Ann. Bull. RILP, No. 14 1-27 (1980)

PHONETIC MANIFESTATION OF SYNTACTIC STRUCTURE IN ENGLISH*

Ilse Lehiste**

Abstract

The topic of this lecture is the phonetic manifestation of syntactic structure in English. An orally produced sentence has both a phonetic structure and a syntactic structure. In much of my research I have tried to establish the phonetic structure of spoken utterances independently of syntactic structure, and then investigate the possible relationship between units established on the basis of phonetic criteria and units posited on the basis of a syntactic analysis of the same utterance.

Linguistic units can be established by delimiting their boundaries. I look first at words, which enter into syntactic units on the one hand, and into rhythmic units on the other hand. There exist ways in which word boundaries can be signalled in English; some such ways are described and illustrated. But the phonetic realization of sounds and syllables within the words is variable and depends on a number of factors, such as the position of the word relative to a syntactic boundary or to a pause. In particular, the temporal structure of a word is determined by its position within the larger utterance.

It has been claimed that English is a stress-timed language, characterized by isochrony. This means that stressed syllables follow each other at regular intervals. A review of recent studies of isochrony shows that to a certain degree, isochrony is present in the production of spoken English, and that it is psychologically real: listeners expect English sentences to be produced in such a way that stressed syllables follow each other at regular intervals. This fact appears to be utilized for the purpose of signalling the presence of syntactic boundaries: in otherwise ambiguous sentences, the speaker can indicate which of two syntactic structures he has in mind by increasing the interstress interval that contains the syntactic boundary.

Rhythmic structure and syntactic structure are thus interrelated. In English, the controlled timing of articulatory gestures takes the rhythmic structure of speech into account. Stressed syllables carry the greatest amount of information; attention has to be focussed on the stressed syllables, and this is facilitated by setting up an expectation as to when the next stressed syllable is likely to occur. Producing sentences in such a way that stressed syllables occur at isochronous intervals contributes to

*Text of the open lecture at the University of Tokyo, Faculty of Medicine, Tokyo, Japan. May 15, 1980
**Professor of Linguistics, The Ohio State University. Guest Professor at the Research Institute of Logopedics and Phoniatrics. April to May, 1980

optimal perception by the listeners whose attention is cyclically directed to the points in time at which the stressed syllables can be found. A disruption of the expected pattern can then be used to convey crucial information about syntactic structure.

The topic of this lecture is the phonetic manifestation of syntactic structure in English. A basic premise of my approach to this question is the notion that a spoken utterance is structured in two ways at the same time: its constituent units may be recognized by their phonetic properties, or they may be identified at the level of morphosyntax and semantics. Both aspects are simultaneously present. I submit that there also exists a code that connects the two kinds of structures--phonetic structure and syntactic structure.

One way of approaching the problem is to start from syntax and semantics and look for the phonetic manifestation of syntactically determined units, such as noun phrases and verb phrases. Another way would be to establish the phonetic structure of spoken utterances independently of syntactic structure, and then investigate the possible relationship between units established on the basis of phonetic criteria and units posited on the basis of a syntactic analysis of the same utterance. This has been my approach in much of my recent research.

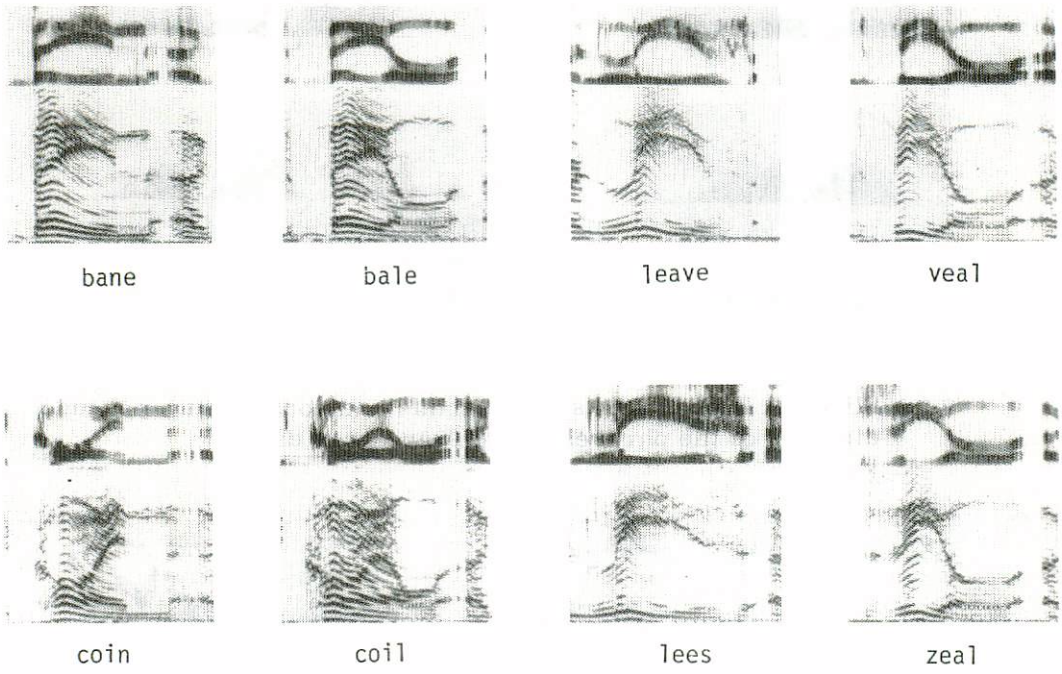
A sentence can be viewed as consisting of words, and words can be defined as lexical items. A study of the phonetic characteristics of words produced as part of a sentence reveals that lexical words and phonological words are not necessarily co-terminous. In English, and in a large number of other languages, there exist so-called clitics: lexical items (usually function words) that have no inherent stress, so that when they appear in a spoken sentence, they get phonetically attached to a neighboring word that does carry a certain degree of stress. The indefinite article in English almost always combines with the following noun into a phonological word: there is no phonetic difference in the syllable structure of the sequences among and a man. These examples could be multiplied; they constitute evidence for the parallel existence of lexical words and phonological words.

Words have a number of phonetic properties, segmental and suprasegmental. Let us consider first some segmental characteristics.

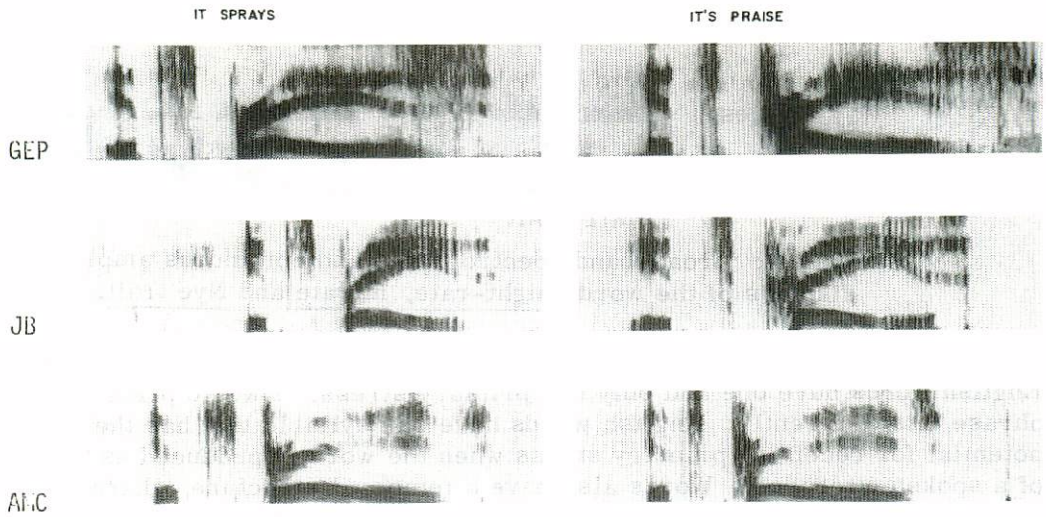
Already in 1960 (Lehiste 1960, Lehiste 1964) I established the word-initial and word-final allophones of various English phonemes. Some examples are presented on Slide 1.

