
Contextualized Computer-based L2 Prosody Training: Evaluating the Effects of Discourse Context and Video Input

DEBRA M. HARDISON
Michigan State University

ABSTRACT

Two types of contextualized input in prosody training were investigated for 28 advanced L2 speakers of English (L1 Chinese). Their oral presentations provided training materials. Native-speakers (NSs) of English provided global prosody ratings, and participants completed questionnaires on perceived training effectiveness. Two groups received training input using *Anvil*, a web-based annotation tool integrating the video of a speech event with visual displays of the pitch contour, and practiced with *Real-Time Pitch (RTP)* in *Computerized Speech Lab* including feedback from a NS. Two groups used only *RTP* to view their pitch contours and practiced with the same feedback. Within these pairs, one group received discourse-level input and the other individual sentences. Each group served as its own control in a time-series design. All had comparable levels of performance prior to training. Results indicated that although all groups improved as a result of training, discourse-level input produced better transfer to novel natural discourse. The presence of video was more helpful with discourse-level input than with individual sentences. Speech samples collected 1 week after training revealed sustained improvement. Questionnaire results support the use of computer-based tools and authentic speech samples. Findings strongly suggest that meaningful contextualized input is valuable in prosody training when the measurement is at the level of extended connected speech typical of natural discourse.

KEYWORDS

Prosody, Training, Discourse, Video, Context

INTRODUCTION

Current directions in the field of second-language (L2) pronunciation emphasize both segmental and suprasegmental features of the language and, importantly, encourage teachers to base decisions concerning pedagogical focus and appropriate sequencing of learner attention on learner needs and the functional use of language (e.g., Celce-Murcia, Brinton, & Goodwin, 1996; Morley, 1991). These

concerns have established criteria of high quality, authenticity, and adaptability for computer-based speech training materials or tools to meet. Studies have shown that computer programs providing a visual display of a pitch contour are effective tools for training L2 learners to produce more native-like prosody (e.g., de Bot, 1983; Hardison, 2004; Weltens & de Bot, 1984). This auditory-visual feedback is significantly better for L2 speakers than auditory-only (de Bot, 1983), and such training can generalize to novel sentences and improved segmental accuracy (Hardison, 2004). Typically these studies use decontextualized scripted sentences for testing and training. Although this approach provides control over content and offers direct comparison of the same stimuli produced in a pretest and posttest, the question arises as to whether computer-assisted training can improve the production of L2 prosody in the natural discourse-level speech that is more often required of a speaker in the language environment.

This paper reports the findings of a study conducted with advanced L2 speakers of English (graduate students) to evaluate the effects on their production of discourse-level prosody of different types of contextualized training using speech segments from the participants' own series of oral presentations on a variety of familiar topics. This provided a speech context referred to by Flowerdew and Tauroza (1995) as "conversational lecture" (p. 442). Their first language (L1) was Mandarin Chinese (Taiwan). Two computer-based tools were used to compare the contributions of different types of contextualized input in training; that is, a comparison was made between training with and without the visual context of the speech event, and with discourse-level input versus isolated sentences in a 2x2 design. The tools were (a) web-based *Anvil* (Kipp, 2001),¹ developed as an annotation tool for multimodal dialogue, that provides a screen display integrating the audio and video components of a speech event with the associated time-aligned pitch contour created in *Praat*,² a public domain phonetic tool, and (b) Kay Elemetrics *Real-Time Pitch (RTP)* program in conjunction with the *Computerized Speech Lab* that produces a pitch contour in real-time and allows on-screen comparison of a learner's utterance with that of a native speaker (NS) for feedback including overlay of one contour on the other. Although a screen display best represents intonation, intonation is inextricably linked to tempo, stress, and rhythm especially with discourse-level speech; therefore, this paper will refer to the study as one of prosody rather than intonation, per se. The evaluation of training effectiveness was based on NS global ratings of prosody in natural speech samples gathered from each participant's series of presentations before and after training and a posttraining questionnaire.

The content of training was dictated by the speech produced by the individual participants; therefore, it involved authentic material and addressed participants' specific needs (Morley, 1991). Avery and Ehrlich (1992) note that Mandarin speakers often have difficulty with English intonation and linking. The absence of the latter in L2 English produces a "choppy" type of speech. This generally describes how the use of prosody by the current study's participants was perceived by NSs of English. More objectively, their prosodic difficulties fell into the category described by Chun (2002) as discourse functions of intonation that

contribute to cohesion in speech, including the marking of thought groups with appropriate pausing and pitch movement, and the use of stress and intonation to mark information focus and express contrast and emphasis. Wennerstrom (1998) found that effective intonation use by NSs of Chinese, specifically in terms of increased pitch at rhetorical junctures to signal topic shift in an informal lecture, contributed to higher production ratings on a scale similar to that used in the Speaking Proficiency English Assessment (SPEAK) test.

