ABSTRACT

Although electronic visual feedback for teaching suprasegmentals has been discussed in the literature in the last few years, insufficient information has been published on how to interpret visual feedback graphs, a necessary skill for selecting appropriate phonetic material to be used with visual feedback and for explaining the visual display to students. The purpose of this paper is to identify some of the problems in interpreting visual feedback displays of suprasegmentals and to discuss some strategies for using this type of feedback on suprasegmentals more effectively. The problems discussed are interpreting the pitch display, identifying syllables, determining syllable boundaries and syllable duration, and interpreting pitch and intensity patterns when the parameters of duration, pitch, and display space are varied. An understanding of such problems should help the teacher to become more effective in selecting utterances to be used with visual feedback so that students will be able to identify suprasegmental patterns and learn them with the greatest ease possible.

KEY WORDS

Teaching suprasegmentals; electronic visual feedback; computer assisted instruction; acoustic phonetics; interpreting pitch and intensity displays; intonation contours, syllable duration; stress and rhythm.
In current language teaching methodology, suprasegmentals (stress, rhythm, and intonation) are given very high priority in the pronunciation curriculum (Gilbert 1984, 1987; Pennington and Richards 1986; Wong 1987; McNerney and Mendelson 1987; Chun 1988; Dickerson 1989). This is because suprasegmentals provide the "backbone" of utterances (Cruttenden 1986); they highlight the information speakers regard as important while at the same time revealing their feelings (Bolinger 1986), and they are important in communicating discourse meaning (Brazil, Coulthard, and Johns 1980).

It has been found that suprasegmentals can be most effectively taught through the use of equipment which extracts pitch and intensity from the speech signal and presents the information on a video screen in real time, providing instantaneous visual feedback on stress, rhythm, and intonation. A dual display allows the native speaker target to be presented on the upper half of the screen and the learner's attempts at replicating the target on the lower half. The effectiveness of such equipment has been demonstrated experimentally (James 1976; deBot 1983). It has been shown that visual feedback combined with the auditory feedback available to normal-hearing individuals through the conduction of sound through air and bone is more effective than auditory feedback alone.

Recently, hardware and software for microcomputers have been developed so that visual feedback as a tool for teaching suprasegmentals is now more accessible to language teachers. Chun (1989) discusses some of the ways in which such hardware and software can be used to teach suprasegmentals. She also discusses the equipment and software in light of their functions, their manageability (or 'user friendliness"), and their cost, and she makes recommendations for the development of courseware for teaching suprasegmentals.

Electronic visual feedback as an aid for teaching suprasegmentals is already being used in English as a Second Language (ESL) programs throughout the United States, and the ways in which it being used have been discussed at professional conferences and in scholarly publications in the past few years (Lane, Mitchell, Molholt, Pennington, Perdreau, Cessaris, and Fisher 1988; Molholt 1988; Pennington 1989; Anderson-Hsieh 1990, 1991, 1992). The kinds of English suprasegmental problems that are addressed and remediated though visual feedback are incorrect accenting of syllables, inappropriate timing of syllables, inappropriate direction, slope, and height of pitch movements, insufficient linking of words in connected speech, and vowel epenthesis (insertion) errors, which are disruptive to normal English rhythm.
Chun (1989) has shown how electronic visual feedback can also be used to teach suprasegmentals to learners of other languages. Through electronic visual feedback, native speakers of American English practicing Mandarin tone are able to compare their pitch patterns to native speaker models on the display and they can thus easily correct their tendency to use, for example, falling intonation for level tone. Also when native speakers of American English use electronic visual feedback to practice French intonation, they can very easily see, for example, when practicing information questions, their tendency to use the English pattern instead of the French pattern, in which intonation falls steadily as the utterance is spoken.