An interesting characteristic of English is the fact that selection of allophones, together with the suprasegmental feature of duration, can be used to distinguish between minimal pairs that consist of the same segmental phonemes, but differ in the position of word boundary. For example, it sprays differs from it's praise primarily due to the different allophones of /p/: in it's praise, /p/ is in initial position, and therefore strongly aspirated. In sprays, /p/ has an unaspirated allophone that occurs in

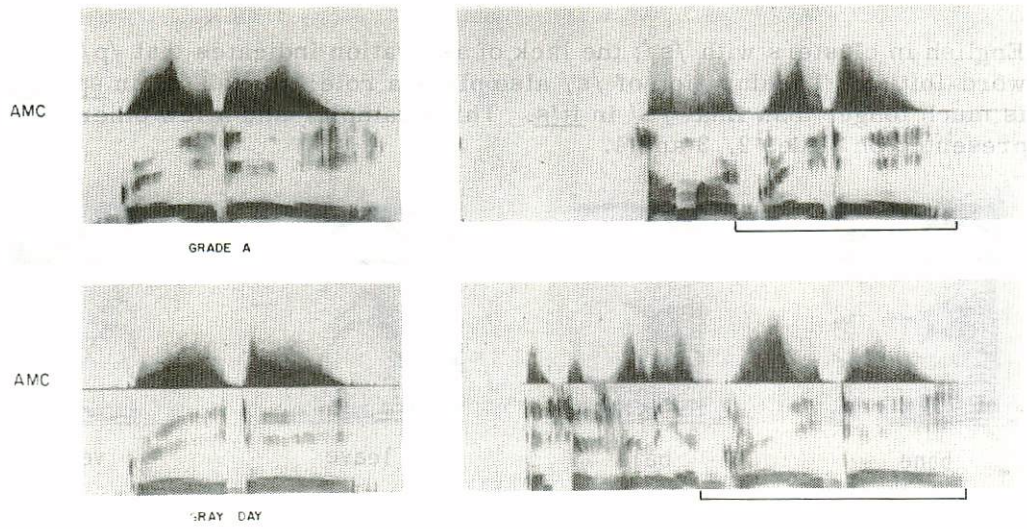
English in clusters with /s/; the lack of aspiration indicates that /p/ is not word-initial. The duration of /s/ also plays a role: initial /s/ in sprays is much longer than final /s/ in it's. This example, and a few other, are presented on Slides 2, 3 and 4.



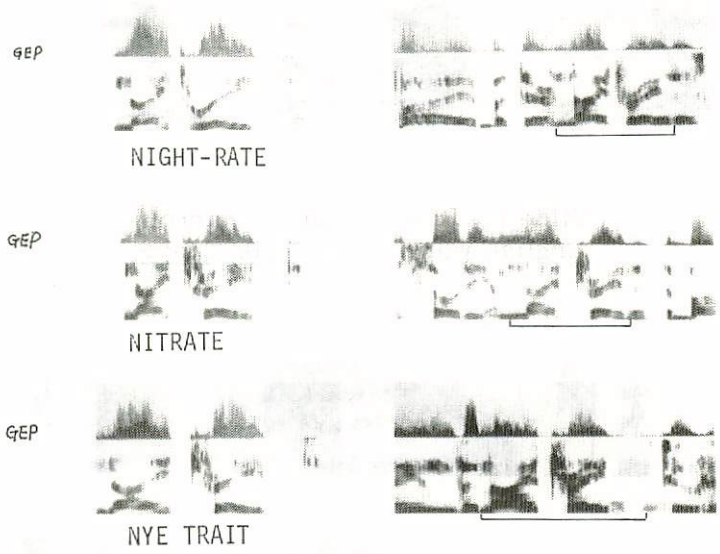
Slide 1: Broad-band and narrow-band spectrograms of the English words bane, bale, leave, veal, coin, coil, less, zeal.



Slide 2: Broad-band spectrograms of it sprays and it's praise, produced by three informants.



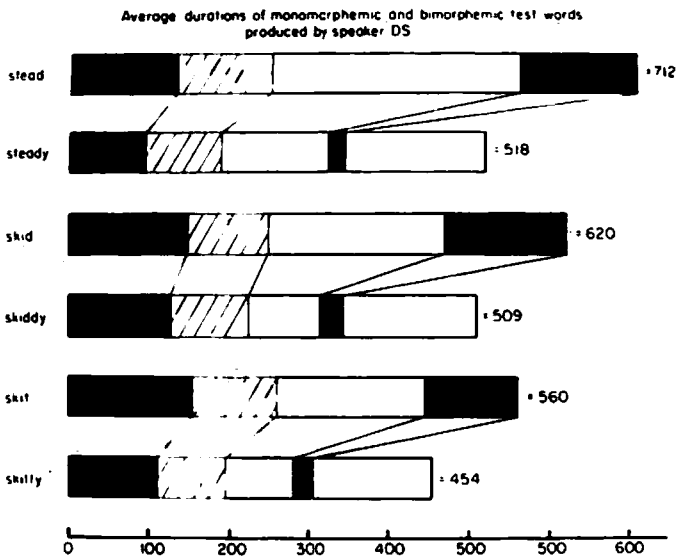
Slide 3: Broad-band spectrograms and continuous amplitude displays of the phrases grade A and gray day.



Slide 4: Broad-band spectrograms and continuous amplitude displays of the words night-rate, nitrate and Nye trait.

Words also have some suprasegmental characteristics. For example, English words have one and only one primary stress. I would prefer to phrase this differently: English words have one syllable that has the potential for receiving primary stress when the word is produced as part of a spoken utterance. Words also have a temporal structure, characterized by a tendency toward maintaining a certain basic duration. As a result, there exists an inverse relationship between the number of syllables in a word and the duration of these syllables. A monosyllabic word and a disyllabic word tend to have the same duration. I investigated this is a series

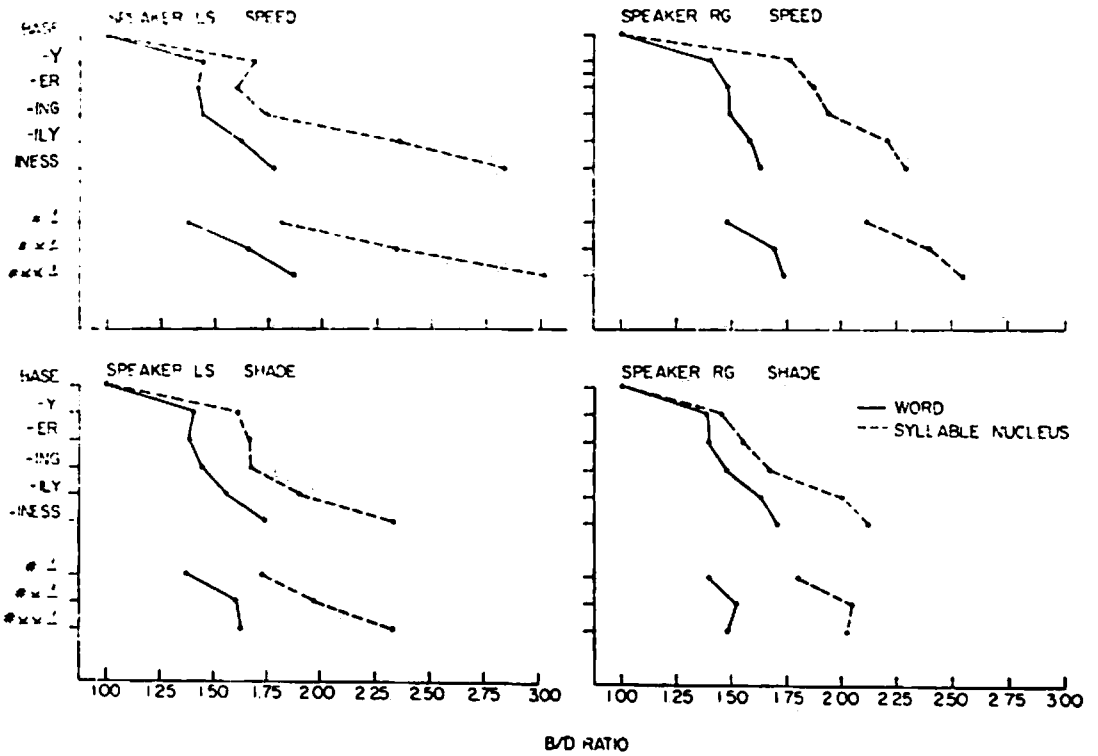
of studies published in 1971 and 1972 (Lehiste 1971, 1972). First I had two speakers produce a set of ten words including the words stead, steady, skid, skiddy, skit, and skitty. The words were produced approximately 110 times each. Among other things, I measured the average durations of all segments and I found the disyllabic word was, in each case, actually shorter than the corresponding monosyllabic word.



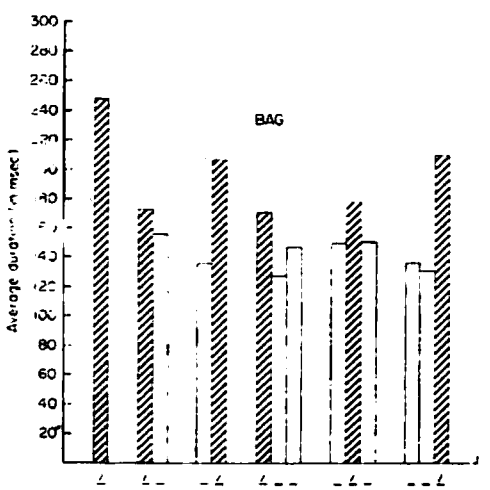
Slide 5: Average duration of productions of the words stead, steady, skid, skiddy, skit, and skitty.

Then I selected four base words and provided them with a series of suffixes, for example speed, speedy, speeder, speeding, speedily, and speediness. Two speakers produced twenty tokens of each set. As part of the statistical analysis of these words, I computed the ratio of the durations of the base word (pronounced by itself) and the durations of the same segments occurring in the derived word. For example, the mean duration of speed was divided by the mean duration of the speed part of the word speedy. The ratios were always larger than 1--that is, the duration of the base part of the derived word was always less than the duration of the same set of segments in the base word.

