PITCH DECLINATION IN JAPANESE: AN INSTRUMENTAL STUDY

by

NEIL GALLAIFORD

B.Ed., University of Alberta, 1979

A THESIS SUBMITTED IN PARTIAL FULFILMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department of

Linguistics

We accept this thesis as conforming to the required standard

[redacted]

Professor Henry J. Warkentyne

[redacted]

Professor Gunter H. Schaarschmidt

[redacted]

Professor Thomas E. Hukari

[redacted]

Professor Paul M. Baker

UNIVERSITY OF VICTORIA

April, 1983

ABSTRACT

Supervisor: Professor Henry J. Warkentyne

Pitch declination involves a progressive lowering of pitch and a narrowing of pitch range over the length of utterance in speech. Because this pattern is attested in a great number of diverse languages, it has been proposed as a universal feature of intonation. The present study was undertaken to discover if pitch declination is a feature of the intonation of Japanese.

Previous, non-acoustic, studies of Japanese have suggested that pitch does tend to decline in longer phrases in the language. Using a method similar to that employed by Sorensen and Cooper (1980, 1981) in their work on pitch declination in English, an acoustic analysis of neutral declarative sentences spoken by native speakers was carried out. After extracting the fundamental frequency (F0) contours of the utterances, the F0 peaks associated with the lexical items in each sentence were identified and compared. It was found that the fundamental frequency of these peaks tends to decrease over the length of the sentence. Japanese and English are similar in this respect. The tendency of pitch to decline over the course of the shorter sentences in the data was strong; longer sentences tended to contain a pitch rise somewhere within them. On either side of the rise, pitch declined from peak to peak. This suggests that

the domain of pitch declination is shorter in Japanese than it is in English as described by Sorensen and Cooper. Longer sentences are broken up into two (or more) 'intonation phrases' each of which is characterized by declining pitch. These intonation phrases do not seem to bear any necessary relationship to the syntactic structure of the sentences. The acoustic-phonetic data presented here provide a better basis for the description of the behaviour of pitch in Japanese utterances than do the auditory analyses on which previous studies have been based.

[redacted]

Professor Henry J. Warkentyne

[redacted]

Professor Gunter H. Schaarschmidt

[redacted]

Professor Thomas E. Hukari

[redacted]

Professor Paul M. Baker

TABLE OF CONTENTS

APPENDICES

LIST OF TABLES

LIST OF FIGURES

DEDICATION

This thesis is gratefully dedicated to Gina M. Benvie without whose encouragement and unfailing support it could not have been written.

Chapter I
INTRODUCTION

This study has two objectives. On the basis of instrumental measurement of the pitch contours of a corpus of Japanese sentences, it undertakes to formulate a linguistic analysis of some aspects of the intonation of Japanese. It also attempts to show that the intonation of Japanese sentences can best be described in terms of a pitch pattern which has wide currency in the world's languages. Pitch declination, also known variously as downdrift, downstep and the declination effect, refers to a gradual lowering of pitch and a narrowing of pitch range from the beginning to the end of an utterance (Pierrehumbert 1980:116). This pattern has been found in languages as diverse as Dutch (Cohen and 't Hart 1967), Danish (Thorsen 1980), Serbo-Croatian (Lehiste and Ivić 1978), French, Finnish and English (Pierrehumbert 1980). In some of the tone languages of Africa, pitch declination is so marked that High tones near the end of a sentence can actually be phonetically lower than Low tones near the beginning (Gleason 1961:298). It has been suggested that pitch declination is so widespread as to constitute a universal feature of intonation, and hence, of language (Bolinger 1964). If it can be shown that pitch declination is the characteristic unmarked intonation pattern for Japanese, this will lend strength to the claim for its universality.

The analysis of Japanese intonation presented in this thesis is based on acoustic-phonetic data obtained through instrumental measurement of the fundamental frequency contours of neutral declarative sentences produced by native speakers of standard colloquial Japanese. Fundamental frequency (F0) is the primary acoustic correlate of pitch[1] (Lehiste 1970:54). Intonation is a term whose definition ranges from the very broad--Intonation is the melody of speech--to the narrow--Intonation is systematic variation in the fundamental frequency of speech which cannot be accounted for in terms of the word level phenomena of tone or accent alone and which is not the result of segmental phonetic factors. In later sections of this thesis we shall discuss the form, function and domain of pitch declination as an intonational feature of language and of Japanese in particular. For these purposes, the narrow definition of intonation above is the more relevant.

Japanese lexical items are defined in part by a pattern of High and Low tones derived from accentual information contained in the lexicon. In order to understand how High and Low tones are mapped onto actual fundamental frequency

[1] The terms pitch and fundamental frequency can and should be distinguished in certain contexts, but, in general, they can be used interchangeably with each implying the other and they will be so used here.

values in speech, we must be able to separate the F0 variations in the acoustic signal which are the consequence of word level accent from the variations which are due to intonation. The identification of intonational pitch patterns is as essential to the hearer as it is to the linguist who is trying to model the production/perception processes involved in the coding/decoding of tonal features.

Consider for a moment the task of the hearer in recovering tonal values from the acoustic speech signal, primarily from fundamental frequency. In a language which lacked intonation, the recovery of tonal values, whether resulting from tone or word accent, would be a relatively simple task. There would be some localized variation in fundamental frequency due to segmental effects, but tonal values could be coded in a one to one relationship with clearly delineated pitch ranges. The phonological description of the tonal features of such a language would quite closely match the phonetic description of actual utterances. Categorical distinctions between accented and unaccented syllables realized as a sequence of High and Low tones would be matched in speech by fundamental frequency values occupying the upper and lower portions of the total pitch range of the speaker. If prominence or sentence accent were present in the language, a three-way distinction in tonal values could be used, High and Low for word accent, Extra-High for prominence. The speaker's pitch range would then be simply divided into three subranges corresponding to the three tones, one-third to each. In a language with no intonation, many more categorical distinctions in tone or accent, and rather fine nuances of prominence relations could easily be coded into F0; the human ear is after all a very sensitive instrument and the brain is capable of very fine discriminations in pitch (on the order of plus or minus 1 Hertz, Lehiste 1970). Linguists studying tone languages have found, however, that the number of categorical distinctions utilized in even the most elaborate tone systems represents a tiny fraction of the number possible in terms of the human capacity for distinguishing among vocal pitch values (Anderson 1978:146). The reason for this is obvious: languages make use of pitch for other than tonal distinctions.

All languages have intonation and to borrow Pierrehumbert's (1980:117) analogy, intonation has the effect of continuously varying the graph paper on which tonal values are mapped as fundamental frequency. Under this condition, it is easy to see that the hearer's capacity for recovering a large number of tonal distinctions from fundamental frequency is considerably reduced. If intonation were entirely random (which, of course, it is not) then the task of the hearer in recovering tonal values would be an almost impossible one. Only to the extent that intonation is systematic, that it consists of recognizable patterns which can be factored out by the hearer, can tonal values and prominence relations be recovered from the stream of speech. Pitch declination is one of the more common of these patterns. In Chapter 2, the possible physiological basis for the prevalence of this pattern will be explored. Chapter 3 consists of a discussion of the linguistic role that pitch declination might play in language and its relationship with syntax. Our understanding of this relationship is bound up with the question of what the domain of pitch declination is, and given its domain, what its form is. These three aspects of the phenomenon, its form, its domain and its function must be kept in mind as we look at the non-acoustic and acoustic evidence (in Chapters 4 and 5) for pitch declination being the basic intonation pattern of Japanese.

Chapter II
THE PHYSIOLOGICAL BASIS FOR PITCH DECLINATION

2.1 PHYSIOLOGY AND UNIVERSALITY

In his discussion of intonation as a universal of human language, Bolinger (1964) rightly states that for intonation to be considered as universal there must be meaningful uses of pitch that are so widely shared as to suggest a common origin. 'Common origin' might refer to 'a single primordial language or a common denominator in human nature that could give rise to similarities [p 833].' While any evidence for a single primordial language is probably impossibly obscured by the passage of time, 'human nature' (and by this we take Bolinger to mean the underlying physiological and psychological characteristics of the organism) is more open to investigation.

What physiological and psychological explanations can be found for the pattern of pitch declination? Bolinger suggests that 'the only indefectibly universal aspect of fundamental pitch is its association with the muscular tension of the whole organism [p 843].' Rises in pitch are associated with increased tension, as for instance at the beginning of utterances, and falls in pitch with relaxation both physiological and psychological, as at the end of utterances. When rises and falls in pitch are used intonationally in language they involve a 'simulation' of tension and relaxation which serves a linguistic purpose. The idea that pitch patterns in language involve simulation of actual psychophysiological states of tension/relaxation removes the necessity for us to find pitch patterns which are completely alike in detail across languages--languages may differ in the way they exploit the tension/relaxation dichotomy. Whatever the psychophysical basis we might claim for intonation it is the way that it is used linguistically in languages that is our concern; if we can find widespread patterns which have plausible physiological explanations so much the better.

2.2 THE ARCHETYPAL NORMAL BREATH GROUP

Lieberman (1967, 1980) proposes that pitch declination is fundamentally related to the human breathing cycle. Speech is accomplished during the expiratory phase of the cycle: air flows out of the lungs through the glottis and phonation is a consequence of the opening and closing of the glottis under conditions of positive air flow from the lungs. The fundamental frequency of phonation is directly proportional, *ceteris paribus*, to subglottal air pressure. In the absence of counteracting articulatory gestures such as increased tension in the vocal cords, as subglottal air pressure gradually falls during expiration the fundamental frequency of phonation also falls. Appealing to the principle of least articulatory effort, Lieberman maintains that in the 'archetypal normal breath-group', which roughly corresponds to the speech uttered during a single expiration, the vocal cords will be held at the same tension throughout the entire expiration and thus the normal breath-group will be characterized by a fall in pitch attributable to the falling subglottal pressure. Lieberman allows, however, that the acoustic pattern which is to be associated with the 'normal breath-group' can be the result of other articulatory gestures on the part of the speaker. In particular the speaker can relax the vocal cords while maintaining a relatively constant subglottal pressure. Still, Lieberman claims that hearers 'perceive intonational signals as though they had been produced by means of the primary or archetypal articulatory gestures [1967:27].' These gestures are the ones that human infants utilize in their cries. Infant cries are characterized by a rapid increase in subglottal pressure initially and then a gradual fall towards the end. The pitch contour of these cries closely follows the subglottal air pressure contour. The innate referential breath-group of the infant thus 'furnishes the basis for the universal acoustic properties of the normal breath-group that is used to segment speech into sentences in so many languages [1967:42].'

Ohala (1978) cites evidence that pitch declination cannot be attributed simply to falling subglottal pressure during expiration. He observes that the magnitude of pitch declination as it is found in speech, 'usually exceeds what could be accounted for by [subglottal pressure] variation [p 32].' Ohala believes instead that the effect is due to active laryngeally caused changes in vocal cord tension. Lieberman allowed that speakers could use different sets of articulatory gestures (including changes in vocal cord tension) to simulate the archetypal pattern characterized by falling subglottal pressure and decline in pitch. We can speculate that what begins as an innate physiologically determined pattern in the cries of infants is exploited linguistically and by means of other articulatory gestures in adult languages.

2.3 THE PERCEPTUAL CONSEQUENCES OF PITCH DECLINATION

Whatever the physiological basis for pitch declination, if any, there is some evidence that it has a perceptual consequence. In experiments reported by Pierrehumbert (1979), English-speaking subjects were asked to judge which of two stressed syllables in a nonsense utterance was higher in pitch. It was found that in making this judgement, the subjects corrected for an expected declination in pitch. They judged the two peaks to be equal even though the second peak was actually lower in pitch than the first. The subjects' perceptions were also affected by the overall pitch range of the sample and by the length of the utterance.

This suggests that there is a normalizing process involved in the perception of pitch in utterances. A peak does not have to be as high later in a sentence in order to express a given degree of prominence since the hearer corrects for a certain decline in pitch based on the overall pitch range and length of the utterance. Pierrehumbert later applied a formulation of the declination effect to the synthesis of neutral English declarative intonation with very good results (Pierrehumbert 1981).

Chapter III

THE LINGUISTIC ROLE OF PITCH DECLINATION

3.1 THE PHRASING FUNCTION

All the research discussed above focussed on the possible physiological mechanism(s) responsible for pitch declination and on its perceptual consequences. What linguistic role does pitch declination play in language? In dismissing Lieberman's claim that pitch declination is an automatic feature of speech, Ohala states that even if pitch declination were not purposeful to start out with, speakers would soon enough make it so since it is so useful in signalling clause and sentence boundaries (Ohala 1978:32). It seems likely that intonation's primary linguistic function is in signalling the organization of utterances into phrases, clauses, sentences, and the like. Intonation not only marks the boundaries between phrases, clauses, etc., it also signals to what extent the phrases are related to each other and what element within the phrase is the center of attention (Gardiner 1977, cited in Ladd 1980:163). Pitch declination would be a useful mechanism whereby intonation could perform this phrasing function. If the phrases of a language were characterized by a gradual decline in pitch, breaks in the line of declination would signal the presence of a boundary between phrases. Pause, the other primary cue to phrasing, can be dispensed with to the extent that the intonation pattern which characterizes phrases is recognized by the hearer (Bolinger 1978).

3.2 PITCH DECLINATION AND SYNTAX

Sorensen and Cooper (1980, Cooper and Sorensen 1977) investigated the relationship between pitch declination (which they refer to as F0 declination) and syntactic boundaries in English. They formulate a pitch declination rule (the Topline Rule) which predicts the pitch of intermediate F0 peaks given their location in the sentence and the pitch of the first and last peaks.[2] Their Topline Rule (1980) applies to utterances characterized by a single line of F0 declination; it predicts the pitch values of intermediate peaks with an accuracy of better than plus or minus 5% (Sorenson and Cooper 1980:415). Sorensen and Cooper found that, while the F0 of the final peak was almost invariant, the value of the first peak and consequently the slope of the line of F0 declination depended on the length of the utterance; the initial peak was higher in longer utterances. This suggests that a certain amount of preplanning on the

---

[2] Not all the F0 peaks in the utterance are on the Topline. On a graph of time versus fundamental frequency, the Topline is a line whose reference points are the last F0 peak in the utterance and a 'key point' which occurs halfway between the first and last peak on the frequency axis and one quarter of the way between the first and last peak on the time axis. The first peak, therefore, is not on but above the Topline. (Sorensen and Cooper 1980:410)

part of the speaker is necessary for the coding of fundamental frequency contours in speech. Having discovered the form which F0 declination takes in English, Sorensen and Cooper address the issue of whether the domain of F0 declination is syntactically defined.

In a sentence consisting of more than one clause there are a number of ways in which pitch declination could be manifested. If we refer to the domain of F0 declination as the 'intonation phrase', leaving unspecified the possible syntactic characteristics of such a phrase, a sentence containing more than one clause could constitute a single intonation phrase or be made up of two or more intonation phrases.[3] If all the intermediate F0 peaks of a sentence fall on a line of declination defined by the Topline Rule with reference to the first and last peaks in the sentence, then the sentence must constitute a single intonation phrase. If each clause in the sentence is an intonation phrase, i.e. if the syntactic domain of F0 declination is the clause, then each clause would be characterized by a line of declination defined with reference to the first and last peaks in each clause. If the clauses are of equal length and the clause-initial and clause-final F0 peaks of each are at the same frequency, the slope of the lines of

---

[3] A common English example of the difference might be
a) We'll go on a picnic tomorrow if the weather holds out.
b) If the weather holds out, we'll go on a picnic tomorrow.
The a) sentence is often only one intonation phrase while the b) sentence will often have an intonation phrase boundary after 'holds out'.

declination should be identical. This situation would involve a complete resetting of F0 declination at the clause boundary. Partial resetting might also be possible. Under conditions of partial resetting, each clause would be characterized by F0 declination, but the initial F0 peaks and the slope of the lines of declination would not be the same. In utterances made up of more than one intonation phrase, relationships among the phrases could be signalled by the amount of resetting of F0 declination at the boundaries.

