ABSTRACT

One of the difficult problems facing teachers of phonetics is the lack of tools for training prosody (i.e., intonation, stress, and syllable rhythm). Our Interactive Prosody Trainer is a low-cost, microcomputer-based system for interactively teaching speech prosody. Based on a digital signal processing chip and an easy-to-use graphical interface, this device does two different, interrelated jobs. On output from the host microcomputer, it synthesizes models of utterances from stored LPC and prosodic features. This device does two different, interrelated jobs. On output from the host microcomputer, it synthesizes models of utterances from stored LPC and prosodic features (i.e., intensity, or the syllable timing of a model utterance). Using analog F0 and intensity data (SAID) required users to match, as closely as possible, either the F0, the intensity, or the syllable timing of a model utterance. Unfortunately, neither one can play back the model and/or the response for auditory comparison by the user, and both require the teacher to be present to evaluate the student's response. And yet they are both expensive. In other words, these devices are helpful aids for teaching the prosody, but they require extensive one-to-one interaction with a trained professional, and a majority of training programs can afford to buy only one.

Interestingly enough, Lane & Buiten [3] showed over 25 years ago that an interactive computer workstation could teach prosody effectively. Their so-called "Auto-Instructional Device" (SAID) required users to match as closely as possible, either the F0, the intensity, or the syllable timing of a model utterance. Using analog F0 and intensity extractors, plus a DEC PDP/1 micocomputer to calculate the match between model and response, the device cycled users successively among the three prosodic features until all three had converged to an acceptable degree. As successful as the SAID was in training fluent prosody, it is perhaps surprising that its principles have never been extended to modern microcomputers and digital signal processing technology. That time has now come.

1. INTRODUCTION

The teaching of prosody has not taken adequate advantage of modern speech technology. Beyond what is found in books or in the heads of phonetics teachers, there are just two audio-cassette resources and some not-very-interactive hardware. One of the audio-cassettes is the demonstration tape accompanying Cuetenden's text [1]. The author has simply read the examples from the text, as illustrations of the points he makes—a hardly a compelling pedagogy. The other cassette material is Allen's audio-imitation tutorial [2], used successfully for several years in his own phonetics classes but not disseminated widely.

Two (very similar) hard-ware products exist for training prosody, namely the Visipitch (Kay Elemetrics Corp.) and the PM Analyzer (Voice Identification, Inc.). Both devices extract fundamental frequency (F0) and intensity in real time and display them on a computer monitor. Both permit the visual comparison of a student's response to a teacher's model. Unfortunately, neither one can play back the model and/or the response for auditory comparison by the user, and both require the teacher to be present to evaluate the student's response. And they are both expensive. In other words, these devices are helpful aids for teaching prosody, but they require extensive one-to-one interaction with a trained professional, and a majority of training programs can afford to buy only one.

At this point, the student has several choices. He may wish to produce another response, which he does by pressing and holding down the Talk key. Or, he may wish to perform another response, which he does by pressing a left-pointing arrow key the appropriate number of times (once for the previous response, etc.). Data from earlier responses are then presented on the screen, and he can again listen to either the model or the response. Finally, he may wish to move on to another utterance, which he does by pressing the New key.

As he works, all of his data are saved in a file, which is then written to disk at the session's completion. Later, the instructor will review this file, evaluating the student's progress.
calculated as weighted averages of the differences between the data values. The use of weights permits focusing of attention on important speech features, such as vowel nuclei, without corruption by consonantal gestures.

Adjusting the timing of the response is another case of a difficult task made easier by the use of reiterant speech. Consonant/vowel (CV) boundaries are relatively easy to detect from the intensity trace of repeated /ma/ syllables. These boundaries can then be aligned successively with pre-stored CV boundaries of the model. Data points are deleted or interpolated in the F0 and intensity traces, as required, to match up the timing for each segment of the utterance. Finally, a score for the accuracy of timing can be generated as a weighted sum of the adjustments required to align the response to the model. Again, weighting these sums permits the user to focus on important temporal aspects of the model without becoming distracted by irrelevant features that happen to have been mis-timed.

Since the target users are persons with little or no experience in computer use, the interface must be simple and easily learned. As described in §2 above, a session with the Interactive Prosody Trainer requires the user to listen to a spoken phrase, mimic it using reiterant speech, and then view a graphical display showing his response compared to the original. He is then able to listen to the model phrase with its original prosody or with the prosody of his own response. This is similar to the SAID procedure [1], referred to earlier, in which learners alternately mimicked the F0, intensity, or timing of model phrases. In addition, however, our device permits the user to listen to the target phrase with his or her own prosody; in direct comparison with the model. This form of feedback is crucial for successful prosodic training.

4. REFERENCES