The use of technology for the visualization of prosodic features has been recommended by a number of researchers (e.g., Anderson-Hsieh, 1992, 1994; de Bot & Mailfert, 1982; Leather, 1990; Molholt, 1988; Pennington & Esling, 1996). Although the visual display of the contour is easily interpretable by nonspecialists, which is another important criterion in the assessment of a learning tool (Chun, 1998), and overlays of NS contours on those of the learners constitute valuable feedback (Hardison, 2004), the scope of research inquiry has usually been limited in terms of computer-assisted learning to individual scripted sentences. Therefore, the question that arises is whether training with such sentences transfers to improved discourse intonation in the learners' subsequent speech. This question motivated the present investigation of the effects of contextualized input in training: (a) Does training with discourse-level materials (linguistic context) produce better transfer to natural discourse-level speech than training with individual sentences? (b) Does the addition of the video recording of a speech event such as an oral presentation facilitate training by presenting the auditory-visual context for prosody?

In addition to a comparison of training approaches in the improvement of L2 prosody, the current study also took as important objectives an increase in the participants' self-confidence in speaking in English, and an awareness of prosodic features of the language, which might promote effective self-monitoring strategies beyond the training program. Based on the results of an earlier study using visual pitch displays with American learners of French (Hardison, 2004), I hypothesized that participants in all experimental groups would show some degree of improvement across training but that contextualized input (i.e., with video and/or linguistic context) would provide an advantage in terms of transfer to improved discourse-level prosody. Through questionnaire responses, I expected participants to report increased confidence in and awareness of the prosodic components of their oral communication.

EXPERIMENTAL DESIGN

A total of 28 L2 speakers of English (4 groups of 7 each) volunteered to participate in this study. Each participant gave a series of oral presentations on topics of daily life, each lasting approximately 10 minutes. The rest of the participants served as the audience. Each individual's presentations provided the source for his/her training materials. Two groups received training input that involved both the video of segments of their presentations with the associated pitch contour (using *Anvil*) and practiced with the *Real-Time Pitch* program (*RTP*); the other two groups received training input involving only the visual pitch display using *RTP*

and practiced with this program. Within each of the pairs of groups, the training materials for one group involved discourse-level segments from their presentations, while those for the other were individual sentences. Thus, there were four experimental groups.

For all groups, a time-series approach was used with a total of five presentations for each participant. This provided more contexts from which training materials could be taken and allowed the participants to serve as their own controls (Hatch & Lazaraton, 1991). Following the initial presentation, the members of each group made two subsequent presentations, each separated by 1 week to establish the usual pattern of performance prior to training.³ These sessions were recorded (video recording for the groups whose training involved *Anvil* and audio recording for the groups using only *RTP*). After 2 weeks of training (10 sessions, each about 45 minutes), two subsequent presentations were made by participants—one immediately following the training to determine if a change had occurred and the final one after a delay of 1 week to determine if improvement had been sustained. These posttraining presentations were audio recorded for rating purposes only. For each participant, 5 weeks separated the first and last recordings.

Because it was not possible to train all 28 participants during the same 2-week period of time, staggered sessions were scheduled. As one group began training, another participated in their first presentations. To establish some pretraining equivalence across groups in terms of global performance on English prosody, I met with each participant individually and then assigned them to one of four experimental groups. This assignment was based on my assessment only and ensured a comparable range of abilities within each group. Their pretraining use of prosody was later evaluated by NS raters. The above approach was favored over attempting to control for pretraining differences in the later analysis of data, given the overall design of the study, and permitted more confidence in a between-groups comparison of training type. The initial meeting also allowed me to establish a rapport with each individual.

Following the last recording, participants were asked to complete an anonymous questionnaire (see Appendices A, B and C) on their perceptions of the value of their training. Global prosody ratings on a 7-point scale were obtained from native speakers of English for speech samples gathered at different points throughout each session: the first presentation to evaluate pretraining equivalence of groups, the third presentation to allow comparison between the first and third presentations to monitor usual production performance as a baseline before training, the fourth presentation (posttraining) for comparison with the third presentation (pretraining) to assess the effects of training, and the fifth presentation for comparison with the fourth presentation to evaluate sustained improvement.