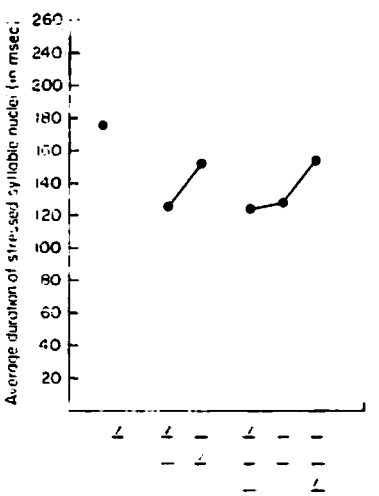
An interesting finding was the fact that the number of phonemic segments in the suffix had no systematic effect on the duration of the stem. This becomes clear from the behavior of stem forms before the suffix -ily. This suffix was pronounced with a syllabic /l/ by both speakers in all productions. Thus the stems of words like speeding [spi:dɪŋ] and speedily [spi:dlɪ] were followed by two segments each, but the -ing suffix was monosyllabic and the -ily suffix was disyllabic. In all cases, the disyllabic suffix produced greater reduction in the duration of the stem than the monosyllabic suffix, although both consisted of the same number of segments.



Slide 6: Ratios of base form/derived form in sets of words built on the bases speed and shade.



Slide 7: Average durations of the base form bag, produced in various positions.



Slide 8: Average durations of stressed syllable nuclei in various positions.

It was thus established that the number of syllables in a word influences its durational structure. It appeared interesting to study whether the position of a syllable in the word influences the duration of that syllable. I conducted a study for the purpose of establishing what happens to a syllable when the same syllable occurs in any position in words of one, two or three syllables, carrying any possible stress pattern (Lehiste 1975b). There were 68 words formed of the four monosyllables bag, back, big, and bick. (While three of the four words are meaningful, the disyllabic and trisyllabic words formed from them were nonsense words, e. g. bagbigbag.) Three informants read the words in three different frames, for a total of 612 productions. Among the results was the finding that word-final syllables are longer than equivalent word-initial or word-medial syllables.

Similar results were obtained around the same time by several researchers for different languages (Lindblom and Rapp 1973; Nooteboom 1972; Klatt 1973; Oller 1973).

Given the observation that syllables are longer when they are in final position in a word, one might question whether they are longer precisely for that reason, or whether there might not be some other causes. In a large number of instances, the word boundary is simultaneously the boundary of a higher-level unit of some kind, be it determined syntactically or phonologically. It might well be that observed word-final lengthening is actually lengthening before some kind of boundary.

I prefer the term "pre-boundary lengthening" to the earlier term "pre-pausal lengthening", since it is broader and covers more instances. "Pre-pausal lengthening" was used when it was first observed (by Gaitenby 1965) that syllables before a sentence boundary are considerably lengthened. Gaitenby worked with isolated sentences which are, of course, always followed by pauses. Even in connected speech, a pause may frequently be present at the end of a sentence. But syntactic boundaries within a sentence are not normally marked by pauses, while lengthening may nevertheless be noted; thus the term "pre-boundary lengthening" appears to me more appropriate.

Pre-boundary lengthening has been observed at the boundaries of syntactic constituents of spoken sentences by several scholars, notably William E. Cooper, who in his 1975 dissertation (Cooper 1975) found different degrees of preboundary lengthening before different kinds of subordinate clauses. Cooper's work is a good representation of the approach that starts from syntax and looks for the phonetic manifestation of syntactically determined relationships within sentences (Cooper 1976; Cooper, Lapointe, and Paccia, 1977; Cooper, Sorensen, and Paccia, 1977; Cooper, Paccia and Lapointe, 1978).

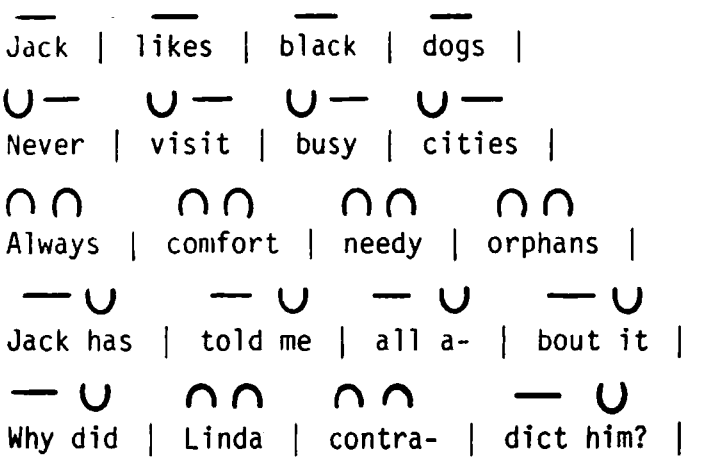
In these studies, the units whose boundaries were both marked by and perceived with reference to preboundary lengthening were determined syntactically. An alternate explanation is available for phenomena of this kind, however. It is possible that what is perceived is not just segmental lengthening before a boundary, but the resultant change in the overall

rhythmic structure of the sentence.

Rhythm is the property of spoken language; it is not inherent in syntax. The term rhythm refers to the timing of stressed syllables and the intervals (usually filled with unstressed syllables) that separate them. There exists a long tradition in British phonetics according to which spoken utterances are divided into metric feet that are independent of the syntactic structure of the sentence (Abercrombie 1964). Furthermore, the metric feet are claimed to be of approximately the same duration, which means that in English, stressed syllables follow each other at approximately equal intervals.

I have carried out a series of investigations into the nature of the phenomenon called isochrony. (Lehiste 1973a; Lehiste 1975a; Lehiste 1977). I was particularly interested in the relationship between speech rhythm and its syntactic structure. This led me to investigate the production and perception of various types of linguistic units, both rhythmic and syntactic, and to analyze the phonetic disambiguation of syntactic ambiguity (Lehiste 1973b; Lehiste, Olive and Streeter, 1976). I shall give an outline of these studies to illustrate the development of this line of research into the phonetic manifestation of syntactic structure in English.

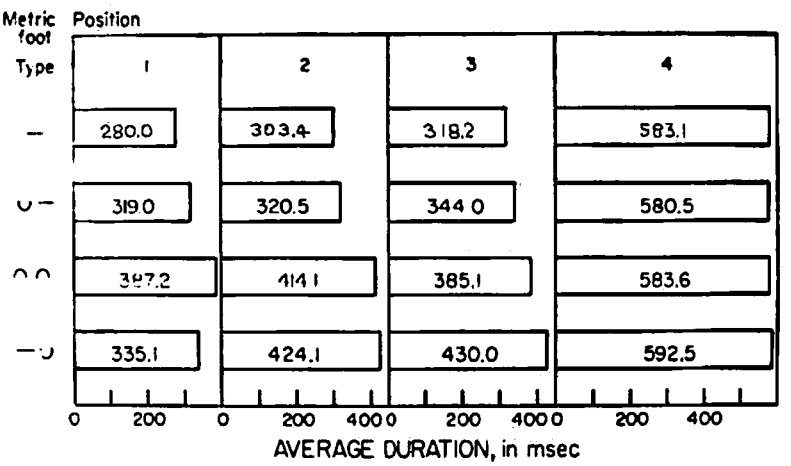
To investigate the possible presence of isochrony in production, I constructed 17 sentences, each consisting of four metric feet. The term "metric foot" refers to isochronous rhythmic units lasting from one major-word-level stress to the next. Abercrombie had claimed that metric feet of different types are nevertheless of equal duration. He set up a monosyllabic foot type, three disyllabic types, and several trisyllabic types. The monosyllabic type consists of a single stressed syllable. The disyllabic metric foot types are made up of three quantitatively different syllable types: a short syllable, consisting of one time unit; a medium syllable, consisting of 1½ time units; and a long syllable, having the value of 2 time units. These syllables combine into disyllabic feet as follows: short+long = 1+2; medium+medium = 1½+1½; long+short = 2+1. Thus all disyllabic feet have the same duration, namely three time units. The sentences that I constructed had four metric feet, and each foot type occurred in each position approximately the same number of times.



Slide 9

Slide 9 presents four of the sentences used in the study, analyzed according to their metric structure. The sentences are: Jack likes black dogs (all metric feet consist of monosyllabic words); Never visit busy cities (all metric feet are disyllabic short-long); Jack has told me all about it (all metric feet are disyllabic long-short); and Why did Linda contradict him? (two of the metric feet are long-short, two medium-medium).

The sentences were produced by two speakers, who read the randomized sequence of 17 sentences ten times each. The set of materials thus consisted of 340 utterances, comprising 1360 metric feet. The material was analyzed acoustically, and measurements were made of all metric feet.



Slide 10: Average durations

Some of the results confirmed earlier expectations. For example, the final foot turned out to be much longer than the other three. This shows that preboundary lengthening is also found at the level of rhythmic units--metric feet--and not necessarily only at the level of words or syllables. The initial foot appears shorter than the others, but this is partly due to the impossibility of measuring the duration of voiceless initial plosives from the acoustic signal. The relevant positions are thus positions two and three. As becomes clear from the slide, the same foot types have remarkably similar durations in positions 2 and 3; the differences are probably within the error of measurement in most instances, and below the perceptual threshold in every instance. If isochrony is defined as similar duration of successive feet of the same type, there is indeed a considerable amount of isochrony present. On the other hand, there are clear differences in the average durations of different foot types in the same position, i. e. within either position 2 or position 3. This difference is in several cases greater than can be attributed to measurement error. There seems to be a hierarchy of durations, proceeding from monosyllabic feet to short-long to medium-medium to long-short metric feet. While the difference in average durations between the first two and second two types is probably below the perceptual threshold, differences between the two groups should be perceptible. In this sense the feet are not isochronous.

