# TEACHING PRONUNCIATION FROM THE TOP DOWN

MARTHA C. PENNINGTON
*University of Hawai'i at Manoa*

## INTRODUCTION

In traditional educational curricula, instruction in pronunciation was often a key aspect of a foreign language course. Today, however, pronunciation tends to be de-emphasized in language courses, and explicit instruction in this aspect of language is considered by many to represent outmoded educational practice. Advocates of modern comprehension-based or communicatively oriented language curricula generally take the view that pronunciation should not be taught explicitly but should rather be allowed to develop naturally as a by-product of attempts by students to communicate.

In fact, however, there has been little systematic research to discover if certain teaching practices or conditions of training can have positive effects on pronunciation. Hence, we have no firm basis for asserting categorically that pronunciation is not teachable or that it is not worth spending time on pronunciation training in a language course. New ways of teaching pronunciation are still to be developed and tested, and there is a need for careful experimentation to determine the effectiveness of methods, both for learners in general and for learners characterized in terms of specific traits (e.g., visual orientation or field dependence). There is a great need for carefully controlled studies of the effects of pronunciation training of different types, as stressed in Pennington & Richards (1986: 221):

> In particular, the kinds of information needed are (a) clear specifications of the aspects of pronunciation being taught, (b) precise descriptions of the instructional procedures used, and (c) valid measures of the effects, positive or negative, of the procedures used. Stricter attention to the question of research design is necessary before the results of particular instructional programs and teaching procedures can be evaluated. With more complete information of this kind, it will be easier to determine not only the relative value of teaching pronunciation as opposed to other aspects of language proficiency but also the amount of attention which should be paid to pronunciation within the context of a whole language teaching program.

In this paper, a theoretical and pedagogical foundation for such research efforts is provided. Pronunciation is examined from a contextual, "top-down" perspective from which segmental articulation assumes less importance than more general properties of speech such as rhythm and voice quality. Pronunciation is described as conveying many different types of messages to a hearer related to the information structure of a discourse, the speaker's attitude and mood, and other social and psychological features of the speaker or of the relationship between the speaker and hearer. Moreover, various aspects of pronunciation are shown to relate to specific gestures.

The aim is to present a more descriptively enlightening and pedagogically useful characterization of second language phonology than traditional treatments, in which phonology was identified with discrete articulations and in which suprasegmental features were relegated to the periphery of language per se, i.e., to the **paralinguistic** and in some cases the **extralinguistic** domains of communication. Suggestions for teaching pronunciation are set in a context of research and theory, and a focus on the non-segmental characteristics of speech is advocated. This discussion makes reference to the use of video and computer media in pronunciation training (see Pennington forthcoming (a) for further discussion), as well as to the use of more traditional types of audiovisual aids. The paper concludes with a set of research questions on pronunciation instruction derived from this investigation.

## COMPONENTS OF PRONUNCIATION

In traditional language teaching and in some modern approaches, pronunciation has been primarily identified with accurate production of phonemes. Perhaps because of their association with graphemes, phonemes have tended to be thought of as building blocks for constructing words segment by segment. In this same tradition, words were viewed as the building blocks for phrases, phrases for sentences, and sentences for discourses. While it may be relevant to certain aspects of linguistic performance, this linear, "bottom-up" conception of how people create language is limited in its explanatory power and its relevance for language teaching. A holistic, "top-down" conception of both perception and

production, now recognized as essential for explaining linguistic performance, also underlies many important developments in modern language pedagogy.

According to this "top-down" conception, language is produced and interpreted with reference to a larger meaningful context. In listening, language users employ pragmatic, semantic, syntactic and phonological context to interpret the full import of messages conveyed by a speaker (Chaudron 1983, Dirven & Oakeshott-Taylor 1983). Likewise, speakers construct messages not item by item, but rather with a view to the larger pragmatic, semantic, syntactic and phonological context (Brown & Yule 1983). From this perspective, pronunciation—far from being a static, building-block phenomenon—is a dynamic process in which many types of contextual elements interact to produce effects in the communication process (Pennington & Richards 1986).

When viewed as the phonological aspect of speech in real communication, pronunciation is correctly seen as quite essentially a non-segmental, non-discrete and non-autonomous phenomenon. The pronunciation of individual phonemes is greatly influenced by such factors as speaking style (Beebe 1980, Schmidt 1977), rhythm (Adams 1979, Gilbert 1984) and speaking rate (Fowler 1981, Hieke 1984). Moreover, the development of individual phonemes in second language acquisition is closely tied to the development of a variety of properties affecting stretches of speech (Pennington forthcoming (b)), referred to as **suprasegmental** properties or features. For purposes of exposition, the suprasegmental category is here broken down into a three-way subclassification scheme which includes **prosody, phonological fluency,** and **voice quality**.[1] In a "top-down"

---

[1] As is typical of other taxonomies of linguistic units, the three- way distinction of prosody, phonological fluency and voice quality is a descriptively convenient idealization. In reality, these three aspects of suprasegmental phenomena do not necessarily (depending on how exactly they are defined) represent mutually exclusive properties. In some descriptive traditions, for example, **prosodic** is synonymous with **suprasegmental**, so that the domain of **prosody** includes the speech characteristics here termed **phonological fluency** and **voice quality**. In the present treatment, **prosody** is taken in its more literal (and literary) sense and restricted to "metrics", i.e. to the linguistic categories of stress, rhythm, pitch and intonation. **Phonological fluency**, while certainly related to stress and intonation, is the manifestation of the conceptually and physiologically distinct processes of weakening and coarticulation which cause the borders of syllables, words, phrases and sentences to coalesce in running speech. **Voice quality**, while also a property of connected

approach, the relevant domain of phonology can also be usefully expanded to include the **gestures** that regularly accompany specific aspects of pronunciation (see, for example, Bolinger 1983, 1985, 1986).