METHOD

Participants

The 28 participants (L1 Chinese; 24 female, 4 male) were graduate students at an American university. None reported any vision or hearing problems, and none

was making presentations in their academic classes or serving as teaching assistants during the period of this study. All had TOEFL scores over 550 (paper-based test) and were not taking any English classes. Although some exhibited a few difficulties at the segmental level, these difficulties were not sufficient to disrupt comprehension. The students were majoring in a variety of disciplines primarily within the natural sciences, math, and economics.

Materials

Recording and Selection for Training

Each presentation was about 10 minutes in length. Participants were permitted to refer to small note cards and use the blackboard or overhead projector to present diagrams, pictures, or maps (no text). Little reference to notes was needed since they were very familiar with the content of the presentations. The topics included their families, hometowns, fields of study, and career goals. The objective was to have the same general topic for all participants to avoid an influence of topic on NS ratings of their speech samples.

Recordings were made in a classroom with overhead lighting. Those who were assigned to the groups that would use only *RTP* in training wore a lavalier microphone connected to a Sony DAT recorder carried in a pocket or pouch around the waist. For those assigned to the groups that would receive training including *Anvil* (i.e., with video input), a Sony Hi-8 video camera was used with an Electrovoice lavalier microphone for the first three presentations which provided the materials for the following training sessions. The camera was focused on the speaker who was allowed to move around during the presentation. For the two presentations following training, audio recording was done following the above procedure.

For all participants, I selected segments from the first three of their own presentations to use in their training focusing on utterances where prosody was problematic, representative of the individual's oral L2 production, and compatible with training using pitch contour displays. The majority of these problems involved a lack of target-like linking of related constituents, inappropriate pausing that disrupted the cohesiveness of thought groups, and misuse of stress and intonation to mark information focus appropriate to a given context.

For the two groups of participants whose training involved only *RTP*, the above segments were transferred from DAT tapes to computer hard disk and saved to files either as individual sentences or discourse-level chunks (i.e., sentences that constituted a cohesive unit) depending on the training-group type. For the two groups that were videotaped for training, the segments I chose were transferred to CDs as separate *QuickTime* movies and saved as files as described above.⁴ These files were then opened in *Anvil* version 4.0.8 running on Windows XP.⁵ In this program, the pitch contour is created in *Praat*. Using *QuickTime Pro*, the sound track was extracted from the movie and saved as a file in .wav format for use in *Praat*.⁶ The *Anvil* screen in this study provided the participant with the video clip selected from their presentation and the time-aligned pitch contour.⁷

Questionnaires

Participants in all groups were asked to complete a questionnaire that included closed-ended items using a 5-point Likert scale dealing with the effectiveness of the training overall and of features specific to the type of training they received (e.g., use of both *Anvil* and *RTP*, use of discourse context vs. single sentences), materials used in training, and the effects of training on their awareness of prosodic features and perceived confidence and speaking ability in English (see Appendices A and B). There were also open-ended items allowing respondents to offer comments on their experience with these computer-based tools and suggestions for training (see Appendix C).

Training Procedure

For each participant, there were 10 training sessions each about 45 minutes over a 2-week period.

Training Including Integrated Video and Pitch Contour (*Anvil* and *RTP*)

The *Anvil* and *RTP* programs run on separate computers in my lab. At the beginning of each participant's training period, a segment that I had selected from the participant's presentations and saved to a file (either discourse-level chunk or individual sentence depending on the group) was played using *Anvil*. This provided an audiovisual display of the speech event and the time-aligned pitch contour. The segment was played as many times as the participant requested and then production of the segment was practiced using *RTP* and a headset condenser microphone (AKG C420) that permitted freedom of head movement. The pitch contour of this utterance was displayed in real time in View Screen A. A NS version (produced by the author) was then provided in View Screen B (see Figure 1 below). One can be overlaid on the other in different colors for direct comparison.⁸ The screens were cleared and the segment was practiced again until the participant felt comfortable with the production and was ready to go on. Although there is a limit to the amount of speech one can see on a screen at one time, in "walking display" the pitch contour in *RTP* continues to draw on the screen moving from left to right.

Training with Pitch Contour Display Only (*RTP*)

For the groups working only with *RTP*, the procedure was similar but with the omission of the *Anvil* screen display (i.e., video and time-aligned pitch contour). The preselected speech segments had been saved as audio files in the *RTP* program. These files were opened on the screen and played, followed by production of the NS version for feedback comparison as described above and subsequent practice by the participant.

