1985

# Evaluation of Foreign Accent Using Synthetic Speech (Perception).

Deborah Marie Rekart

*Louisiana State University and Agricultural & Mechanical College*

8610662

**Rekart, Deborah Marie**

EVALUATION OF FOREIGN ACCENT USING SYNTHETIC SPEECH

*The Louisiana State University and Agricultural and Mechanical Col.* PH.D. 1985

University
Microfilms
International 300 N. Zeeb Road, Ann Arbor, MI 48106

**PLEASE NOTE:**

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark _✓_.

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages _✓_
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) ________ lacking when material received, and not available from school or author.
12. Page(s) ________ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Dissertation contains pages with print at a slant, filmed as received _____
16. Other ____________________

# EVALUATION OF FOREIGN ACCENT
# USING SYNTHETIC SPEECH

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

Speech Communication, Theatre, and
Communication Disorders

by
Deborah M. Rekart
B.A., University of Dayton, 1972
M.A., The University of Rochester, 1974
December 1985

# ACKNOWLEDGMENTS

There are people to thank every step of the way in the realization of this dissertation. First, I would like to thank Kim Robinson, who participated with Ray Daniloff and me in the initial conceptual development of this study seven years ago at Purdue University.

I would like to express my appreciation to Peter Ganley and Carl Thompson at the Veteran's Administration Hospital in Biloxi, MS, for allowing me to use their sound spectrograph to make spectrograms of the three speakers' productions on which the synthetic speech was based.

My thanks to Doug Whalen at Haskins Labs for his technical help in producing the synthetic speech for the pilot study and directing me to the studies of rhythm by Hoequist. Thanks for more technical help at Kresge Lab go to Rick Florenz and Larry Hughes, and particularly to Mick Miller, who was instrumental at every stage in the completion of the synthetic speech stimuli.

I would like to thank Hugh Buckingham, Jo Ann Monjure and especially Jack Damico, who provided me with over forty volunteers by providing them with an incentive to participate in my experiment.

Paul Hoffman deserves special thanks for his

significant help with the experimental design and his work on some of the graphs, and most of all, the statistical analysis (and reanalysis) of this experiment. I would also like to thank Jack Damico and George Yule for their helpful contributions to the Discussion chapter and Joan Besing for her help on graphs. I would like, too, to thank the members of my committee for their help and support.

Ray Daniloff's arrival at L.S.U. gave me the opportunity to finish my Ph.D. and complete the study we had envisioned many years ago at Purdue. My deepest gratitude goes to him, not only for his scholarly expertise, but also for his friendship and guidance throughout my tenure as his student. He has had a profound effect on my life, and I will always be grateful.

TABLE OF CONTENTS

LIST OF TABLES

## LIST OF FIGURES

ABSTRACT

A meaningful sentence loaded with appropriate phonemic and syllabic forms was synthesized as a "standard" stimulus, and 60 "accented" versions of the sentence were made to simulate varying degrees of a moderate and a strong Spanish accent by manipulating the following Spanish cues singly and in combination: (1) fundamental frequency, (2) voice onset time for syllable-initial voiceless stops, (3) duration of medial stressed vowels, (4) F1, F2 and F3 for full vowels, and (5) F1, F2, and F3 for reduced vowels.

Two tapes for each level of accent were prepared on which 30 accented stimulus sentences were each paired with the standard sentence in four randomized sequences. Forty-two English speakers rated how different each accented sentence was from the standard sentence on a 10-point scale; they also gave a confidence rating on a 5-point scale for each item.

It was demonstrated that synthesized sentences can be reliably rated for cue modifications indicative of a moderate Spanish accent in English. Statistical analysis revealed that an increase in the number of cues (from 1 to 2, to 3, to 4, to 5) resulted in the perception of increased

accentedness in both the moderate- and strong-accent condition. In addition, subjects' confidence in their judgments increased along with an increase in number of cues. A factor analysis showed that the suprasegmental cue, fundamental frequency (intonation), was the most perceptually prominent cue signalling a moderate Spanish accent in English. The segmental cue, stressed vowel quality, was the next most prominent cue. The presence of these cues also resulted in an increase in the subjects' confidence in their ratings of stimuli. The two strongest accent-bearing cues signalling the strong accent were both segmental, stressed vowel quality and VOT, but the strong-accent data was determined to be generally unreliable, possibly because of errors in its generation.

# CHAPTER I

# REVIEW OF THE LITERATURE

## Second Language Learning

Most adults who learn a foreign language speak it with a "foreign accent." According to Flege (1979,1980,1984a,To appear), foreign accent detection is probably based on the simultaneous perception of divergences from L2 (i.e., the language being learned) phonetic norms at the subsegmental (i.e., phonetic), segmental, and suprasegmental levels.

It it hypothesized in the present study that (1) many differences considered as segmental are actually subsegmental or phonetic; so that foreign accent is a phonetic rather than a phonological phenomenon; (2) adults can and do learn new "patterns of pronunciation" which increasingly approximate L2 pronunciation; however, (3) the listeners of an L2 speaker analyze the pronunciation through their own sound system, and accoringly hear "phoneme substitutions" rather than phonetic approximations. Such claims are radical considering the Contrastive Analysis model that has dominated the second-language-learning literature for the past 30 years.

It has been claimed that foreign accent occurs because the ability of humans to learn new patterns of pronunciation diminishes near puberty for neurophysiological reasons (Scovel, 1969). Thereafter, contrastive analysis predicts direct substitution from L1 (native language) to L2 (Ioup, 1984: 11). According to Lado (1957: 72), "the learner transfers the sound system of his native language and uses it instead of that of the foreign language." For example, a Spanish-speaker learning English will use Spanish /i/ as in ir ('to go') for both English /iy/ and /I/ in eat and it respectively (Lado, 1957: 73).

Consider the subsegmental (subphonemic, phonetic) variations involving a Spanish-speaker learning English who uses unaspirated Spanish [p] for both aspirated and unaspirated realizations of English /p/ (Lado, 1957: 73). According to contrastive analysis, the English speaker/listener would hear English [b] for Spanish [p] because the voicing of English [b] occurs at a VOT of 1 ms, whereas the VOT of Spanish [p] occurs at 4 ms. Thus, phoneme /b/ would perceptually substitute for /p/.

One major problem with contrastive analysis is that it fails to account for the gradual nature of phonological learning (Flege, 1979: 35). For example, Flege (ibid.) states that "according to Dickerson (1975), the number and

type of substitutions produced by Japanese speakers for English /l/ ([ľ], [ř], [r], [ď]) ([ž], [dz], [š], [θ]) depend, among other factors, on the level of proficiency in English. Dickerson's data argue that the range of variants produced for a single target-language sound is not random. Rather, a learner's pattern of errors reflects the same kind of patterned variation which characterizes gradual sound change within a homogeneous language community...Variable substitution patterns suggest that learners only gradually learn to implement new speech sounds in a foreign language."

Ioup (1984: 12) comments that "in a study of the English spoken by native speakers of five Asian languages, Beebe (1980) found that the data exhibited very few direct substitutions of native language phonemes for different target sounds, in contrast with predictions of contrastive analysis. Instead, the learners tended to approximate, to various degrees of accuracy, the target-language sound."

Ioup (ibid.) elaborates: "Beebe (1980) postulates a sequence of stages of "inter-language" (IL) phonology. (1) At the first stage, learners substitute native language-phonemes for difficult target sounds. (At this stage, learners exhibit the type of transfer detailed in contrastive analysis. Perhaps it is only at this stage that a literal application of contrastive analysis is

appropriate.) (2) As they progress, they attempt various approximations of the target norm. This stage is characterized by a proliferation of variants for each attempted sound. (3) Gradually, the number of variants is reduced as the approximations become closer to the target norm."

Experimental studies have shown that Beebe's stages (2) and (3) do indeed exist. The values of phonetic parameters measured in the speech of second-language learners are often intermediate to those typical of monolingual speakers of the native and target language (Pinkerton, 1972; Suomi, 1976; Niemi, 1979; Flege, 1980; Port and Mitleb, 1980; Eisendoorn, 1980; Williams, 1980; Flege and Port, 1981; Mitleb, 1981. (Flege, 1981: 452). Caramazza et al. (1973) found that French speakers of English labeled stops in a VOT continuum differently than monolingual speakers of French. Williams (1980), in a study of Puerto Rican children learning English as L2, found that they learned to produce English stops with increasingly large (English-like) VOT values and also to produce Spanish stops more like English stops. In a study by Flege and Hillenbrand (1984), native French speakers who had an average of 12.2 years' experience speaking English in an English-speaking environment produced /t/ in French words

with VOT values intermediate to the values typically observed for French and English. It seems that not only does L1 influence L2, but L2 also affects L1 along the "inter-language" continuum of phonetic values.

Studies of VOT perception testify to the decline in some discriminative powers as the infant develops. While infants from diverse linguistic backgrounds respond to contrasts in prevoiced, voiced and voiceless initial consonants, adult speakers of some languages, including English, recognize only the distinction between the voiced and voiceless categories.

R. C. Tees showed that 6- to 8-month-old infants from an English-speaking background readily distinguished phonemic contrasts in Hindi and Salish (a North American Indian language). When they were tested again at the age of 12 months, the same infants, like English-speaking adults, did not detect the contrast to which they had earlier been sensitive (cited in Eimas, 1985).

Apparently the restricted linguistic environment of one's native language does not inactivate unused perceptual mechanisms completely. Furthermore, with enough experience the perception of non-native distinctions begins to operate at the phonemic level: after considerable experience with spoken English, native speakers of Japanese can distinguish

the phonemes /r/ and /l/ categorically and almost as accurately as native English speakers. The fact the perceptual mechanisms available to us as infants can be activated in adulthood confounds the strong hypothesis that early experience with language immutably alters some of the mechanisms of speech perception (Eimas, 1985: 52).

In a study of voice onset time of initial /p/ and /b/, Gass (1984: 71) found that, over time, L2 learners of English showed a greater amount of target language influence and lesser amount of native language influence with increased proficiency. Furthermore, production showed a trend from lesser variability to greater variability and back to lesser variability.

It seems that second-language learners are perceiving as well as producing inter-language forms along a continuum between L1 and L2 values. The native listener may well categorize the forms according to his own system. As early as 1939, Trubetzkoy viewed the phenomenon of "sound substitution" reported by native speaker/hearers of non-native productions as stemming from mistaken phonological interpretations of sounds in L2 which were based on phonological categories found in L1 (Flege, 1979: 17). Trubetzkoy compares the process of identifying sounds in a foreign language to the action of a sieve: a listener

picks out only the phonetic dimensions which are phonologically relevant in his native language. Only these dimensions contribute to the recognition of foreign language sounds.<1>

In summary, within the first year of life, human linguistic experience tends to narrow down a wide range of perceptual contrasts to those important to the language community. The ability to make other contrasts remains, and it can be reactivated (at least in part) with new linguistic experience. Second-language learners tend to produce forms based on their perception along a continuum of values ranging from L1 to L2 (Abramson and Lisker, 1973). As experience "improves" perception, the forms produced approximate the target norms more closely.

According to Flege and Hillenbrand (1984: 692), it is the formation of detailed phonetic representations in L1 that enables listeners to detect accent in the speech of non-native speakers: These phonetic representations are seen to arise via a prototype construct like the one developed by Rosch (1973, 1978) in which objects in the world are categorized in comparison to internal prototypes (icons) which represent a category's core properties. Prototypes are often developed through experience with many categories, and specific exemplars of a category are

accepted or rejected based on how closely they conform to the prototype. The prototype construct has been applied to speech research by Oden and Massaro (1978) (see the section on speech perception models in this chapter), who argue that phonemes are identified by comparing speech stimuli to phonetic category stereotypes stored in long-term memory (Repp, 1976; Miller, 1977; Samuel, 1982) (Flege and Hillenbrand, 1984).

## Prosody

### Suprasegmental versus Segmental Features

It has been claimed that "the little evidence available suggests that suprasegmental errors are largely responsible for the perception of foreign accent" (Flege, 1979: 2). Consider the following evidence:

Ioup (1984: 9) reports that the majority of errors she found for two groups of non-native English speakers, Koreans and Egyptians, involved "prosody and vowel coloration which gave a characteristic identification to each group's accent."

Metcalf (1972) contends that the English of California Chicanos differs minimally from the local Anglo-English. However, intonation often includes separate peaks of loudness and pitch change within a phrase which Metcalf found to be "less rapid falling-off of pitch and loudness at the end of a declarative sentence than in Anglo dialects."

Palmer (1976) counted the number of directly observable "segmental" and suprasegmental errors in a sample

of English prose recorded by native French speakers. He found a stronger correlation between native English speakers' ratings of intelligibility and suprasegmental errors than with segmental errors (Flege, 1979: 3).

It is widely believed that suprasegmental errors persist longer than so-called segmental errors. For example, Trubetzkoy (1939/1969: 54) cites the case of Russians learning Czech. Russians find it difficult to dissociate stress and duration, since in Russian the major physical correlate of stress is duration. Although Russians soon learn that stress always falls on the first syllable in Czech words, they continue to implement Czech stress by means of duration even after a good command of Czech had been acquired (Flege, 1979: 4).

Niemi (1984) reports than an error analysis of Finnish speakers of English shows that the non-local, perhaps rhythmically induced, minor timing adjustments that accompany stress change are difficult to internalize even at a late stage of L2 learning.

In fact, Huggins (1972) found that "even when the words themselves become unrecognizable [via time-compression of speech], prosodic features continue to be transmitted, supporting the notion that prosodic features may well be the most resistant parts of the speech waveform to any natually

occurring form of distortion" (Wingfield, 1975: 153).

The perceptual relevance of prosodic variables is demonstrated in misperceptions occurring in the reading aloud of English by Indian speakers. When stress placement was wrong in these realizations, this prosodic feature tended to override "segmental" information in recognition by native speakers of English (Bansal, 1966, cited by Huggins, 1972).

Segmental versus suprasegmental processing has been explored by several researchers. Wood (1974; 1975) used a 2-choice speeded classification task adapted from Garner's (1974) study of visual perception. Subjects classified syllables according to pitch (high/low) or according to consonant ([b]/[g]) under two conditions of variation. In the control condition, two stimuli varied along the target dimension, with the value along the nontarget dimension held constant. In the orthogonal condition, all four stimuli were presented with subjects classifying syllables according to consonant- or pitch-distinction. Decision times in the two conditions for each target feature were compared.

Wood hypothesized that the extent to which random variation along the nontarget dimension in the orthogonal condition slows reaction time over that obtained in the control condition is assumed to reflect the degree to which

analysis of the target feature is dependent upon the analysis of the nontarget feature.

It was found that the interference caused by orthogonal variation in the nontarget dimension was asymmetric: when the subject was targeting for pitch, there was no reliable increase in classification time in the orthogonal condition, but when the target dimension was consonantal, variation in pitch interfered with processing, resulting in reliably higher RTs in the orthogonal than in the control condition. These results indicate, according to Wood, that although listeners could ignore variation in consonantal quality while making pitch judgments, they could use redundant variation in consonantal quality and pitch to speed pitch judgments.

Goldin (1975) showed that the above asymmetry is not limited to the dimensions of place of articulation and pitch. Asymmetrical interference was also obtained when the two dimensions were (1) place and loudness or (2) voicing ([b]/[p]) and pitch.

The results of Goldin and Wood imply that there is a unidirectional dependency in the processing of segmental and suprasegmental information, with the analysis of segmental information dependent on the prior analysis of suprasegmental information. Analysis of suprasegmental

features, however, is accomplished independently of segmental analysis.

J. Miller (1978) attempted to determine whether a pattern of asymmetric interference would also hold when a contrast in vowel quality rather than in consonantal identity was paired with a contrast in a suprasegmental feature. Subjects were 12 paid university students. Stimuli consisted of 4, 3-formant, synthetic CV syllables, [ba] and [bae], each at 2 levels of FO. There were three conditions of variation: (a) 2 control tapes with low [ba] and [bae] and high [ba] and [bae]; (b) 2 orthogonal tapes with high [ba] and low [bae]; (c) 2 correlated tapes which contained instances of all four stimuli. Stimuli for a second experiment consisted of 4 similar CV syllables except that [ba] and [bae] differed in loudness levels rather than in pitch. Each subject classified the syllables according to each dimension: vowel and pitch (Experiment 1) or vowel and loudness (Experiment 2) under each of the 3 conditions: correlated, control and orthogonal. Results of the two experiments were essentially the same, i.e., unlike the asymmetric interference effect for consonant and pitch, the interference effect for vowel and pitch was both mutual and asymmetrical.

The mutual, asymmetrical interference effects found

in this study along with the asymmetric interference effects reported by Wood and Goldin, indicate that there is not a single type of interaction in the processing of segmental and suprasegmental information. While the analysis of the consonant information and pitch (or loudness) was found to be done in a hierarchical or serial fashion, with consonantal analysis dependent on analysis of the suprasegmental information, the analyses required to determine vowel quality and pitch (or loudness) were found to operate in a mutually dependent fashion. This fact is not surprising when one considers that the suprasegmental information is carried by vocalic segments in the speech signal.

Classification of Suprasegmentals

Following Lehiste (1970), suprasegmental features can be subdividedd into three categories: quantity features, tonal features, and stress features.

The physiological mechanism ultimately responsible for quantity phenomena is the process involved in the timing of articulatory movements. The physical correlate of the timing of articulatory sequences is the time dimension of the acoustic signal. The perceptual correlate of the time

dimension is the perception of duration.

The general term <u>tonal features</u> refers to all aspects of the linguistic use of fundamental frequency and its physical and perceptual correlates. The physiological correlate of tonal features is the vibration of the vocal folds in phonation. The acoustic correlate of vocal fold vibration is the fundamental frequency of the sound wave generated at the glottis. The perceptual correlate of tonal features is perception of pitch. Pitch, when it functions distinctively at the word level, is referred to as tone; when it functions at the sentence level, it is referred to as intonation.

Differences in stress are due to differences in physical effort. The effort is reflected directly in the activity of the muscles involves in respiration, and indirectly in subglottal pressure. Stress is acoustically manifested in terms of intensity and amplitude variation. The perceptual correlate of stress features is perception of loudness and stress.

## Perception of Duration

There have been many studies investigating the perception of duration. Huggins (1978) measured JNDs for

increases/decreases in duration of a [p]-closure in a naturally produced English sentence. Measurements were made on two versions of a sentence which differed only in the duration of the stressed vowel followed by the [p]-closure. The perceptual effect of a lengthened closure was that the speaker had hesitated in otherwise fluent speech. If the [p]-closure was greatly shortened, all subjects reported an apparent increase in speaking rate in the two unstressed syllables preceding the shortened closure. These results were interpreted to mean that the whole temporal pattern was spontaneously restructured to assimilate the shortened closure.

In English, there is a tendency for stressed groups to assume approximately the same duration (Classe, 1939). Peterson and Lehiste (1960) hypothesized that if such isochronous groups include a varying number of syllables, their duration must vary according to the number of syllables included in the group, and spacing between main stresses would tend to remain constant.

Peterson and Lehiste devised an experiment in which the number of syllables was kept constant in a series of English frame utterances while the spacing between main stresses was varied systematically. (To control tempo, subjects were asked to speak in time to a periodic pulse

produced over headphones so the monosyllabic test words bearing sentence stress coincided with the pulse.)

Results of the experiment revealed that the duration of syllables with main sentence stress changed less than the duration of unstressed syllables: when rate of utterance was increased by a factor of two, the stressed words decreased in duration by a factor of approximately 1.5.

When speech tempo increases, segmental durations shorten (Peterson and Lehiste, 1960; Lehiste, 1970; Kozhevnikov and Chistovich, 1965). In addition, temporal relationships in a "precursor sequence" have been shown to influence the perception of subsequent segments. Summerfield (1975) varied precursor tempo in synthetic sentences and found that tempo change produced a shift in VOT boundary.

In English the perception of the vowel /ae/ as opposed to the vowel [E] depends, among other things, on its physical duration. Nooteboom et al. (1978) state: "By shortening an /ae/-like sound we can evoke /E/ responses. We may refer to the vowel duration which results in 50 /ae/ responses and 50% /E/ responses in a binary forced-choice experiment as the <u>phoneme boundary</u> between /ae/ and /E/. We may embed the test vowels used for measuring this phoneme boundary in a sequence of vowel segments."(p. 85).

Ainsworth (1972) showed that the measured value for the phoneme boundary depended on the tempo of the vowel sequence: the faster the tempo, the shorter the value for the phoneme boundary. (This phenomenon, known as perceptual normalization, also occurs with VOT as a cue to the voiced/voiceless distinction.) Thus, categorical boundary specification in the identification of English vowel phonemes can be affected by changing the tempo of the sequence of synthetic vowels in which the test vowel is embedded (Ainsworth, 1972, cited in Nooteboom et al., 1978: 85).

In a language like Dutch where length is phonemic, if a vowel segment is ambiguous as to its phonemic perception, duration is systematically affected by some aspect of the local context which follows (i.e., local speech rate) (Nooteboom, 1975, cited in Nooteboom et al., 1978). Evidence for backward perceptual normalization, both within a syllable and over more than one syllable, has been obtained in a binary forced-choice listening test in which response percentages of identification of Dutch short /a/ and long /a:/ were measured as a function of acoustic vowel duration (Nooteboom, 1975). The same set of test stimuli, consisting of spectrally homogeneous and monotonous vowel segments with durations varying in steps of 5 ms, from 60 to

130 ms, was embedded in six different contexts (i.e., a, as, at, atə ,atəxə, atəxərə). (The context frames were obtained by removing an initial vowel /a:/ from recordings of naturally spoken words.) On the basis of acoustic measurements of similar speech material it was predicted that if backward perceptual normalization were to take place, response distributions would shift towards shorter durations from context condition (1) to (6) in the order given.

The standard deviations of the distributions of phoneme boundaries (50% points) for the ten subjects in each context condition were calculated in milliseconds. The mean values for mean phoneme boundaries and their standard deviations (in parentheses) in conditions (1) to (6) respectively were $\bar{X}$=100(s=8.4); 97(6.7); 91(6.9); 88(5.8); 85(5.5); and 83(4.3) ms. (Nooteboom et al., 1978: 87). The authors (ibid.: 88) state: "The effect of the /t/-stop closure duration was investigated separately in a nonsense word (atəxə-a:təxə). The phoneme boundary between /a/ and /a:/ in this frame word was measured in four conditions, viz. with /t/-stop closure durations of 60, 70, 80 and 90 ms." The positive linear relation between duration and postvocalic silent interval and phonemic boundary confirms the hypothesis that local speech rate is a determining

factor in backward perceptual normalization (i.e., delay of judgment of phonemic identity until more information is available) (Nooteboom et al., 1978).

In another experiment, Nooteboom et al. (1978) attempted to show that perceptual normalization may reflect contextual regularities due to the syntactic position of a word in a sentence. Response distributions of the vowels in Dutch words /tak/ (branch) and /ta:k/ (task) were measured. "The following two frame sentences were used: (1) Kees kreeg een nieuwe t-k (Kees got a new --); (2) Kees kreeg een nieuwe t-k op zijn schouders (Kees got a new -- on his shoulders)" (Nooteboom et al, 1978: 89). The acoustic realization of the first frame sentence was obtained by removing the final phrase of a recording of the second sentence so that the speech material preceding the test vowel was identical in both conditions (the first frame sentence = final condition; the second = embedded condition). An identical set of /a/-like vowel segments was used for constructing stimulus sentences by inserting each vowel segment between /t/ and /k/ in both frame sentences via a computer facility for editing the speech wave form.

Nooteboom et al. (1978: 89-90) report that "the average values of phoneme boundaries and the standard deviations in milliseconds of the distributions of these

phoneme boundaries over ten listeners were, (a) final condition $\bar{X}$=94(s=6.7)ms; (b) embedded condition: $\bar{X}$=79(s=3.4)ms. It could be argued that the shift in phoneme boundary is caused by a difference in semantic probability of the words 'taak' and 'tak' in the two sentences. Alternatively it is possible that the shift is brought about by an adjustment to the acoustic information which follows; that is, a rather long /k/ plus silence in the sentence final condition, and a short /k/ plus a rather brief syllable /op/ in the embedded condition."

If the latter hypothesis is correct one would expect that a shift in phoneme boundary might be obtained by changing temporal aspects of the immediate environment without changing the recognition of the frame sentence and its meaning. In the frame sentence used for the embedded condition a silent gap was introduced between the /k/ of /t-k/ and the rest of the sentence having one of the following durations: 0, 100, 200, 400, and 800 ms. The /a/-/a:/ phoneme boundary was measured for each of these five silent-gap durations. (Nooteboom et al., 1978: 90) Results showed that a silent-gap duration of approximately 200 ms is sufficient to obtain a phoneme boundary which is nearly identical to the one measured in sentence final condition. From the onset of the test vowel to the end of

the 200 ms gap, a time interval of about 350 ms elapses which is the time-span over which perceptual normalization can take place (ibid.).

In another experiment, Nooteboom et al. (1978) investigated the predictive nature of duration, showing that changing the duration of only three preboundary vowels in an utterance may reliably alter its perception. The first and second of the three preboundary vowels were equally effective; the third had no effect at all (possibly because by the time the subject hears the third vowel, his perception has already been determined by the other two.

Verbrugge, Strange, Shankweiler and Edman (1976) used natural speech, splicing a precursor of three syllables spoken at a regular 1/s rate in front of a destressed test syllable excised from sentence context. They found that recognition errors were greater when the test syllable was preceded by the precursor sequence than when there was no precursor. This result was attributed to misleading tempo effects established by the precursor sequence, i.e., by the mismatch between the slow precursor tempo and fast (destressed) test vowel tempo.<2>

Martin (1979) used a 6-syllable "sentence" of 3 stressed syllables containing the V /a/ and 3 unstressed syllables (approximately schwa) with grammatical structures:

(a) DASaLASaGASa (trochaic), and (b) aBASaGASaLAS (iambic). The following distortions were made to each of the structures. In trochaic sentences, the vowel in one stressed syllable (serial positions 1 or 3) was lengthened or shortened by 50, 90, and 130 ms; an unstressed vowel in positions 2 or 4 was lengthened or shortened by 50 ms. In iambic sentences, the stressed syllables in positions 2 or 4 and unstressed syllables in positions 1, 3, and 5 were distorted in the same way. Nine paid student volunteers participated in three sessions approximately 2 hours long. They received a bonus to their $2.50/hour base pay of .05 for each rapid response (less than 150 ms) minus .05 for each miss/false alarm. Reaction time was recorded to assigned target segments /b,d,g/ in syllables 1-5.

Martin reports the following results: (a) In general, time distortions of a sentence increased segment target reaction time; (b) vowel changes in the first syllable increased reaction time to later targets and the reach of effects spanned up to 4 syllables. Target reaction time was increased by both shortening and lengthening; (c) effects of stimulus expectancy increased directly with time into the sentence, whereas effects of processing time decreased; (d) tempo-change effects persisted throughout the experiment despite practice, motivation, and familiarity

with stimuli.

According to Martin, "the results suggest that rhythmic expectancies may be established once the first syllables are heard, that perception makes immediate use of both rhythmic and segmental information as it becomes available" (p. 1293).

From the beginning of this century (cf. Roudet, 1910: 237), many investigators have showed that speech segment durations are systematically affected by the position in an utterance. One of the most important effects appears to be a shortening of speech segment durations as a function of the number of syllables yet to be produced in the word or phrase (Nooteboom, 1973, cited in Nooteboom and Cohen, 1975: 133). Klatt and Cooper (1975: 69) found that the JND for a change in duration to a single segment (in a randomized set of different sentences) was 25 ms or more. "Just noticeable differences between sentence environments suggest the following constraints on the perception of durational information: (a) sensitivity to changes in duration is best for segments in non-word final syllables; (b) sensitivity is influenced by a backward masking effect of any of the following words" (Klatt and Cooper, 1975: 69).

In one experiment, Lehiste (1980) developed four

sets of test words. Two sets of monosyllabic, disyllabic and trisyllabic words were composed of syllables big/bag in one list and bick/back in the second list. (All possible stress placements were represented in the lists which contained 34 words each.) The third list was made up of 34 English words matching the nonsense words in syllable length and stress placement. The fourth set was divided into two parts: part (a) contained 10 words in which the unstressed syllable be was combined with the stressed syllables big/bag (e.g., bebig) in disyllabic and trisyllabic words; and (b) with 10 similar words where unstressed syllable be was combined with stressed syllables bick/back (e.g., bebick). Test words were placed in three frames: (a)a short frame, "Say X instead"; (b) a long frame with the test word near the end, "Sometimes it's useful to say the word X instead; and (c) a long frame with the test word near the beginning, "The word X is sometimes a useful example." The lists of words were read by three graduate students in the three frames in an anechoic chamber.

Figure 1 gives average durations of test words in different frames for the three speakers. Lehiste hypothesizes that if the duration of words depends on the number of syllables that remain to be produced in the utterance, then test words should have the same duration in

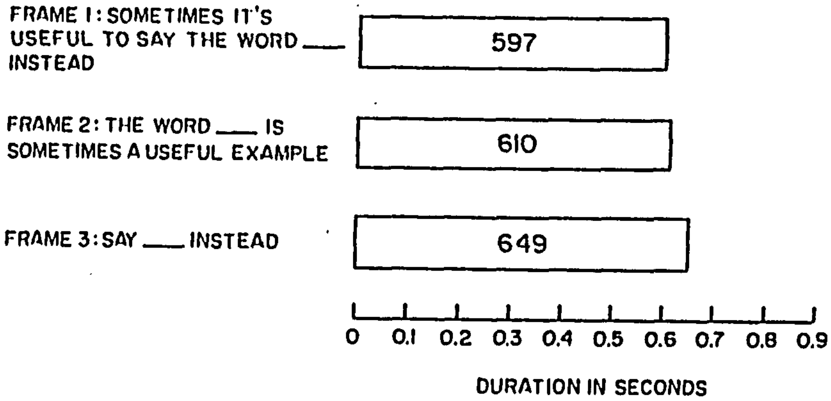


Figure 1. Average duration of test words in different frames averaged for three speakers.
(from Lehiste, 1980)

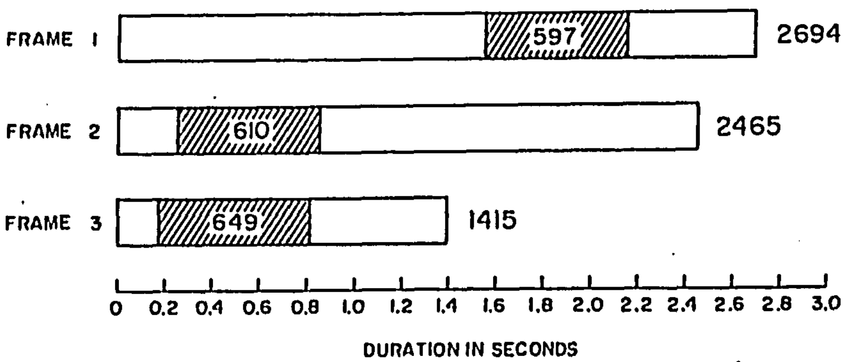


Figure 2. Average duration of test utterances, averaged for three speakers.
(from Lehiste, 1980)

the two frames where only "instead" follows the test word. However, that is not the case.

Figure 2 shows that the average durations of the words is inversely correlated with the length of the total utterance, so the test words appear longest in the shortest utterance and the shortest in the longest utterance. Thus, the duration of test words depends on the total duration of the utterance rather than on the position of the test word within the sentence. These results indicate that the speakers integrated the test words into the utterance at the level at which the time program for the whole sentence was generated.

In summary, it seems that the speaker generates the overall temporal pattern of an utterance, which determines the duration of the segments in that utterance. The listener establishes "rhythmic expectancies" upon hearing the first syllables of an utterance. These may be challenged by alterations in context, e.g., mismatched precursor sequence, at which time the hypotheses are revised on the basis of available rhythmic and segmental information.

## Perception of Stress

Factors such as intensity, duration and FO variation have been investigated in connection with the perception of stress.

Fry (1955) used synthetic noun/verb pairs of test words (e.g., object, digest, permit), in which he varied duration and intensity of both syllables. Intensity values were as follows: V1/V2 equal; V1: +5, +10, -10 dB. Duration ratios between the two syllables were based on observed durations in actual productions of words (e.g., permit: 0.50, 0.75, 1.00, 1.50, 2.00). A listening test was administered to 100 subjects.

In words in which duration and intensity were operating in the same direction, there was excellent agreement among the subjects: when the vowel was long and of high intensity, listeners agreed that the vowel was strongly stressed; when it was short and of low intensity, it was judged as weakly stressed.

When the effects of duration and intensity were studied separately, duration provided the overriding cue. When intensity was constant, increasing the duration ratio (V1/V2) increased the noun judgments (i.e., first syllable stressed) by 70%. The whole range of intensity change

produced an increase of only 29% in the number of noun judgments.

Morton and Jassem (1965) used synthetic syllables /sisi, s s , sasa/ produced on PAT (parametric synthesizer by Lawrence, 1953). Fundamental frequency, intensity and duration were varied systematically.

It was found that variations in FO produced far greater effects than variations in either intensity or duration; a syllable was marked stressed if it was different from the "context" fundamental, and a raised fundamental was more efficient than a lowered one. In general, more intense and longer syllables were more likely to be marked as stressed.

These results appear to confirm the "all-or-none" effect of FO changes observed by Fry; changes of 25 and 58% in FO (i.e., step-ups from 120 Hz. to 151 or 190 Hz., or step-downs from 120 Hz. to 96 or 76 Hz.) were equally effective in producing stress judgments.

In the above studies, it appears that FO provided relatively stronger cues for the presence or stress than did intensity. Duration also appeared to play a larger role than intensity.

Cutler (cited in Cutler, 1975) spliced target words into a sentence context, whose prosody suggested that the

target word would or would not be stressed. Reaction time to a target phoneme in the target word was faster when subjects expected a stressed word than when they did not. Thus, stress seems to depend on the preceding prosodic pattern, allowing the listener to anticipate the forthcoming stressed word (Darwin, 1975: 179).

In a second experiment, a quarter-second period of silence was inserted before a pair of monosyllabic words embedded in a fluently spoken sentence. Reaction time to a phoneme target at the beginning of either of the words was influenced: reaction time was slower to the second word, which was initially spoken with stress and faster to the unstressed first word. Cutler concluded that local temporal disturbances influence phoneme-monitoring reaction times.

Allen (1972a,b) showed that subjects can tap with less variability to stressed syllables than to unstressed ones implying that listeners attend more to stressed syllables than to unstressed syllables.

Huggins (1972) found subjects to be more tolerant of timing distortions which preserve stressed syllable separations.

Cutler and Foss (1973), following an earlier finding that RT to an initial stop consonant was faster when it appeared at the beginning of a content word than when it

appeared at the beginning of a function word, showed that this difference was attributable to differences in the stress with which function and content words are normally spoken. The authors concluded that allowing processing to be directed towards the stressed parts of the sentence allows the focus of the speakers sentence to control the listener's perception.

Shields, McHugh and Martin (1974) measured RT to nonsense disyllables beginning with a target phoneme. The disyllable was pronounced with the stress either on the first or second syllable. When the word occurred as part of a fluent sentence, subjects were faster to react to a target in a stressed syllable than in an unstressed one provided that the target did not occur too close to the end of the sentence. However, when some target words were spliced out and presented in isolated list form, the differences between stressed and unstressed syllables disappeared.

Thus, it appears that the listener's attention is focused on stressed syllables in an utterance, and the preceding prosodic pattern allows the listener to anticipate forthcoming stress.

Perception of Rhythm

Faure, Hirst and Chafcouloff, 1980) tested the hypothesis that the alternation between more and less prominent syllables (i.e., stressed and unstressed respectively) contributes to the perception of rhythm.

Two British subjects recorded sentences of the following type in an anechoic chamber: (1) The manager is the one who purchased it. (2) The teacher is interested in buying some books. (3) The teacher was interested in buying some books, but the manager was the one who purchased them at the university.

Three phoneticians independently marked stressed syllables and pauses; those which were marked by two of the three were considered for analysis. The duration between onsets of succeeding stressed syllables was measured.

Two explanations can be given to account for the data: (1) Although it is not true that stressed syllables are separated by even "roughly equal" intervals of time, the results tend to confirm the impression that "the speed at which the unstressed syllables are uttered and the length of them will depend upon the number occurring between the strong beats" (Gimson, 1962: 238); (2) There is no reduction at all, but stressed syllables are simply longer than

unstressed syllables. Evidence presented in Table 1 (from Faure et al.: 75) seems to point to an average duration of stressed and unstressed syllables: 0.22 and 0.14 sec. respectively, which lends support to the first explanation.

Nakatani and Schaffer (1978) used reiterant speech (i.e., "speech obtained when the same syllable, usually /ma/, is substituted for every syllable of a meaningful sentence") to eliminate powerful meaning and sound cues for word perception in order to study only prosodic cues (e.g., rhythm) for word perception.

Subjects consisted of 7 talkers (4 female, 3 male). Stimuli were trisyllabic Adj-N phrases in the context of sentences, e.g., "The remote stream was perfect for fishing." becomes "The mama ma was perfect for fishing." Adj blocks consisted of phrases with disyllabic Adj + monosyllabic N; N-blocks consisted of phrases with a monosyllabic Adj + disyllabic N, each spoken by 7 talkers. The stimuli consisted of two blocks of 63 trials, each composed of ambiguous phrases (same stress pattern, different parsings, e.g., 10-1 noisy dog vs. 1-01 bold design) and unambiguous phrases (parsing unique to stress pattern).

Reiterant phrases were played to listeners who judged whether they heard each phrase as "ma mama" or "mama

Table 1. Mean duration of unstressed syllables

| Number of syllables | Duration of interval | Duration of stressed syllable (estimated) | Total duration of unstressed syllables | Number of unstressed syllables | Mean duration of unstressed syllables |
|---|---|---|---|---|---|
| 1 | 22.00 | 22 | 0 | 0 | — |
| 2 | 35.75 | 22 | 13.75 | 1 | 13.75 |
| 3 | 50.03 | 22 | 28.03 | 2 | 14.01 |
| 4 | 68.50 | 22 | 45.50 | 3 | 15.00 |
| 5 | 80.71 | 22 | 58.71 | 4 | 14.68 |
| 7 | 102.00 | 22 | 80.00 | 6 | 13.33 |

(from Faure et al., 1980)

ma" (i.e., monosyllabic Adj + disyllabic N or disyllabic Adj + monosyllabic N) by putting a slash between the appropriate syllables, i.e., ma/ma ma or ma ma/ma on the response sheet.

All phrases were parsed significantly better than chance (50%), but the stress pattern made some phrases significantly easier to parse than others (i.e., null or secondary stress on the first of last syllable meant that the syllable was part of a disyllabic word since monosyllabic words carried primary stress).

The authors conclude: (1) most, but not all, listeners used stress pattern as a prosodic cue for word perception; (2) When the stress pattern was not a sufficient cue for word perception, listeners must have used other prosodic cues for word perception since ambiguous phrases were parsed better than chance.

In another experiment, Nakatani and Schaffer (1978) used hybrid speech synthesis (Olive and Nakatani, 1974). From two phrases A and B with the same stress pattern, different parsings, a new phrase X (hybrid) is created, in which rhythm comes from B and other prosodic features come from A. How often listeners parse X like A or B is a measure of how weak/strong rhythm is as a cue for word perception.

The stimuli consisted of two sets of parents (i.e.,

pairs of ambiguous phrases x 3 stress patterns (i.e., 101: "noisy dog," 111: "malformed nose," 121: "foolproof lock") x 2 talkers. Hybrids were generated by taking rhythm, pitch, amplitude and spectrum from either parent in all possible combinations. Each offspring was paired with its "ma mama" parent on one stimulus tape and its "mama ma" parent on the second tape; the parent served as the anchor phrase on each trial. Each tape was played for two groups of 11 subjects each with either "ma mama" or "mama ma" parsing instructions.

The authors hypothesized that if rhythm is a strong cue, and A has the rhythm of the anchor parent, then A should sound like the anchor and have a parsing discriminability score not much better than chance; parsing B would be easily discriminable from that of the anchor and have a higher discriminability score. The difference between the high score for B and the near-chance score for A is a measure of the strength of rhythm as a cue for parsing ambiguous phrases. Results showed that rhythm was the only prosodic cue for word perception in the ambiguous reiterant phrases; pitch and amplitude were not.

Nakatani and Schaffer (1978) then address the question: What specific aspect of rhythm differentiated two ambiguous phrases? Further analysis of the data showed that

(1) the monosyllabic Adj in "ma mama" was longer by about 50 ms than the first syllable of the disyllabic Adj in "mama ma"; (2) the difference between word-final and word-initial syllables averaged only 5 ms; (3) word-initial /m/s were about 11 ms longer than word-medial /m/s in syllables 2 and 3. The elongation was only about a fifth of that of monosyllabic words; (4) /ma/ syllable lengthened by about 50 ms as stress level rose from 0 to 2, and lengthened by about 50 ms as stress level rose from 2 to 1. In summary, monosyllabic word elongation was the primary cue for parsing ambiguous phrases; word-initial consonant elongation was probably a secondary cue.

The authors conclude that stress pattern and rhythm are primary cues for word perception. In their experiment, stress had to be a direct cue because, as parsing of hybrid phrases showed, the reiterant phrases had no segmental cue that enabled subjects to parse them. Rhythm may also be a primary cue for stress perception, since in reiterant phrases there were noticeable differences in length of syllables with 0, 1, and 2 stress levels.