In their analysis of fall-rise pitch patterns at syntactic boundaries, Cooper and Sorensen (1977) found that there were significant differences in these patterns depending on the type of syntactic boundary. The sites for the peak and valley measurements which defined the fall-rise pattern were carefully matched for phonetic content.[4] They found that the fall-rise pattern from the last peak in the first clause to the initial peak of the second was more pronounced across a main clause-main clause boundary than across a main clause-embedded clause boundary. The data did not support a hypothesis of complete resetting of the declination function in either case since the initial peaks in the first clause

---

[4] Example (Cooper and Sorensen 1977:685)

a) Anthony was sur<u>prised</u> and <u>Ray</u>mond became upset.
(Main clause-main clause)

b) Anthony was sur<u>prised</u> An<u>drea</u> became upset.
(Main clause-embedded clause)

The underscored syllables contained the F0 peaks.

were consistently higher than the initial peaks of the second clause. Cooper and Sorensen do not rule out the possibility of partial resetting but their experiment did not test for it. They do conclude, on the basis of this and other experiments, that the magnitude of the fall-rise patterns at syntactic boundaries, whether they involve purely local effects or (partial) declination resetting, should provide a good measure of boundary strength (Sorensen and Cooper 1980:425).

The fact that (partial) declination resetting may occur at syntactic boundaries does not necessarily imply that intonation phrases have a syntactic domain. Much as pauses occur other than at syntactic constituent boundaries so are intonation phrases not necessarily co-extensive with syntactic units. Intonation and syntax perform separate functions. Intonation has to do with the organization of utterances, syntax with the organization of sentences; intonation performs a communicative, syntax a grammatical function (cf. Daneš 1960). It is true that utterances are often sentences, and so we might speak of a 'sentence' intonation; it is true that intonation phrases are sometimes co-extensive with clauses, but the communicative function performed by intonation might demand otherwise and so intonation phrase boundaries may occur anywhere in an utterance--an intonation phrase may constitute part of a sentence or, less often, of more than one sentence may be included in a single intonation phrase. Gardiner (1980:4) puts the situation most clearly:

> ...it cannot be said that there is any necessary relation between the rules of syntax or grammar and the structure (rules) of intonation...whatever [the intonation marks] as separate phrases are separate phrases, no matter whether the words are organized in syntactically well-formed phrases [or not].

Of course clauses and such can be communicative units as well as syntactic ones and thus they can be marked by phrase intonation and be bounded by intonation phrase boundaries.

As to intonation being a good predictor of syntactic boundary strength, recall that one of the functions of intonation is to indicate the relationship between phrases in an utterance. An intonation phrase is identified by a recognizable pitch contour--the relationship among intonation phrases can be indicated by the relative pitch level of these contours. Under these circumstances, there may very well be some correlation between the strength of syntactic boundaries and the 'strength' of intonation phrase boundaries which co-occur with them. 'Strength of intonation phrase boundary' should be interpreted as being a reflex of the prominence relations between intonation phrases. Sorensen and Cooper found that the magnitude of a fall-rise pattern in F0 at a main clause-main clause boundary was greater than at a main clause-embedded clause boundary. If each clause were an intonation phrase we might see this correlation as an indication of the relative prominence of the two phrases: two main clauses may be more equal in prominence while an embedded clause may tend to take less prominence.

Sorensen and Cooper's main contribution to our understanding of pitch declination is their specification of the form of F0 declination in English. Their method of comparing F0 peaks will be applied to the Japanese sentences under analysis in Chapter 5. As we shall see, intonation phrases in Japanese do not have a syntactic domain, they are not determined by the syntax. Intonation phrase boundaries in Japanese often delineate phrases which have no definable syntactic organization, a finding consistent with Gardiner's view reported above. In addition, there is some evidence that prominence relations among Japanese phrases are signalled by the degree of pitch change at phrase boundaries.

Chapter IV

NON-ACOUSTIC EVIDENCE FOR PITCH DECLINATION IN JAPANESE

Before proceeding to the instrumental analysis which is at the heart of this study, it will be worthwhile to review the literature on pitch and accent in Japanese. Most of the authors of the studies reported below did not have access to acoustic-phonetic information about the behaviour of pitch in Japanese utterances; the studies were based on auditory impression. We know from Pierrehumbert's experiments that there is a normalizing process which is involved in the perception of pitch in speech--the hearer 'corrects' for an expected decline in pitch. It would not be surprizing then if we failed to find evidence for pitch declination in the non-acoustic literature. Certainly, with the exception of Haraguchi (1977), we find no explicit reference to it. There are, however, many indications in this body of work that something like pitch declination does exist in Japanese.

4.1 THE JAPANESE PITCH-ACCENT SYSTEM

Since intonation and accent both contribute to the ultimate pitch patterns of Japanese utterances, it is essential that we be able to account for the pitch variations which are the result of accent alone. The domain of accent seems to be what McCawley (1968) calls the 'minor phrase'. A minor phrase is relatively short and consists typically of one content word plus its encliticized postpositions, if any. Sometimes minor phrases involve compounds which in many cases have resisted elegant analysis (McCawley 1968:130). By and large, however, the accent patterns of minor phrases can be accounted for by a few fairly simple rules.

Every minor phrase in Japanese has a characteristic pitch pattern which is derived from its accentuation. A minor phrase contains at most one accent which is signalled by a precipitous fall in pitch after the first mora[5] of the accented syllable. When lexically accented forms are combined in minor phrases there are accent manipulation rules which, among other things, ensure that all accents but one will be suppressed. Pitch assignment rules,[6] which operate

---

[5] Morae are the rhythmic units of Japanese. McCawley (1978:114), somewhat facetiously, suggests that a mora is something a long syllable has two of while a short syllable has only one. The morae of Japanese can be a single vowel, a consonant plus a vowel, or a consonant plus /y/ plus a vowel, a syllabic nasal, or a non-nasal syllable final consonant (Jorden 1955:3fn). Thus Tokyo /tookyoo/ has two syllables but four morae.

[6] It would probably be better to call them tone assignment

after all accent manipulation rules have applied, will assign High and Low pitches to the morae of the minor phrases in the following way. All morae up to and including the first mora of the accented syllable will be High and everything thereafter will be Low except that the first mora of the phrase will be Low unless it is itself accented (McCawley 1968: 137). Unaccented minor phrases will therefore have a Low first mora followed by a sequence of High morae up to the end of the phrase.

In isolation, minor phrases have pitch patterns which are maximally determined by the tones assigned to their constituent morae. There will be a rise in pitch from the first to second morae of the phrase if the first syllable is not accented; there will be a precipitous fall in pitch after the first mora of the accented syllable. Consider the following otherwise homophonous forms (the * marks the accented syllable):

ha* si ga 'chopsticks + Subj marker'

ha si* ga 'bridge + Subj marker'

ha si ga 'edge + Subj marker'

The first minor phrase is accented on the first syllable and so will have the tonal pattern High-Low-Low. The second is accented on the second syllable and so will have the pattern

---

rules since intonation will contribute to the ultimate realization of pitch in utterances.

Low-High-Low. The third is unaccented and so will have the pattern Low-High-High. The Subject marker ga is lexically unaccented and receives its tone on the basis of the accent of hasi. Without the ga, the second and third minor phrases above are homophonous in all respects. Except for the fact that McCawley makes no reference to intonational effects on pitch assignment and so assigns pitches to morae on the basis of accent pattern alone his description of the accent patterns of minor phrases is adequate and is supported in the main by acoustic studies of Japanese pitch-accent such as Weitzman's (1970).

4.2 PITCH IN LONGER UTTERANCES

McCawley finds it necessary to revise his rules somewhat because 'there are actually two kinds of phrase, "major phrases" and "minor phrases" [137].' Each of the sentences in Appendix D, for example, are major phrases made up of three minor phrases. It is in McCawley's discussion of the pitch shape of major phrases[7] that we find an indication that pitch declination might apply to utterances in Japanese. McCawley notes that when minor phrases combine into major phrases only the first minor phrase will contain

[7] Weitzman's acoustic study of word accent in Japanese reveals in fact that pitch declines within minor phrases (when they are uttered in isolation), such that sequences of Highs in minor phrases are realized with gradually declining pitch. The sequence Low-High-Low is distinguished from the sequence Low-High-High by a precipitous fall in pitch between the second and third morae (Weitzman 1970).

high pitches, 'subsequent minor phrases will only go up to mid pitch [138]'. This decline in pitch could perhaps be seen to be a consequence of pitch declination except that McCawley explicitly states that the mid pitch occurs on the second and <u>subsequent</u> minor phrases in a major phrase. Assuming that the non-low pitches in non-initial minor phrases do not differ from each other in the way that the first and second do, i.e. the mid pitches are similar, (an assumption that we cannot help making given McCawley's description) then pitch declination cannot be the reason for this phenomenon. If pitch declination were operating then the non-low pitches should continue to decline throughout the major phrase.

McCawley uses a cyclic[8] accent reduction rule to account for the appearance of mid pitches in non-initial minor phrases. In his formulation, the rule applies first to minor phrases, reducing all accents but the leftmost, and then to major phrases, again reducing all accents but the first. His pitch assignment rules then assign pitches on the basis of the strength of the accents--only sequences of morae preceding a primary accent (the leftmost in the phrase) are assigned high pitch. (The very first mora will be low if it is unaccented.) All secondary accents in the

---

[8] Shibatani (1972) criticizes McCawley's use of cyclic rules in accounting for the behaviour of pitch in major phrases without however questioning his description of the data. Whatever criticisms we might have of McCawley's description will apply equally well to Shibatani's.

phrase and pitch level 3 to those of the second.[9] So far this agrees with McCawley's description above. In addition, however, Jorden allows that there are higher and lower 'alternants' within each pitch level and states that:

> in any given pause group [corresponding to McCawley's major phrase], the first sequence of one or more syllables to be characterized by a given pitch level usually has the highest alternant of that level, and each successive sequence has a slightly lower alternant [29].

This running-down pattern both from level to level and from higher alternant to lower alternant is said to characterize the 'normal phrase accent' in Japanese [p 20]. Sequences of minor phrases which depart from this normal phrase accent pattern by having occurrences of pitch level 4 (or 5) in a non-initial constituent have a intonation phrase boundary before the constituent with the higher pitch and this departure from the normal pattern is significant syntactically in that it indicates the constituent structure of the utterance, and it has meaning, i.e. 'emphasis' (or strong emphasis in the case of pitch level 5) (Jorden 1955:20).

Like McCawley, Jorden reports a step down in pitch from one level to another between the first two minor phrases in a major phrase, or pause group as she calls it. Unlike McCawley, Jorden says that pitch continues to decline throughout the major phrase, not from level to level but from alternant to alternant. Why should the fall in pitch

---

[9] In Jorden's notation, pitch level 1 is low and level 5 is extra-high.

from the first to second minor phrase be considered more significant than the fall from the second to the third, or the third to the fourth? We could speculate that the initial fall is quantitatively different, i.e., greater, but it does not seem to be functionally different from the gradual descent from alternant to alternant. A rise in pitch indicates the beginning of a new intonation phrase, the lack of a rise indicates the continuation of the same intonation phrase. The initial fall and the following gradual descent both have the purpose of knitting together the elements of the phrase. Measurement of the F0 patterns of Japanese utterances will tell us whether there is in fact a quantitative difference between initial falls and other falls but there seems to be no reason to treat them as functionally different.

In Chew (1961) we have another even more explicit indication that pitch declination is operating in Japanese.

> Within the same accent [=intonation] phrase each successive high pitch and low pitch is lower than the preceding one [p 7].

Chew's example (p 8) is worth reproducing here in its original format which graphically illustrates his point. The pronunciation of the phrase /a no mise de ka u hoo ga i i to omot ta kedo/ 'I thought it would be better to buy at that store.' is graphically represented as:

```
     no mi se
                         u ho
/a          de ka              i
                                     o ga          mo
                                           i to o
                                                    t ta ke do/
```

In a footnote to this example Chew makes a very telling point:

> Assigning different pitch levels to /no mi se/ and /u ho/ in this sentence, while assigning the same pitch level to /u ho/, /i/, and /mo/ in accordance with Jorden's [and McCawley's] interpretation of Japanese intonation, seems inconsistent to me. They are either all different (a phonetic reality), or they are all phonemically the same. [p 159]

As far as Chew is concerned, the decline in pitch is not limited to the first two minor phrases but, as he has illustrated, continues through the third and fourth minor phrase and beyond. Furthermore it is not a stepping down from 'phonemic' pitch level to phonemic pitch level (as in Jorden from level 4 to level 3 or in McCawley from high to mid) but rather it is a phonetic effect imposed on all (phonemically) high (and low) morae in the utterance.

Chew uses this 'normal accent phrase' pattern as a tool for determining the location of intonation phrase boundaries and as an indication of their deletion in the case of the emphasis transforms he later describes (his Section 14). (We will return to this topic in a later section of this paper.) If the pitches of a non-initial minor phrase are at the same height as they would be in utterance-initial position then an intonation phrase boundary must occur directly to the left of the non-initial minor phrase. (This would amount to complete resetting of pitch declination.) Therefore, a break in the normal running-down pattern signals the onset of a new intonation phrase. Furthermore these intonational processes indicate the constituent structure of the utterance. Chew gives the following two sentences to show how the process works (p 8):

```
(1)  /a
                 si mi
       ni no sa
                          o ka u/  'He buys my brother's fish.'

(2)  /a          si mi
              sa
       ni wa
                          o ka u/  'My brother buys fish.'
```

In the second sentence, /si mi/ is pronounced at the same pitch level as /a/ and this indicates that there must be an intonation phrase boundary between /ani wa/ and /sasimi o/. In the first sentence, /ani no/ and /sasimi o/ are closely linked and together with the verb constitute only one intonation phrase. In the second sentence, /ani wa/ is not so closely linked with /sasimi o/ and the intonation phrase boundary between them emphasizes /ani wa/'s status as Topic, even though it is already marked with the /wa/ Topic marker. (It is probably a matter of de-emphasis rather than emphasis.)

Chew's statement (p 8) that intonation phrase initial pitch values are the same as utterance initial pitch values seems to rule out the possibility of partial declination resetting. In the example above, where the intonation phrase boundary is near the beginning of the sentence, it is not unlikely that resetting would result in the pitch values of the second phrase being the same as those of the first. One wonders though whether intonation phrase boundaries occurring later in an utterance would necessarily result in complete resetting of declination. It may be possible that declination resetting itself is enough to signal the beginning of a new intonation phrase--the amount of resetting, relative to utterance initial pitch values, might depend on position in the utterance and/or prominence relations between intonation phrases. It is unfortunate that the only example Chew gives is of an utterance with an intonation phrase boundary between the first and second minor phrases. Intonation phrase boundaries can occur at any minor phrase boundary (cf Chew 89ff), but Chew provides us with no graphical representation of sentences where the intonation phrase boundary occurs later in the utterance. The acoustic data in Chapter 5 below reveal that non-initial F0 peaks have a range of values, sometimes higher, sometimes lower, and sometimes similar to utterance initial peak values. Chew's insistence that intonation phrase initial minor phrases take on utterance initial pitch values may be too strong a statement of the process involved.

Of the descriptions we have discussed, Chew's gives us the clearest indication that declining pitch is a characteristic of intonation phrases in Japanese. Haraguchi (1977), who works within an autosegmental framework, explicitly formulates a rule, which he calls the Downdrift rule, to account for the behaviour of pitch in Japanese utterances. Haraguchi claims that downdrift affects the realization of High and Low tones to the extent that a High tone in a non-initial minor phrase will be lowered to the level of an immediately preceding Low. Succeeding Lows (and Highs) will be lower still. It is only in 'deliberate and relatively slow speech [p 30]' that pitch will rise anywhere in the utterance except at the very beginning. The pitch contour of an utterance is then a continuously falling one, with precipitous falls marking the accents. The question of the effect of rate of speech on pitch declination will be addressed in the analysis of the acoustic data which follows but in the light of that analysis it seems that Haraguchi has put his case too strongly. We have no reason to doubt that declination could apply to the extent that non-initial Highs are realized at the same pitch as immediately preceding Lows in some styles of (rapid) speech but the rule of pitch declination will probably need to be flexible enough to account for all ordinary styles and rates of speech. The requirement of Haraguchi's Downdrift rule that non-initial Highs downdrift to the level of immediately preceding Lows would not allow us to account for cases where pitch rises do occur and where downdrift still seems to apply though not to the extent that Haraguchi claims.