RESULTS AND DISCUSSION

Prosody Ratings

Three native speakers of American English volunteered to rate the native-like

quality of the prosody in the speech samples produced by the participants in this study. Because of staggered training sessions, it was possible to have the same raters for all analyses. For each sample, the raters were given a 7-point scale ranging from "1" (definitely not native) to "7" (definitely native-like) and asked to provide a global rating of prosody.⁹ They were not told whether the speech samples they were rating were taken from recordings before or after training or the type of training the group had received. They were not given any specific guidelines on what features to look for in order to avoid bias. The speech samples were randomized. For all ratings, interrater reliability was determined using a method involving the calculation of mean interrater correlations with correction for use with ordinal data by Fisher Z transformation (Hatch & Lazaraton, 1991) and ranged from .82 to .89.

Several analyses were conducted. First, in order to compare the training approaches, the raters assessed the pretraining equivalence of the four groups of participants in the use of English prosody. This assessment was done to corroborate my earlier evaluation. Therefore, for the groups that were audio recorded (for training with *RTP* only), raters listened to the recording of each participant's first presentation from the DAT tape using headphones. For those who were videotaped (for training with *Anvil* and *RTP*), the raters listened to the audio track only.¹⁰ The mean ratings for the four experimental groups were compared with a single-factor ANOVA and revealed no significant difference $F(3, 27) = 1.23, p = .32$ confirming an acceptable level of pre-training equivalence.

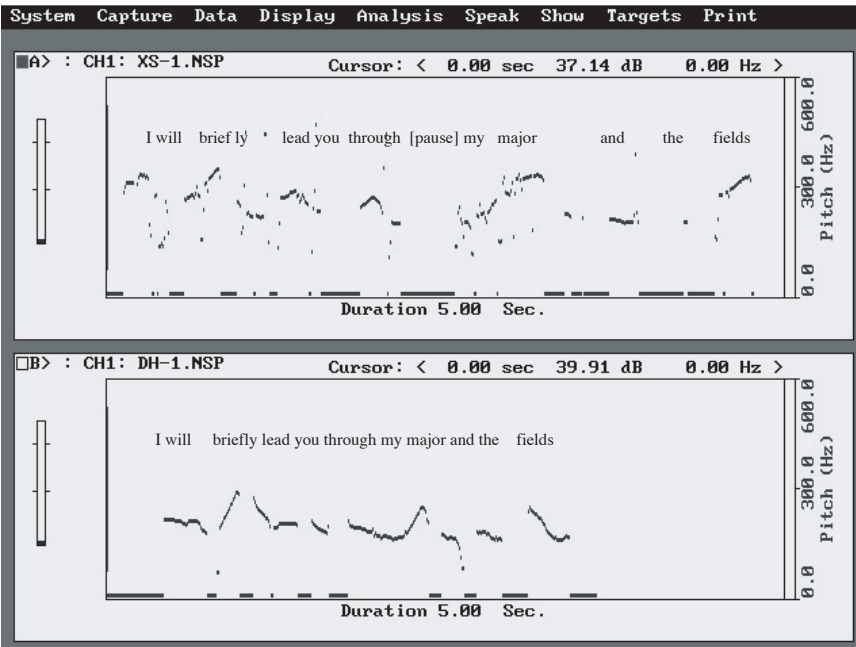
The second analysis compared the mean prosody ratings for the first and third presentations to determine if any significant change had taken place over a period of 2 weeks. Recall that in a time-series design, the participants serve as their own controls. The 2-week period of time for pretraining control was selected because it matched the duration of the training period and was compatible with the participants' schedules. Mean ratings for the third presentations also showed no significant difference across the groups; therefore, the data were collapsed across groups for comparison between the first and third presentations. A paired *t* test revealed no significant change during the 2-week period of time, $t = -1.06, df = 27, p = .30$.

The third analysis assessed the effects of the different training approaches with a 2 (Input type: discourse-level, individual sentences) x 2 (Display type: *Anvil* & *RTP*, *RTP* only) x 2 (Time: pre-, posttraining) ANOVA with repeated measures on the last factor. Pre- and posttraining data were the ratings from the third and fourth presentations, respectively. Analysis revealed main effects of time, $F(1, 51) = 15.21, p < .01$, input type $F(1, 33) = 12.15, p < .01$, and display type, $F(1, 28) = 6.38, p < .05$ indicating that all groups improved as a result of training and discourse-level input produced better transfer of improvement in prosody to natural discourse. There was a significant Input Type x Display Type interaction, $F(1, 32) = 8.62, p < .01$ pointing to the advantage of the presence of video (*Anvil*) with discourse-level input over that of individual sentences. The use of the video likely facilitated recall of the speech event, which, in turn, may have enhanced the value

of the training materials and the ability to transfer this training to novel natural discourse.