In final position, however, the differences between metric foot types are minimal and clearly below the perceptual threshold. While all feet in final position are longer than feet of the same type in other positions, they tend to have the same duration, and the difference associated with foot types appears to have been neutralized.

Several other linguists have studied isochrony, both before and after my investigation which I have just described. The first instrumental investigation of isochrony seems to have been that of Classe (1939). Classe failed to find absolute isochrony, but nevertheless did not dismiss isochrony out of hand; he argued that it constitutes the basis of the rhythmic system of English, "although frequently it only remains as an underlying tendency of which some other factor at times almost completely obliteratures the effects". Not having found perfect isochrony in objective measurements, he held open the possibility that isochrony might be a subjective phenomenon: "In speech, long groups, provided other circumstances are not too unfavorable, will tend to be made subjectively isochronous by the reader or listener because of his speech habits".

Classe's formulation is rather careful: he claims neither perfect isochrony in production nor in perception, just a tendency to speak in rhythmic units that are perceived as isochronous.

Bolinger (1965) had six speakers record two rather lengthy sentences, identified the accents, and measured the intervals between the accents. His results gave little support to the idea of isochronous rhythm in production. Of the 53 intervals, 13 has approximately twice the length of the shortest interval. The lengths of the intervals appeared to be determined by syllable structure, nearness to initial or final position, and relative semantic importance, besides the number of syllables. Bolinger concluded that such factors seem to have a good deal more influence than rhythm has in determining the length of accentual groups.

O'Connor (1965) recorded a limerick with as strict a rhythm as possible, and accompanied the speech with a click, produced by hand, at each stress. There were 15 stress groups with an average duration (measured from click to click) of 518 msec. The duration difference between the shortest and longest stress group (488 and 566 ms) was 88 ms. O'Connor concluded that physical isochrony was clearly not present even under these very favorable conditions. In 1968, O'Connor studied a set of seven utterances, each containing three monosyllabic feet. The first and third foot remained constant, the second varied in segmental length from three to nine segments. Four speakers read the sentences once each, and a fifth speaker read the entire set 10 times. Duration measurements showed that the variable foot had a clear tendency to greater duration as segmental length increased. There was no evidence that the duration of the frame items adjusted itself to compensate for the changes in the variable foot. On the basis of these results, O'Connor expressed continued doubt as to the existence of isochrony in production.

Uldall (1971, 1972) analyzed a reading by David Abercrombie of "The

North Wind and the Sun". The reading, which lasted for 45 seconds, was later divided into rhythmic feet by Abercrombie. There were 56 metric feet, ranging in duration from 260 to 870 msec. In spite of the large differences between the extremes, Uldall acknowledged a tendency to isochrony for this speaker in the moderately slow reading style employed by him. More than half (57%) of the filled feet, i. e. feet not containing any pauses, fell between 385 and 520 msec. The average duration of all filled feet was 520 msec; the average duration of all monosyllabic feet was 440 msec, all disyllabic feet 510 msec, and all trisyllabic feet 540 msec. The four-syllable feet, of which the text contained six, had an average duration of 760 msec and thus differed considerably from the average.

One of the more detailed recent instrumental studies of isochrony in production is that by Lea (1974). Using eight talkers reading 31 sentences, six talkers reading the Rainbow script (a standard text used in speech experiments), and two talkers reading a monosyllabic script, Lea tested several hypotheses concerning isochrony. The first of these corresponds to the standard notion that stressed syllables follow each other at regular intervals. This presupposes that the number of unstressed syllables between two stresses has little or no effect on the interstress time intervals. Pike (1945) had described isochrony as the crowding together of unstressed syllables as their number increases between two stresses. Lea's results show that interstress intervals are indeed substantially affected by the number of intervening unstressed syllables; the average time intervals appeared to increase almost linearly with the number of intervening syllables. The original definition of isochrony was therefore rejected.

Lea hypothesized that his results might be interpreted as a manifestation of an alternating stress and unstress pattern: "As one inserts more and more unstressed syllables between two stresses, he tends to make one of the intervening syllables more like a stressed syllable, to re-establish something like the ideal alternation pattern." This suggests that whenever three or four syllables intervene between stresses, one of them will acquire some characteristics of a stressed syllable. This was indeed the case for most of the interstress intervals with three or four intervening syllables in Lea's materials: of 38 such instances, 21 included a syllable which was perceived as stressed by at least one of his listeners, and of the 17 remaining cases, 10 had syllables that were declared stressed by Lea's stressed syllable location algorithm.

Histograms published by Lea of the number of occurrences of various sizes of interstress intervals show both a fairly large amount of clustering around certain mean values and a large amount of variability. For example, the average interstress interval for eight speakers producing 31 sentences was 532 msec, with a standard deviation of 230 msec. For one of the speakers, the mean was 480 msec, the standard deviation 198 msec; for another speaker, the mean for the Rainbow script was 470 msec, the standard deviation 131 msec, and for the Monosyllabic script, 502 and 184 msec. The regularities thus are quite apparent, even though absolute isochrony could not be found.

All of these studies have shown that interstress intervals vary in duration to a greater or lesser degree. Most investigators have therefore either rejected the claim that English is a language characterized by isochrony, or have attempted to reinterpret the experimental findings to take into account the fact that perfect isochrony cannot be found in production. An example is provided by Halliday (1967), who claims that isochrony is phonological. This makes it possible for him to overlook phonetic differences in the realization of phonologically isochronous units. In Halliday's system, the phonological units are, in descending order, the tone group, foot, syllable and phoneme. It is the foot that is characterized by phonological isochronicity, which Halliday describes as follows: "There is a tendency for salient syllables to occur at roughly regular intervals of time whatever the number of weak syllables, including zero, in between" (p. 12). Rees (1975) proposes the tone group as the domain of isochrony (in Halliday's system, the tone group is the basic phonological unit of intonation). A sentence may consist of more than one tone group; the possibility remains open that two or more consecutive tone groups might have their salient syllables fall at regular intervals, and that thus isochronicity might appear to embrace the whole sentence, but according to Rees, such instances are rare.

I would like to repeat here the reasons for being interested in isochrony in considering the phonetic manifestation of syntactic structure. The question is whether the phonetic realization of a sentence can be completely predicted from its syntax. I maintain that this is not true; spoken utterances have a rhythmic structure that is basically independent of syntax. To anticipate the outcome, this rhythmic structure is integrated into the grammar of English at a syntactic level; but in order to describe the manner in which rhythm and syntax interact, it is necessary at first to establish the phonetic reality of the rhythmic structure of English sentences.

In deciding whether English is indeed a language characterized by isochrony, one must take into account the constraints on production and perception that might influence isochrony. Two questions that need to be settled are (a) what the limits of the ability of a speaker to produce regular rhythm, and (b) what are the limits within which a listener perceives regular rhythm. The constraints on production will be discussed first.

Allen has explored the control of speech timing in a series of studies (Allen, 1972, 1973, 1975). In the 1975 paper he summarizes several previous studies of the temporal variability of speech, and reports that the variability of production of speech segments matches the variability of other rhythmic activities such as finger-tapping. Short speech segments have variabilities of about 10%, longer stretches of speech about 4%, while the overall range for standard errors for motor rhythms is about 3-11% of the length of the time intervals being produced.

The variability in interstress interval durations reported, for example, by Lea, was considerably larger than 10%. However, the materials used in previous studies were quite heterogeneous, and some of the differences may have been due to the different type and length of the sentences. I

would like to quote Classe (1939) again at this point: Isochrony exists, but only under favorable circumstances, namely, "the groups concerned must not contain very different numbers of syllables; the phonetic structure of the component syllables must not differ too widely; the grammatical connection between the groups and the grammatical structure of these groups must be similar" (p. 85). To test Classe's theory under conditions outlined by Classe, I conducted several studies in which I analyzed relatively more homogeneous material (Lehiste 1973a, Lehiste 1975a).

I mentioned already my study involving sentences consisting of four metric feet. The range of differences between the metric feet durations was from 10 msec to 133 msec. Clearly there exist differences in the size of the interstress intervals in production. However, most of the differences were so small that it seemed reasonable to assume they would be below the perceptual threshold, which for metric foot durations in the range of 300 – 500 msec would be about 10% of the duration of the metric foot. If the differences are indeed below the perceptual threshold, they are perceptually irrelevant and from the point of view of perception, the rhythm of the sentences must be considered isochronous.

Just noticeable differences in duration have been established in a series of earlier investigations (cf. Lehiste 1970 for a review). In attempting to compare the differences which emerged in my studies with previously established thresholds for duration, I wondered whether these thresholds were directly applicable to my data, since the published just noticeable differences had been established on the basis of comparing two stimuli, and I was concerned with sequences of four intervals. Therefore I ran a few additional experiments.