In order to become a competent speaker and listener, a language learner needs to attend to not only the strictly mechanical, articulatory aspects of pronunciation, but also to the meaningful correlates of those articulatory features in the immediate linguistic context, as well as in the larger context of human communication. For pronunciation is one aspect of an interlocking meaning complex of language-specific conventions for interactional and transactional speech. These conventions relate to the presentation of one's knowledge and attitudes towards a variety of features of the communicational situation—e.g., the speaker's relationship towards the topic of discourse and to individuals in the speech situation as well as in the larger speech community. By following such conventions, speakers structure the information in a discourse and provide minute detail for its interpretation by others in the speech community.

Thus, it would seem reasonable in a language class to provide instruction and practice in the actual mechanics of English phonology as well as in producing the appropriate pronunciation in specific types of meaningful contexts. In the next sections, a foundation for such instruction and practice is provided through an examination of the phonological categories of prosody, phonological fluency, voice quality and the co-occurring gestures. The examination begins in each major section with a brief description of the suprasegmental property under consideration. This description forms the starting point for a discussion of types of meaning associated with that aspect of phonology and an examination of relevant teaching approaches.

## PROSODY

speech, is a generalized, non-local characteristic of a language or speech variety derived from its underlying phonological basis. This phonological basis comprises the general features of articulation -- e.g., tongue fronting, lip rounding, nasalization -- which individual phonemes share. Following Laver (1980) and Esling and Wong (1983), **voice quality** derives from articulatory settings in any part of the vocal tract, not merely at the glottis, and is not restricted to the expression of paralinguistic and extralinguistic meaning. In fact, the present discussion, following the lead of other phonologists (e.g., Ladd 1980, Bolinger 1986), does not make use of the traditional, and ill-defined distinction, between **paralinguistic, extralinguistic, and linguistic types of meaning.**

### Description

**Prosody** is defined with reference to the patterns in individual words of **stress** (e.g., English), **pitch**, (e.g., Japanese), and **tone** (e.g., Chinese), as well as the **rhythmic and intonational patterns** of longer utterances. Prosodic patterns on both of these levels are an important aspect of pronunciation to master for nativelike speech in a second or foreign language.

Stress is the amount of energy expended in producing a syllable. For the hearer, stress is manifested as perceptual prominence or strength. All languages operate on the general principle that utterances are timed on the basis of (an alternation of the weak and) the strong syllables, though the manifestation of strength and weakness varies from language to language. Languages exploit different combinations of parameters such as duration, intensity and pitch, to achieve strength, or prominence. Moreover, languages vary in the unit of duration that is relevant to achieving their characteristic rhythmic pattern. For "stress-timed" languages such as English and Mandarin Chinese, the relevant unit is the **stress foot**, which is one stressed syllable and its adjacent unstressed syllables (if any); in "syllable-timed" languages such as Spanish and Cantonese, it is the syllable; and in "mora-timed" languages such as Japanese and Estonian, it is the mora, or segment (Hoequist 1983).

Intonation is the pattern of pitch changes that occurs over grammatical units, the sentence being the basic category of reference. A **tone group** is that part of a sentence over which a particular intonation pattern extends. A tone group contains one prominent syllable, the **tonic syllable.** General intonational patterns are similar from language to language (Abe 1955, Bolinger 1978, Ohala 1983), but the details show considerable variation from one language to another (Delattre 1963). Categories of pitch change derived from research into Dutch intonation, based on work by 't Hart & Collier (1975), indicate the following as relevant for perception:

a) **direction of pitch change** (rise, fall, or level);
b) **range of pitch change** (difference between high and low levels);
c) **speed of pitch change** (how abruptly or gradually the change         happens;
d) **place of pitch change** (in sentence, word, or syllable)

<div align="right">de Bot & Mailfert (1982:72)</div>

Collier (1984) maintains these basic categories of pitch movements, while adding a parameter for the characteristic declination of pitch from the

beginning to the end of an utterance. Such categories of pitch movement can be considered to represent the parameters of intonation which are relevant for describing intonational differences within and between languages.

### Types of Meaning

In English, stress differences are relevant at the level of individual syllables and words, and sometimes differentiate pairs of grammatically related words as in:

SUBject   subJECT
OBject    obJECT

There are few pairs in English of the type <u>differ/defer,</u> that is, grammatically (and semantically) unrelated words which differ only phonologically in stress placement. In other languages, it is common for unrelated words to differ only in their non-segmental (suprasegmental) characteristics. In Japanese, Chinese, and many African languages, for example, lexical items are differentiated solely by pitch pattern.

Pennington & Richards (1986:211) summarize the kinds of meaning which may be associated with intonation contours:

> In every language, characteristic intonation contours carry both referential and affective meaning.... In their referential function, intonation contours provide an interpretation for a sentence by indicating which part of the information is viewed as new versus known, salient versus less salient, or topic versus comment. Intonation and stress are highly context-dependent, so that the patterns of stress and pitch that characterize isolated words or phrases are typically modified when these words or phrases occur in the context of longer utterances. For example, pitch level tends to be reduced in later parts of a discourse as predictability of information increases. Thus, intonation is an essential component of the 'prosodic continuity' that makes connected stretches of speech—as opposed to individually spoken words or syllables—coherent and interpretable by the listener.

