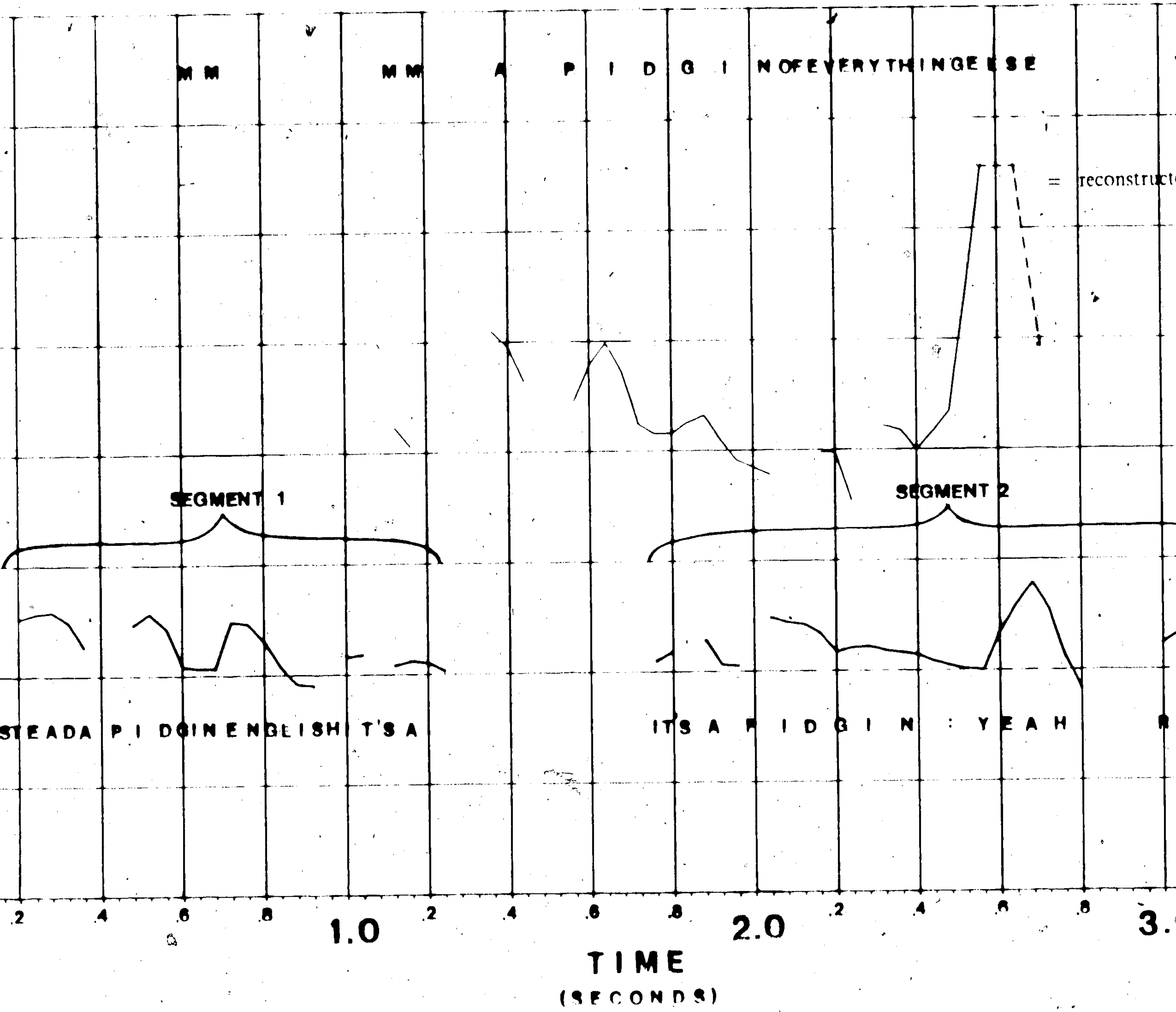
ure 4.1 F_0 plot of Example 1 with lexical text and segments. ——— = D's F_0-series ——— = S's F_0-se



CANADIAN THESES

THÈSES CANADIENNES

NOTICE

The quality of this microfiche is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

AVIS

La qualité de cette microfiche dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, examens publiés, etc.) ne sont pas microfilmés.

THIS DISSERTATION HAS BEEN MICROFILMED EXACTLY AS RECEIVED

LA THÈSE A ÉTÉ MICROFILMÉE TELLE QUE NOUS L'AVONS REÇUE

TC -
ISBN 0-315-23207-2

CANADIAN THESES ON MICROFICHE SERVICE – SERVICE DES THÈSES CANADIENNES SUR MICROFICHE

PERMISION TO MICROFILM – AUTORISATION DE MICROFILMER

• Please print or type – Écrire en lettres moulées ou dactylographier

AUTHOR – AUTEUR

Full Name of Author – Nom complet de l'auteur: Eugene Hauck Buder, Jr.

Date of Birth – Date de naissance: 21 April 1957

Canadian Citizen – Citoyen canadien: ☐ Yes Oui ☒ No Non

Country of Birth – Lieu de naissance: U.S.A.

Permanent Address – Résidence fixe: 50 Fair Oaks, St. Louis, Mo. U.S.A. 63124

THESIS – THÈSE

Title of Thesis – Titre de la thèse: Coherence of Speech Rhythms in Conversations: Autocorrelation Analysis of Fundamental Voice Frequency

Degree for which thesis was presented / Grade pour lequel cette thèse fut présentée: M. Ed.

Year this degree conferred / Année d'obtention de ce grade: 1985

University – Université: University of Alberta

Name of Supervisor – Nom du directeur de thèse: Dr. Carl A. Union

AUTHORIZATION – AUTORISATION

Permission is hereby granted to the NATIONAL LIBRARY OF CANADA to **microfilm** this thesis and to **lend** or **sell copies of the film.**

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

L'autorisation est, par la présente, accordée à la BIBLIOTHÈQUE NATIONALE DU CANADA de **microfilmer** cette thèse et de **prêter** ou de **vendre des exemplaires du film.**

L'auteur se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans l'autorisation écrite de l'auteur.

ATTACH FORM TO THESIS – VEUILLEZ JOINDRE CE FORMULAIRE À LA THÈSE

Signature: Eugene H. Buder, Jr.

Date: 15 October 1985

THE UNIVERSITY OF ALBERTA

COHERENCE OF SPEECH RHYTHMS IN CONVERSATIONS: AUTOCORRELATION ANALYSIS OF FUNDAMENTAL VOICE FREQUENCY

by

(C) EUGENE HAUCK BUDER, JR.

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF EDUCATION

IN

ANTHROPOLOGY AND EDUCATION

DEPARTMENT OF EDUCATIONAL FOUNDATIONS

EDMONTON, ALBERTA

FALL 1985

re 4.6 F_0 plot of Example 1 with waveform model. (wave symbol) = 260 msec wave

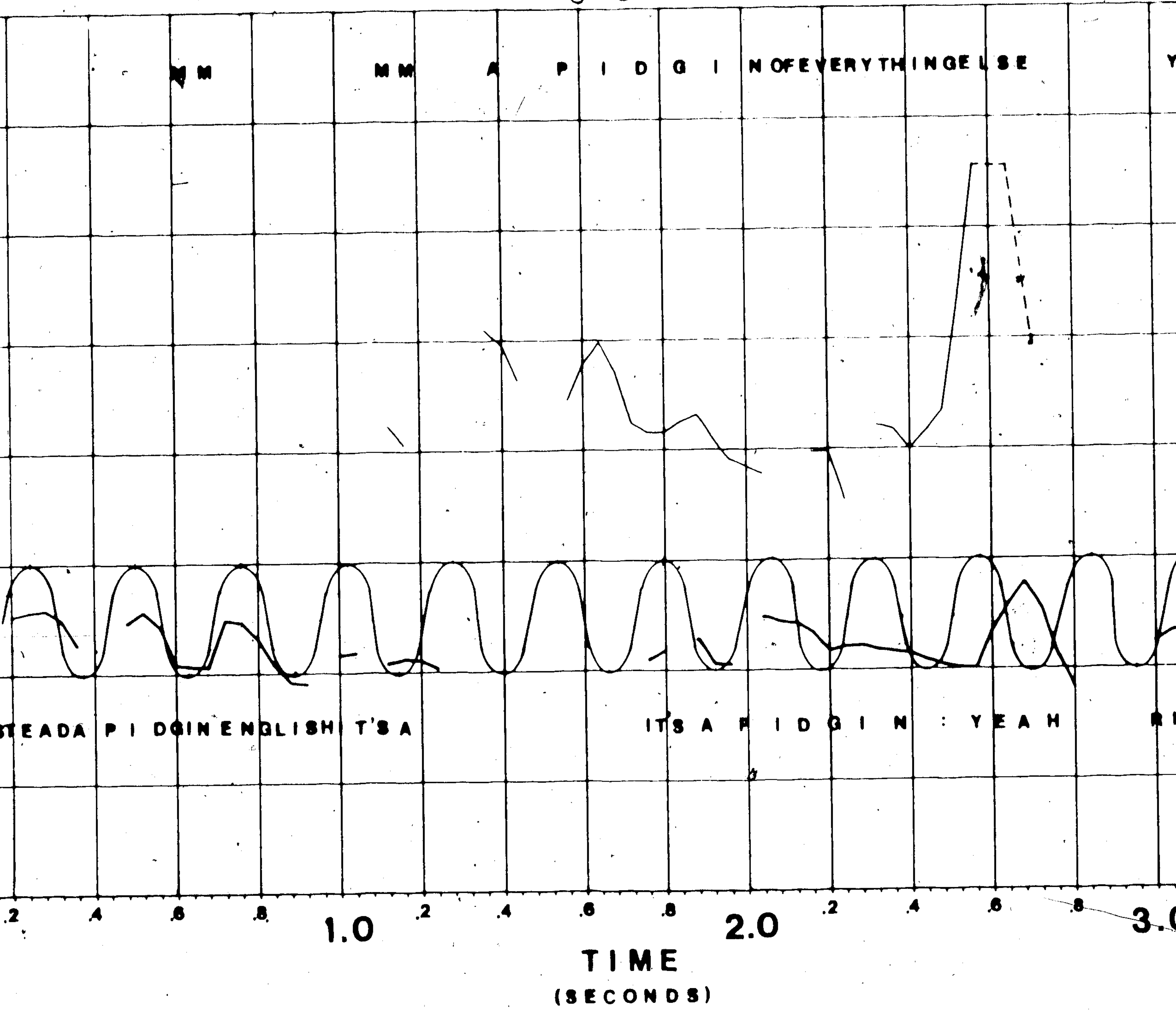
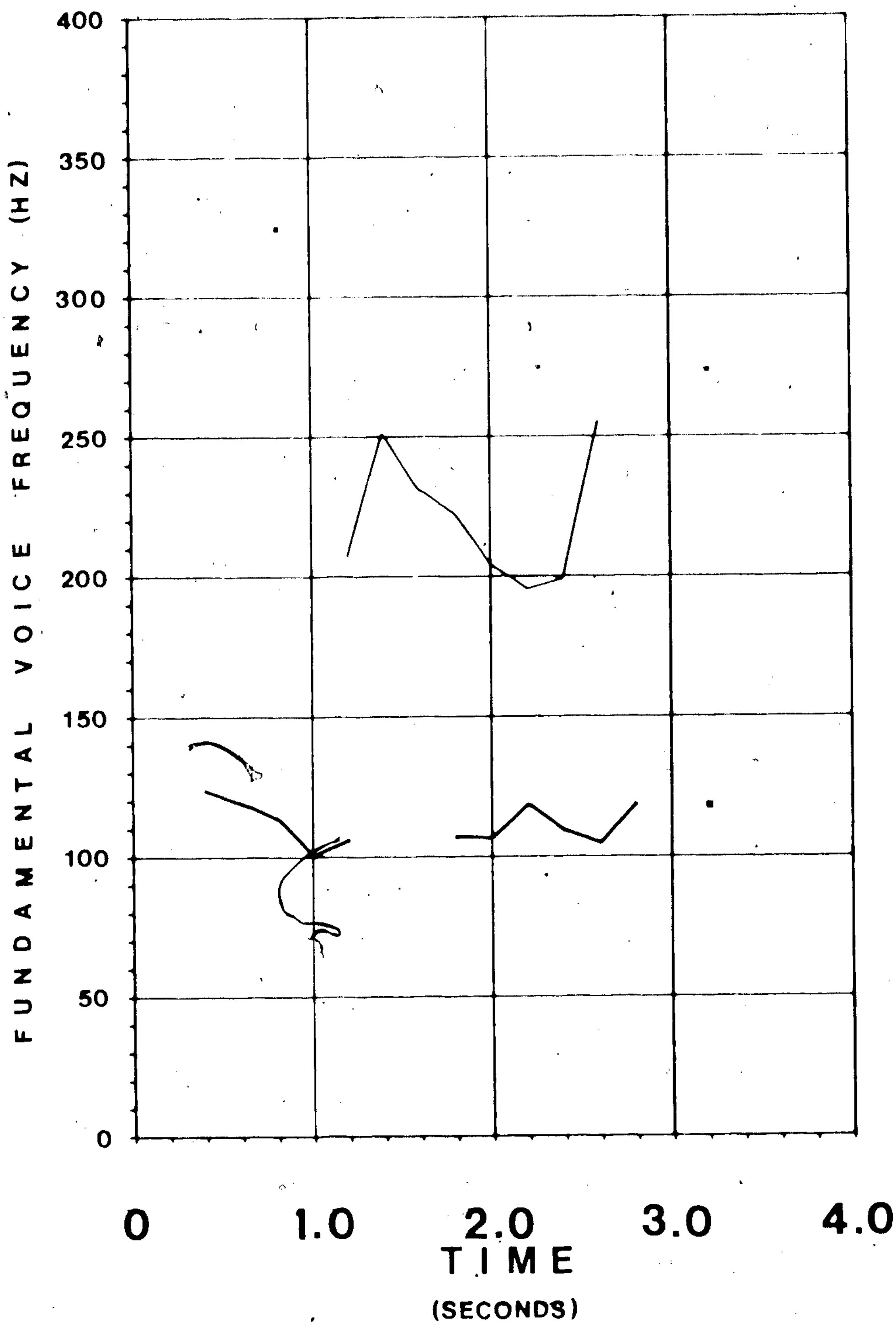
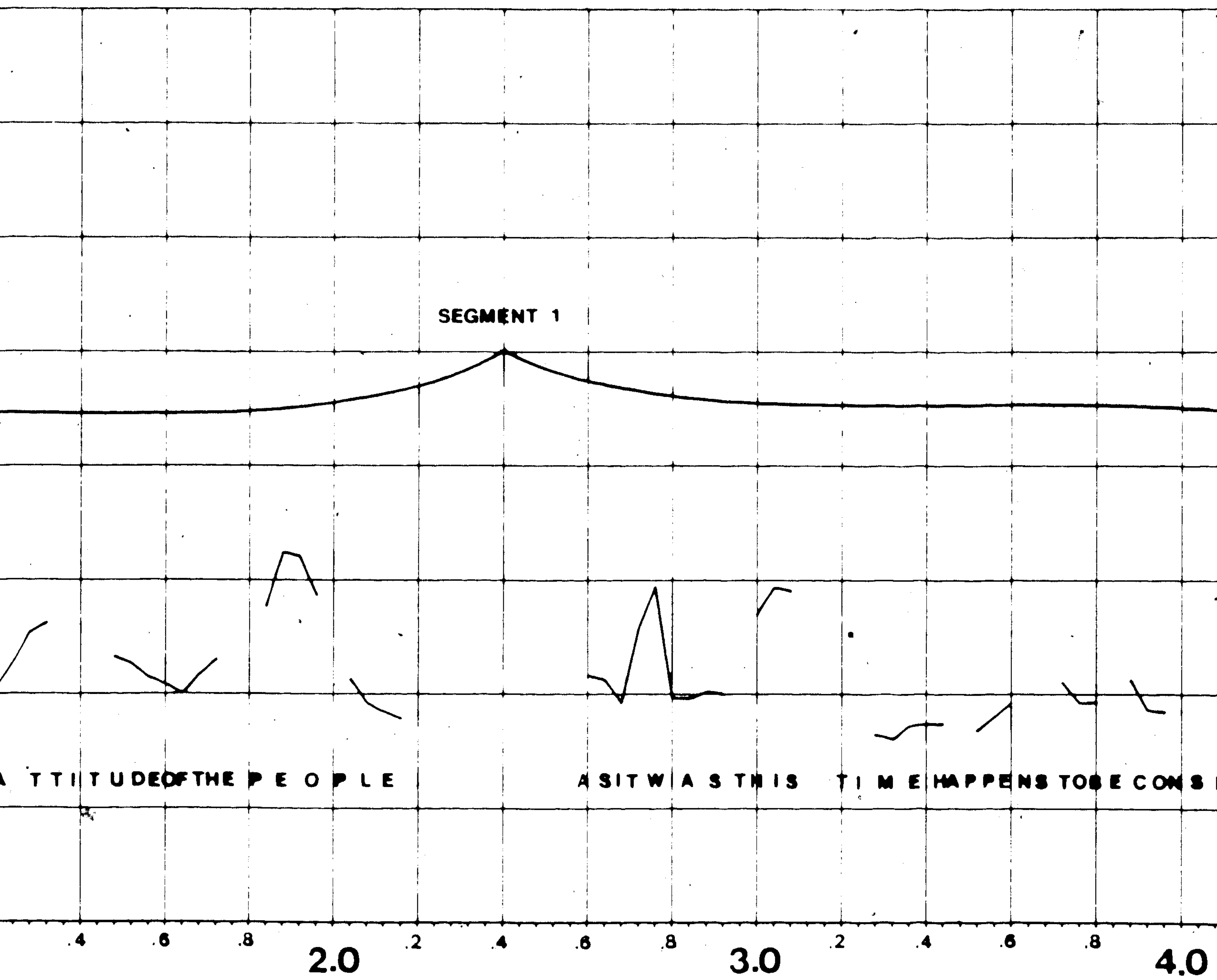


Figure 4.8 F_0 plot of Example 1 at 5 samples per second.



text and segments. ——— = D's F_0-series ——— = J's F_0-series



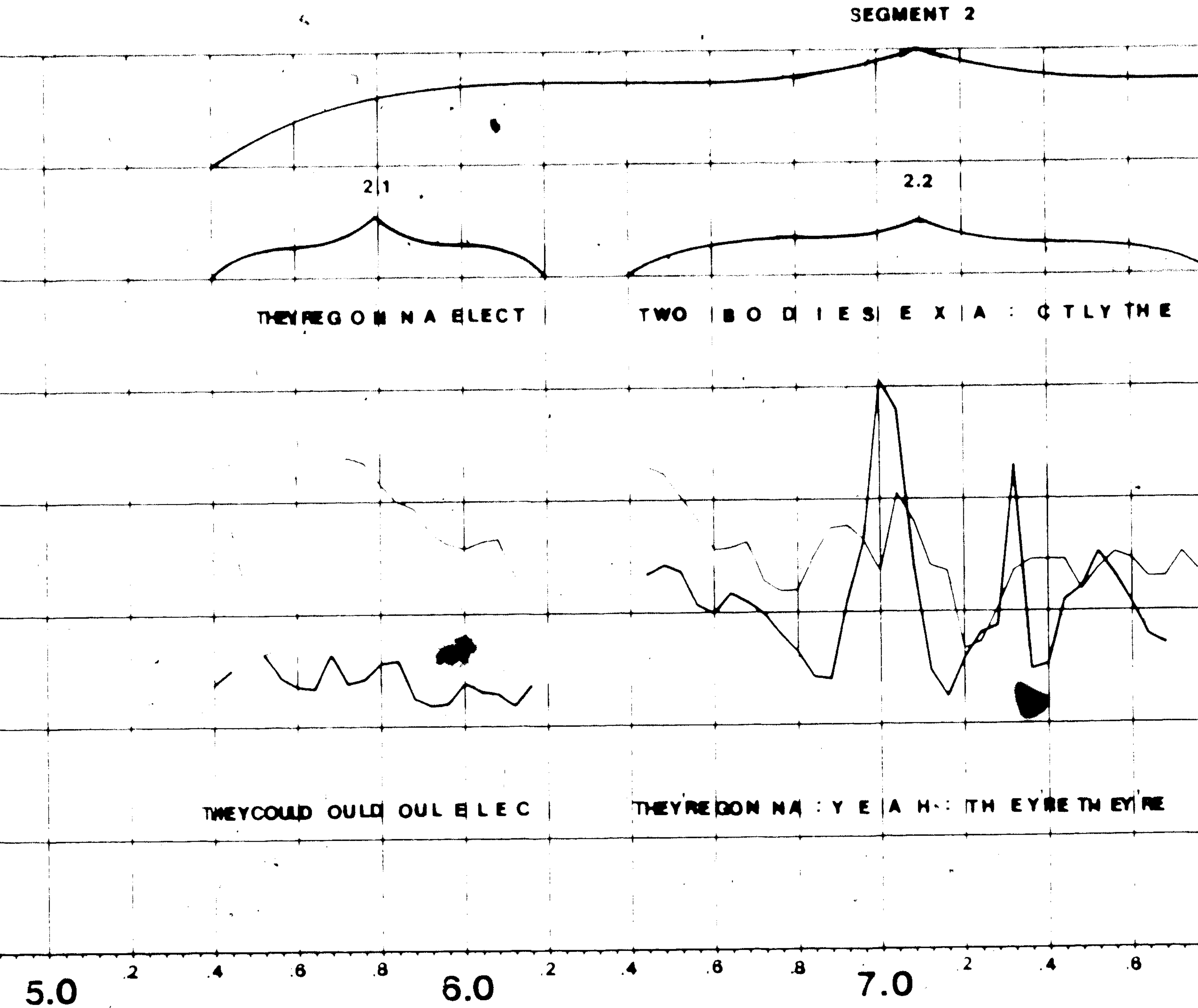
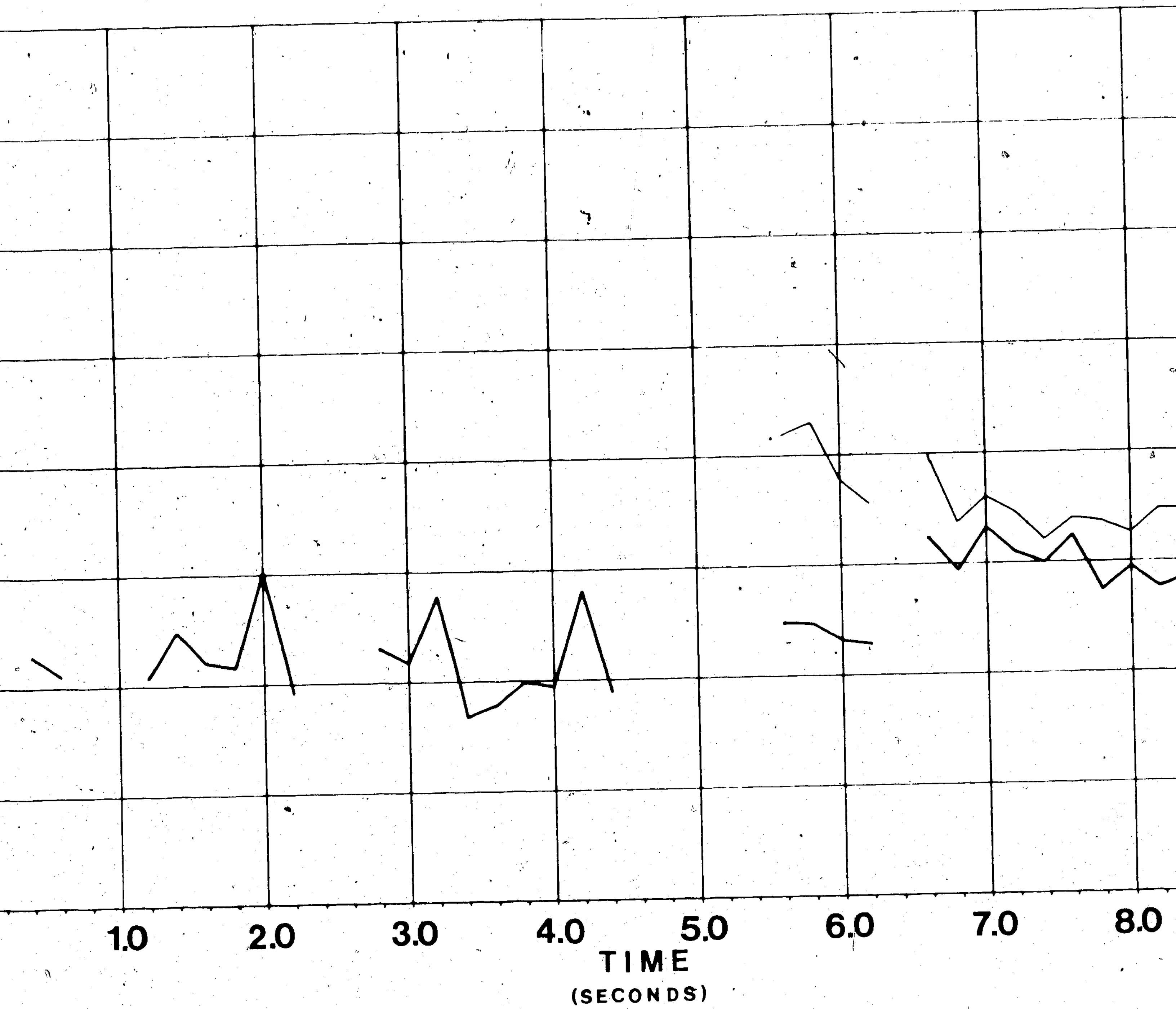
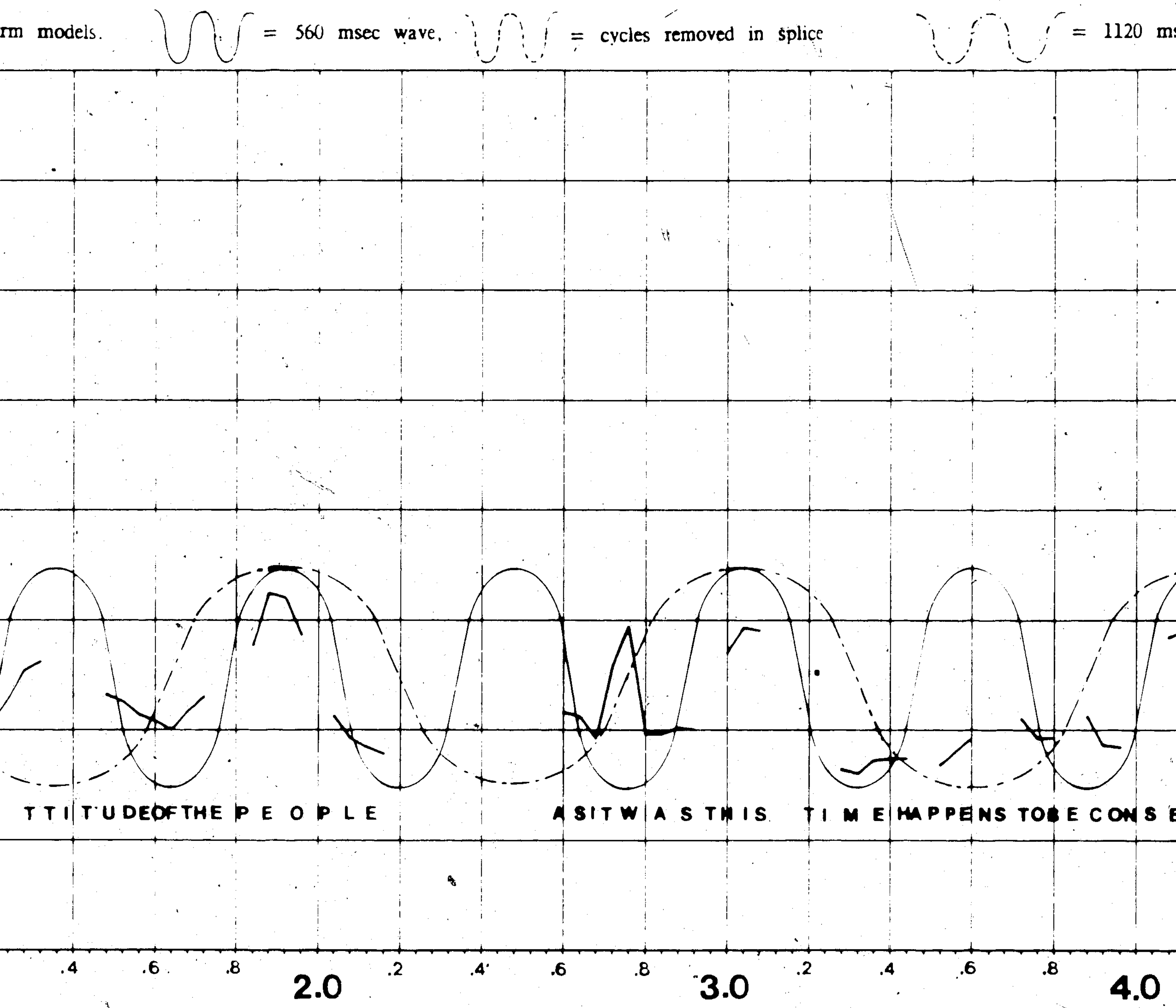
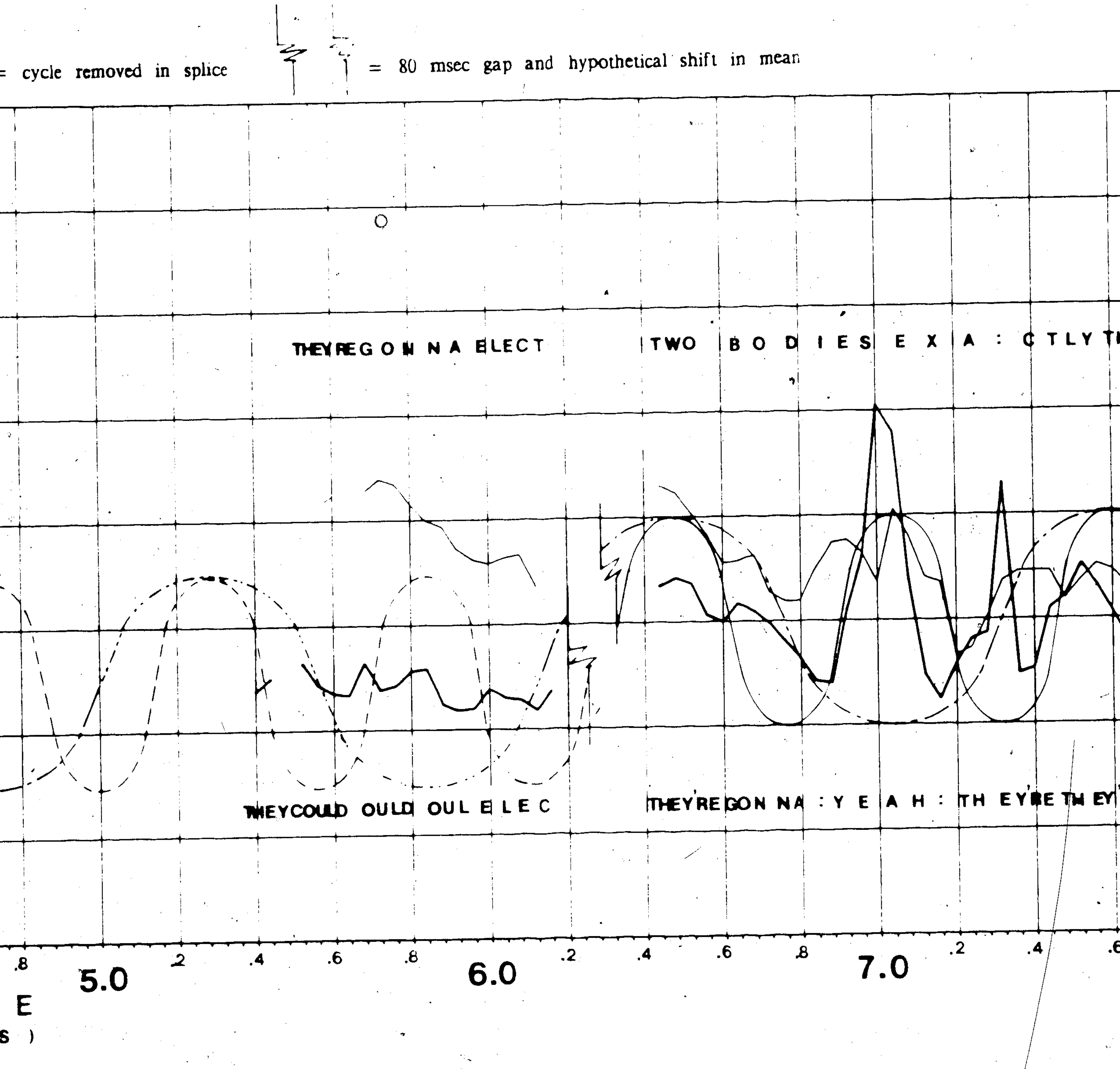