The technique used in experiments to test the predictive nature of rhythm is a phoneme monitoring task, a reaction time technique developed by Foss and Lynch (1969). While listening to a sentence which they must subsequently

recall, subjects must press a key whenever they hear a word beginning with a given phoneme (usually /b/).

Aaronson (1968) asked subjects to monitor a list of digits, spoken at a constant rate, for the occurrence of a target digit. There was a decrease of about 100 ms in RT to the target over the first three serial positions which occurred whether the subjects subsequently had to recall the list of digits or not. Reaction times remained steady in succeeding serial positions for subjects who had only to monitor the lists, but they increased after about the third item until the end of the list for subjects who had to recall the list as well as monitor. These results were interpreted to mean that the initial decrease in RT may reflect subjects' accommodating to the rhythm of the list.

It seems that the listener accommodates to the rhythm of an utterance on the basis of an initial pattern of stressed and unstressed syllables. In addition, rhythm has been found to be the only prosodic cue for stress perception.

## Perception of Pitch

Faure, Hirst and Chafcouloff (1980: 74) tested whether or not it is possible to perceive rhythmic patterns

in English without any accompanying variation in pitch. They attempted to neutralize the effect of pitch change in an utterance without having any effect on intensity, duration or vowel timbre. Eight ambiguous pairs (e.g., blackbird/black bird) were placed in sentence-final position in the frame: "I know very well/well it's a X." Two lists were formed with members of the 8 pairs plus other similar expressions (e.g., frying pan, empty vase). One subject recorded the lists with each expression placed in the frame: "It's a X."

A recording of one of the lists was played to a second subject in an anechoic chamber via headphones, who had to repeat each sentence as faithfully as possible. Responses were recorded, and afterward the subject listened to the sequence: "It's an empty box--I know very well it's an empty box." The subject then replied in the same way to each sentence on the recording which was played back to him. After one week, the subject repeated the procedure.

Ten English-speaking subjects (8 British, 2 American) heard a recording of the two repetitions of the 16 sentences in the context: "It's a X." in random order in Test 1. In Test 2, they heard a similar recording of the expressions in the context: "I know very well it's a X." The subjects' task was to distinguish the ambiguous expressions.

Sentences in Test 1 were decoded more easily than those in Test 2 (92% correct identification versus 58.1%). These results showed that without the acoustic information provided by pitch change, listeners obtain no better than random scores in a test on which they score 100% when pitch change information is available" (ibid.: 77). Thus, rhythmic pattern depends above all on the recognition of a sequence of stressed and unstressed syllables, and without pitch variation, stress judgments are virtually impossible. It is the recognition of a pattern of recurrent stressed syllables against a background of unstressed syllables that accounts for the fact that widely different intervals between stressed syllables appear to the listener to be "approximately equal."

Svensson (1974) showed that subjects can extract considerable information about the stress pattern of speech by virtue of its being hummed. Thus, information which is a potentially useful indicant of syntactic stress can be extracted independently of segmental information (Darwin, 1975: 179).

The phenomenon of "primary auditory stream segregation" (Bregman and Campbell, 1971) illustrates the predictive nature of the pitch contour. A random sequence of six notes (3 high; 3 low), when played rapidly, will

perceptually segment into a high and a low tune, despite the lact of any greater rhythmic cohesion within than between tunes. The analogy with speech may lie both in the use of frequency continuity of formants to help in their tracking (cf. Dorman, Cutting and Raphael, 1975) and in the use of continuity of pitch to help in attending to one voice against competing sounds (Nooteboom et al., 1978).

Vowel duration and pitch height or pitch movement can easily result in stress perception, but pitch is the stronger of the two cues (e. g., Fry, 1958; Rigault, 1962; Jassem, Morton and Steffen-Batog, 1968; Janota, 1970, cited by Nooteboom et al., 1978). The most powerful cue to the perception of accent appears to be not an increase in pitch level, as is often believed, but rather a rapid pitch rise, pitch fall, or combination of the two (Bolinger, 1958; Cohen and 't Hart, 1967). The timing of pitch movement with regard to vowel onset and end is particularly important to accent perception (e.g., a rise has to be early in the vowel, a fall late in the vowel) (Collier, 1972; Van Katwijk, 1974). The perception of a particular pitch movement in an utterance as an accent can be influenced by its position in the overall pitch contour (Van Katwijk, 1974).

Nooteboom et al. (1978) report that two types of

pitch movement, a pitch fall immediately following a potential boundary and a pitch rise immediately preceding a potential boundary, equally affected sentence perception in Dutch. The authors state, "This suggests that the pitch movements we used are less effective than syllable lengthening as prosodic boundary markers. We introduced conflicting cues, namely the syllable lengthening of one reading and pitch cues of the other. Syllable lengthening gave a correct score of 90%, easily overriding the cue value of pitch movements" (ibid.: 98)

Prosodic boundaries, which often coincide with syntactic boundaries, can be marked by syllable lengthening, speech pauses and particular pitch contours.

Ladefoged and Broadbent (1960) introduced a paradigm where clicks were inserted into recordings of either sentences or strings of random digits. They found that listeners had great difficulty in reporting exactly where the clicks occurred, and they tended to shift the subjective location of the clicks to boundaries of percetual units. These results indicated to the authors that the decoding of speech information must involve operating on units which are larger than the duration of a single speech sound.

Fodor and Bever (1965), using the same paradigm, found that clicks at major syntactic boundaries were located

correctly more often than those placed within constituents, and the direction of errors showed a strong tendency to 'migrate' toward those boundaries.

Garrett, Bever and Fodor (1966) obtained the same results even when a splicing technique was used to remove the normal intonation pattern of the sentence. They concluded that perceptual segmentation based on clausal boundaries can be demonstrated even in the absence of intonation cues which might normally accompany such boundaries.

Wingfield and Klein (1975) altered the "click" technique in two ways: (1) intonational marking was put in direct conflict with objective sentence structure, and (2) source localization was used instead of extraneous clicks.

Wingfield (1975) performed two experiments to investigate the syntax-intonation interaction. In Experiment 1, pairs of sentences were constructed that shared a common sequence, but where the major syntactic boundary (MSB) occurred in different positions within the common sequence, e.g., (a) To avoid any attempts to <u>influence voting, machines were installed</u>. (b) Due to our new mayor's <u>influence, voting machines were installed</u>.

Identical word sequences of each sentence (underlined) were cross-spliced into the linguistic frame of

the other sentence, e.g., (a) with a lexically defined MSB between voting and machines was now heard in the intonation pattern of (b) where the MSB occurred between influence and voting. Forty sentences were presented over dichotic headphones. A sentence would begin in one ear and be switched without warning to the other ear. The listener's task was to repeat the full sentence and to indicate at which point the source switched ears. Half the sentences were in normal intonation and the other half were in anomalous conflicting intonation.

Wingfield found that for normal intonation, source localization was significantly more accurate when the switch occurred at the MSB (i.e., major syntactic boundary) (93% correct) than when it occurred within a constituent either before the MSB (62% correct) or after it (59% correct). Of the errors, 44% migrated to the MSB. For anomalous intonation, there were no significant differences in accuracy of localizing switches occurring at the MSB (59% correct) than when they occurred before (53% correct) that point. It was when the switch occurred at the point marked by intonation associated with the MSB that localization was most accurate.

These results were interpreted to mean that the integration of a single message from two sources is more

influenced by segmentation imposed by the listener than by segmentation in terms of the source. Wingfield (ibid.) concludes, "Prosodic features such as intonation do not directly determine perceptual segmentation so much as add information about clausal boundaries which do."

Darwin (1975) investigated the interaction between semantics and intonation. In one experiment, he used as stimuli pairs of 50-word passages from short stories by H. E. Bates. Four recordings of each pair of passages were made: (1) 2 of the original passages, (2) 2 readings of the first part of one passage followed smoothly by the second part of the other passage.

There were 4 dichotic conditions: (1) normal (the 2 original passages aligned), (2) semantic change (other 2 original recordings), (3) intonation change (latter 2 passages switched at midpoint), (4) semantic and intonation change (2 original passages switched at midpoint). Fourteen subjects were instructed to shadow the passage on one ear with as little lag as possible between hearing the speech and saying it. They were told not to "chunk the speech into phrases."

There were significantly more intrusion errors (i.e., subject shadowed any words from the unattended ear over the break point) in both conditions when intonation

changed than when only semantics changed, and there were significantly more omissions (i.e., subject missed at least two words over the break point) when semantics alone changed than in the other two conditions. Darwin concluded that for subjects who shadow continuously, an abrupt change in the intonation contour between the ears causes intrusion errors to occur, but without any switch in intonation contour, gives omission errors.

In a second experiment, 4 sentences were derived from pairs of sentences sharing a common 6-12 word string with two words of shared material preceding and following the MSB (major syntactic boundary). Two of the 4 sentences were originals with appropriate intonation, while the other 2 were generated by cross-splicing the common string of words between the sentences. Subjects were instructed to write down each sentence as soon as they heard it. (To make the task more difficult, they were told to listen for the occurrence of stop consonants and circle them in the answer.)

On the average, 55.5% of the normally intoned sentences were recalled correctly, i.e., significantly more errors were found on cross-spliced sentences than on normally intonated sentences. For crossed intonation sentences, subjects made more errors within one word on

either side of the intonationally suggested boundary when it preceded the MSB. Darwin interpreted these results to mean that intonation is used dynamically to revise syntactic hypotheses, and the disruptive effect of an inappropriate intonational cue to the boundary causes the listener to backtract through the previous clause.

When Collier and 't Hart (1975) asked subjects to think of sentences fitting a particular pitch contour, they often placed the major syntactic boundary earlier in the sentence than the intonationally suggested boundary but very rarely later (Nooteboom et al., 1978).

## Conclusion

All the above evidence points to a model of speech perception in which prosodic feature processing (particularly of FO and duration) may result in an initial hypothesis with regard to the overall rhythm of the utterance (Wingfield, 1975). Hypotheses regarding semantic and syntactic structure proceed from processing segmental information. These hypotheses are then dynamically revised as more information becomes available regarding clause boundaries which is indicated by the intonation of the utterance.

## Models of Speech Perception

Pisoni and Sawusch (1975: 20) state, "Perhaps one of the most neglected topics of research in speech perception is prosody...For the most part, prosody and grammar have been ignored in speech perception theories. The way prosodic information might be used in the recognition process has not been considered in any detail. This is an important problem that must be considered in any theoretical account of speech production. Prosody may serve as the interface between low-level segmental information and higher levels of grammatical structure."

Based on the previous literature review on prosody, it seems that (1) prosodic aspects of speech are processed independently of segmental information (Svensson, 1974, cited in Darwin, 1975: 179); (2) the predictive nature of the pitch contour (Bregman and Campbell, 1971), anticipation of stressed syllables (Martin, 1972), and backward perceptual normalization (Nooteboom et al., 1978) indicate that prosodic information is processed early, held in memory and then used for subsequent segmental, lexical and syntactic processing; (3) there are three independent

prosodic factors that interact in speech recognition: syllable duration, pitch movement, and vowel duration. Syllable duration is more effective than pitch movement in "chunking" the speech stream into coherent "breath groups"; however, both contribute to the process (Cohen and 't Hart, 1967). Furthermore, pitch movement is more effective than vowel duration (Fry, 1958; Rigault, 1962; Jassem, Morton and Steffan-Batog, 1968; Janota, 1970, cited in Nooteboom et al., 1978) in stress perception (i.e., prediction of rhythm (Faure et al., 1980)).

Thus, in a model of speech perception, prosodic feature processing results in an initial structural hypothesis (Wingfield, 1975: 149). Based on the results of backward recognition masking experiments (Nooteboom and Cohen, 1975: 136), the first stage of speech recognition must operate on acoustic spans of longer than the syllable (Massaro, 1972: 36), possibly the breath group (Lieberman, 1965) as indicated by syllable duration and pitch movements. Auditory information may be preperceptual or precategorical, i.e., may have undergone some initial processing in terms of pitch and durations, but not yet be interpreted in terms of linguistic categories (Nooteboom and Cohen, 1975: 137).

Pisoni's Information Processing Model (see Figure 3 from Pisoni, 1982: 12) appears to incorporate the

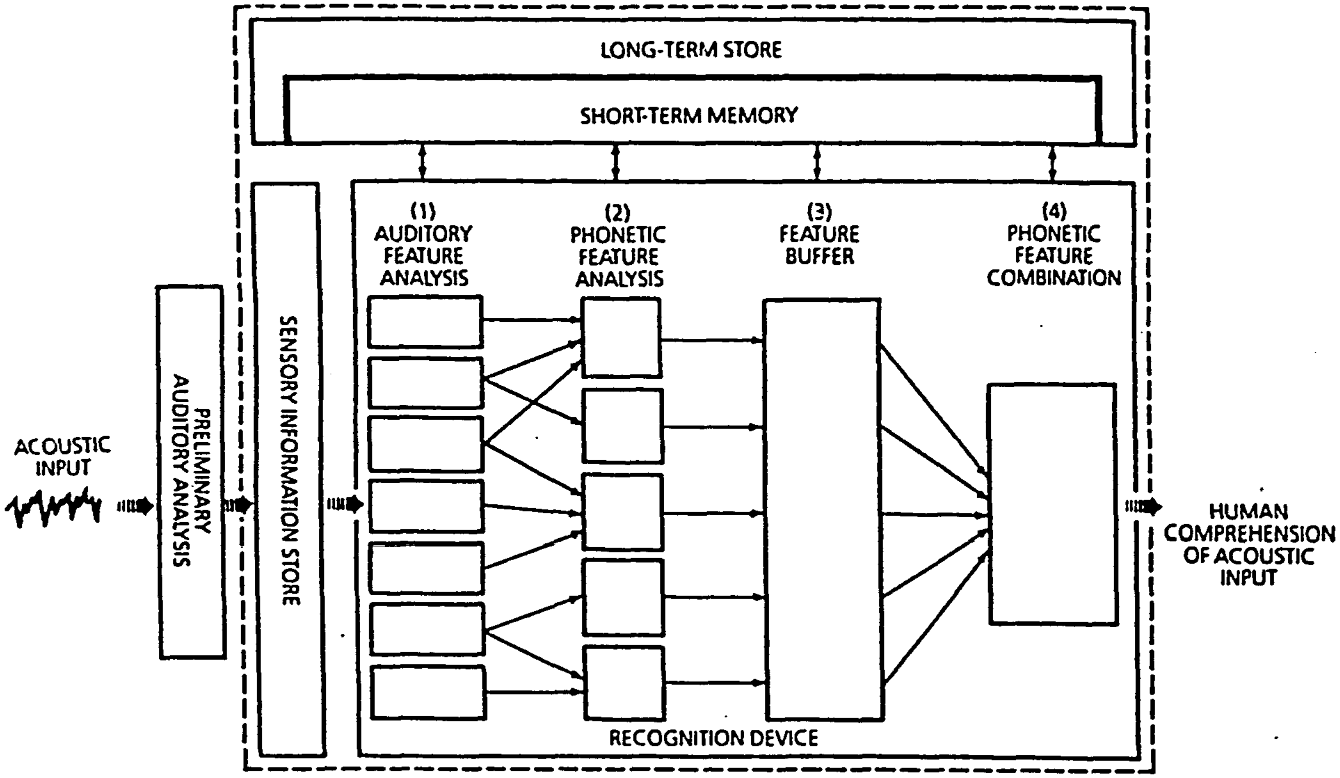


Figure 3

(from Pisoni, 1982)

above-described features better than competing models (Analysis-by-synthesis, Motor theory of speech perception, filter and template matching theories, Fant's auditory theory, and Stage theory). Auditory input enters the system and is processed in progressive stages. The output of Preliminary Auditory Analysis is assumed to be a spectral display in terms of frequency, time and intensity. Sensory input is processed automatically through several levels of analysis. Sensory information is maintained in a relatively gross unanalyzed form in the Sensory Information Store. Information is further processed by the Recognition Device which is shown at four distinct stages: (1) Auditory Feature Analysis, (2) Phonetic Feature Analysis, (3) Feature Buffer and (4) Phonetic Feature Combination. Information from any or all of these stages of processing is placed in Short-term Store where the listener can selectively rehearse, encode or make conscious decisions about it. Information in Long-term Store is also assumed to be used in the recognition process (Pisoni and Sawusch, 1975: 27).

At this stage, Oden and Massaro (1978: 172) propose a "fuzzy logical" model for the identification of speech sounds that they claim is superior to Sawusch and Pisoni's (1974) model. Oden and Massaro's model assumes "(a) the acoustic cues are perceived independently, (b) feature

evaluation provides information about the degree to which each quality is present in the speech sound, (c) each speech sound is defined by a proportional prototype in long-term memory that determines how the featural information is integrated, and (d) the speech sound is identified on the basis of the relative degree to which it matches the various alternative prototypes."<3>

Oden and Massaro (1978: 189) state that the phoneme prototype modifiers (introduced to account for more extreme degrees of presence of a feature, e.g., voicing) may help interpret the learning of dialects (and likewise, perception of foreign accent). "It is at first difficult to understand people whose dialects are strongly different than one's own. However, after a period of listening, it becomes much easier and automatic. This process of "educating your ear" might be a matter of changing the modifiers on various phonemes, that is, restructuring the prototype of perceptual units in long-term memory."

In Stage 2 of Pisoni and Sawusch's model, phonetic Feature Analysis, a set of decision rules is used to map the multiple auditory features extracted from Stage 1 into phonetic distinctive features. These features are subsequently maintained in Stage 3, the Feature Buffer, which is a storage mechanism to maintain decisions about the

feature composition of a particular syllable. Feature information is then used in Stage 4, Phonetic Feature Combination, where individual features can be recombined to form discrete phonetic segments. Information about the feature specifications of these phonetic segments is in the form of a rough distinctive feature matrix which is then made available to higher level of processing for phonological and syntactic analysis (Pisoni and Sawusch, 1975: 28).

Phonological processing consists of integrating segmental and prosodic information from the Recognition Device. The result of this analysis is a tentative string of phonetic segments, including word and morpheme boundaries. This tentative string is fed to the syntactic processor, along with prosodic information directly from the Recognition Device (ibid: 30).

The purpose of the syntactic processor is to construct a preliminary parse tree. The Augmented Transition Network (ATN) (Woods, 1973) makes specific use of memory in syntactic processing and provides a processor for applying syntactic rules, whatever their origin (see Kaplan, 1972) (Pisoni and Sawusch, 1975: 31).

The semantic processor works on information from the syntactic processor in conjunction with prosodic information

from the Recognition Device. The semantic processor draws upon knowledge of context, specific lexical meaning, and other forms of semantic information stored in long-term memory (Pisoni and Sawusch, 1975: 31).

While this type of speech perception model comes closest to one that processes prosodic features, it does not put enough emphasis on the processing of prosodic features to account for the accumulation of data on prosodic processing found in the literature.

## English versus Spanish Rhythm

Pointon (1980: 293) states: "Ever since Pike in The Intonation of American English used Spanish as his example of a typical syllable-timed language, to contrast syllable timing with stress timing, as exemplified by English, it has been generally accepted that Spanish is in fact syllable-timed, together with the other Romance languages except for Portuguese."

### Spanish Rhythm

The earliest experimental work on Spanish rhythm was carried out by Navarro Tomas (1916-1922) in which he measured absolute durations of individual segments in word lists (including short phrases) read by himself and of syllables in a passage from a poem read by three separate informants. The results obtained were unsatisfactory for the following reasons: only one informant was used for the word lists, and neither individual words and short phrases nor verse give an accurate representation of the language as a whole (Pointon, 1980: 293-294). Nevertheless, many

authors feel that Navarro Tomas demonstrated, contrary to previous assumptions, that in Spanish all syllables have similar durations. Segmental durations appear to vary because of stress and phonetic context. Accordingly, Spanish syllables can be strong or weak depending on the presence or absence of stress respectively (Manrique and Signorini, 1983: 117-118).

Gili Gaya (1940) analyzed a recorded prose passage which was divided into "phonic" or breath groups (defined by Jones (1932) as a complete sentence or longest portion of a sentence that can be effortlessly uttered in a single breath), which were in turn divided into three sections: initial (up to and including the first stressed syllable), final (the last stressed syllable to the end), and medial (everything stressed and unstressed in between). He came to the conclusion that there is a striking lack of variation in the durations of syllables in medial position (where he did not divide stressed and unstressed syllables). Pointon (1980: 295) separated the stressed and unstressed syllables in the medial section and found that their average durations were 23.5 cs and 16.9 cs respectively, a 50% increase of stressed over unstressed syllables. He says, "Gili Gaya makes the point repeatedly that the more complex the syllable structure, the greater the duration of that

syllable" (ibid.).

Delattre (1966) made spectrograms of five minutes of spontaneous speech by an undetermined number of native speakers of English, French, German and Spanish, whose geographical origins were also undetermined. Delattre divided consonants from vowels by including formant transitions within the consonant; in previous kymographic and oscillographic measurements, only closures [sic] were measured for consonants. The syllables were divided into groups showing three separate binary parameters: stress, final (in the same group), and "close" ([i,u]) versus "open" ([a]). (No account was taken of syllable structure.) Delattre found that closed syllables were considerably longer than open ones, except in unstressed non-final position where there was a ratio of only 1.06:1 between the figures. The maximum difference in duration, between stressed final closed syllables and unstressed non-final open syllables, showed a ratio of 1.77:1. The ratio of stressed to unstressed syllables was 1.30:1, and the only departure from this figure was in non-final open syllables where the ratio was 1.11:1. Unstressed, open syllables had virtually no duration difference whether they were final of non-final: a ratio of 1.02:1 (Pointon, 1980: 295-296).

Olsen (1972) analyzed one half hour of spontaneous

speech by a speaker of Mexican Spanish, dividing the material into "sense groups," (defined by Jones (1932: 254) as the shortest possible groups between which pauses may be made but are not essential) categorized by rhythmical pattern, sound sequence, structural sequence and length of sequence. It is not clear whether Olsen used the traditional cut-off points between vowel and consonant or whether he used Delattre's method of including the transitions as part of the consonant durations. "Olsen comes nearest of all previous investigators to giving us a picture of the rhythm of spoken Spanish, but by limiting himself to sense groups and ignoring pauses between them, he cannot give us more than a long series of short units which we can only look at individually, and not add together into strings to try and find a larger pattern" (Pointon, 1980: 296-297).

As Pointon (1980: 297) points out, some difference may be expected between Gili Gaya's figures and those produced by Delattre and Olsen, partly because it is certain that Gili Gaya included formant transitions in the figures for vowels and also because Delattre and Olsen took the sense group as their basic unit, whereas Gili Gaya used the breath group. Since Gili Gaya recorded a person reading a literary text, and both Delattre and Olsen used spontaneous speech, the reading style may be expected to lead to a

slower tempo, and greater durations for each syllable. Gili Gaya and Delattre conflict over the effect of stress on a syllable. Delattre writes, "non-final open syllables are only slightly longer in stressed syllables than in unstressed syllables" (p. 189) whereas elsewhere stressed syllables are approximately one third longer than unstressed syllables. One can conclude from Gili Gaya's figures that stressed syllables are about 40% longer than unstressed ones, except for final, closed syllables, where stressed syllables are only slightly longer. The conflict may be the result of the dialects of the informants, i.e., Mexican versus Castilian (Pointon, 1980: 297-298).

Based on the above research, only a very general claim can be made about the duration of Spanish syllables: that stressed syllables tend to be longer than unstressed ones, and that closed syllables tend to be longer than open ones (ibid.). Gili Gaya's figures clearly show that a classical syllable-timed rhythm is not being used, and that not only the stress, but also the number of segments per syllable and the nature of the consonant(s) in the syllable play part in determining its duration (ibid.: 300). Manrique and Signorini (1983: 126) agree, concluding from their study of Argentinian Spanish that "these data do not support the claim that Spanish is 'isosyllabic.' The fact

that in Spanish the syllablic type, CV, is so frequent, 55.9%, offers one plausible explanation for the impression of Spanish 'isosyllabicity'."

Bolinger (1961) commented that stress in Spanish, like in English, is associated with pitch variation. In recent experiments (Manrique et al., 1982; Massone, 1982; Manrique, 1982), it was determined that in Argentine Spanish, stressed syllables were higher in pitch, longer and sometimes louder than unstressed ones, except at the end of the phrase where the main acoustic correlate of stress seemed to be an extra elongation of the stressed vowel (Manrique and Signorini, 1983: 126). Dauer (1983: 58) states that "although stress can be used to distinguish words in Spanish [For example, /'tomo/ versus /to'mo/], its effects are not as great as in English. Stressed syllables in Spanish are on an average 1.3 times longer than unstressed syllables, whereas in English, they are on an average 1.5 times longer. The effect is especially noticeable in non-final open syllables, where in Spanish stressed syllables are only 1.1 times longer than unstressed syllables compared to 1.6 in English." Delattre (1966: 196) concludes "syllabic stress affects syllabic duration for every position or type of syllable, but this conditioning is relatively strong in English [and] weak in Spanish."

Dauer (1983: 57-58) repeats the commonly held belief that syllable-timed languages do not regularly have reduced variants of vowels in unstressed position. However, acoustic data obtained in a study by Manrique et al. (1982) has demonstrated that unstressed vowel areas shift toward the area of [a] for Spanish, which occupies a central position with high F1 in a F1-F2 plot. Thus, although vowel reduction in Spanish does not present the same characteristics as English, i.e., shift toward schwa, it implies a change in quality as well as duration (ibid.)

Theoretical Studies of Spanish Rhythm. The three most fully articulated theoretical studies of Spanish syllable structure are Saporta and Contreras (1962), Hooper (1976) and Harris (1983). Various non-generative studies have also treated the structure of the Spanish syllable, starting with Malmberg (1948). Granda (1966) provides a more detailed diachronic/synchronic analysis, which in part resembles that of Saporta and Contreras (1962), but is less lucidly explicated (Nuñez-Cedeño, 1984: 933).

In Saporta and Contreras (1962), the organization of words into syllables is specified basically by a phrase structure grammar which generates hierarchical structures of terminal elements. In Hooper (1976), on the other hand, syllables are organized linearly rather than hierarchically;

the descriptive device is a template consisting of a string of positions bounded at each end by the symbol $, with which are associated conditions of "consonantal strength."

Harris (1983) demonstrates that both of these works are observationally and descriptively inadequate. Furthermore, he concludes that "neither a phrase structure component nor a linear template is among the formal mechanisms that play a role in the specification of syllable structure. Rather, intrasyllabic organization is to be ascribed to (a) a set of rules that apply to strings of phonemes supplied by the lexicon, collecting groups of segments into a labeled constituent, and (b) a set of filters that mark constituents as deviant under specified conditions (Harris, 1983: 3-4).

For Harris, the syllable has two immediate constituents: a rule-constrained optional ONSET (one or more initial consonants), and an obligatory RHYME (which minimally contains a vowel and optimally a "coda" of one or more terminal consonants) in Spanish). It is presumed that onset and rhyme operate independently of each other. He supports his claim of a binary-branching syllable by arguing that the maximum permissible length of a syllable in Spanish is five segments and that antepenultimate stress is impossible if the penult is checked by a consonant

(Nuñez-Cedeño, 1984: 933-935). Of central interest is Harris' pioneering description of the absence of antepenultimate stress when the penultimate syllable has a branching rhyme (ibid.).

Concentrating only on non-verbs, Harris proposes that, universal conditions of the theory notwithstanding prosodic tree construction of Spanish is carried out according to two rules: one at the FOOT level, where rhymes are gathered from right to left into binary units labeled weak/strong and one at the WORD level, where the feet are binary units labeled weak/strong (ibid.).

There are two opposing views of sentential stress in Spanish found in the literature. The first, which is syntactically based, assumes that sentential stress is assigned on the basis of syntactic criteria. Proponents of this view include Chomsky and Halle (1968) and Bresnan (1971). The second view, which is informational, holds that sentential stress depends not on syntax, but on the informational content of the utterance. This view has been accepted by Bolinger (1954, 1958, 1961, 1972), Daneš (1960) and Hultzen (1956), among others (Contreras, 1980: 45).

The facts presented by Contreras (1980: 52-53) suggest that since "subject" is not topologically defined in Spanish, a theory that assigns an unmarked order to Spanish

sentences on a purely syntactic basis is to viewed with suspicion. Conversely, a theory that assigns word order and sentential stress on the basis of the "thematic," or "informational," structure of the sentence becomes more attractive.<4> Additional support for this alternative comes from the fact that the thematic organization is relevant for semantic interpretation and interacts with the operation of syntactic rules.

## English Rhythm

Many researchers have claimed that there is a tendency towards isochrony in English (beginning with Jones, 1932/1960). The idea that English is stress-timed has a long history in metrical theory going back to the eighteenth century (Steele, 1775: 115), yet no one has been able to prove conclusively that there is a tendency towards isochrony in English (Dauer, 1983: 52).

Many researchers, beginning with Classe (1939), have measured interstress intervals in English, and all have shown that they are objectively longer when they contain more syllables (cf. Lehiste, 1977 for a review). Some researchers claim that a tendency towards isochronism exists (Classe, 1939; Bolinger, 1965; Uldall, 1971; Lehiste, 1977)

Others have rejected isochronism in the production of English on the basis of their data (Shen and Peterson, 1962; O'Connor, 1965; Faure, Hirst and Chafcouloff, 1980; Nakatani, O'Connor and Aston, 1981) (Hoequist, 1983c: 367).

Dauer (1983: 53) measured interstress intervals "from the onset of the first stressed vowel to the onset of the next etc. from readings in five different languages including English and Spanish. He concluded (ibid.: 54), "there is no more of a tendency for interstress intervals to clump together in English than in other languages."

Hoequist (1983,a,b,c) criticizes all previous research employing the technique of ISI (interstress interval) measurement, because "it wrongly assumes that points speakers and hearers use to pick out boundaries of rhythmic units are the same points as used by investigators making measurements on visual representations of the acoustic signal." He cites the following research to support the above claim:

Tuller and Fowler (1980) made EMG measurements of lip muscle activity of speakers producing isochronous strings (strings of equal duration) of syllables beginning with different consonants. Isochrony could be found in the onset of muscle activity in the lips for such strings as /duk suk duk suk/ even though the acoustic signal displayed

systematic deviations from isochrony (Hoequist, 1983c: 368).

Fowler and Tassinary (1981) had speakers produce syllable sequences with differing syllable onsets that all had the same first segment, e.g., sad, strad, sad, strad. If speakers were aligning some part of the initial consonant gesture, the acoustic onsets should have been isochronous, but they were not (Hoequist, 1983c: 368).

Morton et al. (1976) showed that a sequence of spoken digits with evenly spaced acoustic onsets was judged uneven by listeners. If allowed to adjust the intervals between words to make them sound even, listeners produced strings that systematically varied from acoustic isochrony (Hoequist, 1983c: 368).

Fowler (1979) showed that the acoustic anisochronies produced by speakers who are uttering what they consider to be evenly spaced syllables are exactly those anisochronies (inequalities in duration) required by listeners to hear the utterance as isochronous (Hoequist, 1983c: 368).

People are capable of aligning some part of the syllable (referred to as the perceptual center or P-center) consistently, but the part aligned does not correspond to the acoustic onset of the syllable or even necessarily to an acoustically obvious portion of the syllable such as the peak amplitude or peak F0, both of which were tested and

ruled out by Morton et al. (1976) (Hoequist, 1983c: 368).

Hoequist (1983b) compares syllable duration in stress- and syllable-timed languages using Liberman and Prince's (1977) theory of stress and rhythm for English. Three basic characteristics are cited for English: (1) compensatory shortening occurs within a stress foot (stressed syllable(s) followed by unstressed syllable(s)); (2) temporal compensation occurs within words and across foot boundaries (e.g., accent adjacency, which consists of (a) a tendency to reduce stress (and thus duration) of any syllable immediately adjacent to the syllable carrying lexical stress, and (b) a further tendency to reduce medial unstressed syllables whether accent-adjacent or not); (3) accented syllables in English are shorter as a function of the number of following syllables in a word and successive increments of shortening become less as the total number of following syllables increases. (But see Lehiste, 1980.)

The following characteristics are cited for Spanish: (1) general durational effects (e.g., final lengthening and lengthening under accent); (2) overall tendency to lengthen syllables towards the end of a word. (Initial syllables are shorter than medial syllables, which in turn are shorter than final syllables); (3) lack of most temporal compensation effects found in English; (4) restricted range

of possible syllables and syllable types; (5) no discernible shortening trend as a function of the number of following syllables; (6) lack of timing constraints on duration such as the stress foot. Durational controls seem to be either general and applicable to non-syllable-timed languages or indirect and based on the language's structure (e.g., absence of vowel reduction).

Stress- and syllable-timed languages are found to be similar in the following respects: (1) accent is accompanied by considerable syllable lengthening; (2) differences in the amounts of shortening are too small to effect cross-language differences; and (3) shortening due to the number of preceding syllables occurs, although in Spanish it is restricted to stressed syllables.

In the section, "Prosody," it was claimed that syllable duration is the most prominent cue indicating prosodic boundary markers. In English, syllables before a boundary are longer; in Spanish, they tend to be longer in word-final position (Hoequist, 1983). A second difference between English and Spanish deals with the occurrence, specific to English, of "local speech rate" whereby an initial stressed syllable is shortened as a function of the number of succeeding syllables (but see Lehiste, 1980). Thirdly, vowel reduction to schwa in English contrasts with

a different type of reduction, i.e., to [a], in Spanish (at least in Argentinian Spanish) (Manrique et al., 1982).

## English versus Spanish Intonation

### Spanish Intonation

Intonational differences have been documented not only for diverse dialects of Spanish (e.g., Castilian versus Mexican), but also for closely related dialects (e.g., Buenos Aires versus Cordoban versus Tucuman). Presentations such as Stockwell, Bowen and Silva-Fuenzalida (1956) and Cardenas (1960) (and others) are based on a universal pan-Spanish, that is, on informants from many different Spanish-speaking areas so that dialect differences are blurred (Kvavik, 1976: 408).

Experimental research on Spanish intonation is very scarce. Delattre, Olsen and Poenack (1962: 235) give specific intonational contours for Mexican Spanish and American English in a "spontaneous speech style using records of Margaret Mead and Diego Rivera. They conclude, "in a comparison of Spanish and American continuation: that of general direction--mainly rising for Spanish, mainly falling for American" (Kvavik, 1976: 410). Note that Mexican Spanish is widely known to be atypical of Spanish

dialects with its "sing-song" intonation.

Navarro Tomas (1966) presupposes three levels plus five final "tonemas" (i.e., sense group finalities); the melodic groups as a whole have five basic shapes: fall, semifall, level, semirise, and rise. His research is based on a reading style of educated Castilian speakers and verified with kymographic evidence (Kvavik, 1976: 406-407). Note that Castilian Spanish is fairly representative of Spanish dialects in terms of intonation.

Kvavik (1976: 412-413) examines intonational differences and similarities of sentence-initial and final intonation between Castilian and Mexican. "The phonetic information is based on five and one-half minutes of conversational speech of four male Castilian and four male Mexican speakers" (ibid.: 411). Kvavik's results show that the Mexicans range from -1 to +8 Hz above their normal tone, while the Castilians range +1 to -9 Hz below the normal tone for the level finality.

An analysis of Mexican sense-group terminations shows that there are: SIMPLE or unidirectional configurations, which have specific frequency intervals associated with them, as well as COMPLEX configurations (rising-falling or circumflex, falling-rising, terraced). COMPLEX intonations are observed fairly frequently in the

conversational style of educated Mexican speech (ibid.: 415).

## English Intonation

*Theoretical Studies*. According to Lieberman (1967: 171-172), there have been three theoretical approaches to English intonation since 1900. The first approach, typified by the British school of thought identified with Daniel Jones, makes use of suprasegmental "tunes" that, on the acoustic level are quite similar to the breath-group. The second approach, largely identified with British phoneticians, has described pitch contours by means of "tones" that occur on specific vowels. The sequence of tones that may rise and fall determines the intonation pattern of the utterance. The third approach, developed by American linguists who tried to apply the segmental techniques of taxonomic phonemics to intonation, has analyzed intonation in terms of segmental pitch levels, stress levels and junctures. The segmental elements have been grouped into suprasegmental "phonemic phrases," "phonemic clauses" and "suprasegmental morphemes."

Henry Sweet's *New English Grammar* (1892) is the basis of all subsequent "tone" analyses. Sweet set up three

degrees of stress or loudness: strong, half strong, and weak. Intonation is either level, rising or falling. The level tone may be either high or low and the other tones may begin in a high or low pitch. The non-level tones can pass through different intervals; the greater the interval, the more emphatic the tone becomes. There are also "compound" intonation patterns: falling-rising and rising-falling, which could start either at a high or low pitch. Sweet seemed to equate intonation with the pitch or fundamental frequency and stress with perceptual loudness. Lieberman (1965) has shown that Sweet's intonation notation can be used to transcribe the intonation patterns of English utterances quite accurately (Lieberman, 1967: 172-173).

L. E. Armstrong and I. C. Ward's *Handbook of English Intonation* (1926) defines "stress" in terms of breath force and "intonation" in terms of perceived pitch. Armstrong and Ward (1926: 3) comment "a word spoken in isolation may have a certain stress pattern. However, "in connected speech this word stress is often dropped."

Two "tunes" are defined: Tune I essentially starts on a medium pitch and continues on this pitch with some upward variations on stressed syllables until the end of the sentence, when the pitch falls rapidly. Tune II starts at either a high or medium pitch and gradually falls, but it

ends with rising or level pitch. Intonations are divided into "sense-groups" and each "sense-group" is an intonation group that may have the contour or Tune I or II. In short sentences that have only one "sense-group" Tune I is used for statements and imperatives; Tune II is used for some interrogative yes-no questions and for sentences in which the speaker wishes to imply uncertainty. Longer sentences may consist of many sense-groups that do not terminate the sentence. According to Lieberman (1967: 177-178), Armstrong and Ward isolated the acoustic and perceptual manifestations of the breath-group. The two tunes are both suprasegmental where Tune I is equivalent to the unmarked breath-group, while Tune II is equivalent to the marked breath-group.

Daniel Jones (1932), in An Outline of English Phonetics, adopted the Tune I and Tune II of Armstrong and Ward and expands on it. He differentiates between sense-groups and breath-groups, and defines intonation as perceived pitch. Prominence, stress and intonation are related as follows:

> In every spoken word or phrase there is at least one sound which is heard to stand out more prominently than sounds next to it (Jones, 1932: 55).

The prominence of a given sound may be increased or

> diminished by means of the three sound attributes, length, stress or intonation, or by combinations of these. A common and effective means. . .is to increase stress. In English increase in stress is generally accompanied by a modification of intonation and sometimes by an increase of length (ibid.: 228).
>
> It is important not to confuse stress and prominence. The prominence of a syllable is its general degree of distinctness, this being the combined effect of the timbre, length, stress and intonation of the syllabic sound. The term "stress" refers only to the degree of force of utterance, it is independent of length or intonation, though it may be combined with these (ibid.: 228).

Jones' subjective evaluations of the acoustic correlates of stress, intonation and prominence have, for the most part, been substantiated by psychoacoustic and acoustic experiments (cf. Lieberman, 1967: chapter 2).

The principal effects of Leonard Bloomfield's work (1933) were to channel subsequent studies toward the isolation of "pitch phonemes" and explicit characterization of their role in defining syntactic constructions (Lieberman, 1967: 181).

The two most significant aspects of Pike's analysis (1945) are his isolation of the "pauses" that occur at the end of "rhythm units" and his observation that the acoustic modifications that reflect emotion affect the entire utterance rather than only a part of it. Pike's approach resembles Armstrong and Ward's analysis (Lieberman, 1967: 187).

G. L. Trager and H. L. Smith's analysis (1951) resembles Pike's analysis. They use four phonemic pitch levels with number 1 corresponding to the speaker's lower relative pitch and number 4 to his highest. There are three terminal junctures that correspond to Pike's two "pauses." The terminal juncture [#] phonetically means a fall in pitch and corresponds to Pike's "final pause." The terminal symbols [//] and [/], represent a rise in pitch and a sustaining of pitch respectively and correspond to Pike's "tentative pause," which either sustained the height or the pitch level that preceded it or involved a "slight drift unwards." "Internal juncture," /+/ has the principle function "to indicate phonetic cues that separate words or parts of certain compound words during fluent speech. There are four "phonemic" stress levels, /´^`ˇ/: primary stress, /´/, is the loudest, defined in psychoacoustic terms, secondary stress, /^/, occurs in compound words or phrases

where "internal juncture," /+/, occurs (Lieberman, 1967: 188).

One of the fundamental differences between Pike and Trager and Smith is that Pike differentiated between the "pitch contours" that could potentially occur in a sentence and the "rhythm units" which were the pitch contours that actually occurred. Trager and Smith discuss only those prosodic patterns that actually occur. The primary function of the prosodic features is to divide the sentence into linguistic units. Each linguistic unit is always represented by a prosodic pattern or class of prosodic patterns that is actually present in the acoustic speech signal (Lieberman, 1967: 189).