In Figure 1, Screen A shows a speech sample produced at the beginning of a participant's training. This individual was part of the group that worked with *Anvil* and *RTP* in training with discourse-level input. The full sentence was "I will briefly lead you through my major and the fields in it." The choppiness of the utterance is represented by the breaks in the pitch contour and the succession of peaks. There is also a long pause within the phrase "through my major." Within the words "I," "briefly," "lead," "through," "major," and "fields," there is a rise and fall of pitch. For comparison, the NS version of the same utterance appears in Screen B. Stress and intonation were produced in keeping with the context of this utterance in the presentation from which it was taken; therefore, "briefly," "major," and "fields" show some degree of prominence.

Figure 1
Speech Sample of a Participant at the Beginning of the Training



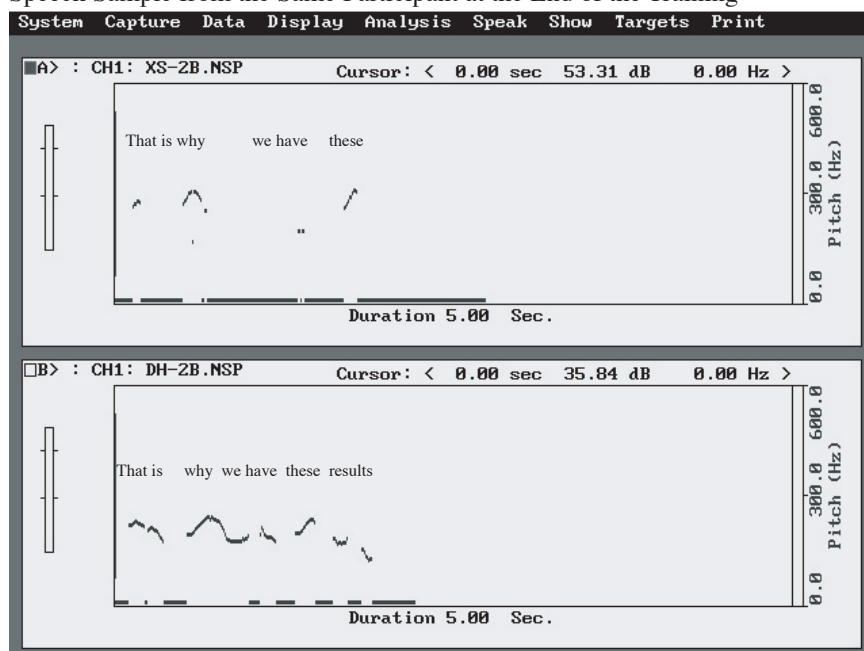
Note: Screen A shows the participant's utterance, and screen B shows the native speaker's version. The text follows the corresponding areas of the pitch contour. Note the rise and fall of pitch throughout the displayed portion of the sentence "I will briefly lead you through [pause] my major and the fields in it," especially the pitch movement within the monosyllabic words "I," "lead," and "through."

In Figure 2, Screen A shows a sample from the same participant at the end of the training period. The full sentence was "That is why we have these results in this study." There were still some interruptions in phonation; however, the produc-

tion was more fluent than the earlier example, without inappropriate pausing, and showed prominence of “That,” “why,” and “these” comparable to the NS version in Screen B. Again, the stress and intonation are appropriate for the context from which this utterance was extracted. Note that in different contexts one could produce this same string of words so that emphasis was placed only on “That” and on either “these” or “results.” This underscores the importance of the context of an utterance in prosody training.

Figure 2

Speech Sample from the Same Participant at the End of the Training



Note: Screen A shows the participant’s utterance, and screen B shows the native speaker’s version. The text follows the corresponding areas of the pitch contour.

The above findings strongly suggest that meaningful contextualized input is valuable in prosody training. That context may be the video component of a speech event with the time-aligned pitch contour as well as linguistic context; that is, when the measurement of native-like prosody is at the level of extended connected speech as is typical of natural discourse, training using discourse-sized units of language facilitates improvement.

To assess sustained improvement, comparison was made between the mean prosody ratings based on the audio recordings obtained from the immediate post-training presentations and those obtained 1 week later. There was no significant difference within this period of time although questions remain regarding long-term retention and perhaps further progress once speakers have been made aware of the features of prosody to which they should attend.