SEGMENT 2
2.1
2.2
THEYREGONNA ELECT
TWO BODIES EXACTLY THE
THEYCOULD OULD OUL ELEC
THEYREGONNA : YEAH : THEYRE THEYRE
5.0
.2
.4
.6
.8
6.0
.2
.4
.6
.8
7.0
.2
.4
.6

Figure 4.13 F, plot of Example 2 at 5 samples per second.




rm models.
= 560 msec wave,
= cycles removed in splice
= 1120 ms
TTITUDEOFTHE PEOPLE
ASITWASTHIS TIME HAPPENS TOBE CONS E
.4 .6 .8 2.0 .2 .4 .6 .8 3.0 .2 .4 .6 .8 4.0


= cycle removed in splice
= 80 msec gap and hypothetical shift in mean
THEYREGONNA ELECT
TWO BODIES EXA:CTLYT
THEYCOULD OULD OUL ELEC
THEYREGONNA : YEAH : THEYRETHEY
.8
5.0
.2
.4
.6
.8
6.0
.2
.4
.6
.8
7.0
.2
.4
.6
E
S)

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR EUGENE HAUCK BUDER, JR.

TITLE OF THESIS COHERENCE OF SPEECH RHYTHMS IN CONVERSATIONS: AUTOCORRELATION ANALYSIS OF FUNDAMENTAL VOICE FREQUENCY

DEGREE FOR WHICH THESIS WAS PRESENTED MASTER OF EDUCATION

YEAR THIS DEGREE GRANTED FALL 1985

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

(SIGNED) ....

PERMANENT ADDRESS:

....

....

....

DATED 19

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled COHERENCE OF SPEECH RHYTHMS IN CONVERSATIONS: AUTOCORRELATION ANALYSIS OF FUNDAMENTAL VOICE FREQUENCY submitted by EUGENE H. BUDER, JR. in partial fulfilment of the requirements for the degree of MASTER OF EDUCATION in ANTHROPOLOGY AND EDUCATION.

....
Supervisor

....

....

John T. Hogan

Date 9th October 1985

Dedicated to

Eugene Hauck and Jutta Zelle Buder

Abstract

Adequate description of speech rhythm is required to make testable some claims regarding the prosodic coherence of conversations. This study is an attempt to demonstrate that time-series analysis of prosodic features in conversation is a research technique that may be used in testing specific hypotheses about conversational coherence. The purpose of the study was not to test such hypotheses, but to demonstrate that such hypotheses are testable with the technique.

In the methodology developed here, fundamental voice frequency (F_0) is extracted from conversational speech and examined for periodicity using autocorrelation analyses. Statistical tendencies towards periodicity in measured fluctations in F_0 are thereby demonstrable, and the prosodic coherence between speech segments can be described in terms of the period and phase relations between waveform models of these tendencies.

Two examples, from a body of data representing a number of recorded conversations, are presented in this study, and a working hypothesis (that simultaneous onset of locution between speakers occurs consistent with a period established by previous speech) is explored with these examples. Though "onset" was not discovered to be the salient physical event of simultaneity in both examples, the results demonstrated that the timing of simultaneous locutions is describable with reference to the periods indicated by the statistical technique. The study thus demonstrates the utility of time-series analysis applied to F_0.

Acknowledgements

My studies in relationships between language and music began at Harvard under the suggestion of Professor David Layzer and with the supervision of Professor Sheldon White. These men tolerated numerous erratic thoughts and [illegible] arts. I received invaluable tutoring in psychology and philosophy from Dr. Sheldon Wagner and Dr. John Daugman. I followed my interests in development and music by studying children's spontaneous rhythm notations and completing an undergraduate thesis under the careful guidance of Professor Jeanne Bamberger at the Massachusetts Institute of Technology. During this period I also had several opportunities to discuss biological frameworks for the description of conversation with Dr. Francisco Varela, and with Dr. Michael Moerman, at the Naropa Institute.

My thinking on human interaction was galvanized by an encounter with Drs. Michael Mair and Carl Urion, following a conference in New York City. Our meeting resulted in correspondence, travels, and a decision to pursue the analysis of conversation under the supervision and guidance of Professor Carl Urion at the University of Alberta.

Faults and failures in the progress of this methodology, and in the following report on that progress, are my own responsibility. What genuine progress has been made I attribute to my association with Dr. Urion, with Dr. Mair, and with Dr. Tom Maguire, Dr. John Hogan, and Dr. Anton Roszypal, for their generous and well-considered advice and assistance. I have also relied on expertise and support from Dr. Paul Bouissac, Dr. Phillipe Martin, Dr. Michael Rodda, Mr. Frederick Ulmer, Dr. Parth Bhatt, and Dr. Alan VanderWell. In completing the present study, I depended heavily on software and hardware developed by Dr. Dwight Harley of the Division of Educational Research Sevices at the University of Alberta, and Mr. Ray Wong of the Compu-Ware corporation in Edmonton, and also on the graphic skills of Mr. Paul Klimczak. I offer my thanks and appreciation to these and to the dozens of other friends, teachers, colleagues, and associates who were involved in furthering this project, with apologies to those not being named here.

Finally, I wish to thank my parents, to whom this thesis is dedicated, my grandmother Eugenia Hauck Buder, my sisters Annette, Beatrice, Stella, and brother-in-law Peter, for their continual support, and White Cloud, Greta, Laura, Julianna, Armand, Celeste, and Madeleine Urion, for their hospitality and fellowship.

Table of Contents

Chapter Page

List of Figures

1. Introduction

Description of language behavior requires the identification of descriptive and analytic units, such as "phone," "syllable," "locution," "utterance," and, comprehensively, "discourse" and "conversation." A major issue is the coherence of the component units, such that their integration into a "conversation" or "discourse" as one text, allows us to speak of these complex systems as complete and and bounded whole units, that is, to regard the conversation itself as an object of analysis. There are several analytic domains in which coherence can be found; for example, in syntax, the notion of anaphora can be seen to demonstrate one level of coherence in the text of a discourse.

There is another major issue when performance phenomena are considered. Attempts have been made based on speech act theory to demonstrate coherence of language units by appeal to their function in social settings; most studies of discoursal coherence rely heavily on notions developed in speech act theory (Coulthard 1977). A problem with such an approach is that, usually, properties internal to a speech act are explained, with reference to "effect" and "purpose," but coherence between speech acts is not demonstrable (Urion 1980, 1985).

A number of scholars have looked at prosody as a domain in which conversational coherence may be studied, for example Chapple (1980,1981), Mair (1977,1980), Urion (1978,1985), Byers (1976), Erickson & Shultz (1982), Scollon (1981), Ulmer (1983), Goldberg (1978), Menn and Boyce (1982), Jaffe, Anderson, and Stern (1979), Gregory and Hoyt (1982), Gumperz (1984), and others. The disciplines in which these scholars work range from anthropology, linguistics, sociology, and psychology, to ophthamology and neurology, and "prosody" in this context has a general meaning--it may refer to "suprasegmentals," "speech rhythm," "speech melody," and "intonation."

There are several problems with this approach. Probably the most compelling one is that most studies are impressionistic. They are valuable as initial statements of imperative to study the area, but the question usually arises as to whether or not the observer imposes a structure upon the data. Another question, often entered but not resolved, is whether or not the prosodic structures claimed to exist are perceptually salient, or fortuitous artifacts of the physiological aspects of speech. Yet another problem is that the components of speech that constitute "prosody" are extremely difficult to isolate and to define unambiguously: "stress," for example, is an obvious component in the establishment of speech rhythm, but its precise definition--in physical terms--is extremely difficult to make.

Those problems are taken account of, but the arguments not entered, in the present study. Instead, the problem is conceived of as attempting first to isolate one basic and fairly transparent measurable aspect of prosody. In this case, it is fundamental voice frequency, perceived as "pitch," (see, e.g. Fry [1968]) that is the object of description. The second component of the problem is to develop a way to take that aspect of the acoustic record of a conversation represented by the fundamental frequency (F_0) and to see whether or not its fluctation is indeed periodic, over time, between two speakers.

In this case, audiotape recorded conversations between two dyads are analyzed in the following way. A record of the fundamental frequency traces of the converations was made and a working hypothesis was formulated--that a period established in speech would predict a point at which two people begin speaking simultaneously. Time-series analyses of the F_0 values were made of instances in which simultaneous onset of speech was observed, and the implications of the method were explored.

1.1 Methodological Rationale

The rationale for this methodological exploration is as follows.

Though pitch and rhythm are recognized to have interactional relevance in a number of perspectives on discourse, there is a need for rigorous criteria in the description of prosodic phenomena in conversations. The role of rhythmic timing relations in maintaining the organization of processes of speech perception and production has received some clarification recently, supporting the notion that speech production in connected discourse is timed according to perceived timing relations in preceding speech. These observations organize the present investigation of rhythmic pitch variations (or periodicities in the cycling of F_0-series) by suggesting that underlying periodic tendencies in speech timing remain constant in period and phase across utterances by different speakers, and supplying a rationale for the regular occurrence of simultaneous-onset speech in casual conversations.

Continuous measuring of F_0 should provide an acoustic record of some of the prosodic aspects of a conversation, making it possible to determine whether or not cycling can be found at specifiable periods in these data. If such cycles, or metric rhythms of alternation, exist, they may reflect the organization of participant's prosodic patterns to cohere with each other. Given a means of specifying and describing the period and phase of F_0 cycling, it may be possible to demonstrate that speech segments cohere prosodically. The simplest form of such coherence would be indicated by an invariance between speech segments of the period and phase of this cycling.

If fundamental voice frequencies from spontaneous conversations between two people are measured as time-series data (i.e. digitized at equal intervals of time), autocorrelation analyses do specify periodicity. Graphic modelling of these periods on plots of the F_0 data also indicate the phase of that periodicity, usually by the regular placement of F_0 peaks in the data, enabling the visual modeling of the tendency as a waveform.

Waveform modeling of F_0 fluctuations in dyadic conversations allows the testing of claims regarding the prosodic coherence of segments in connected discourse. Relations of prosodic coherence may obtain between speech onsets and preceding speech, such that the onsets are timed to cohere with preceding speech. The present operational definition of prosodic coherence would predict that the onset occurs in the period of, and with a specified phase relationship to, the waveform model of preceding speech.

The timing of speech onsets to occur in the period and phase of preceding speech implies that certain moments in time, relative to the conversation, might be generally appropriate for a speech onset by any participant in the conversation. This perception, if mutual in all participants, might explain the regular occurrence of simultaneous-onset speech which can be found in casual dyadic conversations in North American English. This line of reasoning guides the following investigation of some properties of prosodic coherence in conversations, and is summarized as the "projection of simultaneous onset (PSO)" hypothesis: Simultaneous onsets are predisposed to occur in the period and phase of pitch peaks in preceding speech.

1.2 Literature

In Chapter 2 some observations of types of prosodic coherence are reviewed. These references demonstrate the need for an acoustic phonetic description of rhythmic coherence, and indicate the potential which inheres in the study of F_0 for cyclicity. Studies of rhythmic mechanisms in speech perception and production justify assumptions that both processes are rhythmically organized. I also mention some alternative descriptions of simultaneous speech from the traditions of conversation analysis and social psychology.

1.3 Methods

Methodological requirements followed from certain technical and design needs to faithfully record conversational F_0 in stereo, to handle large amounts of data in real time, and to process and analyse data interactively, allowing for aural, numeric, and graphic exlorations of the data at all stages. A methodological overview in the form of a block diagram and a listing of procedural stages are provided in Chapter 3, followed by a more detailed commentary on the procedures.

1.4 Results

Current results consist of two examples, which clearly do not establish predominancy of prosodically coherent conversational speech. Conceived of as single-case studies though, these examples do illustrate the methodological and descriptive principles invoked. They also demonstrate the plausibility of the PSO hypothesis.

In Chapter 4, I claim that F_0 in speech by a male participant cycles with 280 msec periods in Example 1, and with 560 and 1120 msec periods in Example 2. In each example, a simultaneous onset by both participants occurs after a silence and roughly in phase with peaks occurring cyclically in the man's speech with the specified period.

1.5 Conclusions

Specific conclusions are stated in a summary of results at the beginning of Chapter 5. Some theoretical implications are drawn regarding the possibility of statements referring to the conversation as one text. Brief mention is also given to the units of analysis implicit in the foregoing descriptions of the prosodic coherence of the conversation. Methodological implications concerning time-series theoretic distinctions such as stationarity/non-stationarity and deterministic/stochastic are indicated. Though decisions regarding these matters are not forced by the autocorrelation analyses, the present investigation does point out some directions for further research regarding the nature of potential time-series models of the processes of interest. General directions for further research are indicated in physiological, cognitive and linguistic, and socio-cultural fields of research.

1.6 Limitations

This study is esentially methodological. Many implications for further study follow from this one, but as it is designed, there can be no clear claim made here that the cycling of F_0 "organizes" conversational interaction. Other implications, including those involving perception, remain tantalizing.

2. Literature.

A history of time-series analytic methods (developed most extensively in econometrics) which have been applied to studies of social interaction can be found in Glass, Wilson and Gottman (1975), Sackett (1977), Tronick, Als and Brazelton (1977), Gottman and Ringland (1981), and Gottman (1979,1981). Time-series theoretic conceptions of patterns of communication have been anticipated in work by Chapple (1971) and Wiener (1961). Jaffe and Feldstein (1970) advocate the modeling of speech activity/inactivity timing by stochastic processes. Some related psychophysiological applications can be found in Porges et al. (1980), and Lewis et al. (1984). Periodic components in rigorous measures of social and communicative behaviors have been described by Kimberly (1970), Aschoff, Fatranska, and Giedke (1971), Cobb (1973), Hayes and Cobb (1979), Warner (1979), Gregory and Hoyt (1982), Gregory (1983), and in long term F_0 fluctations by Voss (1975). No research programs that I am aware of have applied time-domain analyses to F_0 data in series with the intent of discerning syllable and phrase-length rhythms.

In the following survey, I do not attempt a complete representation of any particular disciplinary perspective, nor can I claim to represent fully the extent or range of disciplinary perspectives on the phenomena under consideration. Descriptions of the interactional significance of speech prosody vary widely in focus, method, and interpretation. The distinction between participants' competencies (the rule-governed behaviors which are the objects of syntactic and phonological description), and performance features (associated with the organization of connected speech in conversational interaction) is sometimes blurred in explanations of speaker choices and discoursal utterance functions.

In order to make any claim about perception or cognition based on the physical event of speech, the properties of the latter must be demonstrated. There is thus a need to organize the description of interactional processes without appealing to assumptions regarding individual's cognitive processes, intentions, or semantic, lexical, grammatical or phonological choices. This is synonomous with the possibility of an acoustic phonetics of interactional prosodics, of which the present study is intended as an example.

2.1 Observations of Prosodic Coherence

Some work done in a variety of fields relating to human interaction and vocal discourse suggests that people time their productions according to perceived timing in the productions of their interactants and that they do so in a variety of sound and movement parameters and at a variety of temporal periods. The following review of such work touches on some descriptions of rhythmic coherence (in pitch and in speech activity/incativity measures) that parallel the results reported here.

Mair (1977) discusses his investigations of F_0 and head movement traces from dyadic conversations. He summarizes some of his observations by suggesting that a "stable state of play" is delivered by an "S-shaped fall" in speech melody. This state of play is by definition supraindividual; there is only one, it is continuous, and it develops in rhythmic intervals. Mair (1980) elaborates on neurological implications found in a selected fragment of his data. He begins by describing the supraindividual beat:

> This fragment has a kind of chronometer ticking in it, or more accurately, two chronometers. One is of roughly syllable duration . . . and the other is of suprasegmental, or prosodic span . . . there is a supra-individual beat running through the fragment (1980:6).

See Ulmer (1983) for a review of the conceptual bases of Mair's work and a set of independent observations of speech and movement along similar lines.

In an ethnographic description of a bilingual speech event, Urion (1978) attributes participants' command of the floor and topic in part to "quality of talk" or "skill in the use of prosodic features." He indicates that "the properties that describe the skill . . . may be described with reference to physical parameters [such as] timbre, volume, intonation, stress" (1978:158). Urion refers especially to a stress beat realized in the intonation of one participant's speech:

> There is a rhythm to discourse of which she is obviously aware: she manipulates with the consciousness that she can establish that rhythm in her hearers . . . she typically manipulates speed of delivery; stress is calculated to establish a beat. The beat is so distinct that, after a joke she told at the beginning of the meeting, the entire group joined in laughter that peaked in volume, twice, in a restatement of that beat (1978:159).

Urion specifies his observations by transcribing this woman's speech and the following laughter in standard music notation. His use of triplets in 2/4 time suggest two, possibly three or more levels, of periodicity in the rhythmic structure of the woman's speech with ensuing laughter.

Erickson and Shultz (1981), in discourse studies relating social identity to the social organization of communicative performance, include a continuous coding of rhythmic "points of emphasis" in speech and body movement in their transcriptions of counselling interviews. These data enable them to specify particular moments as "appropriate" or "predictable" places for the occurrence of various organizational events defined in terms of a "temporal context": "the pattern of timing to which all conversation partners are contributing by the reciprocal and complementary pacing of their behavior in speaking and listening" (1981:72). Points of emphasis, when verbal, are determined by perceived intonation and loudness. These points are marked in the transcripts by their occurrence at the left margin (as if each beat caused a carriage return).

Erickson and Shultz very clearly indicate that "when conversation takes place there are rhythmic cycles and wave patterns in verbal and non-verbal behavior that are both intuitively apparent and mechanically measurable," and that "the verbal and nonverbal speaking behavior of speakers maintains the underlying rhythmic interval within and across speaking turns" (1981:74,89). Some of the terms which these authors use to describe these rhythmic phenomena have precise analogues in mathematics and time-series analysis and resemble closely the models derived here by statistical analysis of F_0 fluctuations:

> The successful intercalation of communicative behavior between conversational partners seems to involve <u>staying</u> <u>in</u> <u>phase</u> with these wave patterns. . . . Notions from wave theory, such as phases and peaks, are additional metaphors by which rhythmic periodicity and interpersonal synchrony and entrainment can be described. The extent to which these metaphors should be taken literally in the study of face-to-face communication is not presently clear (1981:74, and footnote 7).

Scollon (1982), in a preliminary attempt to describe cultural institutions through the study of discursive practices, posits the criterion of "tempo" as an analytic discourse universal. He cites data collected from a wide variety of speech situations to examine the general validity of Erickson's claims. Scollon finds that "talk in apparently all texts is timed to an underlying tempo." He also recognises the implication that beats occur during silences, and claims to distinguish these beats according to "tone group closure" of surrounding utterances. He finds a highly variable range of tempos both within and across situations with periods ranging from 500 to 1000 msec and states that the variability within speakers is as great as across situations. He concludes that tempo functions as a means of "negotiating the interaction" between speakers.