Intonation contours are associated with meanings such as finality, continuation or questioning, which are in turn associated in languages with grammatical units such as, respectively, indicative final clause, indicative non-final clause, and interrogative final clause. Moreover, pitch contours

may have specific meanings which can be identified when they co-occur with specific contextual elements. For instance, in English, there is a certain pitch contour, which, when spoken on a three-syllable vocalization of nasal sounds (the particular nasal selected does not seem to matter), mimics the intonation of the statement, "I don't know." In English as well as in Japanese, a single nasal syllable, when spoken with several different intonations, conveys meanings ranging from surprise, to pleasure, to routine acquiescence.

The meaning of a particular intonation contour may vary depending on the context or the language in which it occurs. For example, use of final high or rising intonation may be a component of "baby talk", "foreigner talk", femininity, condescension, friendliness, or continuation, depending on the culture and the circumstances of communication. Use of final low or falling intonation, in contrast, may signal "adult-talk" (e.g., information withheld from children, or other kinds of "privileged" or in-group messages), masculinity, authority or finality (Laver & Trudgill 1979, Bolinger 1983, 1985; Pennycock 1985). In both of these cases, the relationship of the speaker and hearer, in addition to other aspects of the surrounding situational and linguistic context, will be important determinants of the meaning of the contour used on a particular occasion.

The degree of markedness of a certain type of prosodic contour will vary from language to language. For example, a very regular rhythmic contour, which would be the norm for any language termed syllable-timed, would be quite marked in English and would have a special connotation such as lack of interest. As another example, a "stepping" pitch contour such as the type that is common for Japanese, is a highly marked contour in English, used for what has been called **stylized intonation**, which marks routinized speech acts or highly predictable information (Ladd 1978, 1980). Likewise, a clausal contour with progressively decreasing pitch, while relatively common in Japanese (Haraguchi 1977, Higurashi 1983), is rare in English, where clause final position is reserved for prosodically highlighted, focused information.

### Teaching Approaches

Training in prosody may be more valuable and essential than work on individual sounds, or phonemes, for achieving accurate perception and production at the segmental level. As remarked by Abberton, Parker &

Fourcin (1978:34):

> From a formal point of view, voice and the associated prosodic features of rhythm and intonation provide the essential framework within which vocal tract segmental features are temporally organized. This is a major justification for working on prosody before articulation and segmental phonology.

Prosodic exercises may have students work individually or in small groups—as in the English stress exercises of Byrd et al. (in press)—to list together and then practice words or phrases which have the same prosodic patterns. Following research by Adams (1979), a training program for rhythm is suggested that moves from highly metrical material such as poetry to non-metrical prose. Visual aids—e.g., Woods'(1979) picture of a line of marching soldiers to represent the rhythm of a French sentence or his use of graphic illustrations of linked "beads" of different sizes to represent stress and rhythm—can be helpful for training intra- and interlingual rhythmic differences. Such training in rhythm prepares the student for work aimed at achieving phonological fluency, as described below. Based on the success of the language training program developed by de Bot (1983), which used a combination of an explicit presentation of a componential analysis of intonation and visual feedback on performance, it seems promising to train students—and perhaps also language teachers—using a visually reinforced presentation of intonational components in addition to visual feedback on perception and production of intonation contours.

Computers have many uses in enhancing the analysis and the presentation of prosodic patterns. With the aid of some peripheral equipment, computers can provide accurate and almost instantaneous analysis of speech input, e.g., to compare a student's prosodic patterns (both stress and intonation) with reference patterns stored on disk. The display screen and printer also make possible the use of computer graphics for presentation and feedback on performance.

## PHONOLOGICAL FLUENCY
### Description
**Fluent speech phenomena** are those aspects of language which can be

observed under conditions of sustained oral production—i.e., "running speech"—in a natural context. Hieke (1985:139) describes the following **quantitative parameters** of fluent speech as the most important:

> (1) **speech rate** (the amount of speech produced over a period of time, usually measured in syll/sec.);
> (2) **length of runs** (the average number of syllables occurring    between pauses);
> (3) **rate of articulation** (the total phonation minus pause time);
> (4) **stalls**, including silent pauses of more than 130 msec., **filled pauses** (filled with such vocalizations as uh, ah, hm), and **progressive repeats** to "buy time" while searching the memory for a certain lexical item.

Hieke (1985:140) also mentions a number of **qualitative parameters:**

> Fluent speech is the cumulative result of dozens of different kinds of processes. These can be classified according to the severity with which they cause alterations, i.e., the degree to which portions of running speech are absorbed. **Absorption** affects speech in increments, the mildest of which may be classified into forms of **linking**. A subtype of it [is] **consonant attraction**.... There are two other major classes, the second incorporating various kinds of **levelling**, ... and the third encompassing types of **outright loss**.

These qualitative parameters are the processes that produce **phonological fluency** in English. Hieke (1984) found that while native English speakers realized 77.3% percent of all potential link points for consonant attraction in spoken discourse, native German speakers at a fairly advanced level of proficiency realized only 53.5% percent of those links in English discourse. In a similar vein, Odlin (1978) demonstrated a correlation between overall proficiency level and use of contractions in the speech of ESL learners.

Languages differ in the degree to which they admit coarticulatory effects and "contaminations" of one sound by its neighboring sounds in fluent speech. English is apparently at an extreme degree in this respect among languages. Fowler (1981) describes a **compression effect** which operates in English to maintain even foot duration—syllables are either stretched out or squeezed to accommodate to the stress- timed rhythm of the language. In many other languages, the integrity of individual sounds and word boundaries is maintained through glottal stopping, vocal tension or brief pausing.