I reproduced the temporal patterns of the four-measure sentences as non-speech stimuli. The durations of the measures were replicated as noise-filled intervals separated by clicks. The stimuli were produced on a Glace-Holmes speech synthesizer. Thirty listeners judged both the actual sentences and the sequences of filled intervals, deciding in each case which of the four units was longest or shortest.

In the case of spoken sentences, listeners had considerable difficulty in identifying the measures which were actually the longest or shortest. With nonspeech materials, the corresponding intervals were identified with much greater success. I reasoned that if listeners cannot identify the actually longest or shortest measures in spoken English sentences, the measures must seem to them to have equal duration: if you cannot tell them apart, they must be alike. Isochrony would then be a perceptual phenomenon. The fact that listeners did better with nonspeech materials suggests that the phenomenon is language-bound: isochrony would then characterize spoken language, in this case English, rather than being a general feature of the perception of rhythm. At least if there is a gradient, it is slanted in favor of perception of spoken language.

The just noticeable differences for duration reported in the literature had been established by using nonspeech stimuli, like pure tones or white

noise. Since listeners did better with nonspeech stimuli and did not do as well when they listened to speech, one might assume that the just noticeable differences for spoken language are actually larger than the just noticeable differences for duration that have been reported in psychophysical literature. Just noticeable differences established for nonspeech can then be considered the baseline against which the perceptibility of durational differences in speech may be measured. I ran an experiment to establish such a baseline for sequences of four intervals and presented a report at the 8th International Congress of Phonetic Sciences in Leeds, in 1975. (The paper has just been published in the Journal of Phonetics, 1979).

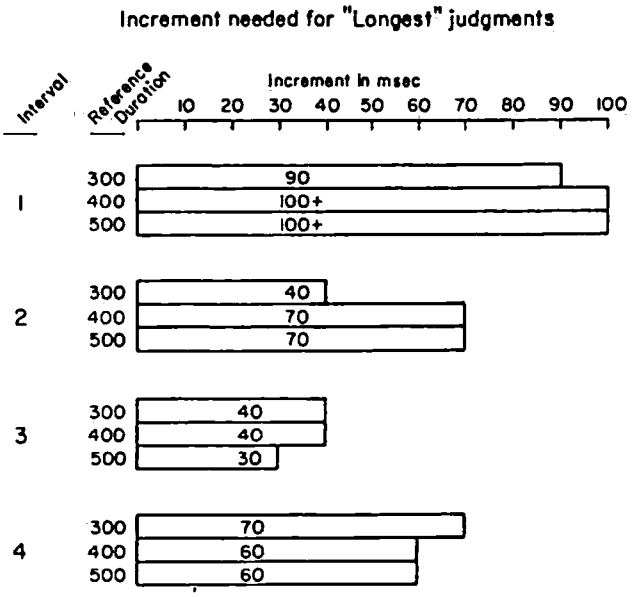
I chose three basic reference durations: 300, 400 and 500 msec. These durations corresponded to the range observed in actual productions of metric feet in my four-measure sentences. For each reference duration, the length of each of the four intervals, one at a time, was decreased and increased in nine 10-msec steps. Three of the four intervals were always of the same duration; one of the four was either shorter or longer. The sequences were again produced using a Glace-Holmes speech synthesizer, randomized and presented to thirty listeners, who were asked to identify first the longest intervals, and on a second presentation, were asked to identify the shortest intervals. To control for response bias, tokens of sequences consisting of equal intervals were also included in the test. In the cases when all intervals had the same duration, the listeners tended to hear the first interval as longer than the others; making "shortest" judgments, the fourth interval was more frequently judged "shortest", even though all intervals had objectively equal duration.

Slide 11 shows the increment needed for "longest" judgments. The results showed that in the first position, a 90 msec increment produced agreement among the listeners that the first interval was longest, when the reference duration of 300 msec. With reference durations of 400 and 500 msec, even an increment of 100 msec was insufficient to get agreement among the listeners that it was the first interval that was longest.

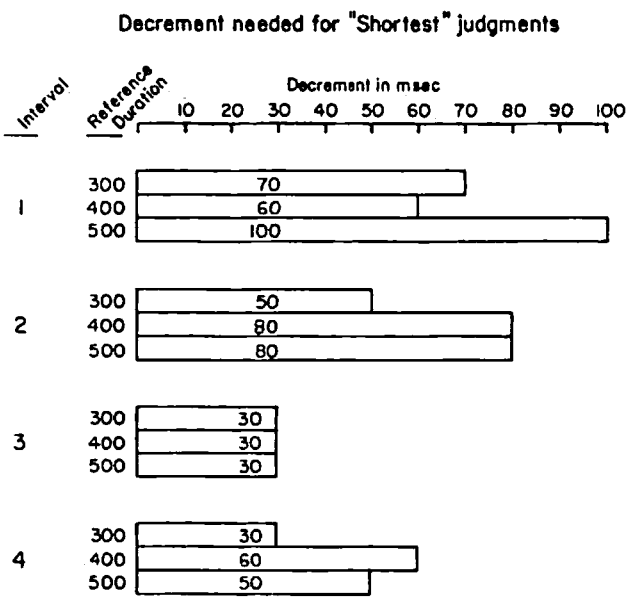
The listeners seemed to be most sensitive to changes that occurred in the third position: here, an increment of 30 msec produced agreement among listeners, when the reference duration was 500 msec. All other increments had to be larger than 30 msec.

Slide 12 shows the decrement needed for "shortest" judgments. Again it is the first whose duration the listeners found it hardest to judge. And again it is the third interval that the listeners can judge with greatest precision: a decrement of 30 msec was sufficient to achieve agreement that the third interval is shortest.

If my assumption is correct and the difficulty in identifying actually longer or shorter intervals may be interpreted to mean that in such cases the intervals must sound alike in duration (if you cannot tell that one of them is different, they must be perceptually the same), then quite large differences in duration are not really perceptible. Most of the differences that I had observed in the production of four-measure sentences were actually smaller than the differences that emerged as the limits of perception



Slide 11: Increment needed for "longest" judgments



Slide 12: Decrement needed for "shortest" judgments

in this particular study. There is then very good reason to believe that the differences found in productions of sentences would be even less likely to be perceptible, since my results had shown that listeners perform less accurately when they are judging durational differences within real speech.

The just noticeable differences in my sequences of four noise-filled intervals ranged from 30 to more than 100 msec. I claim that in listening to speech, the hearer cannot perform any better and is likely to perform somewhat worse, so that the just noticeable differences for speech would be somewhat larger. Differences in the duration of interstress intervals found in several production experiments are of this order of magnitude. For example, Uldall (1971) found that the average duration of monosyllabic feet was 400 msec, disyllabic feet 510 msec, and trisyllabic feet 540 msec. O'Connor (1965) reported a durational difference between the longest and shortest stress group of 88 mec; all other differences had smaller values. I believe it quite likely that such differences are simply not perceived. Thus sentences that are not produced with absolutely isochronous intervals between stresses may still be perceived as if the interstress intervals were identical.

It is, furthermore, quite likely that the listener imposes a rhythmic structure on sequences of interstress intervals in spite of the fact that their durational differences are above the perceptual threshold. This seems to be a fairly general phenomenon. It is well known that when we hear a sequence of short sounds (pulses or clicks), separated from each other by time intervals that are no smaller than about 0.1 seconds and no longer than three seconds, we will impose some rhythmic structure on the sequence. Allen (1975) reviewed the literature dealing with the phenomenon. Two related aspects of the problem can be identified. One of them involves imposing a rhythmic structure on a sequence of identical pulses, so that one of the pulses sounds subjectively stronger. The other involves underestimation of the duration of long time intervals and overestimation of short intervals, as a result of which we may hear sequences of only approximately equal time intervals as more equal than they really are. Allen notes that listeners have a general tendency to adjust their perception of time interval durations toward some central, or average, duration; this, in addition to the tendency to impose a rhythm on any sequence of intervals, contributes to the perception of regular rhythm in languages with stress accent.

Suggestions have in fact been made by several researchers in recent publications that listeners may partially compensate for existing durational constraints and perceive speech as being more rhythmic than it really is. Barnwell (1971) attempted to develop an algorithm for segment durations and isolated a number of durational constraints in English which counteract an observed tendency to make words approximately the same size. He concluded that all words cannot be forced into the same size, but only moved in the proper direction, and conjectured that this may have some importance in the perception of rhythm in English. I would like to quote his insightful conjecture in full: "The implication is that speech attempts to be very rhythmic, but fails because of durational constraints. Hence it may be that

what is heard as rhythmic may be really the interpretation of changes in duration in the direction of true rhythm. Hence, just as the perception of pitch is not necessarily directly related to Fo, so the perception of rhythm may not be directly related to true time intervals in speech. "

Several perceptual studies by Huggins (1972a, 1972b) can be interpreted as supporting the idea that isochrony is a perceptual phenomenon. Huggins studied the effects of durational changes in segments on the perceived "normality" of sentences. He found that there was a small negative correlation between changes involving two adjacent segments when the change to the second segment was in the direction towards re-establishing isochrony (restoring the onset time of the next stressed vowel to its former distance from the onset of the preceding stressed vowel).