There seems to be considerable agreement among all of the above researchers that longer phrases in Japanese are characterized by a gradual fall in pitch. The nature of this fall, its extent, and the way we should account for it are matters of considerable disagreement. The acoustic-phonetic data presented in the following chapters will provide us with a better understanding of each of these points.

Chapter V
THE PRESENT STUDY

This chapter contains a description of the language data, the experimental procedure, the acoustic measures, and the statistical procedures and results, on which this analysis of Japanese intonation is based.

5.1 THE LANGUAGE DATA

The language data which were subjected to acoustic analysis consisted of sentences spoken by native speakers of Japanese. The bulk of the data was supplied by a young Japanese woman (TI) who is a native speaker of the Tokyo dialect of standard Japanese. This young woman was a visiting student at the University of Victoria. Two other native speakers were also recorded. One was a middle-aged Japanese woman (MW) who has lived in Canada for a number of years, and the other was a young man (MF), who is also a visiting student from Japan. All the subjects were recorded individually in a sound insulated room. The recording was done using a Sony ECM-22P microphone, a Revox model A77 dual track tape recorder and 1.5 mil Scotch 176 Audio Recording Tape run at 7.5 ips.

In all studies of intonation it is necessary to be aware that intonation has linguistic and non-linguistic aspects. The so-called attitudinal aspect of intonation, the use of pitch to signal emotional states, or speakers' attitudes toward their utterances or toward their interlocutors, is non-linguistic; it belongs to the realm of what Abercrombie (1967:103) calls 'vocal gesture'. In its linguistic aspect, intonation communicates linguistic information about the utterance. Many studies have focussed on the relationship of grammatical sentence types to particular intonation contours (Hadding-Koch 1961, for example). We are focussing on intonation as a cue to phrasing. This is the function which Gardiner (1977) identifies as intonation's primary linguistic function. Lehiste (1970:95) cites a number of studies which attempted to separate the linguistic and attitudinal aspects of intonation. She concludes that, while there is agreement among researchers that the two have to be kept separate, 'there is no effective and universally applicable methodology for achieving this separation [p 96].' With this in mind, the sentences which were used in this study were constructed[10] and presented to the subjects in a way calculated to minimize the effect of non-linguistic variables on their intonation.

---

[10] The sentences were constructed with the help of a native speaker of Japanese who was not otherwise involved in the study.

The test sentences are to be found in Appendix A. They are all simple sentences containing only one clause and their lexical content is unremarkable. They consist of ordinary nouns and verbs, and common sentence modifiers. In general, they conform to the basic Subject-Object-Verb word order of Japanese. No discourse context was provided for the sentences; the sentences were presented as a list and the subjects were instructed to consider each as independent of the others. Since the normal form of Japanese writing is kanji script, the subjects read a kanji transcription of the sentences rather than the romanized versions which appear in Appendix A. Every attempt was made to make the subjects as comfortable as possible and the recording sessions were kept short to avoid fatigue or boredom. The subjects were told to read the sentences in as natural a way as possible. After they finished recording the sentences, the subjects were asked to listen to the tapes and were encouraged to re-record items which seemed to them to be unnatural or 'incorrect'. They were not, however, given any instructions as to how to change their rendering of the sentences. All of the above procedures were followed in order to reduce the interference of non-linguistic factors on the intonation of the test items.

Although the lexical items from which the sentences were constructed are common ordinary words, some attention was paid to their phonetic makeup. As far as possible, the lexical items contained voiceless consonants. This was simply a convenience in that the presence of voiceless segments is a considerable aid in the later segmentation of F0 traces of recorded utterances. A second criterion applied in the selection of lexical items was that accented forms were given preference over unaccented forms. In the case where all the lexical items, nouns, verbs, and modifiers, which are involved in a sentence are accented, there will be a minor phrase boundary after each. In the case of unaccented lexical items, some of the phrase boundaries may be absent (McCawley 1968:177). Since we were interested in the behaviour of the pitch peaks of minor phrases when they are involved in longer utterances we wanted to maximize the occurrence of minor phrase boundaries and so avoided selecting unaccented forms for inclusion in the test sentences.

It was not possible to ensure that the peak syllables of each sentence contained similar vowels. Vowel quality has been shown to have an effect on the intrinsic fundamental frequency of vowels (Lehiste and Peterson 1961, Peterson and Barney 1952): in general, high vowels tend to have higher F0 than low vowels. Sorensen and Cooper found that vowel quality affected the level of F0 peaks in their study (Sorensen and Cooper 1980). They also found that when they attempted to standardize the vowels in their test sentences (eg. The man in the van sent pansies to his fans. p 413), their subjects used an atypical 'sing-song' rhythm and so they rejected this approach. To avoid this problem and to make the selection of lexical items easier, no attempt was made to standardize the vowels in these sentences.

5.2 SPECTRUM GENERATION AND FUNDAMENTAL FREQUENCY EXTRACTION

The recorded test items were subjected to further processing to extract fundamental frequency on the one hand and to generate digital spectra for the purpose of segmentation on the other. The components of the system included an FJ Electronics Fundamental Frequency Meter (Type FFM 650), a Digital LSI-11 mini-computer equipped with a 12-bit analog to digital (A/D) converter , and an IBM 4341 mainframe computer running under VM/CMS with a variety of peripheral output devices. (See schematic in Appendix B.) The same tape recorder was used for playback as for recording.

In order to generate a spectral representation of the test items, one of two identical output signals from the tape recorder was low-pass filtered by means of an analog filter set at 5000 Hertz (Hz). The filter attenuates frequencies above the cutoff frequency at a rate of 18 decibels per octave. The voltage of the filtered signal was then sampled (at a sampling rate of 10kHz) by the A/D converter and the resulting time series was stored initially in the core memory of the LSI-11 mini-computer. The analog filter setting and the sampling rate are based on Shannon's sampling theorum which states that

> any band-limited signal can be exactly reconstructed from samples taken periodically in time if the sampling rate is twice the highest frequency of the signal (Schafer and Rabiner 1975:663).

Since the greatest portion of the energy of speech is concentrated in a frequency region well below 5000Hz, this frequency was selected as the cutoff frequency for the limited signal. This procedure of filtering and then sampling the speech signal output of the tape recorder can accurately represent, as a time series, the components of the signal which are below 5000Hz.

Once the sampling routine has begun, the A/D converter periodically samples the input signal but the samples are not stored. In order to capture a segment of speech (i.e. a sentence) the samples have to be stored. Capturing does not begin until the input signal contains a voltage equal to, or greater than, a pre-set trigger voltage. In this case the trigger level was set so that capturing would begin at the onset of the first vowel in the sentence. Successive sample points were then captured until a duration of time set by the operator had elapsed. The maximum number of sample points which can be stored by the mini-computer at one time is 12,500. This means that, at a sampling rate of 10kHz, the maximum duration of the signal which can be captured is 1250 milliseconds (ms) or 1.25 seconds. Since some of the test items were longer than this, they had to be captured in two sections. A delay time option, which specifies that capturing not begin until a certain time has elapsed after the trigger voltage has been sensed, was used for these longer items. For the first section of a longer item, the delay time was set to zero and capturing began immediately after the trigger voltage was sensed. For the second section, a delay time equal to the duration of the previous section, 1250ms, was specified and capturing did not begin until this delay time had elapsed.

The software package under which the sampling and storage routines operate includes an editing routine which provides for digital to analog (D/A) playback. The D/A playback was used for auditory checking of the captured signal. This auditory checking was necessary to ensure that the relevant portions of the input signal were captured. (Later, whenever the digital spectra were ambiguous as to the boundaries between segments, D/A playback was used to resolve the ambiguity auditorily.) If the test item was properly captured, the time series was given a filename and was stored on floppy diskette. Later this digital representation of the audio signal was transferred to the IBM mainframe computer for further processing.

The complex speech signal can be viewed as a combination of sinusoidal waves which vary over time and which differ in frequency, amplitude and phase. Under certain circumstances, a time series derived from the complex speech waveform can be transformed into a spectral representation which separates the component frequencies of the complex waveform while conserving information about their relative amplitudes. The audio time series generated by the mini-computer were subjected to a Fast Fourier Transform (FFT) procedure on the mainframe which does just that. The FFT creates a digital spectrum of the original waveform which shows the distribution of energy of the speech signal in the frequency domain. Inspection of the digital spectra so created allowed us to determine the characteristics of the speech signals as they varied over time. Since different speech sounds have different distributions of energy in the frequency domain, the spectrum of a speech signal allows us to identify individual speech sounds. We used this information to reliably segment the captured test items into speech sounds and words. Once the segmentation of the test items into speech sounds and words was done, this information was applied to the fundamental frequency contours, obtained via a separate procedure detailed below, so as to relate fundamental frequency values to specific speech segments.

A second output from the tape recorder was passed through the Fundamental Frequency Meter. This instrument uses a combination of high and low pass filters to isolate the fundamental frequency component of the input signal. The instrument then uses a sample and hold circuit to sense the zero crossings of the fundamental and produces as output a continuous voltage which is proportional to the reciprocal of the period of the fundamental (FJ Electronics Reference Manual). The output of the F0 meter can be directed to a mingograph but in this case it was directed to the A/D converter of the mini-computer. The output signal was sampled at a rate of 1kHz and a time series was generated for each test sentence which represented the fluctuations in the voltage of the signal. The capture of these F0 time series was initiated by the same triggering device, set at the same level, which initiated the capture of the corresponding audio signal. The triggering device sensed the other output of the tape recorder routed as before through the analog filter. This procedure ensured that the two time series representing different aspects of the same sample began at the same temporal point. The F0 time series were stored locally on floppy diskette before being transferred to the mainframe for further processing.

Both the digital spectra and the F0 time series could be printed in digital form; both could be displayed digitally or graphically on a video display terminal. In addition the F0 traces could be hardcopied using a Tektronics hardcopy device. Sample F0 traces are included in Appendix C.

Since the continuous voltage output of the F0 meter and, of course, the resultant time series are only proportional to the frequency of the fundamental, the system had to be calibrated so that the values to be displayed would be in Hertz. The function used for the calibration was empirically determined using input signals of known frequency. These signals were passed through the system and the time series values generated from the output of the F0 meter were compared with the known frequency values of the input signals. The relationship between the two was determined and a conversion function derived. This function was applied to all of the F0 time series before display and so the output is scaled as frequency (in Hertz) over time. After the conversion function was applied but before the values were displayed they were rounded to the nearest 5Hz. The voltage from the F0 meter fluctuates slightly--the rounding of the F0 values smoothes the F0 contours without significantly affecting the relative frequencies of the F0 peaks.

A comparison of the digital spectra of the audio time series with the F0 time series was made and segmentation of the latter was carried out on the basis the segmental composition of the former. At this point, the test items were represented by segmented F0 time series and the selection of the relevant fundamental frequency values could be made.

5.3 THE MEASUREMENT OF FUNDAMENTAL FREQUENCY PEAKS

Since the fundamental frequency of speech varies over time, the F0 of a particular speech segment, mora, word, or sentence is not a single datum. All work with F0 requires some form of data reduction--the selection or abstraction of a single significant value or of a relatively small number of values (Sorensen and Cooper 1980:403). In the present study the values which seemed most useful for the characterization of intonation in Japanese were the F0 peaks associated with each minor phrase in an utterance. This is essentially the same method of data reduction used by Sorensen and Cooper in their studies of English intonation (see section 3.2, above). In the past, F0 peaks have yielded valuable information about intonation (Sorensen and Cooper 1980:403). Ease of measurement was also a consideration, since F0 peaks are relatively easy to locate on an F0 trace.

Even though F0 peaks are relatively easy to locate, care had to been taken to avoid the selection of spurious peaks. In order for a data point to be selected as representing an F0 peak, it had to be consistent with its neighbours. Sometimes momentary peaks appear in the F0 trace which do not represent actual F0 values; this is especially true when speech intensity is low. These spurious peaks are an artifact of the measurement procedure and as such are not valid data and should not be selected. Another consideration in the selection of peak values was the fact that F0 often changes very rapidly in the region of an F0 peak. Consequently a time window of 10ms was applied to the contour in the region of a peak and the frequency value which appeared most often in that span (the mode) was assigned to the peak. If peak F0 was relatively steady, the mid point of the steady state was selected as the location of the peak.

Relatively short term perturbations in F0 are to be found in the region of segment boundaries, particularly when the boundary is between a voiceless obstruent and a vowel (Hombert 1978, House and Fairbanks 1953). It is unlikely that these perturbations, which are the result of coarticulation, represent significant features of intonation. Inspection of the F0 traces of the the Japanese sentences under analysis revealed that these momentary perturbations were generally confined to a period within 20ms of the preceding and 10ms of the following segment boundary. No data points were selected from these regions.

Once the fundamental frequency peaks had been identified they were recorded along with their locations. The location of the first peak in the utterance was set to zero. The location of each subsequent peak was recorded as the elapsed time (in ms) from the first peak. For the purpose of comparison among sentences, the length of each was considered to be the time between the first and last peak. This method of locating F0 peaks was used by Sorensen and Cooper in their studies of English intonation (Sorensen and Cooper 1981).

5.4 PITCH DECLINATION IN JAPANESE

Even a cursory examination of the F0 traces of the sentences under consideration reveals that pitch does indeed tend to decline over the length of the sentence. The general pattern is unmistakable. In order to provide a statistical basis for the description of the intonation contours of these sentences, a number of procedures available in the Statistical Package for the Social Sciences (SPSS) were applied to the acoustic data. Several variables were coded for each sentence. The fundamental frequency of the F0 peaks (PEAKF0), their location in the sentence (SENTPOS), and their associated time of occurrence (TIMEPK) were coded along with the length of each utterance in words[11] (LENUTT) and in time (TIMEUTT).

Since we were unable to control for vowel quality in constructing the sentences, the height of the vowels (HILO) during which the F0 peaks occurred was included as a variable in order to find out what influence, if any, vowel height had on the F0 of the peaks. Vowel height was coded in the following way. The vowels /i/ and /u/ were considered to be high, all other vowels were considered to be low. The two high vowels are important in Japanese in that they pattern similarly with respect to a high vowel devoicing rule

---

[11] 'Words' in this context refers to what we have up to now been calling minor phrases. Minor phrases pattern in most respects as words and 'word' is a less cumbersome term than minor phrase. SENTPOS is also coded in words, i.e. SENTPOS=1 refers to first word in a sentence, SENTPOS=2 refers to the second and so on.

which devoices /i/ and /u/ when they occur between two voiceless consonants (Haraguchi 1977:35). Acoustic evidence for the effect of vowel height on pitch suggests that /i/ and /u/ have the highest intrinsic F0 while /a/ and /o/ have the lowest (Lehiste and Peterson 1961). The other relevant vowel /e/ is somewhat closer in intrinsic F0 to the low vowels than to the high vowels and does not partake in the high vowel devoicing rule. For these reasons it was coded as low. In the sentences of speaker TI, F0 peaks occurred in syllables containing low vowels 133 times, in syllables containing high vowels only 39 times. No consistent effect of vowel height on the fundamental frequency of the peaks in these sentences was detected.