Pike notes that there are potential pitch contours that may be realized by a speaker to make the meaning of the sentence clearer. Trager and Smith, however, tie the pitch contours directly to the Immediate Constituent structure, implying that these pitch contours are always physically realized. The suprasegmentals always provide "acoustic" cues that tell the listener how to divide the sentence for syntactic analysis. The Trager and Smith analysis of intonation is the logical extension of Bloomfield's hunt for "objective" facts related here to the level of immediate constituents (ibid.: 189).

The Trager-Smith analysis correctly noticed that the intonation of an utterance could reflect its immediate constituent structure. Trager and Smith refined the segmental analysis of intonation; the "phonemic clause" was defined in terms of segmental pitch phonemes that were used to transcribe the utterance (ibid.: 190).

Trager and Smith saw that intonation patterns of certain utterances could change their "meanings," but they assumed that the difference in meaning was somehow part of the intonation itself. Since intonation contours can have many different meanings, it must be possible to form many different intonation "morphemes." The only elements of the Trager-Smith system that seem to have a reasonably consistent physical basis are the "morphemes 231# and 232// (or 232/), which are equivalent notations for the unmarked and marked breath-groups respectively (ibid.).

Experimental Studies. Lieberman (1965: 52) showed that the pitch levels and stresses of the Trager-Smith system often have no physical basis. He found that when two competent linguists independently transcribe a set of sentences that include "emotional" as well as "normal" utterances, 60% of the pitch levels and junctures of the two Trager-Smith transcriptions vary. The Trager-Smith levels

did not correspond to discrete relative ranges of fundamental frequency even for the transcriptions made by a single linguist who carefully transcribed the tape-recorded sentences of a single talker. Lieberman concludes, "The pitch levels of the Trager-Smith system do not even reflect the relative pitch levels of a single utterance of a single talker when it is transcribed by a single linguist.

When the fundamental frequency and amplitude contours of the complete sentences were accurately reproduced as modulations of a fixed vowel, each linguist changed 50% of the pitch levels and junctures of his transcription vis-à-vis his transcription of the complete sentence where he heard the words (ibid.: 53)

When the linguist heard the complete speech signal he was able to transcribe four degrees of stress. However, when he heard the fixed vowel, he was unable to transcribe accurately more than two degrees of stress: stressed and unstressed. These results suggest that only two degrees of stress may have acoustic correlates independent of vowel quality (ibid.).

Stetson's (1951) study is "in short, misleading in its conception of the articulatory maneuvers involved in inspiration and production on speech. Its data also are often erroneous." Lieberman (1967: 193) believes that its

principle value is that "it focused some attention on the lingustic role of subglottal articulatory maneuvers. "The only aspect of Stetson's theory that seems to have a firm basis is the subglottal air pressure function that is associated with the breath-group, that is, a positive air pressure function normally associated with an expiration bounded by inspirations" (ibid.).

Lieberman (1967) has claimed that intonation is produced in terms of an unmarked and a marked breath-group, and certain segmental features that interact with the breath-group because of the inherent constraints imposed by the human speech production apparatus and auditory system. The breath-group is a suprasegmental feature whose scope is usually a sentence (Lehiste, 1970: 97).

Lieberman (1967) contends that for the unmarked American English breath-group, the tension of the laryngeal muscles appear to remain relatively steady throughout the sentence. The fundamental frequency of phonation is thus a function of the subglottal air-pressure, and it falls during the last 150-200 msec of phonation. The marked breath-group contrasts with the unmarked breath-group during the last 150-200 msec of phonation; the tension of the laryngeal muscles increases in the marked breath-group during this terminal phase. The increased tension of the laryngeal

muscles counters the falling subglottal air pressure, and the marked breath-group thus has a terminal nonfalling fundamental frequency contour (Lehiste, 1970: 98).

Lieberman's hypothesis, stated above, has been challenged by Ohala and Hirano (1967). Their electromyographic data provided no confirmation of the hypothesis that, for American English, during other than yes-no questions, the laryngeal tension remains relatively steady and that pitch variation is a function of the subglottal air pressure. On the contrary, the muscles studied participated actively in pitch control, and variations in subglottal pressure could account for only a fraction of the observed pitch changes" (Lehiste, 1970: 98).

In recent years, studies like those of Maeda (1976), Pierrehumbert (1979) and Cooper and Sorenson (1981) have proposed that fundamental frequency contours of isolated sentences and sentences in connected speech can be described in terms of global declination lines. The total FO contour is characterized in terms of a line that extends across the total sentence or a major constituent. Global "baselines" or "toplines" are constructed to approximate the trend in time of either the valleys or the peaks of FO in a linear relationship that has a negative slope. Furthermore, the falling declination lines are produced as the salient

acoustic cues that listeners use to segment the flow of speech into sentences or to mark out the utterances' constituent (grammatical) structure. Declination theories thus claim that baselines or toplines that gradually fall throughout the course of a sentence describe linguistically salient aspects of the FO contour. These claims differ from those of earlier studies like those of Armstrong and Ward (1926), Trager and Smith (1951), and Lieberman (1967), which claim that the FO contour of a sentence consists of two parts: a nonterminal part, and a terminal part, that have different slopes. Lieberman (1967) accounted for the two-part contours via a "breath-group theory of intonation" (Lieberman, Katz, Jongman, Zimmerman and Miller, 1985: 649).

The technique that has been used to determine the descriptive adequacy competing theories is subjective, involvinge visual inspection of an FO contour. Maeda (1976) introduced a "visual abstraction" procedure in which baselines drawn through the valleys of a plot of FO were estimated by eye. This procedure is present or implicit in subsequent theories involving declination (e.g., Vaissiere, 1974; Delgutte, 1978; Garding, 1979)" (Lieberman, et al., 1985: 649).

Lieberman et al. (1985) tested the visual abstraction procedure using 12 subjects who each fit the FO

contours of 19 spoken short simple sentences with baselines. These baselines were found to be poorly replicated by the fitters. An objective all-points least-squares best-fit procedure was tested on this corpus and on a set of sentences that had been produced in both spontaneous and read speech by 6 speakers. The all-points linear regression line was a better descriptor of the FO contours than either baselines or toplines (Lieberman, et al., 1985: 649).

Previous studies (cf. Lieberman and Tseng, 1981; Umeda, 1982) have noted greater variability and less declination in spontaneous speech than in read speech. Lieberman and Tseng (1981), Lieberman et al. (1982), and Tseng (1981) note that spontaneous speech FO contours more often fit the breath-group model of intonation than declination theories (Lieberman et al., 1985: 654).

At the descriptive level, the differing claims of the breath-group and declination theories depend upon how FO falls in the course of a simple declarative sentence. Declination theories claim that there is a gradual descent over the course of an intonation group (Pierrehumbert, 1979; Cooper and Sorenson, 1981). The breath-group model (Lieberman, 1967) claims that the FO contour consists of two elements: a nonterminal segment or phase and a terminal segment or phase. The terminal is basically defined as a

final rapid rise or fall in FO; terminals appear to function as linguistic cues to mark out the utterances' phrase structure (Trager and Smith, 1951; Lieberman, 1967; Vanderslice and Ladefoged, 1972; Atkinson, 1973). The nonterminal portion of the breath-group, in its most basic form, tends to be level. It can, but does not necessarily, have a downwards slope (as declination theories claim) (Lieberman et al., 1985: 654).

Lieberman et al. (ibid.) explain: "Declination theories claim that the downwards slope of the FO contour throughout the sentence is a psychologically "real," universally present, acoustic cue that listeners track in order to segment the flow of speech into syntactically meaningful segments (e.g., Maeda, 1976) or to assign FO peaks to phenomena such as linguistic stress (e.g., Pierrehumbert, 1979). The breath-group theory, like earlier theories (e.g., Armstrong and Ward, 1926; Jones, 1932; Trager and Smith, 1951), instead claims that the terminals are used by listeners to segment the flow of speech into syntactically meaningful segments."

The linear fit of the all-points lines to the FO contours provides the basis for objective comparison of the difference in declination between spontaneous and read speech. Lieberman et al.(1985: 654) compared the slopes of

the all-points linear regression lines according to whether they were derived from spontaneous or read speech samples. The authors (ibid.: 649) conclude: "There was more variation in the FO contours of sentences from spontaneous speech; the slopes of the FO contours derived from spontaneous speech tended to depart from the declination model more often than FO contours derived from read sentences. 35% of the spontaneous sentences did not show declination; 45% of these sentences better fit the breath-group model of intonation. Their FO contours could be described by a level all-points linear regression line followed by a falling terminal segment.

Sample

Fundamental frequency of standard English and moderate and strong Spanish accents in English for the sentence, "Kansas State's new basketball tickets cost ten dollars now," was ascertained via a pitch extraction algorithm (Gold and Rabiner, 1969). Consider the following differences in intonation contours:

STANDARD ENGLISH

['kaenzəs stets nu 'baeskətbɔl 'tIkəts kɔst tEn 'dalərz naU]

MODERATE ACCENT

['kaenzəs stets nu 'baeskətbɔl 'tIkəts kɔst tEn 'dalərz naU]

STRONG ACCENT

['kaenzəs stets nu 'baeskətbɔl 'tIkəts kɔst tEn 'dalərz naU]

Normalized values in Hertz per point in time (Time, Hertz) are given below for a production of the sentence by each speaker. (The sentence has been divided into three sections for ease in subsequent synthesis).

STANDARD ENGLISH

['kaenzəs stets]:

0,120,75,120,85,150,300,150,355,140,780,140

[nu 'baeskətbɔl tIkəts]:

0,140,35,120,180,120,345,155,690,110,825,110,895,120,1125,120,1175,120;5,120;

[kɔst tEn 'dalərz naU]:

0,120,350,120,360,130,575,130,765,150,805,150,1000,80,1280,80;

MODERATE ACCENT

['kaenzəs stets]:

0,145,70,145,85,145,195,133,495,133,500,139,740,139;

[nu 'baeskətbɔl tIkəts]:

0,116,35,145,90,145,95,116,180,116,345,150,465,150,505,116,605,116,690,104,720,104,840,116,845,116,905,139,970,139,1025,127,1110,127;

[kɔst tEn 'dalərz naU]:

0,127,440,127,445,116,570,116,575,127,765,127,925,110,930,110,995,110,1000,127,1280,104;

STRONG ACCENT

['kaenzəs stets]:

0,121,70,121,195,146,230,146,230,146,235,133,290,133,295,133,355,121,495,121,500,140,740,140;

[nu 'baeskətbɔl 'tIkəts]:

0,140,90,109,175,109,180,121,345,152,460,152,465,146,505,121

,600,121,605,133,690,121,840,121,845,133,1110,133;

[kɔst tEn 'dalərz naU]:

0,139,355,139,360,127,575,127,925,121,930,109,995,109,1000,1
21,1040,121,1280,97;

## Segmental Differences in Spanish and English

There is increasing evidence that detection of accentedness may be triggered by differences between native and non-native speakers in segmental articulation. According to Flege (1984a), listeners do not automatically or irreversibly "filter out" the subcategorical differences in segmental articulation which may distinguish similar sounds in L1 and L2 (e.g., production of /t/ with short-lag rather than long-lag VOT values.)

### Voice Onset Time

Voice onset time (VOT) is a temporal measure of the relative onset of glottal and supraglottal articulatory events. It is measured as the time in milliseconds, from the release of stop constriction and the beginning of periodic vibration of the vocal folds. Cross-language research (e.g., Lisker and Abramson, 1964) has shown that pairs of homorganic stops in most languages differ in terms of VOT. The VOT values associated with a single category (e.g., /p/) are relatively stable within a language and

serve to distinguish it from members of other categories (e.g., /p/ versus /b/) (Flege, to appear: 112).

In languages like Spanish, the VOT values for /b,d,g/ are assigned negative values, indicating that voicing begins before stop release. When /b,d,g/ are realized with "lead" VOT values (i.e., when voicing begins before stop release), /p,t,k/ are usually produced with "short-lag" VOT values. In short-lag stops as in Spanish (which are assigned small positive VOT values) phonation occurs soon after stop release. The voiced stops /b,d,g/ in English are produced with no lag or short lag VOT values. English /p,t,k/ are realized as "long-lag" stops produced with significantly longer (positive) VOT values than in languages like Spanish (ibid.). Long-lag stops are interpreted as "aspirated" and short-lag stops as "unaspirated."

Production studies. Lisker and Abramson (1970) found that the /b-p/ phoneme boundary of monolingual native English speakers occurred at about 25 ms. Their subjects produced /b/ with VOT values shorter than 25 ms, and /p/ with VOT values greater than 25 ms. Lisker and Abramson (1970) and Abramson and Lisker (1973) found that the /p-b/, /t-d/ and /k-g/ phoneme boundaries of bilingual native-Spanish speaking subjects occurred at about 20 ms.

These subjects prevoiced /b/ and produced /p/ with short-lag VOT values of 0-20 ms.

Flege and Hammond (1982) performed an experiment in which English-speaking subjects who were familiar with Spanish-accented English were asked to read several sentences containing a word-initial /t/ with a "typical Spanish accent." Subjects' imitations of Spanish accent produced /t/ with VOT values about 40 ms. shorter than native English subjects reading the same sentences without instructions to imitate a foreign accent. This study indicates that English speakers perceive shorter VOT values as cues to a Spanish accent. However, it should be noted that subjects may have been increasing tempo in their imitations, which would result in shortened closure durations.

Amount of experience seems to have little effect on the extent to which adults approximate the L2 norms for VOT in prevocalic stops (Suomi, 1980; Port and Mitleb, 1983; Flege and Hillenbrand, 1984). For example, Flege and Port (1981) found that a group of Saudi Arabians who lived an average of 9 months in the U.S. produced stops with the same VOT values as Saudi Arabians who had lived an average of 39 months in the U.S. Both groups produced /p,t,k/ with VOT values that were significantly shorter than those produced

by native English speakers.<5>

Williams (1977b) measured stops produced by bilingual adults who began learning their L2 (Spanish or English) around the age of 6 and whose accent in both L1 and L2 was demonstrated to be native-like. The bilingual subjects produced /p,t,k/ in Spanish words with VOT values that did not differ from those of monolingual native speakers of Spanish. They also produced /p,t,k/ in English words with VOT values that did not differ significantly from those of monolingual native speakers of English.

In summary, L2 speech production studies have indicated that (1) relatively inexperienced L2 learners often produce L2 stops with inappropriate VOT values that closely resemble values for similar stops in L1. (2) Inexperienced L2 learners seem to produce L2 stops with VOT values that are too short for L2 when /p,t,k/ are realized with shorter VOT values in L1 than L2, and with VOT values that are too long for L2 when the VOT of /p,t,k/ is longer in L1 than in L2. (3) L2 learners who are relatively more experienced have been observed to modify the VOT value in similar L2 stops; however, they seem generally to produce L2 stops with "compromise" VOT values (Williams, 1980) which are intermediate in value to stops produced by monolingual native speakers of L1 and L2. (4) It appears that even

highly experienced L2 learners seldom match L2 native speakers in producing the VOT in similar L2 stops.

Perception studies. The perceptual effect of VOT closely mirrors differences observed in the production of voiced and voiceless stops (even at a very early age).

Results from a study by Flege and Hammond (1982) suggested that native English speakers are able to distinguish short-lag tokens of /t/ produced in English words by native speakers of Spanish from tokens of /t/ produced with long-lag VOT values in English words produced by native speakers of English. Flege (1984a) found that listeners were able to distinguish the /t/ produced in English words by native French and native English speakers. Thus, native speakers seem to be able to detect "distortions" of sounds as well as detect subtle phonetic differences between dialects of their native language.

Elman et al. (1977) studied the perception of stop consonants by Spanish learners of English. Monolingual English speakers heard naturally produced stops with VOT values of about 20 ms as /b/ whereas monolingual Spanish speakers heard them as /p/. Spanish-English bilinguals heard considerably more of the short-lag stops as /b/ than the Spanish monolinguals.

Williams (1977a,b,1979,1980) examined the perception

of English /p,b/ by native Spanish speakers who had learned English. Each of several groups of Puerto Rican children showed enhanced discrimination of stimuli straddling the English phoneme boundary. The magnitude of the peak increased as a function of length of residence on the U.S. mainland. An enhanced discrimination of stimuli straddling the Spanish phoneme boundary between /p/ and /b/, on the other hand, seemed to diminish with age.

Williams (1979;1980) examined the perception of a /pa/ to /ba/ continuum by monolingual English-speaking children and native Spanish-speaking children. Puerto Rican children aged 8-10 yrs. and 14-16 yrs. had phoneme boundaries at significantly shorter VOT values (8 ms and 5 ms respectively) than the English children (20 ms). Children who had lived on the U.S. mainland more nearly resembled the English monolingual children (a mean boundary of 10 ms) than those who had lived less than 6 mo. on the U.S. mainland (3 ms).

Williams (1977b) also examined the perception of adults who pronounced Spanish and English with native-like proficiency. Three showed a phoneme boundary near the Spanish monolingual boundary of -4 ms, while the remaining five showing a phoneme boundary close to the phoneme boundary established for monolingual native English speakers

(25 ms). In a discrimination test, the subjects who had a Spanish-like phoneme boundary showed enhanced sensitivity to pairs of stimuli straddling both the Spanish and English phoneme boundaries. Subjects with an English-like phoneme boundary, on the other hand, showed enhanced discrimination only for pairs of stimuli which straddled the English phoneme boundary. (This suggests to Flege (to appear) the possibility that as L2 learners acquire sensitivity to new acoustic distinctions they may lose sensitivity to the acoustic dimensions distinguishing stops in L1.)

Williams' (1977a) monolingual Spanish speakers revealed a sharp peak in percent correct discrimination at about the Spanish phoneme boundary (-7 ms) and a second, less prominent, peak near the English phoneme boundary (around 20 ms).

The two-peak pattern is also evident in the discrimination data reported by Abramson and Lisker (1973) for native speakers of Spanish who learned English. A close inspection of individual data indicates a prominent discrimination peak near the English boundary (about 20 ms) and less prominent peaks near the phoneme boundary (-7 ms) for monolingual Spanish speakers in Williams (1977a).

Data reported by Williams (1977a) indicates a close match between production and perception for monolingual

native Spanish speakers. The /p-b/ phoneme boundary for several groups of Latin Americans was about -7 ms. In addition, these subjects produced Spanish /b/ with lead VOT values shorter than -7 ms, and /p/ with short-lag VOT values that were longer than -7 ms.

Elman et al. (1977) examined the perception of stop consonants by Spanish learners of English. Monolingual English speakers heard naturally produced stops with VOT values of about 20 ms as /b/ whereas monolingual Spanish speakers heard them as /p/. Spanish-English bilinguals heard considerably more of the short-lag stops as /b/ than the Spanish monolinguals.

It follows from the results of these and other studies that monolingual English speakers will perceive the production of voiceless stops by Spanish speakers of English as "accented." It is further hypothesized that they will be able to perceive degree of accentedness ("moderate","strong") depending on how much shorter than English the VOT values are.

Sample

In a brief preliminary study, two Spanish speakers (from Cali, Colombia) and one native speaker of English (all

males) were taperecorded speaking the sentence, "Kansas State's new basketball tickets cost ten dollars now." The voiceless stops of interest were [k] in ['kaenzəs], [t] in ['tIkəts], [k] in [kɔst], and [t] in [tEn]. From the second of three repetitions, the voice-onset time for each of these stops was measured on spectrograms in milliseconds from the release of stop constriction to the beginning of periodic vibration of the vocal folds (i.e., voicing). Values in milliseconds are given in Table 2 rounded to the nearest 5 ms.

---

TABLE 2. Voice-Onset Time Values

| STOP | STANDARD ENGLISH | MODERATE ACCENT | STRONG ACCENT |
|---|---|---|---|
| ['kaenzəs] | 70 | 45 | 20 |
| ['tIkəts] | 50 | 30 | 10 |
| [kɔst] | 40 | 35 | 10 |
| [tEn] | 125 | 30 | 20 |

---

These data show successive approximations to standard English values of moderately and strongly accented productions.

Vowel Quality

Most often the study of vowel quality in different languages is explained as the substitution of an L1 segment for a target segment in L2.

Sawyer (1975) states that one source of difficulty in English for speakers whose L1 is Spanish is attaining the proper quality distinctions in a vowel system which has twice as many phonemic contrasts. The following vowel substitutions are noted by Sawyer in her corpus of San Antonio bilingual English:

/i/ for /I/. The high-front lax vowel occurred in San Antonio English in <u>six</u>, <u>this</u>, <u>slip</u>, <u>wind</u>, etc. The Latin bilinguals achieved the /I/, indicating partial mastery of the /i/ vs. /I/ contrast. The monolinguals usually used the phones [i] and [i:] in such words as <u>pig</u>, <u>tin</u>, <u>him</u>, <u>chicken</u>, <u>skillet</u>, and <u>kitchen</u>. The same phones also occurred from time to time in the speech of the Latin bilinguals. In one reading, two bilinguals "corrected" themselves in pronouncing <u>slip</u>; [i] also occurred in <u>kitchen</u> in the speech of the bilinguals.

/ae/. The low-front open vowel occurred as in San Antonio English in <u>bath</u>, <u>calf</u>, <u>dance</u>, <u>cattle</u>, <u>pallet</u> in the speech of all the Anglos, usually as the dialect variant

[aeI], although [ae] occasionally occurred also in catch.

/ae/ occurred in the speech of Latin American bilinguals without the high off-glide, as [ae^], but the monolinguals used [E], indicating interference with the Spanish phoneme /e/. [E] occurred in daddy, man, candy, at, handle, pants and bag. At other times the low-central Spanish /ɑ/ occurred in Saturday, apple, Latin, bath, pantry, aunt.

/ɑ/. The low-central or low-back open vowel occurred as in San Antonio English in Mama, father, hospital and vomit in the speech of both Anglos and Latins. In words with historic short "o" such as rock, pot, slop, crop and slop, [ɑ] occurred generally. But all the Anglo speakers also had the more rounded phone [ɒ] in some of the words as well as in palm, wasp, squash, wash, double, water, and swamp.

Latin informants frequently substituted [ɑ] of the low-central Spanish phoneme where [a] occurred in Anglo San Antonio English (e.g., right, night, etc.). The same phone was also found for /ae/ and /ə/, illustrating the Latin informants' difficulty in mastering a highly differentiated vowel system.

In Latin English the interference in the lower part of the vowel spectrum produced a rather chaotic situation.

[ɔ], an allophone of Spanish /o/, occurred in <u>on</u>, <u>hospital</u> for three informants; [ɑ] occurred in <u>on</u> for one informant; and [ɑ] in <u>calm</u> for one informant. [ɑ] occurred in <u>water</u> and <u>wash</u> for all Latin informants although the backed, rounded phone [ɔ] was customary in San Antonio Anglo speech.

Turning to another study, Flege and Hammond (1982) used a delayed mimicry paradigm to probe the specific knowledge of native English speakers concerning Spanish-accented English. The subjects, 50 University of Florida students enrolled in first-year Spanish classes, were asked to read sentences with what they considered a "typical Spanish accent." The frequency of sound substitutions in key words was tabulated.

Partial results include /i/ for /I/ 42% of the time and /u/ for /U/ 20% of the time. According to the authors, the fact that /i/ for /I/ substitutions were produced twice as frequently as /u/ for /U/ substitutions suggested that the subjects were more aware of the former than the latter.

Brennan and Brennan (1981b) found that the frequency of /i/ for /I/ (but not /u/ for /U/) substitutions was directly related to how negatively native English speakers rated native Spanish speakers of English on "status" dimensions such as "wealth" and "level of education." In other words, it seems that the production of [i] for [I] is

more stigmatized than that of [u] for [U].

Well over half the vowel pronunciation errors in a large corpus reported by Hammond (1982) were related to stress. He stated that Spanish learners frequently substituted Spanish vowels (e.g., /a,i,o/ for the schwa-like vowels in the unstressed syllables of English.

In a second preliminary study, I attempted to demonstrate that Spanish speakers do not, in fact, substitute Spanish vowels for English vowels when they speak English. The taperecordings of the two Spanish speakers and the native English speaker from the previous preliminary study speaking the sentence, "Kansas State's new basketball tickets cost ten dollars now," were used. The stressed vowels of interest were [æ̃] in ['kæ̃nzəs], [ae] in ['baeskətbɔl], [I] in ['tIkəts] and [ɔ] in [kɔst]. The unstressed vowels of interest were [ə] in ['kaenzəs], ['baeskətbɔl] and ['tIkəts] and [ɚ] in ['dalɚz]. Table 3 gives the normalized formant values for these vowels averaged over three repetitions as spoken by the three informants, the native English speaker, the Spanish speaker with a moderate Spanish accent in English (CG), and the Spanish speaker with a strong Spanish accent in English (WA):

TABLE 3. Formant Frequencies for English Vowels

STANDARD ENGLISH (HB)

| | F1 | F2 | F3 |
|---|---|---|---|
| [ae͠] | 666-715-480 | 1994-1540 | 2695-2850 |
| [ae] | 620-650 | 1100-1600 | 2150-2470 |
| [I] | 400-460 | 1600-2000 | 2600-2000 |
| [ɔ] | 575 | 1010-2010 | 2300 |
| [ə] | 385 | 1550 | 2620 |
| | 620 | 1220 | 2550 |
| | 460-575 | 1850-1010 | 2155-2385 |
| [ɚ] | 470-270 | 1050-1310 | 2620-1696 |

MODERATE ACCENT (CG)

| | F1 | F2 | F3 |
|---|---|---|---|
| [ae͠] | 804-359 | 1916-1736 | 2454 |
| [ae] | 532-712-671 | 984-1505 | 2674-2720 |
| [I] | 399-179 | 2182-2524 | 3299-3206 |
| [ɔ] | 579 | 1169-1447 | 2662 |
| [ə] | 579 | 1562 | 2720 |
| | 532 | 1916 | 2587 |
| | 359 | 2141-1869 | 2720-2807 |
| [ɚ] | 532 | 1424 | 2540-2587 |

STRONG ACCENT (WA)

| | F1 | F2 | F3 |
|---|---|---|---|
| [ãe] | 822-654 | 2056-1962 | 2800 |
| [ae] | 560-794-561 | 1402-2148 | 3180 |
| [I] | 374-187 | 1308-1028 | 2616-3180 |
| [ɔ] | 607 | 1869-1638-1869 | 2791 |
| [ə] | 654 | 2056 | 3271 |
| | 467 | 2056 | 3271-3180 |
| | 467 | 2524 | 3299 |
| [ɚ] | 561 | 1869 | 3271 |

---

The Spanish speakers' vowels differ in formant structure from that of the native English vowels in the following ways:

[ãe]. The F1 contour for the moderate accent shows a much greater decline (804 to 359 Hz) than standard English (715 to 480 Hz), while that for the strong accent shows less decline (822 to 654 Hz). The F2 contour for the moderate accent shows less decline (1916-1736 Hz) than standard English (1994-1540 Hz) and the contour for the strong accent shows tha least decline (2056-1962 Hz).

[ae]. The F1 contour for standard English is almost level (620-650 Hz) while that for the moderate and strong accents shows a "hat pattern," i.e., moderate

(532-712-712-671 Hz) and strong (560-794-794-561 Hz). The F2 contour for standard English shows an incline of 500 Hz (1100-1600 Hz), while F2 for moderate and strong accents show one of 700 Hz, i.e., moderate (984-1505 Hz) and strong (1402-2148 Hz). Notice that F2 for the strong accent is also higher than that for standard English and moderate accent.

[I]. The F1 contour for standard English is almost level (400-460 Hz), while that for moderate and strong accents declines, i.e., moderate (399-179 Hz) and strong (374-187 Hz). F2 for standard English shows an incline (1600-2000 Hz). F2 for the moderate accent also shows an incline but is higher in value (2182-2524 Hz). F2 for the strong accent is lower than standard and shows a decline (1308-1028 Hz).

[ɔ]. F1 for standard English, moderate and strong accents are similar, level and 575, 579, and 607 Hz respectively. F2 is different, however. F2 for standard English shows an incline of 1000 Hz (1010 to 2010 Hz), while the moderate accent shows only a slight incline (1169-1447 Hz). F2 for the strong accent is an "inverted hat pattern," i.e., 1869-1638-1638-1869 Hz.

[ə] in [ˈkaenzəs]. F1 for standard English is lower (385 Hz) than for the moderate (579 Hz) and strong (654 Hz)

accents. F2 for standard English and the moderate accent are similar (1550 Hz and 1562 Hz respectively), while F2 for the strong accent is about 500 Hz higher (2056 Hz).

[ə] in ['baeskətbɔ l]. For standard English, F1 is 620 Hz and F2 is 1220 Hz. For the moderate accent, F1 is lower (532 Hz) and F2 is higher (1916 Hz). For the strong accent, F1 is even lower (467 Hz) and F2 even higher (2056 Hz).

[ə] in ['tIkəts]. The F1 contour for standard English shows an incline (460-575 Hz), while for the moderate and strong accents, it remains level at 359 Hz and 467 Hz respectively. The F2 contour for standard English shows a decline from 1850 to 1010 Hz. F2 for the moderate accent also shows a decline but is higher (2141-1869 Hz). F2 for the strong accent remains level at a higher value (2524 Hz).

[ɚ] in ['dalɚz]. For standard English, the F1 contour shows a decline from 470 to 270 Hz, while for the moderate and strong accents, F1 remains level at 532 Hz and 561 Hz respectively. F2 shows a slight incline for standard English (1050-1310 Hz), while F2 for the moderate and strong accents remains level at 1424 Hz and 1869 Hz, respectively.

It seems clear from these data that accented vowels are clearly distinct from standard English vowels. Next, it

will be demonstrated that accented vowels are also different from standard Spanish vowels.

The same two Spanish speakers (CG and WA) were taperecorded speaking the following Spanish sentences:

1. Matanzas está en Cuba.
2. Juego al basketból.
3. Un hombre de Costa Rica se llama un "tico".
4. Este libro vale cinco dólares.

These sentences contain vowels in similar environments, i.e.,

| ENGLISH | SPANISH |
|---|---|
| <u>Stressed Vowels</u> | |
| ['kãenzəs] | [ma'tãnsas] |
| ['baeskətbɔl] | [baskɛ̃t'bol] |
| ['tIkəts] | ['tiko] |
| [kɔst] | ['kosta] |
| <u>Unstressed Vowels</u> | |
| ['kãenzəs] | [ma'tãnsas] |
| ['baeskətbɔl] | [baskɛ̃t'bol] |
| ['dalɚz] | ['dolares] |

Table 4 shows normalized values in Hertz for F1 and F2 for subjects CG (moderate accent in English) and WA (strong accent in English):

TABLE 4. Formant Frequencies for Spanish Vowels

| | CG (Moderate) | | WA (Strong) | |
|---|---|---|---|---|
| | F1 | F2 | F1 | F2 |
| [ma'tãnsas] | 490 | 1337 | 654 | 1682 |
| [baskƐt'bol] | 535 | 1158 | 607 | 1495 |
| ['tiko] | 267 | 2315 | 280 | 2336 |
| ['kosta] | 490 | 1337 | 327 | 1308 |
| [ma'tãnsas] | 401 | 1337 | 374 | 1682 |
| [baskƐt'bol] | 356 | 1604 | 374 | 2056 |
| ['dolares] | 356 | 1248 | 374 | 1588 |

Compare these values to values for corresponding vowels in English:

F1 for [æ̃] in ['kæ̃nzəs] for both speakers is over 800 Hz, while F1 for [ã] in [ma'tãnsas] is only 490 Hz for CG and 654 Hz for WA. F2 is also lower: F2 for CG and WA when speaking English is close to 2000 Hz; when speaking Spanish, F2 is 1337 and 1682 Hz respectively.

The steady state of F1 for English [ae] in ['baeskə tbɔl] for CG is 712 Hz and for WA 794 Hz. F1 for Spanish [a] in [baskƐt'bol] is lower, 535 Hz and 607 Hz respectively. F2 for English [ae] for CG is 984-1505 Hz and for WA is 1402-2148 Hz; F2 for Spanish [a] is 1158 and 1495

Hz respectively.

F1 for English [I] in ['tIkəts] for CG is 399-179 Hz and for WA is 374-187 Hz. Spanish [i] in ['tiko] has a lower F1: 267 Hz for CG and 280 Hz for WA. F2 for Spanish [i] for CG is 2315 Hz and for WA, 2336 Hz. F2 for English [I] for CG is 2182-2524 Hz, but for WA is lower, 1308-1028 Hz.

F1 for [ɔ] in [kɔst] for CG is 579 Hz and for WA is 607 Hz. [o] in ['kosta] has a lower F1 of 490 Hz for CG and 327 Hz for WA. F2 for [ɔ] for CG is 1169-1447 Hz and for WA is 1869-1638-1869 Hz. F2 for [o] for CG is 1337 Hz and for WA is 1308 Hz.

[ə] is ['kaenzəs] has an F1 of 579 Hz for CG and 654 Hz for WA. Spanish [a] in [ma'tansas] has a lower F1 of 401 Hz for CG and 374 Hz for WA. F2 for [ə] in English has an F2 of 1562 Hz for CG and 2056 Hz for WA. [a] in Spanish has an F2 of 1337 Hz for CG and 2056 Hz for WA.

[ə] in English ['baeskətbɔl] has an F1 of 532 Hz for CG and 467 Hz for WA. Spanish [E] in [baskEt'bol] has a lower F1 of 356 Hz for CG and 374 Hz for WA. F2 for [ə] is 1916 Hz for CG and 2056 Hz for WA. F2 for [Ê] is 1604 Hz for CG and 2056 Hz for WA.

[ɚ] in English ['dalɚz] has an F1 of 532 Hz for CG and 561 Hz for WA. Spanish [a] in ['dolares] has a lower F1

of 356 Hz for CG and 374 Hz for WA. F2 for English [ɤ] for CG is 1424 Hz and 1869 Hz for WA, while F2 for Spanish [a] is 1248 Hz for CG and 1588 Hz for WA.

On the basis of this brief investigation, the claim can be made that since Spanish vowels have a lower F1 than their counterparts in English, Spanish vowels are not being substituted for English ones when the Spanish-speaker is speaking English. Rather, approximations to the corresponding English vowels are being produced.

Segment Duration

Divergences from L2 norms for stress implementation may help cue foreign accent. The ratio of stressed to unstressed vowel duration appears to be greater in English than in Spanish (see section on English versus Spanish rhythm). Hutchinson (1973) found that foreign accent judgments by native English-speaking listeners were related to the ratio of stressed to unstressed vowel duration in the English spoken by adult native speakers of Spanish. The Spanish learners who increased the duration ratio to the greatest extent were judged to pronounce English better than those who produced stressed and unstressed vowels with ratios typical of Spanish.

A study by Jonasson and McAllister (1972) suggested that the temporal specification of vowels and consonants does importantly affect acceptability judgments. The duration of consonants and vowels in a naturally produced Swedish word was manipulated by computer editing techniques. Twenty Swedish-speaking subjects made categorical judgments and acceptability judgments using a three-point rating scale. The listeners gave the highest ratings to stimuli whose temporal specifications most nearly conformed to the segment durations measured in spoken Swedish.

Eisendoorn (1983b, 1984) changed the durations of vowels produced by Dutch native speakers in English CVC words to match the average duration of vowels produced by native English speakers in the same words. However, this temporal manipulation had little apparent effect on acceptability judgments.

Many studies have demonstrated that listeners possess great sensitivity to the temporal properties of vowels in their L1 (Abramson, 1962; Jonasson and McAllister, 1972; Nooteboom,, 1973). Flege and Hillenbrand (1985a) found that L2 learners were considerably more sensitive to variations in vowel than fricative duration as perceptual cues to the contrast between English /s/ and /z/ in a "piece" to "peas" continuum in which both vowel and

fricative duration were varied in equal steps.

Hutchinson (1973) found that Spanish-speaking learners of English produced only a small duration difference between stressed and unstressed syllables in English words upon their arrival in the U.S. The difference they produced resembled in magnitude the difference observed in Spanish. However, the L2 learners produced a much greater temporal difference between stressed and unstressed English syllables after a six-month intensive English course. Hutchinson noted a high positive correlation between the degree to which stressed and unstressed syllables differed in duration and global foreign accent ratings.

Sample

In a third preliminary study, spectrograms of the sentence, "Kansas State's new basketball tickets cost ten dollars now," spoken by the native English speaker and two Spanish speakers from the previously described studies were measured. Ratios of segment duration to overall utterance duration were calculated to determine where there were significant differences:

TABLE 5. Segment Duration Ratios

| WORD: | ['kaẽnzəs] | | [stets] | [nu] | ['baeskətbɔl] | |
|---|---|---|---|---|---|---|
| STD. ENG. | .04 | .02 | .045 | .0192 | .053 | .0126 |
| MOD. ACC. | .03 | .02 | .041 | .0245 | .024 | .0086 |
| STG. ACC. | .04 | .02 | .06 | .0376 | .04 | .0154 |

| WORD: | ['tIkəts] | | [kɔst] | [tɛ̃n] | ['dalɤ^z] | [naU] |
|---|---|---|---|---|---|---|
| STD. ENG. | .019 | .0176 | .034 | .027 | .06 | .089 |
| MOD. ACC. | .014 | .0139 | .026 | .03 | .028 | .066 |
| STG. ACC. | .026 | .0215 | .059 | .016 | .020 | .084 |

As can be seen in Table 5, the duration of unstressed vowels did not differ significantly among English and accented versions.

The following medial vowels (cf. Hoequist, 1983b) were of interest: [e] in [stets], [ae] in ['baeskətbɔl], [ɔ] in [kɔst] and [a] in ['dalɤ^z]. Measurements of the vowels were made in milliseconds from the beginning to the end of periodic vibration; then measurements were standardized due to differences in duration of utterance among the three speakers. Table 6 lists the standardized durations in milliseconds of the above stressed vowels over three repetitions:

TABLE 6. Durations of Stressed Vowels

| VOWEL | STANDARD ENGLISH | MODERATE ACCENT | STRONG ACCENT |
|---|---|---|---|
| [stets] | 140 | 127 | 188 |
| ['baeskətbɔl] | 165 | 77 | 124 |
| [kɔst] | 125 | 80 | 185 |
| ['dalɚz] | 190 | 90 | 64 |

Stressed vowels were much shorter for the moderate accent than for standard English pronunciation. The fact that vowels in the strongly accented utterance were longer than both standard English and moderately accented versions in two cases may be due to a lack of fluency, i.e., words were produced almost as isolated forms. It is widely accepted that the duration of words produced in isolation is greater than for words produced in fluent speech.

## Synthetic and Natural Speech in Perception Experiments

A pioneering study by Labov (1966) showed that there exists a close correlation between social variation and style variaton within a language. This suggested to Labov that most speakers of that language think or feel that particular variants are "better," i.e., endowed with superior qualities. He argued that one cannot directly question listeners concerning their ranking of social and style variations.

Lambert et al. (1960: 44) argue that "spoken language is an identifying feature of members of a national or cultural group and any listener's attitude toward members of a particular group should generalize to the language they use." The evaluational reactions to a spoken language therefore should be similar to reactions elicited in interaction with those perceived as members of a group that uses it (Anisfeld et al., 1962: 223).

However, reactions to phonological variables are inarticulate responses below the level of conscious awareness, which occur as a part of a general reaction to many variables. There is no vocabulary of socially

meaningful terms with which informants can evaluate speech. Therefore, it is necessary to proceed by eliciting evaluative behaviors sufficiently sensitive to reflect the influence of many variables, yet subject to quantification.

Natural Speech

Investigators such as Osgood (1964), Lambert, Anisfeld and Yeni-Komshian (1965), Anisfeld and Lambert (1964), Lambert, Hodgson, Gardner and Fillenbaum (1960), and Triandis, Loh and Levin (1966) as cited in Fishman (1965) all found significant differences in respondents' attitudes toward various sociolinguistic groups based upon replies to semantic differential scales. According to Fishman (1968), when used to measure attitudinal changes in connection with speech, the semantic differential technique also indicated the absence or presence of dialect differences.

Experiments in the perception of foreign accent often employ the "matched guise technique" developed by Wallace Lambert et al. (1960) to measure differences in attitude toward French and English speaking Canadians. In the matched guise technique, listeners are asked to rate speakers for given traits (e.g., amibition, honesty, friendliness, good looks, socio-economic status, amount of

education, etc.). Unknown to the listeners, a single bilingual speaker records the material in both languages or dialects to be compared. Different ratings of a single speaker are thought to reveal underlying attitudes of the listeners about linguistically different groups.

In previous studies done primarily by psychologists, the evaluative reactions of listeners have been found to be differentially biased toward speakers of different varieties of the same lanaguage: European French and Canadian French (d'Anglejan and Tucker, 1973), standard English and Jewish-accented English (Anisfeld et al., 1962), standard English and French Canadian-accented English (Webster and Kramer, 1968), standard British English and regional accents, e.g., Scottish (Giles, 1970; 1972), standard English and Mexican-American English (Ryan and Carranza, 1975; Brennan et al., 1975; Arthur et al., 1974; Ryan, Carranza and Moffie (1977), white and black English of various social classes (Tucker and Lambert, 1969; Ryan et al., 1977; Eisenstein, 1979; Baird, 1969), native and non-native Spanish (Gynan, 1983), standard English and the English of Spanish-English bilinguals in New York City (Terry and Cooper, 1969), standard English and Puerto Rican Spanish-accented English (Fishman et al., 1969), standard English and the English of Cuban immigrants (Asher and

Garcia, 1969), standard and Italian-accented English (Oyama, 1976), and standard English and the English of Serbo-Croatians (Dimitrijević and Djordjević, 1971). In each of these experiments, listeners discriminated between two or more language types represented by the given samples of natural speech.