Questionnaire Results

All 28 participants in this study returned completed questionnaires. The results are given in Tables 1-4 according to training-group type for Part One of the questionnaires. Part One involved all closed-ended items using a 5-point Likert scale familiar to students (Strongly Agree, Agree, Neutral, Disagree, Strongly Disagree); Part Two offered them the opportunity to respond to two open-ended questions. The questionnaire items varied somewhat across the groups because the nature of the training differed. There were no responses in the “Disagree” or “Strongly Disagree” areas, and very few in the “Neutral” category. These results were not too surprising given the voluntary nature of students’ participation, their level of motivation, and lack of attrition during the study. Raw scores are given rather than percentages because of the number of respondents (7) per group.

Table 1
Responses from Training Group using *Anvil* and *RTP* with Discourse-Level Input
(*n* = 7)

Questionnaire item	Student responses		
	Strongly agree	Agree	Neutral
1. Training was effective overall.	6	1	0
2. Using my own speech was helpful.	7	0	0
3. Seeing the presentation video was helpful.	5	2	0
4. Use of <i>Anvil</i> and <i>RTP</i> ^a was helpful.	5	2	0
5. Training raised my awareness of intonation, etc.	7	0	0
6. The pitch display in real time was helpful.	4	2	1
7. The NS ^b pitch overlay was helpful.	6	1	0
8. I am more confident speaking in English.	4	3	0
9. Training improved use of intonation, etc.	4	3	0
10. My speaking in English overall has improved.	4	3	0
11. Training with context is better than individual sentences.	5	1	1
Total	57	18	2

^aRTP = *Real-Time Pitch* program; ^bNS = native speaker.

Table 2
Responses from Training Group using *Anvil* and *RTP* with Individual Sentences
(*n* = 7)

Questionnaire item	Student responses		
	Strongly agree	Agree	Neutral
1. Training was effective overall.	4	3	0
2. Using my own speech was helpful.	6	1	0
3. Seeing the presentation video was helpful.	3	2	2
4. Use of <i>Anvil</i> and <i>RTP</i> ^a was helpful.	4	3	0
5. Training raised my awareness of intonation, etc.	4	3	0
6. The pitch display in real time was helpful.	4	3	0
7. The NS ^b pitch overlay was helpful.	5	2	0
8. I am more confident speaking in English.	4	2	1
9. Training improved use of intonation, etc.	4	2	1
10. My speaking in English overall has improved.	3	3	1
Total	41	24	5

^a*RTP* = *Real-Time Pitch* program; ^bNS = native speaker.

Table 3
Responses from Training Group using *RTP* with Discourse-Level Input (*n* = 7)

Questionnaire item	Student responses		
	Strongly agree	Agree	Neutral
1. Training was effective overall.	4	2	1
2. Using my own speech was helpful.	5	2	0
3. Training raised my awareness of intonation, etc.	6	1	0
4. The pitch display in real time was helpful.	5	2	0
5. The NS ^a pitch overlay was helpful.	6	1	0
6. I am more confident speaking in English.	5	2	0
7. Training improved use of intonation, etc.	4	3	0
8. My speaking in English overall has improved.	4	2	1
9. Training with context is better than individual sentences.	5	2	0
Total	44	17	2

^aNS = native speaker.

Table 4
Responses from Training Group using RTP with Individual Sentences (*n* = 7)

Questionnaire item	Student responses		
	Strongly agree	Agree	Neutral
1. Training was effective overall.	3	2	2
2. Using my own speech was helpful.	5	2	0
3. Training raised my awareness of intonation, etc.	3	4	0
4. The pitch display in real time was helpful	4	3	0
5. The NS ^a pitch overlay was helpful.	5	2	0
6. I am more confident speaking in English	3	2	2
7. Training improved use of intonation, etc.	3	2	2
8. My speaking in English overall has improved.	3	3	1
Total	29	20	7

^aNS = native speaker.

The responses clearly indicate a positive impression of the training program, the use of their own speech samples, and these computer-based tools. Both discourse-level input in training and the presence of the video clips (in *Anvil*) were favorably evaluated by the respective groups. The frequency of “Strongly Agree” responses differed significantly across experimental groups ($\chi^2 = 30.5$, $df = 3$, $p < .001$) taking into account in analysis only those questionnaire items the groups’ shared (i.e., the 8 items shown in Table 4), ranging from a total of 42 to 29 on those items. The highest number obtained for the group whose training involved *Anvil* and RTP with discourse-level input, followed by the group using RTP with the same type of input, the group with *Anvil* and RTP with individual sentences, and finally, the lowest number for the group who received RTP training with individual sentences.