Some data pointing towards the same conclusion were presented by Coleman in a dissertation written at the University of Washington (Coleman 1974). The dissertation constitutes a study of acoustic and perceptual attributes of isochrony in spoken English. Coleman investigated isochrony within sentences with controlled phonetic and grammatical context. Ten speakers read 16 real and 16 nonsense word samples placed in a carrier sentence. Each word contained two interstress intervals, and each interval contained one stressed syllable and from zero to three unstressed syllables. Forty listeners made perceptual judgments comparing the duration of the first and second interstress intervals in each sentence. The interstress interval increased significantly as syllables were added; but the listeners tended to hear the interstress intervals as more isochronous than they really were.

The most recent study dealing with the perception of isochrony is that of Donovan and Darwin (1979). In several experiments, Donovan and Darwin tested the listeners' ability to perceive the true rhythm of speech. For example, in one experiment subjects adjusted the times between four noise bursts to match the overall rhythm of either a sentence or a control sequence of nonspeech sounds. The matched durations were reliably more isochronous than the actual durations when the subjects matched sentences; no such tendency toward perceptual isochrony was found when the subjects matched the non-speech rhythm. Donovan and Darwin's conclusions state that their results have broadly confirmed Lehiste's proposal that isochrony is partly a perceptual phenomenon. They make two points in addition: first, that it is a perceptual phenomenon which is not independent of intonation, and second, that it is confined to language and reflects underlying processes in speech production. Donovan and Darwin question the value of seeking direct links between syntax and segmental durations rather than indirect ones via an overall rhythmic structure which is also determined by the pragmatic and semantic context of a sentence. (This last statement is very much in harmony with the view that I have been advocating in the whole sequence of studies which I am summarizing in this talk.)

I believe to have shown that there exists a tendency to hear spoken English as possessing a certain degree of isochronicity. First of all, many actual differences in the duration of interstress intervals may be below the

perceptual threshold. Second, listeners tend to impose a rhythmic structure on stretches of sounds and thus subjectively to perceive isochrony even in sequences where the durational differences should be above the perceptual threshold. There is nevertheless some evidence that speakers also have a tendency to aim at isochrony in production. This emerges from the way in which they treat durational constraints in production.

If speakers do indeed aim at isochrony in production, they have to make certain adjustments in the duration of speech sounds and their sequences. For this purpose, various well-known constraints on duration in production must be modified in the direction toward isochrony; such constraints that have not been modified must be compensated for in perception to arrive at perceived isochrony. There is some evidence for both processes, in spite of the counterevidence presented earlier (O'Connor 1958; Lea 1974).

The factors that constrain the duration of segments have been reviewed by Klatt in 1976. I shall therefore limit myself to a few examples. It is generally known that in English, the duration of vowels depends on the nature of the postvocalic consonant. The duration of consonants is affected by their membership in clusters. Besides segmental and suprasegmental factors, the duration of segments is influenced by their positions within a word. Earlier in this talk I discussed the characteristics of word-initial and word-final consonants; you may remember that word-initial consonants are characterized by greater length in English. Non-final segments in words of more than one syllable are shorter, the farther away they are from the end of the word: their duration appears to depend on the number of syllables that remain to be produced. Segments in word-final syllables, on the other hand, tend to be lengthened.

The influence of the position of the word within a sentence on the duration of segments and syllables has also been known for some time. I have already talked about pre-boundary lengthening--the fact that syllables and words receive extra length in final position within a linguistic unit.

Several of the listed factors seem to be relatable to isochrony through the intermediate stage of preserving the durations of words resp. adjusting the durations of different words so that their durations approach a common average. The shortening of consonants in clusters seems to be directed toward achieving this purpose. Another example of a similar phenomenon is the reduction in duration of monosyllabic stems when various suffixes are added at the end (cf. the examples presented earlier: speed-speedy-speeding-speedily, etc.). In each case, the duration of the word as a whole is changed less than it would have been if new segments had been added without adjustment in the duration of the segments already present.

In addition to examples of compensation in perception presented earlier (Barnwell, 1971; Huggins, 1972a, 1972b; Coleman, 1974; Donovan and Darwin, 1979), I would like to refer again to the results of my own study concerning the perception of the duration of sequences of four invervals. In that study the duration of the fourth interval was regularly underestimated: when all intervals had objectively equal durations, the fourth

was most frequently judged as "shortest". It seems that listeners expect the last interval (corresponding to the last word before a pause) to be longer than the other intervals, and if the extra length is not present, the listeners hear the interval as shorter than what they would normally expect.

Having now established that there exists a tendency toward isochrony in production as well as in perception, I would like to review the steps that have led me to the conclusion that isochrony is integrated into the syntax of English in at least one quite specific way.

My studies of the relationship between speech timing and syntax started with an explanation of the strategies speakers use for disambiguating syntactically ambiguous sentences. In a paper published in 1973 (Lehiste, 1973b) I reported the results of an investigation involving 15 ambiguous sentences, produced by four speakers, and listened to by 30 listeners. Some examples might be the following: "The hostess greeted the girl with a smile" (either the hostess smiled or the girl smiled); "The old men and women stayed at home" (either the men only or both men and women were old); "I know more beautiful women than Mary" (either I know women who are more beautiful than Mary, or I know more women who are beautiful than Mary does) etc.

The sentences were recorded by four speakers, two of whom were linguists, the other two non-linguists. The ambiguities were then pointed out, and the speakers were asked which of the possible meanings they had had in mind. The sentences were then re-recorded twice, the speaker making a conscious effort to convey one or the other meaning.

Slide 13 illustrates the material that was recorded and presented to listeners. The slide contains broad-band spectrograms of the sentence Steve or Sam and Bob will come. The letters B_o indicate that the original production was meant by the speaker to convey the meaning we had designated as meaning B: "Steve or (Sam + Bob)". Underneath the original sentence is the consciously disambiguated version of the same reading, followed by the consciously disambiguated version expressing the other possible reading (which we had designated as meaning A): "(Steve or Sam) + Bob".

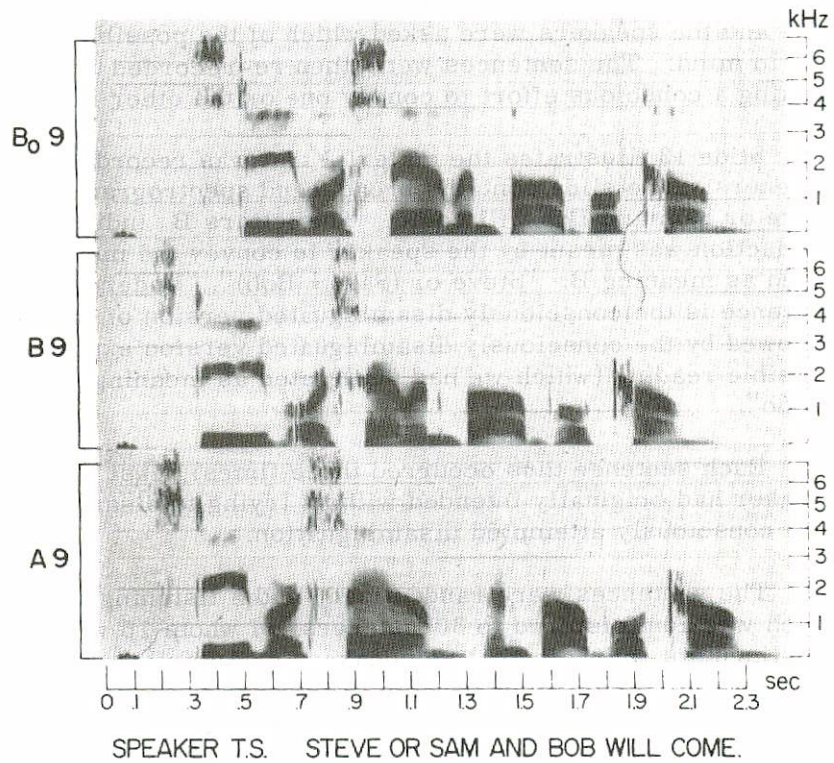
Each sentence thus occurred three times: first with the meaning the speaker had originally intended without trying to disambiguate, and then with consciously attempted disambiguation.