It is not entirely clear what Scollon's timing points are, though he implies that speech intonation and stress are the primary determinants of "beat." By separating "tempo" (number of beats per minute) from "density" (number of words per beat), he implies two levels of temporal sequence. He intends this organization to be consistent with musical conventions and the practice of notating rhythms as fractions of metric beats.

Ulmer (1983) has made careful observations of fundamental voice frequency and accompanying head and hand movements in both partners in dyadic conversations, following Mair's observations of sound-movement copatterning. He demonstrates a distinction between intrapersonal and interpersonal synchronies. Following Scollon, he also employed musical notation to specify beat.

Byers (1976) examined continuous acoustic waveform measures of interaction in various (exotic) cultural settings. Studying these measures with concurrent frame-by-frame film analyses, Byers identifies .1-sec intervals between speech and movement onsets in episodes of interaction in these groups and in rhesus macaques.

He describes the timing of listener response to a story teller in a Kalahari bushman community as beginning an exact number of tenths of seconds after the storyteller's final stress peak. He also discusses an example of simultaneous speech which occurs in a shouting register between Yanomamo men. Referring to the regular alignment of speech onsets with .1-sec intervals from the partner's stress peaks, and the consequent alignment of participants' stress peaks, Byers describes the two as engaged in a unique instance of "full phase-locked synchrony [which] is sustained for minutes at a time between two interactants without an exogenous pacer" (1976:156). Byer's work contrasts with research based on relaxation oscillators models since he favors the idea that interaction rhythms can be found to occur at universally preferred periodicities [1].

Chapple (1971) models temporal aspects of interaction by describing the activity-inactivity cycles in human communicative behavior in terms based on relaxation oscillator concepts. He considers that people as oscillators are capable of entraining and being entrained, and that this happens when the rhythms of those being entrained adjust to match those of the one doing the entraining. Chapple's notion of synchrony between participants' cyclic activities involves specification of phase relationships between those cycles.

Relating these concepts to the rhythms associated with the taking of turns, Chapple (1971) extracts measures of dominance from instances of "double action," such as simultaneous speech. Deriving these measures formally in terms of activity-inactivity cycles, he defines some components of interest in the occurrence of double actions: "the latency of onset before the interruption begins, how long it lasts, its persistence, and who outacted the

---

[1]My examples parallel Byer's claims in marking voice onset and stress (pitch) peaks as timing points involved in the pacing of vocal interaction. I did not however find a propensity for periods of 100 msec and integral multiples thereof in the F_0 values of vocal interaction in casual conversations between English speakers.

other and so obtained an instance of dominance" (1971:150).

This description is based on the assumption that each instance of simultaneous speech constitutes an interruption and as such presents a bid for possession of the floor, and on the general notion that activity-inactivity cycles in conversations are normally inversely phase related so that one is inactive while the other is active[2]. These formal notions would prevent us from recognizing the phenomenon of simultaneously onset speech, which is of interest in the present study. As Chapple states, there is no latency of onset when "both persons . . . begin to act simultaneously after both have been inactive," but indicates that such occurrences are very rare.

The action/inaction patterns Chapple models are not precisely comparable to the present results because of the discrete distinction between action and inaction and the inherent arbitrariness in setting sampling rates for these properties (Hayes, Meltzer, and Wolf 1970). In F_0 data on the other hand, all gradations on the scale of fluctuations per unit time are of ordinal value, patterns are not restricted to a simple alternation of ons and offs, and sampling can be continuous if the analytic methods can handle the amount of data produced.

Cobb (1973) used autocorrelation to investigate Chapple's hypotheses by studying vocal activity measures in casual conversations. He studied measurements of each participant's separately recorded voices (taken every 500 msec of proportion of time vocalizing) in 16 eight-minute conversations. The autocorrelations of separate participant's voice measures were identically periodic, consistent with inverse phase lock. This derives from the fact that one is most often inactive vocally while the other is active and vice versa.

Cobb found periods ranging from 46 to 132 seconds in length, in some but not all of the conversations. He regarded these results to be supportive of Chapple's notion of deterministic activity-inactivity cycles, as opposed to the notion that speech rhythms are stochastic Markov processes in which state transition probabilities are defined conditionally to states presently occupied. Cobb predicts that the latter model, advocated by Jaffe and

----

[2] The observation of "inverse phase" in activity/inactivity measures can be thought of as an acoustic correlate of "turn-taking."

colleagues (Jaffe and Feldstein 1970, Jaffe,Breskin, and Gerstman 1972), would tend to produce voice activity durations randomly distributed in exponential fashion.

Jaffe has been known as a critic of subjective observations of rhythmicity based on his generation of apparently rhythmic data by a random-walk process (Jaffe, Breskin, and Gerstman 1972). Jaffe,Anderson and Stern (1979) describe speech rhythm as having a mathematically tractable stochastic element (in the form of an autocorrelated Markov process), and identifies a problem sounding very much like Chapple's: "the rhythmic coupling (entrainment) of human communicators, each of whom brings his or her idiosyncratic rhythmic propensities into the interaction" (Jaffe, Anderson, and Stern 1979:394).

Jaffe et al claim that conversation rhythms can be adequately described with a four-level hierarchy of rhythms beginning with "stressvowel-transsyllabic" alternation as the fastest, imbricated in "phrase-pause", "talk-listen", and "dialogue-break" levels of alternation; alternations are between activity and inactivity (1979:403). A set of observations are made with regards to timing relations at each of these levels:

> Stern and Gibbon (1978) report scalar timing from phrase onset in the playfully rhythmic speech uttered by mothers to their babies which can possibly be regarded as a dialogue situation. At any rate, it remains to be determined whether there is timing constancy across pauses when dialogue takes place between two adults (1979:408).

Later in article the authors claim more boldly that "the phrase rhythm of speaker when he has the floor is 'synchronized' with that of speaker B when <u>he</u> has the floor so that the phrase rhythm is smooth and unbroken, <u>even while posession of the floor alternates</u>" (1979:418).

2.2 Fundamental Voice Frequency and Interaction

2.2.1 F_0 and Discourse

Brazil (1981) maintains an interest in the interactive significance of intonation (perceived F_0) in their phonological taxonomies of tone units. The discourse analytic perspective which this work represents is based on exchange theoretic concepts (Coulthard and Brazil 1981), which are in turn related to speech act theoretic notions (Coulthard 1977). Aspects of tonal prominence are categorized according to semantic relations of the accompanying lexical items to interactionally given paradigms. Tones may be "referring"--to information given in the preceding discourse, or "proclaiming"--of new information (e.g. Sinclair and Brazil 1982).

Inasmuch as the goal of analyses by Brazil and his colleagues is the specification of speaker option-systems, the objects of description are cognitive and hence explanatory at the level of the individual. Their claim to explain interactional phenomena is built into the notion that the semantic paradigms to which tone selections may refer are situation-specific and may be mutually established by all participants. Most other systematic descriptions of tonal units and intonation contours are similarly phonological in nature and less interactional in focus. An example is Ladd's argument for the necessity of taking phonological function into account in developing taxonomic criteria for the description of intonation (1983).

Mair (1977,1980) centralized the role of speech melody in his studies of human face-to-face interaction, developing a theory largely predicated on observations of changes in head and eye movements at points of change in F_0. Mair claims that the contours of speech melody function in the development of shared cognitive models. In particular, an upgoing melody is thought to "destabilize" the model and a downgoing melody to "stabilize" it.

Menn and Boyce (1982) correlated changes in normalized clause-peak F_0 across utterance pairs with independently motivated (and nearly exhaustive) discourse categorizations of the pairs. Category labels included "topic change," "back-channel," and "Wh-questions," as well as "consonant," "dissonant," and "disagree," and the dyadic conversations were between children and adults. The consistency of results confirms the utility of normalizing pitch data, as has been suggested by Jassem (1975). On the basis of their results, these authors conjectured that relative pitch values consistently reflect a "pragmatic" function:

Overall, the discourse correlate of an increase in normalized clause-peak F_0 from the first to the second member of a pair appears to be summarizable under the pragmatic rubric of "degree of disturbance in discourse flow" occasioned by the second member of the pair (Menn and Boyce 1982:341).

2.2.1.1 F_0 and Arousal

Scherer has remarked that "F_0 might be an interesting indicator of the roles of arousal in interpersonal interaction" (1982:156). A general picture emerges in these ascriptions of organizational significance to F_0 variation: fluctuations, or more precisely, elevations, in F_0 reflect or exert discoursal influences according to "interactionally new information" (Brazil and Coulthard, 1981 my paraphrasing), "destabilization of shared cognitive models" (Mair 1980), and "disturbance in discoursal flow" (Menn and Boyce 1982). If Scherer's suggestion that F_0 is an index of arousal is true, then these interpretations of the interactional effects of F_0 height may be related to the discrepancy-arousal processes hypothesized by Capella and Greene (1982) to account for patterns of "mutual influence" (reciprocity and complementary) in expressive behaviors of human vocal interaction (see also Capella 1981).

2.2.1.2 Pitch in Conversation Analyses

Goldberg (1978) described shifts in amplitude as indicating affiliation-disaffiliation with previous utterances in the context of conversational "sequence types" (e.g. "question-answer," "closing"). Gumperz has indicated a belief that "prosody is one of the most important devices that accomplishes cohesion in spoken interaction" (Gumperz, Kaltman, and O'Conner 1984:6). Gumperz et al. speculate, on the basis of experimental evidence such as Cooper and Sorenson's (1981) work on declination, that "speakers use prosody to predict what is likely to come next in extended discourse" (1984:7).

Acoustic F_0 or perceived pitch have been primary variables of interest in recent conversation analytic studies of conversational data, specifically in relation to "turn-competitive incomings" (French and Local 1983), "conversational repair" (Konefal and Fokes 1985), "perception of sentence and paragraph boundaries" (Krieman 1983), and "turn-taking" (Schaffer 1983) and "topic management" (Schaffer 1984).

2.2.1.3 An Ethological Perspective

Ohala (1983,1984) provides an more nearly universal perspective on pitch data by adopting an ethological approach. He has hypothesized an underlying relationship between:

> (a) cross-language similarities in the intonation contours for statements versus questions, (b) cross-cultural similarities in the vocal expression via intonation of attitude and affect, (c) cross-language use of tone, vowels, and consonants in "sound-symbolic" vocabulary, (d) cross-species use of F_0 in threatening or non-threatening vocalizations, (e) cross-cultural and cross-species use of certain facial expressions (involving distinct mouth shape), and (f) the existence of sexual dimorphism in the vocal anatomy of humans (and certain non-humans) (1984:1).

Ohala explains the relationship between these pitch phenomena as based on an "innate frequency code" by which high acoustic frequency is associated with the primary meaning of "small vocalizer" and low frequency with "large vocalizer."

2.2.2 Experimental Approaches

2.2.2.1 Perceptual Normalization

Jassem (1975) has found that intonation curves of the same sentence produced on different occasions by different speakers can be normalized according to personal "pitch" and "compass" which he defines on the mean and standard variation of F_0 in the utterances. Further discussion of the notion of speaker-independant curves can be found in Jassem and Dobrowgowska (1980). Normalization by the removal of individual differences could result in a better picture of pitch variation attributable to the interaction. I did not perform such normalizations in order to preserve the properties of untransformed data in the interest of developing initially non-parametric time-domain analytic statistics.

Pierrehumbert (1979) has demonstrated perceptual normalization for declination by systematically varying F_0 on the last stressed syllable of nonsense sentences and investigating judgements of pitch height:

> Judgements were found to reflect normalization for expected declination; in general, when two stressed syllables sounded equal in pitch, the second was actually lower. The pattern of normalization reflected certain major features of production patterns: A greater correction for declination was made for wide pitch range stimuli than for narrow pitch range stimuli. The slope of expected declination was less for longer stimuli than for shorter ones (1979:363).

There are many segments of the F_0-series I collected which are affected by trend, and presumably the bulk of this trend is due to declination. Since the presence of overall changes in mean obscures the autocorrelation analyses in many such segments (particularly those produced by women), the removal of trend is also suggested on methodological grounds. Pierrehumbert's work provides the perceptual analogue justifying such treatment in the future. Other segments of speech data were amenable to autocorrelation analysis without prior normalization, and as the trends attributable to declination were longer than those fluctuations of interest in the present study, their presence did not affect the analysis for periodicities.

2.2.2.2 Lexical Tone and F_0 Fluctuations

Though evidence exists for language independent prosodic structures (see Vassière 1983), lexical tone, based on F_0, is a language dependent feature. Some recent studies are relevant in this regard. Connel, Hogan, and Rozsypal (1983) found that perception of Mandarin Chinese lexical tones remains stable over a wide range of pitch contour changes, suggesting that intonation is free to operate similarly in this tone language as in non-tone languages. Eady's (1982) results demonstrate on the other hand that while F_0 fluctuations may be determined primarily by stress placement in American English, the use of lexical tones in Mandarin Chinese creates a greater average rate of F_0 fluctuations.

It remains to be determined whether F_0 fluctuates periodically in tone languages as I have found it to in North American English, but since phonological and syntactic rules regarding stress application seem to co-occur with a phonetic periodicity in F_0 in at least some stretches of normal connected discourse, it may be that semantic and lexical aspects of tone tend to operate within rhythmic constraints as well.

2.2.3 F_0, Rhythm, and Autocorrelation

In the present study I examine F_0 fluctuations for periodicity using autocorrelation analyses. Although the results are interpreted as indicative of language stress rhythms, this is not meant to imply that stress rhythms may be defined as periodic pitch variation; it is well known for example that other parameters (amplitude, duration) may vary to create stress, and use of the term rhythm does not usually denote the presence of only one periodicity in the intervals between various types of stress. Some evidence exists that rhythmic stress placement is not categorically perceived, as in the disambiguation of lexical items, without pitch variation (Faure, Hirst and Chafcouloff 1980).

It is important to recognize that autocorrelation examines for correlation between points of a series, high correlations resulting when points covary. A covariance of F_0 points occurs when segments are roughly alike in shape, as in a repetition of identical contours. At the temporal domains presently of interest (constrained by choice of sampling rates and length of segments selected for autocorrelation), the contours shown to be autocorrelated with one another most likely represent F_0 peaks on stressed syllables.

Periodic autocorrelation of F_0 is therefore a working operationalization of stress rhythm, but in accord with the goal of acoustic phonetic description, this statistical description of periodic tendency does not focus on any particular features of the F_0 trace, such as pitch peaks which may function phonologically as stress markers. The PSO hypothesis does on the other hand capitalize on the fact that pitch peaks associated with linguistic stress do usually appear in periodically autocorrelated F_0 data [3], and implies furthermore that an F_0 peak does represent an underlying rhythmic tendency that expresses itself in F_0 peaks and in speech onsets.

2.3 Rhythm and Interaction

The description of periodic tendencies in F_0 productions invites a larger issue in psycholinguistics; whether an argument can be established for a timing principle which structures processes of both speech production and perception. Such a principle might structure participants' roles as speakers, listeners, and as conversationalists by specifying mutually defined rhythms which can then be observed as structuring the interaction itself.

The assumption that processes of production and perception operate via operationally-identical rhythmic mechanisms centralizes the role of rhythm in the coordination of these processes required for participants to create prosodically coherent passages of conversation. This is because the assumption simultaneously accounts for cognitive mechanisms and isolates an element which can be defined independently: rhythm, or periodicity. The notion that rhythmic tendencies underlie processes of speech production and perception has been investigated by physiologists, psychologists, and linguists.

2.3.1 Psychological and Physiological Speech Production Rhythms

Rhythmic mechanisms form the basis of cognitive structures in theoretical frameworks proposed to explain patterns in development by Piaget (1976), and in attention and memory by Jones (1976). Jones refers to research on the perception of speech and pitch sequences to support a theory that "perceptual rhythms . . . can be synchronized to corresponding nested

----

[3]Exceptions to this feature of English conversation are bound to be found in other languages; Dutch, for instance, systematically marks stress with a dip in intonation (Ladd 1982).

time zones within world pattern structure" (1976:232).

Lashley (1951), in a now famous article entitled "The Problem of Serial Order in Behavior," suggested that rhythmic neurophysiological mechanisms might be the natural link between the perception and production of connected speech. Lenneberg (1967) reviewed Lashley's ideas and suggested they are supported by experimental evidence indicating a time constant of 160 +/-20 msec in speech perception and production tasks.

Efforts to specify rhythmic articulatory cycles in the production of speech were not at first rewarded. Ohala (1975), in a histographic study of jaw openings in the production of speech read from a text, found a normal distribution inconsistent with the notion of a physiologically preferred rate of speech articulatory movements. Stone (1981) obtained clearly periodic jaw opening velocity measurements during the utterance of nonsense syllable phrases, and was able to relate variations in these measurements to three levels of syllabic stress.

Paralleling some of Chapple's idea's (e.g. Chapple 1981), Kelso and Tuller (1984) have assembled evidence for common dynamic principles for speech and movement coordination based on the limit cycle behavior of relaxation oscillator models. Studying "sentences" produced by repeating the syllable "ba" in place of the actual segmentals (thus eliminating segmental but not stress differences), Kelso and Tuller found a very stable limit cycle behavior in primary articulator movements centering on a frequency of 5 Hz (200 msec periods). (I will return to a discussion of rhythmic principles in speech production and perception in references to linguistic rhythm research below.)

2.3.2 Cognitive Rhythms

The existence of cognitive planning rhythms in speech has been debated most recently in an exchange between Powers (1983,1984) and Beattie (1984). Powers (1983) studied spontaneous speech for an alternation between hesitant and fluent passages of speech, thought to reflect an alternation between planning and execution phases of speech production. He reports that a panel of judges examined graphical representations of monologues in which speech/silence ratios were expressed as degrees of slope, and failed to discern a "rhythmic" pattern of alternation between regions of steep and shallow slope. Jaffe et al. (1972) criticised this method because of the indistinguishablity of actual results from results based on randomly generated data. Beattie (1984) criticizes Powers in part for taking data from monologues, implying that cognitive rhythms might be more reliably elicited by interactional effects. See also Power's reply (1984).

2.3.3 Linguistic Rhythms

Linguistic investigations of rhythm in English tend to focus on rules for stress placement, following Fry (1968), have emphasized its perceptual aspects. Claims that perceived rhythms are structured according to rhythmic articulations, and not merely as the result of an imposition of perceptual organizations on temporally disorganized acoustic intervals, have received support from some reasonable psychological assumptions concerning the nature of articulation (the "p-center" hypotheses discussed below).

2.3.3.1 Stress-beat and Perceptual Isochrony

Allen (1972[a] and [b]) investigated perceptual localization and differentiation of stress "beats" in sentences recorded from spontaneous conversations. He noted an inverse relationship between the variance of subjects' perceptual localizations of a beat (measured by subjects' finger taps and placements of clicks while listening to playbacks of the sentence), and the stress level of that beat (as reliably agreed upon by a panel of phonologists). Placement of the beats "centered around a ballistic release of the initial consonants into the stressed vowels," suggesting an articulatory locus (1972[a]:72). He did not test for the perception or production of periodic temporal intervals between stress beats; each presentation of the stimulus sentence tested for the location of a single beat.

Allen (1975:78) argues that rhythmic performance universals are behavioral tendencies with perception playing the crucial role of selection: "we perceive spoken language as rhythmic because it is fairly regular in its sequential sound patterns often enough that we can impose upon it simple rhythmic structures." He notes traditional psychological findings which demonstrate that as long as acoustic events are presented with some regularity they will be perceived as rhythmic. He also presents an argument that performance universals (e.g. tendencies to articulate rhythmically) interact differently with the phonological, syntactic and lexical constraints specific to different languages, to produce the varying categories of rhythmic units in different languages. Allen infers from this that there are learned aspects to speech rhythm.

Lehiste (1977), building on Allen's ideas regarding speech perception, claimed that "although isochrony is based on production, perception seems to play a decisive role" (1977:253). Lehiste based these conclusions in part on the idea that segmental speech production intervals are only isochronous within tolerance limits defined by perceptual thresholds for discrimination of interval durations in the 300 to 500 msec range; subjects could not reliably identify intervals smaller than 30 msec. The interpretive orientation of Lehiste's psycholinguistic work is syntactic: explicit deviations from isochrony are thought to signal the presence of syntactic boundaries.

Martin (1972:487) has expanded on Lashley's suggestions regarding the rhythmic organization of processes of speech production and perception by defining rhythm as "relative timing between adjacent and non-adjacent elements in a behavior sequence (hence non-sequential dependencies)." Using this definition, which is based on the idea that the cognitive structuring of rhythm is hierarchical, not concatenative, Martin applies rules for generating rhythmic behavior sequences to patterns of stress positioning in speech.

Such conceptions have been central in Martin's psycholinguistic investigations using reaction time paradigms to study the perceptual expectancies created by the rhythmic productions of segmentals (using nonsense syllables to account for phoneme-specific articulatory effects). See Martin (1979) and Buxton (1983) for reviews of work that has been done along these lines.

2.3.3.2 Perceptual Centers

While attempting to create rhythmic research stimuli by aligning spoken digits according to various acoustic criteria, Morton, Marcus, and Frankish (1976) considered that listeners do perceive something happening at regular timing intervals, when they perceive speech to be rhythmic, even if there is no isochrony between the various acoustic components of the spoken words. They had required as stimuli a list of spoken words (the list of digits one through nine) at precisely rhythmic intervals, but when they attempted to align the recorded words according to points in their acoustic energy waveforms (word onset, vowel onset, and peak vowel intensity), the list was not heard as "phenomenally regular." Conversely, when a list of words was heard as phenomenally regular no obvious acoustical point of alignment could be discerned. On the assumption that a perceptually rhythmic list of words was nonetheless aligned according to some psychological criterion, they postulated the existence of "perceptual centers" (p-centers) initially defined as "psychological moment of occurrence" (Morton et al. 1976:405).

Fowler (1983) reviews psychological, phonetics, and phonological references and reports experimental results all pointing to the conclusion that vowel productions serve as fundamental units of produced and perceived segmental rhythm. This work suggests that listeners can account for segmental differences (specifically for the amount of consonantal vowel overlap) and are able to identify the onsets of vowel articulations as points of temporal alignment. She hypothesizes on this basis that the points of onset of vowel articulation (which correspond differently according to consonantal coarticulation to points of vowel onset measured by acoustic criterion) are the syllable "perceptual-centers," suggesting that an alignment of these points at rhythmic intervals will produce a perception of rhythmicity where none may be found in the acoustic spectral energy trace.

The concept of perceptual centers has been useful in demonstrating an underlying universality among languages in different "rhythm categories": Hoequist (1983) found p-center effects in directions predicted by the criterion of articulatory vowel onset in

English ("stress-timed"), Spanish ("syllable-timed"), and Japanese ("mora-timed"), suggesting a universal principle of syllable timing.

Dauer (1983) compared stress and syllable timed languages (English, Thai, Spanish, Italian and Greek). He was able to account for variations in the rhythmicity of stress placement among these languages by relating them to language specific differences in "syllabic structure, vowel reduction, and phonetic realization of stress," leading him to conclude that "a tendency for stresses to recur regularly appears to be a language universal property" (1983:51).

Apparently speech rhythmicity inheres in processes of syllable production and perception such that produced segmental rhythms are perceived even when actual acoustic intervals are not rhythmic. The existence of just such a matching between production and perception may also be evident in patterns of suprasegmental coherence between participants' vocalizations in conversation.

The p-center hypotheses enable us to assume perceptual adjustments for variations in production in the organization of speech timing, and may vindicate Lashley's original proposal that a common rhythmic mechanism links neurological processes of production and perception. It is possible that segmental rhythms obtain between vowel productions at the same time as suprasegmental rhythms can be found between F_0 peaks. These rhythms may coexist at different temporal domains and their relative perceptual salience may vary.

In my results, the statistically predominant periodicities in F_0 varied from 240 to 2600 msec in length (this is not a statement about speech data in general but only a rough characterization of a small number of segments). Although the lower end of this range seems to include average syllable durations, fluctuations in F_0 sampled every 40 msec are not typically found in syllable length durations, and my results are primarily relevant to discussions of stress rhythm. Some examples not reported here did suggest simultaneous structuring of F_0 at multiple temporal levels by exhibiting autocorrelation at periods which were not integral multiples of each other.

Findings of periodicity corresponding to stress rhythm intervals reported here do not of course preclude findings of even closer temporal coherence of syllable durations (and phase relations) across utterances. The existence of this latter type of coherence is suggested by the work on p-centers, but is not an object of investigation in the present analysis. The basic principle that people are able to perceive precisely timed rhythmic productions by compensating for specific effects associated with transient aspects of speech performance is nonetheless of clear importance to a discussion of rhythmic coherence across utterances by different speakers.

2.4 Perspectives on Simultaneous Speech

Implications of the coherence of speech rhythms for the occurrence of simultaneous starts have been recognized by Mair (1977) and to some extent Byers (1976). Their regular appearance in casual conversations may be explained, as in the PSO hypothesis, as a consequence of participants timing their speech according to the rhythms in preceding speech. Implications for the study of simultaneous speech in this research have been anticipated in some of the existing approaches to conversational speech data, but proceed from entirely different lines of reasoning and methodological bases.

2.4.1 Conversation Analysis

Sacks, Schegloff, and Jefferson (1974:720), in the context of a discussion of rules for turn-taking, dismiss simultaneous starts as produced by "self selectors [of the turn] aiming for earliest possible start" where earliest possible start is defined with reference to the syntax of preceding speech. Jefferson (1973) treats instances of simultaneous speech as rule-governed in relation to lexico-semantic structures and utterance "sequence types" (e.g. question-answer, closing). She specified certain points at which the precision timing of "overlap" could be considered as signalling recognition of the speaker's message, and hypothesized further that certain terms may be designedly placed so that the other speaker may exercise the option of overlapping those terms (e.g. address terms in closing sequences). She regards these examples of "precision timing" as evidence of the "intense organization" of conversational resources, but makes no attempt to consider prosodic or acoustic aspects of this timing. (Indeed no objective measures of timing are made other than to note pause durations in the transcriptions.) She does suggest briefly that there are shared units of time, such as "beats," in passages of conversational speech data, but her formal interest in simultaneous starts seems to be in determining rules for the outcome. Auer (1983) has suggested that overlap be classifed according to its occurrence before and after "recognition points" in the present speaker's speech specified by lexical and syntactic criteria, as in Jefferson (1973).