In their comments on Part Two of the questionnaire, many respondents indicated they wished that such training had been available earlier in their study of English and that they became aware of the number of features of the spoken language that a learner needs to master to improve speaking skills. Many also commented that they feel it is important to work with the pronunciation of individual sounds as well as intonation, stress, and rhythm. These comments with regard to the increasing of awareness and perceived value of focused training are compatible with those from a different population with a much lower L2 proficiency from an earlier study (Hardison, 2004). Therefore, it seems reasonable to conclude that tools such as these and focused training programs have the potential of contributing to the production of L2 prosody at various stages of interlanguage development. It remains a question as to whether the raising of awareness alone, in the absence of any explicit training, would result in significant improvement.

CONCLUSION

This study focused on one measurable aspect of a speech event—prosody—and found that L2 speakers benefit from linguistic context (i.e., discourse-level speech) and visual context (i.e., video of the speech event) in terms of training input. In this study, the discourse-level training materials more closely resembled the type of connected speech on which the measurement of improvement was based, that is, the type of speech more often produced in the natural language environment. There was a significant improvement in the use of English prosody in novel natural discourse as a result of focused training using selected samples from the participants' own oral presentations. Although this source of training materials was not directly compared with more customary scripted speech, it is likely that such materials are more meaningful for learners especially when the speech event is recreated in training through the use of auditory-visual input.

NOTES

¹ For additional information on Anvil, see <http://www.dfki.de/~kipp/anvil> (last retrieved January 8, 2004). There is a link to a demo screen shot. Directions are given for those who wish to obtain the address for downloading the files (includes the manual). The tool is free for research purposes.

² The *Praat* program was created by Boersma and Weenink and is available at <http://www.fon.hum.uva.nl/Praat> (last retrieved January 8, 2004).

³ In a time-series design, there is no particular number of observations or period of time between observations. The objective is to establish the individual's usual pattern of behavior. This type of design can be implemented when the establishment of an appropriate control population is questionable or unfeasible.

⁴ Digital video recording was not available for this study. The transfer of selected segments from videotape to CDs was done with iMovie.

⁵ Installation of *Anvil* on Windows 98 requires modification of the batch file because different versions of Windows handle the installation of the video-processing package JMF (Java Media Framework) differently. The easiest installation was on Windows XP.

⁶ Transcription may be added to the pitch contour in *Praat*.

⁷ The *Anvil* screen provides users with the option of commenting on other elements of a speech event such as gestures. Users configure the "Annotation Board" to meet their needs. Instructions are given in the manual, which can be downloaded from the distributor (see Note 1). In the current study, the focus was on prosody in the context of the speech event.

⁸ The majority of the participants in this study were female; therefore, their frequency range was comparable to that of the NS author. For the male participants, the overlay of one contour on the other placed both in the same view screen, but there was generally some vertical separation because of the lower male frequency range.

⁹ In Hardison (2004), filtered speech (renders the segmental content unintelligible) was used to investigate the influence of hearing segmental information on the rating of prosody.

A significant difference in the prosody ratings between filtered and unfiltered speech was found prior to but not following training. This approach was considered but rejected for the present study. The participants in the two studies are substantially different in terms of their level of proficiency, and those in the current experiment did not have any segmental-level pronunciation problems that were disruptive to comprehension. In addition, from a technical standpoint, it was not feasible to attempt to load the speech samples for the two groups that were videotaped into the CSL filtering program.

¹⁰ The video portion of the recordings obtained for two of the experimental groups was not shown to the raters to avoid any bias.

