-led approaches are not beneficial, only that direct comparison is not well-founded.

This study's objective was not to determine whether computer-assisted training or a particular software program was better than any other approach; the objectives centered on the generalization of computer-assisted prosody training to segmental improvement and to novel sentences, the learners' allocation of attentional resources throughout training, and their responses to the use of this type of technology as pitch contour display is available in various products and web sites. Therefore, my control group was not designed as a control for training *approach* but for training *itself.* Several authors have alluded to the benefits of computer-assisted training but it had not been tested thoroughly nor investigated for its ability to generalize -- also a hallmark of successful L2 speech training.

Learners did appear to allocate their attentional resources hierarchically, and the following questionnaire comments support my observation during training further: "It's hard to get all of the elements together that are necessary for producing accurate sentences" and "I gained an awareness of all the aspects of learning to speak a language fluently." As the training program was explicitly designed to deal with prosody, and the visual feedback best represents intonation, this feature took precedence and the learners' attention initially; however, as they became more confident with this aspect of their language production, they were able to notice other elements such as liaison and the production of specific sounds.

In addition, the results of Experiment 2 revealed that prosodic cues facilitated the recall of the lexical content of sentences to which the learners had been exposed frequently during training. This finding is compatible with exemplar-based learning models in which all attended perceptual details of events are stored as traces in memory. Those exemplars whose prosodic and/or lexical content attracted the most learner attention in the study were the easiest to recall.

Taken together, the results of these two experiments demonstrate the effective pedagogical application of speech technology. This training strengthened the association between the prosodic and lexical components of sentences through frequent exposure and through practice opportunities involving sentences of familiar content with informative feedback, and the training also resulted in improved production at both the segmental and suprasegmental levels. Further experiments involving more learners at different levels of proficiency and studies of retention would contribute to our understanding of the potential of this approach. In addition to the quantifiable results, my observations during training and the learners' responses to a questionnaire indicated an increase in their confidence in using the language and the raising of their awareness of its various components. One learner commented that, "with practice the sound patterns are easier to recall and produce." The obvious pedagogical implication of this statement is consistent with the results of both experiments in this study; that is, frequent input, use of contextualized vocabulary with applications to daily life, a range of syntactic and prosodic structures, practice opportunities, and auditory and visual feedback contribute to learning.

APPENDIX

Sample Testing and Training Sentences

Il y a beaucoup de fleurs dans le jardin. "There are many flowers in the garden."

Ils ont lu des romans intéressants. "They read some interesting novels."

Mon amie Marie est très sérieuse. "My friend Mary is very serious."

Ma famille m'a envoyé des cadeaux. "My family sent me some gifts."

Ce sont des chocolats belges. "These are Belgian chocolates."

Caroline préfère le vin rouge. "Caroline prefers red wine."

Elle a une soeur et deux frères. "She has one sister and two brothers."

Est-ce que vous avez voyagé au Canada? "Did you travel to Canada?"

Quel âge a Jean? "How old is John?"

Quels livres avez-vous choisi? "What books did you choose?"

Excusez-moi, pourriez-vous m'aider? "Excuse me, could you help me?"

Jean est malade? C'est dommage. "John is ill? That's too bad."

Marianne va travailler demain? "Marianne is going to work tomorrow?"

Bonne idée! Allons au magasin aujourd'hui. "Good idea! Let's go to the store today."

Vous n'allez pas croire cette histoire! "You're not going to believe this story!"

J'en ai marre! "I'm fed up!"

NOTES

1. This study was supported, in part, by the Center for Language Education and Research (CLEAR) at Michigan State University. I gratefully acknowledge the assistance of Jayne Niemann, coordinator of elementary and intermediate French courses at MSU, for help in stimulus selection and data collection, and the native speakers whose voices were recorded and those who served as raters. Portions of this paper were presented at the *American Association for Applied Linguistics Conference* in Salt Lake City in April, 2002, and the *International Conference on Spoken Language Processing, Special Session on the*

Integration of Speech Technology and Language Learning, in Denver, 2002. I am grateful to attendees for their questions and to Martha Pennington for comments on an earlier draft of this paper.

2. Episodic models based on multiple-trace memory theory also incorporate the storage of an abstract or prototype representation (e.g., a phoneme). See Hardison (2000) for details on the application of this theory to second-language speech development.

3. Di Cristo (1998) defines the term General French as the variety used by educated people and professional radio and television speakers characterized by the absence of dialectal marks. He refers to it as equivalent to General American for American English.

4. I do not consider the assignment of the participants to the control group a challenge to the concept of random assignment as schedule conflicts were the sole determining factor and this is irrelevant to the study. It was important to ensure that the control participants were equally motivated as those in the experimental group given the role of motivation in learning.