Bennet (1981) criticizes "structural/syntactic" specification of "possible completion points" (as in Sacks,Schegloff, and Jefferson [1974]) as inadequate for explaining subjective descriptions of qualitative aspects of interruptive overlapping speech. One such aspect noted by Bennet is whether or not a speaker "changes rhythm when interrupted" (1981:174-175). French and Local (1983) note that there are prosodic cues by which participants' may constitute "incomings" as "competetive for the turn". They suggest that loudness, pitch, and tempo operate "irrespective of lexico-syntactic or illocutionary characteristics" (1983:17).

These conversation analytic approaches have, in their treatment of simultaneous speech, progressed from assertions of syntactic and semantic significance to phenomenological explorations of prosody in their treatment of simultaneous speech. None of these authors attempt to describe acoustic phonetic patterning of suprasegmentals in simultaneous speech or simultaneous onset.

2.4.2 Social Psychology

Meltzer,Morris, and Hayes (1971) and Morris (1972) have performed some rigorous investigations of the relation of acoustic amplitude to "interruption outcome," obtaining some very stable results. These authors oriented their research design and interpretations by an interest in turn-taking. The dependent variable was interruption outcome. Their results indicated a very strong role for speech amplitude in determining interruption outcomes, but their interest in turn taking prompts them to specify "attacker" and "defender." They eliminate from consideration categories of simultaneous speech such as "back-channels" which are listener responses not thought to indicate a desire for the floor (Yngve 1967)[4].

Natale, Entin, and Jaffe (1979) also considered simultaneous speech in analyses of conversations, relating interruptive behavior to other participant attributes ("speech anxiety", "confidence as a speaker", and "social anxiety") Their findings were in accord with their predictions. They interpreted significant positive correlation of a person's "need for social approval" and their interruptive behavior as indicating that "speech interruptions may not always represent a contest for the conversation floors" (1979:875-876).

Natale et al. found that "only a minor, albeit significant, proportion of speech interruptions was affected by the personality measures used in [the] study." While agreeing that the timing of interruptions may be primarily stochastic in nature (as found for speech in general by Jaffe and Feldstein [1970]), Natale et al. concluded that external variables ("personality, task orientations") modify the timing of speech: "Transitional probabilities of the vocal parameters cannot be assumed to be constant" (1979:876). My results are in agreement with Natale, Entin, and Jaffe's suggestion that speech timing processes have non-stationary components (see the section on non-stationarity in Chapter 5 below), but proceed from an entirely different line of reasoning. The interest in simultaneous speech is coincidental.

----

[4]See also West and Zimmerman (1977) for findings which relate interruptive behavior to the sex of participants, and Beattie (1981) for evidence that setting may affect the realization of sex-related effects.

3. Method

3.1 Overview

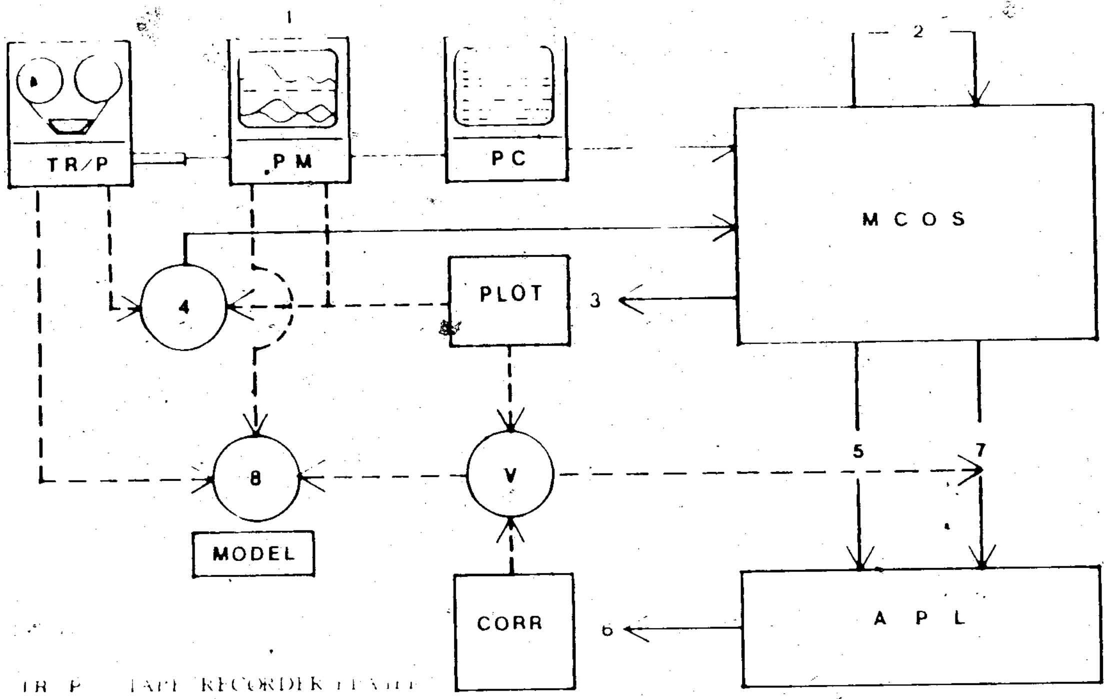
The procedures I used to produce the present results involved the techniques of audio recording, and of numerical and graphic processing in F_0 extraction and time-series analysis. Recording of the conversations was done in stereo and emphasized the fundamental voice frequencies. Collection and treatment of F_0 data were performed by machine with error correction by hand. Analysis consisted primarily in the examination of correlograms taken of selected segments of F_0 data, and visual modelling of the indicated period and phase properties on plots of these data. Main headings in the body of this chapter refer to these three sets of procedures; the recording of the conversations, the collection and treatment of F_0 data, and analytic procedures.

The faithful recording of F_0 in dyadic conversations required a particular type of microphone design to be developed specifically for this purpose. I describe this apparatus and the procedures relating to participant selection, protocols, and recording, below, along with some notes regarding the six three to four minute conversations recorded.

Figure 3.1 represents an overview of the numerical and graphical procedures used in this research, indicating electronic and computer systems (labelled with acronyms), procedural stages (listed numerically), and the methodological paths by which stages are connected. As an introduction to the following discussion of the methodology, I will review the eight stages indicated in Figure 3.1. This list is not presented as an exhaustive chronology of the procedural steps producing the present results, but it does indicate a rough ordering of the main procedures involved.

The first four stages of the numerical and graphic procedures have to do with the collection and treatment of data.

1. Pitch processing. This procedure involves three instruments: an audio tape deck is used to play back the recorded conversations, the Pitchmeter is used to extract measures of F_0



1 PITCH PROCESSING

2 = SMOOTHING

3 = PLOTTING

4 = EDITING

5 = SEGMENTING

6 = AUTOCORRELATION

7 = SPLICING

8 = MODELING

TR/P = TAPE RECORDER PLAYER

PM = PITCHMETER

PLOT = F PLOTS

MCOS = MAINFRAME COMPUTER OPERATING SYSTEM

PC = PERSONAL COMPUTER

CORR = AUTOCORRELATION RESULTS

APL = INTERACTIVE APL SYSTEM

MODEL = HYPOTHETICAL WAVEFORM MODEL

V = VISUAL COMPARISON OF F_0-PLOTS AND CORRELOGRAMS

⟶ ELECTRONIC PROCEDURES

◯ - - -> MANUAL, AURAL, AND VISUAL PROCEDURES

Figure 3.1 Overview ... treatment, and analysis procedures from these recordings, and the personal computer is used to read the digital output of the Pitchmeter into time-series form via a "data capture" system. These series can then be read into a mainframe computer's operating system for further treatment and analysis.

2. Smoothing. In the present investigation the rate of data capture was set to produce series with 200 samples per second (sps). A smoothing program developed for this project eliminates many erroneous samples by an averaging process which excludes zeros and mis-reads, and was used in this study to convert the "raw" series into 25 sps series.

3. Plotting. Plots of the synchronized series representing both participants in conversation were drawn by computer. I ordinarily performed this stage at least twice for each conversational episode transcribed; first viewing the plots on sheets of inexpensive hardcopy to perform the editing procedures (see stage 4), and then replotting the error-free series in full size color plots for graphic analyses and the fitting of hand drawn curves.

4. Editing. Prior to analysis, the computer-transcribed files of F_0 data had to be checked for errors. This was performed by comparing F_0-pl[illegible] the audio and video representations of the conversations available through the tape player and Pitchmeter. The bulk of the corrections consisted of the removal of cross-talk between channels--a more complete inventory of errors is provided below.

The second four stages represent the procedures involved in the analyses of the data. The first two of these stages, segmenting and autocorrelating, establish the extent of cyclicity in selected parts of the F_0-series, and the consequent procedures of splicing and modeling provide clarifications of the period and phase relations between segments.

5. Segmenting. Segments of the F_0-series can be selected simply be referring to the index numbers on the x-axis of the F_0-plot (the coordinate representing time). These index numbers correspond to line numbers in the computer files containing the series plotted.

6. Autocorrelation. Autocorrelation analyses performed in the interactive programming language APL and plotted as correlograms allowed extensive data exploration for the

trial-and-error discovery of significantly periodic segments. Selected correlograms were used to indicate segments of comparable period with intervening silences--candidates for splicing--and to guide the final stages of modeling.

7. Splicing. Where segments with comparable period were separated by stretches of silence, and in some cases where these stretches contained speech with either indeterminate or incomparable period, portions of the intervening stretches equal in length to the specified periods (or integral multiples thereof) in the surrounding segments could be spliced from the series, producing series in which the segments of comparable period were adjacent to each other in known phase relationships, and with fewer zeros (i.e. less "missing data"). Improved correlograms of the speech episodes of interest were obtained with this technique. Splicing is also justified on theoretical grounds as testing the very assumption that period and phase continue across silences in the production of utterances which are prosodically coherent with each other.

8. Modeling. Based on satisfactory autocorrelation results regarding the periodicity in speech segments and the continuance of that period across stretches of sound and silence separating those segments, hand drawn waveforms with the specified period were aligned visually to the F_0-plots. This wave model provides visual confirmation that periods detected by autocorrelation of F_0 can be found to correspond roughly to the intervals between F_0 peaks, and that these intervals can also be found between those peaks and points of simultaneous onset.

3.2 Recording of the Conversations

To study naturally occurring instances of spontaneous speech for periodicities in the cycling of F_0, I required high quality recordings, necessitating a laboratory setting with recruited "subjects" as participants. The recordings needed also to be stereo-specific (each subject recorded onto a separate audio channel) so that pitch processing would be possible on each participant's voice separately during episodes of simultaneous speech. The requirements of naturalness and high-fidelity stereo are compromised in the design of the microphones developed for this research.

3.2.1 Microphones

A high degree of focus on individual speakers' voices at F_0 range frequencies was required of the microphones, and short of immobilizing or separating the participants, only physical proximity of a microphone to the individual speaker could achieve such a focusing[5]. Another consideration, pointed out to me by Dr. Carl Urion, was that optimal transcription of F_0 by the Pitchmeter is obtainable when the microphone is placed directly against the fleshy part of the throat to the side of the larynx (the source of phonation, or F_0 production). The microphone housings used in the recordings analyzed below evolved from a pair of custom microphones designed by Dr. William Pierce. These were cloth straps made to be worn around the neck, with sections of foam rubber housing the actual microphones. Focus on the individual's own larynx was further improved by shielding the microphones inside the straps with quarter-inch thick rubber.

Once accustomed to the snugness, the participants did not complain about the neck straps, and other than a somewhat unusual appearance and some restriction to movement of the head, the naturalness of the conversations did not seem too much intruded upon by the neck straps.

Present shortcomings in the recording process necessitated hand editing of the data, as described in the section dealing with data treatment below. Although the microphones were effective in recording even very quiet phonations, there was still enough overlap from one channel of the tape to the other (through the microphones) to contaminate the data from the individual speakers. This "cross-talk" between channels was rare enough that instances were easily identified and edited out of the transcriptions, but prevalant enough that measures of

----

[5] Bone-conducting microphones of the type used by Hayes and Meltzer (Hayes & Meltzer 1967) and others are not adequate for the recording of frequencies as low as F_0.

raw pitch analysis output could not be relied upon[6].

3.2.2 Procedures

3.2.2.1 Participants

Four volunteers, two men and two women, participated in the study. The participants were graduate students previously acquainted with each other and whose native language was Englsh. They had varying degrees of naivetè with respect to the purpose of the recordings, though none knew that there was to be any interest in simultaneous speech.

3.2.2.2 Protocols

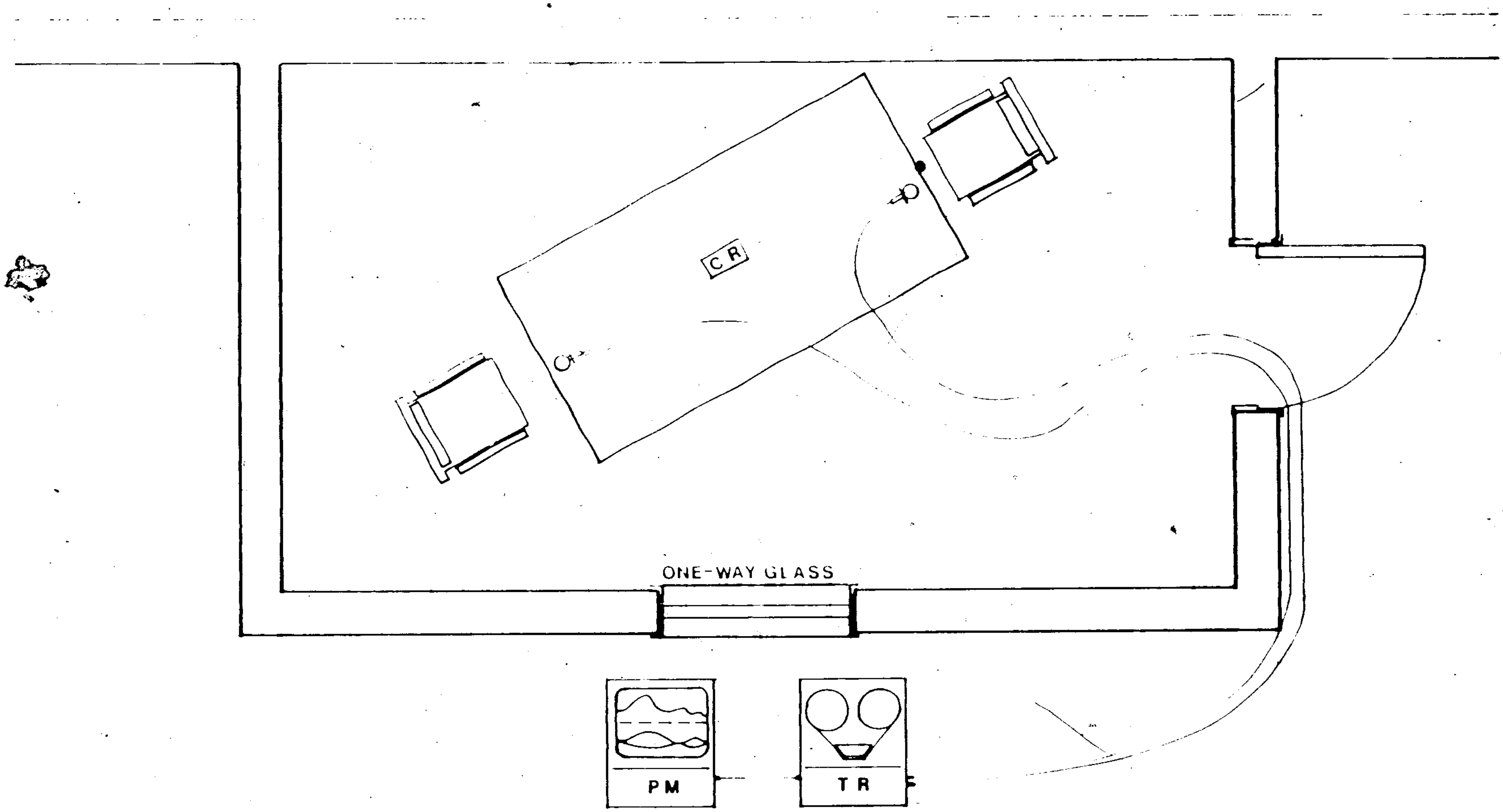
After meeting the participants in a waiting room near the recording studio, I introduced the research saying that I was interested only in statistical aspects of certain acoustical parameters of the conversations and that there would be no content or personality analyses *per se*. The participants were told simply that they would be wearing neck strap microphones and asked to converse freely on any topic for about three minutes at a time. The participants then signed release forms assuring them of their rights to anonymity and a full explanation of the research in layman's terms (see Appendix 1). The conversationalists were then selected in an arbitrarily predetermined order two at a time and escorted to a recording booth just large enough to accommodate them, a table and two chairs (see Figure 3.2).

3.2.2.3 Recording Procedures

The participants fitted the neck strap microphones to their throats, and a small cassette recorder was placed on the table as a backup in case intelligibility was too low in

---

[6] There were occasionally instances in which a female participant's F_0 was masked by a male's. An example of how this problem was dealt with is discussed in Example [illegible] Chapter 4. Analyzability of direct unedited output of automatic F_0 analysis might be attainable with the more recently improved design of microphones in an acoustically dampened setting which permits some distance between participants.



CR = CASSETTE RECORDER TR = TAPE RECORDER PM = PITCHMETER

= MICROPHONE (housed in foam rubber when worn by participant)

Figure 3.2 Recording procedures.

the recordings from the neck strap microphones; this turned out not to be the case. A sound check was performed by asking each subject to count to ten once softly, once in a normal voice, and once loudly. The video output of the Pitchmeter (used to extract F_0 from the sound signal) was also monitored in the setting of recording levels for optimal clarity of signal. A final check was made during which participants alternated digits counting to ten*. After the sound check I asked the participants to remain silent until after I had closed the door to the booth and tapped on the one-way glass separating the booth from the recording studio. Conversations were ended when I opened the door to the booth.

3.2.3 The Conversations

I recorded six conversations, each of the participants speaking with each of the other three. There were some obvious discomforts associated with the task, and some of the conversations seemed tentative or forced. Others were more animated, even lively. In the earlier conversations recorded, participants drew on a topic which was being discussed in the waiting area. In later conversations there was a longer initial phase during which participants prompted each other into picking a topic (more often than not producing repeated bits of simultaneous and simultaneous-onset speech).

One particularly notable feature of the conversations was a lack of tolerance for silence. The participants in these conversations behaved precisely as they were asked; they talked. There were few pauses over one or two seconds in duration. Although the subjects confessed to loquacity, and although there are certainly sociolinguistic and cultural dimensions to the tolerance for silence in conversation (see Hudson 1980:116-118 for some references), it seemed here that these subjects felt a necessity to fill time with talk, a feeling no doubt engendered by the setting.

----

* The recording levels set as a result of these checks were often different between the two channels, preventing later comparison of intensity results on an absolute scale.

The conversations each contained between 16 and 25 simultaneous locutions, occurring on average roughly six to eight times a minute. Most of the simultaneities consisted of "interjections" (e.g. "mm", "uh-huh", and "oh yeah") by one participant during the other's speech[8]. Some instances of simultaneous locutions consisted of laughter by one or both participants [9].

Each conversation contained at least one example of simultaneous-onset speech following a silence. Some of the conversations contained many more of these than others: in this particular set the conversation between the males contained the most occurrences of simultaneously onset speech, and that between the females the fewest. Each conversation also contained episodes of simultaneous locutions lasting one second or longer, and most contained four- to six-second episodes during which short segments of simultaneous speech occurred repeatedly. Such episodes were the preferred habitat of simultaneous-onset locutions.

Details regarding the choice of subjects to participate in recorded conversations are of interest as limitations on the generalizeablity of the results. No attempt was made to control for individual performance aspects of the speech phenomena under consideration here. The research setting and selection of speakers were "unnatural" influences on the conversations, and these aspects of the recording may even have systematically affected the prosodics in the conversations recorded. This is permissable because the design of the present research does not encompass variation in the uses of prosody, but only the description of some prosodic patterns realized in instances of simultaneous locution.

The results may be conceived of as single-case studies. I anticipate that the loss of immediate generalizeability will be recovered upon the development of descriptions that are phonetically precise and formally economical and therefore applicable to a wide variety of

---

[8] I also observed precise timing of these interjections and surrounding speech such that there was neither a perceived silence nor any simultaneous speech during speaker switches. Further study of such interjections may produce another source of data on projected prosodic coherence between speakers.

[9] Laughter containing phonation, in contrast to interjections, was almost always simultaneous with phonation in speech or laughter by the other participant in these conversations.

conversational phenomena.

3.3 Collection and Treatment of F_0 Data

Collection of data involved processing the audio tapes through the PM 100, 200, and 300 programs developed by Dr. Phillipe Martin and hardwired into a microprocessor controlled device (the "Pitchmeter"), and an IBM PC, to produce digital expressions of F_0 in series stored as files in a mainframe computer. Some treatments of the data are required along the way to correct for errors. These procedures eventually result in F_0-series with samples occurring at a rate of 25 samples per second (sps) which can be displayed as plots. Collection and treatment procedures are represented as stages 1 through 4 in Figure 3.1 and in the list at the beginning of this chapter.

3.3.1 Data Collection: Pitch Processing and Data Capture

The stage referred to in Figure 3.1 as pitch processing actually involves two separate procedures which occur simultaneously. The first involves use of the Pitchmeter to extract digitized F_0 measures from the audio tape, and the second involves the "capture" of these data in the form of a stored list of readings occurring at equal intervals of time--a time series.

3.3.1.1 The Pitchmeter

Pitch, intensity, and oscillographic displays are available on a video monitor, and the Pitchmeter also has analog and digital outputs of these signals. The data are updated every glottal period, that is, with each closing of the glottis, and can be output in "real-time" (virtually instantaneously). The Pitchmeter provides as well a number of voice statistics not used in the present study.

The F_0 analysis is performed by "filter tracking":

> Seven (7) bandpass filters . . . are used. The output of every filter channel is continuously monitored by the processor in order to evaluate the <u>periods</u> between two consecutive positive zero crossings of the input signals. The

system then takes, as the fundamental period, the zero crossing period coming from the lowest frequency bandpass filter if this period lies between fixed limits (in order to eliminate nonperiodic signals and noise) (Pitch Instruments n.d.:2, also see Leon and Martin 1972).

I first viewed the conversations on the "Prosodic Feature Display" program (PM 200) in continuous sweep mode. The video display of this program consists of F_0 on the top half of the screen and intensity and the oscillographic envelope on the bottem half. The sweep time, the range of the F_0 display, and the input intensity are all variable. Fixed displays were then used primarily in later stages of analysis, as in the checking of machine transribed data. To inspect both speakers' prosodic traces simultaneously, I used the PM 100 "Teaching Intonation" program to display one person's trace fixed on the top half of the screen and the other's on the bottem. Some manipulations would usually be necessary to synch the two halves of the screen but the resulting display was crucial to the editing of cross-talk between channels during simultaneous speech.

3.3.1.2 Data Cap

To accomplish automatic transcription of data from real-time ouput of the Pitchmeter to the time-sharing operating system of the mainframe computer, it was necessary to read the real-time output of the Pitchmeter using a personal computer. An IBM compatible hardware/software data capture system (programmed in Pascal) was developed for this purpose by Mr. Ray Wong of Compuware Corp., Edmonton Alberta. This system was installed to interface the Pitchmeter to an IBM PC/XT, which was then used to write samples of the digital output of the pitchmeter onto diskette. The diskette files could then be read to the mainframe computer operating system as data files.

3.3.2 Data Treatment:Smoothing and Editing

Pitch analysis is vulnerable to fast fluctuations in frequency caused by periodic components in the input from supra-glottal sources (e.g. frication). These fluctuations are smoothed in the Pitchmeter by a median algorithm with variable N, but some, which may approach syllable duration, remain as error in the digital output of the Pitchmeter. Other readings of F_0 are sporadic due to low intensity of recording levels and/or input sensitivity.

Some samples are occasionally lost in the transcription if a PC read occurs during a PM update. One way of excluding many of these errors was to sample at a much greater rate than needed to represent the fluctuations of interest (and provide sufficient numbers of data points for time-series analysis), and then take medians over succesive numbers of points appropriate to produce the desired temporal sampling rate.

3.3.2.1 Smoothing

I chose to make transcriptions for the analyses presented below by sampling the Pitchmeter output 200 times a second (5 msec. between samples) and averaging over every 8 of these samples to produce an F_0-series with 25 samples per second (40 msec. between samples). This effective rate was chosen to include all F_0 fluctations of syllable length and larger (prominent F_0 fluctuations associated with articulations of phonemic length would also be included at this rate), and because at this rate the sampling of the processes of interest can be regarded as continuous for the purposes of time-series analysis.

Programs used to smooth the 200 sps series into 25 sps series were written for this research project by Dr. Dwight Harley of the Division of Educational Research Services, University of Alberta. The Fortran program was designed to insert zeros where more than half the points were false values or zeros, and to reject false values and zeros from the median averaging. The median averaging itself eliminated many of the outlier effects caused by supraglottal sound sources. The resultant data were then reindexed to synch with those of the partner's data.

3.3.2.2 Sampling Rate

Although this sampling rate of 25 samples per second (one every 40 msec) puts some clear limitations on the temporal precision of results, statements concerning the temporal occurrence of F_0 features could be made on the basis of these series within tolerance limits of +/-20 msec. Comparisons of these supra-segmental data with the faster timed segmental aspects of the speech signal cannot be consistently made in these series.