Languages also differ in the degree to which sounds weaken under

conditions of weak stress. In English, vowels tend to lose their distinctive articulation, or **color** (Donegan 1978), under conditions of casual speech. In an instrumental analysis of fluent English speech, Delattre (1981) found that all vowels tended to centralize towards the position of the schwa vowel. Weakenings under conditions of reduced stress in other languages did not show a comparable tendency. Prosodic patterns in a language are related to and limited by the typical length of a **sense group**, or **thought group**, i.e., a group of words that completes one meaning or thought. Welkowitz et al. (1984) report evidence of language-specific preferences for length of pauses and vocalizations in Canadian and Chinese speakers. Duez (1982) found that the length and distribution of pauses varied considerably in three different speech styles in French, while Reynolds (1984) recorded a difference in type of pause-filling stalling device preferred by Japanese males and females in making formal speeches. Adams (1979) discovered that English teachers who were native speakers of one of several Asian languages paused more frequently and used longer pauses than any of the native English speakers tested under the same conditions.

### Types of Meaning

Junctural differences—e.g., in whether or not linking is present and if so, at what point and to what degree—can mark differences in referential meaning. Well-known examples are: <u>light housekeeper</u> versus <u>lighthouse keeper</u> and <u>an aim</u> versus <u>a name</u>, which can be distinguished in some speech styles. In the following example, the length of a pause, if one is present, and the concomitant differences in intonation and in articulation of the phonemes, would give a different connotation in each case:

Example 1.   a.  Oh. No!  (pause of 1-2 sec. between words)
             b.  Oh, no!  (pause of .2 -.3 sec. between words)
             c.  Oh no!    (linking between words)

In example 1a, the physical separation of the two words connotes a functional separation into two messages which may or may not be related (i.e., the utterance of <u>No!</u> could be in response to the message which the speaker signifies by <u>Oh.</u> has just been received and understood, or it could be a

reaction to something unrelated that is happening in the environment—e.g., a child about to touch a hot stove). Examples 1b and 1c each are more likely to represent a unified message. Example 1c in contrast to 1b perhaps connotes greater spontaneity or excitement; 1b may be the conventionalized emotive expression of a negative reaction.

Increased rate in English may be associated under certain conditions with authority and competence (Laver & Trudgill 1979), while in other situations it will be interpreted as disinterest or anger. Increasing the number of tone groups in an utterance through frequent pausing and careful articulation has a general connotation of emphasis. When speech rate is decreased in this way, it may indicate, in appropriate circumstances, a high degree of significance attached to a message, formality, lack of confidence, reluctance to speak, lack of enthusiasm or depression. A behavioral complex of slow speech and a low and monotonic pitch "in young school children was judged by teachers to be predictive of school failure," as reported in Robinson (1979:236).

### Teaching Approaches

Fluent speech phenomena can be instructed by highlighting contextual processes affecting stretches of speech as well as individual sounds, e.g., by providing classroom activities to practice producing and perceiving reduced pronunciation (Brown & Hilferty 1986) or the pronunciation of sounds which may be confused in context (Pica 1984). Confusions which surface only under conditions of fluent, contextualized speech can form the basis for classroom exercises. In 2 English, these might include "maximal pairs"[2] such as the following, each of which could become indistinguishable in rapid speech:

> Example 2. a. That's not my chair.
>     That's not much air.
>   b. We really don't owe them that much.
>     We really don't know him that much.

---

[2] Pennington (1986) described a test administered to 73 Japanese learners of English which included a subtest of multiple choice "maximal pair" discrimination items such as those in Example 2. Scores on this subtest correlated more highly with overall phonological proficiency, as measured by three raters based on interview data, than measures of minimal pair discrimination or listening comprehension. Hence, classroom instruction on these types of items may help to improve aural as well as oral phonological proficiency.

A useful direction for developing exercises to work on the production and decoding of phonologically fluent utterances is provided by Hieke's (1987) recovery strategies for the resolution of lexical items and boundary markers in connected speech.

General principles can be illustrated and practiced, such as that linking between two neighboring words or a succession of two equally stressed syllables occurs frequently or infrequently in the language under consideration (Gilbert 1984). The extent to which the effects of natural phonological processes affecting phonemes (Bjarkman 1986) operate within the language can also be addressed, for example, the degree to which close sounds such as stop consonants or high vowels open as a result of contextual influence or the degree to which low sounds tend to have closer variants in rapid speech, as a result of physiological constraints on rapid achievement of target positions. Common fluent-speech processes in the native language might be compared with those in the target language, e.g., through imitation by students of a phrase pronounced in a nativelike and a targetlike way.

As in other aspects of pronunciation instruction, visual display generated by computer analysis or ordinary visuals offer salient reinforcement. Woods (1979), for instance, uses pictures of linked beads of different sizes to represent stressed and unstressed English syllables. Under conditions of phonological compression as defined above, the beads are seen to compress from round to ovoid shapes. Computer-assisted peripherals such as the Kay Elemetrics Visi-Pitch can show students a visual display of the timing and coarticulatory properties of contextualized syllables, words and phrases, and allow comparison with a stored display. In the future, it may be possible, with computer-assistance, to have a running display of feedback on aspects of oral performance such as those identified by Hieke (1985) to contribute most significantly to oral fluency.

## VOICE QUALITY
### Description
A certain **voice quality** is achieved by means of mechanical settings of

the tongue, the jaw, the lips and the vocal cords to achieve characteristic modes of articulation, lip shapes and vocal pitch. These settings vary from language to language, so that one language might have a relatively compressed intensity range as compared to another—e.g., French as compared to English (Delattre 1963)—or a comparatively slight degree of opening of the jaw or lips—e.g., Japanese as compared to English (Kaneko 1957).

### Types of Meaning

Voice quality differences in languages may signal differences in language functions, affect and speaker roles (Laver & Trudgill 1979, Esling & Wong 1983). Voice quality differences within a language help to indicate to a language learner the range of variability—e.g., in pitch or the articulation of individual phonemes—which is considered acceptable among native speakers.