The primary variables, the ones which allow us to characterize the general pattern of intonation in these sentences, are PEAKFO and TIMEPK. The peaks themselves are designated as P1, P2,...Pn, where P1 is the first peak in a sentence, and Pn is the last. The designations T1, T2,...Tn refer to the times of occurrence of P1, P2,...Pn, respectively. The F0 peaks of all the three-word sentences of speaker TI are shown in the form of a scattergram plot in Figure 1.
The Condescriptive Statistics sub-routine of SPSS yielded means and standard deviations for the frequency and time of occurrence (t) of the peaks of each sentence. Table 1 presents these measures for the three-word sentences of subject TI. The mean PEAKFO and TIMEPK values in Table 1 are

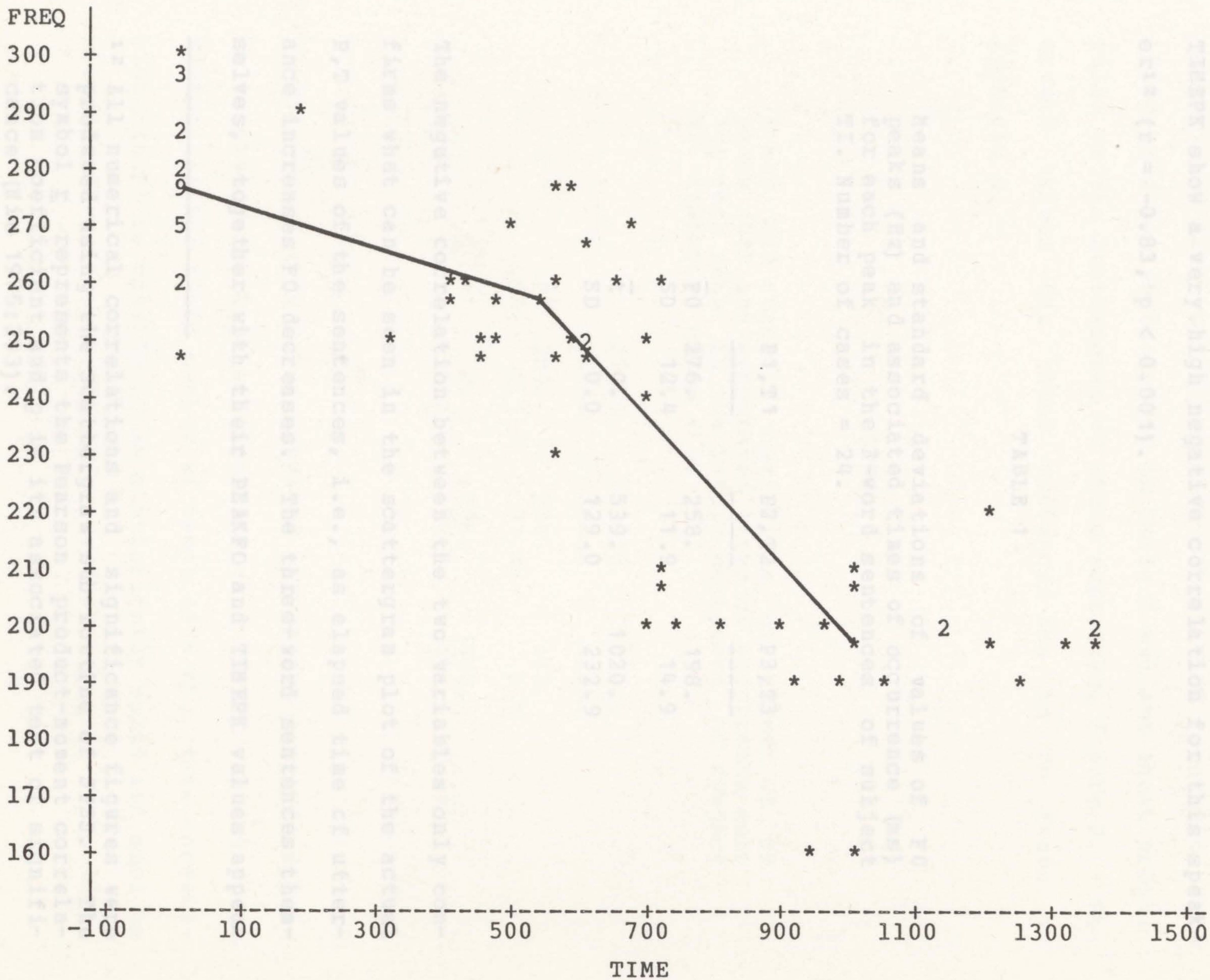


Figure 1: Scattergram. Three-word Sentences of Speaker TI.

included in the scattergram plot in Figure 1 and the line connecting them represents the proposed line of pitch declination for these sentences. The two variables PEAKFO and TIMEPK show a very high negative correlation for this speaker[12] ($r = -0.83$, $p < 0.001$).

TABLE 1

Means and standard deviations of values of F0 peaks (Hz) and associated times of occurrence (ms) for each peak in the 3-word sentences of subject TI. Number of cases = 24.

	P1,T1	P2,T2	P3,T3
$\overline{F0}$	276.	258.	198.
SD	12.4	11.9	14.9
$\overline{t}$	0.	539.	1020.
SD	0.0	129.0	232.9

The negative correlation between the two variables only confirms what can be seen in the scattergram plot of the actual P,T values of the sentences, i.e., as elapsed time of utterance increases F0 decreases. The three-word sentences themselves, together with their PEAKFO and TIMEPK values appear

---

[12] All numerical correlations and significance figures were produced using the Scattergram sub-routine of SPSS. The symbol r represents the Pearson product-moment correlation coefficient and p is its associated test of significance (Nie 1975:293).

in Appendix D.

Subject MW has a lower-pitched voice; all of her peaks were somewhat lower in F0 than those of TI. The mean frequency and time of occurrence of the F0 peaks in MW's three-word sentences appear in Table 2 and are shown graphically on the scattergram of those sentences, Figure 2. The actual values for these sentences are to be found in Appendix G.

TABLE 2

Means and standard deviations of values of F0 peaks (Hz) and associated times of occurrence (ms) for each peak in the 3-word sentences of subject MW. Number of cases = 6.

	P1,T1	P2,T2	P3,T3
$\overline{F0}$	257.	230.	183.
SD	6.1	11.4	6.9
$\overline{t}$	0.	708.	1320.
SD	0.0	54.9	153.1

The correlation between PEAKF0 and TIMEPK for this speaker is also very high and negative ($r = -0.94$, $p < 0.001$). Because of the small number of cases ($N = 6$), this correlation cannot be considered as statistically valid and must be viewed with caution. Still, the fact that it is so high,

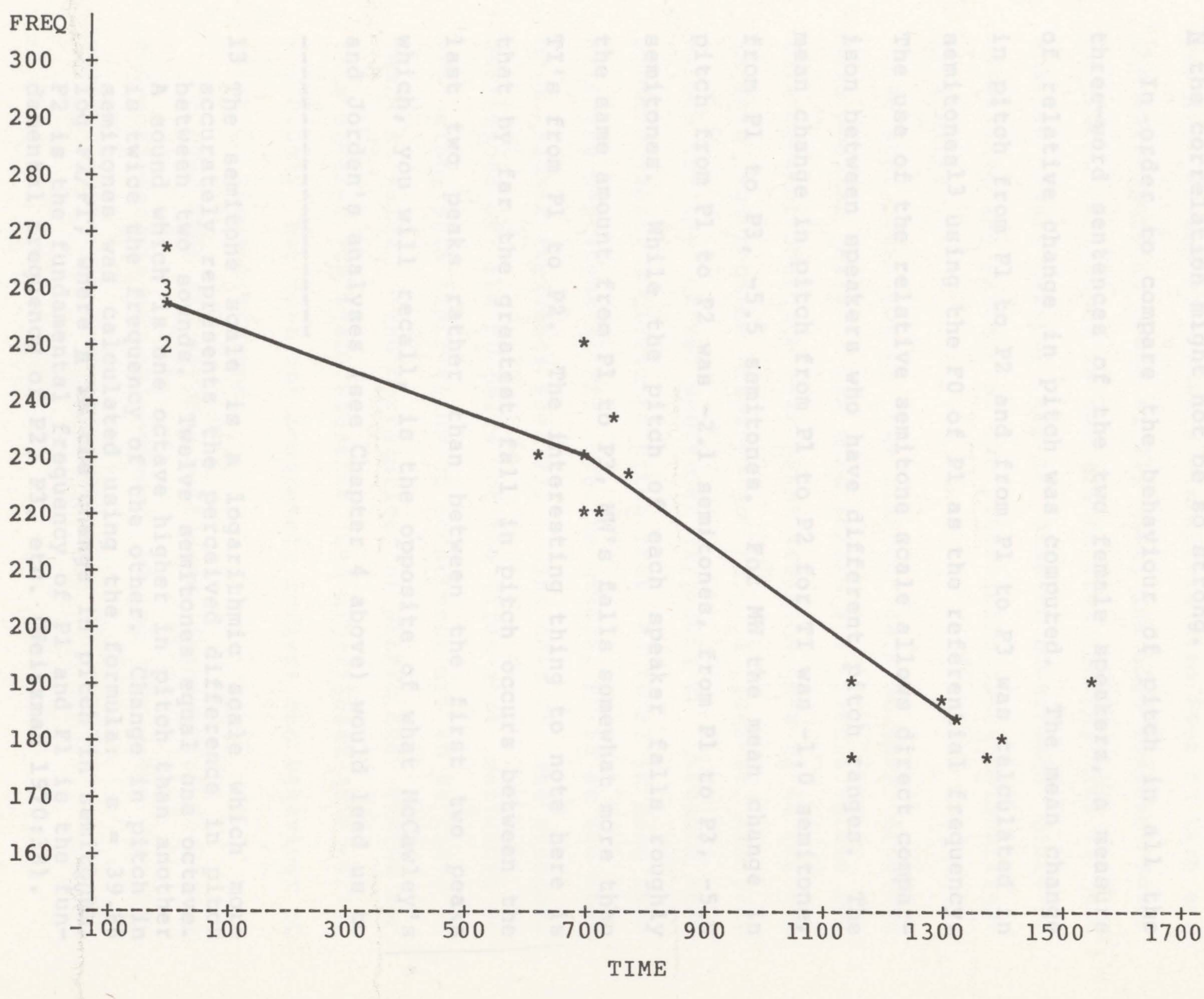


Figure 2: Scattergram. Three-word Sentences of Speaker MW.

and that the pattern of F0 peaks for this speaker conforms quite closely to the pattern for Speaker TI, suggest that PEAKFO and TIMEPK are correlated, even if, for some greater N the correlation might not be so strong.

In order to compare the behaviour of pitch in all the three-word sentences of the two female speakers, a measure of relative change in pitch was computed. The mean change in pitch from P1 to P2 and from P1 to P3 was calculated in semitones[13] using the F0 of P1 as the referential frequency. The use of the relative semitone scale allows direct comparison between speakers who have different pitch ranges. The mean change in pitch from P1 to P2 for TI was -1.0 semitone, from P1 to P3, -5.5 semitones. For MW the mean change in pitch from P1 to P2 was -2.1 semitones, from P1 to P3, -5.9 semitones. While the pitch of each speaker falls roughly the same amount from P1 to P3, MW's falls somewhat more than TI's from P1 to P2. The interesting thing to note here is that by far the greatest fall in pitch occurs between the last two peaks rather than between the first two peaks which, you will recall, is the opposite of what McCawley's and Jorden's analyses (see Chapter 4 above) would lead us to

---

[13] The semitone scale is a logarithmic scale which more accurately represents the perceived difference in pitch between two sounds. Twelve semitones equal one octave. A sound which is one octave higher in pitch than another is twice the frequency of the other. Change in pitch in semitones was calculated using the formula: $s = 39.86 \log F2/F1$, where s is the change in pitch in semitones, F2 is the fundamental frequency of P1 and F1 is the fundamental frequency of P2, P3, etc. (Weitzman 1970:38).

expect. The mean changes in pitch in the three-word sentences of both MW and TI are presented in Figure 3. The pitch of P1 is set at zero semitones as a reference point. For MW, P2 and P3 are more distant in time from P1 than are the corresponding peaks of TI, they are also somewhat lower in pitch. If PEAKFO is correlated with TIMEPK, this is just what we would expect, and if time of occurrence could be normalized the pattern of the pitch falls might be strikingly similar.

We must turn now to the limited data we have from the male speaker (MF). The mean PEAKFO and TIMEPK values for this speaker indicate that pitch declines from P2 to P3, but it does not decline from P1 to P2. A look at the peak values for the actual sentences (Appendix H) reveals that pitch actually declines in pitch from P1 to P2 in less than half the sentences of MF. In the others, P2 is at the same or a higher frequency than P1. The mean F0 and standard deviations of the F0 peaks of the three-word sentences of MF, along with their times of occurrence are to be found in Table 3 below. The correlation of PEAKFO and TIMEPK for this speaker was again negative and significant ($r = -0.70$, $p < 0.001$) but does not accurately reflect the behaviour of pitch in these sentences.

In over half (5) of MF's sentences, fundamental frequency did not decline at all between P1 and P2. Why is this speaker different from the others? The sentence data for

Figure 3: Mean change in pitch in semitones

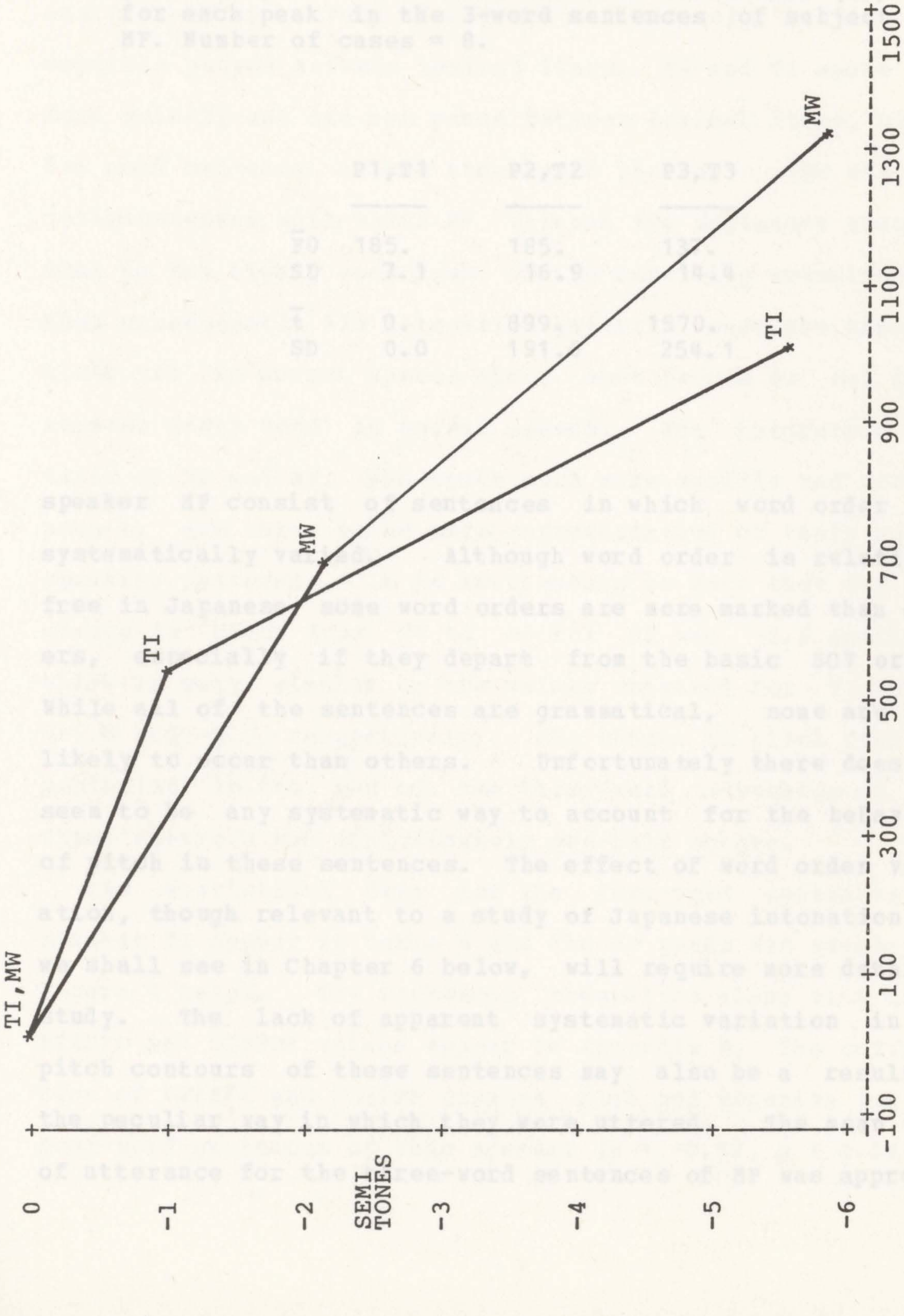


TABLE 3

Means and standard deviations of values of F0 peaks (Hz) and associated times of occurrence (ms) for each peak in the 3-word sentences of subject MF. Number of cases = 8.