The sentences were randomized, and a listening tape was prepared which was administered to 30 listeners, of whom 15 were linguists and 15 non-linguists, all being native speakers of American English. The speakers and listeners were separated into two groups because of a haunting suspicion that disambiguation might be something only linguists do and other linguists perceive; in other words, that we might be dealing with an artifact introduced by the habit of linguists to expound fine points of syntactic theory by considering ambiguous sentences. In actual fact, it turned out that the two groups performed about equally. The overall correct score obtained

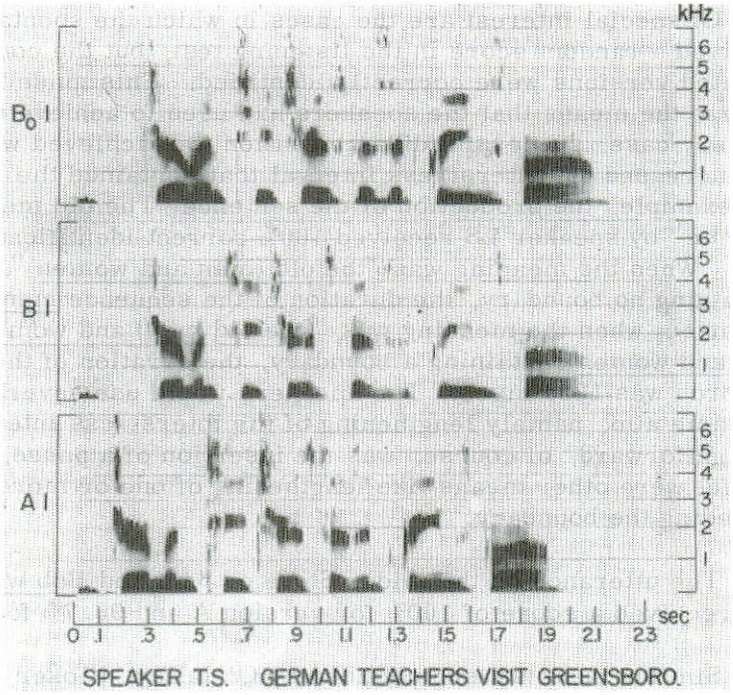
by listeners who were linguists was 62. 86%, while the non-linguists achieved a score of 61. 52%. The speakers serving as listeners did somewhat better than the whole group. There were also noticeable differences in the ability of speakers to produce successful disambiguations.

It turned out that not all sentences can be successfully disambiguated. For ten out of the fifteen sentences, the listeners performed at better than chance level. The set of successfully disambiguation sentences consisted of those sentences for which difference in meaning was correlated with a difference in sentence constituent structure (syntactic bracketing). The sentences that were generally not disambiguated have only one bracketing, although the constituents may bear different labels.

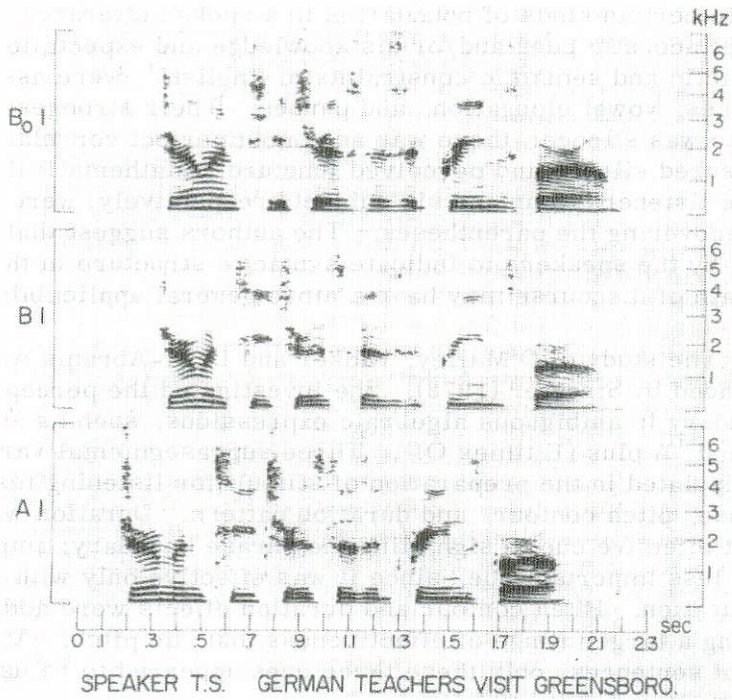
Slide 14 presents broad-band spectrograms of three versions of the sentence German teachers visit Greensboro. The disambiguation score for this sentence was 57. 9%. The sentence has only one surface bracketing, although it is ambiguous at a deeper syntactic level: the word German can either designate the object of the verb 'to teach' or it can constitute a modifier of the noun 'teacher'. Whatever phonetic differences are present in the three productions, they were insufficient for signalling this difference in the meaning. Slide 15 gives narrow-band spectrograms of the same three utterances.



Slide 13: Broad-band spectrograms of three versions of the sentence "Steve or Sam and Bob will come", produced by speaker TS



Slide 14: Broad-band spectrograms of three versions of the sentence "German teachers visit Greensboro", produced by speaker TS.



Slide 15: Narrow-band spectrograms of three versions of the sentence "German teachers visit Greensboro", produced by speaker TS.

Of special interest are the cases in which the spontaneous version received a random score in the listening test, but the consciously disambiguated versions were correctly identified. This made it possible to analyze the means that the speakers had used to achieve disambiguation. In every case, successful disambiguation was achieved when the speakers had increased the interstress interval that contained the relevant boundary. For example, the production of the sentence "The old men and women stayed at home" by speaker LS received 100% correct identification for both meanings. When the meaning was "the old (men and women)", (men and women) containing no boundary, the duration of the sequence men and women was 690 msec; when the meaning was "(the old men) and women", the sequence men and women containing a boundary, the duration of the same sequence of words was 1225 msec. The speakers would use several ways to achieve the same aim, namely lengthening of the interstress interval; the most straightforward, of course, was the insertion of a pause, but equally successful were other means like lengthening of one or more segmental sounds preceding the boundary.

The utterance on the slide, Steve or Sam and Bob will come by speaker TS, received a score of 100% for version A and 96.7% for version B.

Similar results were obtained by O'Malley, Kloker and Dara-Abrams (1973) in a study of parentheses in spoken algebraic expressions. They found that the speakers used "junctures" to indicate the presence of parentheses. The "junctures", which the authors define as "an abstract linguistic unit that is postulated to account for the ability of a native listener to locate certain kinds of boundaries in a spoken utterance on the basis of direct acoustic cues and/or his knowledge and expectations about the lexical, syntactic and semantic constraints of English", were associated with pitch changes, vowel elongation, and pauses. Their strongest perceptual correlate was silence: there was an almost perfect correlation between measured silence and perceived juncture. Mathematically experienced and naive listeners (four and six subjects respectively) were equally successful in recovering the parentheses. The authors suggest that the acoustic cues used by the speakers to indicate syntactic structure in this restricted domain of discourse may have a more general applicability.

The study of O'Malley, Kloker and Dara-Abrams was replicated and extended by Streeter (1978). She investigated the perception of phrase boundary in ambiguous algebraic expressions, such as "(A plus E) times O" and "A plus (E times O)". Three suprasegmental variables were manipulated in the preparation of stimuli for listening tests, namely amplitude, pitch contour, and duration pattern. Duration was found to be the most effective cue in signalling the phrase boundary; amplitude appeared to be a less important cue, since it was effective only with appropriate values of duration. Pitch contour and duration effects were additive, duration having a larger range of effectiveness than did pitch. At least for Streeter's set of sentences, only these three cues appeared to be used by listeners for phrase boundary perception.

Both O'Malley, Kloker and Dara-Abrams and Streeter located their boundary signals at given points in a linear sequence, without relating them to the general rhythmic structure of the utterances. Pauses were the primary cue in naturally produced sentences; in synthesized sentences, duration was the most effective cue. While preboundary lengthening is certainly frequently present before a syntactic boundary, it is not an absolutely necessary cue. In a study jointly carried through with Olive and Streeter (Lehiste, Olive and Streeter, 1976) I showed that increase in the interstress interval is a sufficient boundary signal, even in the absence of intonation and specific segmental lengthening. We processed ten of the sentences used in my 1973 study (those that had been successfully disambiguated) through an analysis-resynthesis program, changed fundamental frequency to monotone, and manipulated systematically the duration of interstress intervals. A listening test, similar to the one I had used earlier, was given to thirty subjects. Disambiguation was achieved when the relevant interval reached a certain duration, the actual value of which depended on the particular sentence. I would like to emphasize here that we did not insert any pauses; neither did we introduce prepausal lengthening. The interstress interval was increased by increasing the duration of each sampling period by the same factor; the durational relationships of the segments to each other remained unchanged. Thus the disambiguation was produced solely by increasing the interstress interval, and the results of that study show that this is indeed a sufficient cue for signalling the presence of a boundary.

I have now come to the point where I can draw the conclusion that connects the rhythm and the syntax of an English sentence. Increase of an interstress interval can be used to signal a syntactic boundary precisely because this increase constitutes a deviation from the expected rhythmic pattern. The listener expects isochrony--expects the stresses to follow each other at approximately equal intervals. A deviation from the pattern--namely an increase in the interstress interval large enough to be perceived--could not signal the presence of a boundary unless the pattern exists in the first place. In principle, of course, a deviation from the pattern could be used to signal anything. In English, it appears to be part of the knowledge of both speakers and hearers that an increase in the interstress interval signals the presence of a syntactic boundary. It is in this sense that isochrony is integrated into the grammar of English at the syntactic level.

I have talked very little about the ways in which intonation and stress function in English to manifest phonetically the syntactic structure of a sentence. There are two reasons for this. First, I have myself done very little in the study of English intonation and have nothing original to contribute. Second, I believe that intonation and stress play a relatively minor role in signalling syntactic structure. Their role is primarily semantic and pragmatic. Let me explain.