Two languages may share a voice-setting feature, and yet in one of those languages the feature may be relatively common while in the other it may be relatively marked. While labialization is an unmarked feature of such languages as German, French and Scandinavian languages, in English it tends to be associated with "baby talk" (Pennycock 1985). Similarly, laryngealization, or creaky voice, which is typical of the speech of the upper class British speaker (Robinson 1979) or the traditional Japanese male speaker, is used more typically in American English for certain marked functions, such as commiseration or sarcasm.

A certain voice quality may be associated with different meanings; often one meaning is relatively positive while the other is less positive. For example, in many cultures, a softened voice and relatively high pitch will be interpreted as deferential and therefore as polite. In other cultures, these features are associated with timidity or exaggerated femininity. In this and in other cases, misunderstandings might arise if the student simply transfers the voice-set from the native culture into the target culture. Sometimes, the addition of another voice setting feature will disambiguate a certain voice quality. For example, a harsh phonatory setting, when added to high pitch overall, will indicate anger less ambiguously than high pitch alone (Laver & Trudgill 1979).

Voice quality can be manipulated by the speaker to achieve a variety of intentions or purposes. For example, a deep, emphatic voice may be used by a

man or a woman to signify either threat or authority. In each culture, certain voice qualities are associated with certain social identities or roles. Royal's (1985) work on Cairo Arabic, for example, showed an association in a westernized, affluent neighborhood of fronter articulation with female speech and of backer articulation (exhibiting, for example, a high degree of pharyngealization) with male speech. Trudgill (1974) also noted differences in voice quality for male and female speakers and for speakers of differing socioeconomic status in Britain.

### Teaching Approaches

Teaching approaches for working on voice quality might include recognition and imitative practice based on tape recordings of different accents or speech styles. Students might be asked to identify or to imitate different types of speakers—e.g., farmer, teacher, actress, businessman—using voice quality. Inferencing exercises in which students must correctly identify situational properties of utterances based on voice quality and other suprasegmental properties can also be devised. Two or more different tokens of the same sentence can be played, each with a different voice quality—e.g., one high-pitched and rapid and the other low-pitched and slow. The students must then reconstruct the general context in which such an utterance might occur. The same sort of exercise can be done at a more advanced level by recording conversations on the basis of which students draw inferences about the exact purpose of the communication and the exact relationship between the participants.

**Silent dictation,** that is, dictation in which words are mouthed but never actually spoken, helps student to focus attention on the visible settings and movements which produce the sounds of a language. A mirror can be a valuable addition to a pronunciation lesson, as it allows students to see the movements of jaw and lips that are occurring as words are pronounced. Video can also be used to help students visualize the movements of the jaw and lip shapes that accompany the pronunciation of individual sounds or running speech in the target language (Ecklund & Wiese 1981) and to compare and contrast these with the jaw movements and lip shapes in the native language. Eventually, we may see computer-assisted graphics which illustrate the continuous movements associated with running speech and which perhaps also supply on-line feedback to correct lip shapes and jaw

movements as speech is input via a microphone.

## GESTURES

### Description

**Gestures** that accompany speech are part of the meaningful system that has been referred to as **kinesics**. There are significant cultural differences in gestural complexes and their application during interaction, as stressed by Birdwhistell (1970). While there is a substantial amount of variation in nonverbal communication from culture to culture, there are also some similarities; moreover, there appear to be some universal tendencies (Ekman et al. 1969, Ekman & Friesen 1974, Eibel-Eibesfeldt 1974) which are realized to a greater or lesser extent from culture to culture.

Analogous to the theory of natural phonology formulated by Stampe (1972), we might posit a natural basis for gestures. According to the theory of natural phonology, all children are born with the same universal basis for sound production. Differences in the linguistic repertoires of children speaking different languages are the result of reinforcement or inhibition of certain innate tendencies, e.g., towards final devoicing of consonants. Similarly, research suggests that children are born with similar expressive potentials as regards facial expression and other gestures, but that different cultures encourage or suppress these gestures in social interaction (Ekman et al. 1969, Ekman & Freisen 1974). Thus, a speaker must learn to inhibit certain gestures and to enhance certain others in order to develop native or native-like competence in the use of culturally correct gestures in the culturally appropriate combinations with other gestures and with speech.

### Types of Meaning

Certain gestures may have referential meaning in that they may unambiguously replace a linguistic item—e.g., the head nod or head shake in this culture. In most cases, however, individual gestures and gestural complexes signify much more generally. Bolinger (1983, 1985, 1986) has pointed out that certain gestures tend to be associated with certain aspects of intonation:
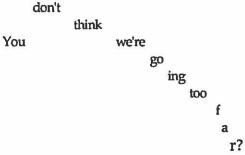
> If intonation is part of a gestural complex whose primitive and still surviving function is—however elaborated and refined—the signaling of emotions and their

degrees of intensity, then there should be many obvious ways in which visible and audible gestures are coupled to produce similar and reinforcing effects. This kind of working-in-parallel is easiest to illustrate with exclamations. An <u>AH!</u>  of surprise, with a high fall in pitch, is paralleled by a high fall on the part of the eyebrows, and also, especially if the surprise is `surprised realization', by a head movement consisting of a drop from a high position. How tight the correlation is can be felt by trying to reverse the direction of movement in one of the three actions. A similar coupling of pitch and head movement can be seen in the normal production of a conciliatory and acquiescent utterance such as

$$\begin{array}{l} \text{I} \\ \quad\text{l} \\ \quad\text{l} \\ \text{wi} \end{array}$$

with the accent at the lowest pitch—we call this a **bow** when it involves the head, but the intonation bows at the same time. It is not essential to the interconnections of the entire gestural complex that all movements be in parallel. Different parts of the complex may respond to different parts of the message. An example is what happens with questions that lead to a conclusion. Most typical are the ones that play an inconclusive facial gesture—raised eyebrows—against a conclusive intonation—namely a fall—on a piece of declarative syntax. Such utterances are usually understood as statements presented for confirmation, e.g. something like

$$\begin{array}{l} \qquad\quad\text{don't} \\ \qquad\text{think} \\ \text{You} \qquad\quad \text{we're} \\ \qquad\qquad\quad\text{go} \\ \qquad\qquad\qquad\text{ing} \\ \qquad\qquad\qquad\quad\text{too} \\ \qquad\qquad\qquad\qquad\text{f} \\ \qquad\qquad\qquad\qquad\text{a} \\ \qquad\qquad\qquad\quad\text{r?} \end{array}$$