	P1,T1	P2,T2	P3,T3
$\overline{F0}$	185.	185.	137.
SD	7.1	16.9	14.4
$\overline{t}$	0.	899.	1570.
SD	0.0	191.6	254.1

speaker MF consist of sentences in which word order was systematically varied. Although word order is relatively free in Japanese, some word orders are more marked than others, especially if they depart from the basic SOV order. While all of the sentences are grammatical, some are less likely to occur than others. Unfortunately there does not seem to be any systematic way to account for the behaviour of pitch in these sentences. The effect of word order variation, though relevant to a study of Japanese intonation, as we shall see in Chapter 6 below, will require more detailed study. The lack of apparent systematic variation in the pitch contours of these sentences may also be a result of the peculiar way in which they were uttered. The mean time of utterance for the three-word sentences of MF was approximately 50% greater than for the three-word sentences of TI and MW. An inspection of the F0 traces and the digital spectra of MF's sentences reveals that the reason for this was that MF spoke very slowly and deliberately, leaving perceptible pauses between lexical items. MW and TI spoke much more quickly and did not pause between lexical items, uttering each sentence as one three-word phrase. The atypical deliberateness with which MF uttered the sentences suggests that he was highly conscious of the recording situation and that consequently his intonation patterns were not those he would use in normal speech since most people do not pause between every word in normal speech. The intonation patterns of TI and MW, who spoke much more rapidly and without pauses, are taken to be more representative of their normal speaking patterns. It is interesting to note that the mean change in pitch from P1 to P3 for MF was -5.5 semitones which is very similar to the values obtained for TI and MW (-5.5 and -5.9, respectively). The change in pitch from the beginning to the end of the three-word sentences of all three speakers was approximately one-half octave. (3.43)

The statistical data for the four-word sentences of speaker TI appear in Table 4 and the F0 peaks are graphed in Figure 4 below. The sentences themselves along with their PEAKF0 and TIMEPK values appear in Appendix E. The correlation of PEAKF0 and TIMEPK remains high and negative for the four-word sentences of this speaker ($r = -0.84$, $p < 0.001$).

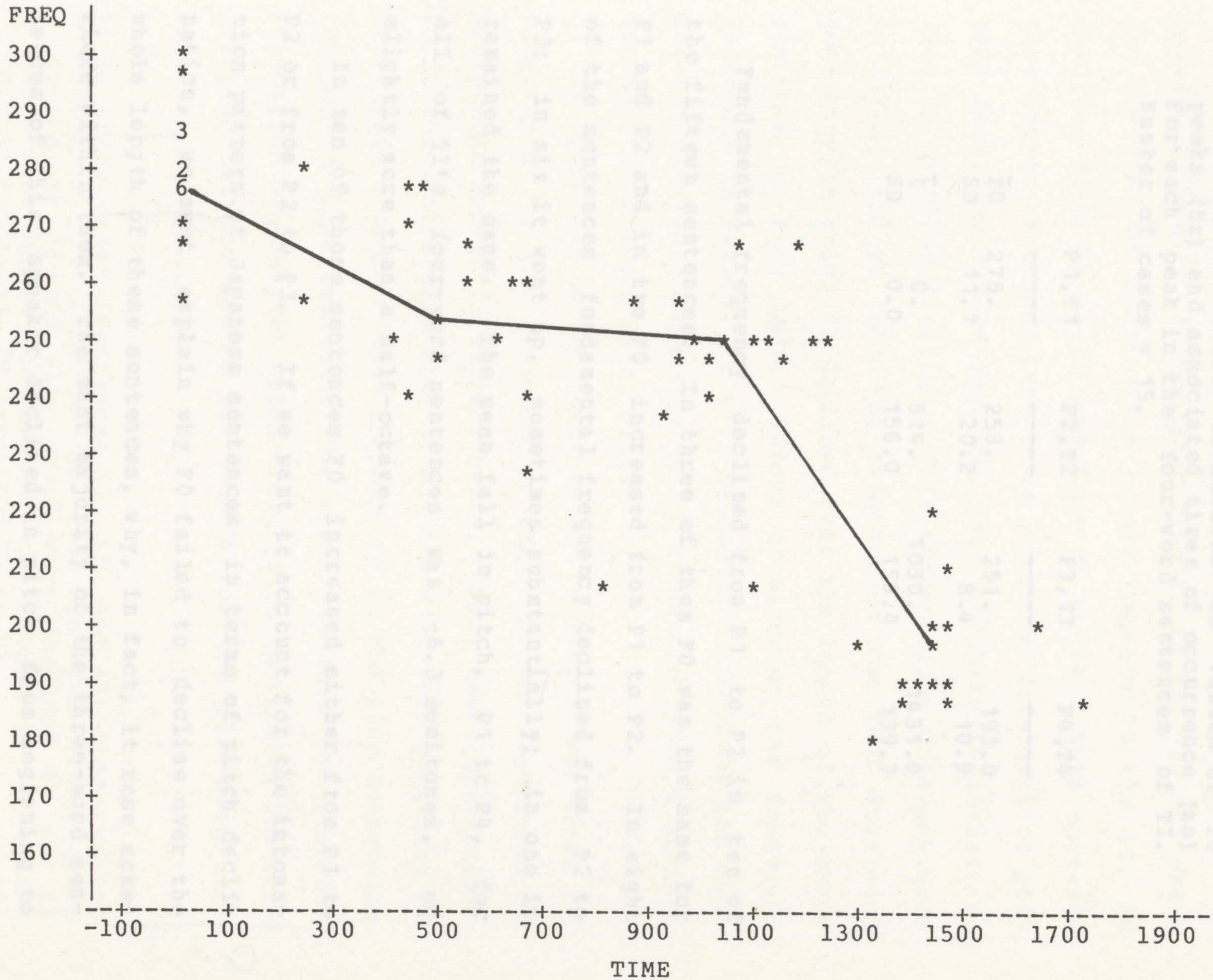


Figure 4: Scattergram. Four-word Sentences of Speaker TI.

TABLE 4

Means and standard deviations of values of F0 peaks (Hz) and associated times of occurrence (ms) for each peak in the four-word sentences of TI. Number of cases = 15.

	P1,T1	P2,T2	P3,T3	P4,T4
$\overline{F0}$	278.	253.	251.	195.0
SD	11.1	20.2	8.4	10.9
$\bar{t}$	0.	514.	1030.	1431.0
SD	0.0	156.0	150.4	139.7

Fundamental frequency declined from P1 to P2 in ten of the fifteen sentences. In three of them F0 was the same for P1 and P2 and in two F0 increased from P1 to P2. In eight of the sentences fundamental frequency declined from P2 to P3; in six it went up, sometimes substantially; in one it remained the same. The mean fall in pitch, P1 to P4, for all of TI's four-word sentences was -6.3 semitones, or slightly more than a half-octave.

In ten of these sentences F0 increased either from P1 to P2 or from P2 to P3. If we want to account for the intonation pattern of Japanese sentences in terms of pitch declination, we must explain why F0 failed to decline over the whole length of these sentences, why, in fact, it rose somewhere within them. The vast majority of the three-word sentences of this speaker declined in pitch from beginning to end.[14] In the four-word sentences, we see breaks in the line of declination, with rises in F0 occurring between intermediate peaks. If the break in the line of declination occurs between P1 and P2, F0 declines from P2 to P4. If the break in the line of declination occurs between P2 and P3, F0 declines from P1 to P2 and from P3 to P4. This suggests that longer sentences tend to be segmented into two, or perhaps more, intonation phrases, each of which is characterized by pitch declination. This segmentation does not have to occur, and we have the evidence of the four four-word sentences in which F0 declined over the length of the whole sentence to show this. Before considering the four-word sentences in detail, it will be useful to look at the data for the five-word sentences of TI since they show a similar tendency to include pitch rises.

The mean values of fundamental frequency and times of occurrence for the F0 peaks of the five-word sentences of subject TI appear in Table 5 below. These means are included in the scattergram plot of all the five-word sentences of TI, Figure 5 below. The disjuncture in the line connecting the points that represent the means reflects the proposed intonation boundary at which declination is reset. The correlation between PEAKFO and TIMEPK for the five-word sen-

---

[14] In three cases (Sentences 13, 32, and 45) there was no change in F0 from P1 to P2. In two other cases (Sentences 34 and 35), which were included to see if there were direct contrasts in the intonation of certain related sentences, F0 rose from P1 to P2.

tences of TI is again negative and high, but it is not so high as for her three and four-word sentences ($r = -0.69$, $p < 0.001$). The overall fall in pitch from P1 to P5 for these sentences is -5.6 semitones, again almost one half-octave.

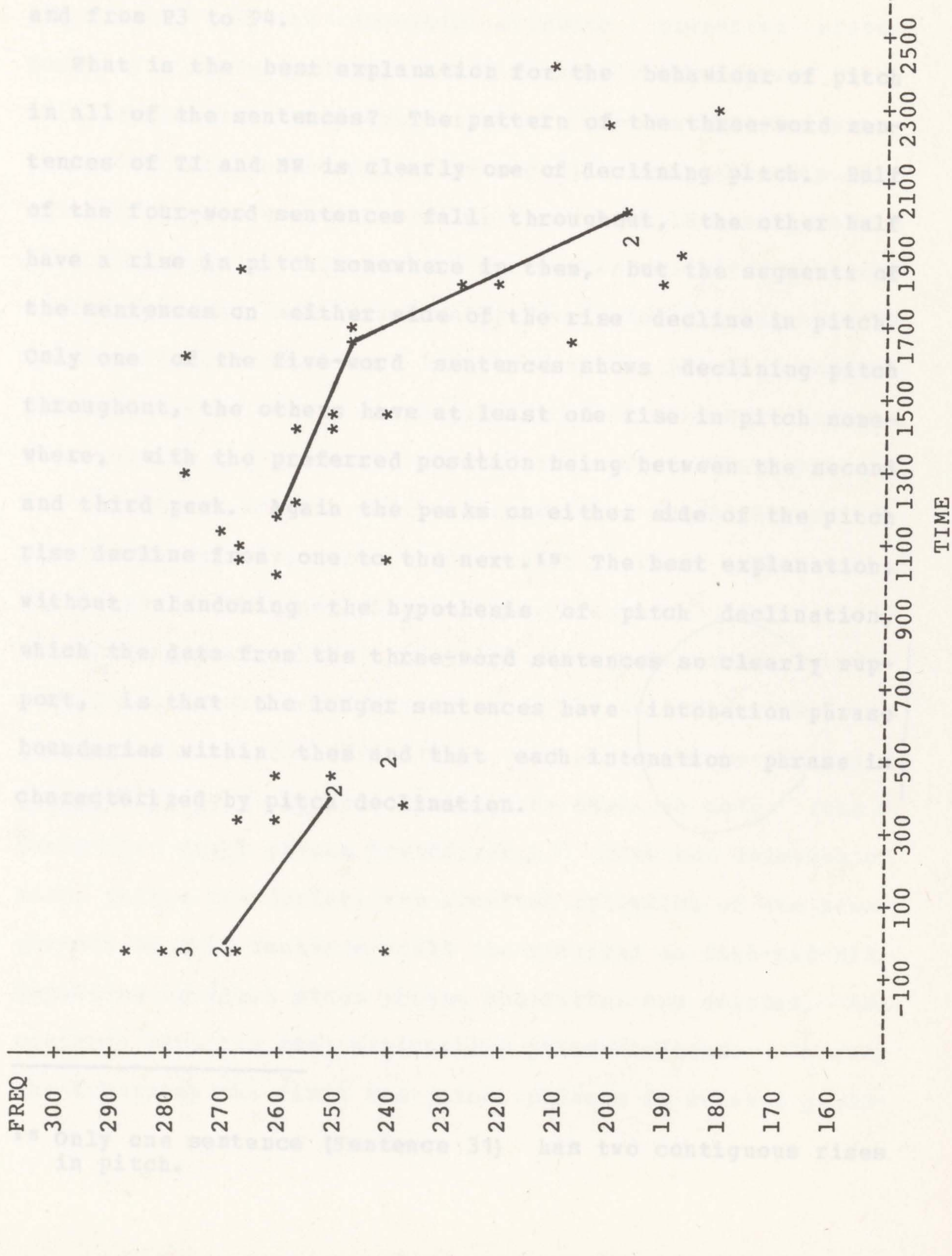
TABLE 5

Means and standard deviations of values of F0 peaks (Hz) and associated times of occurrence (ms) for each peak in the 5-word sentences of subject TI. Number of cases = 8.

	P1,T1	P2,T2	P3,T3	P4,T4	P5,T5
$\bar{F0}$	271.	250.	260.	246.	195.0
SD	13.7	11.0	11.3	18.7	10.0
$\bar{t}$	0.	428.	1169.	1652.	2036.0
SD	0.0	66.1	134.5	182.2	210.0

A look at Figure 5 reveals that there is a distinct tendency for F0 to rise between P2 and P3. A look at the actual sentence data (Appendix F) reveals that five of these sentences show a rise between P2 and P3, one neither rises nor falls. Of the others, in Sentence 12, F0 falls throughout; in Sentence 3, F0 rises from P1 to P2, falls from P2 to P3, and rises again from P3 to P4. Sentence 12 shows that pitch <u>can</u> decline throughout the length of even a longer sentence. The others show that it usually does not. Sentences 5, 20, 28, 36, and 38 all have rises in pitch from

Figure 5: Scattergram. Five-word sentences of TI.



P2 to P3--pitch declines from P3 to the end in all of them (actually Sentence 38 has the same pitch for P2 and P3). Sentence 31 has two contiguous rises in pitch, from P2 to P3 and from P3 to P4.

What is the best explanation for the behaviour of pitch in all of the sentences? The pattern of the three-word sentences of TI and MW is clearly one of declining pitch. Half of the four-word sentences fall throughout, the other half have a rise in pitch somewhere in them, but the segments of the sentences on either side of the rise decline in pitch. Only one of the five-word sentences shows declining pitch throughout, the others have at least one rise in pitch somewhere, with the preferred position being between the second and third peak. Again the peaks on either side of the pitch rise decline from one to the next.[15] The best explanation, without abandoning the hypothesis of pitch declination, which the data from the three-word sentences so clearly support, is that the longer sentences have intonation phrase boundaries within them and that each intonation phrase is characterized by pitch declination.

---

[15] Only one sentence (Sentence 31) has two contiguous rises in pitch.

5.5 A RECONSIDERATION OF THE NON-ACOUSTIC EVIDENCE

Previous descriptions and analyses of the behaviour of pitch in longer Japanese utterances must be re-considered in the light of the acoustic evidence presented above. McCawley's analysis (Chapter 4.2 above) which implied that, while there was a step-down in pitch from the leftmost minor phrase in a major phrase to the second, the pitches in non-initial phrases were all similar (i.e. mid) is easily falsified by our data. In many of our sentences, pitch declines over three, four and occasionally five minor phrases. It is hard to imagine that each of these minor phrases contains pitches which are a consequence of different strength accents. If the first and second minor phrases have primary and secondary accents which are realized as high and mid pitch respectively, one wonders what sort of accents the third and fourth minor phrases have.

McCawley does allow that there is a rule which, under certain circumstances, deletes minor phrase boundaries. He gives as an example (1968:179) the major phrase: doo % ittara % ii desu ka 'how would it be best to go'. (The % marks the minor phrase boundaries.) With no deletion of minor phrase boundaries, the accented syllables of the minor phrases of this sentence would be realized as High-Mid-Mid. Depending on which minor phrase boundaries are deleted, the sentence also has two alternative pronunciations. In one, the % between the first two minor phrases is deleted yielding *doo* *ittara* % *ii* *desu* *ka*. McCawley's accent reduction rule (p 173) applies first at the minor phrase level reducing the accent in *ittara* to a secondary accent and then again at the major phrase level reducing it still further to tertiary accent which is 'not pronounced'. This means that only *do* in the first phrase and the first *i* in *ii* *desu* *ka* have non-low pitches, high and mid respectively; all other morae are low. In the other alternative pronunciation, the first *and* second % are deleted and the operation of the cyclic accent reduction rule ensures that only *do* is realized with high pitch--all other morae are low. The pitch falls from the first mora of the phrase to the second and there are no subsequent rises (or presumably falls) in pitch after this. There can be no rises in pitch because McCawley's pitch assignment rules make explicit reference to minor phrase boundaries in their assignment of low pitch to the *first* mora in a minor phrase. Recall from Chapter 4.1 above that the first mora of a minor phrase is low if it is unaccented. Since pitch can only rise, according to McCawley, between the first and second mora after a minor phrase boundary, and since these boundaries have been deleted in this example, there can be no pitch rises after *do*.