But to teach suprasegmentals most effectively through electronic visual feedback, the teacher must use phonetic material that clearly and unambiguously illustrates the patterns being taught. This is not always easily achieved, due to the fact that the relationship between the acoustic signal and the ways in which it is perceived auditorily is not a simple one. Spaaij and Hermes (1992) have addressed this problem by developing a visual intonation display system — the Intonation Meter — that more closely represents the way in which intonation is perceived. Instead of displaying the unaltered pitch contour with all of its interruptions due to voiceless sounds, the Intonation Meter fills in the interruptions with dotted lines so that the signal is continuous. It also marks vowel onsets, the points at which pitch movement is perceived in Dutch. Because this better matches the way in which intonation is perceived, it is felt that students should have an easier time learning the intonation. However, most of the equipment on the market does not have these capabilities, and teachers must rely on their knowledge of acoustic phonetics to explain the display.

The purpose of this article is to provide information that should be helpful to users in interpreting visual feedback on suprasegmentals and to suggest some strategies for more effective use of the visual display. The problems to be discussed are interpreting the pitch display, identifying syllables, determining syllable boundaries and syllable duration, and Interpreting patterns of pitch and intensity when the parameters of duration, pitch range, and display space are varied. The equipment used to illustrate pitch and intensity graphs is the Visi-Pitch. A split screen display was used for all of the visual feedback graphs presented in this article.
INTERPRETING THE PITCH DISPLAY

The pitch display is usually used alone for teaching tone and intonation. The phonemic aspects of tone that are taught through the visual display are tone height and direction, although the slope and duration of the pitch are also addressed. In teaching intonation contours, the teacher is also concerned with the height, direction, slope, and duration of the contour, but the contours being taught often involve whole sentences rather than isolated words as is the case most often when teaching tone. Of particular importance in teaching English intonation are the accents (or nuclear tones) and the direction of pitch at syntactic boundaries and the ends of utterances.

For the pitch function to operate, voicing, or vocal cord vibration must be present. Therefore, in order to teach English intonation through visual feedback, it is important for the teacher to understand how voicing is used in English. In English all of the vowels, liquids, and nasals are normally voiced. The obstruent class of consonants, on the other hand, consists mostly of pairs of consonants which are identical with each other except for the feature of voicing, one sound in each pair being voiceless and the other voiced. This is true for the stops, the affricates, and all of the fricatives except /h/, which does not have a corresponding voiced phoneme. It is also important to note the tendency in English for voiced obstruents to be partially devoiced at the beginnings and ends of words and for liquids and glides to be partially devoiced when they occur after voiceless obstruents.

Because only voiced sounds are represented in pitch contours, the quality of the contours is dependent on the proportion of voiced sounds that appear in the utterance being displayed. This is illustrated in Figure 1 in which the sentence "Mary will win all nine awards," displayed in the upper half of the graph, contains only voiced sounds, while the utterance displayed in the lower half of the graph "Jack sat at Pat's desk" contains a preponderance of voiceless sounds.

It can be easily seen that the quality of the intonation contour displayed in the upper half of the graph is superior to that of the contour displayed in the lower half. The difference in quality between the two contours is especially noticeable in utterance-final position. In both utterances, the final syllable is accented, but in the upper half of the graph the final accent is realized as a long fall while in the lower half of the graph, the final accent is realized as a truncated fall. This difference in the intonation contours occurs not because it was the speaker's intention to convey a different intonational
meaning, but rather because the word "awards" in the upper half of the graph ends in voiced consonants which carry the pitch movement, while the word "desk" ends in two voiceless consonants, which cut off the pitch movement.

"Mary will win all nine awards."

"Jack sat at Pat’s desk."

Figure 1. Pitch display

"I’ll be in my office at nine."

"I’ll be in my office at six."

Figure 2. Pitch display
This difference in the effect of the final consonants on the intonation contour is even more dramatically displayed in the two sentences illustrated in the visual feedback graph in Figure 2. In the upper half of the graph, the word ‘nine” is realized as a long fall, but in the lower half of the graph, the word 'six” is truncated at such a high level that it does not even reach the baseline of intonation.