Here the eyebrows may be said to do the asking and the intonation the stating.

(Bolinger 1985:98-99)

Analogously, a frowning eyebrow gesture and a retracted chin position are used to reinforce the meaning of a substantially lowered 3 (deepened) voice.[3]

Although there has been little systematic research into this co-occurrence of gestures with specific aspects of pronunciation, it does appear

---

[3] It may be recalled that Ted Baxter, a character on the Mary Tyler Moore Show, often made use of this combination of gestures and deepened voice when striking a stereotypically authoritative news anchorman pose and pontificating about "weighty" matters on which he thought himself to be an authority.

that certain gestures regularly accompany statement, question and continuation intonation patterns (Bolinger 1986). McNeill's (1987) extensive work associating gestures with English grammatical patterns provides a foundation for further work in this area. It appears from some recent research that prosodic peaks and kinesic peaks tend to coincide (Bennett 1980; Hadar et al. 1983; Bolinger 1983, 1985; McNeill 1987). For example, Hadar et al. (1983) confirmed a relationship between head and body movements and the presence or absence of stress or juncture. The work of Duncan (1974, 1976) has shown that cultures vary in the ways that two speakers use synchronization and non-synchronization of rhythms during conversation.

### Teaching Approaches

Techniques to help students synchronize body movements and other gestures with intonation might include rocking in time to lyrical, metrical or other prose material. Acton (1984), for example, included work on rhythmic body gestures accompanying speech in a pronunciation course aimed at changing fossilized articulatory habits. Schnapper (1979) developed activities to increase use and perception of gestures through a type of discovery exercise in which students have to guess what another person is doing based on gestures and movements alone. Inferencing exercises in which the relationship between two interactants or the probable topic of conversation must be guessed based only on gestures—e.g., by watching a video with the sound turned off—may help to sharpen awareness of different gestural complexes. Imitative or creative exercises in which students simulate different roles or role relationships through gestures alone may help to increase the "active vocabulary" of gestures in the new language. "Mirror-me" exercises in which one member of a pair imitates all of the gestures and movements of the other member are a common improvisional theatrical technique which can be applied in the language teaching classroom. Pennycock (1985:275) describes a variety of such imitative games for working on gestures that might accompany speech.

Video programs can be specially developed for work on speech and gesture complexes. These might include exercises in which students match videotaped segments illustrating gestures such as raised eyebrows or lowered chin with a variety of sentences spoken on a high or low pitch. A more advanced exercise might require groups of students to work together to

synchronize stretches of recorded speech to the videotaped movements and gestures which originally accompanied the speech at the time that it was recorded. In addition, student speeches or discussions can be videotaped and then later analyzed for appropriate and effective use of gestures. Commercial videotapes or television programs can be analyzed by class members for use of gestures accompanying speech.

Computer-assistance can be of value in this aspect of training, too. Teacher-designed or commerical videodiscs (which have the advantage that any frame can be accessed in any order) provide more flexibility than traditional videotape for modeling gestures. In addition, computer graphics programs provide a basis for developing sequential lessons on gestures and their speech accompaniments. Computer-assisted video programs or graphics material can be designed to compare and contrast the complexes of gestures and speech which occur within one language or across two or more languages.

## CONCLUSION

In this paper, I have sought to provide a fresh perspective on pronunciation as a curricular area in language teaching. The position taken is that it is impossible to become a competent speaker of a language without attending to the whole meaning complex within which articulation is situated. This complex has been described as including prosody, phonological fluency, voice quality and certain related aspects of kinesics.

The categorization of suprasegmental phenomena into prosody, phonological fluency and voice setting, and the inclusion of gestures in the discussion of phonology, represent a significant departure from traditional accounts. In fact, a "top-down" approach to phonology is essentially the opposite of traditional approaches in which linguists sought to describe all of phonology in terms of discrete segments or features and to exclude from linguistics proper any aspects of communication which did not relate directly to the meaning of individual lexical items or grammatical units. It is felt that the "top-down" approach to phonological description is not only technically more correct for describing connected speech than a "bottom-up" analysis, but also more consistent with modern communicative language pedagogy and with attempts to break the general construct of oral proficiency into reliable

and valid metrics for linguistic analysis and assessment. Most importantly, this type of description would appear to have greater explanatory power than a description in terms of individual phonemes, if the goal is to explain phonological proficiency in the context of language use.

It is expected that pronunciation teaching will become more effective when training and practice opportunities are provided which exploit the multiplex associations of phonological phenomena in human communication, rather than maintaining the traditional separation of pronunciation—e.g., for lab drilling—from its many meaningful associations. Because sounds are not produced in isolation in normal communication and because pronunciation displays so many symbolic functions in human interaction, it is reasonable to hypothesize better results in language education based on a contextualized, associational approach to teaching pronunciation. At the same time, the need to test a variety of specific approaches under controlled conditions must be stressed. Studies should be conducted which compare the results for learners when pronunciation is attended to and taught in context, rather than being ignored or taught as an isolated, mechanical phenomenon in which instruction focuses on individual segments.