In the Asher and Garcia (1969) study, the subjects were Cuban immigrant students (26 boys and 45 girls aged 7 to 19 from the San Francisco Bay area), whose taped readings of four sentences was judged by a group of 19 native high school students for accent. The sentences were as follows: (1) I had two hot dogs and a glass of orange juice yesterday. (2) The girls were jealous because we had a better party. (3) Pat and Shirley are measuring the rug to see if it shrank. (4) It started to snow when we were about to leave for the mountains. The judges, all of whom learned their English in the San Francisco Bay area, were instructed that a voice would utter four sentences after which each judge would classify the pronunciation in one of the four categories A-D: A (native speaker), B (near native speaker), C (slight foreign accent), D (definite foreign accent). None of the 71 Cuban children was judged to speak like a native, although the earlier arrivals had much less accent than the later ones. Longer stays in the U.S. wre

also associated with better pronunciation.

The study by Arthur, Farrar and Bradford (1974) uses the matched guise technique for isolating language differences as a variable influencing attitudes. In Phase 1, subjects were 17 Los Angeles Mexican-Americans 15-40 years who could speak both standard English and Chicano English. Each subject read a brief paragraph, first using standard English and then using Chicano English. From the 17 pairs of readings, four pairs (2 male, 2 female speakers) were chosen for use in the experiment on the basis of "clear separation of dialect guises, and fluency and proficiency in reading in both guises."

In Phase 2, 6 male and 6 female raters were chosen randomly from a class at UCLA; all identified themselves as Anglos and native speakers of English. Before the subjects listened to each of the voice guises, they were told that these were voices of Mexican-Americans from the Los Angeles area and asked to rank each voice on a 10-point scale with polar labels: "Completely standard English" and "completely non-standard English."

In Phase 3, the subjects were 25 Anglo-American UCLA students, who listed personality traits which they regarded as important for friendship/success and ranked a list of traits in order of their importance for friendship and

success. Based on the frequency of subjects' trait labels, 14 6-point bipolar scales of labelled terminal poles were developed. A fifteenth scale (Upper Class-Working Class) was added.

In Phase 4, 48 Anglo-American subjects (24 male, 24 female) were selected randomly from two undergraduate linguistics courses at UCLA and given the following explanation: "You will hear 10 voices all reading the same passage. All are voices of Chicanos or Mexican Americans from the Los Angeles area...Please pay close attention to each of the individual speech styles on the tape..."

(1) When students were asked to rank the 8 voices in terms of their departure from standard English, the results strongly supported the assumption that Anglo-American raters could hear clear dialect differences between the standard and Chicano guises. (2) For those traits with favorable and unfavorable poles, the standard English guise was consistently rated more favorable than the Chicano English guise. This supports the hypothesis that Anglo-American college students, operating within a university environment, judge speakers of a Chicano or Mexican-American dialect of English more negatively than speakers of a more standard variety of English. (3) Dialect differences were one of the voice characteristics that raters consistently attended to.

Note that the raters also displayed clear and different attitudes towards the four different individuals speaking regardless of the dialect they were speaking.

Ryan and Carranza (1975) explored reactions toward accent with taped readings of standard passages. They found that Mexican American speakers of strongly accented English were not viewed as favorably as speakers of standard English by Anglo, Black and Mexican American high school females.

The purpose of a study by Brennan, Ryan and Dawson (1975) was to determine whether two psychophysical scaling methods, magnitude estimation (ME) and sensory-modality matching (SMM), could provide reliable and practical measures of subjective accentedness as judged by naive listeners.

Subjects were 72 Notre Dame students enrolled in introductory psychology classes and received credit for participation in the experiment. All students were judged by the experimenter to be speakers of standard English. The readings of 8 speakers with a "wide range of accentedness" were selected as stimuli from taped readings of an English passage by 35 male Spanish-English bilingual college students. Each speech sample was known to have specific features of Spanish phonological interference (Ohweiler, 1972, unpub. ms.).

Readers were taped in 9 randomized orders in a soundproof room. The test passage read: "Scientific technology will no doubt play an undisputed role in these realizations--the eradication of war, poverty, pestilence and other vital social problems. The time to commence this massive undertaking is now; let virtue be our guide and faith our instrument."

Each experimental session had 3 parts: (1) Stimulus familiarization: The subject was instructed to listen carefully to 8 readers he would later be asked to rate for accentedness, and the tape was played. (2) Magnitude Estimation (ME) task: The test series was played for a second time and the subject was instructed to assign to the first speaker any number that seemed appropriate for the amount of accentedness in his speech sample. For successive speakers, the subject was to give numbers proportional to his impression of accentedness relative to the first speaker. (3) In the Sensory Modality Matching (SMM) task, the subject listened to 8 speakers reading the test passages but in a different order from that in the ME task. Each of 72 subjects received a randomly assigned set of two readings. Written directions instructed the subject to squeeze a Lafayette hand dynamometer (used to measure magnitude of hand grip in kilograms) with a force matching

the accentedness of each speaker. For the first speaker, the subject was to make whatever match seemed appropriate for him; then for successive speakers, he was to squeeze the dynamometer proportional to his impression of accentedness relative to the first speaker.

In Experiment 1, all subjects were given the ME task first, followed by the SMM task; while in Experiment 3, all subjects performed the SMM task, then the ME task. Sessions were counterbalanced for order in Experiment 2 so that half of the subjects performed the ME task before SMM, while the other half were required to do the SMM task before ME.

The results showed that nonlinguistically trained listeners give reliable judgments of the accentedness of speech samples, and that they are in agreement as to what constituted various levels of accentedness. Magnitude estimation and sensory modality matching correlate highly with the occurrence of accented pronunciations produced by the speakers, as perceived and tallied by the judges (EBR and graduate student). Results suggest that ratings of degree of accentedness will change consistently with the amount of actual change in accentedness. Informal questioning of subjects at the end of the experiment revealed the inability of untrained listeners to articulate directly which pronunciation features were important in

their judgments.

Galvan et al (1977) attempted to identify specific speech characteristics in the English of a group of bilinguals which might account for their relative positions on a scale of least to more accented. They found that phonetic differences alone do not account for the reliable scaling of accent (unlike Brennan et al.).

Ryan, Carranza and Moffie (1977) chose as subjects 103 students from Notre Dame, over 90% of whom were male, who received credit in an introductory psychology course for their participation. Eight taped readings "with a wide range of accentedness" from Brennan et al. (1975) plus two low-accent readings were put into 6 randomized orders; in each of 3 testing sessions 2 randomized orders were used.

The subjects were asked to rate each taped speaker on the basis of his voice cues alone. The test booklet consisted of (1) a section on which to rate first impressions of the speaker on two attributes (likelihood of being a friend; eventual occupation); (2) a section on which to rate the speech of speakers on three attributes (accented/unaccentedness, pleasant/unpleasant, and fluent/non-fluent). Each judgment was placed along a 7-point rating scale with left-right positions of the most favorable ends of the scale counterbalanced. Responses were

scored with integers from 1-7 (with a value of 1 indicating the most negative judgment and 7 indicating the most positive).

The results showed that the college students made rather fine discriminations among varying degrees of accentedness in rating a speaker's personal attributes and speech. The high correlation between accentedness ratings and each of the other ratings indicate that small increments in accentedness are associated with gradually less favorable ratings of status, solidarity and speech characteristics. This study thus provides additional support for the Ryan and Carranza (1975) proposition that Spanish accent features in spoken English are negatively stereotyped, and indicates that the greater the prominence of these features, the stronger the stereotyping.

The strong agreement between the group ratings of accentedness in this study and scale values based on more tedious scaling methods can be taken as support for future use of the more convenient rating scale. Group administration (impossible in sensory modality matching) results in substantial savings in the time and effort involved in data collection.

## Problems with Listener Evaluation of Natural Speech

Arthur et al. (1974: 262) found that "in addition to the shift in attitudes associated with dialect change, the raters displayed clear and different attitudes towards the four different individuals speaking in this study regardless of the dialect they were speaking. . .The consistently significant differences between mean scores. . .suggest that the raters were responding to individual voice characteristics in addition to dialect differences."

Giles (1970) studied adolescents' reactions to the matched-guises of one male speaker on a variety of regional accents. He reflects: "The speaker attempted to control for accent broadness throughout the thirteen different guises by *subjectively* (emphasis mine--DMR) producing intermediary broad guises" (Giles, 1972). Giles (1972: 268) employed the same speaker in order to make comparisons with the earlier study. He states, "It can be seen. . .that a speaker, *if proficient at the skill*, (emphasis mine--DMR) may be able to encode at least three distinct levels of accent broadness, the form of which may be reflected in judges' evaluations." Therefore, it seems that only subjective control of natural speech parameters is possible. In addition, listeners react to parameters other than those under investigation which are

present in the natural speech signal.

## Synthetic Speech

The use of speech synthesis systems for stimulus generation in speech perception research has become an established technique over the past two decades. The degree to which the intelligibility of such speech synthesis systems matches that of natural speech is therefore a question of some interest both in establishing their validity as research tools and in assessing their effectiveness as communications devices (Clark, 1983: 37).

Results of intelligibility comparisons between synthetic and natural speech samples by Clark (1983: 37) "indicate that the intelligibility of synthesized speech sources is equal or superior to natural speech for vocalic segments, but that synthetic stop and fricative class consonants have lower intelligibility than their natural counterparts." However, "despite the poorer performance of the synthesized sources i.: consonants compared to natural speech, the general pattern of ranking of the consonant class intelligibility scores is the same as for the natural speech" (ibid.: 48) (i.e., "The order from most to least intelligible is: liquids, nasals, voiced stops, voiceless

stops, voiced fricatives and voiceless fricatives" (ibid.: 46).

Several recent studies using rule-generated synthetic speech have shown lower performance levels for perception of synthetic speech relative to natural speech. For example, Pisoni and Hunnicutt (1980) performed experiments on the intelligibility of speech generated by the MITalk unrestricted text-to-speech system. In their first experiment, Pisoni and Hunnicutt asked subjects to identify a single target word from a set of six phonemically confusable alternatives using the Modified Rhyme Test (House, Williams, Hecker and Kryter, 1965). Phoneme recognition for the synthetic speech was 93.1% compared to 99.4% for natural speech. In their second experiment, Pisoni and Hunnicutt presented listeners with either (1) meaningful or (2) syntactically correct but anomalous sentences. Recall for the synthetic meaningful sentences was about 6% lower than for the natural speech. However, for the anomalous sentences, recall performance for the synthetic speech was about 19% lower than for natural speech.

Jenkins and Franklin (1981), using the VOTRAX and FOVE synthesizers, reported that when subjects were asked to recall the gist of simple stories, recall for the synthetic

stories was not demonstrably poorer than recall for the natural stories. This result is consistent with Pisoni and Hunnicutt's finding that the identification of meaningful sentences was not as severely impaired as the identification of syntactically correct but anomalous sentences.

Pisoni (1981) found that when isolated synthetic and natural words were presented in a lexical decision task, response times were, on the average, 140 ms slower than response times for natural words and nonwords. This study shows that for isolated words, significant decrements in performance can be shown for synthetic speech relative to natural speech.

Pisoni and Koen (1982) have found differences in intelligibility of natural and synthetic words presented in noise at several signal-to-noise ratios. Intelligibility of synthetic words was affected by noise more than the same naturally produced words. The effects of noise produce a greater decrement on recognition of the synthetic items probably because they contain fewer redundant acoustic cues to support recognition of the phonetic structure.

The above findings appear to indicate that the processes used to perceive and understand synthetic speech are heavily dependent on the contextual environment in which the synthetic speech is presented (Lindblom, 1982). When

meaningful sentences or simple passages are used, intelligibility and comprehension of synthetic speech do not seem very different from that of natural speech, but this is not the case when isolated words or meaningless sentences are presented (Luce et al., 1981: 6).

Luce et al. (1981: 5) state: "For a naive listener, synthetic speech often seems difficult to understand. Problems may arise in the recognition of words and the interpretation of the meaning of sentences because of the distracting mechanical quality of the speech signal."

The authors add that "several investigators (Allen, 1976; Nickerson, 1977) have suggested that prosodic differences between synthetic and natural speech constitute the major difficulty in the comprehension of synthetic speech, particularly fluent synthetic speech" (ibid.).

One explanation advanced by Luce et al. (ibid.) for the difficulties observed in perception of synthetic speech may be found at the relatively early stages of perceptual analysis and encoding at which words are recognized from their phonetic representation of the message (Pisoni, 1981). Synthetic speech is often generated by rules that manipulate only a limited number of the potential cues to the phonological representation of the message. Thus, perception of synthetic speech may be adversely affected by

only a partial specification of the acoustic cues necessary for a natural perception of phonetic segments (Luce et al., 1981: 5).

Capacity Demands in Short Term Memory for Synthetic and Natural Speech

According to Luce et al. (1981: 5) difficulties in the perception and comprehension of synthetic speech may arise from a general constraint on the processing of information in short term memory. In particular, synthetic speech may require more processing capacity than natural speech for maintainance of information in short term memory and subsequent transfer of information to long term memory. Thus, perception of synthetic speech may be analogous to the perception of natural speech presented in high levels of noise. Earlier research has demonstrated that difficulties in encoding of speech perceived in noise produce subsequent difficulties in rehearsal processes in short term memory and therefore recall of information from long term memory (Dallett, 1964; Rabbitt, 1968) (Luce et al., 1981: 5).

Luce et al. (1981: 22) report that their results "demonstrate that synthetic speech is difficult to perceive and understand, relative to natural speech, in part because

it affects the allocation of limited processing resources in short term working memory."

Luce (1982) found that "subjects perform more poorly for synthetic passages on comprehension questions designed to probe the content of a given passage. However, the subjects hearing synthetic passages perform better than those hearing natural passages on questions that probe retention of the surface structure of the passages. These comprehension results suggest that the subjects' attention is somehow directed toward the superficial (surface) properties of the actual speech signal in the synthetic speech condition than to the properties of the message in the natural speech condition (see Aaronson, 1976)" (Luce, 1981: 23).

Luce (1981: 240) points out the following problem with synthetic speech revealed in his study of comprehension of fluent synthetic speech produced by rule. "Under conditions of increased demand on the cognitive mechanisms involved in speech processing, synthetic speech may detract from other tasks the listener is engaged in as well as increase the probability of errors in perception and comprehension."

Synthesizing Speech Sounds

Klatt (1980) provides guidelines for synthesizing speech on a parallel formant synthesizer, the type used to generate the synthetic speech stimuli for this experiment, which allows for complete control of parameters (e.g., formant amplitudes) so that they can be varied independently. Following is a partial list of parameters used and their function.

The amplitude of each resonator in the parallel branch is controlled in dB by two voicing sources and a noise source, which are controlled by amplitude parameters AV, AVS, AH and AF. AV (amplitude of normal voicing) controls the "normal" voicing source (in dB). AVS (amplitude of sinusoidal voicing) controls a "smoothed" voicing source (e.g., found in prevoicing or closure of stop consonants). AF (amplitude of frication noise) controls frication sent through parallel resonators A2 to A6 (i.e., amplitude controls). AH refers to amplitude of aspiration noise, which is given in dB.

FO (fundamental frequency) consists of an impulse train corresponding to normal voicing and is generated when FO > 0. Setting either FO or AV to zero turns off the voicing source.

*Synthesis of Vowels.* The parameters that are usually varied to generate an isolated vowel are the amplitude of voicing (AV), the fundamental frequency of vocal fold vibrations (F0), the lowest three formant frequencies (F1, F2, F3) and bandwidths (B1, B2, B3). Formant and bandwidth values are given in Table 7 (Klatt, 1980). The amplitude of the voicing source (AV) is set to about 60 dB for a stressed vowel and falls gradually by a few dB near the end of the syllable.

*Synthesis of Consonants.* The *approximants* [w,y,r,l] are similar to vowels and require the same set of control parameters to be varied in order to differentiate among them. Formant values given in Table 8 (Klatt, 1980) for postvocalic [r,l] depend somewhat on the following vowel. The source amplitude (AV) for a prevocalic approximants should be about 10 dB less than in the vowel.

The *fricatives* include both voiceless fricatives (AF=60, AV=0, AVS=0) and voiced fricatives (AF=50, AV=47, AVS=47). Formants to be excited by the frication noise source are determined by the amplitude controls A2, A3, A4, A5, A6, and AB. (AB is used to synthesize stop bursts and fricatives where the source of turbulence is located above the glottis.)

For velars before nonfront vowels, A2 is set to

Table 7 Parameter values for the synthesis of selected vowels. If two values are given, the vowel is diphthongized or has a schwa-like offglide in the speech of the author. The amplitude of voicing, AV, and fundamental frequency, F0, must also be given contours appropriate for an isolated vowel.

| Vowel | F1 | F2 | F3 | B1 | B2 | B3 |
|---|---|---|---|---|---|---|
| [iʸ] | 310 | 2020 | 2960 | 45 | 200 | 400 |
| | 290 | 2070 | 2960 | 60 | 200 | 400 |
| [ɪᵊ] | 400 | 1800 | 2570 | 50 | 100 | 140 |
| | 470 | 1600 | 2600 | 50 | 100 | 140 |
| [eʸ] | 480 | 1720 | 2520 | 70 | 100 | 200 |
| | 330 | 2020 | 2600 | 55 | 100 | 200 |
| [ɛᵊ] | 530 | 1680 | 2500 | 60 | 90 | 200 |
| | 620 | 1530 | 2530 | 60 | 90 | 200 |
| [æᵊ] | 620 | 1660 | 2430 | 70 | 150 | 320 |
| | 650 | 1490 | 2470 | 70 | 100 | 320 |
| [ɑ] | 700 | 1220 | 2600 | 130 | 70 | 160 |
| [ɔᵊ] | 600 | 990 | 2570 | 90 | 100 | 80 |
| | 630 | 1040 | 2600 | 90 | 100 | 80 |
| [ʌ] | 620 | 1220 | 2550 | 80 | 50 | 140 |
| [oʷ] | 540 | 1100 | 2300 | 80 | 70 | 70 |
| | 450 | 900 | 2300 | 80 | 70 | 70 |
| [ʊᵊ] | 450 | 1100 | 2350 | 80 | 100 | 80 |
| | 500 | 1180 | 2390 | 80 | 100 | 80 |
| [uʷ] | 350 | 1250 | 2200 | 65 | 110 | 140 |
| | 320 | 900 | 2200 | 65 | 110 | 110 |
| [ɚ] | 470 | 1270 | 1540 | 100 | 60 | 110 |
| | 420 | 1310 | 1540 | 100 | 60 | 110 |
| [aʸ] | 660 | 1200 | 2550 | 100 | 70 | 200 |
| | 400 | 1880 | 2500 | 70 | 100 | 200 |
| [aʷ] | 640 | 1230 | 2550 | 80 | 70 | 140 |
| | 420 | 940 | 2350 | 80 | 70 | 80 |
| [oʸ] | 550 | 960 | 2400 | 80 | 50 | 130 |
| | 360 | 1820 | 2450 | 60 | 50 | 160 |

(from Klatt, 1980)

Table 8 Parameter values for the synthesis of selected components of English consonants before front vowels (see text for source amplitude values).

| Sonor | $F1$ | $F2$ | $F3$ | $B1$ | $B2$ | $B3$ | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| [w] | 290 | 610 | 2150 | 50 | 80 | 60 | | | | | | |
| [y] | 260 | 2070 | 3020 | 40 | 250 | 500 | | | | | | |
| [r] | 310 | 1060 | 1380 | 70 | 100 | 120 | | | | | | |
| [l] | 310 | 1050 | 2980 | 50 | 100 | 290 | | | | | | |
| **Fric.** | $F1$ | $F2$ | $F3$ | $B1$ | $B2$ | $B3$ | $A2$ | $A3$ | $A4$ | $A5$ | $A6$ | AB |
| [f] | 340 | 1100 | 2080 | 200 | 120 | 150 | 0 | 0 | 0 | 0 | 0 | 57 |
| [v] | 220 | 1100 | 2080 | 60 | 90 | 120 | 0 | 0 | 0 | 0 | 0 | 57 |
| [θ] | 320 | 1290 | 2540 | 200 | 90 | 200 | 0 | 0 | 0 | | 28 | 48 |
| [ð] | 270 | 1290 | 2540 | 60 | 80 | 170 | 0 | 0 | 0 | 0 | 28 | 48 |
| [s] | 320 | 1390 | 2530 | 200 | 80 | 200 | 0 | 0 | 0 | 0 | 52 | 0 |
| [z] | 240 | 1390 | 2530 | 70 | 60 | 180 | 0 | 0 | 0 | 0 | 52 | 0 |
| [š] | 300 | 1840 | 2750 | 200 | 100 | 300 | 0 | 57 | 48 | 48 | 46 | 0 |
| **Affricate** | | | | | | | | | | | | |
| [č] | 350 | 1800 | 2820 | 200 | 90 | 300 | 0 | 44 | 60 | 53 | 53 | 0 |
| [ǰ] | 260 | 1800 | 2820 | 60 | 80 | 270 | 0 | 44 | 60 | 53 | 53 | 0 |
| **Plosive** | | | | | | | | | | | | |
| [p] | 400 | 1100 | 2150 | 300 | 150 | 220 | 0 | 0 | 0 | 0 | 0 | 63 |
| [b] | 200 | 1100 | 2150 | 60 | 110 | 130 | 0 | 0 | 0 | 0 | 0 | 63 |
| [t] | 400 | 1600 | 2600 | 300 | 120 | 250 | 0 | 30 | 45 | 57 | 63 | 0 |
| [d] | 200 | 1600 | 2600 | 60 | 100 | 170 | 0 | 47 | 60 | 62 | 60 | 0 |
| [k] | 300 | 1990 | 2850 | 250 | 160 | 330 | 0 | 53 | 43 | 45 | 45 | 0 |
| [g] | 200 | 1990 | 2850 | 60 | 150 | 280 | 0 | 53 | 43 | 45 | 45 | 0 |

| Nasal | FNP | FNZ | $F1$ | $F2$ | $F3$ | $B1$ | $B2$ | $B3$ |
|---|---|---|---|---|---|---|---|---|
| [m] | 270 | 450 | 480 | 1270 | 2130 | 40 | 200 | 200 |
| [n] | 270 | 450 | 480 | 1340 | 2170 | 40 | 300 | 300 |

(from Klatt, 1980)

about 60 dB. The values given for F2 and F3 can also serve as "loci" for the characterization of the C-V formant transitions before front vowels.

The plosive parameters in Table 8 refer to the brief burst of frication noise generated at plosive release. Formant frequency values serve as loci for predicting formant positions at voicing onset. In addition to differences in source amplitudes, voiced and voiceless consonants differ in that F1 is higher and B1 is larger when the glottis is open.

The parameters used to generate a nasal murmur include the nasal pole and zero frequencies (FNP and FNZ). A nasalized vowel is generated by increasing F1 by about 100 Hz, and by setting FNZ to be the average of this new F1 value and 270 Hz (the frequency of the fixed nasal pole).

Using a Klatt-based synthesizer, a natural sounding rendition of the English sentence, "Kansas State's new basketball tickets cost ten dollars now," was produced.

## Conclusion

The Oden and Massaro (1978) fuzzy logical model explains the superior performance of experienced listeners in perception of synthetic speech. Experienced listeners

apparently have developed more elaborate category prototypes against which to judge their stimuli. It also explains why in experiments using natural speech (Asher and Garcia, 1969; Scovel, 1981), inexperienced listeners had a tendency to identify sentences produced by native speakers as having been produced by a non-native speaker. That is, the model explains why they are inclined to label speech as "accented" or "distorted" when asked to scrutinize it in an unaccustomed way (Flege, 1984b).

A review of the literature and pilot data yield the following conclusions: (1) the use of synthetic speech is required for control of parameters, (2) high quality synthetic speech is essential, (3) a training session is needed to give listeners experience with synthetic speech and the range of accentedness, (4) context of the stimulus is important for intelligibility and must be long enough to examine prosodic effects.

There are many cues conveying foreign accent. Although primitive speech synthesizers had a limited range of distinctive and redundant features to be manipulated in a crude way, today's high quality synthesizer builds in a fuller range of features, especially coarticulatory and prosodic, in such a way that foreign accent can be manipulated much more precisely than was previously

possible.

The current study will attempt to evaluate the single and combined effects of segmental and prosodic cues upon the perception of Spanish-accented English. The experiment is designed to evaluate cue strength of each of a list of five variables and to assess the additivity or non-additivity of their effect when concatenated.

It is hypothesized that (1) prosodic variables take precedence over segmental variables in the perception of foreign (Spanish) accent. In particular, changes in F0 and/or formant frequency associated with prosodic variables will prove to be more potent than segmental variables. (2) Degree of accentedness conveyed by individual cues will be nonlinearly related to the composite effect of the cues.

# CHAPTER II

# METHODOLOGY

## Pilot Study

### Stimulus Preparation

Previous studies of foreign accent reported in the literature are based on natural speech stimuli which cannot be precisely controlled by the researcher. This fourth brief preliminary study uses synthetic speech in order to maintain precise control of the accent-bearing cues under investigation.

Subjects. The adult male speaker of an inland dialect of standard Latin American Spanish (Cali, Colombia) from the previous preliminary studies, who was judged to have a moderate Spanish accent in English (Oral Interview Evaluation of English as a Second Language), served as one subject. The second subject was the adult male speaker of standard American English also from the previous studies. Neither of the subjects reported any speech or hearing

disorders.

*Speech Sample*. Under controlled laboratory conditions, each subject read aloud a sequence of three sentences three times:

1. They watched the ship named *Titanic* as she slipped below the waves.
2. Kansas State's new basketball tickets cost ten dollars now.
3. We relaxed as we passed through the tiny Mexican shops.

The sentences were approximately 4 sec. in length, and were constructed to exhibit 4 occurrences of the cues to be manipulated (i.e., VOT of syllable-initial voiceless stops; duration of stressed vowels; formant frequency of stressed vowels; formant frequency and duration of reduced vowels). The subjects were asked to imitate the rate at which the sentences were read on a taperecording that was played for them in order to control their relative reading rate. The speech sample was recorded on a Sony WDM-6 cassette recorder with ECM-50 dual microphones.

The three repetitions of the second sentence, "Kansas State's new basketball tickets cost ten dollars

now," spoken by each speaker were analyzed with broad, narrow and cross-sectional spectrograms made on a Kay Elemetrics digital Sona-Graph model 7800. Measurements were made of accent-bearing cues under consideration: (1) voice onset time of syllable-initial voiceless stops (i.e., [k] in "Kansas," [t] in "tickets," [k] in "cost," [t] in "ten"); (2) duration of stressed vowels (i.e., [ae] in "Kansas," [u] in "new," [ɔ] in "cost," [E] in "ten"); (3) F1, F2, and F3 of stressed vowels [I, ae, ɔ ] (i.e., [I] in "tickets," [ae] in "Kansas" and "basketball," [ɔ] in "cost"; and (4) F1, F2, and F3 plus duration of reduced vowels. Durations and formant frequencies were averaged from the 3 repetitions and normalized for use as guidelines in synthesizing the stimuli.

Synthetic Stimuli. All synthetic stimulus sentences were prepared at Haskins Laboratories using the Haskins software serial synthesizer (SYN), developed by Mattingly, which allowed an update of parameter values every 10 ms. SYN computed the PCM waveform for an utterance from synthesizer parameter values, which could be played through the normal PCM system. A synthetic version of the native English production of "Kansas State's new basketball tickets cost ten dollars now," was prepared using values from

spectrographic measurements of the English speaker's production. Fifteen "accented" versions were created by substituting values from spectrographic measurements of the Spanish speaker's production of the sentence for the cues to be manipulated. (Note that duration values had to be rounded off to the nearest 10 ms.) In the first 4 versions, single cues were manipulated, in the next 6, 2 cues were combined, in the next 4, 3 cues were combined, and in the last version, all 4 cues were combined. So, for example, in version 1, VOT of syllable-initial voiceless stops received Spanish values; in version 2, duration of stressed vowels received Spanish values and so forth; in version 5, both VOT and stressed vowel duration were given Spanish values, etc. Quality control was achieved through spectrographic comparison of live voice versus synthetic versions of the sentence.

The 15 accented stimuli were randomized (using a program called RANDOM for the VAX computer) into 2 sequences for a total of 30 stimulus sentences. Each was paired on magnetic audio tape with the unaccented synthetic sentence, following it by 1 sec., with 6 sec. between stimulus pairs. A brief training session consisting of 5 stimulus pairs (unaccented sentence followed by an accented sentence), representing the range of accentedness, preceded the stimuli

to be rated.

Stimulus tape generation was accomplished by outputting randomizations of the synthetic tokens through 12-bit D/A conversion and audio processing equipment associated with a VAX computer at Haskins Laboratories. Digitized rate was 10k samples/sec, and recordings were made on an Otari audio tape recorder (1/2 track, 7-1/2 ips). These recordings on reel-to-reel tape were dubbed onto cassette tape by connecting a Sony WDM-6 cassette recorder to an Ampex ATR800 reel-to-reel recorder at Louisiana State University.

Perception Experiment

*Judges*. Serving as judges were 10 randomly chosen listeners from the communication disorders program at Louisiana State University. All were native speakers of American English with normal speech and hearing and no discernible biases regarding Hispanics as indicated by their responses on a written questionnaire administered before the testing session. (See Appendix A.)

*Procedure*. Experimental sessions were held in a quiet environment. At the beginning of each experimental

session, the judges were given copies of written instructions to follow along as they were read aloud by the investigator and questions were answered. They then heard through Sony MDR-8 headphones at a comfortable listening level the 5 pairs of stimuli intended to familiarize them with the range of accentedness and quality of the synthetic speech. The judges then rated the 30 stimulus pairs, each pair consisting of the unaccented followed by an accented sentence. They indicated how different they thought the accented version of each pair was from the unaccented version on a 10-point scale.

Analysis of Judges' Ratings

(Figure 4) Mean rating values of overall cue combinations for all 1-, 2-, 3- and 4-cue sentences were as follows: 3.125, 4.008, 5.537, and 5.95 respectively. The effect of the number of cues present upon the rated accentedness was assessed with a one-way ANOVA with repeated measures: The significant F score (19.794, df 3, 1.606) revealed that number of cues was indeed the cause of significant differences. A post-hoc Scheffe analysis of differences between simple means for 1-, 2-, 3-, and 4-cue sentences revealed that contrasts of 1 and 3, 1 and 4, and 2

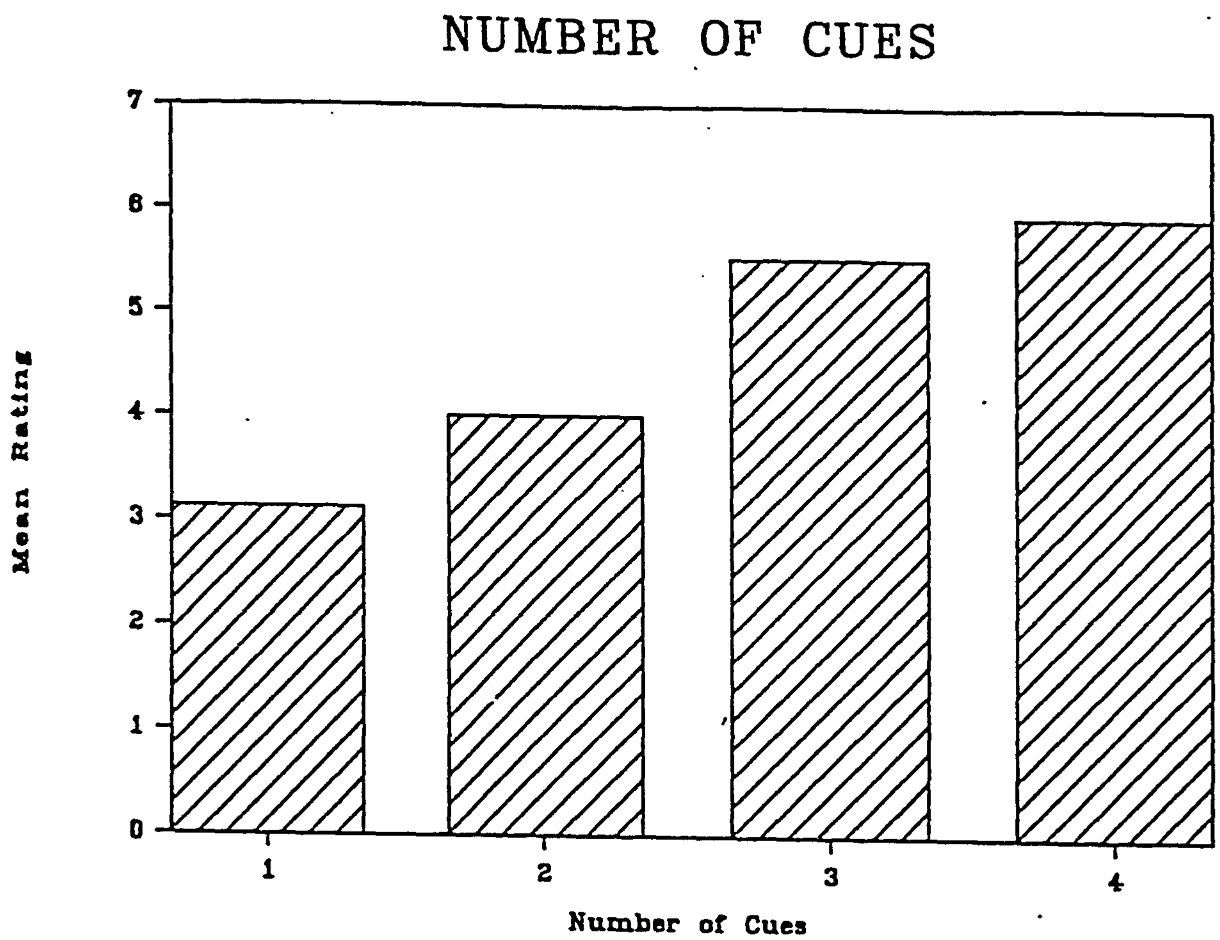


Figure 4

and 4 cues were all significant differences; that is, all cue increments of 2 additional cues yielded significantly different rating.

(Figure 5) Amongst single cues, only formant-frequency perturbations in stressed vowels (marked by *'s in Figure 5) proved to be significantly more potent than all other 1-cue items (Scheffe test). Amongst all comparisons of 2-cue combinations, only those involving formant-frequency perturbations of stressed vowels yielded significant Scheffe values; that is, vowel quality in combination with any other single cue, as a pair, was significantly more accented than any other 2-cue pairing. Similarly for 3-cue combinations, only those triple cue sets with formant-frequency perturbations yielded significant differences in rated accentedness when compared with cue triplets not containing formant-frequency perturbations.

## Conclusion

In conclusion, these results support the claim that judgments of foreign accent (at least for Spanish-accented English) depend to a large extent on vowel quality and are very little affected by other elements of the speech signal (cf. Barry, 1974).

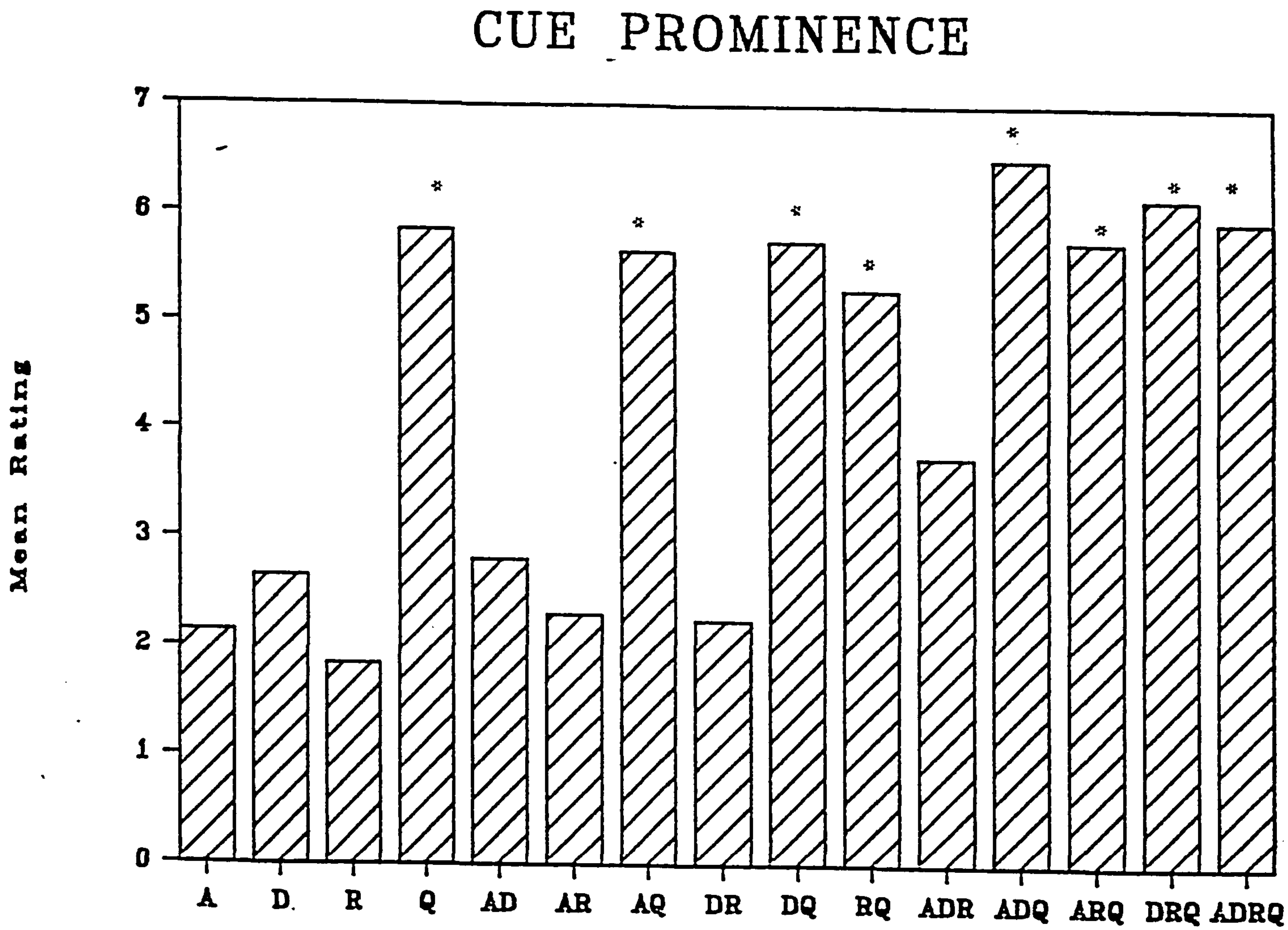
CUE PROMINENCE
Mean Rating
0
1
2
3
4
5
6
7
A
D.
R
Q
AD
AR
AQ
DR
DQ
RQ
ADR
ADQ
ARQ
DRQ
ADRQ

Figure 5

## Methods of Procedure

### Stimulus Preparation

Subjects. Subjects consisted of the same three speakers from previously described studies. Two adult male speakers of the same inland dialect of standard Latin American Spanish (Cali, Colombia) served as subjects. This particular dialect of Spanish was chosen because of its lack of distinguishing dialect markers (e.g., Mexican Spanish intonation, Castilian Spanish [θ], Argentinian Spanish [ž], Caribbean Spanish aspiration/omission of /s/, etc.) One was judged to have a "moderate" foreign accent in English on the Oral Interview Evaluation of English as a Second Language (Louisiana State University), while the other speaker was judged to have a "strong" accent. A third subject was a male adult speaker of standard American English with no observable dialect markers. None of the subjects reported any speech or hearing disorders.

Speech Sample. Each subject read aloud the following sequence of three sentences three times:

1. They watched the ship named Titanic as she slipped below the waves.
2. Kansas State's new basketball tickets cost ten dollars now.
3. We relaxed as we passed through the tiny Mexican shops.

Sentences were approximately 4 sec. in length and were constructed to exhibit the 5 cues to be manipulated (i.e., F0 across the utterance; VOT of voiceless stops in syllable-initial position; F1, F2, F3 for stressed vowels [ae,I,ɔ]; F1, F2, F3 for reduced vowels; duration of medial stressed vowels [e,ae,ɔ,a]). The subjects were asked to imitate the rate at which the sentences were read on a taperecording that was played for them prior to recording in order to control their relative reading rate. The speech sample was recorded on a Sony WDM-6 cassette recorder with ECM-50 dual microphones.