McCawley's accent reduction rule, even with the addition of the minor phrase boundary deletion rule above cannot account for the behaviour of pitch in most of our sentences, particularly the longer ones. Consider the following example: 16) musume ga otoko ni hana o ataeru 'The daughter gives the man flowers.' This sentence would start out with boundaries between each of the minor phrases. Since the pitch of the F0 peaks of each of the phrases declines over the whole utterance and, incidentally, as the F0 trace reveals, the pitch rises from the first to second mora of each phrase thus indicating the continued presence of the phrase boundaries, there is no way for McCawley to account for the intonation of this sentence. In a sentence like: 20) kyoo otoko ga isi o ike ni nageru 'Today the man throws pebbles into the pond.' whose pitch rise from P2 to P3 might signal the presence of a major phrase boundary between otoko ga and isi o, McCawley might conceivably be able to account for the pitch contours of each major phrase,[16] but he cannot account for the fact that the F0 peak in isi o is considerably lower than that in kyoo. If both these phrases are the leftmost in a major phrase then they should both contain equally high pitches.

Eleanor Jorden's analysis (1955) fares somewhat better than McCawley's in the light of the acoustic evidence. Recall that Jorden describes the pitch of a major phrase as declining first of all from, typically, level 4 to level 3 and from alternant to lower alternant thereafter. There

---

16 In this sentence there is no rise in pitch from the first to the second mora of the last minor phrase, the F0 peak occurs within the first mora and pitch falls to the end of the sentence. In McCawley's terms, this could be explained as being a consequence of the deletion of the minor phrase boundary between ike ni and nageru.

seemed to be no functional or qualitative reason, and Jorden does not give one, why the the fall in pitch from the first to second minor phrase is treated differently from subsequent falls. We speculated that there may be a quantitative difference in the falls with the first being much greater than the others. On the contrary, the F0 traces reveal that the greatest fall in pitch is from the next-to-last to the last minor phrase and that, although the mean pitch fall from the first to the second peak is greater than the mean fall from the second to the third,[17] in particular sentences the fall from P1 to P2 is sometimes greater and sometimes less than that from the P2 to the P3. For example, in Sentence 10 the fall from P1 to P2 is -2.0 semitones and from P2 to P3 it is -0.7 semitones. In Sentence 16, the fall from P1 to P2 is -0.6 semitones while the fall from P2 to P3 is -1.4 semitones. Jorden would treat the greater fall in pitch near the end of the major phrase as being a consequence of an 'intonation morph' (1955:27) which affects pitch at the end of a phrase. There seems to be no reason, however, to consider a fall in pitch in one location 'phonemic' and a similar fall in another location a matter of 'alternants'. We only need speak of 'phonemes' in respect to the High and Low tones which are a consequence of Japanese lexical accent. In their realization, these tones take a whole range of pitch values, but the actual values

---

[17] This does not apply to the three-word sentences.

they take are a consequence of intonation and the systematic variation in pitch over major phrases is best described in terms of phonetic rules of the sort suggested by Chew (in Chapter 4.2 above).

Chew's description of the intonation of utterances in Japanese is, in the main, vindicated by the acoustic evidence. It is true that pitch gradually declines from the beginning of the intonation phrase to the end. In Chapter 6 below, we will discuss the use Chew makes of intonation phrase boundaries in his 'emphasis transforms'. For now, it should be pointed out that the occurrence of an intonation phrase boundary in our sentences did not necessarily involve complete resetting of the line of declination as it does in Chew's description. It is often the case in our data that the utterance-initial F0 peak is not the highest peak in the utterance (Sentences 3, 22, 25, 34, 35, and 36) and, more often still, a rise in pitch from one peak to another does not necessary mean that the second peak will reach the level of the utterance-initial peak (Sentences 4, 5, 18, 19, 20, 28, 29). The fact of the rise from one peak to the next is enough to signal the phrase boundary; the amount of the rise probably depends on other things. To put it another way, there is a categorical distinction between falling pitch as a part of a intonation phrase contour, and non-falling (i.e. level or rising) pitch as a signal for the start of a new intonation phrase. The realization of non-falling pitch is gradient--the greater the ultimate rise, the 'stronger' the intonation phrase boundary being signalled and the greater the prominence of the second phrase.[18] Chew's requirement that the pitch rise to utterance-initial level after an intonation phrase boundary is incorrect.

Haraguchi (1977) errs in the other direction. His formulation of the Downdrift Rule (p 30) requires, except in the most deliberate speech, that there be _no_ pitch rises within the intonation phrase, not even the accentual rises from the first to the second mora in each minor phrase. It might be possible to argue about what constitutes a 'natural' rate of speech--this is the rate in which Haraguchi's ever lowering Downdrift Rule applies--and about whether the samples of speech which form the basis for this thesis were spoken at this rate, but about one thing Haraguchi is entirely wrong. He claims that the application of his Initial Lowering Rule (p 31ff), which accounts for the rise in pitch from the first to the second mora of a minor phrase, is 'crucially dependent on the presence of a pause in front of the initial mora [p 30].' He also implies that the presence of pauses is the defining characteristic of 'deliberate' speech. Many of the F0 traces of our sentences reveal that voicing continues, without cessation, from one F0 peak to the next, with a fall and then a rise in pitch in between but, obviously, no pause. Even when a voiceless consonant intervenes between

---

[18] 'Stronger' is used here in the same sense as in Chapter 3 above.

the peaks, the silent interval seems to be no longer than silent intervals associated with voiceless consonants which are not at minor phrase boundaries. It may very well be possible that pitch declination could operate to the extent that non-initial highs will be lowered to the level of immediately preceding lows[19] but there will be relatively rapid rates of (perhaps careful) speech where pitch declination still operates but does not have this extreme effect (of this we have numerous examples).

The non-acoustic evidence, except in some details, generally supports our hypothesis that pitch declination is a common intonation contour in Japanese, perhaps even the most common. The acoustic evidence is weighted in favour of this hypothesis as well. Sorensen and Cooper's studies have shown that the typical form of pitch declination in English has a high first peak, a considerable fall in pitch, and then a gradual descent from the second peak to the end of the utterance. Consequently initial peaks in English are above the declination line defined by the Topline Rule. As well, Sorensen and Cooper found that there was a correlation between the F0 of P1 and the time of utterance (Sorensen and Cooper 1980: 419). They found that the frequency of P1 was consistently higher in longer sentences. In Japanese, pitch

---

[19] In many of our sentences, there is no rise from the first to the second mora of the last minor phrase, and pitch simply declines from the initial mora in the phrase to the end of the sentence. This phenomenon does not occur in other positions and may be a consequence of a narrowing of the pitch range near the end of the utterance.

gradually declines from the first to the next-to-the-last peak and then shows a steep decline to the last peak. In our data, the F0 of the initial peaks only showed a significant correlation with time of utterance in three-word sentences ($r = 0.73$, $p < 0.001$). In longer sentences the two variables were not correlated. It may well be that the F0 of initial peaks is correlated with the length (time) of the intonation phrase rather than the length (time) of utterance. The three-word sentences in our data almost invariably consist of single intonation phrases, time of utterance is the same as time of intonation phrase and the F0 of the initial peaks and the time of utterance are correlated. The four and five-word sentences in our data tend to consist of more than one intonation phrase. If the F0 of P1 is affected by the length of intonation phrase rather than the length of utterance, the lack of correlation between PEAKFC and TIMEUTT is understandable.

Sorensen and Cooper (1981:96) report that they have recently completed a study of pitch declination in Japanese. The precise results of this study and the data on which they are based have not been published sc direct comparisons with the present study are not possible. Sorensen and Cooper do report that their Topline Rule did not result in a good characterization of pitch declination for their Japanese data. It appears that, though English and Japanese both exhibit the general feature of pitch declination, the form which pitch declination takes in each is different. They also found that the frequency of initial peaks was not correlated with length of utterance which indicates to them that the domain of pitch declination in Japanese is shorter than it is for English. Japanese speakers do not take the length of the entire sentence into account in determining the F0 of sentence initial peaks. Sorensen and Cooper's findings are consistent with those of the present study. In the following chapter we will discuss how pitch declination might be involved in the coding of prominence relations as well as in the the phrasing of Japanese utterances.

Chapter VI
CONCLUSION

There seems to be little doubt, on the basis of the acoustic evidence of this study and of the non-acoustic descriptions of the behaviour of pitch in Japanese sentences, that pitch declination exists in Japanese. It has the function of segmenting utterances into intonation phrases. It differs from pitch declination in English as formulated by Sorensen and Cooper (1981). It differs in both form and domain.

The form of pitch declination in English involves a steep initial fall followed by a gradual descent to the final peak in the phrase. In Japanese, pitch declines more steeply at the end of the phrase. There seems to be less correlation in Japanese between the level of the initial peak and the length of utterance than in English. In English many long sentences (consisting of a single clause) have declining pitch throughout. This, together with the observed correlation between the level of the first peak and the length of utterance, and the relative stability of the level of the final peak, led Sorensen and Cooper to speculate that the speaker must have look-ahead mechanism whereby he estimates the approximate length of the utterance and then computes the level of the first peak in order to determine the rate of declination. In Japanese, longer sentences are rarely characterized by a single line of declination, but are broken up into intonation phrases, each of which is characterized by its own line of declination. This suggests that the domain of pitch declination in Japanese is much shorter (cf Sorensen and Cooper 1981:97) than it is in English. This may be a consequence of the fact that pitch is a highly salient feature of lexical items in Japanese. Because Japanese is a pitch accent language which makes use of pitch at the word level to signal accentual distinctions, an intonationally coded decline in pitch which narrowed the pitch range over the length of utterance might tend to make the realization of word level pitch patterns more difficult toward the end of the utterance.[20] The shorter domain of pitch declination in Japanese tends to preserve the accentual patterns of the lexical items.

Sorensen and Cooper believe that pitch declination is syntactically coded in English. Since its domain is very often the sentence, or clause, or some other syntactically definable unit, this seems plausible. In Japanese, because

---

[20] In fact, the F0 traces of our sentences reveal that the verbs, which occur at the end of our sentences, very seldom retain their accentual pitch patterns. Often the syllable which has the highest pitch is the first even though the accent is elsewhere. The accentuation of verbs is much more regular than that of nouns so their accent patterns have a lower functional load and this levelling of their pitch contours does not impede intelligibility.

of its shorter domain, pitch declination is often the characteristic intonation contour of parts of utterances, parts which are often difficult to define syntactically. Consider Sentence 22 as spoken by subject TI:

asita otoko wa azuki o taberu 'tomorrow the man will eat red beans' This sentence contains a pitch rise from the peak in asita to the peak in otoko wa. Our claim is that this rise in pitch signals an intonation phrase boundary. The two intonation phrases in this sentence can be syntactically defined; they are both constituents of the sentence. asita is a sentence modifier, the rest of the sentence is the main clause consisting of the subject (marked in this case with the Topic particle wa), the direct object and the verb. Sentence 5 is very similar in structure to Sentence 22 above, except that it is made longer by the inclusion of an indirect object noun phrase:

kyoo otoko wa kusa o uma ni ataeru 'today the men feed hay to the horses'. This sentence, as spoken by TI, has a rise in pitch from the peak in otoko wa to the peak in kusa o. If this rise in pitch signals the presence of an intonation phrase boundary, it is difficult to see what sort of syntactic constituent a sentence modifier, asita and a subject noun phrase, otoko wa would be, and yet they are the sole elements in the first intonation phrase. There are many examples in our data of sentences whose intonation phrases are not easily definable syntactic constituents of the sentence. This leads us to the question of what function pitch declination performs in Japanese utterances.

In Chapter 3, we cited Gardiner's definition of the primary function of intonation. He states (Gardiner 1977, cited in Ladd 1980:163) that the primary function of intonation is to segment the stream of speech into phrases <u>and</u> to signal to what extent the phrases are related to one another and what element within the phrase is the center of attention. Recall also that Gardiner (1980) claims that the phrases identified by intonation do not have to be syntactically well-formed. The function of intonation, manifested as pitch declination, which we wish to focus on here is that of the signalling of prominence relationships between and within phrases in utterances.

Chew, in Section 14 of his monograph (1961), claims that there are two ways of putting emphasis on a sentence element (minor phrase) in Japanese. One is to change the order of the elements. The other is to insert an intonation phrase boundary directly to the left of the element. This would make it the first element in an intonation phrase characterized by declination. According to Chew, 'there is no order of...elements that does not put some slight emphasis on at least one of the elements [p 89].' In the 'basic order' there is slight emphasis on the first element of the phrase which contains the verb. Warkentyne (1978) makes a similar claim. Chew gives the following example (/ marks the intonation phrase boundaries):

kinoo / gakkoo de / masayuki ga / hon o nakusita 'masayuki lost his book at school yesterday'. In this sentence there is slight emphasis on hon o 'book'. If the word order is changed to a less basic one so that another element is in place of hon o, as in

kinoo / masayuki ga / hon o / gakkoo de nakusita

'masayuki lost his book at school yesterday', there will be special emphasis on gakkoo de which is now the first element in the phrase containing the verb. Moving an element to the beginning of the sentence and marking it with the wa particle de-emphasizes it to the status of a topic.[21]

Without changing the word order, emphasis can be shifted to the left in a sentence by removing the intonation phrase boundaries between the focussed element and the verb, or by placing an intonation phrase boundary in front of the focussed element if there was not one there before (as long as this boundary marks the start of the intonation phrase which contains the verb). Thus, word order and intonation phrase boundary placement interact to put prominence on particular sentence elements. The most important element within an intonation phrase is the leftmost; the most important element in the sentence is the leftmost element in the into-

---

[21] Chew points out (p 90) that the strongest form of de-emphasis is the movement of an element to a position at the end of the sentence, but after the verb. This de-emphasizes it to the status of an 'afterthought'. It is interesting to note that final position is a strong position in English, while both final and initial position are weak in Japanese.

nation phrase which contains the verb. In both cases it is this element which has the highest F0 peak within its intonation phrase--pitch declines from this peak to the end of the phrase.[22] Since there is a preferred position for prominence in a Japanese sentence, the peak on the element occupying this position does not have to be the highest peak in the sentence. If the whole sentence is one intonation phrase, the first peak <u>will</u> be the highest. If the sentence is made up of two, or more, intonation phrases, then the break in the line of declination which signals the boundary before the phrase which contains the verb will be sufficient to indicate prominence on the first element in that phrase. Degrees of prominence can be indicated by the amount of declination resetting at the boundary. One would assume that different, common or uncommon, word orders could achieve similar results. The interplay of intonation and word order in achieving prominence has been reported in other languages. Schubiger (1964) describes this interaction in German, which has a relatively free word order, and in English, which has a relatively fixed word order. Contreras (1980) discusses the interaction of word order and sentential

---

[22] As a preliminary to the analysis of the sentence data in this study, the individual lexical items were recorded in citation form. Although there are no statistics to verify it, the impression is that lexical items in citation form take pitch levels which are approximately those of the items which occur to the left of the verb in the sentence data. Lehiste and Ivic (1978:108) report the same phenomenon for Serbocroatian. Individual lexical items uttered in isolation take sentence prominence by default.

stress in Spanish.