In English, deep falls and truncated falls convey different meanings. According to Bolinger (1986), a deep fall conveys finality or assertiveness, and the deeper the fall, the more final or assertive the meaning; truncated falls, on the other hand, convey a sense of offhandedness or tentativeness. Although native speakers of English listening to utterances with truncated falls resulting from voiceless sounds will tend to fill in the missing intonational information automatically and will give the utterance the same intonational reading as they would a sentence with a long fall, when teaching assertive versus tentative intonation, the teacher should use phonetic material that unambiguously illustrates what is being taught. Clearly the teacher should use words ending in voiced consonants so that s/he will not have to explain whether any truncated falls were intended or whether they occurred because of final voiceless consonants.

In teaching intonation to learners of languages such as French and German, it is also important to select utterances containing a preponderance of voiced sounds so that intonational shapes can be clearly represented. An example of a native speaker of English learning the French continuation contour illustrates this point. The intonational shapes for continuation differ between French and English, French using a rising concave contour and English using a falling inverted S-contour (Delattre, 1963). Since a tendency exists for the learners to inappropriately transfer the inverted S-shape to French, words with voiced sounds should be selected so that the pitch shapes can be clearly represented and any errors therefore more easily corrected.

The contours that the students see on the screen serve as templates that will aid them in learning to produce and recognize patterns of intonation; they should therefore be as clear and easily interpretable as possible. This is not to say that sentences with voiceless sounds should not be used in teaching intonation. Since they occur frequently in utterances in normal speech, at some point the student should be introduced to utterances that contain them, and the concepts of voicing and voicelessness should be explained. However, when first introducing intonation contours and illustrating native speaker models of intonation on the visual display, the preferred material to be presented should contain a predominance of voiced segments, particularly in accented syllables.
Another factor affecting intonation contours is "declination" or "down-drift," a tendency for the baseline of intonation to be higher at the beginning of an utterance than it is at the end. This tendency can be seen in the intonation contour illustrated in the upper half of Figure 1, in which the baseline tilts slightly downward as the utterance progresses. However, because this process is considered to be universal (Bolinger 1986) across the languages of the world, the teacher should not be concerned about it. If the student should notice it, a short explanation should suffice.

IDENTIFYING SYLLABLES

The teacher may want to have his/her students identify syllables for the purpose of sensitizing them to the stress and rhythm of the language or to help them correct any tendency they have to delete syllables or create extra ones through the process of vowel epenthesis. The syllable, which is the basic unit of stress and rhythm, obligatorily consists of a peak — almost always a vowel — and it may also consist of an onset and/or coda, each consisting of one or more consonants. This is illustrated in the syllable structure of the English word "blend" which appears in Figure 3.

![Figure 3. Structure of the syllable](image)
The sonority theory of the ideal syllable (Hooper 1976) holds that the segments that are the most sonorous in terms of the amount of acoustic energy they expend are found in the peak of the syllable - it is the vowels that most often meet this criterion - and the least sonorous sounds - the consonants - occur near the borders. Further, the consonants in such ideal syllables tend to be ordered according to sonority with the most sonorous consonants occurring near the peak and the least sonorous ones occurring next to the syllable boundaries. The order of sound classes according to their sonority, in increasing order of sonority, is obstruents > nasals > liquids > glides > vowels. Also, within the class of obstruents, stops are less sonorous than fricatives, and the voiceless obstruents are less sonorous than the voiced ones. The example of the syllable presented in Figure 3 conforms to this ideal syllable structure. The outermost segments (the /b/ at the beginning of the onset and the /d/ at the end of the coda) are obstruents, which are the least sonorous of sounds, and the second consonants from the syllable boundary (the /l/ in the onset and the /n/ in the coda) are more sonorous than the stops. At the peak of the syllable is the vowel [c], which has the highest sonority value of all of the segments in the syllable.