In a research program on the instruction of second language phonology, the following questions regarding outcomes need to be addressed:

(1) Do language learners develop a more nativelike accent as a result of a "top-down" curriculum in pronunciation?

(2) Do language learners improve in pronunciation more quickly when contextual methods are used?

(3) Is transfer of classroom training to real-life situations facilitated by associating pronunciation with various meaningful correlates?

(4) Do learners show improved attitudes to instruction when pronunciation is approached in a contextualized manner?

(5) Which specific approaches achieve the greatest positive results in the areas addressed in questions (1) - (4)?

(6) How do individual learner variables such as age, educational level and learning style interact with teaching approach?

Basic research is also needed on the acquisition of suprasegmental

features and the specific ways in which their development at different stages promotes or inhibits the acquisition of other discourse structuring devices and of individual phonemes (see Pennington forthcoming (b) for discussion). Until the results of much applied and basic research of this type become available, questions about the effectiveness of instruction in second language phonology cannot receive definitive answers.

Author's address for correspondence:

Martha C. Pennington
Department of English as a Second Language
University of Hawai'i at Manoa
1890 East-West Road
Honolulu, HI 96822

# LIST OF WORKS CITED

Abe, Isamu. 1955. "Intonational patterns of English and Japanese." *Word* 11, 386-398.

Abberton, E., A. Parker & A. J. Fourcin. 1978. "Speech improvement in deaf adults using laryngograph displays." *Speech and Hearing Work in Progress*. University College of London, Department of Phonetics and Linguistics: 33-60.

Acton, William. 1984. "Changing fossilized pronunciation." *TESOL Quarterly* 18: 71-85.

Adams, Corinne. 1979. *English Speech Rhythm and the Foreign Learner*. The Hague: Mouton.

Beebe, Leslie. 1980. "Sociolinguistic variation and style shifting in second language acquisition." *Language Learning* 30: 433447. Birdwhistell, Ray L. 1970. *Kinesics and Context*. Phildelphia: University of Pennsylvania Press.

Bjarkman, Peter C. 1986. "Natural phonology and strategies for teaching English/Spanish pronunciation." Peter C. Bjarkman & Viktor Raskin. *The Real World Linguist*, Chapter Five. Norwood, NJ: Ablex, 77- 115.

Bolinger, Dwight. 1978. "Intonation across languages." Joseph Greenberg (ed.): *Universals of Human Language*, vol. 2, *Phonology*. Stanford, CA: Stanford University Press, 471-524.

Bolinger, Dwight. 1983. "Intonation and gesture." *American Speech* 58, 156-174.

Bolinger, Dwight. 1985. "The inherent iconism of intonation." John Haiman (ed.): *Iconicity in Syntax*. Philadelphia: John Benjamins, 97- 108.

Bolinger, Dwight. 1986. *Intonation and its Parts: Melody in Spoken English*. Stanford, CA: Stanford University Press.

de Bot, Kees. 1983. "Visual feedback of intonation I: Effectiveness and induced practice behavior." *Language and Speech* 26: 331-350.

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

Brown, Gillian & George Yule. 1983. *Teaching the Spoken Language*. Cambridge: Cambridge University Press.

Brown, James D., & Ann Hilferty. 1986. "The effectiveness of teaching reduced froms for listening comprehension." *RELC Journal* 17: 59-70.

Byrd, Patricia, Janet Constantinides & Martha Pennington. In press. *Foreign Teaching Assistant's Manual.* Chapter Three: "Hearing and pronouncing American English." New York: Collier Macmillan.

Chaudron, Craig. 1983. "Simplification of input, topic restatements and their effects on L2 learners' recognition and recall." *TESOL Quarterly* 17: 437-58.

Collier, R. 1984."Some physiological and perceptual constraints on tonal systems." Butterworth, Brian, Bernard Comrie & Osten Dahl (eds.): *Explanations for Language Universals.* Mouton: Berlin, 237-247.

Delattre, Pierre. 1963 "Comparing the prosodic features of English,German, Spanish, and French." *IRAL* 1: 193-210.

Delattre, Pierre. 1981. "An acoustic and articulatory study of vowel reduction in four languages." *Studies in Comparative Phonetics.* Heidelberg: Julius Groos Verlag, 63-93.

Dirven, R., & J. Oakeshott-Taylor. 1984. "Listening comprehension" (Part 1). State of the Art Article. *Language Teaching* 17: 326-343.

Donegan, Patricia J. 1978. *On the Natural Phonology of Vowels.* Ph.D. dissertation, The Ohio State University.

Duez, Danielle. 1982. "Silent and non-silent pauses in three speech styles." *Language and Speech* 25: 11-28.

Duncan, Starkey, Jr. 1974. "Some signals and rules for taking speaking turns in conversations." Shirley Weitz (ed.): *Nonverbal Communication.* New York: Oxford University Press, 298-311.

Duncan, Starkey, Jr. 1976. "Language, paralanguage, and body motion in the structure of conversations." William C. McCormack & Stephen A. Wurm (eds.): *Language and Man: Anthropological Issues.* The Hague: Mouton, 239-267.

Ecklund, Constance L., & Peter Wiese. 1981. "French accent through video analysis." *Foreign Language Annals* 14: 11-17.

Eibel-Eibesfeldt, I. 1974. "Similarities and differences between cultures in expressive movements." Shirley Weitz (ed.):*Nonverbal Communication.* New York: Oxford University Press, 20-33.

Ekman, Paul, & Wallace V. Friesen. 1974. "Nonverbal leakage and clues to deception." Shirley Weitz (ed.), *Nonverbal Communication*. New York: Oxford University Press, 269-290.