5. Liaison refers to the linking between sounds, where a sound is produced at the end of a word when preceding a certain context (e.g., the final consonant of one word pronounced before another word beginning with a vowel sound) as in *Elles‿adorent lire des romans* (They love reading novels) where "s" at the end of the feminine plural subject pronoun *Elles* is pronounced /z/ before the vowel sound beginning the verb *adorent.* In contrast, this sound is not pronounced in the sentence *Elles vont acheter du fromage* (They are going to buy some cheese) where the verb *vont* (from *aller* "to go") begins with a consonant sound.

6. The CSL digital filtering program allows the user to determine various settings. A low pass filter was used to reduce the level of signal components above the frequency level determined by examination of the spectrogram with overlay pitch extraction. There are several types of windows; the CSL manual recommends a Blackman window weighting for this purpose. A filter order can be selected between 3 and 100. The higher the number, the greater the filtering; therefore, to filter as much as possible, a 100^{th} order was selected. The determination of the cutoff was calculated relative to the Nyquist Frequency (half the sampling rate, i.e., 12,800/2=6,400). Cutoff = Shoulder frequency (where filtering will begin or 300 in this case) divided by the Nyquist Frequency (i.e., 300/6,400). Cutoff = .04

7. A reviewer suggested an ANOVA involving Group (experimental, control), Time (pretest, posttest), and Presentation Type (filtered, unfiltered). This was not done for two reasons: (a) this approach omits the segmental accuracy data, as filtered speech has no segmental content and (b) the raw data from the control group do not justify their inclusion in statistical analysis.

8. Parentheses around ə indicate its omission in some cases in connected speech.

9. There are two classes of words beginning orthographically with *h: h muet* and *h aspiré*. The first group (e.g., *hiver* "winter," *heure* "hour") is subject to liaison and elision producing phrases such as *en‿hiver* (in winter) as described in the text, and *l'heure* (the hour) with the elided form of the article. Words in the second group (*h aspiré)* also begin with a vowel sound in Modern French but are not subject to liaison or elision (e.g., *le haricot* "bean" without elision of the article).

ABOUT THE AUTHOR

Debra M. Hardison is assistant professor of second language acquisition at Michigan State University. Research interests include spoken language processing and computer-assisted second-language speech training. Publications appear in *Language Learning, Applied Psycholinguistics,* and various edited

collections. She teaches courses on second language acquisition theory and research, advanced studies in language teaching, and second-language speech.

E-mail: hardiso2@msu.edu

REFERENCES

Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System, 20*(1), 51-62.

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

Bertinetto, P.-M. (1989). Reflections on the dichotomy "stress" vs. "syllable-timing." *Revue de Phonétique Appliquée, 91-93*, 99-130.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages.* Cambridge, England: Cambridge University Press.

Chun, D. M. (1998). Signal analysis software for teaching discourse intonation. *Language Learning & Technology, 2*(1), 61-77. Retrieved July 22, 2002, from http://llt.msu.edu/vol2num1/article4/

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.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*(3), 393-410.

Di Cristo, A. (1998). Intonation in French. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 195-218). Cambridge, England: Cambridge University Press.

Fletcher, J. (1991). Rhythm and final lengthening in French. *Journal of Phonetics, 19*(2), 193-212.

Fónagy, I. (1979). L'accent français: Un accent probabilitaire. *Studia Phonetica 15*, 123-233.

Goldinger, S. D. (1997). Words and voices: Perception and production in an episodic lexicon. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 33-66). San Diego, CA: Academic Press.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context and talker variability. *Applied Psycholinguistics, 24*, 495-522.

Hardison, D. M. (2000). The neurocognitive foundation of second-language speech: A proposed scenario of bimodal development. In B. Swierzbin, F. Morris, M. E. Anderson, C. A. Klee, & E. Tarone (Eds.), *Social and cognitive factors in second language acquisition* (pp. 312-325). Somerville, MA: Cascadilla Press.

Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics.* New York: Newbury House.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93*(4), 411-428.

Hirst, D. J., & Di Cristo, A. (1984). French intonation: A parametric approach. *Die Neueren Sprachen 83*(5), 554-569.

Hirst, D. J., & Di Cristo, A. (1998). A survey of intonation systems. In D. Hirst & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 1-44). Cambridge, England: Cambridge University Press.

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather & A. James (Eds.), *New Sounds 90* (pp. 72-97). Amsterdam: University of Amsterdam.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America, 94,* 1242-1255.

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.

Munro, M. J. (1995). Nonsegmental factors in foreign accent. *Studies in Second Language Acquisition, 17*(1), 17-33.

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. *Journal of Memory and Language, 45,* 526-545.

Pennington, M. C., & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *Modern Language Journal, 84*(3), 372-389.

Pennington, M. C., & Esling, J. H. (1996). Computer-assisted development of spoken language skills. In M. C. Pennington (Ed.), *The power of CALL* (pp. 153-189). Houston, TX: Athelstan.

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

Vaissière, J. (1974). *On French prosody* (Quarterly Progress Report No. 114). Cambridge, MA: MIT Research Laboratory of Electronics.

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

Wenk, B. J., & Wioland, F. (1982). Is French really syllable-timed? *Journal of Phonetics, 10,* 193-216.