Intonation is used in English to signal, first of all, whether an utterance is complete or incomplete. An incomplete utterance terminates in sustained level pitch--pitch that neither rises nor falls. Falling terminal intonation signals the completion of a declarative sentence; rising terminal

intonation signals the completion of a certain type of interrogative sentence (a yes-or-no question). All other uses of intonation carry non-grammatical information. They may signal the difference between old and new information, they may convey something about the attitude of the speaker toward the content of the utterance and so forth; but intonation changes do not change a subject into an object. Stress singles out some constituent of a sentence for essentially semantic reasons--to express a judgment about what is important to the speaker, to indicate that the item is being compared with something else, etc.; but a subject that is emphatically stressed remains, grammatically, a subject. In an unpublished study, Olive and I tried to use intonational changes to disambiguate the same sentences that we had successfully disambiguated by changing the duration of interstress intervals. Listener responses remained random, as far as the syntax was concerned. They reported informally that they heard changes in the degree of emphasis on a particular constituent, but the syntactic role of that constituent remained the same.

The conclusion then is the following. The syntactic structure of a sentence is expressed phonetically primarily through the controlled timing of articulatory gestures that speakers use when they produce spoken utterances. Speech is a rhythmic activity, as are most motor activities performed by human beings. In English, the controlled timing of articulatory gestures takes the rhythmic structure of speech into account. Stressed syllables carry the greatest amount of information; attention has to be focused on the stressed syllables, and this is facilitated by setting up an expectation as to when the next stressed syllable is likely to occur. Producing sentences in such a way that stressed syllables occur at regular, isochronous intervals contributes to optimal perception by the listeners whose attention is cyclically directed to the points in time at which the stressed syllables can be found. Furthermore, a disruption of the expected pattern can be used to convey crucial information about syntactic structure. The syntactic structure of an English sentence is thus primarily manifested in the timing pattern of that sentence when produced orally by a native speaker of the language.

Bibliography

Abercrombie, D. (1964). "Syllable quantity and enclitics in English." In: In Honour of Daniel Jones, ed. by D. Abercrombie, D. B. Fry, P. A. D. MacCarthy, N. C. Scott, and J. L. M. Trimm. Longmans, London, pp. 216-222.

Allen, George D. (1972). "The location of rhythmic stress beats in English: an experimental study II." Language and Speech 15, pp. 179-195.

Allen, George D. (1973). "Segmental timing control in speech production." Journal of Phonetics 1, pp. 219-237.

Allen, George D. (1975). "Speech rhythm: its relation to performance universals and articulatory timing." Journal of Phonetics 3, pp. 75-86.

Barnwell, T. P. (1971) "An algorithm for segment durations in reading machine context." Technical Report 479, Cambridge, Mass.: M.I.T. Research Laboratory of Electronics.

Bolinger, D. L. (1965). "Pitch accent and sentence rhythm." In: Forms of English: Accent, Morpheme, Order. p. 163 ff. Cambridge, Mass.: Harvard University Press.

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

Coleman, Colette L. (1974). A Study of Acoustical and Perceptual Attributes of Isochrony in Spoken English. Ph. D. dissertation, University of Washington.

Cooper, William E. (1975). Syntactic Control of Timing in Speech Production. Ph. D. dissertation, M.I.T., Cambridge, Mass.

Cooper, William E. (1976). "The syntactic control of timing in speech production: a study of complement clauses." Journal of Phonetics 4, pp. 151-171.

Cooper, William E., S. G. Lapointe, and J. M. Paccia. (1977). "Syntactic blocking of phonological rules in speech production." Journal of the Acoustical Society of America 61, pp. 1314-1320.

Cooper, William E., J. M. Sorensen, and J. M. Paccia. (1977). "Correlations of duration for nonadjacent segments in speech: Aspects of grammatical coding." Journal of the Acoustical Society of America 61, pp. 1046-1050.

Cooper, William E., J. M. Paccia, and S. G. Lapointe. (1978). "Hierarchical coding in speech timing." Cognitive Psychology 10, pp. 154-177.

Donovan, A., and C. J. Darwin. (1979). "The perceived rhythm of speech." Proceedings of the Ninth International Congress of Phonetic Sciences, Copenhagen, Volume II, pp. 268-274.

Gaitenby, Jane. (1965). "The elastic word." Status Report on Speech Research SR-2, Haskins Laboratories, New York, pp. 3.1-3.12.

Halliday, M. A. K. (1967). Intonation and Grammar in British English. The Hague, Mouton.

Huggins, A. W. F. (1972a). "Just noticeable differences for segment duration in natural speech." Journal of the Acoustical Society of America 51, pp. 1270-1278.

Huggins, A. W. F. (1972b). "On the perception of temporal phenomena in speech." Journal of the Acoustical Society of America 51, pp. 1279-1290.

Klatt, D. H. (1973). "Interaction between two factors that influence vowel duration." Journal of the Acoustical Society of America 54, pp. 1102-1104.

Klatt, D. H. (1975). "Vowel lengthening is syntactically determined in connected discourse." Journal of Phonetics 3, pp. 129-140.

Klatt, D. H. (1976). "Linguistic uses of segmental duration in English: acoustic and perceptual evidence." Journal of the Acoustical Society of America 59, pp. 1208-1221.

Lea, W. A. (1974). Prosodic aids to speech recognition: IV. A general strategy for prosodically-guided speech understanding. Univac Report No. PX10791. St. Paul, Minn.: Sperry Univac, DSD.

Lehiste, Ilse. (1960). An Acoustic-Phonetic Study of Internal Open Juncture. Phonetics 5 (Supplement), pp. 1-54.
Lihiste, Ilse. (1964). Acoustical Characteristics of Selected English Consonants, Indiana University, Bloomington, and Mouton & Co., The Hague.
Lehiste, Ilse. (1970). Suprasegmentals. Cambridge, Mass.: M. I. T. Press.
Lehiste, Ilse. (1971). "Temporal organization of spoken language." In: Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen. L. L. Hammerich, Roman Jakobson, and Eberhard Zwirner, Eds., Akademisk Forlag: Copenhagen, pp. 159-169.
Lehiste, Ilse. (1972). "The timing of utterances and linguistic boundaries." Journal of the Acoustical Society of America 51, pp. 2018-2024.
Lehiste, Ilse. (1973a). "Rhythmic units and syntactic units in production and perception." Journal of the Acoustical Society of America 54, pp. 1228-1234.
Lehiste, Ilse. (1973b). "Phonetic disambiguation of syntactic ambiguity." Glossa 5, pp. 107-122.
Lehiste, Ilse. (1975a). "The role of temporal factors in the establishment of linguistic units and boundaries." In: Phonologica 1972. Wolfgang U. Dressler and F. V. Mares, Eds., München -Salzburg: Wilhelm Fink Verlag. pp. 115-122.
Lehiste, Ilse. (1975b). "Some factors affecting the duration of syllable nuclei in English." Salzburger Beiträge zur Linguistik 1, pp. 81-104.
Lehiste, Ilse. (1977). "Isochrony reconsidered." Journal of Phonetics 5, pp. 253-263.
Lehiste, Ilse. (1979). "The perception of duration within sequences of four intervals." Journal of Phonetics 7, pp. 313-316.
Lehiste, Ilse, J. P. Olive, and L. A. Streeter (1976). "The role of duration in disambiguating syntactically ambiguous sentences." Journal of the Acoustical Society of America 60, pp. 1199-1202.
Lindblom, Björn, and Karin Rapp. (1973). "Some temporal regularities of Spoken Swedish." Papers from the Institute of Linguistics 21, University of Stockholm.
Nooteboom, S. G. (1972). Production and Perception of Vowel Duration. Dissertation Utrecht.
O'Connor, J. D. (1965). "The perception of time intervals." Progress Report 2, Phonetics Laboratory, University College, Longon, pp. 11-15.
O'Connor, J. D. (1968). "The duration of the foot in relation to the number of component sound-segments." Progress Report 3, Phonetics Laboratory, University College, London, pp. 1-6.
Oller, D. K. (1973). "The effect of position in utterance on speech-segment duration in English." Journal of the Acoustical Society of America 54, pp. 1235-1247.
O'Malley, M. H., D. R. Kloker, and B. Dara-Abrams. (1973). "Recovering parentheses from spoken algebraic expressions." IEEE Transactions on Audio and Electro-Acoustics AU-21, pp. 217-220.
Pike, K. L. (1945). The Intonation of American English. Ann Arbor, Mich.: University of Michigan Press.

Rees, Martin. (1975). "The domain of isochrony." Work in Progress 8, Edinburgh University, Department of Linguistics, pp. 14-28.

Streeter, Lynn A. (1978). "Acoustic Determinants of phrase boundary perception." Journal of the Acoustical Society of America 64, pp. 1582-1592.

Uldall, E. T. (1971). "Isochronous stresses in R. P." In: Form and Substance: Phonetic and Linguistic Papers Presented to Eli Fischer-Jørgensen. L. L. Hammerich, Roman Jakobson, and Eberhard Zwirner, Eds., Akademisk Forlag: Copenhagen, pp. 205-210.

Uldall, E. T. (1972). "Relative durations of syllables in two-syllable rhythmic feet in R. P. in connected speech." Work in Progress 5, Edinburgh University, Department of Linguistics, pp. 110-111.