A further increase in sampling rate would have allowed a greater number of samples per segment and a consequently larger number of lags at which correlations could reliably be taken, but this increase would not result in a significantly better estimate of size of the cycle in the data. Cycles of the size to be found in these F_0-series (with periods of 6 or more points) could not be detected with greater confidence at faster sampling rates, as the effective sample size would be reduced by a proportionate amount if these cycles were expressed as maximal correlations at a greater number of lags (discussed in the section dealing with autocorrelation, below).

Time-series values are thought of not as samples from a population, but as samples from a process occurring over time. The proper sampling rate of a time-series measure of that process is determined more by the density of samples over the process, and not strictly by the density of samples over time. In the present research, the processes of interest are cycles ranging from .2 to 2.5 seconds in period.

For a given episode of the conversation then, the pitch processing, data capture, and smoothing procedures resulted in pairs of data files, each file containing two columns indexing the sample number (or time, in 40 msec units) and a second column with a corresponding F_0 value representing the median of all non-false non-zero pitchmeter readings occurring four 10 msec intervals preceding and four 10 msec intervals following that point in time represented by the index number. I refer to these data files as F_0-series.

3.3.2.3 Editing

Though some errors have been elimated by the smoothing from 200 to 25 sps, these F_0-series can still not be treated as data. There is still cross-talk between channels. The

Pitchmeter has variable sensitivity levels which may or may not be appropriate for the entire episode. Occasional erroneous readings from non-glottal periodic sound sources lasting longer than 20 msec. remain in the 25 sps series. All of these persisting sources of error leave gaps, wrong samples, and samples where there should be gaps in the F_0-series even after the smoothing procedures. These contaminations are most often quite blatant when the F_0-series is plotted and compared to audio playbacks and video analysis with the Pitchmeter. In the recordings made for this study errors consisted mostly in instances of cross-talk.

As a result of the comparison between plots of the F_0-series with audio and video presentations, I edited the raw F_0-series. Approximately one sample in every twenty-five was determined to be erroneous by one of the above explanations and corrected. In an earlier stage of my studies I transcribed entire series by hand. The checking process merely corrects the machine transcriptions where they fail. The time saved by the data capture system is still enormous.

3.3.2.4 Plotting

The plotting system is based on programs written by Dwight Harley of the Division of Educational Research Services at the University of Alberta. In this study, I created plot description files representing unedited F_0-series of both participants in a selected episode. I then compared crude hardcopies of these plots to the audio and video playbacks, and noted the discrepancies according to file line number. I corrected the series and the plot description files, this time sending the plots to a plotting device for high quality color plots of both Fo-series in a selected episode of simultaneous speech. I refer to these as F_0-plots. For research purposes, the x-axis, representing time, was labelled with file line numbers. (In the F_0-plots presented in Chapter 4, the x-axis scale has been redrawn to indicate the actual time units.)

3.4 Analysis

Autocorrelation analyses are now applicable to the F_0-series in which values consistently and accurately represent fundamental voice frequencies as measured by the Pitchmeter and confirmed by visual and auditory examinations of the data. Autocorrelation analyses are used to test for cyclity in the data, and they may specify the period of that cycle.

It has been stated that the problem in this investigation is to describe prosodic coherence as an invariance of period and phase in F_0 cycling across utterances and speakers. In particular, I suggested that simultaneous onsets of speech will be found occurring in the phase and period of preceding speech. Implicit in these claims is the notion that the fluctuation of F_0 in some segments can be described as a periodic waveform, and that a waveform model of F_0 fluctations can be specified by demonstrating period and phase in actual data.

The analytic component of the method I used to support these claims consists of the statistical demonstration of periodicity in F_0-series by autocorrelation, and the subsequent determination of phase by visual analysis. Both period and phase in speech segments, but phase in particular, can be confirmed by autocorrelation analyses performed on spliced segments from which data have been selectively removed as a test of the assumption that period and phase underlying the cycling of F_0 remain unaltered during some silences. Results of these analyses can be displayed visually by the alignment of a hypothetical waveform with the specified period and phase parameters to the F_0-plots of the speech episodes.

3.4.1 Period

3.4.1.1 Segmenting

Periodic cycling of F_0-series can be demonstrated by autocorrelation analysis of selected segments. Proper selection of segments for analysis is vital for detection of underlying transient cycles. The initial segments chosen were usually bounded by the occurrence of a syntactic boundary with a pause. Second and third segmentations were based on the results of autocorrelation of the initial segments; the autocorrelation of segments based on syntactic and/or pausal critieria was not always the clearest possible. (In Examples 1 and 2 presented in Chapter 4 a straightforward segmentation according to pause is adopted.)

The goal of the analyses therefore became in part to determine those segments of the F_0-series containing demonstrable periodicity. This exploratory process of isolating those segments was aided by intuitions based on playback of the audio tape and by visual examination of the F_0-plots. Autocorrelation analytic criteria for cyclicity impose certain limitations on the type of segment in which periodicity is thereby demonstrated. Some of these limitations are discussed in the section on autocorrelation which follows.

3.4.1.2 Autocorrelation

Autocorrelation is a procedure which describes the dependency of points in a series on previous points of that series. This dependency is measured by correlating the points in a series with those a number of lags preceding. Calculation of correlations of a series with itself over successive numbers of lags produces an autocorrelation series. An autocorrelation series in this case is a series of correlations between successive points in the F_0-series displaced by increasing intervals of time. The independent variable of this time-domain analysis is therefore an expression of time in its manifestation as a set of sequential dependencies, but not in its historical aspects. A plot of the autocorrelation series by the lags at which correlations are taken is called a correlogram.

Figure 3.3 is a highly idealized visual representation of the process of autocorrelating a segment with three peaks (the minimum number of cycles which must appear in the data for detection of periodicity by autocorrelation), and a very clear 10 point cycle. The correlogram alternates between high and low correlations with a period (in lags) equal to that in the F_0 curve (see Figure 3.3). A statistical tendency towards periodicity in a series of points can thus be detected by a patterning of the autocorrelation of that series to alternate regularly between positive and negative values.

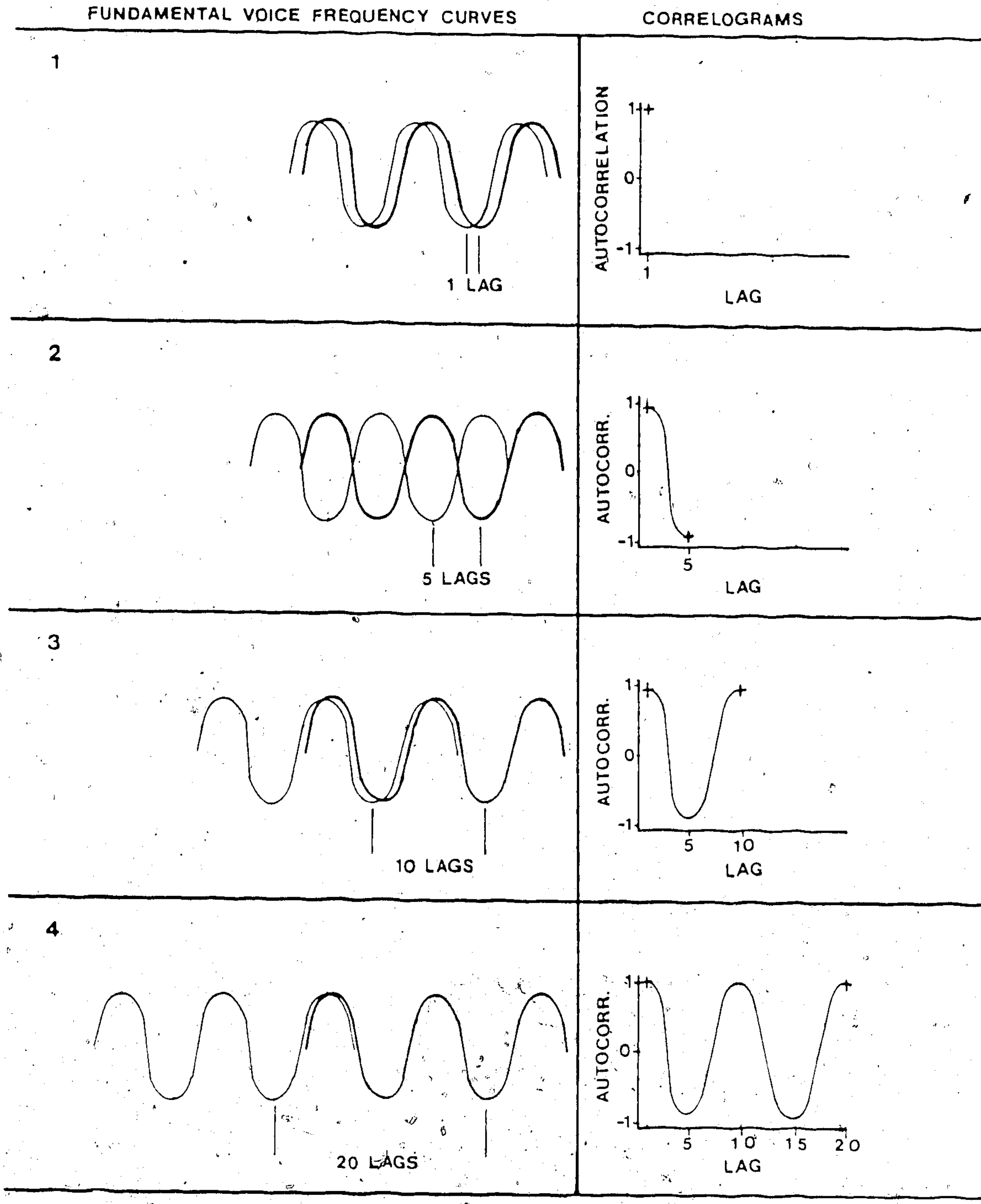


Figure 3.3 Graphic representation of the autocorrelation process. Depicted on the left-hand side of this figure are highly idealized versions of an F_0 segment in which there is no trend and a smooth 10 point period (bold line). Superimposed on this segment in steps 1 through 4 is an image of itself (thin line) at increasing lags (in points) of time. On the right-hand side are correlograms of this segment. In steps 1 through 4, the autocorrelation is taken at increasing lags. Note the alternation in the correlograms at a period equal to the period in the F_0 segment. (This effect also obtains, though less powerfully, when the period in the F_0 segment is masked by noise.)

The need for large enough numbers of samples to produce reliable correlations imposed minimal length restrictions. In many cases changes in phase or period or an absence of periodicity in the F_0-series altogether in parts of the segment would prevent the resolution of the autocorrelation series into a periodic alternation, imposing maximal length restrictions. Minimal length restrictions are also imposed with respect to the number and length of cycles in the F_0-series. Periodicity is established in the correlogram of a segment of F_0 data only if two comparable periods occur between maxima (or between minima). As seen in Figure 3.3, this obtains only when a segment with at least three cycles can be autocorrelated to lags equal to the span of two consecutive cycles, as in step 4 of that figure in which the autocorrelation of a series with three 10 point cycles is taken to 20 lags[10].

The basic tool in the following analyses was an autocorrelation program which could handle missing data, written by Dr.Tom Maguire of the Division of Educational Research Services at the University of Alberta in the interactive programming language APL (Iverson 1962). Conceptions of time-series as descriptions of continuously measurable quantities need to be altered in applications to speech data containing silences. In this time-series application, zeros are technically missing data since they cannot be included in correlations: they are not on the same scale as the other data and would introduce spurious results at all lags of an autocorrelation. The methodological problem of missing data can be solved by assuming that the parameters of tendencies underlying speech segments remain constant across silences and intervening bits of speech.

A rough method for determining the significance of autocorrelations is described by Gottman (1981):

> An approximate standard deviation for the autocorrelations estimate is given by $1/N^{.5}$, under the null hypothes that there is no autocorrelation in the process of which this series is an approximation (Bartlett, 1946). . . . To be significantly

---

[10] Here is a clear distinction between statistical periodicity and perceptual rhythm--a perceptual beat may be established with only one cycle between two peaks.

different from zero (at alpha = .05), [an autocorrelation] must exceed $2/N^{.5}$.
(1981:67)

In a correlogram, these significance levels can be represented as lines above and below the zero line indicating minimum absolute values for a significant difference from zero. These lines demarcate what are known as Bartlett bands (Gottman 1981). (Bartlett bands are shown by the dashed lines in the correlograms presented below in Examples 1 and 2, Chapter 4.) The correlation levels required for significance rise slightly with increasing lag because of the decrease in effective N with increasing lag. These levels also differ from segment to segment and from correlogram to correlogram depending on the length of the segment selected, and the amount of missing data in that segment.

In the analysis, I preferred to refer to levels of significance specified by the Bartlett bands only as a method of comparing degrees of cyclicity between the various types of segments selected. I did not adhere strictly to these criteria in the exploratory phases; I tolerated some deviation from the criteria, even when modelling the waveforms, where this deviation was attributable to the presence of "noise" in the form of an uneven distribution of zeros, and trend (see the following section). In most correlograms a clear alternation between significantly positive and negative values was not established; I proceeded with analyses when an alternation occurred only between local maxima and minima in the correlograms suggestive of cycles in the data. For consideration as a basis for modeling, the correlogram had to peak at least twice at significant values after descending well below the zero line and at equal (within one or two lags, depending on the size of the cycle) intervals between peaks.

3.4.1.3 Noise in the Correlograms: Missing Data and Trend

The uneven distributions of zeros and the presence of trends and shifts in mean in the segments of F_0-series are two factors to which we may attribute irregularities in the correlograms. Stretches of zeros must be treated as missing data, and when long, they seem to affect the smoothness of the series, that is, whether the series are predominantly monotonic, or continuous in slope, between prominent maxima and minima. This condition was not usually obtained in the autocorrelations of series with more than 10-20% missing data, and in some cases the positions of maxima and minima could not be determined because of the lack of resolution caused by this problem.

Interpretations of some autocorrelations were also complicated by the presence of trends or changes of mean in the data. Using the smoothing programs, I took means over every 5 samples of the 25 sps F_0-series, producing series with 5 samples per second. Overall trends in the original series are rendered clearly visible in plots of F_0-series at 5 sps, and can be seen in both examples presented in the following chapter. No attempt was made to compensate for these trends, excepting in the interpretation of autocorrelations taken on series exhibiting visible trend. Further autocorrelation analysis of F_0 fluctuations in these temporal domains will require removal of longer trends. In the present study the question of trend was set aside since segments with relatively little trend were found in sufficient numbers to continue with the analyses [11].

3.4.2 Phase and Modeling

3.4.2.1 Splicing

Assumptions about what happens to underlying tendencies during the silences can be tested by splicing F_0 segments containing comparable periods; this is achieved by the removal of points in numbers both consistent and inconsistent with that period, and examination for systematic changes in the correlograms. This application of autocorrelation analysis, which was suggested to me by Dr. Tom Maguire, reduces the

----

[11]There is again the possibility here that a procedural adjustment results in an inadvertently differential treatment of the sexes, or more precisely of those whose F_0-series tend to fluctuate with a certain rapidity with respect to range. In the speech episodes discussed below, one involves a man with a woman, and the other involves the same man with a different woman. In both, the man's F_0-series contain relatively little overall trend (except for one clear shift in mean in Example 2) while the trend in both the women's F_0-series seems strong enough to obscure the results of the autocorrelation series.

proportion of missing data and simultaneously accomplishes the goal of testing whether period and phase in the cycling of F_0 remain constant across utterances and speakers. Splicing is used in both examples below to examine the research hypotheses.

The assumption that speech rhythm predisposes the timing of the following events is operationally defined here as a constancy of period and phase between utterances and therefore across intervening silences and brief phonations. If this assumption holds, then the removal from that interval of the number of zeros corresponding to a periodicity in neighboring segments of the F_0-series should not disturb that periodicity--in fact, it should enhance the salience of that periodicity to an autocorrelation analysis.

These assumptions guided autocorrelation analyses of F_0-series consisting of selected speech segments spliced together by the removal of numbers of points (mostly silences) equal to one, two, or three of the cycles found in those segments. In the first example, portions of a silence are spliced out to clarify a question of phase relations. In the second example, a speech segment is spliced out of the data along with surrounding zeros in an investigation of phase and period relations which results in a clarification of the period underlying the man's F_0 fluctuations in that episode.

3.4.2.2 Modeling

The final stage of the analysis of the F_0-series was the visual alignment of a waveform to the plots of those segments with demonstrated periodicity. Although the period of this wave is specified by the correlograms, its phase alignment to F_0 peaks and the simultaneous onset is determined visually. The issue of whether or not the PSO hypothesis is suported by a given wave model is therefore a matter of judgement at this stage. The results presented below indicate two examples of speech episodes in which I feel that the temporal occurrence of simultaneous-onset speech is explainable with reference to a wave model. The presence of periodicity in these data is empirically observable in the correlograms.

Such research techniques, in which the fitting of a curve derived from statistical aspects of time-series data is performed visually, have been used before in communications research using interrupted time-series designs. I quote from McReynolds and Kearns' "Single-Subject Experimental Designs in Communicative Disorders":

> Results of single-case experimental designs are generally presented on a graph and direct visual inspection of an individual's performance provides the primary method of data analysis. There are several advantages to this method of data analysis. First, because raw data are plotted and analyzed during each phase of a study, the experimenter is kept in close contact with his results and is in a position to know if the data are conforming to preexperimental expectations. This allows appropriate modifications in the study protocol as problems arise and provides a considerable degree of experimental flexibility. In addition, graphing of raw data permits other members of the scientific community to assess independently the power of obtained results. Finally, when the criterion for acceptance of data is a marked effect evident from visual inspection, variables having a weak treatment effect will be teased out and further explored (1983:26).

Although McReynolds and Kearns are discussing experimental approaches, their comments on visual inspection are highly relevant to the quasi-experimental time-series modelling which I attempt in the present descriptive study.

The technique of fitting a specified waveform to prosodic data has also been considered by other researchers, notably Jassem (1978). In outlining a program of study for the phonetic analysis of pitch phenomena he indicated that

> Procedures for a linguistic interpretation of pitch curves obtained instrumentally have not yet been established. Probably statistical methods, especially those that deal with fitting certain types of theoretical curves to empirical data, could provide a solution (1978:365).

Maguire (1985) discusses the application of time-series analytic methods to detect cyclicity in pitch fluctuations, and describes the problem as consisting in part of the removal of noise: "We can use techniques to accentuate certain parts of our data much like computer enhancement is used in satellite photography, or like stains are used in biology" (1985:7). In the present research, we wish to filter out noise to reveal underlying periodic tendencies, and the waveforms which I fit to the F_0-plot in the chapter on results below are visual models of this tendency.

The visual alignment of a hypothetical wave does not represent a formal attempt to model the F_0-series. Strictly speaking, the only modelling done in this study is in the description of an F_0-series in terms of the periods in its autocorrelation series. The actual waveform of the F_0-series is ambiguously described by indications of periodicity--period being the only parameter specified. In order to examine the indicated periodic patterns in the F_0-series for phase alignment and for consistency with the PSO hypothesis, I required a "measuring stick." More a visual aid then a model, this measuring stick took the form of a sinusoidal wave with peaks occurring in the period and phase of the peaks in the F_0-plot[12].

The methods of visual fitting are demonstrated in the exposition of results. The goal is to check on the correspondence of autocorrelation results to the points specified by the hypotheses. Points specified by the initial PSO hypothesis are 1) the pitch peaks preceding a point of simultaneous onset and 2) the point of simultaneous onset. These

[12] Note that the shapes of the 25 sps F_0-plots are not sinusoidal--peaks are frequently sharper than valleys (i.e. discontinuities in slope tended to appear at the maxima in the F_0-series). Modifications of Pierrehumberts "string" model of F_0 declination by Cooper & Sorenson (1980) indicate that F_0 peaks are not discontinuous but "concave downward." The effect of sharpness in the peaks of the 25 sps F_0-series is probably an artifact of the sampling rate, and would therefore be consistent across the spectrum of F_0 fluctuations: fluctuations or changes in direction occurring above certain critical rates will result in discontinuities in the slopes of series measures. Given the nature of these discontinuities in slope, it seemed appropriate to fit the series with sinusoidal-like (spline) curves with continuously varying slopes. The choice of this type of wave was also appropriate in anticipation of spectral analyses, as sinusoidal components of a Fourier series are the independant parameters of a frequency-domain analysis.

hypothesized points were slightly modified in the examinations of two instances of simultaneous-onset locutions presented below.

4. Results

The results presented in this chapter demonstrate some methodological procedures and descriptive principles. They are not presented as evidence for or against general hypotheses concerning the role of prosodic coherence in organizing interaction. Though consistent with the projection of simultaneous onset hypothesis (that simultaneous onsets occur in the period and phase of preceding speech), the two examples below do not establish the prevalence of prosodically coherent speech, nor do they establish the perceptual significance of the described F_0 patterns. The examples do indicate the possibility for rigorous description of the periodic fluctuation of a measurable parameter of vocal interaction.

As judged by autocorrelation analysis, periods and phase in the cycling of F_0 do remain constant across some segments of speech. Investigations of the projection of simultaneous onset hypothesis proceeded differently in the two examples presented below. Each of the analyses is problematic in a different way, highlighting different technical, methodological, statistical, and design problems, though in both, the operational definition of prosodic coherence as a constancy of period and phase across utterances has explanatory value. In both cases the results are consistent but slightly discrepant with those predicted by the PSO hypothesis. In both a point of simultaneous onset is preceded by simultaneous speech and a pause, and the timing of the simultaneously onset speech is consistent with intervals in preceding nonsimultaneous speech by one of the participants.

In Example 1, the point projected to does not appear to be the point of F_0 *onset* (which was not acoustically simultaneous in the F_0-series). Rather than onset, it is the F_0 ***peaks*** that were physically more nea simultaneous. 260 msec periods are found between pitch peaks in the speech by a male speaker and the consequent occurrence of simultaneous pitch peaks. In Example 2, the point of simultaneous onset is part of an F_0 peak in both participants, and this point is found to occur at 560 msec intervals from pitch peaks in speech by the male speaker, these intervals both preceding and *following* that point. In both examples the female partners' F_0-series showed evidence of comparable periodicities (at roughly twice the wavelength in

Example 1 and at the same wavelength in Example 2), but these periodicities were not unequivocally established in the correlograms.

In the conversations I recorded, not all speech heard as simultaneously onset was preceded by speech with cyclically autocorrelated F_0, and when this was the case it supported the idea of projection at varying levels of temporal precision. I should point out that the PSO hypothesis does not predict that all occurrances of simultaneous onset will occur in precise period and phase relation to previous periodic F_0 movements, but merely suggests that such an occurrance might be found with some regularity based on the notion that conversation is often prosodically coherent. The results presented here do demonstrate clearly that objective study of period and phase in the cycling of F_0 in discourse is attainable in time-series analyses such as autocorrelation.

4.1 Example 1

My first example is from an episode containing 14 seconds of speech by male subject D with simultaneous and non-simultaneous speech by female subject S. The episode seems to be one in which D is making a commentary on something S has said, with S's speech indicating both agreement with D's commentary and an intention to continue speaking once D has finished his commentary. Out of the 14 second F_0 transcription of the episode I chose a set of utterances lasting 3 seconds for closer study. Following approximately 1 second of mutual laughter with phonation and a .8 second pause (absence of phonation), D begins a clause which both he and S finish together (see Figure 4.1).

A very clear and compelling beat is heard as being produced mutually by S's "else. . .yeah" and D's "yeah. . .righ." This impression is not clearly borne out by the acoustic intervals indicated in the F_0-plot of these syllable utterances. The utterances heard--even after repeated listening--as uttered simultaneously, have F_0 onsets 60 to 100 msec apart as measured in the F_0-series. The F_0 peaks of S's "else" and D's "yeah" are also heard as being simultaneously uttered on a single beat but peak in F_0 at apparently distinct times. There is nonetheless demonstrable periodicity in D's speech.

4.1.1 Period

I selected the 75 data points of D's entire F_0-series for autocorrelation over 35 lags. The resulting correlogram exhibits five cycles, three of which are clearly significant (see Figure 4.2). These results strongly suggest a cycle in the data with a period of roughly 6.5 points or 260 msec.

This inference could be made with more confidence if the maximum at 21 lags was significant, or at least positive. Note that the minima in this region are occurring between five and seven lags distant, indicating a cyclicity over this range of points in the series, that is, at the period established in other regions of the correlogram. An overall dip in the autocorrelation series in this region may be caused by trends in the F_0 data, masking the significance of this cycle in the autocorrelation series. This problem is illustrated more clearly in the consideration of S's F_0 data.

I examined separately the two segments of D's speech divided by the central 1.3 second pause; segment 1 consisting of speech up to the first "it's a," and segment 2 beginning with the repetition of "it's a." Though exhibiting dramatically the effect of trend, the correlogram of the first segment peaks at six and thirteen lags (see Figure 4.3 [a]). The second segment has autocorrelation peaking only once at fifteen lags and with an irregular minimum (see Figure 4.3 [b]). These results would certainly suggest that the 6.5 point cycle indicated in Figure 4.2 is due primarily to autocorrelation in segment 1.

Note that this may not mean an absence of periodicity underlying segment 2 considered in the context of segment 1. One such interpretation would be that F_0, in segment 2, cycles with roughly twice the period of the F_0 cycle in segment 1. There may also be a shift in phase involved. For the purposes of modeling, it is generally better to rely on the autocorrelation of both segments together because of the reduced sample sizes in the smaller segments and because the autocorrelation of segment 2 peaks only once, failing to specify an inherent periodicity.