REFERENCES

- Anderson-Hsieh, J. (1992). Using electronic visual feedback to teach suprasegmentals. *System*, 20, 51-62.
- Anderson-Hsieh, J. (1994). Interpreting visual feedback on suprasegmentals in computer assisted pronunciation instruction. *CALICO Journal*, 11, 5-22.
- Avery, P., & Ehrlich, S. (1992). *Teaching American English pronunciation*. Oxford: Oxford University Press.
- 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, 61-77. Retrieved January 8, 2004 from <http://llt.msu.edu/vol2num1/article4>
- Chun, D. M. (2002). *Discourse intonation in L2: From theory and research to practice*. Amsterdam: John Benjamins.
- de Bot, K. (1983). Visual feedback of intonation I: Effectiveness and induced practice behavior. *Language and Speech*, 26, 331-350.
- de Bot, K., & Mailfert, K. (1982). The teaching of intonation: Fundamental research and classroom applications. *TESOL Quarterly*, 16, 71-77.
- Flowerdew, J., & Tauroza, S. (1995). The effect of discourse markers on second language lecture comprehension. *Studies in Second Language Acquisition*, 17, 435-458.
- Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, 8, 34-52. Retrieved January 8, 2004, from <http://llt.msu.edu/vol8num1/hardison>
- Hatch, E., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York: Newbury House.
- Kipp, M. (2001). *Anvil* – A generic annotation tool for multimodal dialogue. In *Proceedings of the 7th European Conference on Speech Communication and Technology* (pp. 1367-1370). Aalborg, Denmark: Eurospeech.
- 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.
- Molholt, G. (1988). Computer-assisted instruction in pronunciation for Chinese speakers of American English. *TESOL Quarterly*, 22, 91-111.

- Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly*, 25, 481-520.
- 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.
- Weltens, B., & de Bot, K. (1984). Visual feedback of intonation II: Feedback delay and quality of feedback. *Language and Speech*, 27, 79-88.
- Wennerstrom, A. (1998). Intonation as cohesion in academic discourse: A study of Chinese speakers of English. *Studies in Second Language Acquisition*, 20, 1-25.

APPENDIX A

[For participants whose training involved the use of both *Anvil* and *RTP*. For each item, the scale appeared as SA A N D SD. Item number 11 appeared on the questionnaire only for the discourse-input training group.]

Questionnaire Part One: Now that you have completed your training program, please respond to each of the following statements by circling one of the options on the scale. SA = strongly agree; A = agree; N = neutral; D = disagree; SD = strongly disagree. Thank you very much for your participation.

1. This training program overall was very effective.
2. It was very helpful to work with my own speech.
3. It was very helpful to see the video of my presentation along with the pitch contour in training.
4. It was very helpful to work with both *Anvil* and *RTP* in training.
5. The training raised my awareness of intonation, stress and rhythm in English.
6. It was very helpful to see the pitch contour of my voice as I produced it.
7. It was very helpful to see an overlay of a native-speaker pitch contour on my production.
8. Since the training, I have been more confident speaking in English.
9. The training program contributed to improvement in my use of intonation, stress and rhythm in English.
10. My speaking in English overall has improved.
11. I think it is better to practice with context (a set of related sentences) instead of individual sentences.

APPENDIX B

[For participants whose training involved the use of *RTP* only. For each item, the scale appeared as SA A N D SD. Item number 9 appeared on the questionnaire only for the discourse-input training group.]

Questionnaire Part One: Now that you have completed your training program, please respond to each of the following statements by circling one of the options on the scale. SA = strongly agree; A = agree; N = neutral; D = disagree; SD = strongly disagree. Thank you very much for your participation.

1. This training program overall was very effective.
2. It was very helpful to work with my own speech.
3. The training raised my awareness of intonation, stress and rhythm in English.

4. It was very helpful to see the pitch contour of my voice as I produced it.
5. It was very helpful to see an overlay of a native-speaker pitch contour on my production.
6. Since the training, I have been more confident speaking in English.
7. The training program contributed to improvement in my use of intonation, stress and rhythm in English.
8. My speaking in English overall has improved.
9. I think it is better to practice with context (a set of related sentences) instead of individual sentences.

APPENDIX C

[For all participants. On the questionnaire, space was provided for comments after each question.]

Questionnaire Part Two: Please take a few moments to respond to the following questions. Your comments are valuable. Place your questionnaire in the envelope provided and return it to me through campus mail. Thank you again for your participation.

1. What elements or parts of your spoken language production in English would you like to see addressed in a training program?
-
2. Would you like to make any other comments about this training program?
-

AUTHOR'S BIODATA

Debra M. Hardison is Assistant Professor of second language acquisition in the Department of Linguistics and Languages at Michigan State University. Her research interests include speech perception and production, auditory-visual integration in spoken language processing, gesture and language, and computer-assisted second-language speech training. Her publications appear in *Applied Psycholinguistics*, *Language Learning*, *Language Learning & Technology*, and various edited collections. She teaches courses on second language acquisition theory and research, advanced studies in language teaching, and second-language speech.

AUTHOR'S ADDRESS

Debra M. Hardison
Michigan State University
A-714 Wells Hall
East Lansing, MI 48824
Phone: 517/353-0800
Fax: 517/432-1149
Email: hardiso2@msu.edu