Considerably more study of the intonation of Japanese sentences will have to be undertaken, before the exact form and function of pitch declination in Japanese can be established. That it characterizes at least some sentence types in Japanese is indicated by the present study. The possible role of pitch declination in signalling prominence relations will not be made clear without careful acoustic research coupled with perceptual tests of perceived prominence. As to the claims for the universality of pitch declination in language, we can say in general terms that Japanese, like English, and many other languages, has a tendency toward declining pitch in neutral declarative utterances, but that the form of pitch declination in Japanese is different in detail from the form of pitch declination in English. Any feature which is purported to be a universal characteristic of all languages will always differ in detail from language to language.

BIBLIOGRAPHY

ABERCROMBIE, David. 1967.
Elements of General Phonetics. Edinburgh: Edinburgh University Press.

ANDERSON, Stephen R. 1978.
In V. Fromkin 1978. P. 133-75.

BOLINGER, Dwight L. 1964.
Intonation as a Universal. In, Horace G. Lunt, ed., Proceedings of the Ninth International Congress of Linguists (1962). The Hague: Mouton. P. 833-48.

BOLINGER, Dwight L. 1978.
Intonation Across Languages. In, J. Greenberg, ed., Universals of Human Language, Vol. 2--Phonology. Stanford: Stanford University Press. P. 471-524.

CHEW, John Jr. 1961.
A Transformational Analysis of Modern Colloquial Japanese. Ann Arbor: University Microfilms.

COHEN, A. and 'T HART, J. 1967.
On the Anatomy of Intonation. Lingua Vol. 19:177-192.

CONTRERAS, Heles. 1980.
Sentential Stress, Word Order, and the Notion of Subject in Spanish. In, Linda R. Waugh and C. H. van Schooneveld, eds., The Melody of Language: Intonation and Prosody. Baltimore: University Park Press. P. 45-54.

COOPER, William E. and Sorensen, John M. 1977.
Fundamental Frequency Contours at Syntactic Boundaries. Journal of the Acoustical Society of America. Vol. 62:3:683-92.

DANEŠ, Frantisek. 1960.
Sentence Intonation from a Functional Point of View. Word, Vol. 16:1:34-54.

FJ ELECTRONICS. no date.
Manual for Fundamental Frequency Meter. Gentofte, Denmark.

√ FROMKIN, Victoria A. (ed.). 1978.
Tone: A Linguistic Survey. New York: Academic Press.

GARDINER, Duncan B. 1977.
Two Assumptions in the Study of Intonation. Unpublished paper read at LSA Annual Meeting, Chicago. Cited in Ladd 1980.

√ GARDINER, Duncan B. 1980.
Intonation and Music: The Semantics of Czech Prosody. Physsardt Series in Prague Linguistics 2 (Series eds. Daniel Armstrong and C. H. van Schooneveld). Bloomington: Physsardt.

GLEASON, H. A. 1961.
Review of Guthrie: African Language Studies I, and Richardson: The Role of Tone in the Structure of Sukura. Language Vol. 37:2:294-308.

HADDING-KOCH, Kerstin. 1961.
Acoustico-Phonetic Studies in the Intonation of Southern Swedish. Lund: Gleerup.

√ HARAGUCHI, Shosuke. 1977.
The Tone Pattern of Japanese: An Autosegmental Theory of Tonology. Tokyo: Kaitakusha.

√ HOMBERT, Jean-Marie. 1978.
Consonant Types, Vowel Quality, and Tone. In V. Fromkin 1978. P. 77-111.

HOUSE, Arthur S. and FAIRBANKS, Grant. 1953.
The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels. Journal of the Acoustical Society of America, Vol. 25:105-13.

JORDEN, Eleanor Harz. 1955.
The Syntax of Modern Colloquial Japanese. Language, Vol. 31-1 (Part 3 - Supplement, Dissertation No. 52).

LADD, D. Robert Jr. 1980.
The Structure of Intonational Meaning: Evidence from English. Bloomington: Indiana University Press.

LEHISTE, Ilse. 1970.
Suprasegmentals. Cambridge Mass.: MIT Press.

LEHISTE, Ilse and IVIĆ, Pavle. 1978.
Interrelationships Between Tone and Sentence Intonation in Serbocroatian. In, Donna Jo Napoli, ed., Elements of Tone, Stress, and Intonation. Washington: Georgetown University Press. P. 100-28.

LEHISTE, Ilse and PETERSON, G.E. 1961.
Some Basic Considerations in the Analysis of Intonation. Journal of the Acoustical Society of America Vol. 33:419-25.

LIEBERMAN, Philip. 1967.
Intonation, Perception and Language. (Research Monograph No. 38). Cambridge Mass.: MIT Press.

LIEBERMAN, Philip. 1980.
The Innate, Central Aspect of Intonation. In, Linda R Waugh and C. H. van Schooneveld, eds., The Melody of Language: Intonation and Prosody. Baltimore: University Park Press. P. 187-200.

McCAWLEY, James D. 1968.
The Phonological Component of a Grammar of Japanese. The Hague: Mouton.

McCAWLEY, James D. 1978.
What is a Tone Language? In V. Fromkin 1978. P. 113-31.

NIE, Norman H., HULL, C. Hadlai, JENKINS, Jean G., STEINBRENNER, Karin, and BENT, Dale H. 1975.
SPSS: Statistical Package for the Social Sciences, 2nd Edition. New York: McGraw-Hill.

✓OHALA, John J. 1978.
Production of Tone. In V. Fromkin 1978. P. 5-39.

PETERSON, G.E. and BARNEY, H.L. 1952.
Control Methods Used in the Study of Vowels. _Journal of the Acoustical Society of America_, Vol. 24:2:175-184.

✓PIERREHUMBERT, Janet. 1979.
The Perception of Fundamental Frequency Declination. _Journal of the Acoustical Society of America_, Vol. 66:2:363-69.

✓PIERREHUMBERT, Janet. 1980.
The Phonology and Phonetics of English Intonation. Unpublished dissertation. MIT.

✓PIERREHUMBERT, Janet. 1981.
Synthesizing Intonation. _Journal of the Acoustical Society of America_, Vol 70:4:985-95.

SCHAFER, Ronald W. and RABINER, Lawrence R. 1975.
Digital Representations of Speech Signals. _Proceedings of the Institute of Electrical and Electronic Engineers_, Vol. 63:662-77.

SCHUBIGER, Maria. 1964.
The Interplay and Cooperation of Word-Order and Intonation in English. In, David Abercrombie, et al, eds., _In Honour of Daniel Jones_. London: Longmans. P. 255-65.

✓SHIBATANI. Masayoshi. 1972.
The Non-Cyclic Nature of Japanese Accentuation. _Language_, Vol. 48:1:584-595.

SORENSEN, John M. and COOPER, William E. 1980.
Syntactic Coding of Fundamental Frequency in Speech Production. In, Ronald A. Cole, ed., _Perception and Production of Fluent Speech_. Hillsdale, New Jersey: Lawrence Erlbaum Associates. P. 399-440.

SORENSEN, John M. and COOPER, William E. 1981.
Fundamental Frequency in Sentence Production. New York: Springer-Verlag.

THORSEN, N. 1980.
A Study of the Perception of Sentence Intonation--Evidence from Danish. Journal of the Acoustical Society of America, Vol. 67:3:1014-30.

WARKENTYNE, H. J. 1978.
Information Focus and Primary Accent. The Study of Sounds, Vol. 13:217-222.

√WEITZMAN, Raymond S. 1970.
Word Accent in Japanese. Studies in the Phonology of Asian Language No. 9. Mieko S. Han, Series ed. Detroit: Management Information Services.

Appendix A

THE SENTENCE DATA

1) hito ga susi o taberu
People eat rice cakes

2) otoko ga isi de atama o tataku
The man strikes his head with a stone

3) asita otoko ga azuki o sukosi taberu
Tomorrow the man will eat some red beans

4) otoko ga isi o ike ni nageru
The man throws pebbles into the pond

5) kyoo otoko wa kusa o uma ni ataeru
Today the men feed hay to the horses

6) hito ga tamago o taberu
People eat eggs

7) uma ga hana o taberu
Horses eat flowers

8) otoko ga musume o tataku
The man strikes his daughter

9) otoko ga tama o musume ni nageru

The man throws a ball to his daughter

10) musume ga otoko no chikakuni imasu

The daughter is near the man

11) musume ga hasi kara ochiru

The daughter falls off the bridge

12) kyoo masaaki ga kusa o uma ni ataeru

Today Masaaki feeds hay to the horses

13) hito ga azuki o taberu

People eat red beans

14) musume ga tama o ukeru

The daughter catches the ball

15) otoko ga kawa no chikaku ni imasu

The man is near the river

16) musume ga otoko ni hana o ataeru

The daughter gives the man flowers

17) uma ga kusa o taberu

Horses eat hay

18) otoko ga musume ni isi o nageru

The man throws a stone at the girl

19) musume ga susi to azuki o taberu

The daughter eats red beans and rice cakes

20) kyoo otoko ga isi o ike ni nageru
Today the man throws pebbles into the pond

21) otoko ga kusa o uma ni ataeru
The man feeds hay to the horses

22) asita otoko wa azuki o taberu
Tomorrow the man will eat red beans

23) otoko ga uma o hanasu
The man releases the horses

24) isi ga iwa kara ochiru
The pebble falls from the large rock

25) asita hito wa azuki o taberu
Tomorrow everyone will eat red beans

26) otoko ga uma o tataku
The man strikes the horse

27) otoko ga atama o tataku
The man strikes his head

28) kyoo masaaki ga musume ni hano o ataeru
Masaaki gives flowers to his daughter today