The three repetitions of the second sentence spoken by each speaker were analyzed with broad, narrow and cross-sectional spectrograms on a Kay Elemetrics Digital Sona-Graph Model 7800. In addition, a computer-implemented pitch extraction routine (Gold and Rabiner, 1969) was run on each repetition. The criterion measures extracted for use

as synthesis control parameters were reevaluated for inclusion as cues for Spanish accent in English. Cue (1) VOT of syllable-initial voiceless stops (i.e., [k] in "Kansas," [t] in "tickets," [k] in "cost," [t] in "ten"), an often-cited phonetic difference between the two languages, was included to determine its relative strength as an accent-bearing cue. Cue (2), duration of utterance-medial stressed vowels (i.e., [e] in "State's," [ae] in "basketball," [ɔ] in "cost," [a] in "dollars"), was included because vowel duration has been cited as a prosodic cue signalling perception of stress and rhythm. Cue (3), F1, F2 and F3 values of unstressed vowels (i.e., [ə] in "Kansas," "basketball," and "tickets," and [ɚ] in "dollars"), was included since schwa occurring in English has been cited as a major factor in the perception of stress-timing as opposed to syllable-timing. Cue (4), F1, F2, and F3 values for the phonetic realization of stressed vowels [I,ae,ɔ] (i.e., [I] in "tickets," [ae] in "Kansas" and "basketball," [ɔ] in "cost,"), has been included because vowel quality differences have been often cited in descriptions of Spanish-accented English. Recall that vowel quality was found to be the major accent-bearing cue in the pilot study above. Cue (5), F0 contour across the entire sentence was also manipulated (see Chapter I, "English Intonation:

Sample" for specific values) because changes in pitch have been cited as the major prosodic cue in signalling perception of stress.

*Synthetic Stimuli.* All synthetic stimulus sentences were prepared at The Kresge Hearing Research Laboratory of the South in New Orleans, LA, using the Klatt-based synthesizer amended by Creighton Miller with a variable time-frame (i.e., 5 ms to 1 ms), which allowed more frequent updating of the speech signal than the original Klatt synthesizer with a 10 ms time-frame, resulting in more natural-sounding synthetic speech.

A synthetic version of the unaccented (native) English sentence, "Kansas State's new basketball tickets cost ten dollars now," was prepared using the time/frequency normalized values from the English speaker's production. Thirty "strongly accented" versions and 30 "moderately accented" versions were created by substituting normalized values from the Spanish speakers determined to have a strong and moderate accents in English respectively for the cues to be manipulated singly and in combination:

STIMULI MODIFICATIONS

S1 None (Standard English pronunciation)

| | |
|---|---|
| S2M | FO (fundamental frequency)--moderate |
| S2S | FO--strong |
| S3M | VOT (voice onset time)--moderate |
| S3S | VOT--strong |
| S4M | VQS (vowel quality, stressed vowel)--moderate |
| S4S | VQS--strong |
| S5M | VQU (vowel quality, unstressed vowel)--moderate |
| S5S | VQU--strong |
| S6M | DUR (vowel duration)--moderate |
| S6S | DUR--strong |
| S7M | FO, VOT--moderate |
| S7S | FO, VOT--strong |
| S8M | FO, VQS--moderate |
| S8S | FO, VQS--strong |
| S9M | FO, VQU--moderate |
| S9S | FO, VQU--strong |
| S10M | FO, DUR--moderate |
| S10S | FO, DUR--strong |
| S11M | VOT, VQS--moderate |
| S11S | VOT, VQS--strong |
| S12M | VOT, VQU--moderate |
| S12S | VOT, VQU--strong |
| S13M | VOT, DUR--moderate |
| S13S | VOT, DUR--strong |

| | |
|---|---|
| S14M | VQS, VQU--moderate |
| S14S | VQS, VQU--strong |
| S15M | VQS, DUR--moderate |
| S15S | VQS, DUR--strong |
| S16M | VQU, DUR--moderate |
| S16S | VQU, DUR--strong |
| S17M | FO, VOT, VQS--moderate |
| S17S | FO, VOT, VQS--strong |
| S18M | FO, VOT, VQU--moderate |
| S18S | FO, VOT, VQU--strong |
| S19M | FO, VOT, DUR--moderate |
| S19S | FO, VOT, DUR--strong |
| S20M | FO, VQS, VQU--moderate |
| S20S | FO, VQS, VQU--strong |
| S21M | FO, VQS, DUR--moderate |
| S21S | FO, VQS, DUR--strong |
| S22M | VOT, VQS, VQU--moderate |
| S22S | VOT, VQS, VQU--strong |
| S23M | VOT, VQS, DUR--moderate |
| S23S | VOT, VQS, DUR--strong |
| S24M | VOT, VQU, DUR--moderate |
| S24S | VOT, VQU, DUR--strong |
| S25M | VQS, VQU, DUR--moderate |
| S25S | VQS, VQU, DUR--strong |

| | |
|---|---|
| S26M | FO, VQU, DUR--moderate |
| S26S | FO, VQU, DUR--strong |
| S27M | FO, VOT, VQS, VQU--moderate |
| S27S | FO, VOT, VQS, VQU--strong |
| S28M | FO, VOT, VQS, DUR--moderate |
| S28S | FO, VOT, VQS, DUR--strong |
| S29M | FO, VQS, VQU, DUR--moderate |
| S29S | FO, VQS, VQU, DUR--strong |
| S30M | VOT, VQS, VQU, DUR--moderate |
| S30S | VOT, VQS, VQU, DUR--strong |
| S31M | FO, VOT, VQS, VQU, DUR--moderate |
| S31S | FO, VOT, VQS, VQU, DUR--strong |

Quality control was determined by spectrographic comparison with naturally spoken repetitions of the sentence. In a future experiment, individual words will be spliced out of context both in the natural and synthetic versions of the unaccented sentence using a waveform-editing routine. The spliced words will be presented in an environment of four different signal-to-noise ratios (i.e., +30, +20, +10, 0 dB) to ascertain the signal-to-noise ratios for 50 per cent scores. A difference between the S/N score for synthetic and natural words of 10 dB or less will be considered acceptably close perceptual quality in the

synthetic versions.

In a second future study, the unaccented natural sentence will be compared to the unaccented synthetic sentence, and the accented natural sentences as spoken by the Spanish speakers with moderate and strong accents in English will be compared to the most accented (i. e., all cues present) corresponding synthetic sentences on a 7-point scale.

Each of the 30 moderately accented stimulus sentences and the 30 strongly accented stimulus sentences was paired with the unaccented (native English) synthetic sentence and randomized by an ONLINE routine developed by C. J. Miller for the Perkin-Elmer 8/32 minicomputer. That is, the standard stimulus was presented, followed by an approximately one-second inter-stimulus-interval of silence, followed by the comparison (accented) stimulus. The total duration of each trial was approximately 9 sec., and there was a 6 sec. interval between trials. In addition, the unaccented version (standard stimulus) was paired with itself in order to determine whether the listener was rating the unaccented sentence as accented or not. Thus, there were 31 trials each for the moderate and strong condition.

Stimulus tape generation was accomplished by outputting randomizations of the synthetic tokens through

12-bit D/A conversion and audio processing equipment associated with a Perkin-Elmer 8/32 minicomputer at Kresge Laboratory. Digitized rate was 10k samples/sec, and recordings were made on an Ampex AG-440 audio tape recorder (1/2 track, 7-1/2 ips.) after anti-alias filtering at 5 kHz (low pass). The tape recorded sentences were generated in four different randomized sequences. That is, each stimulus tape consisted of four randomized sequences of 31 stimulus pairs for a total of 124 stimulus pairs per tape. There were two moderately accented stimulus tapes and two strongly accented stimulus tapes.

The original recordings on reel-to-reel tape (Ampex studio quality recording tape) were dubbed onto cassettes (Sony UCX-S90) using an Ampex reel-to-reel recorder connected to a Sony WDM-6 cassette recorder.

The same procedure was followed to make a training tape for each accent condition by dubbing the stimulus trials consisting of (1) the standard stimulus paired with itself ten times (no fixed inter-trial interval of silence) from reel-to-reel to cassette tape; (2) the standard stimulus followed by the most extremely accented stimulus (M31 or S31) ten times (no fixed inter-trial interval of silence). There was a 10 sec. interval of silence between the first and second sets of 10 trials.

Perception Experiment

Subjects. Serving as judges were 42 L.S.U. students who were monolingual speakers of American English with normal speech and hearing. Subjects participated in two 45-minute sessions on different days; at one session they heard one of the two "moderate-accent" tapes, while at the other session, they heard one of the two "stong-accent" tapes. Half of the subjects heard the moderate-accent tape first; the other half heard the strong-accent tape first. In addition, half of the subjects heard one set of randomizations while the other half heard the other set of randomizations of a given level of accent (moderate or strong) first.

Procedure. In a sound-attenuated booth, each individual judge evaluated the stimuli. Stimuli were played on a Sony WDM-6 cassette recorder at a fixed comfortable listening level and heard over Beyer headphones.

Training. At the beginning of each experimental session, the judges were given copies of written instructions (see Appendix B) to follow along as

they were read aloud by the investigator. Questions were answered. In addition, oral examples were given of Spanish-accented English to demonstrate the cues under investigation.

A training tape was played with 10 repetitions of the standard sentence (no cues present) followed by itself and 10 pairs of the standard sentence followed by the most extremely accented sentence (all cues present) of the moderate or strong accent (depending on which tape was to be played). Thus, listeners could gain an understanding of the range of accentedness and the quality of speech to be presented. The investigator answered questions.

The subjects then heard through Beyer headphones at a fixed comfortable listening level most of one of the randomly-ordered stimulus sequences (29 trials) to familiarize them with the quality of the speech and the range of accentedness. The judges filled out a response sheet during the training session identical in format to the one they were to use during the actual testing situation but shorter. That is, they rated the 29 stimulus pairs by indicating how different the accented version in each pair was from the unaccented version on a ten-point scale. For each stimulus pair, subjects also indicated how sure they were of their judgment on a scale like the following:

NOT SURE SURE

1 2 3 4 5

---/---/---/---/---

Subjects were allowed to stop the tape whenever they wished to ask the investigator (present in the sound booth) questions.

Testing. After the training session, problems and questions were discussed, and the stimulus tape was rewound. Then each judge was given another response sheet and he/she heard through Beyer headphones the same stimulus tape again. Subjects were told that they could stop the tape any time if the task became tedious, if they needed to hear a particular trial again, or if they had any questions. The investigator remained in the sound booth with the subject throughout the experiment to observe and to aid the subject if necessary. The judges rated the 124 stimuli by indicating on a ten-point equal-appearing scale how different the accented version in each pair was from the unaccented version (1=same. . .10=most different). They also indicated confidence in their judgments on the 5-point scale described above.

Analysis of Judges' Ratings

Questions were addressed as follows. All analyses involved a one-way analysis of variance with repeated measures. Post-hoc analyses of judgments of moderately and strongly accented stimuli were carried out separately for questions 1-4.

1. Is the effect on accent judgments of divergences along several dimensions cumulative?

Using a repeated measures anlysis of variance, mean ratings of stimuli containing one cue (i.e., S2M-S6M and S2S-S6S) were compared to mean ratings of stimuli containing two cues (i.e., S7M-S16M and S7S-S16S). Mean ratings of one and two cues were compared to mean ratings of stimuli containing three cues (i.e., S17M-S26M and S17S-S26S). Then, mean ratings of stimuli containing one, two and three cues were compared with mean ratings of stimuli containing four cues (i.e., S27M-S30M and S27S-S30S). Finally, mean ratings of stimuli with one, two, three, and four cues were compared to mean ratings of a stimulus with all five cues (i.e, S31M and S31S).

2. Do certain perturbations influence degree of

accent to a greater extent than others? Do listeners respond to the same or different cue for moderate and strong accents?

Each subject's ratings of the four repetitions of sentences with a single cue modification were averaged. These data, along with the standard stimulus, were analyzed in two analyses of variance--one for moderate stimuli and one for strong stimuli.

Using post-hoc Tukey tests, the stimulus containing FO (S2M) was compared to those with VOT, VQS, VQU and DUR (S3-S6M respectively); the stimulus containing VOT (S3M) was compared to stimuli with VQS, VQU and DUR (S4-S6M); the stimulus containing VQS (S4M) was compared to stimuli with VQU and DUR (S5-S6M); the stimulus containing VQU (S5M) was compared to the stimulus with DUR (S6M). In addition the stimulus with no cue manipulation (S1M) was compared to all other single-cue stimuli. Parallel comparisons were carried out with strongly accented stimuli.

A factor analysis was also carried out on the data for both moderate and strong data to ascertain whether any single factor(s) stood out in determining accentedness.

3. Does divergence from L2 phonetic norms for vowel production affect acceptability judgments to a greater

extent than divergences from the phonetic norms for the consonants (cf. Flege, 1984a)?

Using a t-test, mean ratings of moderate stimuli containing a consonantal cue (VOT) (i.e., S3M, S7M, S10-13M, S19M, and S30M) were compared to those stimuli with vocalic cues (VQS and VQU) singly and in combination (i.e., S4-S5M, S8M-S9M, S11-S12M, S14-16M, S20-S21M, S25-S26M, S29M). The same procedure was carried out for strong stimuli.

4. Does divergence from L2 norms for suprasegmental features (i.e., FO [intonation] and DUR [rhythm]) affect acceptability judgments to a greater extent than divergences from L2 norms for segmental features (i.e., VOT, VQS, and VQU)?

Using a t-test, mean ratings of moderate stimuli containing FO and DUR cues (i.e., FO: S2M, S7M, S10M, S19M; DUR: S6M and S13M) were compared to stimuli containing VOT, VQS and VQU cues (i.e., S4-S5M, S11-S12M, V13M, V21M). The same procedure was carried out for strong stimuli.

5. Are strongly accented stimuli judged as more accented than corresponding moderately accented stimuli?

Each subject's accentedness ratings for the 4 repetitions of the 30 moderately accented stimuli were

averaged to form a mean rating of moderate sentences. A similar average was computed for each subject's ratings of the strong stimuli. The mean ratings were then compared via a correlated-samples t statistic.

6. Are listeners more confident rating stimuli that they consider "more accented?"

Pearson correlation coefficients were calculated to compare the subjects' confidence ratings to their stimulus ratings. A highly positive correlation would indicate that the more accented a stimulus was rated (i.e., the higher the rating given), the more confident the subject was about the rating (i.e., the higher the confidence rating given). A highly negative correlation would indicate that the more accented a stimulus was rated, the less confident the subject was about his judgment.

7. Are listeners better able to perceive a strong accent more readily than a moderate one?

Standard deviations for 1-, 2-, 3-, 4- and 5-cue moderate stimuli were compared to those for corresponding strong stimuli. A large increase in S.D. would indicate less certainty on the part of the listeners.

8. Which of the stimuli were least and most preferred as accented by listeners? Can listeners be grouped according to cue preference?

Individual subject scores (i.e., mean of ratings for each stimulus by each subject) were examined, and any mean of 5.00 or greater was recorded in a table; means of 7.00 or greater were indicated by an asterisk next to the value. The number of values greater than or equal to 5.00 and 7.00 per stimulus were noted.

In addition, means for each subject were examined to determine individual patterns of response. Subsequent analysis of subject means for stimuli containing the two previously determined most potent cues for the moderate accent condition were studied. A 2:1 ratio was determined to indicate subjects' preference for one of the two cues.

Finally, means for stimuli in 1-, 2-, 3-, and 4-cue groups were rank ordered from highest to lowest in order to show which ones the subjects as a group considered to be the most and least accented in each cue group. (See Table 14 in Chapter III).

# CHAPTER III

# RESULTS

## Moderate versus Strong Accent

Each subject's accentedness ratings for the four repetitions of the 30 moderately accented sentences were averaged to form a mean rating of moderate sentences. A similar average was computed for each subject's ratings of the strongly accented sentences. These mean ratings were then compared via a correlated-samples t statistic. The results ($t=.45$, df 41, $p > .05$) showed that there was no statistically reliable difference in the group mean ratings of the moderately accented sentences ($\bar{X}=4.65$, SD=.64) and the strongly accented sentences ($\bar{X}=4.71$, SD=.89).

Each subject's confidence ratings were similarly averaged to form mean confidence ratings for the moderately accented and strongly accented productions. The group mean ratings for moderately accented ($\bar{X}=2.79$, SD=.47) and strongly accented ($\bar{X}=2.80$, SD=.42) sentences did not differ reliably ($t=.21$, df 41, $p > .05$).

In summary, judges' average ratings of the moderately accented and strongly accented sentences did not

differ in either accentedness or confidence.

## Number of Cues

Each subject's ratings of the moderately accented sentences containing one cue were averaged to form a single-cue score. Similarly, individual-subject averages for the sentences containing two, three, four and five cues were averaged. These scores were then analyzed in a repeated measures analysis of variance. The results of this analysis ($F=217$, $df=4,164$, $p < .01$) indicate a significant difference across number of acoustic cues manipulated. Table 9 shows that the average accentedness ratings increase steadily from one to five cues.

---

TABLE 9. Average Accentedness Ratings for sentences containing one, two, three, four and five cues

| | | NUMBER OF CUES | | | | |
|---|---|---|---|---|---|---|
| CUE STRENGTH | | 1 | 2 | 3 | 4 | 5 |
| Moderate | $\bar{X}$ | 3.08 | 4.08 | 5.08 | 6.62 | 7.49 |
| | SD | 1.00 | 1.08 | 0.97 | 1.05 | 1.43 |
| Strong | $\bar{X}$ | 3.60 | 4.51 | 4.69 | 5.76 | 6.58 |
| | SD | 1.13 | 0.89 | 0.82 | 1.38 | 1.30 |

---

Similar analysis for the strongly accented sentences also showed a significant difference across number of cues (F=57.26, df=4,164, p <.01).

As expected, the more accent bearing cues present, the greater the mean overall accentedness. Notice the orderly almost monotonic increase in rated accentedness as the number of cues varied increased from one to five for the moderate accent and the somewhat less regular increase in accentedness for the strong-accent stimulus set (see Figure 6).

It is remarkable that the greater physical size of the cue change for the strong accent produced rated accentedness only slightly greater for 1- and 2-cue stimuli but a much less degree of rated accentedness for 3, 4-, and 5-cue stimuli.

Confidence Ratings Compared to Stimulus Ratings

Inspection of Table 10 shows a trend for the correlation coefficients between rated accentedness and confidence to become more positive and larger in size, going from 1- to 5-cue stimuli at least for the moderate accent. For the strong accent, there is an increase between 1- and

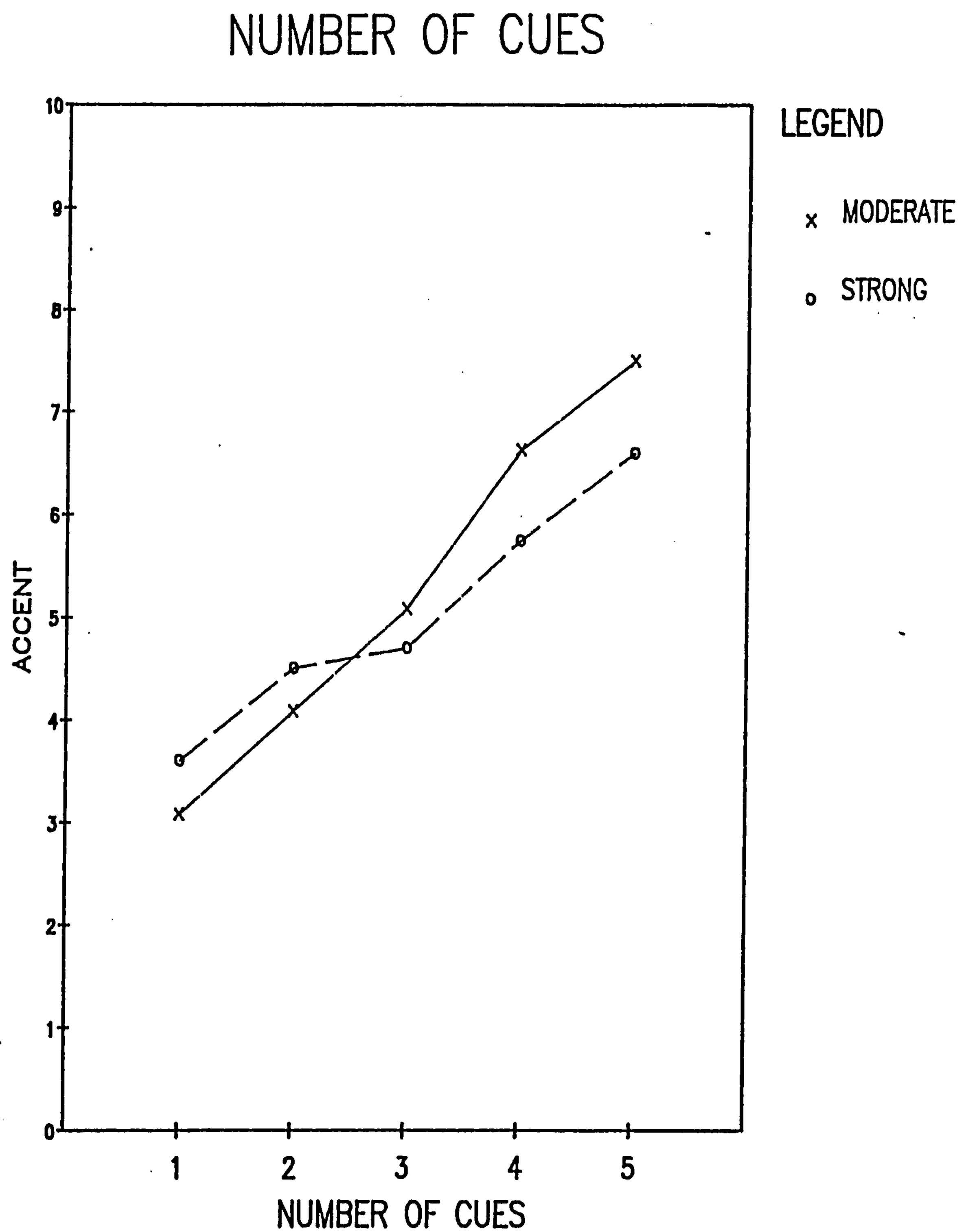
NUMBER OF CUES
LEGEND
MODERATE
STRONG
ACCENT
0
1
2
3
4
5
6
7
8
9
10
1
2
3
4
5
NUMBER OF CUES

Figure 6

2-cue stimuli, then a decrease between 2- and 3-cue stimuli, followed by a large increase between 3- and 4-cue stimuli, and a slight increase between 4- and 5-cue stimuli.

---

TABLE 10. Correlation between stimulus ratings and confidence ratings

| | | MODERATE ACCENT | | STRONG ACCENT | |
|---|---|---|---|---|---|
| No. Cues | Stimulus | Pearson Corr.Coeff. | Group Mean | Pearson Corr.Coeff. | Group Mean |
| 0 | 1 NO CUES | -0.2340 | | -0.0790 | |
| 1 | 2 FO | 0.0357 | | -0.1465 | |
| | 3 VOT | -0.3178 | | 0.0402 | |
| | 4 VQS | 0.2321 | | 0.0923 | |
| | 5 VQU | -0.0857 | | 0.0850 | |
| | 6 DUR | -0.2728 | -0.0817 | 0.0231 | 0.0188 |
| 2 | 7 FOVOT | 0.0495 | | 0.2680 | |
| | 8 FOVQS | 0.2178 | | 0.3643 | |
| | 9 FOVQU | 0.1726 | | 0.0131 | |
| | 10 FODUR | 0.1611 | | 0.1046 | |
| | 11 VOTVQS | 0.1752 | | 0.3851 | |
| | 12 VOTVQU | -0.1134 | | 0.1283 | |
| | 13 VOTDUR | -0.1757 | | -0.0201 | |

| | | | | | |
|---|---|---|---|---|---|
| | 14 VQSVQU | 0.2800 | | 0.2699 | |
| | 15 VQSDUR | 0.1739 | | 0.2989 | |
| | 16 VQUDUR | -0.0040 | 0.0937 | 0.1380 | 0.1950 |
| 3 | 17 FOVOTVQS | 0.3705 | | 0.2246 | |
| | 18 FOVOTVQU | 0.2603 | | 0.1665 | |
| | 19 FOVOTDUR | 0.1844 | | 0.2434 | |
| | 20 FOVQSVQU | 0.5314 | | 0.1054 | |
| | 21 FOVQSDUR | 0.2168 | | 0.2381 | |
| | 22 VOTVQSVQU | 0.2789 | | 0.2005 | |
| | 23 VOTVQSDUR | 0.2531 | | 0.2728 | |
| | 24 VOTVQUDUR | 0.1440 | | -0.0373 | |
| | 25 VQSVQUDUR | 0.2817 | | 0.1588 | |
| | 26 FOVQUDUR | 0.2648 | 0.2786 | 0.0036 | 0.1576 |
| 4 | 27 FOVOTVQSVQU | 0.4153 | | 0.4346 | |
| | 28 FOVOTVQSDUR | 0.2841 | | 0.3170 | |
| | 29 FOVQSVQUDUR | 0.3767 | | 0.2718 | |
| | 30 VOTVQSVQUDUR | 0.4939 | 0.3925 | 0.3011 | 0.3311 |
| 5 | 31 ALL CUES | 0.4293 | | 0.3694 | |

A highly positive correlation indicates that the more accented a stimulus was rated (i.e., the higher the

rating given), the more confident the subject was about his rating (i.e., the higher the confidence rating given). A highly negative correlation indicates that the more accented a stimulus was rated, the less confident the subject was about his rating. Thus, for the moderate accent, the more cues present, the higher the correlation, indicates that the more accented the stimulus was, the more confident subjects were of their rating. For the strong accent, the presence of 4 cues, as opposed to 3 cues, caused a drop in subjects' confidence level.

In the moderate accent condition for 1-cue stimuli, stimulus 4 (VQS) received the most positive correlation, while stimulus 3 (VOT) received the most negative correlation. For 2-cue stimuli, stimulus 14 (VQSVQU) received the most positive correlation, while stimulus 13 (VOTDUR) received the most negative correlation. For 3-cue stimuli, stimulus 20 (FOVQSVQU) showed the most positive correlation, while stimulus 24 (VOTVQUDUR) received the least positive correlation. Finally, for 4-cue stimuli, stimulus 30 (VOTVQSVQUDUR) received the highest positive correlation, while stimulus 28 (FOVOTVQSDUR) received the lowest positive correlation.

In the strong accent condition for 1-cue stimuli, stimulus 4 (VQS) again showed the most positive correlation;

however, stimulus 2 (FO) received the most negative correlation. For 2-cue stimuli, stimulus 11 (VOTVQS) received the most positive correlation, while stimulus 13 (VOTDUR) received the only negative correlation. For 3-cue stimuli, stimulus 23 (VOTVQSDUR) received the most positive correlation, while stimulus 24 (VOTVQUDUR) received the only negative correlation. For 4-cue stimuli, stimulus 27 (FOVOTVQSVQU) showed the most positive correlation, while stimulus 28 (FOVOTVQSDUR) received the least positive correlation.

Comparison of Individual Cues

Each subject's ratings of the four repetitions of sentences bearing a single cue manipulation were averaged. These data, along with the unaccented sentences, were analyzed in two analyses of variance--one for the moderate-strength cues, and one for the strongly accented cues. Table 11 shows the group's mean ratings for these sentences.

TABLE 11. Group mean accentedness ratings for the single-cue sentences

| | | Cue | | | | | |
|---|---|---|---|---|---|---|---|
| | | None | FO | VOT | VQS | VQU | DUR |
| Moderate | $\bar{X}$ | 1.07 | 3.66 | 1.98 | 5.54 | 2.33 | 1.88 |
| | SD | 1.11 | 1.94 | 1.54 | 1.84 | 1.79 | 1.22 |
| Strong | $\bar{X}$ | 2.65 | 3.34 | 3.99 | 4.88 | 2.75 | 3.04 |
| | SD | 2.32 | 1.47 | 2.58 | 1.74 | 1.80 | 2.10 |

Analysis of variance for both the moderately ($F=52.84$, $df=5,205$, $p <.01$) and strongly ($F=9.65$, $df=5,205$, $p <.01$) accented sentences indicate a significant difference.

Clearly, VQS for the moderate group received highest rated accentedness (5.54), FO second (3.66), VQU (2.33), VOT (1.98) and DUR last (1.88). Reassuringly, the control stimulus for the moderate accent was rated at a value of 1.07 as compared to the ideal rating of 1.0. However, the control stimulus for the strong accent had a mean rating of 2.65, lower than that for any of the five single cues but far above 1.0. Post-hoc Tukey tests for the moderate stimuli showed that VOT and DUR were not significantly

different from the O-cue stimulus, whereas FO, VQS, and VQU were significantly different from the O-cue condition. FO was found to be significantly different from all other cues, and VQS was significantly different from VQU, VOT and DUR, which were not significantly different from each other.

Post-hoc Tukey tests for the strongly accented stimuli showed that only VQS and VOT were significantly different from the O-cue condition. VQS was found to be significantly different from FO, VQU and DUR, and VOT was significantly different from VQU.

Factor Analysis

Factor analysis, employing orthogonal Varimax rotation of factors, was used to determine significant factors underlying judgments of accentedness for the moderately cued sentences.

The factor analysis recovered 8 underlying dimensions, but the lion's share of the variance, 69%, was accounted for by the first two dimensions, emphasizing their importance to subject response.

TABLE 12. Factor Analysis of moderately accented stimuli for Factor 1

| HIGH POSITIVE VALUES | | | NEGATIVE VALUES | | |
|---|---|---|---|---|---|
| Stimulus | Description | Value | Stimulus | Description | Value |
| 2 | FO | 0.70623 | 1 | 0 cues | -0.03746 |
| 7 | FOVOT | 0.78549 | 4 | VQS | -0.06213 |
| 8 | FOVQS | 0.72259 | 11 | VOTVQS | -0.03990 |
| 9 | FOVQU | 0.74093 | 13 | VOTDUR | -0.03057 |
| 10 | FODUR | 0.70752 | 14 | VQSVQU | -0.01629 |
| 18 | FOVOTVQS | 0.76668 | 15 | VQSDUR | -0.22394 |
| 19 | FOVOTDUR | 0.70682 | 23 | VOTVQSDUR | -0.22394 |
| 26 | FOVQUDUR | 0.72190 | 25 | VQSVQUDUR | -0.10758 |
| | | | 30 | VOTVQSVQUDUR | -0.18671 |

In Table 12, high positive values for Factor 1 are associated with stimuli containing FO, while negative values are associated with stimuli not containing FO.

TABLE 13. Factor Analysis of moderately accented stimuli for Factor 2

| HIGH POSITIVE VALUES | | | NEGATIVE VALUES | | |
|---|---|---|---|---|---|
| Stimulus | Description | Value | Stimulus | Description | Value |
| 4 | VQS | 0.67527 | 2 | FO | -0.1934 |

| | | | | | |
|---|---|---|---|---|---|
| 11 | VOTVQS | 0.77791 | 7 | FOVOT | -0.06509 |
| 15 | VQSDUR | 0.60331 | 9 | FOVQU | -0.07054 |
| 22 | VOTVQSVQU | 0.63834 | 10 | FODUR | -0.15066 |
| 30 | VOTVQSVQUDUR | 0.62987 | 18 | FOVOTVQS | -0.08257 |
| | | | 19 | FOVOTDUR | -0.14133 |
| | | | 21 | FOVQSDUR | -0.17549 |
| | | | 26 | FOVQUDUR | -0.17549 |

In Table 13, high positive values for Factor 2 are associated with stimuli containing VQS, while negative values are associated (in all but two cases) with stimuli not containing VQS. The conclusion can be drawn from the data in Tables 13 and 14 that the two major factors contributing to perception of accentedness in the moderate accent condition are directly related to FO and VQS.

A factor analysis of the strongly accented stimuli revealed no factor associated with stimuli containing a particular cue or cues. It seems that the prominence of VQS and VOT in single-cue stimuli did not extend to stimuli where they were combined with other cues.

The scatterplot for factors 1 and 2 for the moderate accent is shown in Figure 7. As the two dimensions accounting for the greatest degree of variance (i.e., scatter), dimensions 1 and 2 present the most interpretable

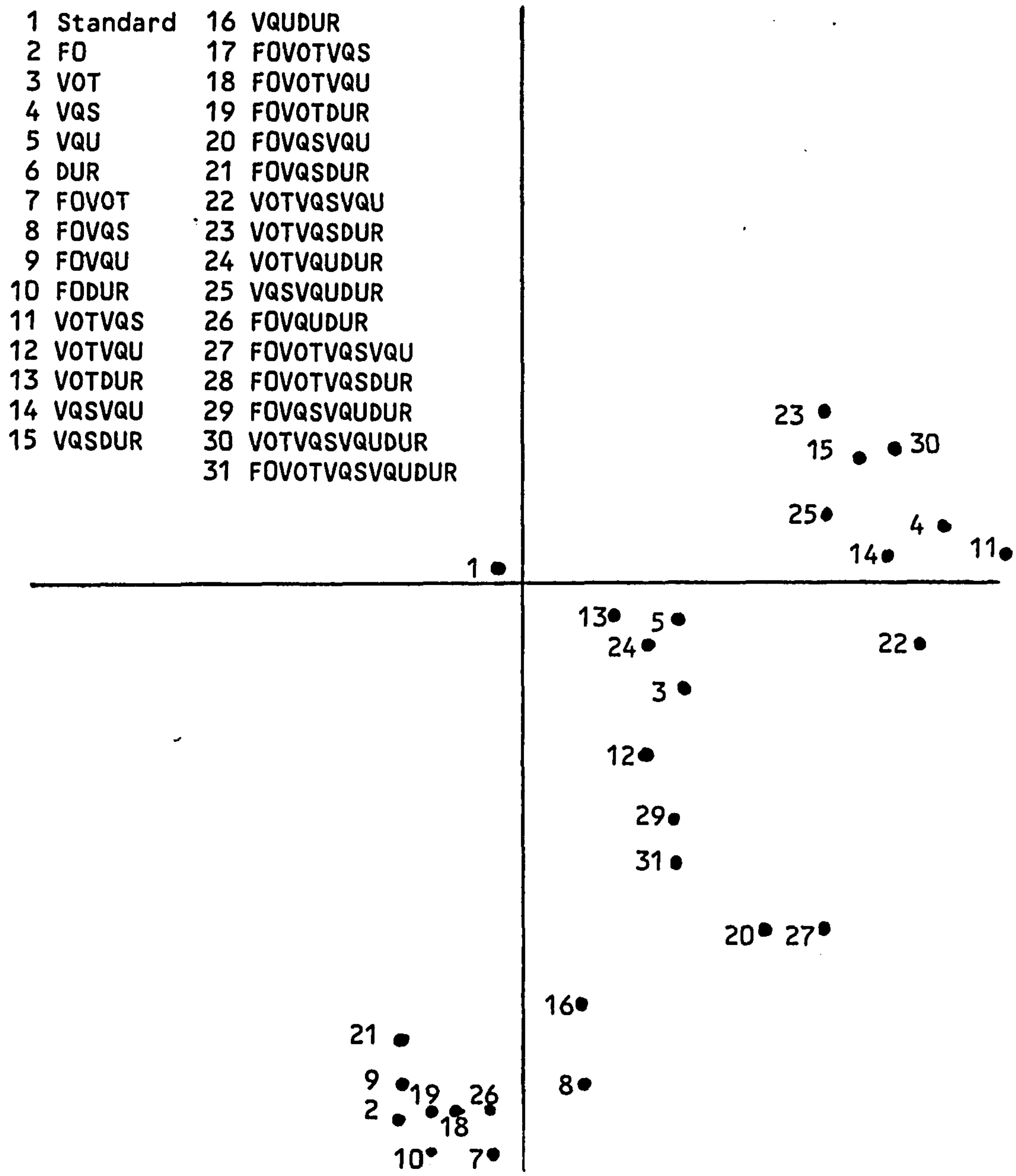
1 Standard
2 FO
3 VOT
4 VQS
5 VQU
6 DUR
7 FOVOT
8 FOVQS
9 FOVQU
10 FODUR
11 VOTVQS
12 VOTVQU
13 VOTDUR
14 VQSVQU
15 VQSDUR
16 VQUDUR
17 FOVOTVQS
18 FOVOTVQU
19 FOVOTDUR
20 FOVQSVQU
21 FOVQSDUR
22 VOTVQSVQU
23 VOTVQSDUR
24 VOTVQUDUR
25 VQSVQUDUR
26 FOVQUDUR
27 FOVOTVQSVQU
28 FOVOTVQSDUR
29 FOVQSVQUDUR
30 VOTVQSVQUDUR
31 FOVOTVQSVQUDUR
23
15
30
25
4
14
11
1
13
5
24
22
3
12
29
31
20
27
16
21
9
19
26
2
18
8
10
7

Figure 7

data. The scatterplot shown in Figure 7 shows three clusters of stimuli. One cluster of stimuli, centered on dimension 1, consists entirely of stimuli containing FO in some combination. Cluster 3, centered about a high positive value of dimension 2, represents stimuli containing VQS and not FO in various combinations with other stimuli. An intermediate group of stimuli, cluster 2, showing intermediate positive values on both dimensions 1 and 2, represent primarily stimuli containing both FO and VQS.

There are exceptions, however. In group 1, stimuli 18 and 21 should be in group 3 but are not. In the intermediate group, stimuli 5, 12, 13, and 16 don't have VQS or FO. In group 3, only stimuli 20, 27, 29, and 31 satisfy the two combined values, whereas 3, 5, 12, 13, 16, and 24 do not.

Group 1 on dimension 1, it would appear, is interpretable as a group of stimuli sharing FO as a primary common cue. Group 2 shares VQS as a primary cue.

Thus, it seems that subjects reliably classified the moderate stimuli along at least two underlying dimensions: dimension 1, related to intonation contour or melody, and dimension 2, related to stressed vowel quality. These data agree nicely with the fact that the strongest single cues (i.e., those with the highest means) are FO and VQS (See

discussion above on Single Cues).

Consonantal versus Vocalic Cues

Each subject's ratings of moderately accented sentences containing VOT were averaged to form a consonantal-cue score. Similarly, individual-subject averages for the sentences containing VQS were averaged to form a vocalic-cue score. These scores were compared via a a t-test ($t=19.05$, df 167, $p > .05$), which showed that there was a statistically reliable difference in group-mean ratings of the consonantal stimuli ($\bar{X}=3.12$, SD=1.24) and the vocalic stimuli ($\bar{X}=5.33$, SD=1.30).

When mean ratings of moderately accented sentences containing VOT (consonantal cue) were compared to those containing VQU (vocalic cue), a t-test ($t=6.63$, df 167, $p > .05$) showed that there was no statistically reliable difference in group-mean ratings of the consonantal stimuli ($\bar{X}=4.44$, SD=1.09) and the vocalic stimuli ($\bar{X}=5.04$, SD=1.22).

Then mean ratings of moderate stimulus sentences containing VOT (consonantal cue) were compared to those containing either VQS or VQU (vocalic cues). A t-test ($t=20.37$, df 167, $p > .05$) showed that there was a statistically reliable difference in group-mean ratings of

the consonantal stimuli ($\bar{X}$=2.92, SD=1.33) and vocalic stimuli ($\bar{X}$=4.76, SD=1.20).

The same procedure was followed for the strongly accented stimulus sentences. First, mean ratings of stimuli containing VOT were compared to those containing VQS. A *t*-test (t=7.55, df 167, p >.05) showed that there was no statistically reliable difference in group-mean ratings of the consonantal stimuli ($\bar{X}$=4.01, SD=1.46) and vocalic stimuli ($\bar{X}$=5.21, SD=1.28).

Next, when mean ratings of strongly accented sentences containing VOT were compared to those containing VQU, a *t*-test (t=5.10, df 167, p >.05) showed that there was no statistically reliable difference in group-mean ratings of the consonantal stimuli ($\bar{X}$=4.83, SD=1.16) and vocalic stimuli ($\bar{X}$=4.29, SD=1.01).

Finally, mean ratings of strong-stimulus sentences containing VOT were compared to those containing VQS or VQU. A *t*-test (t=1.96, df 167, p >.05) showed that there was no statistically reliable difference in group-mean ratings of the consonantal stimuli ($\bar{X}$=4.30, SD=1.76) and vocalic stimuli ($\bar{X}$=4.59, SD=0.98).

In summary, for moderately accented sentences, judges' average ratings of consonantal and vocalic stimuli differed when VQS was involved. However, there was no

difference in the average ratings for strongly accented consonantal and vocalic stimuli.

Comparison of Segmental and Suprasegmental Cues

Mean ratings of moderately accented sentences containing VOT, VQS and VQU (segmental cues) were compared to those containing FO and DUR (suprasegmental cues). A t-test ($t=4.34$, df 167, $p > .05$) showed that there was no statistically reliable difference in group-mean ratings of the suprasegmental stimuli ($\bar{X}=3.44$, SD=1.59) and the segmental stimuli ($\bar{X}=4.18$, SD=1.51).

When mean ratings of strongly accented sentences containing segmental cues were compared to those containing suprasegmental cues, a t-test ($t=1.92$, df 167, $p > .05$) showed again that there was no statistically reliable difference in group-mean ratings of the suprasegmental stimuli ($\bar{X}=4.01$, SD=1.48) and the segmental stimuli ($\bar{X}=4.25$, SD=1.30).

Thus, judges' average ratings of segmental and suprasegmental stimuli did not differ between moderately and strongly accented conditions.

Individual Subject Differences

In order to establish individual differences in preference for certain cues and cue combinations, individual subject ratings for all stimuli were scrutinized (See tables in Appendix C). Arbitrarily, in order to establish a level showing stimulus preference, a mean rating of 5.00 or greater had to be achieved in order to indicate a strong response. Accordingly, values of 5.00 or greater appear in the tables and values of 7.00 or greater are indicated by asterisks.