It is such ideal syllables that are the most easily identified when using visual feedback because each such syllable is usually represented by one major peak on the visual display. This is illustrated in Figure 4 in which the syllables in the two sentences in the visual feedback graph are easily identified because each syllable is represented by one intensity peak. All of the syllables in both sentences have a very simple structure (C V or C V C), and the intensity of all of the consonants is low enough to clearly separate the syllable peaks (the vowels) from each other on the display.
However, in some syllables that conform to the sonority principle, separate small peaks of acoustic energy can appear depending on the sounds that constitute the syllable. The word "pick" displayed in the graph in Figure 5 shows a small peak at the end of the word, which represents the release of the final voiceless consonant /k/. The stretch of baseline between that small peak and the rest of the syllable represents the closure portion of the stop. It is important to note here that the intensity of speech sounds is variable, being related to the force of the air stream, and that a strongly aspirated voiceless stop would probably be realized as a higher peak than the one illustrated in Figure 5. On the other hand, a weakly aspirated stop might be barely visible on the display.

![Figure 5. Intensity and pitch display](image)

In words that violate the principle of sonority, such as "church," separate peaks of intensity may occur that are large enough to be mistaken for separate syllables. This is illustrated in Figure 6. The word "church" violates the sonority principle because the [∫] portion of the affricate [t∫] which occurs at the end of the syllable is higher in intensity than the stop portion [t] which occurs closer to the syllable peak.

![Figure 6. Intensity and pitch display](image)
Another type of syllable identification problem occurs when there are more syllables in an utterance than there are discernible peaks of intensity on the visual feedback graph. Such a phenomenon can occur in words which contain consonants whose intensity or sonority is fairly close to that of vowels. The glides and the liquids are sounds with such intensity. An example of a visual feedback graph illustrating the word "narrower," which has a liquid at the first syllable boundary and a glide at the second word boundary, appears in Figure 7. Although the word 'narrower' has three syllables, the graph shows only one somewhat blunted peak with no discernible depressions to indicate where the consonant boundaries are located. (In such cases, a sound spectrograph would be needed to identify boundaries.)

In the early stages of instruction, the teacher should exploit the ways in which the intensity display matches the teachers' and students' perceptions of syllables as discernibly separate peaks of acoustic energy (in the same way that the teacher might appeal to the students' auditory processing mechanism by tapping on the desk the number of syllables in an utterance). Thus, utterances such as those illustrated in Figure 4, in which each syllable is represented by only one intensity peak, should be used to illustrate syllables and words such as those presented in Figures 6 and 7, which do not show such a one-to-one relationship, should be avoided. However, after the student has had some experience identifying syllables, the teacher can then introduce words whose syllables cannot be so clearly identified. Since such words abound in the stream of speech, they will have to be dealt with at some point, but to present such words at the beginning of instruction may cause the student unnecessary difficulty in identification.
IDENTIFYING SYLLABLE BOUNDARIES

In the previous section, the problems in identifying the number of syllables in utterances were discussed without any mention of syllable boundaries. However, at times it is necessary to determine where one syllable ends and another begins. Such information is important, for example, when teaching English stress, because one of the acoustic features associated with stressed syllables in English is increased syllable duration. Therefore, the teacher may want to compare, for example, a learner’s successive attempts at lengthening a stressed syllable or reducing an unstressed one.

In making such measurements, the first consideration is to select words which show syllables as easily discernible peaks of energy and to avoid words in which syllable peaks cannot be readily discerned, as discussed in the previous section. However, even when separate peaks can be discerned, the intervening “valleys” may not provide sufficient clues to determine exactly where the boundaries lie. One such case occurs when two consonants which occur together at a syllable boundary and which belong to different syllables have nearly the same sonority values. This problem is illustrated in the sentence “Jack sat at Pat’s desk” presented in the visual feedback graph in Figure 8. Although the boundary between the fourth and fifth syllables clearly occurs between the final/s/in “Pat’s” and the initial /d/ in ”desk,” it is impossible to determine on the visual display where the /s/ ends and the /d/ begins. The intensity graph does not offer any clue because the/t/ and the/s/ from the word “Pat’s” and the closure portion of the /d/ are all represented by the same stretch of baseline between the fourth and fifth syllables. Nor can the syllable boundary be determined from the pitch display. Normally the voicing of the /d/ in the word ”desk” begins during the closure phase of the stop, but in the utterance displayed in Figure 8, voicing is completely absent during this stage probably due to the assimilation of the /d/ to the voicelessness of the previous /s/.