29) otoko ga isi de musume o tataku
The man strikes his daughter with a stone

30) otoko ga musume ni hana o ataeru
The man gives flowers to his daughter

31) otoko ga musume ni susi o sukosi ataeru

The man gives his daughter a few rice cakes

32) hito ga sasimi o taberu

People eat raw fish

33) motto usiro ni iru

They are further back

34) asita kuru hitodesu

It's the man who is coming tomorrow

35) asita iku hitodesu

It's the man who is going tomorrow

36) kinoo gakkoo de masaaki ga hon o nakusita

Masaaki lost his book at school yesterday

37) ani no sasimi o kau

He buys my brother's fish

38) kinoo masaaki ga hon o gakkoo de nakusita

Masaaki lost his book at school yesterday

39) ani wa sasimi o kau

My brother buys fish

40) taroo wa kaimono ni ikimasita

Taroo went shopping

41) kaimono ni taroo wa ikimasita

Taroo went shopping

42) taroo wa koogai ni ikimasita

Taroo went to the suburbs

43) koogai ni taroo wa ikimasita

Taroo went to the suburbs

44) taroo wa kuukoo ni ikimasita

Taroo went to the airport

45) kuukoo ni taroo wa ikimasita

Taroo went to the airport

46) kare wa kanojo ni kaki o agemasita

He gave her a persimmon

47) kanojo ni kare wa kaki o agemasita

To her he gave a persimmon

48) taroo ga kaimono ni ikimasita

Taroo went shopping

49) kaimono ni wa taroo ga ikimasita

Taroo went shopping

50) kinoo watakusi wa ikimasita

I went yesterday

51) watakusi wa kinoo ikimasita

I went yesterday

52) tokyo e watakusi wa ikimasita

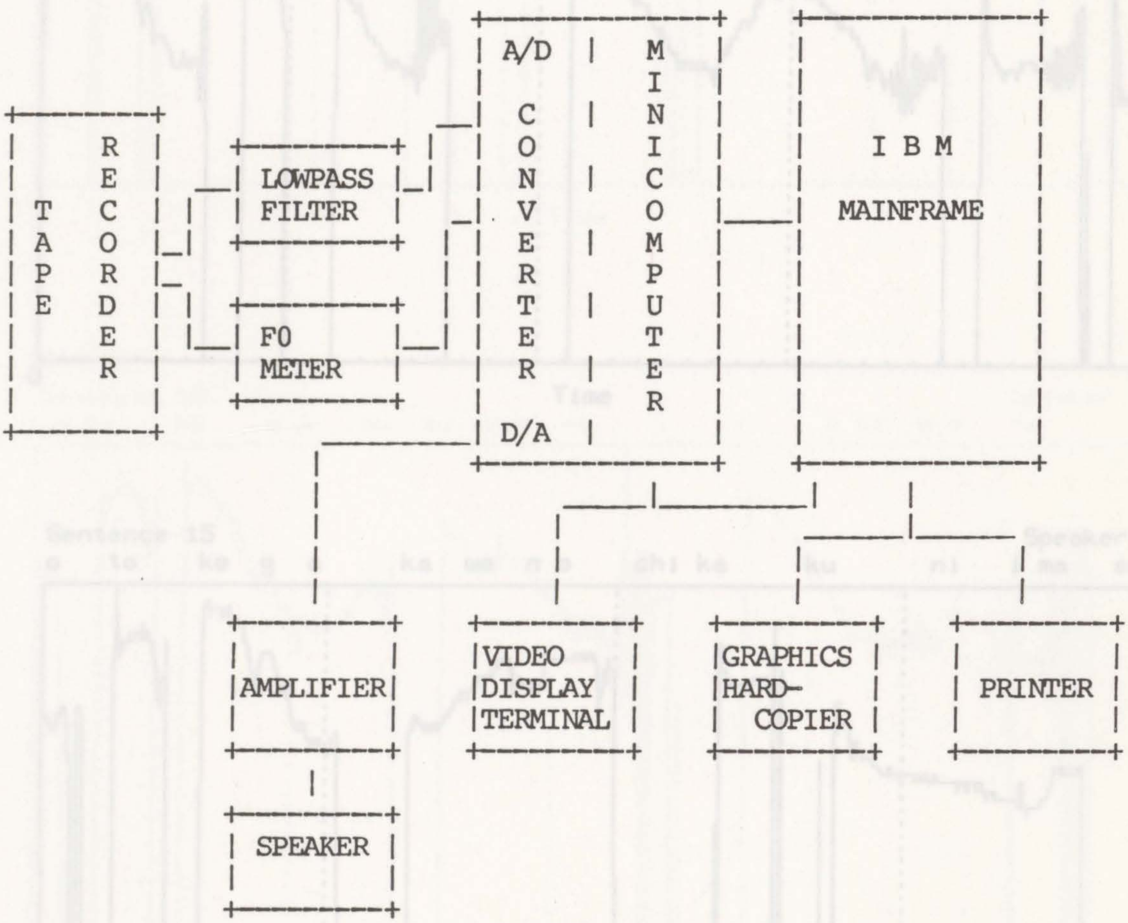
I went to Tokyo

53) watakusi wa tokyo e ikimasita

I went to Tokyo

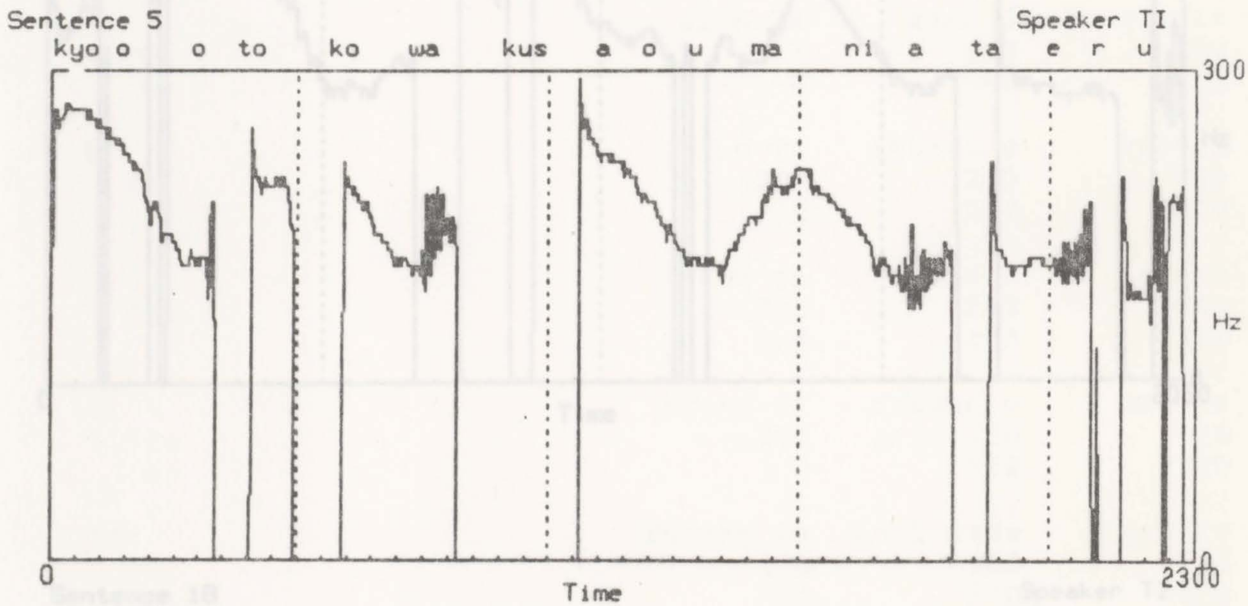
Appendix B

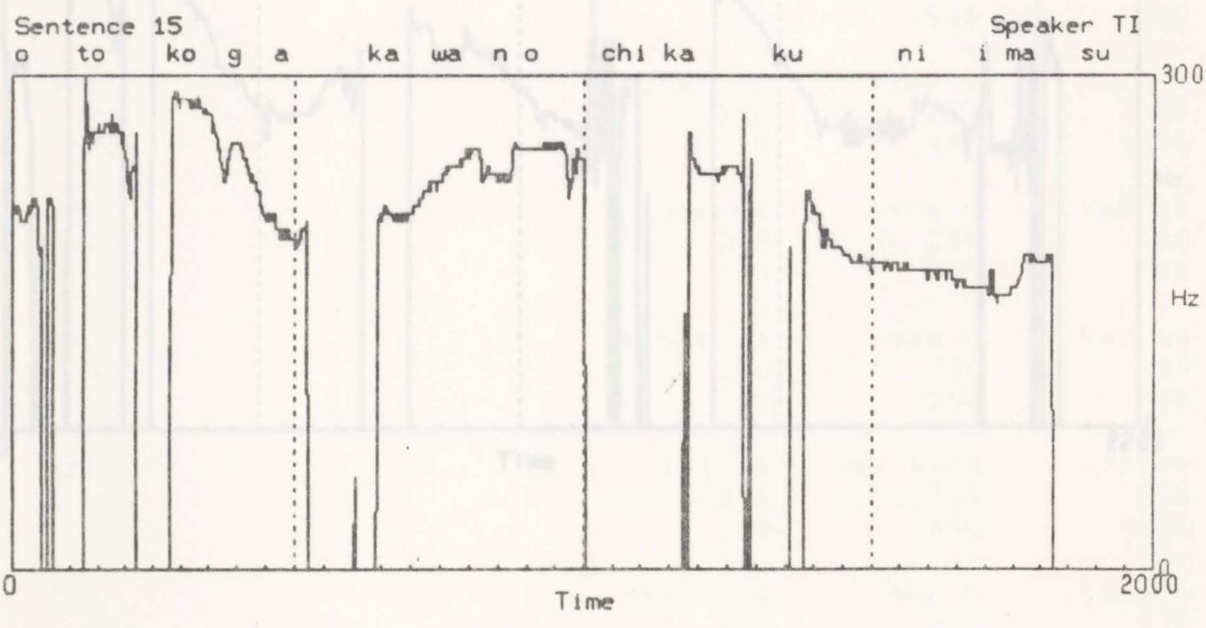
Schematic Diagram of Equipment Used in Measurement of F0.



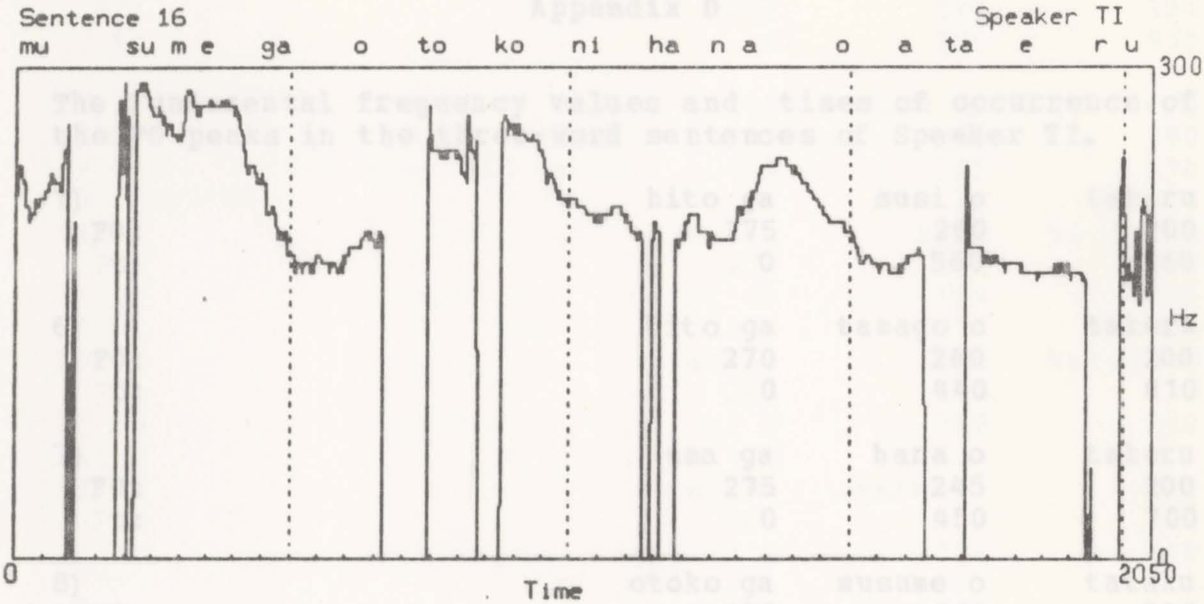
Appendix C

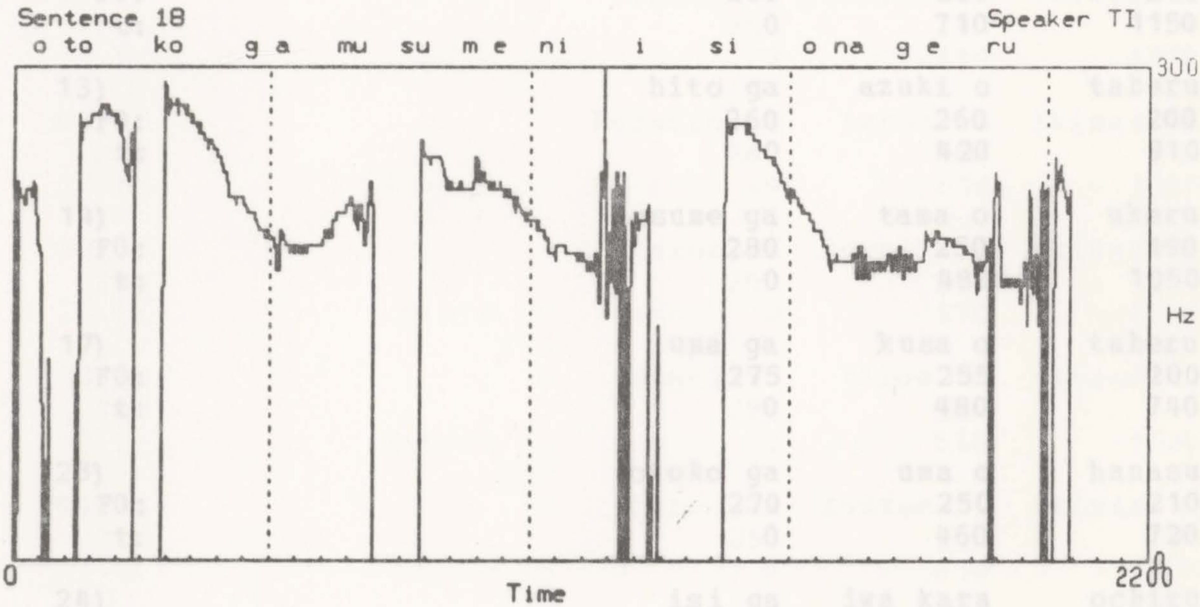
Sample F0 Traces.





Sample F0 Traces.





Appendix D

The fundamental frequency values and times of occurrence of the F0 peaks in the three-word sentences of Speaker TI.

1)	hito ga	susi o	taberu
F0:	275	260	200
t:	0	560	960
6)	hito ga	tamago o	taberu
F0:	270	260	200
t:	0	440	810
7)	uma ga	hana o	taberu
F0:	275	245	200
t:	0	450	700
8)	otoko ga	musume o	tataku
F0:	275	260	220
t:	0	720	1200
11)	musume ga	hasi kara	ochiru
F0:	285	250	200
t:	0	710	1150
13)	hito ga	azuki o	taberu
F0:	260	260	200
t:	0	420	910
14)	musume ga	tama o	ukeru
F0:	280	250	190
t:	0	480	1050
17)	uma ga	kusa o	taberu
F0:	275	255	200
t:	0	480	740
23)	otoko ga	uma o	hanasu
F0:	270	250	210
t:	0	460	720
24)	isi ga	iwa kara	ochiru
F0:	275	255	205
t:	0	400	1010
26)	otoko ga	uma o	tataku
F0:	270	250	210
t:	0	580	1000

27)	otoko ga	atama o	tataku
F0:	275	270	200
t:	0	680	1140
32)	hito ga	sasimi o	taberu
F0:	270	270	190
t:	0	510	990
33)	motto	usiro ni	iru
F0:	285	245	190
t:	0	570	930
34)	asita	kuru	hitodesu
F0:	275	290	205
t:	0	190	730
35)	asita	iku	hitodesu
F0:	245	250	230
t:	0	330	570
37)	ani no	sasimi o	kau
F0:	260	250	160
t:	0	620	940
39)	ani wa	sasimi o	kau
F0:	270	250	160
t:	0	600	1020
40)	taroo wa	kaimono ni	ikimasita
F0:	300	240	195
t:	0	710	1370
41)	kaimono ni	taroo wa	ikimasita
F0:	280	265	195
t:	0	610	1330
42)	taroo wa	koogai ni	ikimasita
F0:	295	275	200
t:	0	570	1360
43)	koogai ni	taroo wa	ikimasita
F0:	295	260	195
t:	0	660	1220
44)	taroo wa	kuukoo ni	ikimasita
F0:	295	245	200
t:	0	610	1370
45)	kuukoo ni	taroo wa	ikimasita
F0:	275	275	190
t:	0	580	1250

Appendix E

The fundamental frequency values and times of occurrence of the F0 peaks in the four-word sentences of Speaker TI. The oblique stroke (/) marks the proposed intonation phrase boundaries.

2)	otoko ga	isi de	/ atama o	tataku
F0:	275	225	250	200
t:	0	680	1200	1650
4)	otoko ga	isi o	/ ike ni	nageru
F0:	270	240	250	200
t:	0	670	1140	1460
9)	otoko ga	/ tama o	musume ni	nageru
F0:	275	275	235	185
t:	0	480	920	1470
10)	musume ga	otoko no	chikakuni	imasu
F0:	280	250	240	180
t:	0	410	1010	1340
15)	otoko ga	kawa no	chikakuni	imasu
F0:	285	260	245	190
t:	0	630	950	1470
16)	musume ga	otoko ni	hana o	ataeru
F0:	275	265	245	190
t:	0	550	1020	1380
18)	otoko ga	musume ni	/ isi o	nageru
F0:	275	245	265	200
t:	0	490	1080	1450
19)	musume ga	susi to	/ azuki o	taberu
F0:	280	205	265	185
t:	0	820	1180	1720
21)	otoko ga	/ kusa o	uma ni	ataeru
F0:	275	275	255	195
t:	0	440	880	1290
22)	asita	/ otoko wa	azuki o	taberu
F0:	265	280	250	190
t:	0	250	980	1400

25)	asita	/? hito wa	/ azuki o		taberu
F0:	255	255	260		205
t:	0	250	660		1110

29)	otoko ga	isi de	/ musume o		tataku
F0:	285	240	255		185
t:	0	430	950		1390

30)	otoko ga	musume ni	/ hana o		ataeru
F0:	285	250	250		190
t:	0	600	1090		1450

46)	kare wa	/ kanojo ni	kaki o		agemasita
F0:	245	270	250		220
t:	0	440	1230		1430

47)	kanojo ni	kare wa	kaki o		agemasita
F0:	300	260	245		210
t:	0	570	1170		1460

Appendix F

The fundamental frequency values and times of occurrence for all the F0 peaks in the five-word sentences of Speaker TI. The oblique stroke (/) marks the proposed intonation phrase boundaries.

3)	asita	/ otoko ga	azuki o	/ sukosi	taberu
F0:	240	260	240	255	205
t:	0	330	1050	1430	1660
5)	kyoo	otoko wa	/ kusa o	uma ni	ataeru
F0:	275	235	260	240	185
t:	0	400	1030	1450	1910
12)	kyoo	masaaki ga	kusa o	uma ni	ataeru
F0:	275	260	250	225	180
t:	0	480	1410	1820	2290
20)	kyoo	otoko ga	/ isi o	ike ni	nageru
F0:	280	250	265	250	190
t:	0	450	1110	1480	1830
28)	kyoo	masaaki ga	/ musume ni	hana o	ataeru
F0:	285	250	270	245	195
t:	0	410	1160	1690	1940
31)	otoko ga	musume ni	/? susi o	/ sukosi	ataeru
F0:	275	240	255	275	195
t:	0	510	1240	1640	1950
36)	kinoo	gakkoo de/	masaaki ga	hon o	nakusita
F0:	265	240	275	265	210
t:	0	490	1290	1870	2430
38)	kinoo	masaaki ga	/ hon o	gakkoo de	nakusita
F0:	270	265	265	220	200
t:	0	350	1060	1840	2280

Appendix G

The fundamental frequency values and times of occurrence for the F0 peaks of the three and four-word sentences of Speaker MW.

40)		taroo wa	kaimono ni	ikimasita
F0:		260	235	190
t:		0	760	1550
41)		kaimono ni	taroo wa	ikimasita
F0:		250	230	190
t:		0	620	1160
42)		taroo wa	koogai ni	ikimasita
F0:		250	250	175
t:		0	690	1370
43)		koogai ni	taroo wa	ikimasita
F0:		265	220	185
t:		0	720	1290
44)		taroo wa	kuukoo ni	ikimasita
F0:		260	220	180
t:		0	690	1400
45)		kuukoo ni	taroo wa	ikimasita
F0:		260	225	175
t:		0	770	1150
46)	kare wa	kanojo ni	kaki o	agemasita
F0:	240	235	220	190
t:	0	550	1260	1460
47)	kanojo ni	kare wa	kaki o	agemasita
F0:	265	215	220	165
t:	0	500	930	1440

Appendix H

The fundamental frequency values and times of occurrence for the F0 peaks of all the sentences of speaker MF.

40)	taroo wa	kaimono ni	ikimasita
F0:	185	185	145
t:	0	990	1790
41)	kaimono ni	taroo wa	ikimasita
F0:	185	190	125
t:	0	870	1450
48)	taroo ga	kaimono ni	ikimasita
F0:	195	165	130
t:	0	930	1550
49)	kaimono ni wa	taroo ga	ikimasita
F0:	185	200	115
t	0	1140	1870
50)	kinoo	watakusi wa	ikimasita
F0:	175	170	145
t:	0	470	1290
51)	watakusi wa	kinoo	ikimasita
F0:	180	195	160
t:	0	960	1160
52)	tookyoo e	watakusi wa	ikimasita
F0:	195	165	130
t:	0	900	1680
53)	watakusi wa	tookyoo e	ikimasita
F0:	180	210	145
t:	0	930	1770

VITA

SURNAME: Gallaiford GIVEN NAME: Neil

PLACE OF BIRTH: Bradford, England

DATE OF BIRTH: April 5, 1951

EDUCATIONAL INSTITUTIONS ATTENDED:

University of Alberta, Edmonton	1974 to 1979
University of Victoria, B.C.	1979 to 1983

DEGREE AWARDED:

B. Education 1979 University of Alberta, Edmonton

HONOURS AND AWARDS:

University of Victoria Fellowship, 1980/81

PUBLICATIONS:

'The Effect of Downdrift on Major Phrase Intonation in Japanese' Working Papers of the Linguistic Circle, Vol. 3:1:37-44, February, 1983.