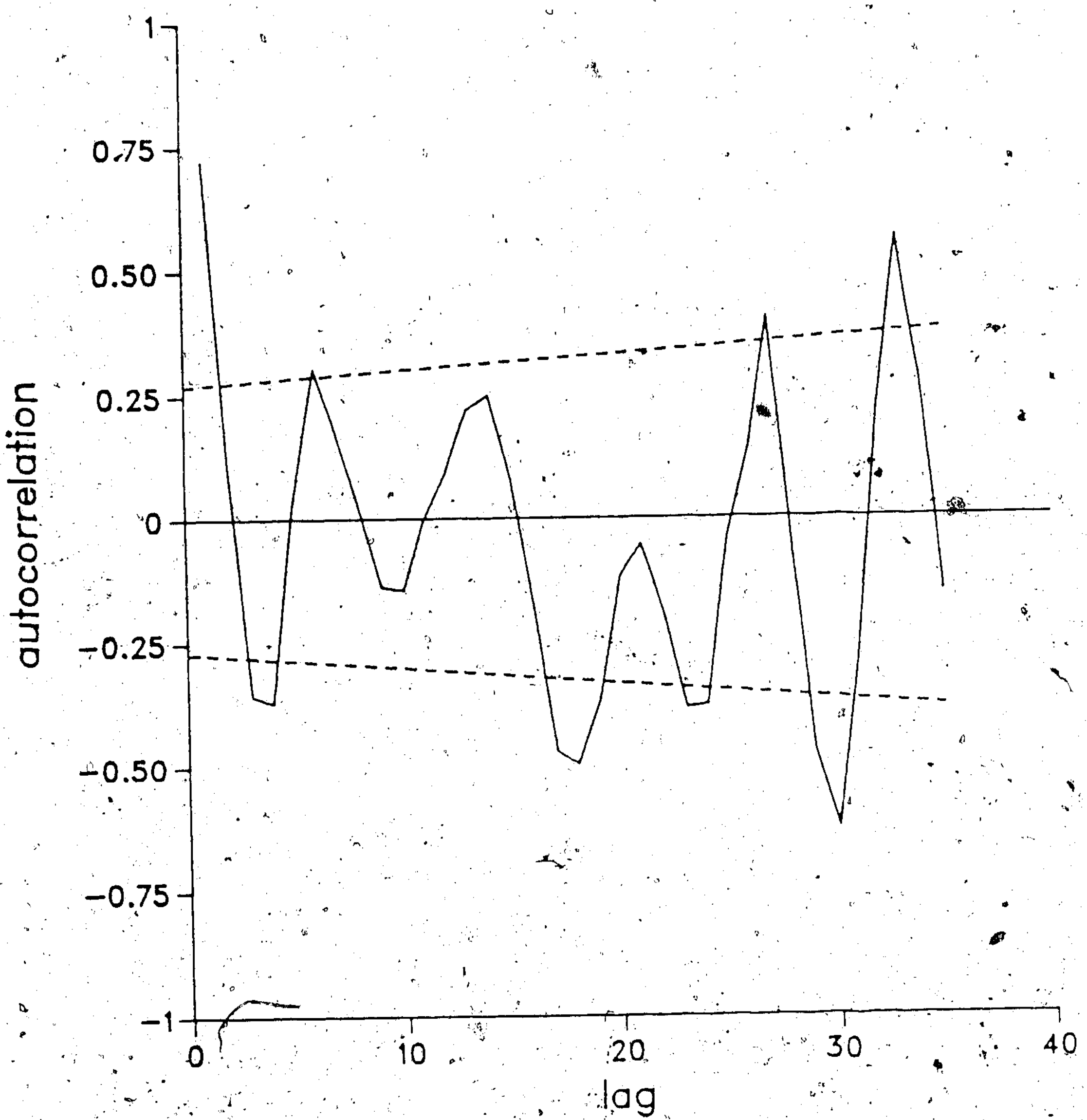


Figure 4.2 Correlogram of D's F_0-series (lags increase by 40 msec intervals).

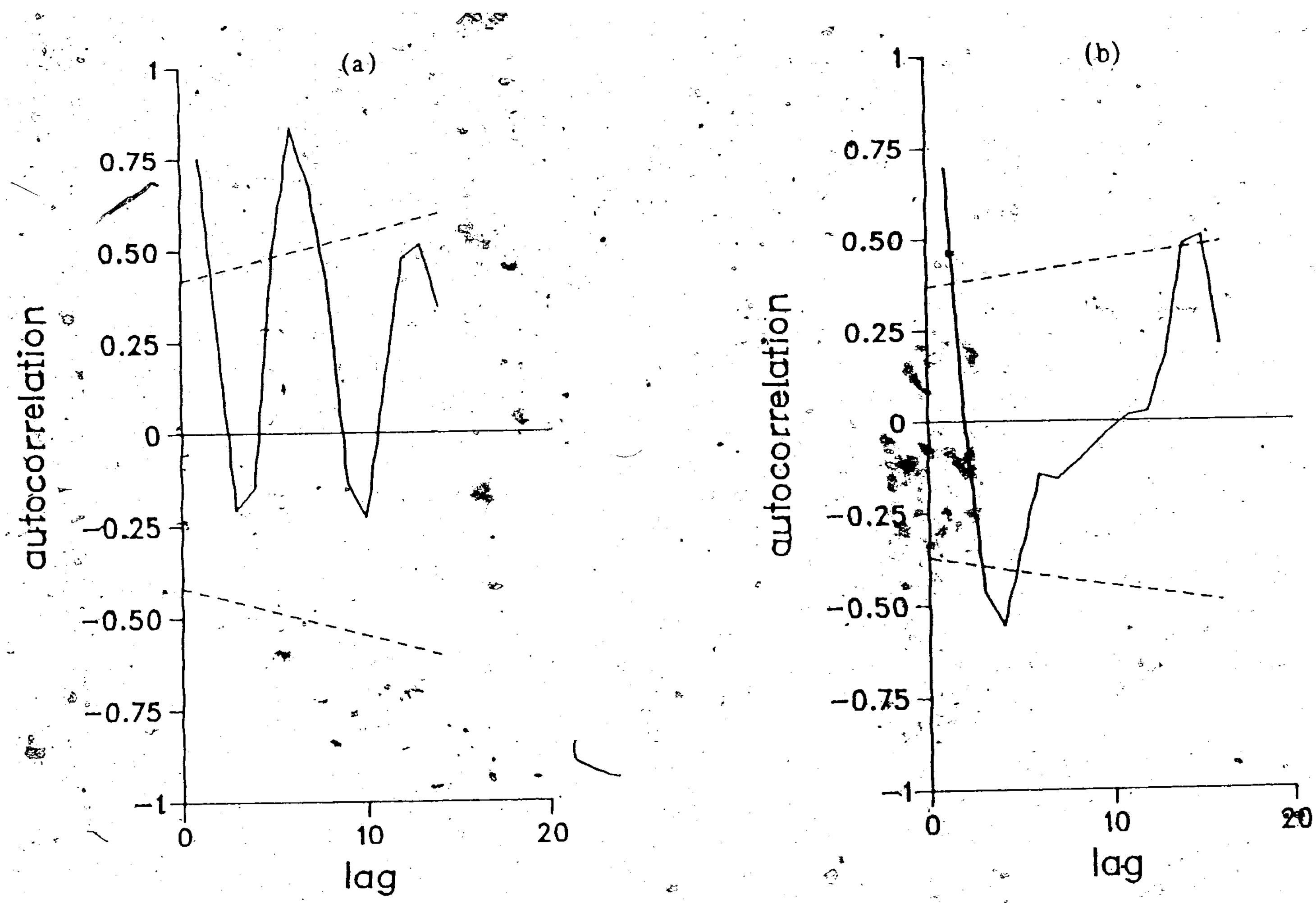


Figure 4.3 Correlograms of segments of D's F_0-series: (a) segment 1 (b) segment 2.

Splicing techniques demonstrate that the cyclicity indicated in D's F_0-series is due to properties of both segments considered together.

4.1.2 Phase

I examined for period and phase relations between segments 1 and 2 by splicing the segments; removing systematically different numbers of points (all zeros) from the intervening gap. Figure 4.4 shows the autocorrelation results of D's F_0-series with 4, 6, and 8 points (160, 240, and 320 msec), spliced out of the gap intervening between the two segments. Based on the hypothetical presence of a 6.5 point period underlying these segments, the second splice, in which an interval roughly equal to the hypothetical cycle is being removed, would be expected to show stronger cyclity relative to the other two splices, in which intervals incommensurable with the hypothetical 6.5 point cycle are being removed.

Though the first two peaks in the correlogram remain relatively undisturbed in the four and eight point splices (Figures 4.4[a] and 4.4[b]), the positions of the third and fourth peaks in the correlograms are displaced in these splices, resulting in predominantly aperiodic correlograms.

A further test of the existence of the underlying tendency is provided by splicing thirteen points (containing two whole, but unrealized, cycles of the hypothesized period) from the gap (see Figure 4.5)[13]. Though the resultant decrease in N widens the Bartlett bands to where the cycles of the series are technically insignificant, the regular alternation in the series is clearly indicative of a periodic cycle in the data.

4.1.3 Modeling

Autocorrelation analyses of this example have indicated a tendency towards a 260 msec periodicity in the F_0 of D's speech and suggest that the periodicity is strongest before his pause. The next stage of the analysis involved the fitting of a wave to the F_0-plot and checking for a

---

[13] In this splice one point of data is sacrificed--I chose the last point of D's first "it's a" just preceding the pause.

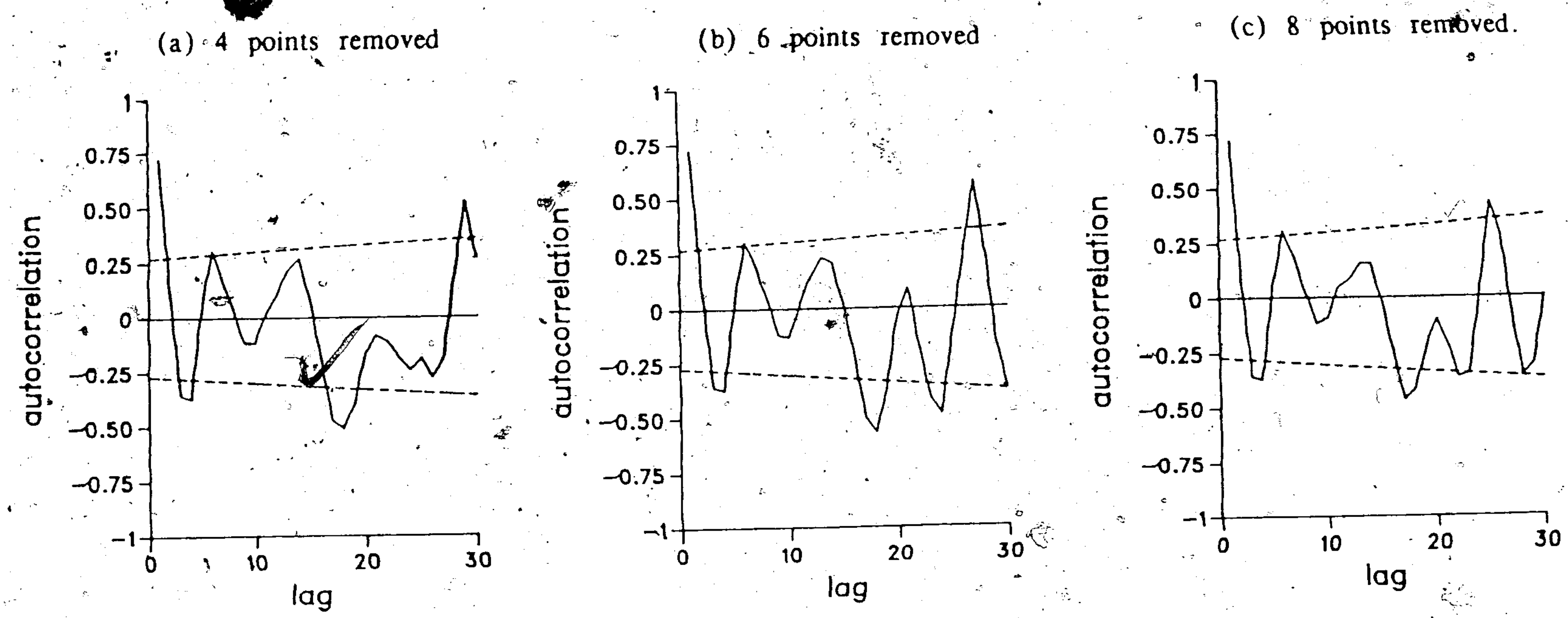


Figure 4.4 Correlograms of D's spliced F_0-series.

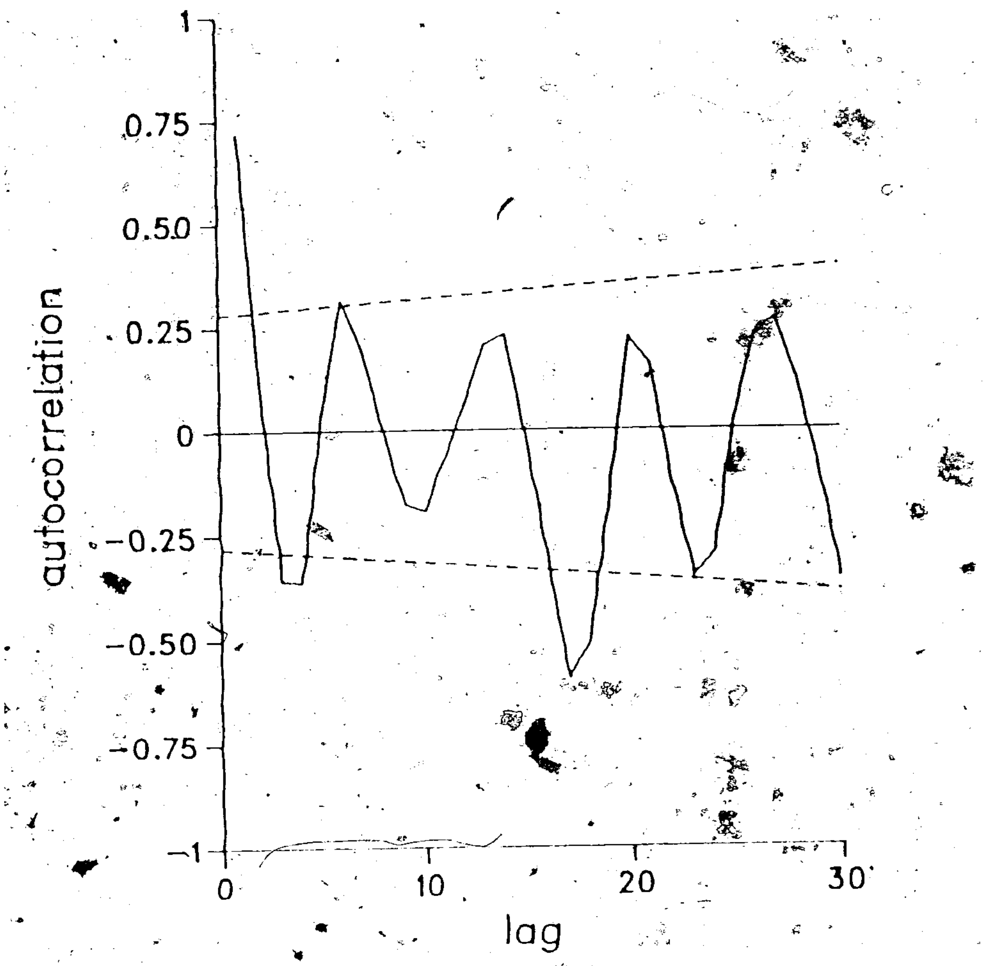


Figure 4.5 Correlogram of D's spliced F_0 series: 13 points removed.

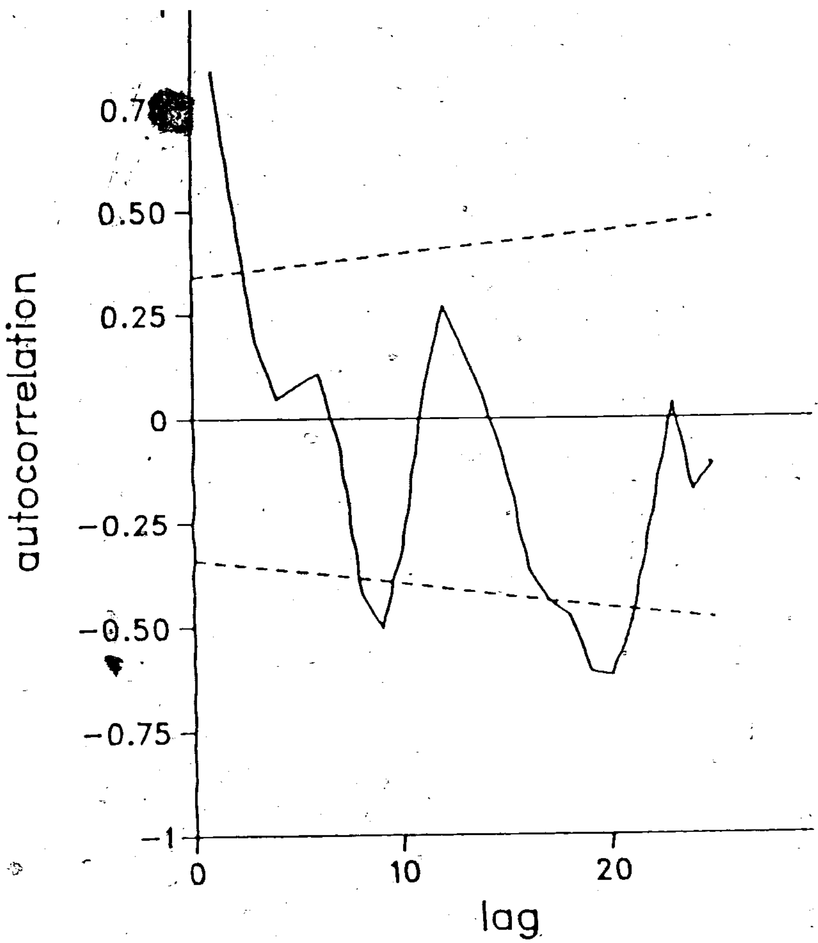


Figure 4.7 Correlogram of S's F_0-series.

period and phase alignment of the simultaneous onset speech with preceding F_0 peaks. This is in effect the "test" of the PSO hypothesis.

The most reliable periodicity was the 6.5 point 260 msec wave detected in the autocorrelation series of the 75 data points of D's F_0-series. I therefore attempted to align a 260 msec wave to the F_0-plot of these points (see Figure 4.6). An alignment of the wave with regularly appearing F_0 peaks was indeed clearest in segment 1. In that segment, the F_0 peaks occur on the syllables "stead-," "pid-," and "Eng-," and these peaks occur 240 to 280 msec apart in the F_0-series.

I conjectured that this periodicity continued to underlie D's immediately following utterances, consistent with the notion that speech segments cohere rhythmically. The peaks of the 260 msec wave, phase aligned with the first three F_0 peaks of segment 1, also align with both of D's pronunciations of "it's." This observation, convergent with the splicing results, suggests that the phase and period of the cycle underlying the timing of D's F_0 fluctuations is not disturbed during the pause. If it is indeed the case that there is an underlying tendency for F_0 peaks to occur every 260 msec in D's speech during this episode, then two peaks of the hypothetical wave occur during the 520 msec silence, as seen in Figure 4.6.

The next peak of the wave is on the F_0 peak of "pid-." It is again appropriate to consider the lexical data here as convergent with the wave analysis: the hypothetical wave has now peaked on both instances of a repeated syllable twice, on the two "pid" syllables and on the two "it's." (Note also that the point of this second "pid" syllable which aligns to the wave is both an F_0 peak and a point of onset.)

The following peak of the wave occurs on the "n" of "pidgin." It seems to be at this point that D begins to "draw out" his phonation, apparently waiting for S to complete her contribution to the clause. This interpretation is relevant as a possible explanation for the "delay" in D's next F_0 peak relative to the hypothetical wave. At the end of his extended "n" D produces a very clear F_0 peak on the syllable "yeah." This peak is clearly *out* of phase with the wave model, occurring 100 msec after the peak of the wave. I nonetheless continued to assume that the wave model applies to following speech in the phase set by preceding speech (D's segment 1).[14]

After D's and S's syllables "yeah" and "else" (which are heard as occurring simultaneously), there is a mutual silence followed by their syllables "righ'" and "yeah" (heard as beginning simultaneously--at a point of simultaneous onset). The hypothetical wave peaks once during this silence, and once again--100 msec after the onset of phonation in D's "righ'," but within only 20 msec. of of the F_0 peak of this utterance. Furthermore, this location is also within 20 msec. of the F_0 peak in S's "yeah." I regard this as an instance of projection of simultaneous *peaks* rather than of simultaneous onsets since there was no acoustic of simultaneous onset, and since the simultaneity of the peaks may correspond to the perception of simultaneity. This point conforms to the notion that a simultaneous onset (or more generally, "simultaneity") will occur in the period and phase of preceding speech.

4.1.4 Periodicity in S's F_0-series

The foregoing analysis has produced results consistent with the notion that participants in conversations can time their F_0 productions so as to preserve periodicity and phase of F_0 cycling. This analysis proceeded on the basis of D's F_0-series, and appealed to S's series only as it coincides with D's final syllables. The phenomenon of interest, this coincidence, is a property of the interaction of the two individual pitch production texts. In the present example there is speech by both participants preceding the heard simultaneous onset. [illegible] D's speech was presented first by itself simply because the findings were clearer in his data.

The autocorrelation analysis of S's F_0-series is not straightforward. There are aspects of the timing of S's speech, even preceding the final two syllables "else" and "yeah," which lend an impression of rhythmicity, both heard (on the audio tape) and [illegible] (on the F_0-plot).

---

[14]This analytic assumption might be validated by the functional interpretation that D is waiting to "join in" with S's speech. This would imply that D's speech is delayed as a *response* to S's peak (which does occur within 20 to 60 msec of the peak in the hypothetical wave). Note that the peak of the wave also aligns with the beginning of D's F_0 rise.

In an autocorrelation analysis, a weak tendency towards cyclicity appears in the data at periods approximately twice as long as those found more strongly in D's F_0-series. Details of the analysis reveal some limitations in the techniques but also indicate directions for improvement.

A technical problem with S's F_0-series concerns actual missing data (see dashed line in Figure 4.1). Data on the F_0 of S's "else" are masked by D's "yeah"; since F_0 is by definition the lowest frequency in the auditory pitch range, the Pitchmeter will usually prefer a male voice. I reconstructed the F_0 on S's "else" by extending the last observation across the two missing points; thus avoiding the temporal specification of a peak which would influence the results.

Another problem with S's F_0 data in this example is with the proportion of silence between her first few utterances. S's first two "mm"s are brief and quiet (indeed their audibility to her interlocutor may be in doubt)[15]. Because of it's relative temporal isolation and short duration, S's first "mm" is difficult to incorporate into an autocorrelation of the rest of the segments. Neither editing, by splicing out points in the silence, nor filling, by interpolation to the following points, are motivated here. Inclusion of this utterance and the following silence into an autocorrelation analysis of the remaining speech would have the effect of reducing sample size to the point where the correlogram would be uninterpretable. This is not to say that the timing and pitch of the utterance are not of interest, but that in the absence of patterns to which these points may be fit, it is difficult to incorporate them into an autocorrelation analysis of the rest of the episode.

Autocorrelation analyses like those applied to D's F_0-series failed to detect periodicity at statistically significant levels in S's speech, and the results indicate a need to account for trend in the data. need to account for trend in the data. An autocorrelation of the 52 point

---

[15] In the context described above (D's comments are seen as having occurred during S's "turn"), these vocalizations might indicate a desire to begin speaking; S seems to be delaying her speech because D is talking, though she clearly has something she also means to say. S's entry during D's pause, with a noun phrase fitting D's speech semantically, syntactically and, it is my impression, prosodically, suggests that her preceding "mm"s can be taken as markers of both imminent speech and hesitation. The availability of these vocal gestures for a close analysis is enhanced by the proximity of the microphones to the larynx.

segment of F_0 readings from S's second "mm" to her "yeah" at the end of the episode is presented in Figure 4.7. The correlogram also clearly exhibits an overall downward trend indicating general changes in F_0 larger than the span of the 25 lags to which the autocorrelation could be taken. This may be attributable to the declination which accompanies S's declaratively voiced "a pidgin of everything." The suggestion of a 11 to 13 lag cycle in the correlogram of S's speech (see Figure 4.7) may be related to the 440-520 msec interval in her "else. . .yeah" (12 points = 480 msec) (these being the syllables heard to coincide rhythmically with D's "yeah. . .righ'"). It is not immediately apparent though how a 480-520 msec wave might fit to S's overall F_0-series, in which the fluctuation seems to consist predominantly in downward trend.

The existence of this trend is confirmed visually by inspection of a plot of the episode series smoothed down to 5 sps (see Figure 4.8). Here an overall rise-fall in the data is apparent (which may itself be regarded as a cycle with a period of roughly 1.2 seconds), and this overall trend undoubtedly lowers correlations at certain lags in the correlograms of S's speech, possibly masking the significance of components with shorter periods, such as a 480-520 msec cycle.

These results are presented here primarily as indicative of a need to further refine the the techniques of autocorrelation analysis of F_0 data. They do not result in a waveform model, but the results are consistent with the notion that temporal intervals underlying both F_0 productions in this episode of simultaneous speech are commensurable--S's with a period roughly twice as long as D's.

4.2 Example 2

The second episode I selected for analysis seems at first to be an instance of one speaker providing an apparent semantic and syntactic continuance of another's preceding speech (although an apparent conflict developes later in the five to six second stretch of simultaneous speech in this episode, in what appears to be a "contest for the floor".) The episode consisted of male subjct D (the same as in Example 1) speaking alone and then in simultaneous speech with female subject J: during the simultaneous speech, there is a mutual silence followed by a heard simultaneous onset. The simultaneous onset occurs in the 7th second of this 10 second episode, about 1 second into the simultaneous speech. A possible functional or "common-sense" (i.e. impressionistic) interpretation of the non-simultaneous speech might be that D makes a statement of a political condition, in the manner of leading up to a conjecture or prediction. As he begins to utter what apparently would have been this prediction, his interlocutor J enters with her understanding of the consequences of D's statements, and D attempts to "recycle" the text of the conversation so that he may continue his line of reasoning (see Figure 4.9).