As expected, for single cues, stimulus 4 (VQS) was an overwhelming favorite in both moderate and strong accent conditions: it received 12/42 (i.e. 12 of a total of42) ratings of 7.00 or above and 26/42 ratings of 5.00 or above for moderate stimuli, and it received 5/42 ratings of 7.00 or above and 21/42 ratings of 5.00 or above for strong stimuli.

Subjects can be classified into three groups: low raters (those with between 1 and 10 ratings of 5.00 and above, moderate raters (those from 11-20), and high raters (those over 20). The majority of subjects (29/42) were found to be moderate raters, while 6/42 were low raters and 7/42 were high raters.

Subjects can also be classified as either conservative or liberal raters on the basis of whether they had only 0-2 of their total mean ratings of 7.00 or higher (i.e., conservative) or only 0-2 of their total mean ratings of less than 7.00 (i.e., liberal). With regard to the strong stimuli, no subjects were found to be liberal raters; however, 20 subjects could be considered to be conservative raters. With regard to the moderate stimuli, 2 subjects were found to be liberal raters, while 7 were found to be conservative. Those who were conservative in their ratings of the moderate stimuli were almost always (5/7) conservative in their ratings of the strong stimuli as well.

A total of 23/42 subjects demonstrated a strong preference for F0, VQS or both. Strong preference is claimed if a given cue appears in most, if not all, of the stimuli receiving a mean rating of 5.00 or greater for a particular subject. For the moderate stimuli, 7/42 showed a strong preference for F0, 10/42 for VQS, and 6/42 for a combination of the two. For the strong stimuli, 3/42 showed a strong preference for F0, 6/42 for VQS, and 4/42 for the two in combination. Eleven subjects showed a strong preference in both accent conditions. Seven of the 11 subjects showed a strong preference for the same cue in both moderate and strong accent conditions (1 for F0, 3 for VQS,

and 3 for both). Four subjects preferred different cues in different accent conditions: 1 preferred VQS for moderate stimuli but FO for strong stimuli, 1 preferred FO for moderate stimuli but VQS for strong stimuli, 1 preferred VQS for moderate stimuli but a combination of the cues for strong stimuli, and 1 preferred the combination of cues for moderate stimuli but VQS for strong stimuli.

Ranking All Cues

Table 14 shows all cues for both accent conditions and their mean ratings ranked from highest to lowest for 1-, 2-, 3- and 4-cue combinations.

---

TABLE 14. A ranking of all one-, two-, three-, four- and five-cue stimuli for moderate and strong conditions

| | MODERATE ACCENT | | | STRONG ACCENT | | |
|---|---|---|---|---|---|---|
| Cues | Description | Mean | S.D. | Description | Mean | S.D. |
| 0 | No cues | 1.070 | 1.112 | No cues | 2.649 | 2.324 |
| 1 | VQS | 5.542 | 1.838 | VQS | 4.875 | 1.738 |
| | FO | 3.661 | 1.944 | VOT | 3.994 | 2.578 |
| | VQU | 2.333 | 1.786 | FO | 3.339 | 1.467 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | VOT | 1.976 | 1.539 | DUR | 3.042 | 2.105 |
| | DUR | 1.881 | 1.219 | VQU | 2.750 | 1.795 |
| 2 | VOTVQS | 5.482 | 2.479 | FOVQS | 6.452 | 2.375 |
| | VQSVQU | 5.452 | 2.608 | VOTVQS | 5.244 | 2.878 |
| | VQSDUR | 5.030 | 2.605 | FOVOT | 5.077 | 2.747 |
| | FOVQU | 4.446 | 2.635 | VQSDUR | 4.940 | 2.815 |
| | FOVQS | 4.333 | 2.582 | VQSVQU | 4.655 | 2.585 |
| | FODUR | 4.095 | 2.406 | FODUR | 4.464 | 2.633 |
| | FOVOT | 4.089 | 2.392 | FOVQU | 4.065 | 2.583 |
| | VQUDUR | 3.482 | 2.377 | VQUDUR | 3.821 | 3.334 |
| | VOTVQU | 2.887 | 2.361 | VOTVQU | 3.417 | 2.513 |
| | VOTDUR | 1.518 | 1.801 | VOTDUR | 2.935 | 2.710 |
| 3 | FOVQSVQU | 7.095 | 1.799 | FOVOTVQS | 5.833 | 2.901 |
| | FOVOTVQS | 6.821 | 1.881 | FOVQSVQU | 5.685 | 2.714 |
| | VOTVQSVQU | 5.565 | 2.431 | FOVQSDUR | 5.387 | 2.967 |
| | VQSVQUDUR | 5.542 | 2.711 | FOVOTDUR | 5.202 | 2.209 |
| | VOTVQSDUR | 5.274 | 2.485 | VOTVQSVQU | 4.798 | 2.744 |
| | FOVQUDUR | 4.810 | 2.524 | VOTVQSDUR | 4.732 | 3.028 |
| | FOVOTVQU | 4.488 | 2.526 | FOVOTVQU | 4.690 | 2.592 |
| | FOVQSDUR | 4.298 | 2.655 | VQSVQUDUR | 4.500 | 2.541 |
| | FOVOTDUR | 4.113 | 2.348 | FOVQUDUR | 3.387 | 2.390 |
| | VOTVQUDUR | 2.786 | 2.283 | VOTVQUDUR | 2.786 | 2.670 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 4 | FOVOTVQUDUR | 7.155 1.889 | | FOVOTVQSVQU | 6.667 2.182 | |
| | FOVOTVQSVQU | 6.792 2.047 | | FOVOTVQSDUR | 5.631 3.079 | |
| | FOVOTVQSDUR | 6.274 2.090 | | FOVOTVQUDUR | 5.429 2.915 | |
| | VOTVQSVQUDUR | 6.262 2.430 | | VOTVQSVQUDUR | 5.238 2.396 | |
| 5 | ALL CUES | 7.494 1.864 | | ALL CUES | 6.595 2.340 | |

For the moderately accented stimuli, among the 1-cue group, stimulus 4 (VQS) had the highest mean (5.542) while stimulus 6 (DUR) had the lowest (1.881). Among the 2-cue group, stimulus 11 (VOTVQS) had the highest mean (5.482) and stimulus 13 (VOTDUR) had the lowest (1.518). Among the 3-cue group, stimulus 20 (FOVQSVQU) had the highest mean (7.095) while stimulus 24 (VOTVQUDUR) had the lowest (2.786). Among the 4-cue group, stimulus 29 (FOVOTVQUDUR) had the highest mean (7.155) while stimulus 30 (VOTVQSVQUDUR) had the lowest (6.262).

For the strongly accented stimuli, among the 1-cue group, stimulus 4 (VQS) again had the highest mean (4.875); stimulus 5 (VQU) had the lowest mean (2.750). In the 2-cue group, stimulus 8 (FOVQS) had the highest mean (6.452) while stimulus 13 (VOTDUR) again had the lowest (2.935). In the 3-cue group, stimulus 17 (FOVOTVQS) had the highest mean

(5.833) while stimulus 24 (VOTVQUDUR) again had the lowest (2.786). In the 4-cue group, stimulus 27 (FOVOTVQSVQU) had the highest mean (6.667) and stimulus 30 (VOTVQSVQUDUR) again had the lowest mean (5.238).

The cues in each group with the highest means are characterized by the presence of FO and/or VQS, while the cues in each group with the lowest means are characterized by a lack of FO. In addition, all the low-mean stimuli for the moderate accent have DUR and/or VOT, while those for the strong accent have VQU and/or DUR.

Notice that in the moderate accent condition, standard deviations for the no-cue and all-cue stimuli are low, and so are those for at least one of the extremes (first or last ranked) for each number-of-cue group. This is not the case, however, for the strong stimuli.

Mean ratings of standard deviations for the 30 moderately accented sentences and the standard sentence were compared to mean ratings for the 30 strongly accented sentences and the standard sentence via a _t_-test. The result ($t=5.33$, df 31, $p > .05$) showed that there was a significant difference in the group mean ratings of the moderately accented sentences ($\bar{X}=2.28$, SD=0.32) and the strongly accented sentences ($\bar{X}=2.69$, SD=0.27). Thus, the judges' ratings of strong stimuli exhibited more variability

than their ratings of moderate stimuli.

# CHAPTER IV

# DISCUSSION

## Nature of Accent

Natural speech production yields a constellation of cues which may be ranked, by individuals, into a hierarchy of perceptual prominence. In an attempt to establish a hierarchy of those cues in order to study foreign accent (German/English), Barry (1974) noted that the use of natural speech presents difficulties in analysis which cannot by overcome by multiple correlation techniques because the incredibly large number of variables becomes impossible to interpret. He says, "natural stimuli just do not allow the certainty that all the parameters that have interacted in the judgments have been included in the correlation" (p. 87). For example, voice characteristics have been shown to interact with judgments of natural speech and confound results of matched guise experiments (Arthur, 1974).

This study approaches the problem of establishing a cue hierarchy for foreign accent (Spanish/English) by using synthetic speech to investigate a small portion (i.e., five cues) of the much larger set of possible cues. The number

of cues yet to be investigated remains unknown. Even though the values of the five cues studied were intended to mirror those of naturally produced accented speech samples, it is not certain that the synthesized speech reflected optimum values of the natural cues.

The choice of which cues to study was determined by an exhaustive survey of the literature and by the limitations of synthetic speech. For example, while the devoicing of /z/ is widely known to signal a foreign accent (Spanish and other languages, Flege and Hillenbrand, 1985a) in English, it is difficult to reproduce acceptable voicing for /z/ in synthetic speech. Likewise, the fricative allophones of voiced stops /b, d, g/, which occur in word-final position in the English of Spanish speakers (Flege and Davidian, 1984: 323), are quite difficult to reproduce satisfactorily in synthetic speech.

Vowel quality was a primary cue chosen for study for several reasons. First, the English vowel system, with 11 monophthongal phonemes and tense/lax distinctions (see Fischer-Jorgensen, 1985 for a discussion), is one of the most difficult aspects of the language for foreign speakers to learn (Bowen and Stockwell, 1965). Vowel quality, as indicative of a Spanish accent in English, has been studied by Sawyer (1975). Second, vowels are one of the most

perceptually salient components of the speech signal (i.e., vowels are higher in intensity due to the open vocal tract and longer in duration than consonants). Third, it has been observed by Heike (1969) that judgments of deviant accent depend to a large extent on vowel quality and are little affected by other elements in the speech signal (cited in Barry, 1974: 65).

Vocal fundamental frequency, which is realized perceptually as intonation at the sentence level, was chosen as a cue for the following reasons. First, the claim was to be explored that "suprasegmental errors are largely responsible for the perception of foreign accent" (Flege, 1979) in light of supporting evidence found in the literature (Ioup, 1984; Metcalf, 1972; Palmer, 1976; Wingfield, 1975; Huggins, 1972). Second, it has been shown that analyses required to determine vowel quality and pitch operate in a mutually dependent fashion (Miller, 1978). Third, fundamental frequency interacts with the less-dominant cue of duration in the perception of stress (Morton and Jassem, 1965) and thus, of rhythm, a major differentiating factor between English and Spanish (Pike, 1946; Delattre, 1966; Dauer, 1983; Pointon, 1980; Hoequist, 1983a,b,c; and others). Some researchers disagree that Spanish is "isosyllabic" (Navarro Tomas, 1916-1922; Manrique

and Signorini, 1983; Gili Gaya, 1940), and some disagree that isochrony exists in English (Shen and Peterson, 1962; Nakatani et al., 1981; Faure et al., 1980; O'Connor, 1965; Dauer, 1983). Rhythmic interference has been cited as a major factor for the English dialect of Spanish speakers by Wolfram (1973) and Hutchinson (1973).

Since the prosodic cue of duration contributes to the perception of rhythm, it was chosen for the present study. Also said to contribute to the stress-timed rhythm of English (i.e., pattern of alternating weak and strong stresses, cf. Faure, Hirst and Chafcouloff, 1980) is vowel reduction (Hoequist, 1983b; Dauer, 1983), so that it, too, became a cue for consideration. The quality of the reduced (neutral) vowel in English has been discussed by, among others, Peterson and Barney (1952) and Jakobson, Fant and Halle (1952).

Voice onset time (i.e., duration value between burst and onset of voicing for the following vowel) for syllable-initial voiceless stops was the final cue chosen for study. The difference between VOT values for Spanish and English have been studied by Abramson and Lisker (1973), and VOT, as indicative of a Spanish accent in English, has been extensively studied by Williams (1977a,b;1979;1980), Elman (1977) and Flege and Hammond (1982).

Individual Cues

In this study, a factor analysis showed fundamental frequency (FO) to be the dominant factor cuing a moderate Spanish accent in English. Compare the three tone levels described by Navarro Tomas (1966) for Spanish with the four pitch levels in English described by Trager and Smith (1951); see also Stockwell and Bowen (1965). Specifically, the levelling of pitch peaks and weakening of sharp terminal fall (see p. 82 of this study) was found to be the most potent cue in signalling a moderate Spanish accent in English, accounting for 41.4% of the variance in the data. This finding offers support for the hypothesis that suprasegmental features make the greatest contribution to perception of foreign accent (Flege, 1979).

For the strong single-cue stimuli, FO ranked third, after VQS and VOT; in fact, its mean rating was not significantly different from the O-cue condition. It would appear that the artificiality of the strong stimuli suppressed the strong cue value for FO, while elevating that of VOT to great prominence. Four factors suggest that the results for strongly accented stimuli should be interpreted with extreme caution. The fact that (1)

confidence-accentedness correlations were reduced, (2) the single-cue variability was nearly doubled in 3/5 cases for strong stimuli (see Table 11), (3) the accentedness rating was 2.65 for the standard stimulus when included with strong-accent single cues, and (4) VOT was elevated to much greater prominence than the widely acknowledged strong FO in the moderate condition, all suggest that the strong stimuli were confoundingly unnatural in perceptual quality. This, in turn, suggests that only the results for moderate accent can be realistically interpreted.

Stressed vowel quality (VQS) was found to be the second most prominent factor in the perception of a Spanish accent in English, accounting for 27.2% of the variance in the moderate data. It was also found to be the most prominent in single-cue strongly accented stimuli. This supports Heike's claim (see above) that judgment of a foreign accent depends to a large extent on vowel quality; however, the finding of FO as the dominant cue for moderate accent contradicts his claim that judgment is affected by other elements of the speech signal. And what is more, the factor analysis supports the interpretation, albeit cautiously, that the two factors interact somewhat, since FO and VQS stimuli combined formed a group intermediate to each single dimension. It is not surprising that vowel quality

is a prominent cue in the English of Hispanics since the Spanish vowel system consists of just five monophthongal vowels, which must be used as a basis for the approximation of the 11-monophthongal-vowel system for English. Note that although listeners had different preferences with regard to VOT and FO depending on level of accent, VQS remained a strongly preferred cue at both moderate and strong levels.

Acoustically, the vowels [a] and [e] are more salient than [i] and [u] due to the degree of vocal tract opening; however, the shift of [ae] to a more [a]-like vowel to approximate a Spanish accent in English was less perceptible than the shift of [I] to a more [i]-like vowel according to informal reports by subjects. The perceptual prominence of the high front vowel over other vowels in this study agrees with the results of Flege and Hammond (1982) and Brennan and Brennan (1981b). Note that [i] was not substituted for [I] or [a] for [ae], but rather, these sounds represented interlangauge approximations to English [I] and [ae] respectively (cf. Beebe, 1980).

The location of the vowel in the sentence may be one explanation for the phenomenon; the [i]-like vowel ("tickets") occurs later in the sentence than either of the occurrences of the [a]-like vowel ("Kansas" and "basketball") and thus, it may have been remembered because

it was more recent. However, it was followed by an additional word string (i.e., "cost ten dollars now") which could have interfered with short term memory for vowel quality.

The finding of FO and VQS as the strongest cues contributing to the perception of a moderate accent is interesting in light of a model of speech proposed by Gunnar Fant (which refers back to an early Bell Systems paper by Crandall). According to this model, the speech wave is a slowly drifting, powerful carrier wave that is periodically frequency and amplitude modulated. This carrier signal is seen to consist of vowels with certain levels of FO and amplitude. Modifications in amplitude and frequency occur when consonants are imposed on the wave. This model is relevant to the present study because subjects could detect changes in the powerful carrier-wave cues (FO, VQS) much more readily than changes in other cues In fact, one reason that there was no difference between segmental and suprasegmental cues, at least for the moderate data, is probably that the two strong cues were placed in competition with each other, and thus, one may have cancelled out the other's dominating effect.

Unstressed vowel quality (VQU), quite logically, ranked third in single-cue mean ratings for moderate stimuli

and fourth for strong stimuli. It was not as strong as vowel quality of stressed and, hence, more prominent vowels (cf. Jones, 1932), possibly because listeners have been shown to attend more to stressed syllables than to unstressed syllables (Allen, 1972a,b; Cutler and Foss, 1973).

It has been claimed that reduced vowels in English are inherently shorter than full vowels (Dauer, 1983), (but note that Faure et al. (1980) found that stressed and unstressed syllables were approximately equal in duration; however, Nakatani and Schaffer (1978) found a positive correlation between an increase in stress level and syllable duration). In Spanish, syllables containing stressed vowels have also been found to be longer than those containing unstressed vowels; however, the difference is not as great as in English (Dauer, 1983; Delattre, 1966). It has also been claimed that syllable-timed languages like Spanish do not have regularly reduced vowels in unstressed position (Dauer, 1983). It was initially hypothesized that the lack of reduced vowels in Spanish would make it difficult for Spanish speakers to produce reduced vowels in English. This difficulty was expected to be manifested in longer duration for reduced vowels in accented speech than in standard English speech (as well as in a difference in formant

structure). This in turn would have a disruptive effect on rhythm, causing it to be less stress-timed and more syllable-timed. However, when durations were measured on spectrograms of standard and accented English production of the stimulus sentence, it was found that there was almost no durational difference for reduced vowels, only a difference in formant structure (i.e., quality) contrary to the findings of Dauer (1980) and Delattre (1966).

Somewhat surprisingly, for the moderate stimuli, VOT proved to be a fairly weak cue in light of the amount of attention devoted to it in the literature (see previous section of this Discussion). It may be that since it occupies less time than any of the other cues (maximum of 95 ms. for the moderate accent condition), they may override it in judgment. Such a brief acoustic cue would also decay more quickly in short-term memory than cues of longer duration (e.g., long, strong acoustic signals such as vowels may remain in precategorical acoustic storage (PAS) for up to 2 sec.) (Pisoni and Sawusch, 1975; Repp, Healy and Crowder, 1979).

It can only cautiously be concluded that for strong single-cue stimuli, VOT proved to be second in perceptual prominence only to VQS, concurring with reports in the literature of its strong effect on perception of

accentedness. Perhaps, an optimal value for perception of VOT change was reached only in the strong-accent condition. This issue needs to be further explored (see Implications section of this chapter).

Similar to the results reported by Huggins (1978), it might be that the shortened closure duration values for accented stimuli may have been perceived simply as an increase in speaking rate in the unstressed syllables preceding the shortened closure (e.g., the last two syllables of "basketball" preceding the [t]-closure of "tickets"). Note that tempo of the stimulus sentences was fairly slow, being based on read speech rather than spontaneously produced speech (Pointon, 1980). In addition, the tempo of the naturally produced strongly accented sentence was slower and more halting than that of the moderately accented sentence. When the strongly accented sentence was reproduced synthetically, however, extra pauses between words were removed because they were considered to be an extraneous cue signalling accent.

It was hypothesized that duration (DUR) would be a fairly perceptible cue in signalling accentedness due to its intended effect of shifting the rhythm of the sentence from stress-timing to syllable-timing. The change in rhythm in the middle of the sentence was intended to contradict

listeners' rhythmic expectancies which were established upon hearing the first syllables of the utterance and induce them to reanalyze the sentence (Wingfield, 1975; Aaronson, 1968). Such a reanalysis should have caused the subjects to become consciously aware of a difference between the accented version of the sentence and the standard pronunciation.

Unfortunately, the data show that DUR was the least perceptible cue for the moderately accented stimuli and one of the least perceptible for the strongly accented stimuli, so it is probable that it was not tested adequately. However, this result agrees with Eisendoorn (1983b, 1984), who found that altering durations of vowels produced by Dutch native speakers in English so that they were more English-like "had little effect on acceptability judgments."

The fact that most rhythmic alterations occurred in the medial part of the 4.5-second sentence may have allowed their image in short-term memory to be masked by the information that followed in the sentence. According to Klatt and Cooper (1975), sensitivity to changes in duration is influenced by a backward masking effect of any of the following words.

In addition, strong cue FO suddenly shifted (i.e., dropped) at the end of accented sentences when there had been almost no previous change in that cue. An answer may

also lie in the interaction between duration and FO in signalling stress, which occurs both in Spanish (Manrique et al., 1982; Massone, 1982) and English (Bolinger, 1961); perturbations in one cue may affect the other and confound judgment. Faure et al. (1980) showed that it is not possible to perceive rhythmic patterns in English without accompanying variations in pitch. Perhaps, then, the levelling of many FO variations to approximate Spanish intonation may have interfered with perception of DUR. Furthermore, while DUR and FO changes result in stress (and thus, rhythm) perception, FO has been shown to be the stronger of the two cues (Fry, 1958; Rigault, 1962; Jassem et al., 1970; Nooteboom, 1978). VQS has also been shown to interact with duration to signal stress as has intensity (Fischer-Jorgensen, 1985; Fry, 1955; Morton and Jassem, 1965). Furthermore, duration adjustments occurred only for four syllables which may not have resulted in enough rhythmic change to be perceived.

Another possible confounding factor might be that the stimuli were biased toward English; that is, accented stimuli consisted of four occurrences of each of the modifications (except FO which was not discrete) of an English sentence. If the stimuli had consisted of English modifications of a basically Spanish sentence (e.g.,

imposing English rhythm on a Spanish sentence), different results may have been obtained. Finally, the fact that there were additional English cues present which were not under investigation may have biased the listener to only respond to the strongest cues perceived.

## Confidence Ratings

The use of confidence rating has been applied to speech perception skills in English-speaking hearing-impaired listeners and second-language learners of English by Yule, Yanz and Tsuda (1984), who state: "Yanz (1984) proposed that confidence ratings may add a new dimension to our knowledge of speech perception by quantifying the way in which listeners cope with uncertainty in their understanding of spoken communication" (ibid.: 4).

This study employed the confidence rating format to correlate listeners' ratings of stimuli with their confidence in those ratings. A major finding was that an increase in number of cues was accompanied by a concomitant increase in mean confidence ratings. As expected, for the moderate accent, confidence ratings climbed monotonically with number of cues; however, for the strong accent, the increase was somewhat less consistent (see Figure 6).

Furthermore, listeners were most confident for highest or lowest ratings in each number-of-cue group, indicating that those stimuli were consciously considered the most or least salient in each group. The most salient 1- and 2-cue stimuli contained VQS and 3-cue stimuli contained both F0 and VQS; least salient 1- and 2-cue stimuli contained VOT and DUR and 3-cue stimuli contained both VOT and DUR.

A change in the pattern of choosing stimuli containing a preferred cue for both moderate and strong stimuli occurred for 4-cue stimuli, which may indicate that subjects experienced a stimulus overload at this point; that is, there may have been so many strong cues, and given the cue-impoverished and rather artificial quality of synthetic speech, that the subjects were experiencing unnatural speech quality. In other words, they may have reacted to the increased quantity of cues but were uncertain of their judgments because they had not encountered such an overload in natural speech perception.

## Cue Strength

In this study, extreme cue values were taken from a single informant (i.e., the Spanish speaker judged to have a strong accent in English) and reproduced in synthetic

speech. The quality of the synthetic speech, while satisfactorily intelligible and clearly accented, was still less than optimum (cf. Luce et al., 1981, for a discussion regarding the cue-poor nature of synthetic speech). The most highly intelligible sounds were vowels, while stops and fricatives, particularly voiced sounds, were the least intelligible. These results agree with those of Clark (1983), who used a 12-formant serial formant speech synthesizer in his intelligibility study for synthetic speech. However, the use of a sentence context in the study undoubtedly resulted in greater segment intelligibility (Luce et al., 1981; Lindblom, 1982). While all single moderate cues clearly exhibited a well-differentiated ranking in degree of accentedness (4.5:1 range), in contrast, single strong cues did not achieve either as large a mean rated accentedness as moderate stimuli (5.54 VQS-moderate versus 4.88 VQS-strong) or as great a range of accentedness (2.23). (See Table 11).

As strong cues were concatenated into groups of 2, 3, 4, and 5, rated accentedness indeed grew, but with greater variability than for moderate-cue stimuli, as Table 10 shows. This indicates that the judgment of an ensemble of cues is a highly natural act, so natural that it overcame abnormal cue strength of the individual cues entering into

the ensemble.

## Subject Differences

An examination of individual subject differences revealed no outstandingly divergent groups. Almost everyone preferred individually one or both of the two strongest cues, FO and VQS. A few select individuals showed extraordinary preferences (e.g., for the moderate accent, 4 subjects preferred VOT and 1 preferred DUR at the single-cue level but these preferences did not extend to combined-cue groups). Four subjects exhibited a preference for different cues for the different levels of accent; however, their preferences were restricted to the two dominant cues, FO and VQS.

Individual differences in raters show two small groups, enthusiasts (i.e., "liberals") (16.6%) and conservatives (48%), in terms of how often they were willing to prefer strongly or not prefer strongly a given moderate cue. It is possible that the small group of subjects classified as high or low raters may have been overreacting in one direction or the other to the unnaturalness of the synthetic speech, and thus, their responses might not be considered trustworthy.

Some difficulty was encountered in the performance of a few of the subjects who participated in this study. One subject refused to use the extreme ends of the 10-point rating continuum because of a lack of certainty about her judgments, so that the data was discarded. Another subject demonstrated inconsistency (by assigning both extreme values (1 and 10) to the four repetitions of most of the stimuli) and inattentativeness to the task (by assigning only one value, i.e., 3, to confidence ratings for almost all stimuli and by doodling excessively on the response sheet). A third subject complained that all stimuli, including the standard sentence, sounded accented; however, her responses were consistent and, therefore, included as data. Recall that even in experiments using natural speech (Asher and Garcia, 1969; Scovel, 1981), inexperienced listeners had a tendency to identify sentences produced by native speakers as having been produced by a non-native speaker. Several subjects commented that they could not hear much difference between the standard sentence and the moderately accented stimuli, but as the results demonstrate, there was a perceptible difference.

The question arises of whether subjects' responses to synthetic acoustic stimuli can be generalized to apply to natural speech and, in particular, to perception of foreign

accent. In defense of an affirmative reply, the acoustic cues chosen for study were selected because they had been observed in naturally produced speech and reported in the literature, and the synthetic cues were modeled after the accented English speech produced by native speakers of Spanish judged to have moderate and strong accents on the basis of their pronunciation of English.

## Underlying Dimensions

Factor analysis of the moderate data shows the melody (FO) and vowel quality (VQS) as dimensions 1 and 2 to be the strongest. Both factors are temporally long and acoustically powerful. Both factors represent phonological interference. Factor 1 (FO) has to do with the implementation of an architypal intonation type, while Factor 2 has to do with the phonetic realization of a phonemic inventory. The stimuli in which the two cues are combined fit into an intermediate space (see Figure 7). The strength of the variance accounted for suggests that for Spanish, these are two major sources of accent in conversational speech.

Failure to find an underlying dimension relating to rhythm was disappointing. A major objective of this study

was to investigate perception of rhythm which necessitated the use of an utterance of greater duration than that of an isolated word. Ladefoged (1975: 91) explains that for the citation form of a word at least one syllable is fully stressed and has no reduction in vowel quality. However, in connected speech, many changes may take place (e.g., smaller words will usually be unstressed, the vowel reduced to schwa, vowel duration shortened, etc.). Shields, McHugh and Martin (1974) showed that when target words were spliced out of fluent speech, the differences between stressed and unstressed syllables disappeared. Armstrong and Ward (1926) comment that word stress for an isolated word is often dropped in connected speech. Thus, to get an accurate representation of English rhythm (i.e., the pattern of alternating strong and weak stresses), it is desirable to study connected speech, not citation forms. In addition, it has been demonstrated by Flege (1984a: 704) that foreign accent detection is somewhat better for relatively long (i.e., phrase length) compared to short excerpts of speech. It is also necessary to use relatively long stretches of speech if one is to study intonation as has been done here. Recall that the domain of intonation is most probably the breath group as described by Lieberman (1965, 1985). Few studies besides the research reported here have used a

context of connected speech which was produced synthetically to study suprasegmental features (Beckman, Hertz and Fujimura, 1983; Pierrehumbert, 1979).

Implications

The eradication of foreign accent as the goal ESL instruction is somewhat unrealistic. More commonly, the aim is to foster comprehensible speech. Having identified intonation and stressed vowel quality as markers of foreign accent, it might be suspected that the inability to understand a foreign speaker may be partially caused by such features. Consequently, teachers of ESL (and of other foreign languages as well) might concentrate on those areas as a means of improving comprehensibility (that is, make the learner's speech approximate what the native speaker expects to hear in terms of FO and VQS).

High quality speech synthesis could be used to create stimulus templates differing in only one variable for use in discrimination exercises. This would be particularly useful in teaching intonation, where utterances with the intonation of the learner's L1 would be played next to utterances with L2 intonation for comparison. Keeping all variables except intonation constant would be impossible

using natural speech. The student could also learn L2 intonation patterns by imitating an utterance in which a fixed vowel such as [a] is frequency modulated to reproduce different intonation patterns in the absence of words (Lieberman, 1967) which could represent a distracting influence. Then, words could be added and the student could imitate an utterance and view the difference in intonation contour visually by means of a visipitch or similar visual display of the intonation contour.

The quality (formant frequencies) of that fixed (stressed) vowel might be altered to mimic a learner's inappropriate vowel production. Then that utterance would be played next to an utterance with the appropriate L2 production of the vowel so that the student could learn to discriminate between the two. The learner would next attempt to imitate the utterance with the appropriate vowel quality and then put that vowel into words, phrases and sentences where coarticulartory effects would operate. Additionally, a fixed syllable such as [ma] could be used to produce reiterant speech stimuli (Nakatani and Schaffer, 1978) which integrate L2 intonation and rhythm for imitation purposes.

This study naturally has implications for future research as well. It demonstrates a procedure to test the

intelligibility of synthetic speech. Much the way visual acuity is tested using lenses that are gradated in curvature, fine nuances intended to improve the naturalness of synthetic speech can be introduced and tested by comparison with a "standard sentence" or each other. Confidence ratings can be employed to offer a subjective reaction on the part of the listener to cues which he may be perceiving subconsciously.

Of course, better quality synthetic speech is necessary for its use in experiments such as this one, so that unnaturalness does not introduce unwanted cues in addition to those under investigation. Improved quality would also allow for the examination of cues that cannot presently be studied using synthetic speech because of the difficulty involved in stimulus production (e.g., devoiced /z/, spirantized stop allophones). In addition, high quality synthetic speech, used in experiments of foreign accent perception, would eliminate unwanted cues such as those pertaining to voice characteristics of individual speakers.

It would be interesting to investigate suprasegmental cues, controlling for the length of the stimuli as Flege (1984a) has done for VOT and the vowel /u/. For example, context could be shrunk from the length of a

sentence to the breath group to the noun phrase in order to determine the optimum length for stimuli. Different utterances need to be employed in a single study, rather than using only one utterance throughout, so as to determine whether the present results are unique to the stimulus.

In a future study, a titration of cues should be performed in order to ascertain the full range of values for a given cue. For instance, the English production of a number of Spanish speakers could be examined for a range of duration values in order to further study rhythm. It is possible that the rhythm of the two speakers employed in this study was not representative of the norm, or it is possible that foreign speakers of English may produce some cues, e.g. duration, more accurately than they produce other cues. Perhaps another Spanish speaker would approximate English vowel quality more accurately than rhythm, so that DUR would become more perceptually salient as an accent-bearing cue. The range of values for a cue could be reproduced in precisely controlled increments using high-quality synthetic speech to determine the point at which English speakers perceive a rhythmic change. It would also probably be desirable in a study of duration not to introduce changes in F0 which could conflict with rhythmic changes. The cue-titration approach should also be used for

further study of VOT in order to learn the optimum values for perception of that cue in various contexts. In addition, the effects of place of articulation, position in utterance, and stress on VOT should be taken into consideration (Westbury and Keating, 1980).

Summary

This study has demonstrated that synthesized sentences can be reliably rated for cue modifications indicative of a moderate Spanish accent in English. An increase in the number of cues was found to result in an increase in the rating of accentedness and in subjects' confidence in their ratings. Individual ratings were found to be representative of group mean ratings. Suprasegmental as well as segmental cues were rated, and for the moderate stimuli, the suprasegmental cue FO was found to be perceptually the most prominent with the segmental cue VQS being the next most prominent cue. The weakness of rhythmic cue DUR was probably an artifact of the experimental procedure. The general unreliability of strong-accent data may most likely be attributed to errors in the production of the strong synthetic stimuli.

Conclusion

The strength of the present study lies in (1) the use of a sentential frame for the stimuli; (2) the manipulation of multiple cues, segmental and suprasegmental, in a precise, controlled manner; and (3) the use of confidence ratings to tap subjective reactions to the stimuli. The weakest components of the study involve (1) the use of one informant each to represent standard, moderately accented and strongly accented English pronunciation; (2) the unnaturalness of the synthetic stimuli; and (3) the use of a single sentence for all stimuli. Despite these drawbacks, this is the first study, to our knowledge, to apply synthetic continuous speech to the perception of foreign accent.

NOTES

<1>This same issue has also been addressed with regard to apraxic speech by Buckingham (unpublished ms.: 124), who says:

> Faulty control of the velum in apraxia is likely to impart some added nasalization to vowels, which in turn, will cause F1 to rise. As it has been demonstrated by Wright, 1975; Ohala, 1975, the auditory correlate of a rising F1 has generally been assumed to be a lowering of the vowel. The change of the acoustic cue in F1 is ultimately caused in this case by an apraxic articulation, but the categorical perception of vowel lowering on the part of hearers will cause them to think that the patient substituted a lower vowel for a higher one.

<2>The technique of cross-splicing introduces two perturbations into prosody: (1) it produces an abrupt change in both pitch and rhythm at the splice points, and (2) it gives an inappropriate placement of the intonational cues to the syntactic boundaries (Darwin, 1975: 179). These claims

cast suspicion on the results of cross-splicing experiments such as that of Flege (1984), who claim that spectral information alone is sufficient to cue accent in /t/ produced by French speakers of English.

<3>This model is based on a lexical prototype model (Rosch, 1973; 1978) in which objects in the world are categorized in comparison to internal prototypes which represent a category's core properties. Prototypes are often developed through experience with many categories. Specific exemplars of a category are accepted or rejected based on how closely they conform to the prototype.

<4>Information structure is the ordering of the text, independently of its construction in terms of sentences, clauses, etc., into units of information on the basis of the distinction into given and new: what the speaker is treating as information that is recoverable to the hearer (given) and what he is treating as non-recoverable (new) (Halliday and Hasan, 1976: 27).

<5>However, amount of experience may affect children's production of L2 stops. Williams (1979, 1980) found that both 8-10 and 14-16-year-old Puerto Rican

children who had lived 3-3.5 years on the U.S. mainland produced /p/ with longer VOT values (i.e., stops that were more English-like by about 10 ms) than children who had lived less than 6 months on the U.S. mainland. At the same time, it was evident that (with length of residence on the U.S. mainland held constant) the younger children produced /p/ with VOT values 5-25 ms longer, and therefore, more English-like than the older children.

## REFERENCES


Aaronson, D. (1968). Temporal course of perception in an immediate recall task, J. Exp. Psychol., 76: 129-140.

Aaronson, D. (1976). Performance theories for sentence encoding: Some qualitative observations, J. Exp. Psychol.: Human Perception and Performance, 2: 42-55.

Abramson, A. (1962). The vowels and tones of standard Thai: Acoustical measurements and experiments, Indiana Research Center in Anthropology, Folklore and Linguistics, Publication 20.

Abramson, A. and L. Lisker (1973). Voice-timing perception in Spanish word-initial stops, J. Phonetics, 1, 1: 1-8.

Ainsworth, W. (1972). Duration as a cue in the recognition of synthetic vowels, JASA, 51: 648-651.

Allen, G. D. (1972a). The location of rhythmic stress beats in English: An experimental study I, Lang. and Speech, 15: 72-100.

Allen, G. D. (1972b). The location of rhrthmic stress beats in English: An experimental study II, Lang. and Speech, 15: 179-195.

Allen, J. (1976). Synthesis of speech from unrestricted text, *Proceedings of the IEEE*, *4*: 433-442.

Anisfeld, E. and W. Lambert (1964). Evaluational reactions of bilingual and monolingual children to spoken languages, *J. Abnorm. Soc. Psychol.*, *69*: 89-97.

Anisfeld, M., N. Bogo, and W. Lambert. (1962). Evaluational reactions to accented English speech, *J. Abnorm. Soc. Psychol.*, *65*: 223-231.

Armstrong, L. and I. Ward (1926). *Handbook of English Intonation*. Leipzig and Berlin: B. G. Teubner.

Arthur, B., D. Farrar and G. Bradford. (1974). Evaluational reactions of college students to dialect differences in the English of Mexican-Americans, *Lang. and Speech*, *17*: 255-270.

Asher, J. and R. Garcia (1969). The optimal age to learn a language, *Mod. Lang. J.* *53*: 334-341.

Atkinson, J. R. (1973). Aspects of intonation in speech: Implications from an experimental study of fundamental frequency. Ph.D. dissertation, U. Connecticut.

Baird, S. (1969). Employment Interview Speech: A Social Dialect Study in Austin, Texas. Univ. Texas at Austin Ph.D. diss.

Bansal, R. K. (1966). The intelligibility of Indian English: measurements of the intelligibility of connected speech, and sentence and word material presented to listeners of different nationalities. Unpublished Ph.D. Dissertation, London University.

Barry, W. (1974). Language background and the perception of foreign accent, J. Phonetics, 2: 65-89.

Beckman, M., S. Hertz, and O. Fujimura (1983). SRS pitch rules for Japanese, Working Papers of the Cornell Phonetics Laboratory, 1: 1-16.

Beebe, L. (1980). Myths about interlanguage phonology. Paper presented at the 14th annual TESOL convention, San Francisco.

Bloomfield, L. (1933) Language. New York: Holt Rinehart and Winston.

Bolinger, D. L. (1954). English prosodic stress and Spanish sentence order, Hispania, 37: 152-156.

Bolinger, D. L. (1958). A theory of pitch accent in English, Word, 14: 109-149.

Bolinger, D. L. (1961). Three analogies, Hispania, 44: 134-137.

Bolinger, D. L. (1965). Pitch accent and sentence rhythm. In I. Abe and T. Kanekiyo (eds.) Forms of Eng-

lish: Accent, Morpheme, Order. Cambridge: Harvard Univ. Press, pp. 139-180.

Bolinger, D. L. (1972). Accent is predictable (if you're a mind-reader), Language, 48: 633-644.

Bregman, A. S. and J. Campbell (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones, J. Exp. Psychol., 89:244-249.

Brennan, E., E. Ryan and W. Dawson (1975). Scaling of apparent accentedness by magnitude estimation and sensory modality matching, J. Psycholing. Res., 4: 27-36.

Brennan, E. and J. Brennan (1981a). Accent scaling and language attitudes: Reactions to Mexican American English speech, Lang. Speech, 24: 207-221.

Brennan, E. and J. Brennan (1981b). Measurements of accent and attitude toward Mexican American speech, J. Psycholing. Res., 10: 487-501.

Bresnan, J. (1971). Sentence stress and syntactic transformation, Language, 48: 257-281.

Buckingham, H. (1985). The case for an apraxia of language: An essay in the theory of apraxia. Unpublished ms.

Caramazza, A., G. Yeni-Komshian, E. Zurif and E. Carbone (1973). The acquisition of a new phonological con-

trast: The case of stop consonants in French-English bilinguals, JASA, 54: 421-428.

Cardenas, D. (1960). Introduccion a una comparacion fonologica del espanol y del ingles. Washington, D. C.: Center for Applied Linguistics.

Chomsky, N. and M. Halle (1968). Sound Pattern of English. New York: Harper and Row.

Clark, J. E. (1983). Intelligibility comparisons for two synthetic and one natural speech source, J. Phonetics, 11: 37-49.

Classe, A. (1939). The Rhythm of English Prose. Oxford: Basil Blackwell.

Cohen, A. and S. Nooteboom (eds.) (1975) Structure and Process in Speech Perception. Heidelburg: Springer-Verlag.

Cohen, A. and J. 't Hart (1967). On the anatomy of intonation, Lingua, 19: 177-192.

Collier, R. (1972). From pitch to intonation. Unpublished Ph.D. dissertation, University of Louvain.

Collier, R. and J. t'Hart (1975). The role of intonation in speech perception. In A. Cohen and S. Nooteboom (eds.) Structure and Process in Speech Perception. New York: Springer-Verlag, pp. 107-123.

Contreras, H. (1980). Sentential stress, word order and the notion of subject in Spanish. In L. Waugh and C. van Schooneveld (eds.) The Melody of Language: Intonation and Prosody. Baltimore: University Park Press, pp. 45-53.

Cooper, W. E. and J. M. Sorensen. (1981). Fundamental Frequency in Sentence Production. NY: Springer-Verlag.

Cutler, A. (1975). Rhythmic factors in the determination of perceived stress. Paper presented at the 89th Meeting of the Acoustical Society of America (Austin, Texas).

Cutler, A. and D. J. Foss (1973). The importance of lexical item stress for lexical access, Paper presented at the 44th annual meeting of the Midwestern Psychological Association, May 10-12, 1973.

Dallett, K. M. (1964). Intelligibility and short-term memory in the repetition of digit strings, JSHR, 7: 362-368.

Daneš, F. (1960). Sentence intonation from a functional point of view, Word, 16: 34-54.

d'Anglejan, A. and R. Tucker. (1973). Sociolinguistic correlates of speech style in Quebec. In R. Shuy

and R. Fasold (eds.) *Language Attitudes: Current Trends and Prospects*. Washington, D. C.: Georgetown U. Press.

Darwin, C. J. (1975). On the dynamic use of prosody in speech perception, in A. Cohen and S. G. Nooteboom (eds.) *Structure and Process in Speech Perception*. *Berlin*: *Springer-Verlag*, *pp*. *178*-*194*.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *J*. *Phonetics*, *11*: 51-62.

Delattre, P. (1966). A comparison of syllable length conditioning among languages, *International Review of Applied Linguistics*, *4*: 183-198.

Delattre, P., C. Olsen and E. Poenack (1962). A comparative study of declarative intonation in American English and Spanish, *Hispania*, *45*: 233-241.

Delgutte, B. (1978). Technique for the perceptual investigation of FO contours with application to French," *J*. *Acoust*. *Soc*. *Am*. *64*: 1319-1332.

Dickerson, L. (1975). The learner's inter-language as a system of variable rules, *TESOL Q*., *9*: 401-407.

Dimitrijević, N. R. and D. Djordević (1971). The reliability of the subjective assessment of the pupils' pronunciation of English as a foreign language,

*Internat'l. Rev. of Appl. Ling*, *9*: 245-265.

Dorman, M., J. Cutting and L. Raphael (1975). Perception of temporal order in vowel sequences with and without formant transitions, *J. Exp. Psychol.: Human Perception and Performance*, *2*: 121-129.

Eimas, P. (1985). The perception of speech in early infancy, *Scientific American*, *252*: 46-52.

Eisentstein, M. (1979). The development of dialect discrimination and stereotyping in adult learners of English as a second language. C.U.N.Y. Ph.D. diss.

Eisendoorn, B. A. G. (1980). Vowel durations as a cue to a Dutch foreign accent in English CVC-words, Progress Report of the Institute of Phonetics, Univ. of Utrecht, Holland, 5, 64-86.

Eisendoorn, B. A. G. (1983b). Production and perception of English vowel duration by Dutch speakers of English. In M. P. R. Van den Broecke and A. Cohen (eds.) *Proceedings of the Tenth International Congress of Phonetic Sciences*, Utrecht (Dordrecht, Foris), pp. 673-676.

Eisendoorn, B. A. G. (1984). Tolerances of durational properties in British English vowels. Ph.D. thesis, Univ. Utrecht.

Elman, J., R. Diehl and S. Buchwald (1977). Perceptual switching in bilinguals, JASA, 62: 971-974.

Fauer, G., D. J. Hirst, and M. Chafcouloff. (1980). Rhythm in English: Isochronism, pitch and perceived stress. In L. R. Waugh and C. H. van Schooneveld (eds.), The Melody of Language. Baltimore: University Park Press, pp. 71-19.

Fischer-Jorgensen, E. (1985). Some basic vowel features, their articulatory correlates, and their explanatory power in phonology. In V. A. Fromkin (ed.), Phonetic Linguistics: Essays in Honor of Peter Ladefoged. New York: Acedemic Press, pp. 79-99.

Fishman, J. A. (1965). Who speaks what language to whom and when? Linguistique, 2: 67-88.

Fishman, J. A., R. L. Cooper, R. Ma et al. (1968/1971 and v.d.). Bilingualism in the Barrio. Indiana Univ. Publications: Language Science Monographs, vol. 7. Bloomington: Indiana Research Center for the Language Sciences, and The Hague: Mouton.

Flege, J. E. (1979). Phonetic interference in second language acquisition, Indiana U. Ph.D. Diss.

Flege, J. E. (1980). Phonetic approximation in second language acquisition, Lang. Learning, 30:

117-134.

Flege, J. E. (1981). The phonological basis of foreign accent: a hypothesis, TESOL Quarterly, 15, 4: 443-454.

Flege, J. E. (1984a). The detection of French accent by American listeners, JASA, 76: 692-707.

Flege, J. E. (1984b). Limits on pronunciation accuracy in adult foreign language speech production, JASA, 76: 708-721.

Flege, J. E. (To Appear). The production and perception of foreign language speech sounds. In H. Winitz (ed.) Human Communication and its Disorders, Vol. 1. Norwood, NJ: Ablex Publishing.

Flege, J. E. and R. D. Davidian (1984). Transfer and developmental processes in adult foreign language speech production, Applied Psycholing., 5: 323-347.

Flege, J. E. and R. Hammond. (1982). Mimicry of non-distinctive phonetic differences between language variation, Studies in Second Language Acquisition ,5: 1-17.

Flege, J. E. and J. Hillenbrand (1984). Limits on phonetic accuracy in foreign language speech production,

*JASA*, *76*, 3: 708-721.

Flege, J. E. and J. Hillenbrand (1985a). Perception of the English /s-z/ contrast by native speakers of English, French, Swedish and Finnish. Unpub. ms.

Flege, J. E. and R. Port. (1981). Cross-language phonetic interference: Arabic to English, *Lang. and Speech*, *24*: 125-145.

Fodor, J. A. and T. G. Bever (1965). The psychological reality of linguistic segments, *JVLVB*, *4*: 414-420.

Foss, D. J. and R. H. Lynch (1969). Decision processes during sentence comprehension: Effects of surface structure on decision times, *Percep. and Psychophys.*, *25*: 375-388.

Fowler, C. A. (1979). Perceptual centers in speech production and perception, *Perception and Psychophysics*, *25*: 375-388.

Fowler, Carol and Louis Tassinary. (1981). Natural measurement criteria for speech: The anisochrony illusion. In J. Long and A. Baddeley (eds.) *Attention and Performance*. Hillside, NJ: Lawrence Erlbaum Assocs.

Fry, D. B. (1955). Duration and intensity as physical cor-

relates of linguistic stress, JASA, 27: 765-768.

Fry, D. B. (1958). Experiments in the perception of stress, Lang. Speech, 1: 126-152.

Galvan, J. et al. (1977). A comparison of the accented speech of seven Spanish-English bilingual speakers, Workpapers in TESL, 1: 69-76. Los Angeles.

Garding, E. (1979). Sentence intonation in Swedish, Phonetica, 36: 207-215.

Garner, W. R. (1974). The Processing of Information and Structure. Potomac, MD: Erlbaum.

Garrett, M. F., T. G. Bever and J. A. Fodor (1966). The active use of grammar in speech perception, Percep. and Psychophys., 1: 30-32.

Gass, S. (1984). Development of speech perception and speech production abilities with adult second-language learners, Applied Psycholing., 5: 51-74.

Giles, H. (1970). Evaluative reactions to accents. Educ. Rev., 22: 211.

Giles, H. (1972). The effect of stimulus mildness-broadness in the evaluation of accents. Language and Speech, 15: 262-269.

Gili y Gaya, S. (1940). La cantidad silabica en la frase,

*Castilla (Valladolid)* *1*: 287-298.

Gimson, A. (1962). *An Introduction to the Pronunciation of English*. London: Edward Arnold.

Gold, B. and L. Rabiner (1969). Parallel processing techniques for estimating pitch periods of speech in the time domain, *JASA*, *46*: 442-448.

Goldin, S. E. (1975) Unpublished studies. Brown University.

Granda, G. de (1966). La estructura silábica, *Revista de Filolgía Española*, *anejo 81*.

Gynan, S. (1983). Attitudes of Spanish speakers toward taped native and nonnative speech samples. U. Texas Austin Ph.D. diss.

Halliday, M. A. K. and R. Hasan (1976). *Cohesion in English*. London: Longman Group Ltd.

Hammond, R. (1982). Second language phonetic interference: Oral reading versus speaking--A preliminary analysis, Paper presented at the Southeast Conference on Linguistics, Gainesville, FL, March, 1982.

Harris, J. W. (1983). *Syllable Structure and Stress in Spanish: A Nonlinear Analysis*. Cambridge, MA: The MIT Press.

Hoequist, Charles E. (1983a). Durational correlates of linguistic rhythm categories, *Phonetica* *40*:

19-31.

Hoequist, Charles E. (1983b). Syllable duration in stress-, syllable- and mora-timed languages, *Phonetica* *40*: 203-237.

Hoequist, Charles E. (1983c). The perceptual center and rhythm categories, *Language and Speech*, *26*, 4: 367-376.

Hooper, J. B. (1976). *An Introduction to Natural Generative Phonology*. New York: Academic Press.

House, A. S., C. E. Williams, M. H. L. Hecker and K. D. Kryter (1965). Articulation-testing methods: Consonantal differentiation with a closed-response set, *JASA*, *37*: 158-166.

Huggins, A. W. F. (1972). On the perception of temporal phenomena in speech, *JASA*, *51*: 1279-1290.

Huggins, A. W. F. (1978). Speech timing and intelligibility. In R. J. Raquin (ed.) *Attention and Performance VII*. Hillsdale, NJ: Erlbaum.

Hultzen, L. (1956). "The poet Burns" again, *American Speech*, *31*: 195-201.

Hutchinson, S. P. (1973). An objective index of the English-Spanish pronunciation dimension. Unpublished M. A. thesis, U. Texas.

Ioup, G. (1984-To Appear). Is there a structural foreign accent? A comparison of syntactic and phonological errors in second language acquisition, Language Learning, 34: 2.

Janota, P. (1970). Perception of stress by Czech listeners, Proceedings of the 6th International Congress of Phonetic Sciences. Prague: Academia.

Jassem, W., J. Morton and M. Steffen-Batog (1968). The perception of stress in synthetic speech-like stimuli by Polish listeners. In W. Jassem (ed.) Speech Analysis I. Warsaw: Panstwowe Wydawnictwo Naukowe.

Jenkins, J. and L. Franklin. (1981). Recall of passages of synthetic speech. Paper presented at the meeting of the Psychonomic Society, Philadelphia, Nov. 1981.

Jonasson, J. and R. McAllister (1972). Foreign accent and timing: An instrumental study, PILUS 14: 11-40. (Dept. of Linguistics).

Jones, D. (1960). An Outline of English Phonetics. (9th ed.) Cambridge: Heffer and Sons.

Kaplan, R. (1972). Augmented transition networks as psychological models of sentence comprehension, Artifi-

cial Intelligence, 3: 77-100.

Klatt, Dennis H. (1980). Software for a cascade/parallel formant synthesizer, JASA, 67, 3: 971-995.

Klatt, D. H. and W. E. Cooper (1975). Perception of segment duration in sentence contexts, in A. Cohen and S. G. Nooteboom (eds.) Structure and Process in Speech Perception. New York: Springer-Verlag.

Kozhevnikov, V. A. and L. A. Chistovich (1965). Speech: Articulation and Perception. Moscow-Leningrad.

Kvavik, K. (1976). Research and pedagogical materials on Spanish intonation: A re-examination, Hispania, 59, 3: 406-417.

Labov, W. (1966). The Social Stratification of English in New York City. Washington, D. C.: Center for Applied Linguistics.

Ladefoged, P. (1975). A Course in Phonetics. New York: Harcourt Brace Jovanovich, Inc.

Ladefoged, P. and D. E. Broadbent (1960). Perception of sequence in auditory events, Quarterly Journal of Exp. Psychol., 13: 162-170.

Lado, R. (1957). Linguistics Across Cultures: Applied Linguistics for Language Teachers. Ann Arbor: Univ. Michigan Press.

Lambert, W. E., M. Anisfeld and G. Yeni-Komshian (1965). Evaluational reactions of Jewish and Arab adolescents to dialect and language variation, J. Pers. and Soc. Psychol., 3: 84-90.

Lambert, W. E., R. C. Hodgson, R. D. Gardner, and S. Fillenbaum. (1960). Evaluational reactions to spoken languages, J. Abnorm. Soc. Psychol., 60: 44-51.

Lehiste, I. (1980). Interaction between test word duration and length of utterance. In L. Waugh and C. van Schooneveld (eds.) The Melody of Language: Intonation and Prosody. Baltimore: University Park Press, pp. 169-176.

Lehiste, I. (1973). Suprasegmentals. Cambridge, MA: MIT Press.

Lehiste, I. (1977). Isochrony reconsidered, J. Phonetics, 5: 253-263.

Lehiste, I., J. Olive, and L. Streeter (1976). Role of duration in disambiguating syntactically ambiguous sentences, JASA, 60, 5: 1199-1203.

Liberman, Mark and Alan Prince (1977). On stress and linguistic rhythm, Linguistic Inquiry, 8. 2: 249-336.

Lieberman, P. (1965). On the acoustic basis of the percep-

tion of intonation by linguists, _Word_, _21_: 40-55.

Lieberman, P. (1967). _Intonation, Perception and Language_. Cambridge, MA: MIT Press.

Lieberman, P., W. Katz, A. Jongman, R. Zimmerman, and M. Miller (1985). Measures of sentence intonation of read and spontaneous speech in American English, _J. Acoust. Soc. Am._, _77_, 2: 649-657.

Lindblom, B. (1982). The interdisciplinary challenge of speech motor control. In S. Grillner, B. Lindblom, J. Lubker and A. Persson (eds.) _Speech Motor Control_. London: Pergamon.

Lisker, L. and A. Abramson (1970). The voicing dimension: Some experiments in comparative phonetics, _Proceedings 6th International Congress of Phonetic Sciences_, Prague, 1967 (Academia Prague), pp. 563-567.

Luce, P., T. Feustel, and D. Pisoni (1981). Capacity demands in short term memory for synthetic and natural speech, Research on Speech Perception, Progress Report No. 7, Jan. 1981-Dec. 1981. Speech Research Laboratory, Dept. of Psychology, Indiana Univ., pp. 3-28.

Maeda, S. (1976). A characterization of American English

intonation. Unpublished Ph.D. dissertation, MIT.

Malmberg, B. (1948). La structure syllabique de l'espagnol, Boletin de Filologia, 9: 99-120.

Manrique, A. M. B. and A. Signorini (1983). Segmental duration and rhythm in Spanish, J. Phonetics, 11: 117-128.

Manrique, A. M. B. (1982). Final-word lengthening in Spanish, 104th Meeting of the Acoustical Society of America, Orlando, FL.

Manrique, A. M. B., A. Signorini and M. I. Massone (1982). Rasgos prosodicos: el acento, Fonoaudiologica, 28: 19-36.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior, Psychol. Review, 79: 487-509.

Martin, J. G. (1979). Rhythmic and segmental perception are not independent, JASA, 65, 5: 1286-1297.

Massaro, D. (1972). Preperceptual images, processing time, and perceptual units, Psychol. Review, 79: 124-145.

Massone, M. I. (1982). Relationship between stress and FO in Spanish, Paper presented at the 104th Meeting of the Acoustical Society of America (Orlando, FL).

Metcalf, A. (1972). Mexican-American English of southern California, Word, 9: 13-21.

Miller, J. (1977). Properties of feature detectors for VOT: The voiceless channel of analysis, JASA, 62: 641-648.

Miller, J. (1978). Interactions in processing segmental and suprasegmental features of speech, Percep. and Psychophys., 24: 175-180.

Mitleb, F. (1981). Timing of English vowels spoken with an Arabic accent, Research in Phonetics (Indiana U. Dept. of Ling.), 2: 193-226.

Morton, J. and W. Jassem (1965). Acoustic correlates of stress, Lang. and Speech, 8: 159-181.

Morton, J., S. Marcus and C. Frankish (1976). Perceptual centers (P-centers), Psychological Review, 83: 405-448.

Nakatani, L. H. and J. A. Schaffer (1978). Hearing "words" without words: speech prosody and word perception. Paper presented at 92nd Meeting: Acoustical Society of America.

Nakatani, L. H., K. D. O'Connor, and C. H. Aston (1981). Prosodic aspects of American English speech rhythm, Phonetica, 38: 84-106.

Navarro Tomas, T. (1916). Cantidad de vocales acentuadas, _Rev. Filología Española_, _3_: 387-408.

Navarro Tomas, T. (1917). Cantidad de vocales inacentuadas, _Rev. Filología Española_, _4_: 371-388.

Navarro Tomas, T. (1918). Diferencias de duración entre las consonantes españolas, _Rev. Filología Española_, _5_: 367-393.

Nickerson, R. S. (1977). Characteristics of the speech of deaf persons, _The Volta Review_, _77_: 342-362.

Niemi, J. (1979). Noun and noun phrase stress: A phonetic study of English supplemented with an error analysis using Finnish speaker-hearers, I. Production, _Lang. and Speech_, _22_: 221-235.

Niemi, J. (1984). Word-level stress and prominence in Finnish and English: Acoustic experiments on production and perception. Joensuu, Finland: University of Joensuu Publications in the Humanities, No. 1.

Nooteboom, S. (1973). Perceptual reality of prosodic durations, _J. Phonet._, _1_: 25-45.

Nooteboom, S. G. (1975). Context effects in the perception of phonemic vowel length, Paper presented at the VIIIth International Congress of Phonetic Sciences, Leeds, August 17-23.

Nooteboom, S. G., J. P. L. Brokx and J. J. deRooij (1978). Contributions of prosody to speech perception, in W. J. M. Levelt and G. B. Flores d'Arcais (eds.) Studies in the Perception of Language. New York: John Wiley, pp. 75-107.

Nooteboom, S. G. and A. Cohen (1975). Anticipation in speech production and its implications for perception. In A. Cohen and S. G. Nooteboom (eds.) Structure and Process in Speech Perception. New York: Springer-Verlag, pp. 124-145.

Nuñez-Cedeño, R. (1984). Review of Syllable structure and stress in Spanish: A nonlinear analysis, Language, 60, 4: 933-937.

O'Connor, J. D. (1965). The perception of time intervals, Progress Report (Phonetics Laboratory, Univ. College London), 2: 11-15.

Oden, G., and D. Massaro (1978). Integration of featural information in speech perception, Psychol. Rev., 85: 172-191.

Ohala, J. and M. Hirano (1967). An experimental investigation of pitch changes in speech, JASA, 42: 1208-1209 (A).

Olive, J. P. and L. H. Nakatani (1974). Rule synthesis of

speech by word concatenation: A first step, *JASA*, *55*: 660-666.

Olsen, C. I. (1972). Rhythmical patterns and syllabic features of the Spanish sense-group, Proceedings of the Seventh International Congress of Phonetic Sciences. The Hague: Mouton, pp. 990-995.

Osgood, C. E. (1964). Semantic differential techniques in the comparative study of cultures, *American Anthropologist*, *66*, 3: 171-200, Part 2.

Oyama, S. (1976). A sensitive period for the acquisition of a non-native phonological system, *J. Psycholing. Res.*, *5*, 3: 261-283.

Palmer, J. D. (1976). Linguistic accuracy and intelligibility, Proceedings of the 4th International Congress of Applied Linguistics, 505-513. Stuttgart: Hockschul Verlag.

Peterson, G. E. and I. Lehiste (1960). Duration of syllabic nuclei in English, *JASA*, *32*: 693-703.

Pierrehumbert, J. (1979). The perception of fundamental frequency declination, *J. Acoust. Soc. Am.*, *66*: 363-369.

Pierrehumbert, J. (1981). Synthesizing intonation, *JASA*, *70*, 4: 985-995.

Pike, K. L. (1946). The Intonation of American English. Ann Arbor: Univ. Michigan Press.

Pinkerton, S. (1972). The learning of English suprasegmental for stress and final syllables by Spanish speakers. Paper presented at the Mid-America Linguistics Conference, Oct. 1973.

Pisoni, D. (1981). Speeded classification of natural and synthetic speech in a lexical decision task, JASA 70: S98.

Pisnoi, D. (1982). Perception of speech: The human listener, Speech Technology, 1, 2: 10-23.

Pisoni, D. and S. Hunnicutt. (1980). Perceptual evaluation of MITalk: The MIT unrestricted text-to-speech system. 1980 IEEE International Conference Record on Acoustics, Speech and Signal Processing, April 1980.

Pisoni, D. and E. Koen (1981). Some comparisons of intelligibility of synthetic and natural speech at different speech-to-noise ratios, Research on Speech Perception, Progress Report No. 7, Indiana University: 243-53.

Pisoni, D. and J. Sawusch (1975). Some stages of processing in speech perception. In A. Cohen and S. Nooteboom

(eds.) Structure and Process in Speech Perception. New York: Springer-Verlag, pp. 16-35.

Pointon, G. E. (1980). Is Spanish really syllable-timed? J. Phonetics, 8: 293-304.

Port, R. and F. Mitleb (1980). Phonetic and phonological manifestations of the voicing contrast in Arabic-accented English, Research in Phonetics, Indiana Univ. Dept. of Linguistics 1: 137-166.

Rabbitt, P. (1968). Channel-capacity, intelligibility and immediate memory, Quarterly Journal of Exp. Psychol., 20: 241-248.

Repp, B. (1976). Identification of dichotic fusions, JASA, 60: 456-469.

Repp, B., A. Healy and R. Crowder (1979). Categories and context in the perception of isolated steady state vowels, J. of Experimental Psychol.: Human Perception and Performance, 5: 129-154.

Rigault, A. (1962). Role de la frequence de l'intensite et de la duree vocaliques dans la perception de l'accent en francais. In Proceedings of the 4th International Congress of Phonetic Sciences (Helsinki), pp. 735-748.

Rosch, E. (1973). Natural categories, Cognit. Psychol.,

4: 328-350.

Rosch, E. (1978). Principles of categorization. In E. Rosch and B. Lloyd (eds.) Cognition and Categorization. Hillsdale, NJ: Erlbaum, pp. 28-50.

Roudet, L. (1910). Elements de phonetique générale. Paris: H. Welter.

Ryan, E. B. and M. A. Carranza (1975). Evaluative reactions of adolescents toward speakers of standard English and Mexican American accented English, J. Personality and Soc. Psychol., 31: 855-863.

Ryan, E. B., M. A. Carranza, and R. W. Moffie (1977). Reactions toward varying degrees of accentedness in speech of Spanish-English bilinguals. Language and Speech, 20: 267-273.

Samuel, A. (1982). Phonetic prototypes, Percept. Psychology, 31: 307-314.

Saporta, S. and H. Contreras (1962). A Phonological Grammar of Spanish. Seattle, Washington: University of Washington Press.

Sawusch, J. and D. Pisoni (1974). On the identification of place and voicing features in synthetic stop consonants, J. Phonetics, 2: 181-194.

Sawyer, J. B. (1975). Spanish-English bilingualism in San

Antonio, Texas. In E. Hernandez-Chavez, A. Cohen and L. Beltramo (eds.) El lenguaje de los chicanos. Washington, D. C.: Center for Applied Linguistics, pp. 77-98.

Scovel, T. (1969). Foreign accent, language acquisition and cerebral dominance, Lang. Learning, 19: 245-254.

Scovel, T. (1981). The recognition of foreign accents in English and its implications for psycholinguistic theories of language acquisition, in Proceedings 5th International Congress of the International Association of Applied Linguistics, ed. by J-G. Savard and L. Laforge. Quebec: Laval U. P.

Shen, Y. and G. G. Peterson. (1962). Isochronism in English, Univ. of Buffalo Studies in Linguistics, Occasional Papers, 9: 1-36.

Shields, J., A. McHugh and J. G. Martin (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech, J. Exp. Psychol., 102: 250-255.

Steele, J. (1775/1969). An essay towards establishing the melody and measure of speech. Menston: The Scolar Press.

Stetson, R. H. (1951). Motor Phonetics. Amsterdam:

North-Holland.

Stockwell, R., J. D. Bowen and I. Silva-Fuenzalida (1956/1963). Spanish juncture and intonation, _Language_, _32_: 641-645. rpt. in M. Joos (ed.) _Readings in Linguistics_ (3rd ed.) N.Y.: American Council of Learned Societies.

Summerfield, Q. (1975). How a full account of segmental perception depends on prosody and vice versa. In A. Cohen and S. G. Nooteboom (eds.) _Structure and Process in Speech Perception_. NY: Springer-Verlag, pp. 31-68.

Suomi, K. (1976). English voiced and voiceless stops as produced by Finnish and native speakers. Jyvaskyla Contrastive Studies, 2, Reports from the Dept. of English, Univ. of Jyvaskyla, Finland.

Suomi, K. (1980). Voicing in English and Finnish stops. Publications of the Depart. Finnish and General Linguistics 10. Finland: Turku Univ.

Svensson, S.-G. (1974). Prosody and Grammar in Speech Perception. Ph.D. diss. MILUS 2 (Monographs from the Institute of Linguistics, Univ. of Stockholm.

Sweet, H. (1892). _New English Grammar_. OIxford: Clarendon Press.

Terry, C. and R. Cooper (1969). A note on the perception and production of phonological variation, MLJ, 53: 254-255.

Trager, G. L. and H. L. Smith (1951). Outline of English Structure, Studies in Linguistics, No. 3. Norman, Oklahoma: Baltenburg.

Triandis, H., W. Loh and L. Levin (1966). Race, status, quality of spoken English and opinions about civil rights as determinants of interpersonal attitudes, J. Pers. and Soc. Psychol., 3: 468-472.

Trubetzkoy, N. (1939/1969). Principles of Phonology. Trans. C. Baltaxe. Berkeley: U. Calif. Press.

Tseng, C. Y. (1981). An acoustic study of tones in Mandarin. Unpublished Ph.D. dissertation, Brown University.

Tucker, G. R. and W. E. Lambert (1969). White and Negro listeners' reactions to various American English dialects, Social Forces, 47: 463-468.

Tuller, B. and C. A. Fowler (1980). Some articulatory correlates of perceptual isochrony, Perception and Psychophysics, 24: 277-283.

Uldall, E. T. (1971). Isochronous stresses in R. P. In L. L. Hammerich et. al (eds.), Form and Substance.

Copenhagen: Akademisk Forlag, pp. 205-210.

Vaissiere, J. (1974). On French prosody, Res. Lab. Electron. Q. Prog. Rep. No. 114, MIT: 212-213.

Vanderslice, R. and P. Ladefoged (1972). Binary suprasegmental features and transformational word-accentuation rules, Language, 48: 819-838.

Van Katwijk, A. (1974). Accentuation in Dutch: An Experimental Linguistic Study. Amsterdam and Assen: Van Gorcum.

Verbrugge, R. R., W. Strange, D. P. Shankweiler, T. R. Edman (1976). What information enables a listener to map a talker's vowel space? JASA, 60: 198-212.

Webster, W. G. and E. Kramer (1968). Attitudes and evaluational reactions to accented English speech, J. Soc. Psychol., 75: 231-240.

Westbury, J. and P. Keating (1980). Central representation of vowel duration, JASA, 67, Suppl. 1, S37 (A).

Williams, L. (1977a). The voicing contrast in Spanish, J. Phonetics, 5: 169-184.

Williams, L. (1977b). The perception of stop consonant voicing by Spanish-English bilinguals, Percep. and Psychophys. 21: 289-297.

Williams, L. (1979). The modification of speech perception

and production in second-language learning, Percep. and Psychophys., 26: 95-104.

Williams, L. (1980). Phonetic variation as a function of second-language learning. In G. Yeni-Komshian, J. Kavanaugh, and C. Ferguson (eds.), Child Phonology, Vol. 2: Perception. New York: Academic Press.

Wingfield, A. (1975). The intonation-syntax interaction: Prosodic features in perceptual processing of sentences. In A. Cohen and S. Nooteboom (eds.) Structure and Process in Speech Perception. New York: Springer-Verlag, pp. 146-156.

Wingfield, A. and J. F. Klein (1975). Syntactic structure and acoustic pattern in speech perception, Percep. and Psychophys., 9: 23-25.

Wood, C. C. (1974). Parallel processing of auditory and phonetic information in speech discrimination, Percep. and Psychophys., 15: 501-508.

Wood, C. C. (1975). Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses, J. Exp. Psychol.: Human Perception and Performance, 1: 3-20.

Woods, W. A. (1973). An experimental parsing system for transition network grammars. In R. Rustin (ed.) *Natural language processing*. NY: Algorithmics Press.

Yanz, J. L. (1984). The application of the Theory of Signal Detection to the assessment of speech perception, *Ear and Hearing*, *5*: 64-71.

Yule, G., J. L. Yanz and A. Tsuda (1985). Investigating Aspects of the language learner's confidence: An application of the Theory of Signal Detection, *Language Learning*, *32*. (Forthcoming)

APPENDIX A

EVALUATION OF FOREIGN ACCENT USING SYNTHETIC SPEECH

Name------------------------------ Course------- Date------

DIRECTIONS: Circle the appropriate response.

1. Are you of Hispanic descent? YES NO

2. Are you a native speaker of English? YES NO

3. Do you know any foreign languages? YES NO

4. Do you speak Spanish or have you ever tried to learn to speak it? YES NO

5. (a) Can you remember having any extensive contact with Hispanics? YES NO
(b) Can you remember anyone close to you having any extensive contact with Hispanics? YES NO

6. If you answered YES to Question #5, how would you characterize the experience with Hispanics?
(a) Your experience: POSITIVE NEUTRAL NEGATIVE MIXED
(b) Other's experience: POSITIVE NEUTRAL NEGATIVE MIXED

7. How would you characterize your attitude toward Hispanics? POSITIVE(like) NEGATIVE(dislike) MIXED(like some, dislike others

8. How would you characterize a "Spanish accent?"
SIMPLY FOREIGN NICE(pleasant) UGLY(unpleasant) NOT SURE

9. Have you ever been diagnosed as having a speech or hearing problem? YES NO

APPENDIX B

EVALUATION OF FOREIGN ACCENT USING SYNTHETIC SPEECH

Name------------------------------Course---------Date--------

DIRECTIONS: For each item, you will hear two synthetic speech productions of the sentence, "KANSAS STATE'S NEW BASKETBALL TICKETS COST TEN DOLLARS NOW." The first production is Standard English; the second production is English that may be "accented" in some way. For each pair of sentences, put an X in one of the slots along the continuum to indicate how different the second sentence is from the first sentence.

1. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

2. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

3. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

4. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

5. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

6. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

7. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

8. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

9. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

10. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

11. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

12. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

13. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

14. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

15. -----/-----/-----/-----/-----/-----/-----/-----/-----/-----

16.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

17.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

18.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

19.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

20.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

21.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

22.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

23.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

24.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

25.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

26.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

27.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

28.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

29.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

30.-----/-----/-----/-----/-----/-----/-----/-----/-----/-----

Name----------------------Date-------

DIRECTIONS: For each item, you will hear two synthetic speech productions of the sentence, "KANSAS STATE'S NEW BASKETBALL TICKETS COST TEN DOLLARS NOW." The first production is considered Standard English pronunciation; the second production may be "accented" in some way. For each pair of sentences, put an X in one of the slots along the first continuum to indicate how different the second sentence is from the first sentence. In addition, indicate how sure you are of your judgment along the second continuum provided.

Example:

| | SAME 1 2 3 4 5 6 7 8 9 10 DIFFERENT | | NOT SURE 1 2 3 4 5 SURE |
|---|---|---|---|
| 0. | ---/---/---/---/---/---/---/---/---/--- | 0. | ---/---/---/---/--- |

| | 1 2 3 4 5 6 7 8 9 10 | | 1 2 3 4 5 |
|---|---|---|---|
| 1. | ---/---/---/---/---/---/---/---/---/--- | 1. | ---/---/---/---/--- |
| 2. | ---/---/---/---/---/---/---/---/---/--- | 2. | ---/---/---/---/--- |
| 3. | ---/---/---/---/---/---/---/---/---/--- | 3. | ---/---/---/---/--- |
| 4. | ---/---/---/---/---/---/---/---/---/--- | 4. | ---/---/---/---/--- |
| 5. | ---/---/---/---/---/---/---/---/---/--- | 5. | ---/---/---/---/--- |
| 6. | ---/---/---/---/---/---/---/---/---/--- | 6. | ---/---/---/---/--- |
| 7. | ---/---/---/---/---/---/---/---/---/--- | 7. | ---/---/---/---/--- |

SAME                                   DIFFERENT          NOT SURE        SURE

1   2   3   4   5   6   7   8   9   10            1   2   3   4   5

8. ---/---/---/---/---/---/---/---/---/---    8. ---/---/---/---/---

9. ---/---/---/---/---/---/---/---/---/---    9. ---/---/---/---/---

10. ---/---/---/---/---/---/---/---/---/---    10. ---/---/---/---/---

11. ---/---/---/---/---/---/---/---/---/---    11. ---/---/---/---/---

12. ---/---/---/---/---/---/---/---/---/---    12. ---/---/---/---/---

13. ---/---/---/---/---/---/---/---/---/---    13. ---/---/---/---/---

14. ---/---/---/---/---/---/---/---/---/---    14. ---/---/---/---/---

15. ---/---/---/---/---/---/---/---/---/---    15. ---/---/---/---/---

16. ---/---/---/---/---/---/---/---/---/---    16. ---/---/---/---/---

17. ---/---/---/---/---/---/---/---/---/---    17. ---/---/---/---/---

18. ---/---/---/---/---/---/---/---/---/---    18. ---/---/---/---/---

19. ---/---/---/---/---/---/---/---/---/---    19. ---/---/---/---/---

20. ---/---/---/---/---/---/---/---/---/---    20. ---/---/---/---/---

21. ---/---/---/---/---/---/---/---/---/---    21. ---/---/---/---/---

22. ---/---/---/---/---/---/---/---/---/---    22. ---/---/---/---/---

23. ---/---/---/---/---/---/---/---/---/---    23. ---/---/---/---/---

24. ---/---/---/---/---/---/---/---/---/---    24. ---/---/---/---/---

25. ---/---/---/---/---/---/---/---/---/---    25. ---/---/---/---/---

26. ---/---/---/---/---/---/---/---/---/---    26. ---/---/---/---/---

27. ---/---/---/---/---/---/---/---/---/---    27. ---/---/---/---/---

28. ---/---/---/---/---/---/---/---/---/---    28. ---/---/---/---/---

29. ---/---/---/---/---/---/---/---/---/---    29. ---/---/---/---/---

EVALUATION OF FOREIGN ACCENT USING SYNTHETIC SPEECH

Name----------------------Date-------

DIRECTIONS: For each item, you will hear two synthetic speech productions of the sentence, "KANSAS STATE'S NEW BASKETBALL TICKETS COST TEN DOLLARS NOW." The first production is considered Standard English pronunciation; the second production may be "accented" in some way. For each pair of sentences, put an X in one of the slots along the first continuum to indicate how different the second sentence is from the first sentence. In addition, indicate how sure you are of your judgment along the second continuum provided.

Example:

SAME DIFFERENT NOT SURE SURE

1 2 3 4 5 6 7 8 9 10 1 2 3 4 5

0. ---/---/---/---/---/---/---/---/---/--- 0. ---/---/---/---/---

1 2 3 4 5 6 7 8 9 10 1 2 3 4 5

1. ---/---/---/---/---/---/---/---/---/--- 1. ---/---/---/---/--- (11)

2. ---/---/---/---/---/---/---/---/---/--- 2. ---/---/---/---/--- (22)

3. ---/---/---/---/---/---/---/---/---/--- 3. ---/---/---/---/--- (22)

4. ---/---/---/---/---/---/---/---/---/--- 4. ---/---/---/---/--- (15)

5. ---/---/---/---/---/---/---/---/---/--- 5. ---/---/---/---/--- (21)

6. ---/---/---/---/---/---/---/---/---/--- 6. ---/---/---/---/--- (24)

7. ---/---/---/---/---/---/---/---/---/--- 7. ---/---/---/---/--- (15)

MODTAPE1

| | SAME | | | | | | | | DIFFERENT | | | NOT SURE | | | | SURE | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | 1 | 2 | 3 | 4 | 5 | |
| 8. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 8. | --- | --- | --- | --- | --- | (28) |
| 9. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 9. | --- | --- | --- | --- | --- | (8) |
| 10. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 10. | --- | --- | --- | --- | --- | (29) |
| 11. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 11. | --- | --- | --- | --- | --- | (19) |
| 12. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 12. | --- | --- | --- | --- | --- | (25) |
| 13. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 13. | --- | --- | --- | --- | --- | (16) |
| 14. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 14. | --- | --- | --- | --- | --- | (17) |
| 15. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 15. | --- | --- | --- | --- | --- | (31) |
| 16. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 16. | --- | --- | --- | --- | --- | (4) |
| 17. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 17. | --- | --- | --- | --- | --- | (9) |
| 18. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 18. | --- | --- | --- | --- | --- | (22) |
| 19. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 19. | --- | --- | --- | --- | --- | (14) |
| 20. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 20. | --- | --- | --- | --- | --- | (29) |
| 21. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 21. | --- | --- | --- | --- | --- | (28) |
| 22. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 22. | --- | --- | --- | --- | --- | (21) |
| 23. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 23. | --- | --- | --- | --- | --- | (28) |
| 24. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 24. | --- | --- | --- | --- | --- | (14) |
| 25. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 25. | --- | --- | --- | --- | --- | (11) |
| 26. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 26. | --- | --- | --- | --- | --- | (9) |
| 27. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 27. | --- | --- | --- | --- | --- | (5) |
| 28. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 28. | --- | --- | --- | --- | --- | (9) |
| 29. | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | 29. | --- | --- | --- | --- | --- | (3) |

| | SAME | | | | | | | | DIFFERENT | | NOT SURE | | | | SURE | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 1 | 2 | 3 | 4 | 5 | |
| 30. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 30. ---/---/---/---/--- | | | | | (7) |
| 31. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 31. ---/---/---/---/--- | | | | | (1) |
| 32. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 32. ---/---/---/---/--- | | | | | (20) |
| 33. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 33. ---/---/---/---/--- | | | | | (29) |
| 34. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 34. ---/---/---/---/--- | | | | | (19) |
| 35. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 35. ---/---/---/---/--- | | | | | (16) |
| 36. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 36. ---/---/---/---/--- | | | | | (3) |
| 37. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 37. ---/---/---/---/--- | | | | | (6) |
| 38. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 38. ---/---/---/---/--- | | | | | (19) |
| 39. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 39. ---/---/---/---/--- | | | | | (28) |
| 40. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 40. ---/---/---/---/--- | | | | | (20) |
| 41. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 41. ---/---/---/---/--- | | | | | (25) |
| 42. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 42. ---/---/---/---/--- | | | | | (14) |
| 43. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 43. ---/---/---/---/--- | | | | | (15) |
| 44. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 44. ---/---/---/---/--- | | | | | (23) |
| 45. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 45. ---/---/---/---/--- | | | | | (10) |
| 46. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 46. ---/---/---/---/--- | | | | | (27) |
| 47. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 47. ---/---/---/---/--- | | | | | (18) |
| 48. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 48. ---/---/---/---/--- | | | | | (26) |
| 49. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 49. ---/---/---/---/--- | | | | | (2) |
| 50. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 50. ---/---/---/---/--- | | | | | (27) |
| 51. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 51. ---/---/---/---/--- | | | | | (22) |

| | SAME 1 2 3 4 5 6 7 8 9 10 DIFFERENT | | NOT SURE 1 2 3 4 5 SURE | |
|---|---|---|---|---|
| 52. | ---/---/---/---/---/---/---/---/---/--- | 52. | ---/---/---/---/--- | (18) |
| 53. | ---/---/---/---/---/---/---/---/---/--- | 53. | ---/---/---/---/--- | (1) |
| 54. | ---/---/---/---/---/---/---/---/---/--- | 54. | ---/---/---/---/--- | (10) |
| 55. | ---/---/---/---/---/---/---/---/---/--- | 55. | ---/---/---/---/--- | (12) |
| 56. | ---/---/---/---/---/---/---/---/---/--- | 56. | ---/---/---/---/--- | (23) |
| 57. | ---/---/---/---/---/---/---/---/---/--- | 57. | ---/---/---/---/--- | (26) |
| 58. | ---/---/---/---/---/---/---/---/---/--- | 58. | ---/---/---/---/--- | (30) |
| 59. | ---/---/---/---/---/---/---/---/---/--- | 59. | ---/---/---/---/--- | (3) |
| 60. | ---/---/---/---/---/---/---/---/---/--- | 60. | ---/---/---/---/--- | (4) |
| 61. | ---/---/---/---/---/---/---/---/---/--- | 61. | ---/---/---/---/--- | (7) |
| 62. | ---/---/---/---/---/---/---/---/---/--- | 62. | ---/---/---/---/--- | (8) |
| 63. | ---/---/---/---/---/---/---/---/---/--- | 63. | ---/---/---/---/--- | (5) |
| 64. | ---/---/---/---/---/---/---/---/---/--- | 64. | ---/---/---/---/--- | (24) |
| 65. | ---/---/---/---/---/---/---/---/---/--- | 65. | ---/---/---/---/--- | (29) |
| 66. | ---/---/---/---/---/---/---/---/---/--- | 66. | ---/---/---/---/--- | (9) |
| 67. | ---/---/---/---/---/---/---/---/---/--- | 67. | ---/---/---/---/--- | (5) |
| 68. | ---/---/---/---/---/---/---/---/---/--- | 68. | ---/---/---/---/--- | (31) |
| 69. | ---/---/---/---/---/---/---/---/---/--- | 69. | ---/---/---/---/--- | (12) |
| 70. | ---/---/---/---/---/---/---/---/---/--- | 70. | ---/---/---/---/--- | (13) |
| 71. | ---/---/---/---/---/---/---/---/---/--- | 71. | ---/---/---/---/--- | (13) |
| 72. | ---/---/---/---/---/---/---/---/---/--- | 72. | ---/---/---/---/--- | (31) |
| 73. | ---/---/---/---/---/---/---/---/---/--- | 73. | ---/---/---/---/--- | (4) |

| | SAME 1 2 3 4 5 6 7 8 9 10 DIFFERENT | | NOT SURE 1 2 3 4 5 SURE | |
|---|---|---|---|---|
| 74. | ---/---/---/---/---/---/---/---/---/--- | 74. | ---/---/---/---/--- | (31) |
| 75. | ---/---/---/---/---/---/---/---/---/--- | 75. | ---/---/---/---/--- | (23) |
| 76. | ---/---/---/---/---/---/---/---/---/--- | 76. | ---/---/---/---/--- | (7) |
| 77. | ---/---/---/---/---/---/---/---/---/--- | 77. | ---/---/---/---/--- | (12) |
| 78. | ---/---/---/---/---/---/---/---/---/--- | 78. | ---/---/---/---/--- | (14) |
| 79. | ---/---/---/---/---/---/---/---/---/--- | 79. | ---/---/---/---/--- | (16) |
| 80. | ---/---/---/---/---/---/---/---/---/--- | 80. | ---/---/---/---/--- | (16) |
| 81. | ---/---/---/---/---/---/---/---/---/--- | 81. | ---/---/---/---/--- | (24) |
| 82. | ---/---/---/---/---/---/---/---/---/--- | 82. | ---/---/---/---/--- | (2) |
| 83. | ---/---/---/---/---/---/---/---/---/--- | 83. | ---/---/---/---/--- | (17) |
| 84. | ---/---/---/---/---/---/---/---/---/--- | 84. | ---/---/---/---/--- | (8) |
| 85. | ---/---/---/---/---/---/---/---/---/--- | 85. | ---/---/---/---/--- | (4) |
| 86. | ---/---/---/---/---/---/---/---/---/--- | 86. | ---/---/---/---/--- | (8) |
| 87. | ---/---/---/---/---/---/---/---/---/--- | 87. | ---/---/---/---/--- | (27) |
| 88. | ---/---/---/---/---/---/---/---/---/--- | 88. | ---/---/---/---/--- | (5) |
| 89. | ---/---/---/---/---/---/---/---/---/--- | 89. | ---/---/---/---/--- | (3) |
| 90. | ---/---/---/---/---/---/---/---/---/--- | 90. | ---/---/---/---/--- | (1) |
| 91. | ---/---/---/---/---/---/---/---/---/--- | 91. | ---/---/---/---/--- | (7) |
| 92. | ---/---/---/---/---/---/---/---/---/--- | 92. | ---/---/---/---/--- | (6) |
| 93. | ---/---/---/---/---/---/---/---/---/--- | 93. | ---/---/---/---/--- | (21) |
| 94. | ---/---/---/---/---/---/---/---/---/--- | 94. | ---/---/---/---/--- | (24) |
| 95. | ---/---/---/---/---/---/---/---/---/--- | 95. | ---/---/---/---/--- | (26) |

| | SAME 1 2 3 4 5 6 7 8 9 10 DIFFERENT | | NOT SURE 1 2 3 4 5 SURE | |
|---|---|---|---|---|
| 96. | ---/---/---/---/---/---/---/---/---/--- | 96. | ---/---/---/---/--- | (18) |
| 97. | ---/---/---/---/---/---/---/---/---/--- | 97. | ---/---/---/---/--- | (6) |
| 98. | ---/---/---/---/---/---/---/---/---/--- | 98. | ---/---/---/---/--- | (30) |
| 99. | ---/---/---/---/---/---/---/---/---/--- | 99. | ---/---/---/---/--- | (23) |
| 100. | ---/---/---/---/---/---/---/---/---/--- | 100. | ---/---/---/---/--- | (13) |
| 101. | ---/---/---/---/---/---/---/---/---/--- | 101. | ---/---/---/---/--- | (26) |
| 102. | ---/---/---/---/---/---/---/---/---/--- | 102. | ---/---/---/---/--- | (11) |
| 103. | ---/---/---/---/---/---/---/---/---/--- | 103. | ---/---/---/---/--- | (1) |
| 104. | ---/---/---/---/---/---/---/---/---/--- | 104. | ---/---/---/---/--- | (10) |
| 105. | ---/---/---/---/---/---/---/---/---/--- | 105. | ---/---/---/---/--- | (20) |
| 106. | ---/---/---/---/---/---/---/---/---/--- | 106. | ---/---/---/---/--- | (17) |
| 107. | ---/---/---/---/---/---/---/---/---/--- | 107. | ---/---/---/---/--- | (21) |
| 108. | ---/---/---/---/---/---/---/---/---/--- | 108. | ---/---/---/---/--- | (18) |
| 109. | ---/---/---/---/---/---/---/---/---/--- | 109. | ---/---/---/---/--- | (13) |
| 110. | ---/---/---/---/---/---/---/---/---/--- | 110. | ---/---/---/---/--- | (19) |
| 111. | ---/---/---/---/---/---/---/---/---/--- | 111. | ---/---/---/---/--- | (25) |
| 112. | ---/---/---/---/---/---/---/---/---/--- | 112. | ---/---/---/---/--- | (30) |
| 113. | ---/---/---/---/---/---/---/---/---/--- | 113. | ---/---/---/---/--- | (2) |
| 114. | ---/---/---/---/---/---/---/---/---/--- | 114. | ---/---/---/---/--- | (10) |
| 115. | ---/---/---/---/---/---/---/---/---/--- | 115. | ---/---/---/---/--- | (20) |
| 116. | ---/---/---/---/---/---/---/---/---/--- | 116. | ---/---/---/---/--- | (30) |
| 117. | ---/---/---/---/---/---/---/---/---/--- | 117. | ---/---/---/---/--- | (15) |

| | SAME | | | | | | | | | DIFFERENT | | NOT SURE | | | | SURE | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | | 1 | 2 | 3 | 4 | 5 | |
| 118. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 118. | ---/---/---/---/--- | | | | | (6) |
| 119. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 119. | ---/---/---/---/--- | | | | | (2) |
| 120. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 120. | ---/---/---/---/--- | | | | | (25) |
| 121. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 121. | ---/---/---/---/--- | | | | | (11) |
| 122. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 122. | ---/---/---/---/--- | | | | | (17) |
| 123. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 123. | ---/---/---/---/--- | | | | | (12) |
| 124. | ---/---/---/---/---/---/---/---/---/--- | | | | | | | | | | 124. | ---/---/---/---/--- | | | | | (27) |

INFORMANT ______________________ # ______

SESSION ______ MOD ______ STG ______

| STIMULUS | RATING 0 | RATING 1 | RATING 2 | RATING 3 | CONFIDENCE 0 | CONFIDENCE 1 | CONFIDENCE 2 | CONFIDENCE 3 |
|---|---|---|---|---|---|---|---|---|
| 1 STD | | | | | | | | |
| 2 Fφ | | | | | | | | |
| 3 VOT | | | | | | | | |
| 4 VQS | | | | | | | | |
| 5 VQU | | | | | | | | |
| 6 DUR | | | | | | | | |
| 7 FφVOT | | | | | | | | |
| 8 FφVQS | | | | | | | | |
| 9 FφVQU | | | | | | | | |
| 10 FφDUR | | | | | | | | |
| 11 VOTVQS | | | | | | | | |
| 12 VOTVQU | | | | | | | | |
| 13 VOTDUR | | | | | | | | |
| 14 VQSVQU | | | | | | | | |
| 15 VQSDUR | | | | | | | | |
| 16 VQUDUR | | | | | | | | |
| 17 FφVOTVQS | | | | | | | | |
| 18 FφVOTVQU | | | | | | | | |
| 19 FφVOTDUR | | | | | | | | |
| 20 FφVQSVQU | | | | | | | | |
| 21 FφVQSDUR | | | | | | | | |
| 22 VOTVQSVQU | | | | | | | | |
| 23 VOTVQSDUR | | | | | | | | |
| 24 VOTVQUDUR | | | | | | | | |
| 25 VQSVQUDUR | | | | | | | | |
| 26 FφVQUDUR | | | | | | | | |
| 27 FφVOTVQSVQU | | | | | | | | |
| 28 FφVOTVQSDUR | | | | | | | | |
| 29 FφVQSVQUDUR | | | | | | | | |
| 30 VOTVQSVQUDUR | | | | | | | | |
| 31 all 5 cues | | | | | | | | |

APPENDIX C

INDIVIDUAL SUBJECT SCORES: MODERATE ACCENT

| | Stimulus | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ss | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 01 | | | | | | | 5.00 | |
| 02 | | | 7.00* | | | | | |
| 03 | 7.25* | | 5.50 | | | 7.50* | 7.75* | 7.75* |
| 04 | 6.25 | | 5.75 | | | 6.25 | 6.75 | 7.00* |
| 05 | 5.00 | 5.25 | 6.50 | 6.50 | | 7.00* | 6.25 | 6.50 |
| 06 | 7.25* | 5.25 | | | | 6.75 | 7.50* | 8.25* |
| 07 | 5.50 | 5.25 | 7.00* | | | 5.50 | 6.50 | 6.00 |
| 08 | | | 5.75 | | | | 5.50 | 5.75 |
| 09 | | | 8.50* | 5.00 | | | 6.50 | 5.75 |
| 10 | | | | | | | | |
| 11 | | | | | | | | |
| 12 | | | 6.00 | | | | | |
| 13 | | | 5.75 | | | | | |
| 14 | | | 5.00 | | | 5.50 | 5.75 | 5.25 |
| 15 | 7.75* | | | | | 7.50* | 6.75 | 7.25* |
| 16 | | | 5.00 | | | 5.50 | | |
| 17 | | | 5.25 | | | | | |
| 18 | | | | | | | | |
| 19 | | | 5.00 | | | | | 6.00 |
| 20 | | | | | | | | |
| 21 | 6.50 | | 8.50* | | | 6.25 | 7.25* | |
| 22 | | | 8.00* | | | | 7.50* | |
| 23 | | | | | | | | |
| 24 | | | | | | | | |
| 25 | | | 6.50 | | | | | 5.50 |
| 26 | | | 7.75* | | | | | |
| 27 | | | | | | | | |
| 28 | | | | | | 5.25 | | |
| 29 | 7.00* | | 6.00 | | | 7.75* | 8.25* | 8.00* |
| 30 | | | 7.00* | | | | | |
| 31 | 5.75 | | | | | 6.25 | 8.00* | 7.75* |
| 32 | | | 7.50* | | | | | |
| 33 | | | | | | | | |
| 34 | | | | | | 7.00* | 6.00 | 6.75 |
| 35 | 5.25 | | | | | 6.75 | 6.25 | 6.25 |
| 36 | | | 8.25* | 8.25* | 5.00 | | 7.50* | 8.25* |
| 37 | | | 6.75 | | | | | |
| 38 | | | | | | | | 5.75 |
| 39 | | | 9.00* | | | 5.50 | | |
| 40 | | | 7.00* | | | | | |
| 41 | | | 5.75 | | | | | |
| 42 | | 5.50 | 8.00* | | | 5.00 | 5.00 | |

| Ss | Stimulus 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
|---|---|---|---|---|---|---|---|---|---|
| 01 | | | | | | | | 5.50 | 6.50 |
| 02 | | 7.50* | | | 7.00* | 7.25* | | 7.00* | |
| 03 | 7.00* | | | | 5.75 | | 5.75 | 7.75* | 8.00* |
| 04 | 6.50 | | | | 5.00 | | | 8.00* | 6.50 |
| 05 | | 6.50 | 6.75 | | 7.50* | 5.25 | 5.25 | 7.75* | 5.25 |
| 06 | 7.25* | | | | | | 6.50 | 7.50* | 7.00* |
| 07 | 6.00 | 7.25* | 5.50 | | 6.00 | 5.50 | | 7.00* | 5.75 |
| 08 | | 7.00* | 5.75 | | 6.50 | 6.25 | | 7.50* | 5.50 |
| 09 | | 8.75* | 5.00 | | 8.75* | 8.00* | 7.50* | 8.00* | 6.25 |
| 10 | | | | | | 6.75 | | | |
| 11 | | 6.00 | | | | | | | |
| 12 | | 5.75 | | | 5.25 | 6.50 | | 6.25 | |
| 13 | | | | | | 6.75 | | | |
| 14 | | 6.25 | 5.00 | | 6.25 | | 6.50 | 7.50* | 7.25* |
| 15 | 8.00* | | | | | | 5.00 | 8.75* | 7.25* |
| 16 | 5.50 | 5.50 | | | 5.75 | 6.75 | | 6.00 | |
| 17 | | | | | | | 5.50 | | |
| 18 | | | | | | | 5.25 | 6.50 | 5.00 |
| 19 | 5.00 | | | | | | | 6.75 | |
| 20 | | | | | | | | 5.00 | |
| 21 | 6.75 | 8.50* | | | 6.00 | 5.75 | 5.00 | 8.00* | 6.00 |
| 22 | | 8.25* | | | 9.00* | 6.50 | | 7.50* | |
| 23 | | | | | 5.25 | | | 6.00 | |
| 24 | | 5.00 | | | 7.00* | | | 5.00 | |
| 25 | | 5.75 | | | 6.00 | 7.50* | | 7.25* | |
| 26 | | 7.75* | | | 7.50* | 7.25* | | 7.50* | |
| 27 | | 6.50 | 5.50 | | 8.50* | 7.00* | | 7.00* | |
| 28 | | 5.50 | | | | | | 7.75* | |
| 29 | 7.50* | 7.75* | 5.00 | | 7.50* | 7.00* | 6.00 | 8.50* | 9.00* |
| 30 | | 7.50* | | | 6.75 | 6.25 | | 6.75 | |
| 31 | 5.50 | 5.00 | | | 7.00* | | 6.50 | 8.50* | 7.50* |
| 32 | | 6.75 | | | 5.00 | 6.50 | | 7.75* | 6.75 |
| 33 | | 5.75 | | | | | | 5.50 | 6.50 |
| 34 | 6.00 | | | | | | | 6.00 | |
| 35 | 5.75 | | | | | | | 7.25* | |
| 36 | | 6.75 | 7.50* | | 6.50 | 6.75 | | 8.00* | |
| 37 | | 5.00 | | | 5.75 | | | 5.75 | |
| 38 | | | | | | | | 7.75* | |
| 39 | 5.50 | 8.50* | | 5.00 | 8.50* | 8.50* | 5.00 | 8.50* | |
| 40 | | 7.50* | | | 6.00 | 7.25* | | 8.00* | |
| 41 | 6.75 | | | | 6.25 | 6.00 | | 7.50* | |
| 42 | | 8.00* | 5.25 | | 6.00 | 5.75 | | 6.00 | 5.25 |

| | Stimulus | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Ss | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 |
| 01 | | 6.75 | | | | | | 6.25 | 7.25* |
| 02 | | 7.50* | | 7.25* | 7.50* | | 7.25* | | 7.50* |
| 03 | 6.75 | 8.75* | 7.00* | | | | 8.00* | 8.75* | 8.75* |
| 04 | 5.50 | 7.25* | 7.25* | | | | | 6.50 | 7.00* |
| 05 | 5.25 | 8.50* | | 5.25 | 5.75 | | 5.25 | 6.00 | 7.00* |
| 06 | 6.50 | 6.75 | 7.25* | | | | | 7.50* | 6.75 |
| 07 | | 6.75 | 5.25 | 7.00* | 6.00 | | 8.00* | 6.25 | 7.75* |
| 08 | | 7.00* | | 7.75* | | | 5.75 | 5.75 | 7.00* |
| 09 | | 9.00* | 7.75* | 9.00* | 8.75* | 6.25 | 8.75* | 7.50* | 8.75* |
| 10 | | | | | 5.00 | | 8.00* | | |
| 11 | | | | 5.00 | | | 5.00 | 5.00 | |
| 12 | | 7.50* | | | 6.25 | | 7.75* | | |
| 13 | | 6.75 | 5.75 | | 5.75 | | 7.75* | | |
| 14 | | 7.25* | 5.50 | 6.00 | | | 5.25 | 7.00* | 7.75* |
| 15 | 7.50* | 8.25* | 7.75* | | | | | 7.25* | 8.50* |
| 16 | 5.50 | 6.75 | 5.00 | 6.00 | 6.75 | | | 5.25 | 6.25 |
| 17 | | | | | 5.00 | | | | |
| 18 | 5.75 | 7.00* | 5.25 | | | | | | 5.75 |
| 19 | | 7.75* | | 6.50 | | | 5.00 | 7.00* | 6.75 |
| 20 | | 6.25 | | | | | | | 5.25 |
| 21 | 5.75 | 8.50* | 6.50 | 7.50* | 6.75 | | 5.50 | 6.75 | 8.75* |
| 22 | | 8.50* | | 8.25* | 7.25* | | 5.50 | | 8.75* |
| 23 | | 6.75 | | | | | | | 6.25 |
| 24 | | 6.25 | | 6.25 | 7.00* | | 7.00* | | 6.00 |
| 25 | 5.75 | 7.25* | 6.75 | | 6.75 | | 7.00* | 6.00 | 5.25 |
| 26 | | 7.50* | | 7.50* | 7.25* | | 7.25* | | 6.00 |
| 27 | | 6.50 | | 7.75* | 5.50 | | 7.00* | 5.50 | 7.00* |
| 28 | 5.00 | 7.25* | | 5.25 | 5.25 | | | 5.25 | 7.75* |
| 29 | 8.00* | 9.00* | 8.00* | 7.75* | 5.25 | | 8.25* | 8.25* | 9.00* |
| 30 | | 5.25 | | 5.50 | 6.50 | | 5.50 | | 7.00* |
| 31 | 5.25 | 9.00* | 6.75 | 6.50 | | 5.25 | 5.75 | 6.50 | 9.00* |
| 32 | | 6.50 | | 6.75 | 7.00* | | 5.00 | | 6.75 |
| 33 | | 6.50 | | | 5.00 | | | | 5.00 |
| 34 | 6.75 | 7.00* | 7.00* | | | | 6.50 | | |
| 35 | 5.75 | 6.75 | 6.25 | 6.25 | 5.25 | | | 6.00 | 8.00* |
| 36 | | 6.75 | | 7.00* | 5.00 | 8.00* | | | 8.25* |
| 37 | | | | 5.25 | 5.75 | | 5.50 | | |
| 38 | 5.75 | 8.50* | | 5.75 | | | | | 8.00* |
| 39 | 5.00 | 8.75* | | 7.50* | 8.00* | | 8.50* | | 8.75* |
| 40 | | 7.25* | | 5.25 | 7.25* | | 8.00* | | 7.75* |
| 41 | 6.75 | 8.25* | 5.00 | 6.25 | 6.50 | | 6.25 | 7.50* | 8.00* |
| 42 | | 6.00 | | 7.00* | 5.50 | | 6.25 | | 6.50 |

| | Stimulus | | | |
|---|---|---|---|---|
| Ss | 28 | 29 | 30 | 31 |
| 01 | | 5.75 | | 7.00* |
| 02 | 7.25* | 7.25* | 7.25* | 7.50* |
| 03 | 6.50 | 8.25* | 7.00* | 8.00* |
| 04 | 5.00 | 6.75 | | 6.50 |
| 05 | 5.50 | 6.25 | 7.50* | 8.25* |
| 06 | 7.50* | 6.00 | 5.50 | 8.25* |
| 07 | 6.75 | 7.75* | 9.00* | 8.75* |
| 08 | 5.50 | 5.75 | 6.75 | 7.25* |
| 09 | 8.25* | 8.75* | 8.50* | 8.75* |
| 10 | 5.25 | 8.50* | 6.25 | 7.25* |
| 11 | 5.25 | 6.25 | 5.50 | 7.00* |
| 12 | 5.50 | 7.50* | 8.75* | 6.75 |
| 13 | | 7.25* | 8.25* | 7.25* |
| 14 | 6.25 | 8.50* | 6.00 | 8.50* |
| 15 | 9.00* | 8.75* | | 9.00* |
| 16 | 6.25 | 7.75* | 6.75 | 8.00* |
| 17 | | | | |
| 18 | 7.00* | 7.25* | | 7.50* |
| 19 | 6.50 | 6.25 | 7.25* | 9.00* |
| 20 | | | | |
| 21 | 7.50* | 8.75* | 6.75 | 8.00* |
| 22 | 6.75 | 6.25 | 7.50* | 6.25 |
| 23 | | 6.50 | | 9.00* |
| 24 | | 7.50* | 6.25 | 7.75* |
| 25 | 6.25 | 7.75* | 5.50 | 8.25* |
| 26 | 7.00* | 7.75* | 7.50* | 7.25* |
| 27 | 5.50 | 8.25* | 8.00* | 8.25* |
| 28 | 7.75* | 7.50* | | 7.75* |
| 29 | 8.50* | 9.00* | 5.50 | 9.00* |
| 30 | | 7.75* | 8.00* | 6.75 |
| 31 | 7.75* | 8.75* | 6.00 | 8.50* |
| 32 | 6.75 | 6.75 | 7.25* | 7.25* |
| 33 | | | 7.25* | 5.50 |
| 34 | 7.00* | 8.50* | | 8.50* |
| 35 | 7.75* | 7.25* | 5.75 | 9.00* |
| 36 | | 5.25 | 9.00* | 7.75* |
| 37 | 6.25 | 5.75 | | 5.00 |
| 38 | 8.00* | 8.50* | 5.25 | 8.75* |
| 39 | 7.50* | 8.50* | 9.00* | 8.50* |
| 40 | 7.50* | 7.75* | 9.00* | 6.00 |
| 41 | 7.75* | 8.25* | 5.25 | 9.00* |
| 42 | 8.00* | 5.25 | 7.00* | 5.75 |

INDIVIDUAL SUBJECT SCORES: STRONG ACCENT

| | Stimulus | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ss | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 01 | | | 5.25 | | | 5.50 | | |
| 02 | | | | | 5.25 | 5.25 | | |
| 03 | | | 7.25* | | | 5.75 | | |
| 04 | | 5.00 | | | | 5.75 | | |
| 05 | | | 6.50 | | 5.25 | 6.25 | 5.75 | 6.00 |
| 06 | | 6.00 | | 5.50 | | 5.50 | 5.00 | |
| 07 | | 8.00* | 7.75* | | 6.00 | 6.25 | 6.50 | 5.75 |
| 08 | | 6.25 | 6.00 | | 6.50 | 6.50 | 6.25 | 5.25 |
| 09 | | | 7.25* | | | 7.75* | 8.50* | |
| 10 | | 5.75 | | | | 5.25 | 5.50 | |
| 11 | | | 5.00 | | | 5.75 | 6.25 | |
| 12 | | 7.75* | | | 6.25 | | 5.00 | |
| 13 | | 6.00 | | | | 5.00 | | |
| 14 | | | 5.75 | | | | 8.00* | |
| 15 | | | | | | | 8.25* | |
| 16 | | | 5.00 | | | 5.25 | 7.25* | |
| 17 | | | | 5.75 | | | 5.25 | |
| 18 | | 6.00 | 5.75 | | | | 5.00 | 5.50 |
| 19 | | | 5.75 | | | 5.25 | 6.25 | |
| 20 | | 6.50 | 6.00 | | | 6.00 | | |
| 21 | | | 6.00 | 8.50* | | 7.50* | 8.25* | 6.75 |
| 22 | | 7.25* | 6.00 | | | 5.25 | 5.50 | |
| 23 | | 6.50 | | | | 5.00 | 5.50 | |
| 24 | | | 7.50* | | | | 8.75* | |
| 25 | | 7.25* | 5.50 | | | 5.75 | 6.00 | |
| 26 | | 6.50 | | | 5.75 | | 7.00* | |
| 27 | | | 6.75 | | | 5.25 | 8.50* | |
| 28 | 5.00 | | | | | 5.00 | 7.75* | |
| 29 | | | 6.75 | | | 5.00 | 8.75* | 5.00 |
| 30 | | | 6.00 | | | | 7.50* | |
| 31 | | | 5.75 | | | | 6.00 | |
| 32 | | | | | | | 7.50* | |
| 33 | | 5.00 | | | 6.75 | | 5.00 | 5.50 |
| 34 | | | | | | 5.00 | 7.50* | 6.50 |
| 35 | | | | | | | | |
| 36 | | | 8.00* | | | 5.75 | 9.00* | |
| 37 | | | | | | | 7.75* | |
| 38 | | | | | | 7.00* | 8.50* | 5.00 |
| 39 | | | | | 6.25 | | 5.25 | |
| 40 | | 8.00* | | 5.50 | 6.75 | 5.75 | 6.00 | 5.00 |
| 41 | | 8.00* | | | | | 8.00* | |
| 42 | 5.25 | 5.25 | | | 5.75 | 6.50 | 7.25* | 7.50* |

| Ss | Stimulus 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 |
|---|---|---|---|---|---|---|---|---|
| 01 | | | | | 5.25 | | 6.00 | |
| 02 | 5.25 | 5.75 | 6.00 | 5.00 | 5.00 | | 5.75 | |
| 03 | | | | | 5.00 | 6.50 | 7.50* | 5.50 |
| 04 | | 5.00 | | | 5.25 | 6.50 | 7.25* | 6.50 |
| 05 | 6.25 | 5.50 | | | | 7.25* | 8.50* | 5.00 |
| 06 | | 5.25 | 5.25 | 5.00 | 5.00 | 7.00* | | |
| 07 | 7.50* | 8.00* | 6.25 | 8.00* | 7.75* | 8.50* | 8.75* | 6.25 |
| 08 | 6.00 | 5.75 | | 6.00 | 6.00 | 6.75 | 6.75 | 5.25 |
| 09 | | 8.25* | | | 7.75* | 5.50 | | 9.00* |
| 10 | | | | | | | 6.25 | |
| 11 | | 7.50* | | | | 5.00 | | 6.50 |
| 12 | | | | | | 5.00 | 5.25 | |
| 13 | | | | | | 5.25 | 5.50 | |
| 14 | | | | | 5.25 | | | 7.00* |
| 15 | | | | | | | | 7.50* |
| 16 | | 6.75 | | | | | | 8.75* |
| 17 | | | | | | | | 5.00 |
| 18 | 5.00 | | | | | 5.50 | 6.50 | |
| 19 | | 7.00* | | | | | | 6.50 |
| 20 | | | | | | 6.50 | 7.50* | 5.75 |
| 21 | 7.50* | 7.25* | | | | 5.25 | | 8.75* |
| 22 | 6.00 | 5.25 | | | 5.00 | | 6.75 | |
| 23 | 5.75 | | | | | 5.00 | 8.50* | |
| 24 | | 8.25* | | | 5.00 | 7.75* | | 8.25* |
| 25 | | | | | 5.00 | | 6.25 | 5.00 |
| 26 | 5.75 | | 5.00 | | | 5.00 | 7.00* | |
| 27 | | 8.00* | | | 7.50* | 7.75* | | 8.75* |
| 28 | | | | | | | | 8.00* |
| 29 | | 7.00* | | | 6.75 | 5.25 | | 8.50* |
| 30 | | 7.00* | | | 5.25 | 6.25 | | 8.50* |
| 31 | | 5.50 | 5.25 | | | 5.75 | | 7.50* |
| 32 | | | | | | | | 7.25* |
| 33 | | 5.75 | 5.25 | | 6.00 | 5.50 | | |
| 34 | 5.50 | | | | | | | 6.75 |
| 35 | | | | | | | | 6.75 |
| 36 | | 7.75* | 5.25 | | | | | 7.00* |
| 37 | 5.25 | 7.25* | | | 5.75 | | | 8.50* |
| 38 | | | | | | | | 9.00* |
| 39 | 5.75 | 5.25 | 6.00 | 5.25 | | | 6.50 | |
| 40 | 5.75 | 6.50 | 6.25 | 5.00 | 5.50 | 5.25 | 5.75 | |
| 41 | 6.25 | | | | | | | 8.25* |
| 42 | 5.50 | | 6.25 | | 6.00 | | 5.75 | |

| | Stimulus | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Ss | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 |
| 01 | | | 5.25 | | | | | |
| 02 | | | | | 5.50 | | | |
| 03 | 5.00 | 6.25 | | | | | | |
| 04 | 5.50 | | | | | | | 5.25 |
| 05 | 6.75 | 6.00 | | 6.75 | | | | 6.50 |
| 06 | 5.75 | 5.75 | 6.25 | | 5.00 | | | |
| 07 | 6.00 | 8.25* | | 6.25 | 5.25 | | | 6.00 |
| 08 | 5.75 | 6.50 | 5.00 | 5.25 | | | 5.50 | 5.50 |
| 09 | 6.25 | 6.25 | 7.75* | 8.00* | 7.75* | 8.00* | | 5.75 |
| 10 | | | 5.50 | | | | | |
| 11 | | 5.75 | 5.75 | | 6.25 | 6.75 | | 5.50 |
| 12 | | | | | | | | |
| 13 | | | | | | | | |
| 14 | | 5.75 | 5.00 | | 5.00 | | | |
| 15 | | 5.25 | 7.50* | 5.50 | | | | |
| 16 | 5.00 | 5.50 | 6.75 | 6.75 | 6.00 | 7.00* | | |
| 17 | | | | | | | | |
| 18 | 5.50 | 5.25 | | | | | | 5.25 |
| 19 | | | 5.00 | 5.50 | | 7.00* | | |
| 20 | | 6.25 | 5.00 | 5.25 | | | | 6.75 |
| 21 | 7.75* | 7.50* | 8.50* | 8.00* | | 5.50 | | |
| 22 | 5.50 | 6.00 | 6.50 | | | | | 6.50 |
| 23 | | 5.25 | | 5.00 | | | | |
| 24 | | | 7.50* | 7.25* | 6.75 | 7.00* | | 7.00* |
| 25 | 7.25* | 6.00 | | | | | 5.00 | 5.50 |
| 26 | | | | 5.25 | | | 5.50 | |
| 27 | | 7.00* | 6.75 | 6.50 | 7.75* | 9.00* | | 8.25* |
| 28 | 5.25 | | 7.50* | 7.50* | | 5.50 | | |
| 29 | 6.25 | | 8.25* | 8.50* | 6.25 | 6.00 | | 6.25 |
| 30 | | | 5.75 | 5.50 | 6.25 | 5.00 | | |
| 31 | | 5.50 | 6.75 | 6.00 | 8.00* | 8.25* | 5.00 | |
| 32 | | 5.50 | 7.50* | 6.75 | | 6.25 | | |
| 33 | 6.50 | | | | 6.25 | 5.00 | | |
| 34 | 5.75 | 5.50 | 5.00 | 7.00 | | | | |
| 35 | | 6.00 | 6.75 | 6.50 | 5.00 | 5.50 | | 6.50 |
| 36 | | | 5.25 | | | 7.00* | 6.50 | 7.00* |
| 37 | | | 5.50 | 5.25 | 8.00* | 6.50 | | |
| 38 | | 7.00* | 9.00* | 6.50 | 7.00* | | | |
| 39 | | | | 6.00 | 5.50 | | 6.25 | 5.00 |
| 40 | 5.75 | | | 5.75 | 5.00 | | | 5.75 |
| 41 | | 5.00 | 7.75* | 8.25* | 6.00 | | | |
| 42 | 5.25 | | 7.50* | | | | | |

| | Stimulus | | | | | |
|---|---|---|---|---|---|---|
| Ss | 26 | 27 | 28 | 29 | 30 | 31 |
| 01 | | 6.25 | | | | 5.75 |
| 02 | | 6.00 | | | 6.00 | 5.50 |
| 03 | | 6.50 | | | 5.00 | 6.50 |
| 04 | | 5.00 | 5.25 | | 5.50 | 6.00 |
| 05 | | 6.00 | | 5.25 | 5.75 | 6.00 |
| 06 | | 5.00 | | | 5.75 | 5.50 |
| 07 | | 8.75* | | 5.75 | 8.50* | 8.75* |
| 08 | | 6.25 | | | 7.75* | 7.00* |
| 09 | | 7.75* | 9.00* | 8.50* | 7.50* | 8.50* |
| 10 | | 7.00* | | | | 6.25 |
| 11 | | 5.50 | 6.25 | 5.25 | | |
| 12 | | 6.00 | | | | 6.25 |
| 13 | | 6.25 | | | | 7.00* |
| 14 | | 6.25 | 5.50 | 7.25* | | |
| 15 | | 6.25 | 8.00* | 8.50* | | 5.50 |
| 16 | | 8.75* | 8.75* | 6.75 | 6.75 | 9.00* |
| 17 | 5.00 | | | | | |
| 18 | | 5.50 | | | | 5.75 |
| 19 | | 6.50 | 6.25 | 6.00 | 5.50 | 5.50 |
| 20 | | | | | 5.00 | 6.50 |
| 21 | 7.50* | 8.00* | 7.25 | 8.25 | 6.00 | 7.75* |
| 22 | | 5.50 | | | | 5.50 |
| 23 | | 6.25 | | | | 6.50 |
| 24 | | 8.75* | 8.00* | 6.75 | 5.00 | 6.50 |
| 25 | | 6.00 | | | 6.25 | 5.75 |
| 26 | | 7.50* | | | 5.50 | 5.50 |
| 27 | | 8.00* | 8.00* | 7.25* | 8.50* | 7.50* |
| 28 | | 8.25* | 8.50* | 8.00* | | 7.25* |
| 29 | | 8.25* | 8.75* | 8.00* | | 8.00* |
| 30 | | | 8.25* | 8.25* | | 6.50 |
| 31 | | 6.00 | 8.25* | 6.50 | 6.25 | 8.25* |
| 32 | | 7.50* | 8.50* | 7.50* | 5.50 | 7.75* |
| 33 | | 5.50 | | | | 7.00* |
| 34 | 5.75 | 6.75 | 7.50* | 6.00 | | 7.50* |
| 35 | 6.00 | 7.50* | 8.50* | 8.00* | 7.75* | 9.00* |
| 36 | | 8.50* | 9.00* | 7.00* | 5.75 | 6.00 |
| 37 | | 7.00* | 6.25 | 7.25* | 6.00 | 5.25 |
| 38 | | 7.75* | 9.00* | 9.00* | 5.50 | 9.00* |
| 39 | | 7.75* | | | 7.75* | 6.25 |
| 40 | | 7.25* | | | 7.25* | 8.50* |
| 41 | | 6.75 | 8.75* | 8.25* | | 6.00 |
| 42 | | 6.75 | | 6.50 | 5.75 | 5.25 |

## CURRICULUM VITAE


| | |
|---|---|
| Deborah M. Rekart | Phone (504)769-1255 |
| 749 Rodney Dr. | S.S. 296-48-2725 |
| Baton Rouge, LA 70808 | Birthdate 4-25-51 |


### EDUCATION

| | | |
|---|---|---|
| University of Dayton<br>Dayton, OH | B.A. Spanish | 1972 |
| The University of Rochester<br>Rochester, NY | M.A. Linguistics | 1974 |
| Louisiana State University<br>Baton Rouge, LA | Ph.D. Speech Science | 1985 |

Dissertation director: Raymond G. Daniloff
Title of dissertation: Evaluation of Foreign Accent Using Synthetic Speech

### PROFESSIONAL EXPERIENCE

Research Assistant: U. Dayton Research Institute
Aerospace Information Systems
Dayton, OH
Dec. 72-Apr. 73

Assisted in feasibility study for a computer-based information retrieval system for U.S. Air Force instructional and training methods and educational data. Responsible for report on computerized question-answering system for natural language.

Consultant: Indiana U. General Assistance Center
(Center for Urban and Multicultural Education
Indianapolis, IN
Apr.-May 78

Co-authored grant proposal dealing with the development of procedures to identify and assess English language deficiencies in National Origin minority group members in the Midwest.

Translator: Association of Master Brewers of the Americas
(Districts of Colombia and the Caribbean Region)
Cartagena, Colombia

March 81
Translated technical papers (Spanish-English/English-Spanish) at annual meeting.

TEACHING EXPERIENCE

Teacher of English as a Second Language (TESOL)
Global Escuela de Idiomas
Valencia, Spain
May-July 1973

Teaching Assistant
Purdue University
Dept. of Foreign Languages and Literatures
1977-78; Fall 1979
Spanish Pronunciation

Dept. of Audiology and Speech Sciences
1978-79
Phonetics

Instructor I, II, III
Louisiana State University
Dept. of English
1980-84
Remedial English; Transformational Grammar

HONORS

Award for excellence in Spanish scholarship
Dept. of Foreign Languages (U Dayton)
Spring 1971

Graduated cum laude (U Dayton)
Dec. 1972

Honor Societies:
Sociedad Nacional Hispanica (Sigma Delta Pi)
Alpha Sigma Tau Scholastic Honor Society (U Dayton)
Phi Kappa Phi National Honor Society (Purdue U)

Graduate Scholarships:
U Dayton (Spring 73)
The U of Rochester (1973-74)

Fellowships:

Fulbright-Hays Fellowhsip (Peru) (1976)
Summer Institute of the Linguistic Society of America
U South Florida (1975)
U Illinois (1978)

Research Assistantship:
Louisiana State U (1984-85;1985-86)

PROFESSIONAL SOCIETIES

American Association of Teachers of Spanish and Portuguese
Association for Computational Linguistics
American Speech, Language, and Hearing Association
Acoustical Society of America

FOREIGN STUDY AND TRAVEL

Junior Year Abroad (Valencia, Spain, 1972)
Travel: Spain (1972, 1973, 1983)
Portugal (1972)
Morocco (1973)
Puerto Rico (1978)
England (1983)
Colombia (1978, 1980)

PUBLICATIONS

Rekart, D. M. "Sibilant development in modern Catalan." In F. Aid et al. (eds.) _1975 Colloquium on Hispanic Linguistics_. Georgetown Univ. Press, 1976, pp. 109-122.

Rekart, D. M. "Teaching the Spanish past tense: theory and practice." _Bulletin CILA_., 28 (Oct. 1978), 33-49. Also ERIC #ED 168 335.

Rekart, D. M. "Comments on Harry Reinert's article, 'One picture worth a thousand words? Not necessarily!'," _The Modern Language Journal_, 67, 3 (Mar. 1978), 24-26.

Buckingham, H. W. and D. M. Rekart. "Semantic paraphasia." _J. Comm. Disorders_, 12 (Mar. 1979), 197-209.

Robinson, K. D., D. M. Rekart, and R. G. Daniloff. "A proposal: Use of analysis by synthesis to investigate naturalness of second language pronunciation." Univ. of Essex Occasional Papers (May 1979), 98-113.

Rekart, D. M. and H. W. Buckingham. "La importancia de la dialectología en el estudio de la afasia: un caso cubano." Proceedings of the IVth Symposium on Caribbean Dialectology. Interamerican Univ. Press. To appear.

Rekart, D. M. and R. G. Daniloff. "Review of Clinical Phonetics," J. Phonetics, 13, 2: 253-263.

PRESENTATIONS

"Sibilant development in modern dialects of Catalan." Second Colloquium on Hispanic Lingusitics, U. South Florida, Tampa, FL (July 1973).

"La importancia de la dialectología en el estudio de la afasia, un caso cubano." (with H. W. Buckingham) IVth Symposium on Caribbean Dialectology, Interamerican Univ., San German, Puerto Rico (Apr. 1979).

"Bilingualism, dialectology and aphasia." (with H. W. Buckingham) 1979 Annual Meeting of the American Speech, Language and Hearing Assoc., Atlanta, GA (Nov. 1979).

"A lingusitic analysis of some syntactic and spelling errors." (with H. W. Buckingham) Evaluation and Error Analysis Seminar, Purdue Univ., W. Lafayette, IN (Mar. 1980).

"Evaluation of foreign accent using synthetic speech," (with P. Hoffman, and R. Daniloff), Acoustical Society of America (April, 1985).

PAPERS SUBMITTED FOR PUBLICATION

"Reliance on Orality in the Production and Recognition of Run-on Sentences" (with Jill Brody)

GRANT PROPOSALS

"Evaluation of foreign accent using synthetic speech" (National Science Foundation, 1979) not funded.

"Evaluation of foreign accent using synthetic speech" (National Science Foundation, Grants for Improving Doctoral Dissertation Research, 1984) not funded.

Proposal for DELTA (speech synthesizer developed by Hertz) (The Frost Foundation, 1985) funded: $40,000.

LABORATORY EXPERIENCE

Kresge Hearing Research Lab. of the South (New Orleans, LA) 1984-present

Haskins Laboratories (New Haven, CT) October, 1984

**DOCTORAL EXAMINATION AND DISSERTATION REPORT**

**Candidate:** Deborah M. Rekart

**Major Field:** Speech & Hearing Science

**Title of Dissertation:** Evaluation of Foreign Accent Using Synthetic Speech

**Approved:**

Raymond G. Daniloff
Major Professor and Chairman

William Cooper
Dean of the Graduate School

**EXAMINING COMMITTEE:**

Paul R. Hoffman

Enrique Diaz

Mike Richardson

Adelaide M. Russo

**Date of Examination:**

12/2/85