Another problem in determining syllable boundaries is related to ambisyllabicity. There are many words containing consonants whose syllable membership is divided between two syllables. Some examples of such consonants are the /n/ in “honest” and the /t/ in petrol, whose closure belongs to the first syllable and whose release belongs to the second (Hogg and McCully 1987). However, the teacher might feel that such distinctions are too fine for his/her purposes and may elect to measure, for example, only the portion of the second syllable in ”petrol” beginning with the release of the /t/.
With careful selection of words for illustrating syllable duration, the problems described above can be avoided and the teacher can use electronic visual feedback more effectively for teaching stress and rhythm. Examples of words in which syllable boundaries can be clearly identified on the visual display are "replace," "defer," and "martial." Since the second syllable in each of these words begins with a voiceless consonant, the pitch display can be used in conjunction with the intensity display to indicate where the second syllable begins. However, although the pitch and intensity displays are suitable for making duration judgments for pedagogical purposes, it is important to note that in research on syllable duration, greater precision in measurement is required which only the sound spectrograph and waveform analysis can provide.

**FACTORS AFFECTING SYLLABLE DURATION**

When making comparisons of syllable duration, it is important for the teacher to understand the factors that affect it. Clark and Yallop (1990) have observed that although syllables are very elastic — their durations being strongly affected by whether or not they are stressed — other factors besides stress are also related to syllable duration.

One of the most obvious factors that influences syllable duration is speech tempo — the faster the speaking rate, the shorter the duration of syllables in the utterance. Another factor that influences syllable duration is the syllable’s inherent segmental composition. Not all sound segments are of equal duration. Vowels are generally longer in duration than consonants, and among vowels, some are longer than others. Tense vowels are slightly longer than lax vowels, diphthongs are longer than pure vowels (Fry 1979) and
full vowels are longer than reduced vowels even when they are not stressed (Bolinger 1986). Among consonants, nasals tend to take up more syllable space than other consonants (Fry 1979).

Another rather obvious factor contributing to the length of syllables is the number of segments which constitute them. For example, although the first syllable in the word "practice" is longer than the second syllable, it cannot be unambiguously said that it is longer as a result of being stressed because of the fact that it consists of more segments than the unstressed syllable. In addition to the effects of sound segments on syllable duration, certain lexical items can have a longer inherent duration than other lexical items with exactly the same phonological form. In the sentence "take the pine away, and she'll just pine away," the syllable "pine" is longer in the second clause than it is in the first, and this is due to a tendency in this word to sound symbolism (Bolinger 1986,43).

But probably the most important factor next to stress affecting syllable duration is the position of a syllable in an utterance. Syllables that occur before pauses tend to be longer than they are in other positions whether they are stressed or not (Clark and Yallop 1990). This is illustrated in the word "dancing," which appears in a different position in each of the two sentences presented in the graph in Figure 9. In the upper half of the display, it appears at the beginning of the sentence and in the lower half of the graph, it appears at the end of the sentence. When "dancing" occurs in sentence-final position, the second syllable is approximately one and one half times as long as it is when "dancing" occurs at the beginning of the sentence in spite of the fact that the syllable is not stressed in either position.

The above factors related to syllable duration have implications for using electronic visual feedback to teach stress and rhythm. First, when comparing the learner's syllable durations with those of a native speaker on the other half of the video display whose speaking rate is appreciably different from that of the normative speaker, relative rather than absolute comparisons should be made. These relative comparisons can be made impressionistically by simply estimating the proportion of a word taken up by the syllable and comparing it with the same proportion in the other utterance. Or if the teacher wishes to be more precise, s/he can compute a duration ratio for each speaker.

When comparing several successive attempts by the same learner to shorten or lengthen a particular syllable, the same principle of relativity applies. If the speaking rate is appreciably different from one attempt to the next although it is unlikely that this
would occur — the comparisons should be proportional. Also, the position of the syllable under consideration should be controlled from one attempt to the next. If the syllable being compared occurs in word final position in the first utterance, it should also be word-final in the learner's ensuing attempts.

Finally, when comparing the duration of stressed syllables to unstressed syllables to demonstrate the effect of stress on syllable duration, the most valid comparison that can be made is one between the same syllable in the same word under the condition of being stressed and under the condition of being unstressed. For example, the first syllable in “retráin,” which is unstressed, could be compared with the same syllable in the same word when it is stressed as in “I said ‘rétrain’ the man, not ‘prétrain’ him.” Although the examples of syllable-related problems discussed above were taken from English, some of the same principles can also be applied when analyzing syllables and timing phenomena in other languages.
Varying Parameters

Certain parameters can be varied on the equipment that is available for visual feedback. On the Visi-Pitch, for example, the parameters that can be varied are pitch range (from 0-400 Hz to 0-1200 Hz), duration (from 0.25 to 15.0 seconds), and display space (full screen single display to split screen dual display). It is important for the teacher to keep in mind that whatever is being displayed on the screen is relative to the particular time, duration, or display space that has been selected. For example, pitch contours displayed in the 0-400 range will show better resolution than those displayed in the 0-800 range and the peaks of intonation will be steeper. It is especially important to keep this in mind when using the dual display for speakers requiring different frequency ranges (the range will generally be lower for men than for women, and lower for women than for children) and when making comparisons of pitch range.

CONCLUSION

The problems related to the interpretation of electronic visual feedback and some suggestions on how to use the feedback more effectively in teaching are discussed in this article. Since one of the major advantages of electronic visual feedback is to illustrate certain generalizations about English suprasegmentals, care needs to be taken in selecting utterances for display that will clearly illustrate those generalizations. For exercises involving the identification of syllables, the principle of “one syllable — one major intensity peak” should be observed as much as possible, especially in the early stages of instruction. Also, when introducing intonation contours to the student, phonetic material should be selected that will provide relatively uninterrupted intonation contours, especially in accented syllables. It is important for the teacher to keep in mind that what is useful for displaying intonation is not always good for displaying syllables. While uninterrupted contours are preferable for illustrating intonation contours, the presence of voiceless segments in utterances can be helpful in identifying syllable boundaries. The principle that should guide the teacher in selecting material is to use his/her knowledge of phonetics and suprasegmentals to display the patterns s/he wishes to teach in the clearest way possible, so that the student can learn suprasegmentals with greater ease.
The teacher should also keep in mind the variation that occurs in speech, particularly in duration and intensity. It is advisable for the teacher to experiment with the material selected for practice with visual feedback well in advance of teaching so that s/he can identify any problems with the visual representations of speech.

This paper was concerned only with the interpretation of visual feedback and some strategies for selecting material that will result in the most effective kinds of displays. Not included in the discussion above were the procedures for teaching suprasegmentals through visual feedback or the processes through which suprasegmentals are learned through visual feedback. There is still a great need for more work in these areas as well as for the development of courseware, as observed previously by Chun (1989).

REFERENCES


AUTHOR'S BIODATA

Janet Anderson-Hsieh, Ph.D. is an Associate Professor in the English Department at Iowa State University. She teaches in the Linguistics, TESL, ESL and International Teaching Assistant Training Programs. Her research interests include interlanguage phonology, the acquisition suprasegmentals by second language learners, pronunciation teaching methodology, and native speaker reactions to foreign accent. Her recent articles have appeared in *English for Specific Purposes, System, Language Learning*, the *JALT Journal*, and *IDEAL*.

AUTHOR'S ADDRESS

Department of English
Iowa State University
203 Ross Hall
Ames, Iowa 50011-1201

Phone: (515) 294-3139 203
Fax: (515) 294-6814
E-mail: jandersn@iastate.edu