The analysis indicated that, during the first three seconds of simultaneous speech including the point of simultaneous onset, D refers very clearly to the rhythm established in his preceding speech. The simultaneous onset is also an F_0 peak in J's series, and is followed in D's series with an F_0 peak only 40 msec later (see Figure 4.9). Both the original (simultaneous onset) and the modified (simultaneous peaks) hypotheses would predict the prosodic coherence of this simultaneous onset with preceding speech. Validation of the projection model in this example by autocorrelation analysis was initially obscured to me by the simultaneous speech preceding this point of simultaneous onset--the segment was edited out to render explicit the pattern underlying D's data.

4.2.1 Periodicity

Three of the syllables of D's first set of utterances (segment 1) are heard as falling on a very regular beat; "peo-," "this," and "-serv-." A weaker, faster beat may be heard on "and," and on "att-" followed by "peo." In this case (in which the periods are two to four times as long as in Example 1), the perceptual rhythm seems to correspond well to the acoustic record: the F_0 peaks on these syllables do seem very regularly placed in time.

I first examined the speech in segment 1. Clearly significant peaks in the correlogram of this segment occur at 28 and 55 lags (see Figure 4.10). There is another peak at 44 lags, and

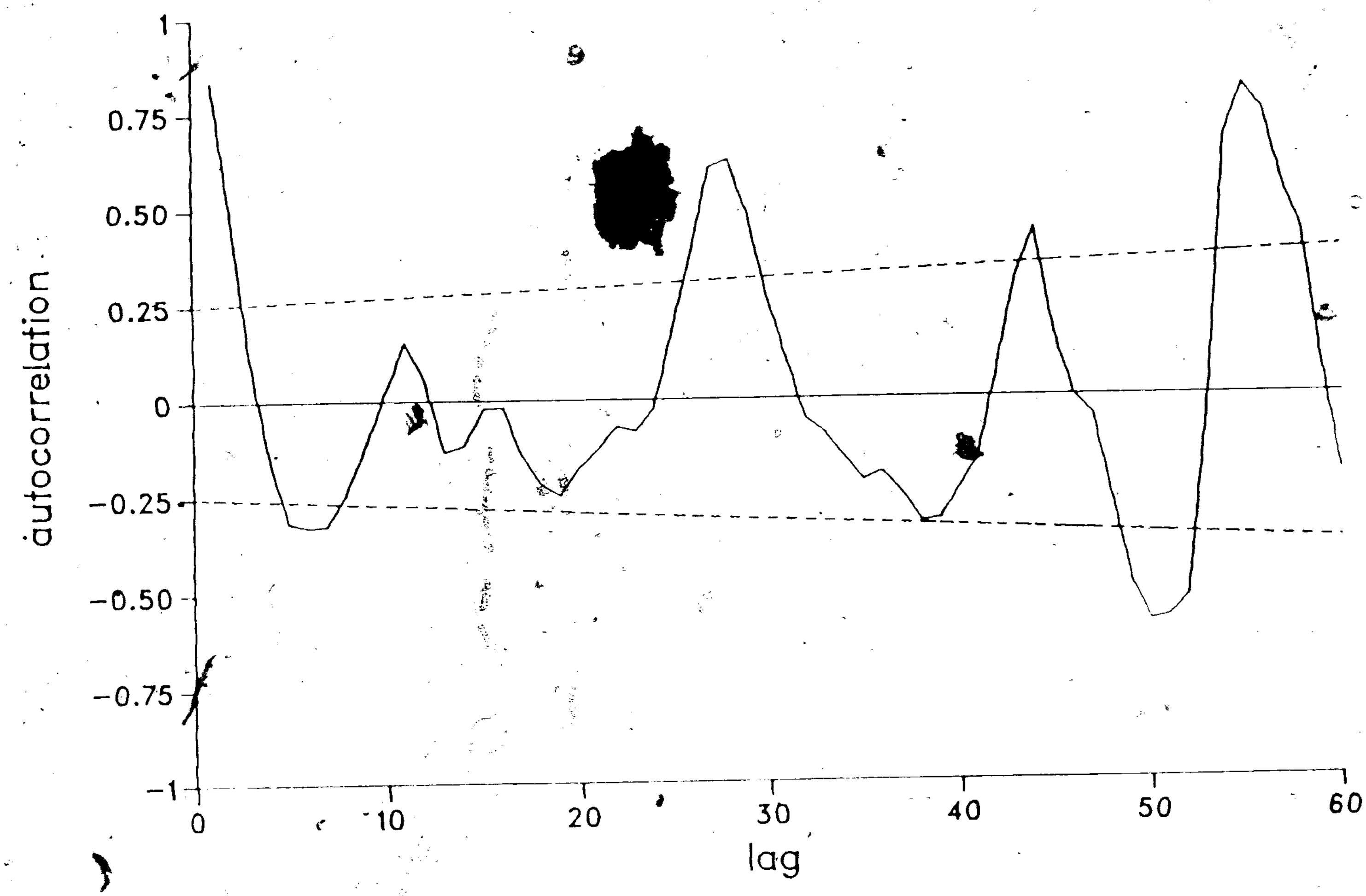


Figure 4.10 Correlogram of D's F_0-series: segment 1.

a local rise in the correlations between 10 and 16 lags. It can be stated with some confidence on the basis of these results that there is a 28 point 1.2 second periodicity in these data, and that there appears also to be a weaker cycle at roughly half that period. Examination of the F_0-plot (Figure 4.9) shows that the F_0 peaks on "peo-," "this," and "-serv-" tend to occur approximately 1.1-1.2 seconds apart, but note that midway between some of these syllables, there are other syllables containing smaller F_0 peaks, e.g. "and" and "att-."

Following segment 1 of D's speech, there is a 1 second pause, then speech by D followed by J 160 msec later, producing a 580 msec stretch of simultaneous speech. D and S both stop simultaneously, pause for approximately .28 sec, and then begin again, both at the same time.

The beat becomes unclear in speech immediately preceding the point of simultaneous onset (in segment 2.1), but seems nonetheless to also continue past this point. I autocorrelated D's F_0-series following the pause (segment 2) and found a very weak suggestion of a 14 lag 560 msec cycle (see Figure 11[a]). A correlogram of the series corresponding to D's "recycling" speech ("They're gonna, yeah, they're, they're" [segment 2.2]) peaked at 13 lags (see Figure 11[b]). An autocorrelation of D's F_0-series from the point of simultaneous onset on (segments 2.2 and 2.3) peaked at 13 and 30 lags with a 12 lag interval between clear minima (see Figure 11[c]).

Although no statistically significant periodicities were established in these autocorrelations, the results are not entirely negative. It may be that the salient period changes within one of the segments, e.g. on D's "yeah." The results do suggest that D's F_0-series expresses an underlying tendency to peak roughly twice a second, and that this tendency underlies his speech throughout the episode both before and after a point of simultaneous onset.

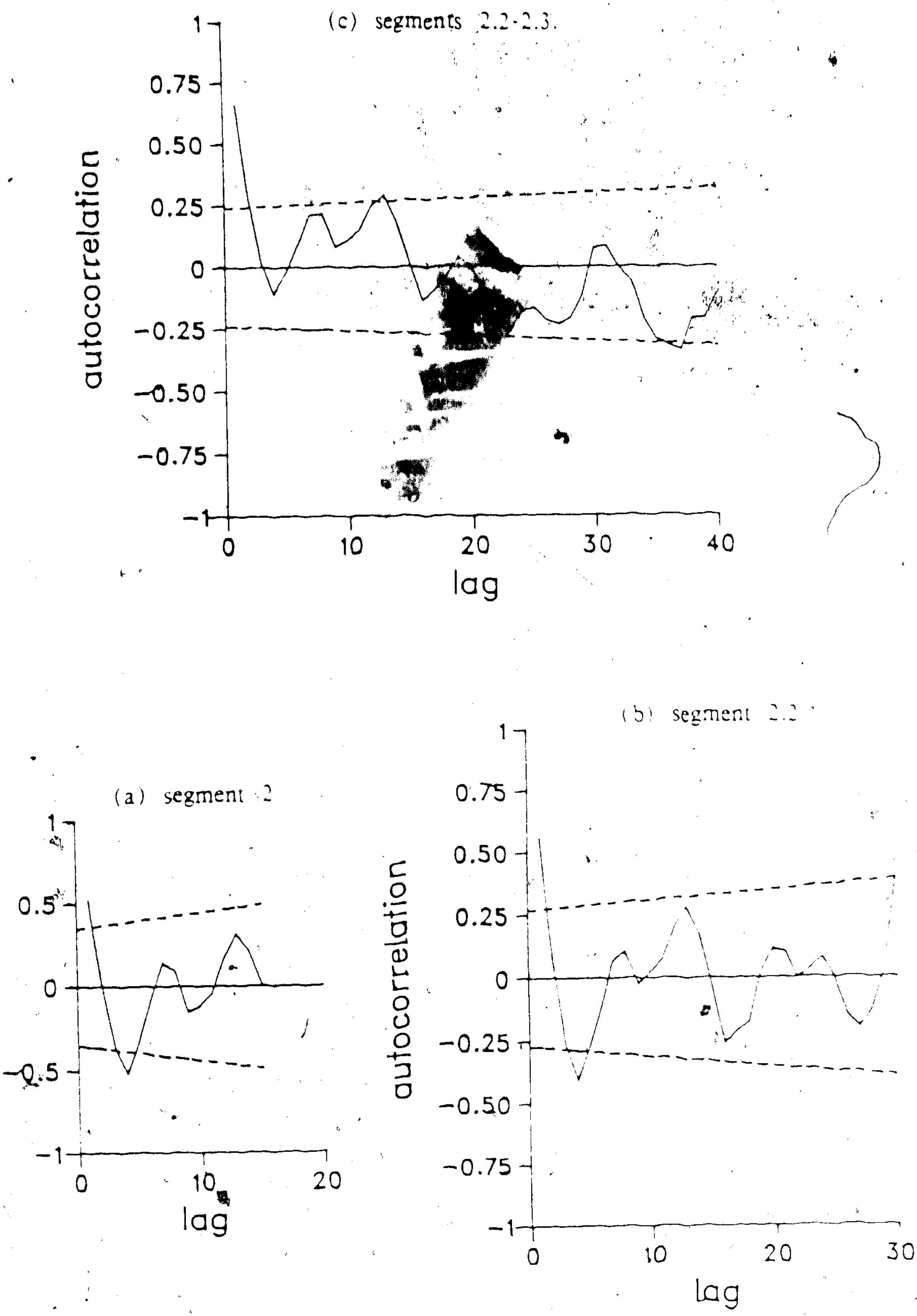


Figure 4.11. Correlograms of D's F_0-series.

4.2.2 Phase and Modeling

D's F_0-series in segment 2.1 seems not to fit a 560 point wave aligned to D's preceding non-simultaneous speech. D's phonation in segment 2.1 begins 1240 msec after his last F_0 peak. There is no obvious peak in D's speech and neither a 28 point nor a 14 point wave aligns clearly with any points in D's speech in segment 2.1. The appropriate modeling of the period and phase relations between segments of D's speech became apparent only when segment 2.1 was spliced out.

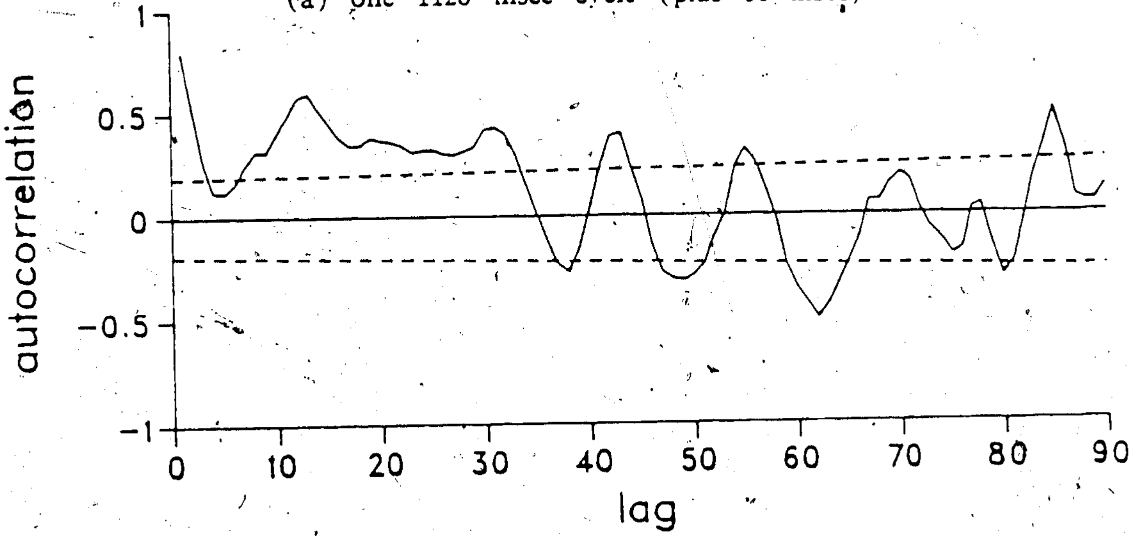
Again making the assumption that there is nonetheless an underlying rhythmicity, with phase and period set by preceding speech, I considered the speech following the simultaneous onset (segments 2.2 and 2.3) for phase relations to segment 1. Speculation that a 28 point wave continues to underlie areas of the F_0-series is supported by the plausible alignments (see Figure 4.9): peaks of the hypothetical wave align with the point of simultaneous onset, with D's final "they're" in segment 2.2, and with his final "me." Furthermore, there seems to be a peak in F_0 on "yeah" occurring at points directly in between the peaks of the 28 point wave, again suggestive of another wave twice as fast.

Consolidation of phase relations between the various segments of D's speech led to support for the existence of a 14 point, 560 msec periodicity. Assuming first that there is a tendency towards a a 28 point periodicity in D's F_0-series which is not expressed in segment 2.1, I spliced out 28 points including segment 2.1 and some surrounding zeros, and autocorrelated the resultant series. Some further experimentation demonstrated the cleanest autocorrelation of this series was obtainable if an additional 2 points were also removed [1]. The result of this analysis is a clearly cyclic correlogram with some disturbance at lower lags, indicating the existence of a 14 point periodicity in the segments of D's speech as spliced (see Figure 4.12[a]).

----

[1] At the present sampling rates it is difficult to tell whether an adjustment of the wave "phase" by this small amount (7%) is caused by an actual phase shift, a slight cumulative lengthening of the period, or simply an error in the alignment of this wave to to the F_0-plot. It may also be a systematic aspect of the data.

(a) one 1120 msec cycle (plus 80 msec) removed



(b) four 560 msec cycles (plus 80 msec) removed

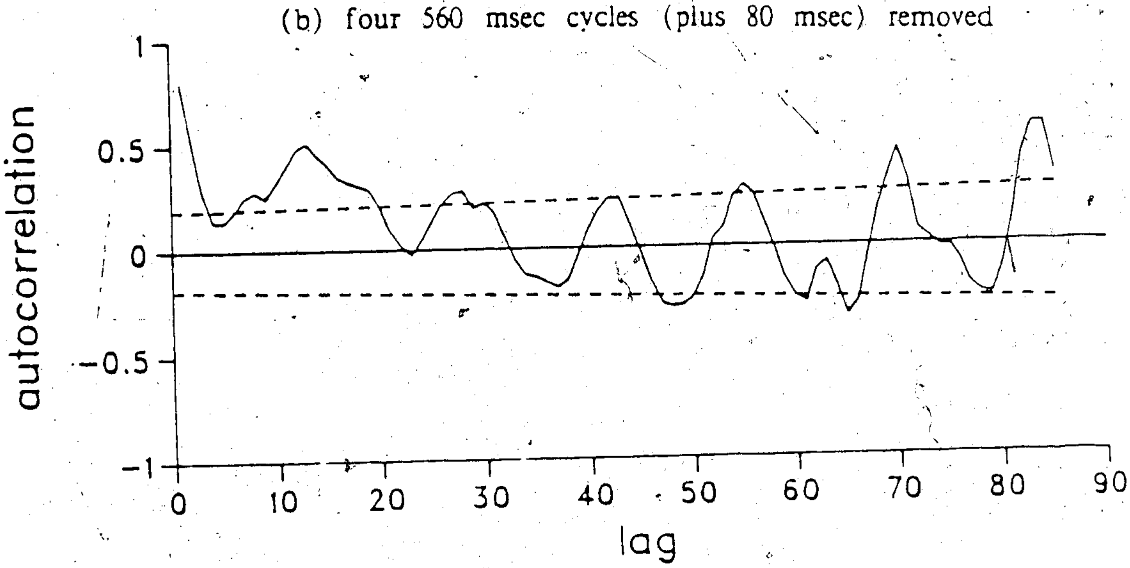


Figure 4.12 Correlograms of D's spliced series.

Further splicing was motivated by the 14 point figure since there were still two runs of 14 or more zeros in the data: one after the initial conjunction "and," and one remaining in the pause after "conservatives" after the 28 point splice, meaning that there seem to be three 560 msec point cycles in the interval between segments 1 and 2.2. Cyclicity of the correlogram of the series spliced on the assumption of a 14 point wave was the clearest of all, with a great deal less disturbance in the earlier lags (see Figure 4.12[b]). Although the first few minima in the autocorrelation fail to reach significance, the maxima are clearly periodic. The vague appearance of an overall fall-rise in the autocorrelation series may again be caused by a trend in the data. Figure 4.13 is a plot of the F_0-series of this episode smoothed to 5 samples per second, and shows clear evidence of a shift in the the mean of D's F_0-series coincident with the point of simultaneous onset.

Figure 4.14 depicts the superimposition of 560 and 1120 msec waves representing the results of the correlograms shown in Figure 4.12. I leave it to the reader to note that these waves do fit visually with the cycling of D's F_0, and with the point of simultaneous onset. Peaks of the 560 msec wave occur on D's syllables and silences as follows: "and," silence, "att-," "peo-," silence, "this," "-pens," "-serv-," silence, silence, "(stutter)," "they're, "yeah," "they're," the "s" of "seems," and on "me."

Excess trend prevents detection of periodicity in J's F_0-series, but in the F_0-plot (see Figure 4.9), J's segment 2.1 contains a peak on "gon-," and 360 msec later, a small local peak on "-lect." It is interesting to note that the simultaneous onset follows this last peak by exactly 360 msec. Furthermore, very small peaks occur exactly every 160 msec in D's F_0-series in segment 2.1, and following a 320 msec gap two more peaks occur in segment 2.2 at 160 msec intervals.

Speculations involving these visual impressions may lead to wave models involving the simultaneous onset in this example in which phase is not shifted during the gap. The presence of these waves may have something to do with the phase shift found in the larger periods. The salience of these periodicities to autocorrelation analysis is marred by the presence of trends and shifts of mean in the data (indicated visually in Figure 4.13). The removal of these non-stationarities may prove especially promising in further study of these and other conversational speech episodes.

5. Conclusions

In this final chapter I will provide a summary of general findings with some discussion of participant perceptions, followed by reference to theoretical implications concerning the conversation as one text and the units of analysis appropriate to its description. Brief mention of some methodological implications concerning time-series modeling of F_0 is followed by a summary of some directions for further research and a view of the conversation as an autonomous social system.

5.1 Summary of Findings

5.1.1 General Findings

My findings can be summarized in four statements. The first three are some claims respecting properties of certain segments of speech, and the fourth indicates the state of the research hypotheses concerning simultaneous onsets. All are illustrated in Examples 1 and 2; and all are to be thought of as possibilities or predispositions, and not as probabilities--they are descriptive rules not experimental results.

1. <u>Correlograms of F_0 specify periodicity.</u>

   Fundamental voice frequencies in conversations tend to cycle with specifiable periodicity. More precisely, some segments of F_0 values at 25 sps are found to be cyclically autocorrelated, meaning that the series tends to be correlated with itself at periodic intervals of time. Such periods were found to be 260 msec in duration in Example 1 and 560 msec in Example 2.

2. <u>F_0 peaks specify phase.</u>

   In these conversations between native speakers of North American English, most segments in which I was able to identify periodicity contained F_0 peaks occurring consistently at that period. These points can therefore be used to specify initial phase regardless of the other characteritics of the F_0 waveform in those segments. The initial phase is thereby set at 90 degrees, meaning that the cycle initiates at it's maximum value. This specification expresses the meaning of the PSO hypthesis in which simultaneous onsets (or peaks) are hypothesized to occur in phase with F_0 peaks.

3. Period and phase may remain constant between utterances. The F_0-series waveform can be crudely modeled by specifying period and phase, and this model may be found to apply across silences with little or no change in these parameters. This means that pitch peaks tended to appear in my data at regular periodic intervals in utterances separated by silence, as if the waveform persisted during silence.

4. Some simultaneous onsets occur in the period and phase of preceding utterances. Heard simultaneous onsets of speech occur at intervals from preceding periodic F_0 peaks equal in length to the intervals between those peaks (or integral multiples of those intervals). At present I regard Example 1 (in which the peaks of F_0, not onsets, occur in the described temporal relations to preceding speech) as indicating the limited temporal precision with which this claim is supportable by investigations of suprasegmentals such as F_0 alone[2].

5.1.2 Participant Perceptions

The waveform models discussed in these claims are not designed to describe participant perceptions. I do however suggest that if participants' perceptions of prosodic rhythms in these segments could somehow be identified (non-experimentally), they would at least be commensurate with the specified periodicities (at probably twice or half the indicated period, if not identical to it), and systematically related in phase.

In Example 1 the rhythm which I hear as created by D's "yeah. . .righ'" and S's "else. . .yeah" occurs on every *other beat* of the 260 msec wave modelled on D's earlier speech. During this earlier speech, the 260 msec wave wave seems to correspond closely to D's segmentals (specifically to pairs of stressed and unstressed syllables--"steada pidgin English it's

----

[2] Future investigations of such rapidly and precisely timed phenomena may profit from the inclusions of segmental parameters by taking into account the "p-center" effects by which segmentals vary from one to the other in the relations of their acoustic waveforms to their moments of psychological occurrance.

a") suggesting that the waveform does correspond to the organization of segmental productions in this instance. In Example 2, the correlograms of D's non-simultaneous speech (segment 4 in Figure 4.9) clearly indicate a 1120 msec period, but when this segment is considered in the context of his consequent speech, the 560 msec period is more salient to the autocorrelation analysis.

Analogies between autocorrelation and perceptual processes in the apprehension of F_0 cycles should remain tentative. There are some obvious differences (e.g. the requirement for three cycles to specify periodicity in an autocorrelation analysis), though the capacity of autocorrelation analysis to detect cycles masked by stochastic processes may prove to be a useful approximation of the perceptual normalizations found to occur in the apprehension of rhythmic structures.

The orientation of the present research program has been towards the development of an acoustic phonetics for discourse analysis, beginning with the elements of prosodic coherence. The only attempt to recover participants' perceptions has been in the study of presumed effects on their consequent productions. This is why simultaneous onsets are of interest: they indicate the possibility that participants have arrived at mutual perceptions of an appropriate time to begin speech relative to the timing of preceding speech.

5.2 Theoretical Implications

5.2.1 The Conversation as One Text

My interest in points of simultaneous onset has been motivated by a larger theoretical interest in identifying phenomena whose properties can best be described by referring to the conversation as one text, the production of which is regulated and coordinated by all participants. The coherence of the prosodic record may be thought of as one such property. This entails an acoustic description of text, in which there is minimal appeal to phonological systems predefined axiomatically to be shared by all participants.

I interpret those periodicities which link utterances by both speakers as properties of the dyadic conversation as one text. These properties are specified as the period and phase of statistical tendencies in the acoustic waveform of conversational F_0, and are therefore demonstrably available to both participants without appeal to the domains of explanation of cognitive systems in terms of rule-governed competencies. They are rather based on principles of speech performance--production and perception of acoustical patterns. The specific claim that statistical waveforms modeling the text of the conversation also model participants' internal representations (psychological, neural, formal, taxonomic or otherwise) is obviated by a careful adherence to the distinct separation between domains of explanation proper to the participants and those proper to the conversation itself.

The design orientation of viewing the conversation as one acoustic phonetic text is motivated by (and interpretable in terms of) several theoretical constructs. The concept of a basic analytic distinction between conversational and cognitive domains has played a central role in the development of Varela's neo-cybernetic epistemology arising from problems in the philosophy of biology (1976). The concept of the conversation as one text is more generally modeled in Varela's notion of an autonomous systems, in which properties of the system as a unit cannot be specified by the properties of the unitary components of the system (1979) [3].

Urion, who has applied some of Varela's heuristics to the consideration of actual conversational data (1980), criticizes the existing approaches of conversation analysis, ethnography of communication and formal discourse analyses for retaining the speech act as the minimal unit of description. In particular he notes that the speech act is a unit which is properly used only in the taxonomizing of rule-governed behaviors, depending therefore on the prior specification of contexts for the applications of those rules, on hypothetical competencies, inferred intentions, and other properties of participants' cognitive systems. Urion has demonstrated formal and practical necessities for maintaining a careful analytic distinction between the individual and interactional domains of description in the analysis of

----

[3]Living systems are defined according to these principles as special types of autonomous systems--*autopoietic systems* (Maturana and Varela 1975).

conversational data for components of usage (1985).

On the basis of his measurements of speech melody (F_0) and movements (of the head and eyes), Mair (1980) believes that these parameters act directly on shared cognitive models conceived of as alternating "plans and percepts." Points of instantaneous change in the the development of the models over time are indicated by "clearly demarcated boundaries and singular shapes [in the] time forms of the speech/movement trajectories." The postulated existence of a "supra-individual beat" is axiomatic to these claims, with the attendant neurophysiological implication that "brains can work synchronously" as if controlled by "one timing device" (1985:5).