Ekman, Paul, Richard Sorenson, & Wallace V. Friesen. "1969. Pan- cultural elements in facial displays of emotion." *Science* 164: 86-88.

Esling, John H., & Rita F. Wong. 1983. "Voice quality settings and the teaching of pronunciation." *TESOL Quarterly* 17: 89-95.

Fowler, Carol A. 1981. "A relationship between coarticulation and compensatory shortening." *Phonetica* 38: 35-50.

Gilbert Judy. 1984. *Clear Speech*. Cambridge: Cambridge University Press.

Hadar, U., T. J. Steiner, E. C. Grant, & F. Clifford Rose. 1983. "Head movement correlates of juncture and stress at sentence level." *Language and Speech* 26: 117-129.

Haraguchi, S. 1977. *The Tone Pattern of Japanese: An Autosegmental Theory of Tonology*. Tokyo: Kaitakusha.

Hieke, A. E. 1984. "Linking as a marker of fluent speech." *Language and Speech* 27: 343-354.

Hieke, A. E. 1985. "A componential approach to oral fluency evaluation." *Modern Language Journal* 69: 135-142.

Hieke, A. E. 1987. "The resolution of dynamic speech in L2 listening." *Language Learning* 37: 123-140.

Higurashi, Yoshiko. 1983. *The Accent of Extended Word Structures in Tokyo Japanese*. Tokyo: Educa.

Hoequist, Charles Jr. 1983. "Syllable duration in stress-, syllable- and mora-timed languages." *Phonetica* 40: 203-237.

Kaneko, Naomichi. 1957. "Specific x-ray observations on the movement of the lower jaw, lips and tongue in pronunciation." *Study of Sounds* 8 (1957): 1-18. Cited from Yasuyo Edasawa. 1984. "Articulatory setting and teaching pronunciation." Doshisha Women's College, Kyoto, Japan: *Asphodel* 18: 289-308.

Ladd, D. Robert, Jr. 1978. "Stylized intonation." *Language* 54: 517- 540.

Ladd, D. Robert, Jr. 1980. *The Structure of Intonational Meaning*. Bloomington, IN: Indiana University Press.

Laver, John. 1980. *The Phonetic Description of Voice Quality*. Cambridge: Cambridge University Press.

Laver, John, & Peter Trudgill. 1979. "Phonetic and linguistic markers in speech." Klaus R. Scherer & Howard Giles (eds.): *Social Markers in Speech*. Cambridge: Cambridge University Press, 1-32.

McNeill, David. 1987. *Psycholinguistics: A New Approach*. New York: Harper & Row.

Odlin, Terence M. 1978. "Variable rules in the acquisition of English contractions." *TESOL Quarterly* 12: 451-458.

Ohala, John J. 1983. "Cross-language use of pitch: An ethological view." *Phonetica* 40: 1-18.

Pennington, Martha C. 1986. "Acquisition of English phonology by native speakers of Japanese." Presentation to the Japan Association of College English Teachers, TESOL Summer Institute, University of Hawaii, Honolulu, HI, August 1986. Repeated at the annual conference of Japan Association of Language teachers, Hamamatsu, Japan, November 1986.

Pennington, Martha C. Forthcoming (a). "Applications of Computers in the Development of Spoken Language." Dunkel, Patricia (ed.), *Research in CAI and CALL: The Search for Effectiveness*. Englewood Cliffs, NJ: Newbury House.

Pennington, Martha C. Forthcoming (b). "Universals, Prosody and SLA: The context of phonological development." Thom Huebner & Charles Ferguson (eds.), *Contributions of Second Language Acquisition to Linguistic Theory*. Stanford, CA: Stanford University Press.

Pennington, Martha C., & Jack C. Richards. 1986. "Pronunciation Revisited." *TESOL Quarterly* 20: 207-225.

Pennycock, Alastair. 1985. "Actions speak louder than words: Paralanguage, communication, and education." *TESOL Quarterly* 19: 259-281.

Pica, Teresa. 1984. "Pronunciation activities with an accent on communication." *English Teaching Forum* July: 2-6.

Reynolds, Katsue. 1984. "Evidence for masculinity constraints: Fillers in Japanese." Paper presented at the annual conference of the Linguistic Society of America, Baltimore, MD, December 1984.

Robinson, W. Peter. 1979. "Speech markers and social class." Klaus R. Scherer & Howard Giles (eds.): *Social Markers in Speech*. Cambridge: Cambridge University Press, 211-249.

Royal, Ann Marie. 1985. *Male-Female Pharyngealization Patterns in Cairo Arabic: A Sociolinguistic Study of Two Neighborhoods*. Texas Linguistic Forum 27. First appeared as Ph.D. dissertation, University of Texas at Austin, 1985.

Schmidt, Richard W. 1977. "Sociolinguistic variation and language transfer in phonology." *Working Papers on Bilingualism* 12: 79-95.

Schnapper, Melvin. 1979. "Your actions speak louder ...." Elise C.

Smith & Louise Fiber Luce (eds.): *Toward Internationalism*. Rowley, MA: Newbury House, 134-140.

Stampe, David. 1972. *A Dissertation on Natural Phonology*. Ph.D. dissertation, University of Chicago.

Trudgill, Peter. 1974. *The Social Differentiation of English in Norwich*. Cambridge: Cambridge University Press.

Welkowitz, Joan, Ronald N. Bond, & Stanley Feldstein. 1984. "Conversational time patterns of Hawaiian children as a function of ethnicity and gender." *Language and Speech* 27: 173-191.

Woods, Howard B. 1979. *Rhythm and Unstress*. Hull, Quebec: Canadian Government Publishing Centre.