5.2.2 Units of Analysis

Three further implications are implicit in [illegible] receiving only brief mention here as they have not been explored systematically.

1. Boundaries of change in period and phase may be useful in segmenting connected speech.
2. Many units so specified may have precisely identifiable temporal contexts, both internally and externally in relation with other such units through larger temporal patterns.
3. These units can function hierarchically with specifiable differences and similarities in their distributions at separate levels or temporal domains.

Properties 1 and 2 are not direct properties of other units for speech analysis such as the speech act. Note also that these units may in no way correspond to units of sound and silence if one accepts the notion that period and phase relations continue across utterances and across the intervening silences.

5.3 Methodological Implications

The primary methodological implications of this study are expressed in the summary of results. The findings demonstrate that time-series analytic tools such as autocorrelation can be used to identify periodicity in the prosodic record of the text, and that periodicity and phase are useful descriptors of conversational prosody.

5.3.1 Trend

The presence of trends in the F_0 data entails adjustments in the methods of analysis. I shall mention two possible responses to the presence of trend. For the purposes of future analysis of F_0 in the temporal domains of the present study (periods 160 msec to 3 seconds in length), larger trends will need to be removed. I anticipate that this may amount to further operationalizing notions of declination by applying tests for trend and shift in mean where called for by the presence of linear trends in the correlograms.

For the purposes of examining data for longer periodic components, this trend may be modeled as cyclical if it is found to be so. Longer periods in speech F_0 remain to be identified through autocorrelation analyses. The role of F_0 in identifying larger discourse unit boundaries (see e.g. Krieman 1982, Menn & Boyce 1982, and Schaffer 1983,1984) as well as the presence of long term cycles in speech activity/intactivity (Warner 1979, Kimberly 1970, Cobb 1973, and Hayes and Cobb 1979) suggest that cyclic autocorrelation may be found over lags taken to 10 or more seconds, from smoothed data.

5.3.2 Non-Stationarity

A time series, which may be realized by a variety of processes (including the Markovian processes postulated by Jaffe and Feldstein [1970]) can only be modeled for any given segment as one type or order of process to the extent that the series is stationary. Gottman (1981:62,70) explains that the criteria for stationarity require the autocovariance to be independent of historical time:

> A stationary process is characterized in part by the fact that the covariance between two random variables at $\underline{t}$ and $\underline{t} + \underline{k}$ is a function only of their relative lag, $\underline{k}$, not of the starting point $\underline{t}$. . . . This means that the correlograms . . . should have the same shape, independent of the starting point in historical time where we begin calculating

Only a cursory examination of the correlograms displayed as results is necessary to confirm that F_0 processes are not stationary in this sense.

General findings indicated a wide variety of trends, deterministic and non-deterministic cycles, and other transient influences on F_0-series, resulting in correlograms which change shape dramatically depending on historical time. Even though I imposed structural similarity on these particular autocorrelations by selecting for display only those with clear periodic components, they are nonetheless sufficiently various to indicate a high probability that the processes are non-stationary [4].

Confirmation that F_0 processes are not stationary would amount to a formal requirement for the specification of temporal context in descriptions of speech timing. The convergence of my results with others on the issue of stationarity is of an analogical nature: Other time-series theoretic investigations of speech have primarily been measures of vocal activity-inactivity. The issue of stationarity in speech timing might be resolved for fundamental voice frequencies using the methods developed here. It is difficult to assess in activity/inactivity measures, primarily because of the arbitrariness of measures.

5.3.3 Determinism

The suggestion that speech timing in some episodes of conversation can by characterized by a single correlogram (the squares of which do not go to zero with higher lags) implies that these episodes can be modeled as stationary deterministic processes. The results in Example 2, indicative of a cycle which persists across gaps several seconds long, and the effectiveness of splicing in general, suggest that the cyclicty of F_0 is not vulnerable to random shifts in the parameters and may be regarded as the realization of a fundamentally deterministic process. It remains to be strictly determined whether the F_0 cycles found in the present study are deterministic or stochastic in nature.

----

[4]This condition is necessary but not sufficient for non-stationarity--see Gottman (1981).

The present investigations of short term autocorrelation series do not determine whether autocorrelations descend to insignificant levels at an increasing lag (a condition indicative of stochasticity and also of stationarity). Through Fourier transforms, the time-series measures can be studied by the frequency-domain statistics expressed as power (related to probability) over the spectrum of F_0 fluctuations. Analyses of power spectra, and of relations between power spectra expressed as coherence and phase spectra, make possible detection of the long term patterning of F_0 data by expressing all possible time-domains of analysis simultaneously. It is probable that deterministic and stochastic components of F_0 fluctuation are analytically separable, and the stochastic components describable as autoregressive or moving average models. There will be little further resolution of these issues until the power spectra of conversational F_0 are examined.

5.4 Directions for Further Research

5.4.1 Description

The claims made in the summary of results were stated as possibilities; they are really little more than preliminary results of single subject case studies. These results represent a highly selective part of a much larger body of data which have been less closely analysed but which indicate the operation of similar principles between utterances by different speakers, across interjections, laughter, and so on. It seems impressionistically clear in heard speech and in F_0-plots that patterns of prosodic coherence between utterances by different speakers appear more often than by chance. One clear direction for further research would be to study a small body of conversational data for the number of segments with periodic F_0, for the periodicity of peaks or other identifiable F_0 features appearing in those segments, and for prosodically coherent simultaneous onsets, utterance junctures, speaker switches, and passages of simultaneous speech.

Some methodological refinements might be developed prior to such an analysis. Data on intensity would undoubtedly contribute to an understanding of the prosodic elements of coherence[5]. Both intra- and interspeaker normalizations of F_0 are justified, as is the removal of trend corresonding to declination. Treatments of trend may also be made clearer in the context of an investigation of the stationarity of F_0 as a process, coupled with an investigation of longer cycles in the data. A design orientation towards properties of the interaction is provided in bivariate time-series analytic procedures in which an understanding of autocorrelation in the univariate data is a necessary first step. Cross-correlations would certainly be appropriate for extended passages of simultaneous speech such as that which begins in Example 2. Analyses of power spectra, and of relations between power spectra expressed as coherence and phase spectra, will provide a variety of perspectives on F_0-series.

5.4.2 Physiological

F_0 fluctations are essentially physiological in nature. Pitch change is closely associated with respiratory mechanisms affecting sub glottal air pressure, though several issues remain unresolved (Lieberman 1967, Ohala 1978, Cooper and Sorenson 1981). Warner, Waggener, and Kronauer (1984) have investigated relations of activity/inactivity cycles to cycles in respiratory ventilation using spectral analytic procedures. Similar analyses may be possible with F_0 cycles.

There has been a history of suggestions that speech rhythms have neurophysiological bases (Lashley 1951, Lenneberg 1967) and many in particular have been developed as principles encompassing speech and movement (Mair 1978, Byers 1976, Erickson and Shultz 1982, Chapple 1981, Ulmer 1983, and Kelso & Tuller 1984). Rodda has indicated that rhythmic patterning is clearly observable in the movements of sign language in casual conversations between fluent signers (Rodda 1985:personal communication). Time-series analytic procedures may be applicable to movement parameters in analyses of the type performed here for F_0 alone,

----

[5] The methodology reported here also produces intensity-series in decibels, and these data are available for incorporation into the present results. Recording conditions prevent comparison of participants' vocal intensities on an absolute scale.

though there are sure to be some technical problems to be solved first (possibly through the use of polarized light [Hadar et al. 1983]). Multidimensional time series analytic statistics are conceivable, though possibly very difficult to interpret.

5.4.3 Cognitive and Linguistic

It is important to remember that F_0 functions in many linguistic and discoursal contexts. It is widely recognized that F_0 is a primary determinant of both stress and tone, and that it can carry semantic, syntactic, and emotive values. It may also fluctuate systematically with respect to the information content of the locution.

The periodicities in F_0 production reported here may have their analogues in perception and rhythmic processes of attention. Further experimental investigations of this notion would probably be necessary to establish an empirical basis for further claims along these lines. Periodic F_0 fluctuations may also be related to linguistic organization at many levels: segmental (Martin 1979, Fowler 1983), phonological (Ladd 1978), syntactic (Lehiste 1973, Cooper and Sorenson 1981) and cognitive (Mair 1980, Coulthard and Brazil 1981). There is a pressing need to investigate the phenomena across a variety of languages in an effort to develop a phonetic (i.e. universally descriptive) set of objective and parsimonious descriptive components of types of physical coherence in social situations.

5.4.4 Social Systems

A conversation is a social system: Conceived of as a microcosmic vehicle for the organization of social interaction, conversation is itself a miniature autonomous society, and it exhibits physical rhythmic patterning. Related perspectives on social systems may be found in Murphy (1971), Urion (1980,1985), Varela (1976,1979), and Warner (1982).

Any person participating in a conversation has the opportunity to create and control that social system by rendering his or her actions coherent with those of other persons. Scholars from several disciplines bearing on studies of conversation sometimes appear to elucidate a fundamental organizational principle of human interaction. That clarity may give promise of manipulative power, and this raises both moral and conceptual issues. It must be observed that even if "rhythmic organization" of interaction is discovered to be fundamental and manipulable, conversation is infinitely more complex than that view would imply. Regarding the moral and conceptual issues, it must be remembered that the autonomy of the individual is paramount to that of the conversation.

References

Allen, G. D.
1972a The location of rhythmic stress beats in English: An experimental study I. Language and Speech 15, 72-100.
1972b The location of rhythmic stress beats in English: An experimental study II. Language and Speech 15, 179-195.
1975 Speech rhythm: its relation to performance universals and articulatory timing. Journal of Phonetics 3, 75-86.

Aschoff, J., H. Fatranska and H. Giedke
1971 Human circadian rhythms in continuous darkness: Entrainment by social cues. Science 171, 213-215.

Auer, J. C. P.
1983 *Review of* Turn-taking in English Conversation, by Bengt Oreström. Linguistics 21, 742-798.

Beattie, G. W.
1981 Interruption in conversational interaction, and its relation to the sex and status of the interactants. Linguistics 19, 15-35.
1984 "Are there cognitive rhythms in speech?": A reply to Power (1983). Language and Speech 27, 193-195.

Bennet, A.
1981 Interruptions and the interpretation of conversation. Discourse Processes 4, 171-188.

Brazil, D.
1981 The place of intonation in a discourse model. *In* M. Coulthard & M. Montgomery (Eds.), Studies in discourse analysis. London: Routledge & Kegan Paul.

Buxton, H.
1983 Temporal predictability in the perception of English speech. *In* A. Cutler & D. R. Ladd (Eds.), Prosody: models and measurements. Berlin: Springer-Verlag.

Byers, P.
1976 Biological rhythms as information channels in interpersonal behavior. *In* P. P. G. Bateson & P. H. Kopfler (Eds.), Perspectives in ethology, Vol. 2. New York: Plenum Press.

Capella, J. N.
1981 Mutual influence in expressive behavior: Adult-adult and infant-adult dyadic interaction. Psychological Bulletin 89, 101-132.

Capella, J. N., and J. O. Greene
1982 A discrepancy-arousal explanation of mutual influence in expressive behavior for adult and infant-adult interaction. Communication Monographs 49, June, 89-114.

Chapple, E. D.
1971 Toward a mathematical model of interaction: Some preliminary considerations. *In* P. Kay (Ed.), Explorations in mathematical anthropology.' Cambridge, Ma.: M.I.T. Press.
1980 The unbounded reaches of anthropology as a research science and some working hypotheses. American Anthropologist 82, 741-758.
1981 Movement and sound: The musical language of body rhythms in interaction. Teachers College Record 82, 635-648.

Cobb, L.
1973 Time series analysis of the periodicities of casual conversations. Unpublished doctoral dissertation, Cornell University. Dissertation Abstracts International 33, 5105B, University Microfilms No. 73-28, 286.

Connell, B. A., J. T. Hogan and A. J. Rozsypal
1983 Experimental evidence of interaction between tone and intonation in Mandarin Chinese. Journal of Phonetics 11, 337-351.

Cooper, W. E. and J. M. Sorenson
1981 Fundamental frequency in sentence production. New York: Springer-Verlag.

Coulthard, M., and D. Brazil
1981 The place of intonation in the description of interaction. *In* D. Tannen (Ed.), Analyzing discourse: Text and talk. University Round Table on Languages and Linguistics 1981. Washington: Georgetown University Press.

Coulthard, M.
1977 An introduction to discourse analysis. London: Longman.

Dauer, R. M.
1983 Stress-timing and syllable-timing reanalyzed. Journal of Phonetics 11, 51-62.

Eady, S. J.
1982 Differences in the Fo patterns of speech: Tone language versus stress language. Language and Speech 25, 29-42.

Erickson, F. and J. Shultz
1982 The counselor as gatekeeper: Social interaction in interviews. New York: Academic Press.

Faure, G., D. J. Hirst, and M. Chafcouloff
1980 Rhythm in English: Isochronism, pitch, and perceived stress. *In* L. R. Waugh & C. H. van Schooneveld (Eds.), The melody of language. Baltimore: University Park Press, 71-80.

Fowler, C. H.
1983 Converging sources of evidence on spoken and perceived rhythms of speech: Cyclic production of vowels in monosyllabic stress feet. Journal of Experimental Psychology: General 112, 386-412.

French, P., and J. Local
1983 Turn-competitive incomings. Journal of Pragmatics, 7 17-38.

Fry, D. B.
1968 Prosodic phenomena. *In* B. Malmberg (Ed.), Manual of phonetics. Amsterdam: North-Holland.

Gandour, J. T.
1978 The perception of tone. *In* V. A. Fromkin (Ed.) Tone: A linguistic survey. New York: Academic Press, 41-76.

Glass, G. V., V. L. Willson, and J. M. Gottman (Eds.)
1975 Design and analysis of time-series experiments. Boulder, Co.: Colorado Associated University Press.

Goldberg, J. A.
1978 Amplitude shift: A mechanism for affiliation of utterances in conversational interaction. *In* J. Shenkein (Ed.), Studies in the organization of conversational interaction. New York: Academic Press.

Gottman, J. M.
1979 Detecting cyclicity in social interaction. Psychological Bulletin 86, 335-348.
1981 Time-series analysis: A comprehensive introduction for social scientists. Cambridge: Cambridge University Press.

Gottman, J. M., and J. T. Ringland
1981 The analysis of dominance and bidirectionality in social development. Child Development 52, 393-412.

Gregory, S. W., and B. R. Hoyt
1982 Conversation partner mutual adaptation as demonstrated by fourier series analysis. Journal of Psycholinguistic Research 11, 35-46.

Gregory, S. W.
1983 A quantitative analysis of temporal symmetry in microsocial relations. American Sociological Review 48 129-135.

Gumperz, J. J., H. Kaltman, and M. C. O'Connor
1984 Cohesion in spoken and written discourse: Ethnic style and the transition to literacy. *In* D. Tannen (Ed.), Coherence in spoken and written discourse (Volume III of R. D. Freedle [Ed.], Advances in discourse processes). Norwood, N. J.: Ablex.

Hadar, U., T. J. Steiner, E. C. Grant, and F. C. Rose
1983 Head movement correlates of juncture and stress at sentence level. Language and Speech 26:2, 117-129.

Hayes, D. P., and L. Meltzer
1967 Bone-conducting microphones. American Journal of Psychiatry 80, 619-624.

Hayes, D. P., L. Meltzer and G. Wolf
1970 Substantive conclusions are dependent upon techniques of measurement. Behavioral Science 15, 265-268.

Hayes, D. P., and L. Cobb
1979 Ultradian biorhythms in social interaction. *In* A. W. Siegman & S. Feldstein (Eds.), Of speech and time. Hillsdale, N. J.: Lawrence Erlbaum Associates.

Hoequist, C. E.
1983 The perceptual center and rhythm categories. Language and Speech 26, 367-376.

Hudson, R. A.
1980 Sociolinguistics. Cambridge: Cambridge University Press.

Iversen, K. E.
1962 A programming language. N. Y.: Wiley

Jaffe, J. J., S. W. Anderson and D. Stern
1979 Conversational rhythms. *In* D. Aaronson & R. W. Rieber (Eds.), Psycholinguistic research. Hillsdale: Lawrence Erlbaum Associates.

Jaffe, J., S. Breskin, and L. J. Gerstman.
1972 Random generation of apparent speech rhythms. Language and Speech 15, 68-71.

Jaffe, J. J. and S. Feldstein
1970 Rhythms of dialogue. New York: Academic Press.

Jassem, W.
1975 Normalization of Fo curves. *In* G. Fant & M. A. A. Tatham (Eds.), Auditory analysis and perception of speech. London: Academic Press, 523-530.

Jassem, W.
1978 On the distributional analysis of pitch phenomena. Language and Speech 21, 362-372.

Jassem, W., and K. Kudela-Dobrogowski.
1980 Speaker-independent intonation curves. *In* L. R. Waugh and C. H. van Schooneveld (Eds.), The melody of language. Baltimore: University Park Press.

Jefferson, G.
1973 A case of precision timing in ordinary conversation: Overlapped tag-positioned address terms in closing sequences. Semiotica 9, 47-96.

Jones, M. R.
1976 Time, our lost dimension: Toward a new theory of perception, attention, and memory. Psychological Review 83, 323-355.

Kelso, J. A. S., and B. Tuller
1984 Converging evidence in support of common dynamical principles for speech and movement coordination. American Journal of Physiology 246, R928-R935.

Kimberly, R. P.
1970 Rhythmic patterns in human interaction. Nature 228, 88-90.

Konefal, J. A., and J. Fokes
1985 Acoustic analysis of children's conversational repairs. Journal of Phonetics 13, 97-105.

Kreiman, J.
1982 Perception of sentence and paragraph boundaries in natural conversation. Journal of Phonetics 10, 163-175.

Ladd, D. R. Jr.
1978 The Structure of intonational meaning. Bloomington: Indiana University Press.
1983 Phonological features of intonation peaks. Language 59, 721-759.

Lashley, K. S.
1951 The problem of serial order in behavior. *In* L. A. Jeffress (Ed.), Cerebral mechanisms in behavior. New York: Wiley.

Lehiste, I.
1973 Rhythmic units and syntactic units in production and perception. Journal of the Acoustical Society of America 54, 1228-1234.
1977 Isochrony reconsidered. Journal of Phonetics 5, 253-263.

Lenneberg, E. H.
1967 Biological foundations of language. New York: Wiley.

Leon, P. R., and P. Martin
1972 Machines and measurements. *In* D. Bolinger (Ed.), Intonation. Harmondsworth: Penguin, 30-47.

Lewis, M., W. MacLean, W. Bryson-Brockmann, R. Arendt, B. Beck, P. S. Fidler, and A. A. Baumeister
1984 Time-Series analysis of stereotyped movements: Relationship of body-rocking to cardiac activity. American Journal of Mental Deficiency 89, 287-294.

Lieberman, P.
1967 Intonation, perception, and language. Cambridge, Mass.: M.I.T. Press.

McReynolds L. V., and K. P. Kearns
1983 Single-subject experimental designs in communicative disorders. Baltimore: University Park Press.

Maguire, T. O.
1985 On the changing role of numerical methods in educational reseach. Unpublished Manuscript, McCalla Lecture, University of Alberta.

Mair, M. W.
1977 Steps towards principles of text regulation. Toronto: Toronto Semiotic Circle, Victoria College, University of Toronto.
1980 A model of the text generator. Paper presented at the International Summer Institute for Structural Semiotic Studies, Toronto, June.

Martin, J. G.
1972 Rhythmic (hierarchical) versus serial structure in speech and other behavior. Psychological Review 79, 487-509.
1979 Rhythmic and segmental perception are not independent. Journal of the Acoustical Society of America 65, 1286-1297.

Maturana, R. H., and F. Varela
1975 Autopoietic systems: A characterization of the living organization. Laboratory Report 9.4. Urbana, Illinois: Biological Computer Laboratory, University of Illinois.

Meltzer, L., W. Morris, and D. P. Hayes
1971 Interruption outcomes and vocal amplitude: Explorations in social psychophysics. Journal of Personality and Social Psychology 18, 392-402.

Menn, L., and S. Boyce
1982 Fundamental frequency and discourse structure. Language and Speech 25, 341-378.

Morris, W. N.
1971 Manipulated amplitude and interruption outcomes. Journal of Personality and Social Psychology 20, 319-331.

Morton, J., S. Marcus, and C. Frankish
1976 Perceptual centers (P-centers). Psychological Review 83, 405-408.

Murphy, R., F.1971 The dialectics of social life: Alarms and excursions in anthropological theory. New York: Basic Books.

Natale, M., Entin, E., and Jaffe, J.
1979 Vocal interruptions in dyadic communicationn as a function and social anxiety. Journal of Personality and Social Psychology 37, 865-878.

Ohala, J. J.
1975 The temporal regulation of speech. In G. Fant & M. A. A. Tatham (Eds.). Auditory analysis and perception of speech. New York: Academic Press.
1978 Production of tone. *In* V. A. Fromkin (Ed.), Tone: A linguistic survey. New York: Academic Press, 5-40.
1983 Cross-language use of pitch: An ethological view. Phonetica 40, 1-18.
1984 An ethological perspective on common cross-language utilization of F_0 of voice. Phonetica 41, 1-16.

Piaget, J.
1976 The psychology of intelligence. Totowa, N. J.: Littlefield, Adams & Co., (originally published in France, 1947).

Pierrehumbert, J.
1979 The perception of fundamental frequency declination. Journal of the Acoustical Society of America 66, 363-369.

Pitch Instruments, Inc.
n. d. Pitchmeter manual. Toronto.

Porges, S. W., R. E. Bohrer *et al.*
1980 New time-series statistic for detecting rhythmic co-occurance in the frequency domain: The weighted coherence and its application to psychophysiological research. Psychological Bulletin 88, 580-587.

Power, M. J.
1983 Are there cognitive rhythms in speech. Language and Speech 26, 253-261.
1984 "Are there cognitive rhythms in speech?" - A reply to Beattie (1984). Language and Speech 27, 197-200.

Sackett, G. P.
1977 The lag sequential analysis of contingency and cyclicity in behavioral interaction resarch. *In* J. Osofsky (Ed.), Handbook of infant development. New York: Wiley.

Sacks, H., E. A. Schegloff, and G. Jefferson
1974 A simplest systematics for the organisation of turn-taking for conversation. Language 50, 696-735.

Sinclair, J. M., and D. Brazil
1982 Teacher talk. London: Oxford University Press.

Schaffer, D.
1983 The role of intonation as a cue to turn taking in conversation. Journal of Phonetics 11, 243-257.
1984 The role on intonation as a cue to topic management in conversation. Journal of Phonetics 12, 327-344.

Scherer, K. R.
1982 Methods of research on vocal communication: Paradigms and parameters. *In* P. Ekman and K. Sherer (Eds.), Handbook of methods in nonverbal behavior research. Cambridge: Cambridge University Press.

Scollon, R.
1981 Rhythmic integration of ordinary talk. *In* D. Tannen (Ed.), Analyzing discourse: Text and talk. University Round Table on Languages and Linguistics 1981. Washington: Georgetown University Press.

Stone, M.
1981 Evidence for a rhythmic pattern in speech production: Observations of jaw movement. Journal of Phonetics 9, 109-120.

Streeter, L. A., R. M. Krauss, V. Geller, C. Olson, and W. Apple
1977 Pitch changes during attempted deception. Journal of Personality and Social Psychology 11, 243-257.

Tronick, E. D., H. Als, and T. B. Brazelton
1977 Mutuality in mother-infant interaction. Journal of Communication, Spring, 74-79.

Ulmer, F.
1983 Melody and synchrony in the text of conversation. Unpublished master's thesis, Educational Foundations, University of Alberta.

Urion, C.,
1978 Control of topic in a bilingual (Cree-English) speech event. Unpublished doctoral dissertation, Anthropology, University of Alberta.
1980 The ethnography of speaking and the description of conversation. *In* W. McCormack and H. Izzo (Eds.), Sixth LACUS Forum. South Carolina: Hornbeam Press.
1985 The analysis of conversation. Papers read at the Sixth International Institute for Semiotic and Structural Studies, Central Institute of Indian Languages, Mysore, India.

Varela, F. J.
1976 Not one, not two. Coevolution Quarterly, Fall, 62-67.
1979 Principles of biological autonomy. New York: Elsevier North Holland.

Vaissière, J.
1983 Language-independent prosodic features. *In* A. Cutler and D. R. Ladd (Eds.), Prosody: Models and Measurements. Berlin: Springer-Verlag, 53-66.

Voss, R. J., and J. Clark
1975 "1/f noise" in music and speech. Nature 285 (Nov. 27), 317-318.

Warner, R. M.
1979 Periodic rhythms in conversational speech. Language and Speech 22, 381-396.
1982 The psychologist as social systems consultant. *In* T. Millon, C. Green and R. Meagher (Eds.), Handbook in clinical health psychology. New York: Plenum Press.

Warner, R. M., T. Waggener, and R. E. Kronauer
1984 Synchronized cycles in ventilation and vocal activity during spontaneous conversational speech. Journal of Applied Physiology 54, 1324-1334.

West, C., and D. H. Zimmerman
1977 Women's place in everyday talk: Reflections on parent-child interaction. Social Problems 24:5, 521-529.

Wiener, N.
1961 Brain waves and self-organising systems. (Chapter 5 of N. Wiener Cybernetics: Or control and communication in the animal and the machine). Cambridge, Ma.: M.I.T. Press.

Yngve, V. H.
1970 On getting a word in edgewise. *In* Papers from the Sixth Regional Meeting, Chicago Linguistic Society. Chicago: Chicago Linguistic Society, 567-578.

Appendix 1. Release Form

I have volunteered to participate in the recording of conversations held with other volunteers. I am aware that these recordings are not surreptitious, that my identity will be protected, and that I have the right to have the purposes of the research explained to me in layperson's terms.

Subject:

My first language is:

Researcher:

Date: