American Journal of Computational Linguistics Microfiche 60

PITCH CONTOUR GENERATION IN SPEECH SYNTHESIS

A JUNCTION GRAMMAR APPROACH

ALAN K. MELBY, WILLIAM J. STRONG,
ELDON G. LYTLE, AND RONALD MILLETT

Translation Sciences Institute
130 B-34
Brigham Young University
Provo, Utah 84602

SUMMARY

Computer based text synthesis systems require a means for generating sentence-level pitch contours. These contours must have a certain degree of "human fidelity" if the synthetic speech is to sound natural and not too machine-like. The pitch contours in currently operational text synthesis systems are still not perfectly natural-sounding and thus computer generation of pitch contours is a topic of current interest. The introduction includes a survey of current work in this area by researchers at MIT, Bell Labs, Stanford, etc., describing their general approaches.

The research described in this paper uses Junction Grammar as a theoretical base, and Linear Predictor Coefficient (LPC) methods as an analysis-synthesis technique. Motivations for these decisions are presente

Section I begins with an explanation of some sentences which are being studied. For example, there is likely a stress on "study" in the sentence "The boys who study get good grades," if the context is "but the boys who don't get bad grades." On the other hand, if the context is "but the girls who study get poor grades," then there is probably stress on "boys." The various readings of "the boys who study..." and other sentences are explained within the Junction Grammar framework. An overview is given of a system for generating pitch contours for a sentence from a Junction Grammar semantico-syntactic representation.

Section I also includes a description of an extension of Junction Grammar which defines an object called an articulation tree, corresponding to each junction tree. A junction tree contains semantico-syntactic information but no lexical information. An articulation tree

contains segmental information about each lexical item and suprasegmental or prosodic information combining the lexical items into prosodic units. Semantic distinctions in junction trees are recoded as distinctions in the prosodic structure of articulation trees and then articulation trees are used to generate pitch contours. Junction trees and articulation trees are included as figures for several sentences.

Section II describes how pitch contours are generated, including the recoding of junction trees as articulation trees, the assignment of initial and final pitch levels and pitch at nuclear syllables, and how the generated contours are combined with analysis parameters and synthesized into speech. It should be noted that the junction trees are entered manually rather than by automatic analysis, in the current implementation.

The text includes several graphs of natural pitch contours as well as contours generated by the computer system.

The pitch contour system produces a synthesis output for each reading of a sentence. Thirty-five sentences, some with natural, some with hand-drawn, and some with machine-generated pitch contours were evaluated for naturalness and "intelligibility" of intonation in four types of tests. Results of testing several subjects showed that the generated pitch contours were judged nearly as natural as human-produced contours, and except for some specific problems involving duration, the generated contours were intelligible in the sense of causing the listener to perceive the intended reading of the sentence. The text includes a quantitative summary of the results of the evaluation.

For the corpus of sentences treated so far, Junction Grammar provides a satisfactory theoretical base for generating pitch contours and defines some specific cases where pitch alone is insufficient to make distinctions and must be used with duration, pause and intensity.

Appendices:

A. Suggested background reading in acoustic speech processing and Junction Grammar.

B. Glossary of terms, e.g. LPC, FO, Hertz, etc.

C. Description of the computer implementation (on a PDP-15 with a VT-15 grapnics display unit).

D. More details on the evaluation procedure.

For the convenience of the reader, a recent paper on Junction Theory presented at a BYU Linguistics Symposium is reprinted at the end of the microfiche.

TABLE OF CONTENTS

INTRODUCTION AND SURVEY OF RESEARCH IN PITCH CONTOURS 6

Section

I. THEORY 13

II. METHOD 31

III. EVALUATION AND DISCUSSION 41

ACKNOWLEDGMENTS 47

REFERENCES 49

APPENDICES

A. BACKGROUND READING 52

B. GLOSSARY 53

C. COMPUTER IMPLEMENTATION 54

D. MORE DETAILS ON THE EVALUATION 58

REPRINT OF ONE OF THE REFERENCES (Lytle, 1976) 69

INTRODUCTION

All computer based text synthesis systems require a means for generating sentence-level pitch contours. These contours must have a certain degree of "human fidelity" if the synthetic speech is to sound natural, that is, not too machine-like. The pitch contours in currently operational text synthesis systems are still not perfectly natural-sounding and thus computer generation of pitch contours is a topic of current interest. This interest is shown, for example, by Allen as he discusses pause and duration in text synthesis and then goes on to say:

> If temporal control presents great problems in the description of speech, then the problems of fundamental frequency (f0), or pitch control, are at least as difficult. Once again, problems arise due to the fact that the f0 is correlated with many factors, including vowel tongue height, previous consonant, breath group contour, syntactic and semantic content of words, whether a sentence is a question, intonation effects, and word boundary glottalization.
>
> (Allen, 1976: 440)

Given the need for further research in pitch control, a question remains of how to approach the problem. The authors feel it is important to work within a linguistic model that interrelates semantic and phonetic phenomena. Later on in Allen's article he makes the following statement (which coincides with our philosophy):

> The current use of sophisticated means for pitch recording, coupled with increased interaction between linguistics and speech researchers, should, however, lead to significantly improved pitch control programs which are based on sound linguistically motivated theory.
>
> (Allen, 1976: 441)

The need for interation between linguistics and speech research is further explained by Umeda (1976: 450):

> The message realization forms one structure as a whole. Its constituents-acoustic realization, higher level

> prosody, and syntax-semantics-interact with each other very closely; a decision made at any level derives immediately from the obtained result at the level above, and affects a decision at the level below.

The remainder of this section consists of a survey of some of the current work in this area in the USA (at MIT, Bell Labs, and Stanford University), in Germany, and in the USSR. Then the section will conclude with an introduction to the present research.

A. MIT

At MIT, Allen (1976) is working on pitch control as an element in his overall plan to produce a system capable of producing synthetic speech from unrestricted English text. He points out that although a syntactic and semantic analysis is needed, no existing automatic algorithm can provide that analysis reliably for entire sentences of unrestricted text. So he has elected to do a local analysis of the sentence first and then tie together the local analyses into a sentence level analysis if possible. The analyzer is thus designed so that if at some point complete sentence analysis is blocked, the partial analyses are still useful in generating the pitch contour and other prosodic controls such as duration and pause. In response to the need for a theoretical framework for relating a text and its pitch contour, Allen is using the ideas of Halliday (1970) (e.g. discourse focus) to investigate such questions as when and why elements of a verb string are stressed. For example, he notes that the sentence "A farmer was eating the carrot" will receive emphasis on "eating" if it is in response to a question about what the farmer is doing. Allen correctly notes that:

> The discovery and coordination of all these effects is a large and continuing effort, and it is clear that

> substantial semantic and discourse-level knowledge is needed to correctly predict prosodic parameters."
> (Allen, 1976: 441)

B. Bell Labs

Several workers at Bell Labs have attacked the problem of controlling pitch in speech synthesis. Olive (1975) describes a system for generating pitch contours for the sentence type "article-subject-verb-article-object" with an optional adjective on the subject or object. His method for generating the pitch contour was to record several sentences of the specified type using random words and to average the natural pitch contours to obtain prototype contours. Then the contour for each word was approximated by a fourth order polynomial to "facilitate linear stretching and compression of the fundamental frequency contour." Olive reports that by using this pitch contour generation system, in conjunction with a word concatenation scheme in which the words are stored in linear predictor coefficient (LPC) code, the synthesized sentences were of high quality

Umeda, at Bell Labs, is also concerned with pitch contours, asserting that "Among acoustic components, pitch (the fundamental frequency of the voice) shows the most direct relation to higher level prosody, stress and boundaries" (Umeda, 1976: 448). Umeda's algorithm for controlling prosodic parameters is based on a syntactic analysis of the input text. The analyzer fits each clause into a template consisting of the following optional slots: sentence modifier, subject, verb, object or complement, tail modifier, and punctuation mark. A point where the above order of template elements is violated is marked as a boundary, and boundaries are later used to assign pauses and intonation (Umeda, 1975).

C. Stanford University

At Stanford University, there is a research project on generative prosodics ir the Institute for Mathematical Studies in the Social Sciences (IMSSS). Researchers on this project are developing a system which, ultimately, is intended to do synthesis in real time for use in computer-assisted instruction at IMSSS (Levine, 1976). Their technique is to compile a lexicon of words in LPC code (Atal and Hanauer, 1971) and then, when a given sentence is to be synthesized, concatenate the code for each word, adjusting durations and pitch contours as needed. While Olive throws away the original pitch contour of each word, the IMSSS approach is to adjust the original contour of the word and then further smooth the contour so that each word will not sound sentence final.

The IMSSS group uses the ideas of Leben (1976), who relates English prosody to tone languages in that he views both tone languages and English as having a suprasegmental melody which is combined with the segmental phonological elements. The IMSSS group (Levine, 1976: 3) defines melody as a sequence of "auto segmental tones (autonomous from the phonological segments) selected from the tonal repertoire of the language." These tones are treated theoretically as discrete fundamental frequency levels, but then they are realized phonetically as continuous contours. In order to assign tones to key syllables, a program analyzes the sentence to be synthesized using a simple phrase structure grammar which brackets phrases, clauses and other complex constituents, and indicates boundaries between major constituents.

D. Germany

Complementary to pitch contour generation, is the study of the perception of pitch contours.

In Germany Isacenko and Schadlich (1970), performed an interesting series of experiments on the perception of German intonation. Natural sentences illustrating different intonation patterns were recorded and monotonised at various fundamental frequencies (e.g. 150 Hertz and 178.6 Hertz). Then the tapes of the monotone versions were cut and spliced at various points. The spliced tapes thus had an artificially simplified intonation of exactly two tone levels. The team found that they could change the way listeners perceived certain ambiguous sentences by changing only the points at which tone switches occurred.

E. USSR

In the USSR, Haavel et al. (1976) have also performed some experiments in manipulating pitch contours while leaving other parameters constant. They are interested in finding ways to "decrease the amount of information necessary for the description of pitch curves without distorting the parameters interpreted by man as prosodic characteristics of a sentence." They base this search on the assumption that man has only a limited short term memory available for storing the pitch contour and so makes decisions concerning the prosody of a sentence by extracting prosodic features which contain considerably less information than that needed to reconstruct exactly the same pitch contour. They conclude from these experiments that decisions such as declarative versus interrogative are based on the position of the rise or fall in pitch and not on the difference in pitch from high to low. They also conclude that in determining emphasis, the position of the peak value of the second derivative of the pitch contour is very significant.

F. Brigham Young University (BYU)

The research in pitch contour generation to be described in this paper addresses basically the same questions as the various projects surveyed above:

(1) What theoretical base might one use to represent syntactic and semantic information?

(2) How does one convert linguistic information, both at sentence-level and discourse-level, to the algorithmic control of prosodic parameters?

(3) What aspects of the pitch contour (e.g. 1st and 2nd derivatives, transitions relative to key syllables, and actual frequency) are significant in causing intonation and emphasis options to be perceived?

(4) What synthesis technique should be used to incorporate the prosodic controls into a working system (e.g. LPC synthesis, formant synthesis, or articulatory synthesis)?

We have chosen to use Junction Grammar (JG) as a theoretical framework within which to look for answers to questions (1) and (2) above. Junction Grammar refers to a linguistic model formulated by Lytle (1974). Subsequently, Junction Theory has been used to formulate a new theory of phonology in which a semantico-syntactic representation (called a junction-tree) is recoded as a general articulatory representation (called an articulation-tree) (Lytle, 1976). Junction Grammar extended to include Junction Phonology was selected for use in the BYU project because it seems to provide some significant insights and a flexible framework for our research.

It should be pointed out that at present there is no completely automatic algorithm for obtaining a detailed and powerful representation of syntax-semantics from general English text. For this reason, other researchers (e.g., Allen at MIT, Umeda at Bell Labs, and Levine at Stanford) have chosen to use a simple representation which can be obtained automatically. The authors' research, however, takes advantage of a larger project (Lytle, 1975) which uses man-machine interaction to obtain a more powerful representation than can be obtained automatically. Therefore, it was decided to use the full power of Junction Grammar representations in hopes of a future automatic analyzer rather than use some restricted version of Junction Grammar and be forced to add to it piece by piece to account for more and more phenomena.

To gain insight into topic (3) above (concerning which aspects of the pitch contour are significant to perception), we experimented with manually specified pitch contours.

In answer to question (4) above (concerning the choice of an analysis synthesis technique), we have chosen to work initially with an LPC synthesis technique (as did Olive at Bell Labs and Levine at Stanford) because an LPC software package was already available at BYU. But long range plans include the use of an articulatory functional model (Flanagan, 1975).

I. THEORY

We now turn our attention to certain linguistic phenomena which we consider especially interesting. First, we will illustrate the phenomena with sample sentences which will be discussed in intuitive terms and then in terms of Junction Grammar junction-trees (J-trees) and articulation-trees (A-trees). The section will conclude with a block diagram of what a fully developed Junction Grammar text synthesis system would look like and a block diagram of the system as currently implemented.

A. Intuitive Presentation of Some Test Sentences

Consider the sentence "John drove to the store." This sentence can be read several different ways depending on the discourse context. Figure 1 shows five possible readings and their context. Whatever system is used to represent the linguistics of this sentence, it should be possible to represent each of these four readings uniquely.

	Sentence	Possible context
1a	John drove to the store.	What happened?
1b	John drove to the store.	Who drove to the store?
1c	John drove to the store.	How did John get to the store.
1d	John drove to the store.	Where did John drive?
1e	John drove to the store? (Are you sure that's what you meant to say?)	John drove to the store, you know.

Figure 1. John drove to the store.

Now consider the question "Did John or Mary come?" Suppose that you heard someone come in but you did not see who it was. Nevertheless, you are sure that it was either John or Mary. In this context, you would put stress on "John" and on "Mary" and a falling pitch at the end of the sentence. Then you would expect a reply of "John" or "Mary." (If you receive as a reply simply "yes" then the person responding either did not understand or is trying to be funny.) On the other hand, suppose a whole crowd came to a party and you have a message which you must deliver to either John or Mary. In this context, you may or may not stress "John" and "Mary" but you would certainly end the sentence with a rising pitch. Then you would expect a yes/no reply, or perhaps a yes/no with additional volunteered information such as "Yes, John is over there in the corner." Again, we would like our system of representation to handle this distinction. The two readings of "Did John or Mary come?" are summarized in Figure 2.

Sentence	Possible Response
2a Did John or Mary come?	John came.
(falling pitch at end)	
2b Did John or Mary come?	Yes, they are both here.
(rising pitch at end)	

Figure 2. Did John or Mary come?

Finally, consider the sentence "The boys who study get good grades." What difference in meaning is there in stressing "study" as opposed to stressing "boys"? The difference can be illustrated by expanding the sentence to "The boys who study get good grades but the others do not." If "study" is stressed, "others" is interpreted as "boys", namely the boys who do not study. If, however, "boys" is stressed, "others" may no longer be interpreted as "boys," but it can be interpreted as "girls" or "men who study" or some other group of students in contrast with boys. Once again, our system of representation needs to handle this distinction, and handle it in a way consistent with the treatment of other distinctions. Three readings of this sentence are summarized in Figure 3.

	Sentence	Possible continuation
3a	The boys who study get good grades... (neutral)	as is usually the case.
3b	The boys who study get good grades...	but the boys who spend all their time playing basketball get poor grades.
3c	The boys who study get good grades...	but for some reason the girls (even the girls who study) get poor grades.

Figure 3. The boys who study get good grades

B. Junction Grammar Representations of the Same Sentences

We now discuss how Junction Grammar represents the above distinctions in its representations. If the reader is not as yet familiar with Junction Grammar, it might be advisable to consult Appendix A before reading this section. As indicated therein, some recent refinements of Junction Grammar are not yet available in published form. We therefore briefly discuss two of them here. One is the specializations of subjunction in J-trees, and the other is the explicit representation of modalizers.

Direction of Subjunction First consider the three major specializations of subjunction shown in Figure 4.

Specializations of DIRECTION:

symbol	mnemonic	function
*·	right	entry of information
·*	left	recovery of information
·*·	double	non-restrictive association

Indication of REMAINDER:

	hyphen	induces a remainder
	equals	induces no remainder

Figure 4. Specializations of Subjunction in J-trees

A right subjunction (*·) often signifies that information is to be entered into the hearer's memory net. For example, when we read the sentence "I saw a lost child with a scraped knee this morning, and I helped him find his mother," we enter (according to Junction theory) into our memory a slot for a child who was lost. The junction between "a" and "child" would be N ("a") *· N ("child"). If we next read the sentence, "The child had been crying for two hours, the poor thing," we would recover the slot for the child and add to it the information that he had been crying. The junction between "the" and "child" in this case would be N ("the") ·* N ("child"). The third type of subjunction (·*·) would be used, for example, in the sentence "John, our mailman, is going to retire in March," to show that "John," and "our mailman" are defining the same person independently (cf. the traditional restrictive non-restrictive distinction).

In the above examples, we considered full subjunctions, (e.g. "John, our mailman") but the same specializations apply to interjunctions, (e.g. "John, who is our mailman"). In a normal, restrictive modification, a left subjunction is used. For example in, "Please give me the yellow book on the second shelf," "yellow" and "book" would be joined as follows (Fig. 5).

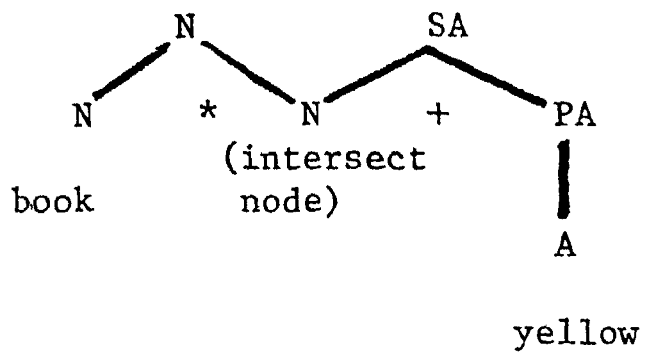


Figure 5. J-tree for"yellow book"

For an explanation of the various nodes in this representation for a simple phrase see Lytle (1975).

In the sentence "Of Tom, John and Rudolph, John drove to the store," the prepositional phrase "Of Tom, John and Rudolph" does not restrict the meaning of "John" in the way "yellow" restricted "book" in the previous example. Actually in this case, "John" restricts the scope of the prepositional phrase. As a reflection of this, the prepositional phrase is interjoined with "John" using a right subjunction as illustrated in Figure 6.

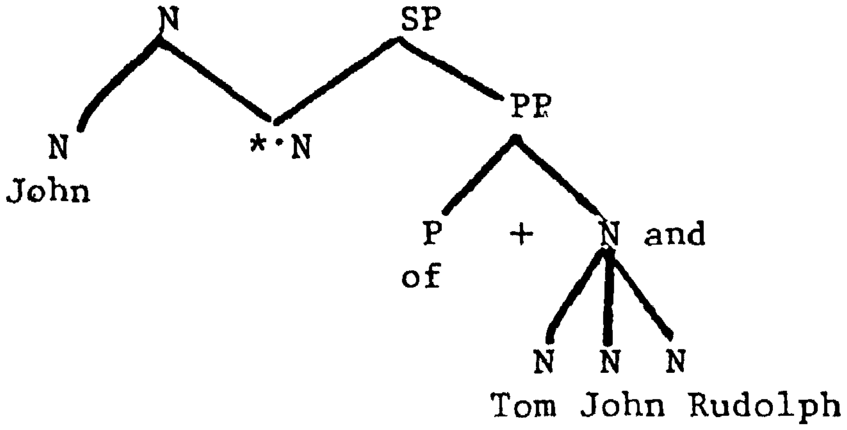


Figure 6. Right interjunction

We call this an example of Frame II modification because the right subjunction is relating "John" to a second frame of reference (i.e. Tom, John and Rudolph). On the other hand, "yellow book" is a frame I modification because it restricts "book" within its own frame of reference (i.e. it determines which book we are talking about).

Remainder. The second type of specialization mentioned in Figure 4 is an indication of remainder. The concept of remainder (Lytle, 1974) is concerned with whether all or only part of a set is referred to. If one desires to indicate whether there is a remainder in a subjunction, he simply replaces the dot with either a hyphen or an equals sign.

The Hyphen option. For example, from the sentence "Please give me the yellow book on the second shelf," we must assume that there are books of some color other than yellow on the second shelf. These other colored books are the remainder and we could diagram "yellow book" more specifically than before as follows (Figure 7).

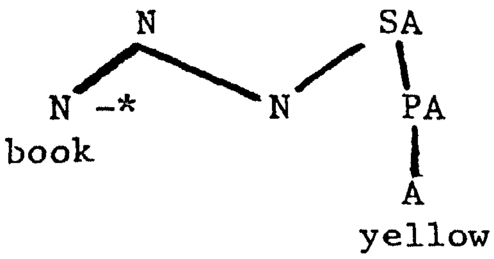


Figure 7. Left Hyphen

The Equals option. One common case of the equals option is for explicit modalizers (e.g. articles). For example, the phrase "The child" could be diagrammed as follows (Figure 8).

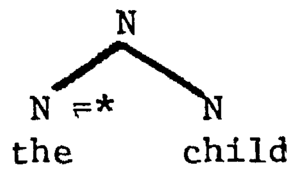
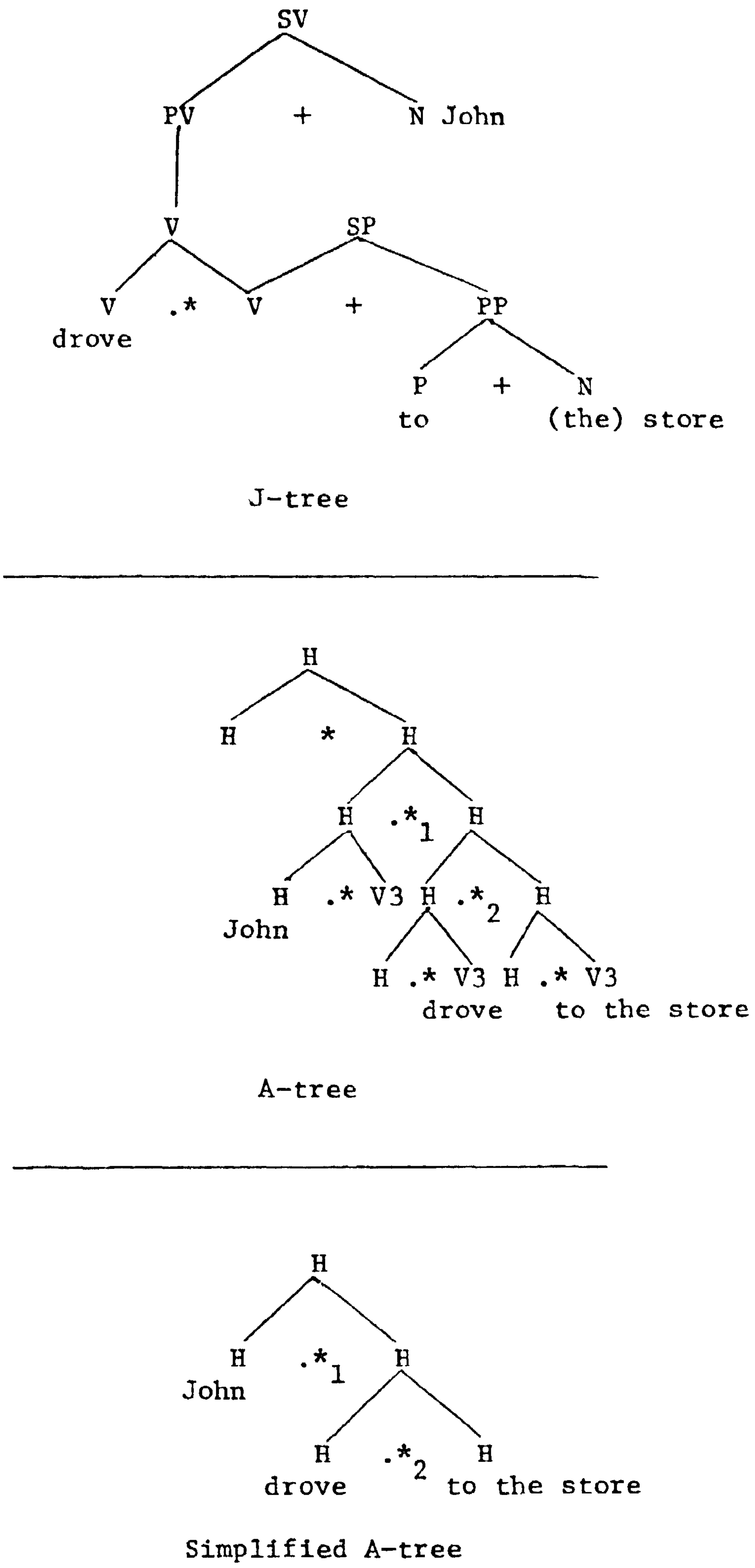


Figure 8. Explicit modalizer .

The identity of "child" is retrieved and placed in the article "the", filling it entirely and leaving no remainder. However, for our purposes, we will leave the modalizers implicit and simply use N (the) cat.

This brief discussion of specialized subjunction and modalizers will suffice for us to reexamine the three sample sentences presented at the beginning of the chapter, but this time in terms of J-trees and A-trees.

"John drove to the Store." Figure 9 shows the J-tree and A-tree for the neutral reading of "John drove to the store" (sentence 1a of Figure 1). The J-tree (a semantico-syntactic representation) is consistent with the version of Junction Grammar described by Lytle (1975). The A-tree (a phonological representation) is consistent with Junction Phonology (Lytle, 1976), except that the internal structure of the V3 nodes is not shown. This A-tree specifies that the sentence is to be pronounced in two units "John" and "drove to the store", and "drove to the store" is further divided into "drove" and "to the store." The subjunctions numbered 1 and 2 indicate the relations between the sub-phrases. In an articulation tree, a left subjunction between H constituents indicates that the right



J-tree

A-tree

Simplified A-tree

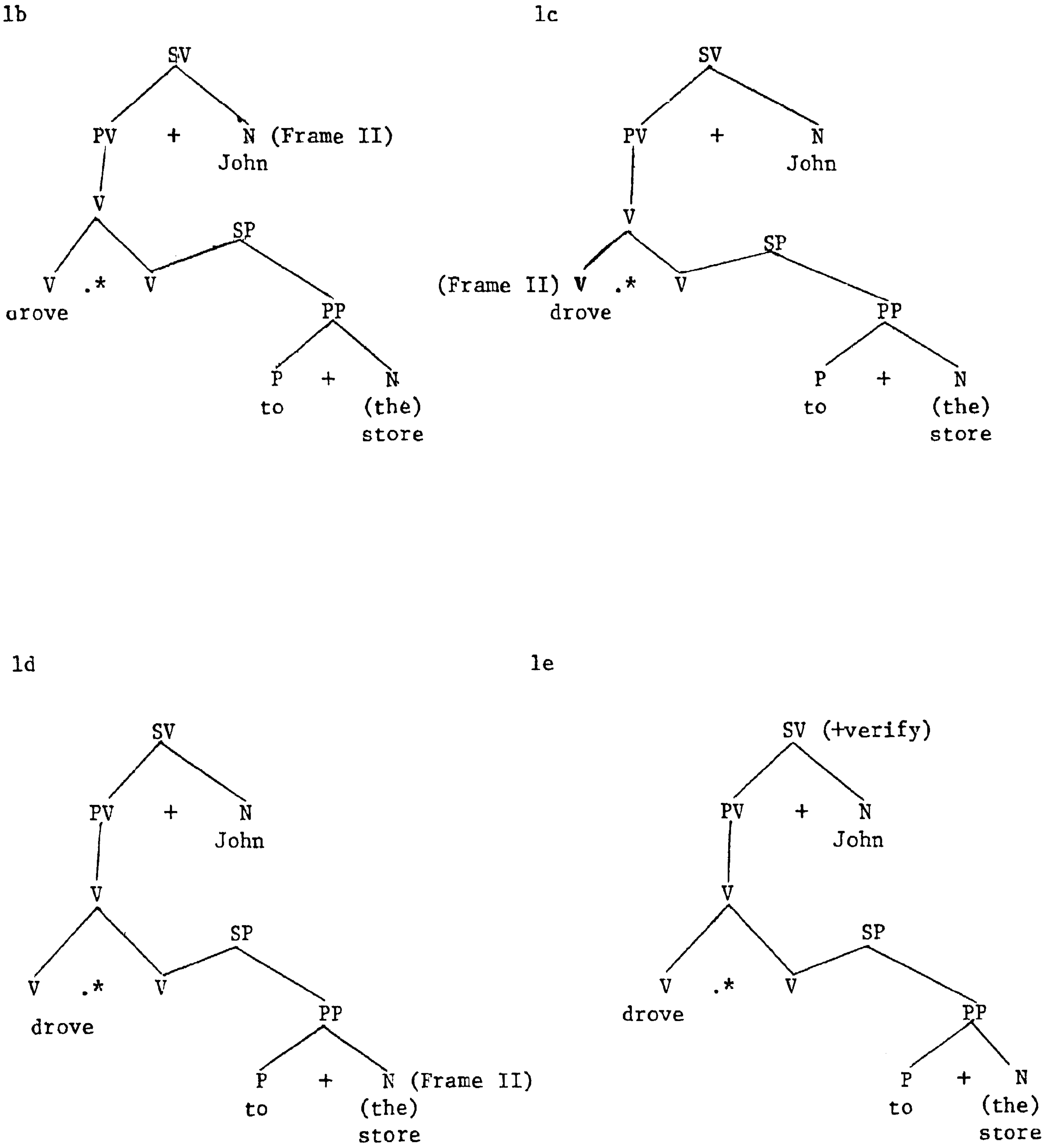
Figure 9. "John drove to the store" Version 1a operand is prosodically subordinate to the left operand. As for the pitch contour, a left subjunction causes a downward pitch shift. Similarly, a right subjunction causes an upward shift. The extra subjunction at the top of the A-tree is available for adding prosodic feature specifications relevant to the entire sentence. The A-tree system of representation is very flexible and a different A-tree could be used if it were decided to group the elements of the sentence differently. At the bottom of Figure 9 is a simplified version of the A-tree, which is used throughout the rest of this paper to make the trees easier to read. But it should be noted that the computer implementation uses the trees in their full form.

Having described the J-tree and A-tree for the neutral form of "John drove to the store," we now consider how the trees differ for the four other versions shown in Figure 1. In versions b, c and d we stress "John," "drove" and "to the store" respectively. This stress is the reflection of an implicit frame II modifier in the J-tree (see Figure 10). For example, according to Junction theory, when the context is "Who drove to the store?", "John" is implicitly modified by a right interjunction which indicates that John has been selected out of a set of possibilities. A possible explicit frame II modifier would be:

> "Of the persons who might have gone to the store, <u>John</u> drove to the store.

At this point, it is worth discussing a very general relationship that has been observed between J-trees and English prosodic stress (Figure 11):

(1) In a full subjunction, any time a remainder is induced (i.e. by *- or -*) in an operand, the other operand receives a stress (e.g. <u>two</u> *- boys). (Continued on page 23.)



J-trees

Figure 10. "John drove to the store" Versions 1b - 1e

(2) In an interjunction, any right interjunction causes a stress on the primary operand, and a left hyphen subjunction causes a stress on the V3 of the subordinate part of the interjunction to which the topic is joined as an enclitic.

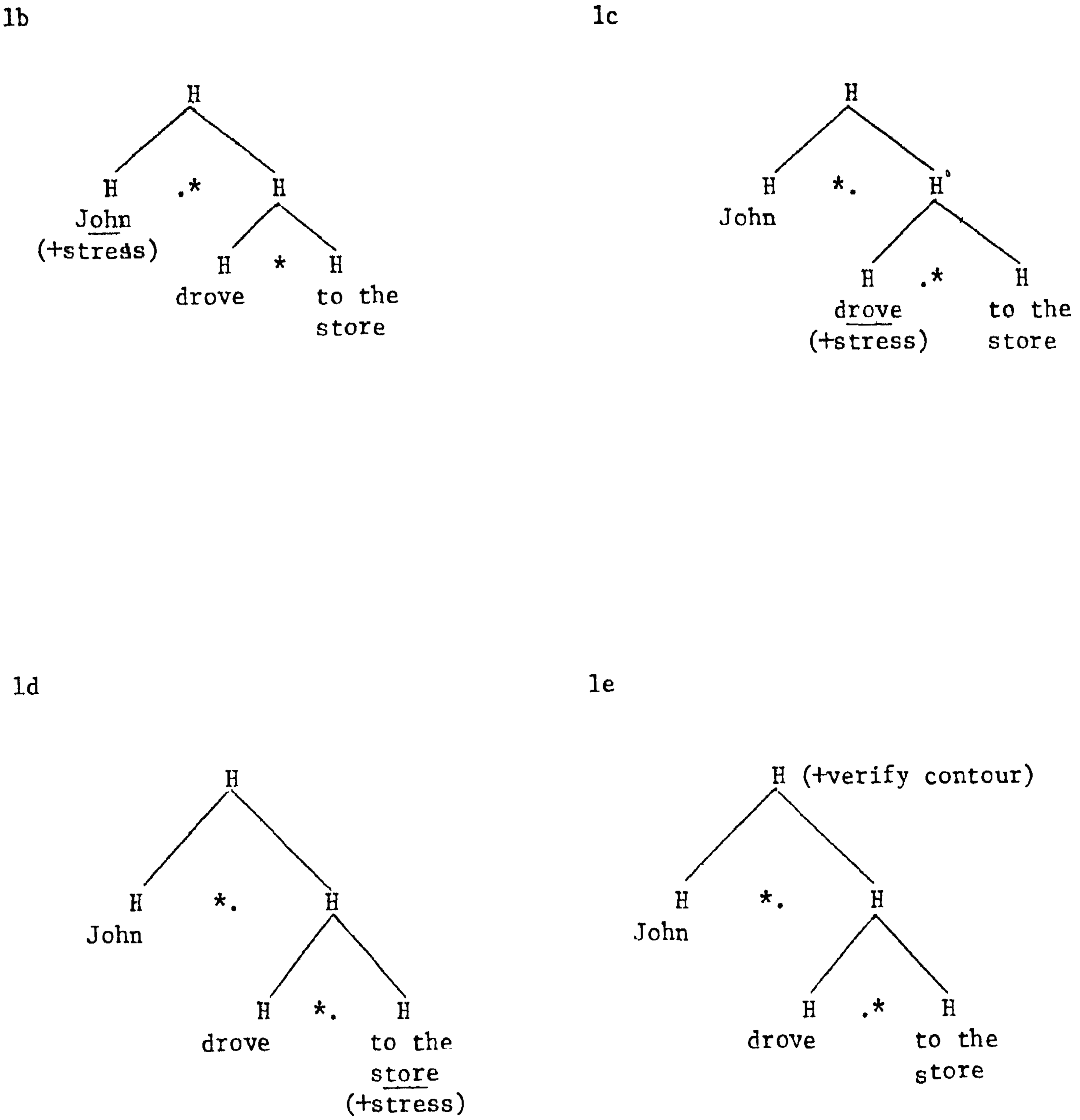
Figure 11. J-trees and English prosodic stress

In the case of the sentence at hand, the implicit frame II modifier, being a right interjunction, causes the primary operand, that is, the element to which the Frame II feature is applied, to be stressed.

Thus we have accounted for the three stressed versions of "John drove to the store." The interrogative version (version le of Figure 1) has a [+ verify] feature on the top of the J-tree. That is, the listener is asking for verification of what was said. This feature is recorded as a prosodic [+ verify] feature in the A-tree. Figure 12 shows the A-trees for these five versions.

Having covered this first example in detail, let us examine the two other sample sentences in a more abbreviated fashion.

"Did John or Mary Come?" Figure 13 shows the J-tree and A-tree for each version of "Did John or Mary come?". As seen in these figures, the semantico-syntactic difference between the two versions is where the interrogative is placed, on the whole sentence or on the conjoined subject. The prosodic difference is that in version 2a, "John" and "Mary" are stressed (stimulated by the interrogation on the OR junction), while



A-trees

Figure 12. "John drove to the store" Versions 1b - 1e

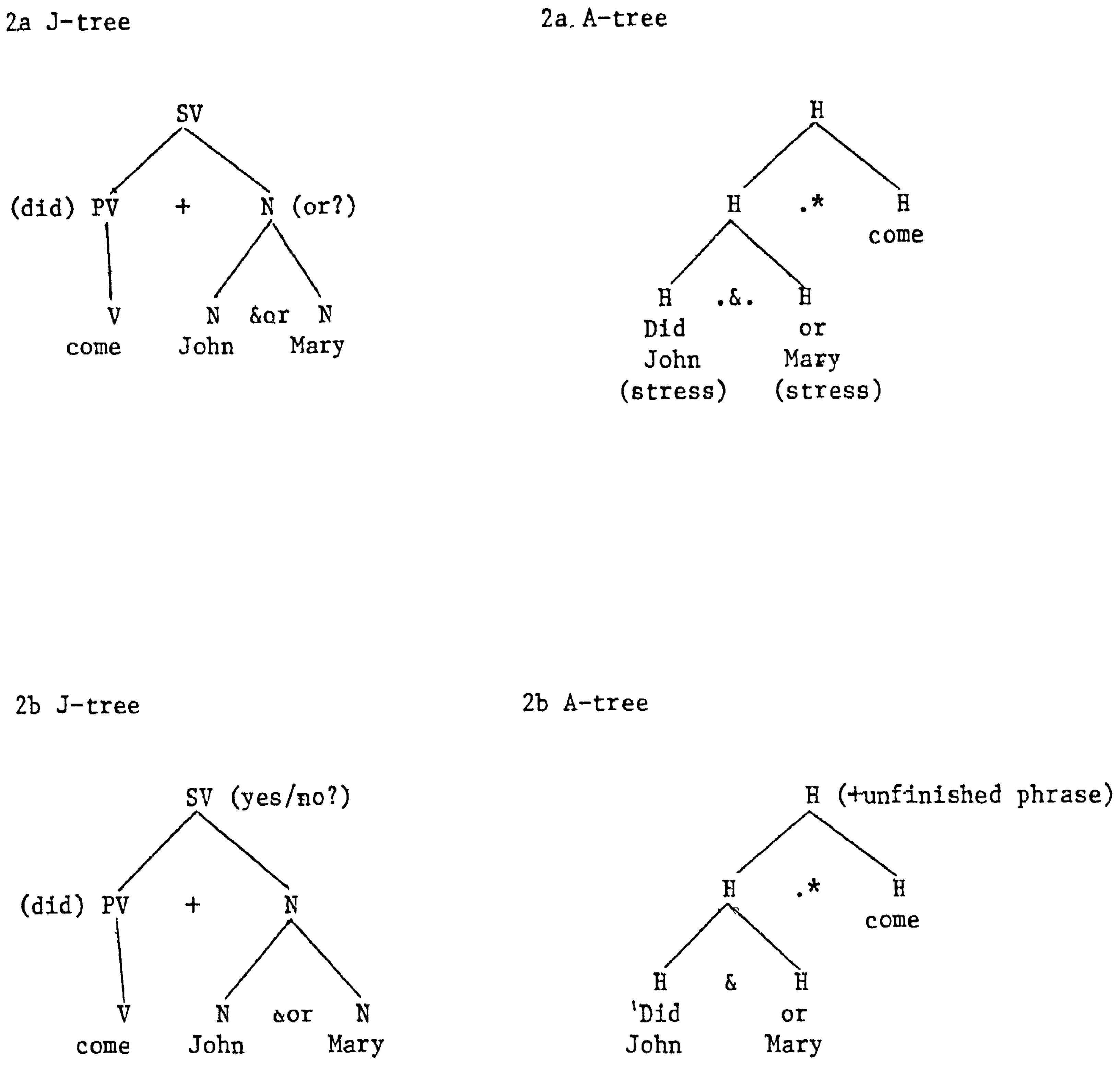


Figure 13. "Did John or Mary come?"

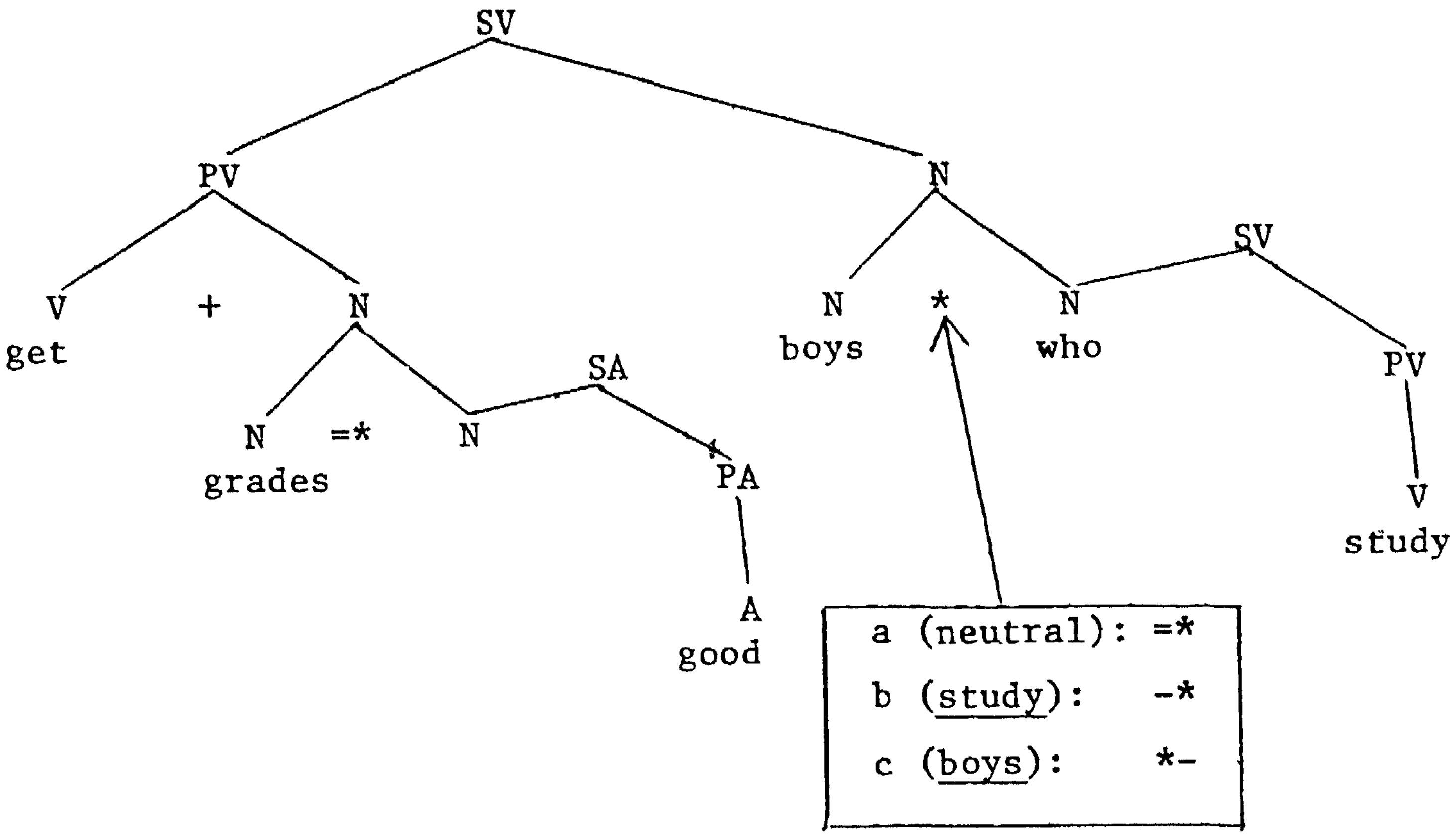
in version 2b, the A-tree is marked [unfinished] because of the [yes-no interrogative] feature on the J-tree. A "finished" version would be "Did John or Mary come or not?".

"The boys who study." Figure 14 shows J-trees and A-trees for the three versions of "The boys who study get good grades." The J-trees differ only in the type of subjunction between "boys" and "who". In the A-tree, "boys" or "who study" is stressed according to the type of subjunction in the J-tree, following the rule stated above. This concludes our discussion of how Junction Grammar handles the three sample sentences presented at the beginning of the section.

C. Text Synthesis Model

We now consider a fully-developed Junction Grammar text synthesis system (Figure 15). This system incorporates the Junction Grammar model of translation so that the input text might be in Spanish and the output in English. In this full system, J-trees adjusted (transfered) for the target language would be needed as well as fully specified A-trees. The A-trees would include the internal structure of the V3 nodes, and the information in the A-tree would be converted into parameters that drive a functional analog of the vocal cords and tract. Clearly, putting together such a system would be a very ambitious project.

A restricted version. At present, we have implemented only a restricted version of the full system, illustrated in Figure 16. In this system we have isolated the pitch contour from other control parameters. Thus, we have chosen to work with an entire sentence as a unit. Essentially, 3 a, b and c J-trees



3 a, b and c A-trees

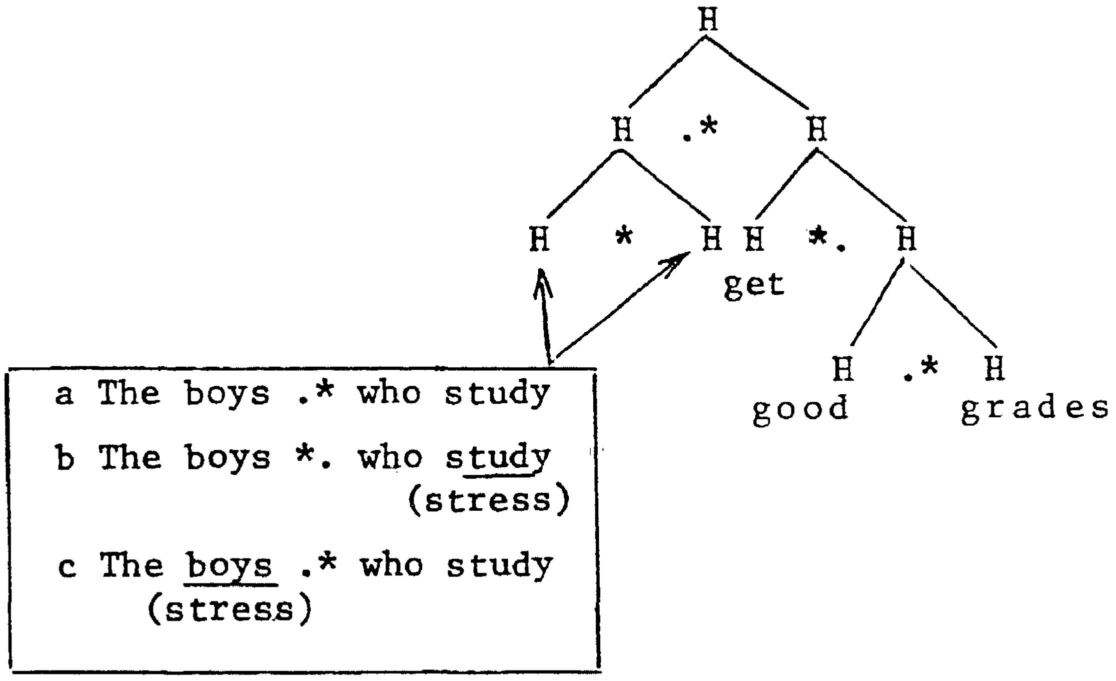


Figure 14. "The boys who study get good grades"

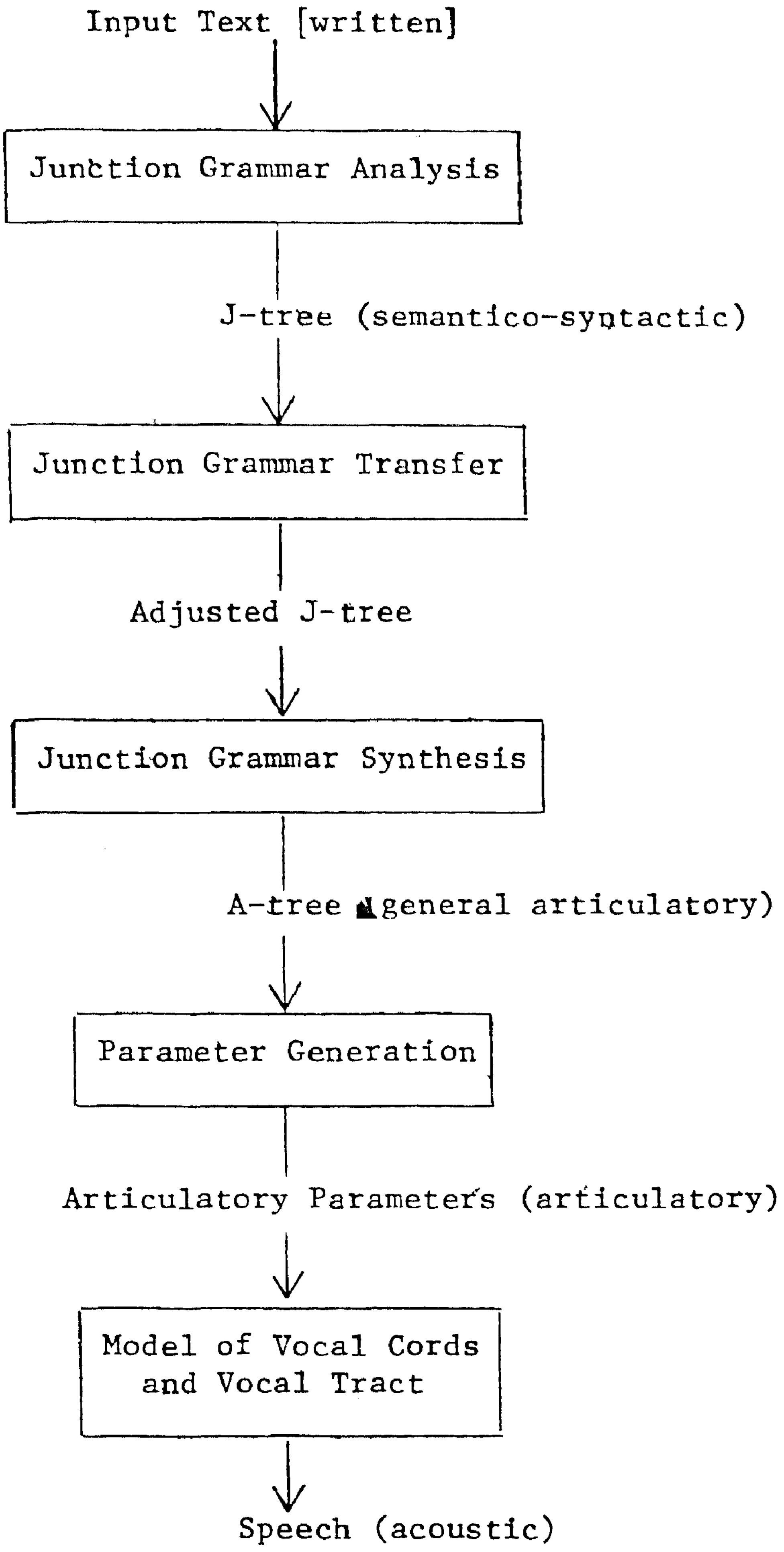

Input Text [written]
Junbtion Grammar Analysis
J-tree (semantico-syntactic)
Junction Grammar Transfer
Adjusted J-tree
Junction Grammar Synthesis
A-tree (general articulatory)
Parameter Generation
Articulatory Parameters (articulatory)
Model of Vocal Cords and Vocal Tract
Speech (acoustic)

Figure 15. A fully-developed Junction Grammar Text Synthesis System

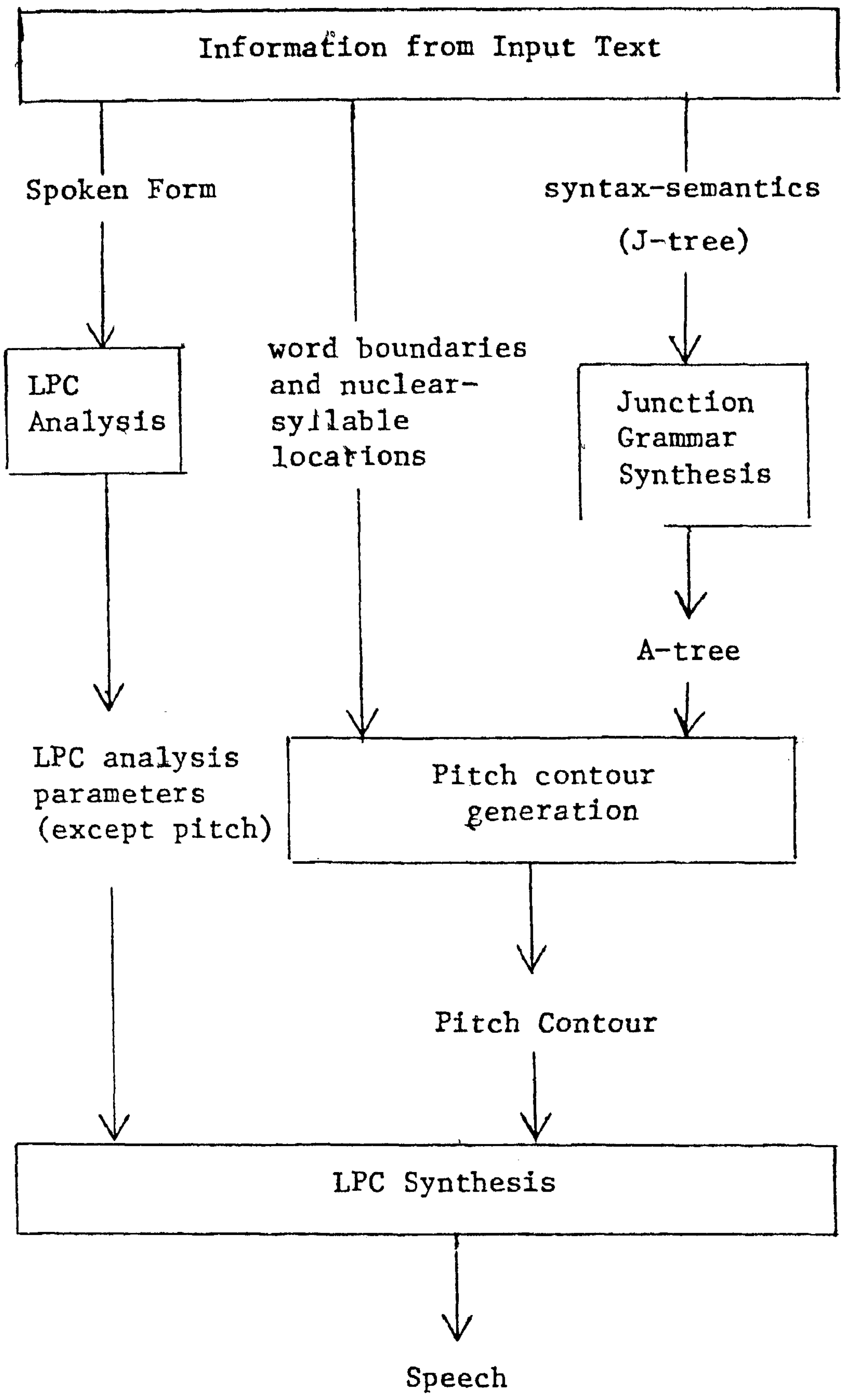


Figure 16. The currently implemented system.

we LPC-analyze the spoken input sentence, enter a J-tree for the sentence, recode the J-tree as an A-tree, generate a pitch contour from the A-tree, replace the natural pitch contour with the generated one, and LPC-synthesize to produce a spoken output sentence.

II. METHOD

The model described in Section I provides a representation for the semantico-syntactic information underlying prosodic contrasts and a very flexible framework for representing phrasing and prosodic features at the general articulatory level. But we have not yet specified how a J-tree is recoded as an A-tree or how the pitch contour is actually obtained from the A-tree. This chapter will describe the computer algorithms that have been implemented to perform these two conversions. Of course, they should not be taken as any kind of final statement concerning the task as they are under continuing development.

A. Recoding a J-tree as an A-tree

The general form of the A-tree is obtained by traversing the J-tree according to the language specific order stored in the J-tree. At each node the algorithm decides whether or not to declare a phrase, thus allowing nested phrases. The criteria for declaring a phrase are:

(1) The topmost node of the J-tree defines a phrase.

(2) If the predicate consists of more than a single verb and a single object, the verb and object will be made into a phrase which will then be joined to the subject.

(3) The contents of each subordinate tree of the J-tree (which is a forest of trees), is phrased under the dominating tree.

(4) Each operand of a conjunction forms a phrase.

The assignment of prosodic features to the A-tree (i.e. [+ stress], [+ unfinished phrase], and [+ verify contour]) is fairly straightforward.

The criteria for assigning [+ stress] to a node are:

(1) A Frame II feature in the J-tree,

(2) A left or right hyphen subjunction (indicating remainder),

(3) The operands of an "OR" interrogative.

Tne directionality of the subjunctions between H-constituents in the A-tree is left except in the following situations:

(1) There is a right subjunction between the A-tree phrases from a simple verb and its complex object in the J-tree,

(2) If a phrase is marked [+ stress], the sub-phrases of the phrase are subordinated to it by adjusting the directionalities of the subjunctions.

B. Background of the A-tree to Pitch Contour Algorithm

With this overview of the J-tree to A-tree conversion algorithm, we describe an algorithm to obtain a pitch contour from an A-tree. The evolutionary phases in the development of this algorithm were:

Plots. We plotted pitch and intensity against time for various readings of several sentences.

Manual Contours. In order to determine which aspects of the pitch contour are essential to natural-sounding synthesis, we programmed a system to allow manual specification of the pitch contour with linear interpolation between specified points and to then permit listening comparison of synthesis outputs with natural versus manual contours.

First Algorithm. Based on these initial experiments, we programmed a simple pitch contour algorithm that imposed on each phrase a contour selected from a fixed inventory of contours and algebraically added in a pitch "bubble" to the syllable of a prosodically stressed V3. In this initial system we were able to create multiple readings of sentences like "John drove to the store" from a single set of LPC analysis parameters, varying only the pitch contour. In other words, we concluded that although the perceptual phenomenon called prosodic or suprasegmental stress is well-known to be based on several acoustic parameters, including pitch (i.e. fundamental frequency), intensity and duration, in at least some cases, changing only the pitch contour is sufficient to cause a word to be perceived as stressed or not stressed. However, after considerable theoretical discussions, we decided to abandon the approach of using a fixed inventory of prototype contours and try a more dynamic approach, which we will now describe.

C. Current A-tree to Pitch Contour Algorithm

Given an A-tree and an option code to indicate initial and final values and bounds on parameters, the algorithm assigns an initial and final pitch based on the option code. Then the A-tree is traversed in left-right order. Upon encountering each V3, we assign a pitch to the core of its nuclear syllable as follows:

(1) The first V3 receives the initial pitch of the sentence.

(2) A left subjunction causes a ratio decrement (about 0.90) to the last assigned pitch.

(3) A right subjunction causes a ratio increment (about 1.12) in relation to the last assigned pitch.

(4) A conjunction causes no change to date, but further research is needed.

(5) An H-constituent dominating multiple V3's receives the average of the most recently assigned pitch level and the highest pitch assigned to any of its operands.

Then the contours between nuclear syllables are defined as valleys whose depth increases with the distance in time between the nuclear syllables it joins. After the initial contour is defined, two types of contour adjustments are added:

(1) Adjustments in the pitch contour caused by stop consonants. We call these stop discontinuities because when the speech waveform becomes voiced again after a stop, the pitch is significantly higher than when the stop began but soon settles down to a value which would be predicted by smooth interpolation of the pitch contour over the unvoiced segment.

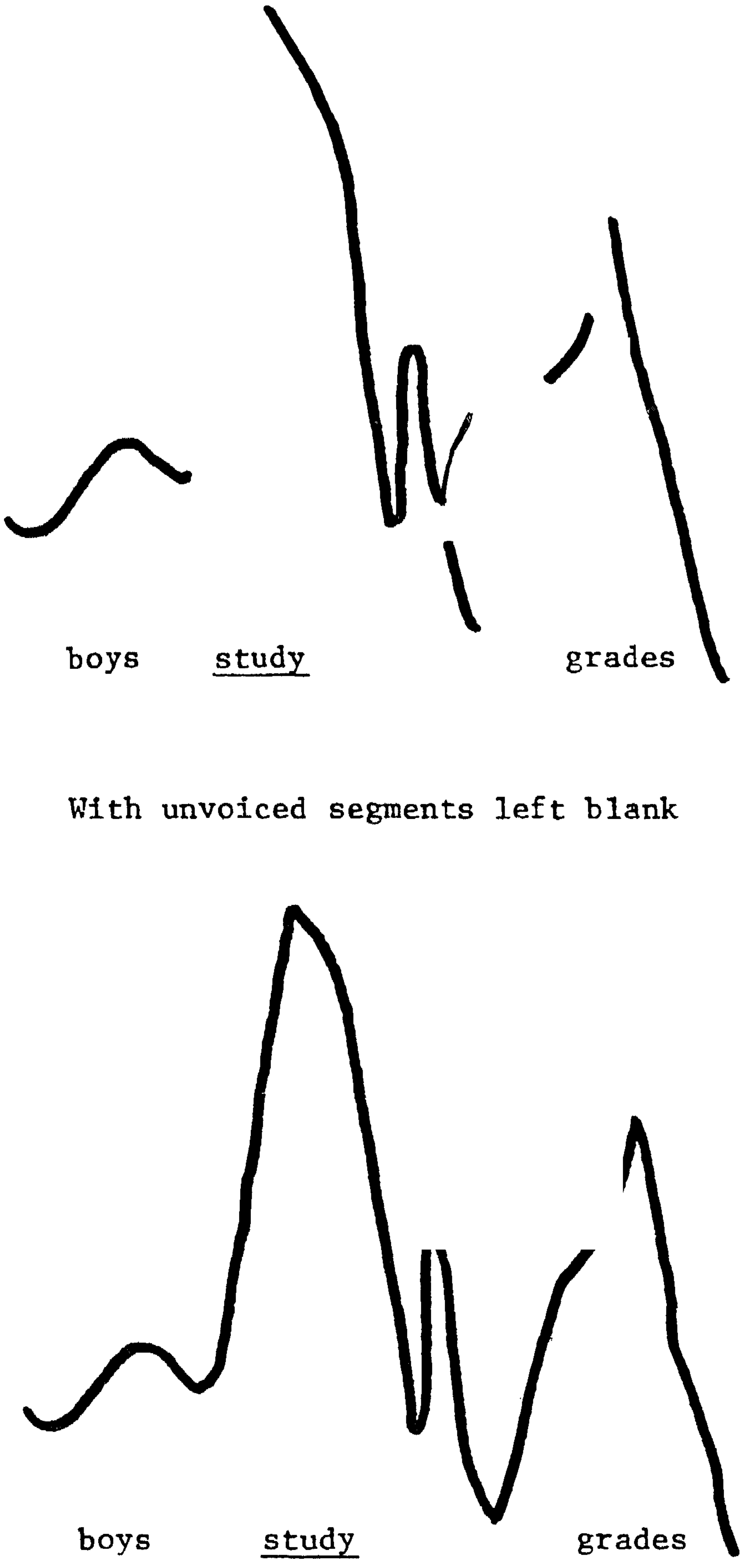
(2) The pitch "bubble" associated with a stressed V3.

Although the above algorithm is not complete, it works reasonably well and does have one already mentioned aspect which we repeat here for emphasis : The pitch contour is generated from the A-tree in a completely _dynamic_ manner. That is, there is no fixed inventory of pitch levels or phrase contours. Each new pitch level is assigned _relative_ to previous values assigned and in accordance with preassigned absolute pitch limits (e.g. 60 Hz and 200 Hz) and the overall structure of the A-tree. This means that, although we have so far restricted ourselves to carefully spoken speech, this system may have the flexibility to eventually allow synthesis of varying speech rates, i.e. very slow and careful or very fast and sloppy speech by appropriate option codes in the

J-tree to A-tree algorithm and the A-tree to pitch contour algorithm.

D. Sample Pitch Contours

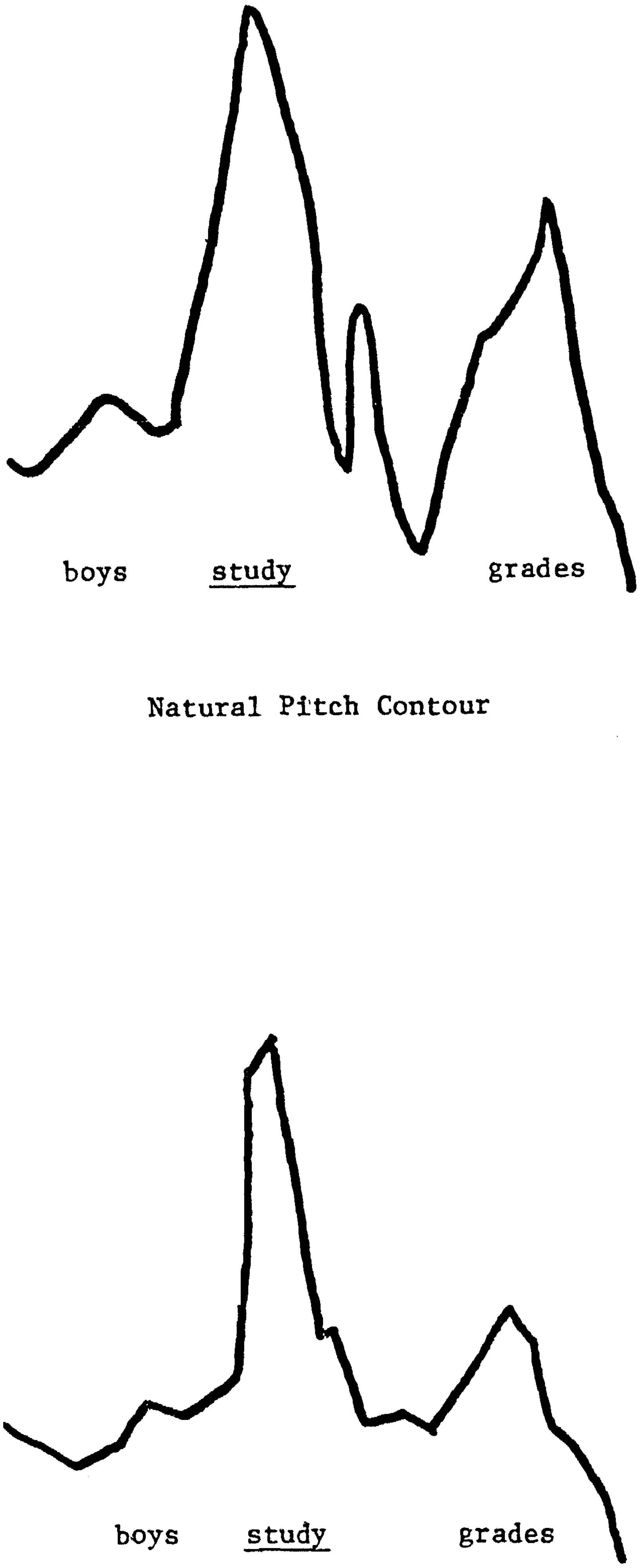
To conclude this chapter we present some graphs of pitch contours for the sentence "The boys who study get good grades." Figure 17 shows a natural, a rule-generated and a manual pitch contour for sentence 3b ("The boys who study get good grades"). Figure 18 shows a natural and a rule generated pitch contour for sentence 3c ("The boys who study get good grades"). Note that these two contours ar imposed on the same set of LPC analysis parameters to produce the two readings. Figure 18 also shows a rule generated and a natural contour for "The cat that the dog chased got away."



With unvoiced segments left blank

With unvoiced segments filled in for easier comparison with rule-generated contours

Figure 17a. Natural contour for sentence 3b ("The boys who <u>study</u> get good grades")



Natural Pitch Contour

Rule-generated Pitch Contour

Figure 17 b. Natural and rule-generated contours for sentence 3b

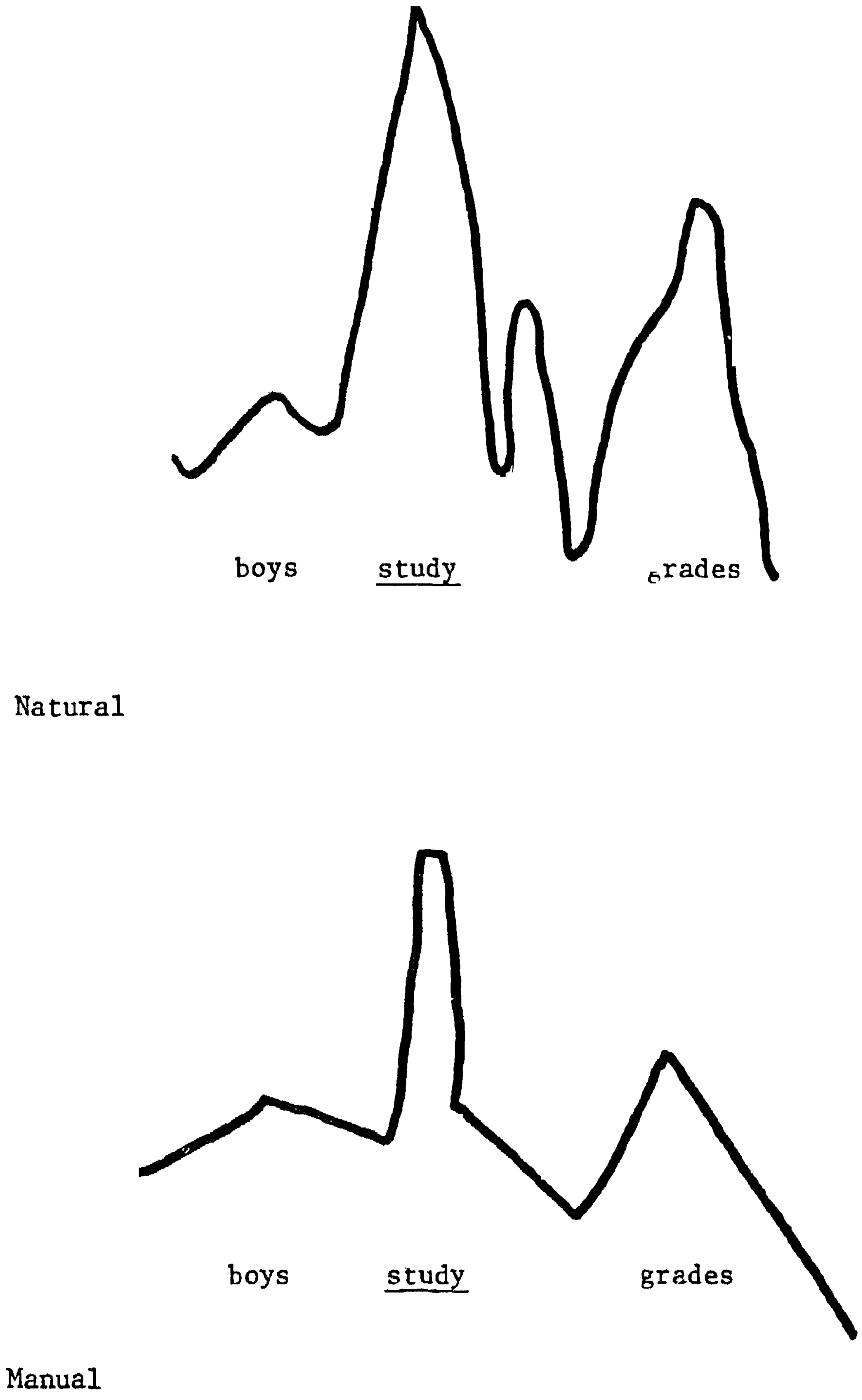
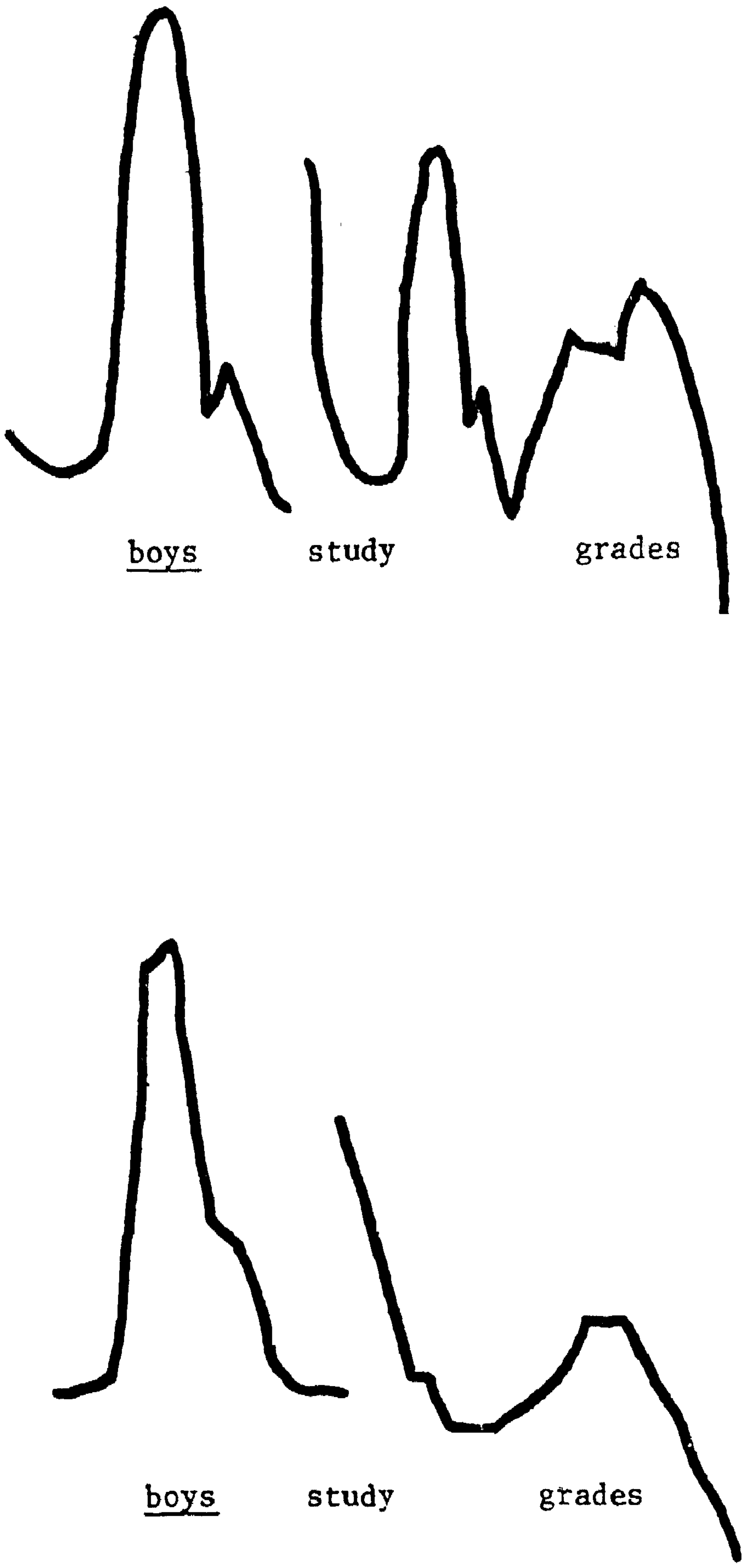


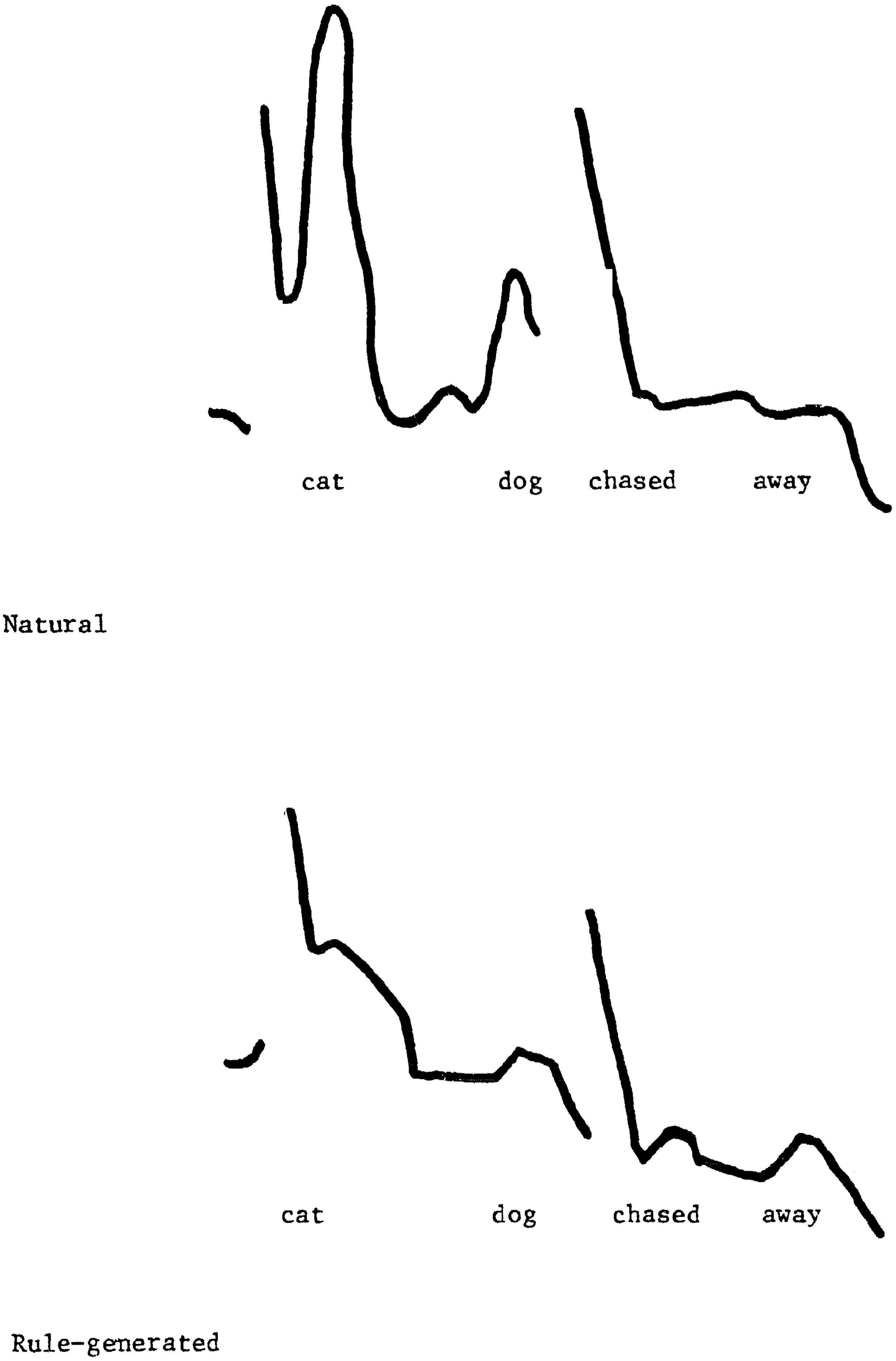
Figure 17c. Natural and Manual Contours for sentence 3b



Natural

Rule-generated

Figure 18a. Natural and rule-generated contours for sentence 3c ("The boys who study get good grades.")



Natural

Rule-generated

Figure 18b. Natural and rule-generated contours for the sentence "The cat that the dog chased got away."

III. EVALUATION AND DISCUSSION

We produced a demonstration tape of LPC synthesized speech using natural, monotone, and rule-generated pitch contours. Figure 19 shows the contents of the tape. Various subjects said that although the sentences with rule-generated pitch contours did not sound as natural as the natural versions, they could clearly perceive the same distinctions in the rule versions as were made in the natural versions. Thus we established two criteria of evaluation: naturalness of intonation, and "intelligibility" of intonation, by which we mean a human listener can correctly perceive which reading of a multiple-reading sentence was intended.

A. Format of the Test

In order to obtain a quantitative evaluation of the system, we devised the following four part test, which was presented to 17 subjects. The sentences in the test consisted of 35 versions made from a dozen sets of LPC analysis parameters by imposing various natural, manual, monotone, and rule-generated pitch contours on them. In the first part listeners were asked to rate readings of 34 sentences on a scale from 1 to 5, where "1" meant the intonation sounded mechanical or monotone, and "5" meant the intonation sounded natural. In the second part, listeners were presented with 24 sentence pairs and asked to indicate whether the first or second sentence sounded more natural.

The third and fourth parts of the test dealt with intelligibility of intonation. In both of these parts, the subjects heard a sentence and indicated which of several possible readings the intonation was intended to convey. The only difference between these last two parts was the method of designating the different readings. In the third part, the

NATURAL vs. GENERATED INTONATION

Natural Intonation	Generated Intonation
1. John drove to the store.	2. John drove to the store. (monotone)
	3. John drove to the store.
	4. *John* drove to the store.
	5. John *drove* to the store.
6. Did *John* or *Mary* come?	
7. Did John or Mary come?	8. Did John or Mary come? (monotone)
	9. Did *John* or *Mary* come?
	10. Did John or Mary come?
11. The boys who study get good grades.	
12. The *boys* who study get good grades.	13. The boys who study get good grades. (monotone)
	14. The boys who *study* get good grades.
	15. The *boys* who study get good grades.
16. They are eating apples.	
17. They are *eating* apples.	18. They are eating apples.
	19. They are *eating* apples.
20. *I* have *one*.	
21. *I* have one.	22. *I* have *one*.
	23. *I* have one.
24. The cat that the dog chased got away.	25. The cat that the dog chased got away.
26. John buys rice?	27. John buys rice?

Figure 19. Contents of Preliminary Test Tape

readings were designated by underlining and using a period or question mark at the end. In the fourth part, the readings were designated by an indication of a typical context for that reading. (Appendix D contains additional details of the test and the results).

B. Test Results

Table 1 gives the results of the first part, where sentences were rated on a scale from 1 (mechanical) to 5 (natural). Natural pitch contours received the highest score as expected, followed by manual contours based on the natural contour, rule-generated contours and monotone "contours" in that order.

Table 1 COMPOSITE AVERAGE SCORES

Natural	Manual	Rule-generated	Monotone
4.14	3.76	3.61	1.24

A paired t-test applied to the average scores for natural and rule contours for each listener showed a statistically significant overall preference for natural contours.

In part 2, in a balanced subset of 12 paired comparisons where natural, manual and rule versions were paired in all possible ways, the natural contours received 87 votes, the manual ones received 76 and the rule contours received 41. Several subjects mentioned after the test that the natural, hand and rule versions of the second sentence, ("The cat that the dog chased got away") were indistinguishable in naturalness of intonation. Using a non-parametric sign test technique, we postulated that if there were a significant preference for one pitch contour method over another, the listeners would be consistent in their choice, regardless of the order of presentation. Specifically, if four or fewer subjects out of 17 changed their minds, we can conclude a preference for a given pair and its reverse.

Using this criterion, we found that for the first sentence, the natural version was significantly preferred but for the second sentence, there was no clear preference for the natural over the rule version.

In parts 3 and 4, we tested for "intelligibility" of intonation by presenting sentences and asking which of several possible readings was intended. We evaluated the results of this part by preparing confusion matrices. (Figure 20.) Each one deals with readings of a single sentence, showing reading transmitted and pitch contour method (N=natural, R=rule) compared to reading received by the listeners. All readings are listed in Appendix D.

A simple Chi-Square test shows that for a given row of one of these confusion matrices, 24 correct votes out of 33 or 34 are sufficient to show significance at the .05 level. Results for part 4 were similar.

C. Transmission Problems

Some of the sentences were not well transmitted by the above definition. A consideration of these indicates the kinds of problems that arose. For example, since the first word of any normal declarative sentence receives some extra stress, the listeners had difficulty distinguishing "John drove to the store" from "<u>John</u> drove to the store." Another problem sentence was "Did John or Mary come?" Although the two

JOHN DROVE TO THE STORE

version sent	version received			
	1a	1b	1c	1d
1 a N	15	19	0	0
1 a R	28	3	2	1
1 b R	1	31	1	1
1 c R	1	0	33	0

I HAVE ONE

version sent	version received	
	5a	5b
5 a N	34	0
5 b N	1	33
5 a R	33	1
5 b R	4	31

THE BOYS WHO STUDY GET GOOD GRADES

version sent	version received		
	3a	3b	3c
3 b N	2	31	0
3 c N	0	0	34
3 b N	3	30	0
3 c R	15	3	16

Figure 20. Confusion matrices for part 3

rule versions were clearly distinguishable (one with falling and one with rising terminal intonation), the listeners made many incorrect choices. This may have been due to either of the following two factors:

(1) As with the other sentences, all the rule versions were based on a single set of analysis parameters, and duration was held constant. In this sentence, duration plays a greater role than in others, and this may have influenced judgment.

(2) There may have been some confusion about what the versions meant, and there may have been confusion with a possible third reading in which "John" and "Mary" are stressed and yet the intonation is rising at the end.

D. Termination Problems

Another problem mentioned by several subjects was that the intonation on some versions (rule and hand versions only) was natural up until the very end of the sentence. We have determined that this is a problem in shaping the contour from the last nuclear syllable to the final pitch of the sentence, assigning an appropriate final pitch, and determining the interaction between the pitch of the last nuclear syllable and the sentence final pitch. Further research is needed in this area.

E. Discussion

This paper is the report of an attempt to generate pitch contours in speech synthesis using Junction Grammar as a theoretical base. Since the various readings of each sentence were made by imposing different pitch contours on the same analysis parameters without changing durations, some versions were less than natural. However, this was to be expected and we feel that it was even desirable in that it pointed out some specific cases in which duration adjustments are necessary. The evaluation also pointed out the need for further research on the shaping of the contour from the last V3 to the end of the sentence. We also realize the need to incorporate some refinements into the system in order to

(1) make degrees of adjustment for fricatives and stops,

(2) improve the naturalness of the contours between nuclear syllables,

(3) make adjustments for the inherent pitch of vowels (Flanagan and Landgraf, 1968).

Based on the results of the evaluation test, we feel it is appropriate to continue use of the Junction Grammar framework and to attempt to develop a word concatenation version with duration, pause and intensity calculations, to attempt better shaping of the contour after the last nuclear syllable, and to examine many more sentence types in order to further test the adequacy of this framework for dealing with the problem of generating prosodic control parameters in speech synthesis.

ACKNOWLEDGMENTS

The author would like to express deep appreciation to co-author Eldon Lytle for many theoretical discussions, to W.J. Strong for sharing his acoustical expertise and LPC analysis-synthesis programs, and to Ronald Millett for his excellent suggestions in our innumerable discussions

during this research and for doing the FORTRAN coding of the J-tree input, display mechanism, the conversion algorithms from J-tree to A-tree, and from A-tree to pitch contour.

REFERENCES

Allen, J. (1976) "Synthesis of Speech from Unrestricted Text," *Proc. IEEE*, Vol. 64, No.4, pp. 433-442, April 1976

Atal, B.S. and S.L. Hanauer (1971) "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Amer.*, Vol. 50, No.2, pp. 637-655

Flanagan, J.L. et al. (1968) "Self-Oscillating Source for Vocal-Tract Synthesizers," *IEEE Transactions on Audio and Electroacoustics*, Vol. AU-16, No.1, March 1968 (see esp. p.60)

Flanagan, J.L. et al. (1975) "Synthesis of Speech From a Dynamic Model of the Vocal cords and Vocal Tract," *The Bell System Technical Journal*, Vol. 54, No.3, pp. 485-505, March 1975

Haavel, R. (1976) "Temporal characteristics of the Pitch Contour," *Acustica*, Vol. 34, pp. 143-157

Halliday, M.A.K (1970) "Functional diversity in language as seen from a consideration of modality and mood in English," *Foundations of Language*, Vol. 6, pp. 322-361

Isacenko, A. and H. Schadlich (1970) *A Model of Standard German Intonation* (The Hague: Mouton)

Leben, W. (1976) Manuscript of an article to appear in *Linguistic Analysis*, No.2, 1976

Levine, A. (1976) Report on Prosodic Research at IMSSS, Stanford University, a preliminary draft of a forthcoming technical report, received March 1976

Lytle, E.G. (1974) *A Grammar of Subordinate Structures in English*, (The Hague: Mouton)

Lytle, E.G. et al (1975) "Junction Grammar as a base for natural language processing," *American Journal of Computational Linguistics*, microfiche No.26

Lytle. E.G. (1976) "Junction Theorv as a base for dynamic phonological representation," Brigham Young University Linguistics Symposium, March 1976

Melby, A. et al. (1975) "Modifying Fundamental Frequency Contours," a paper presented at the 90th Meeting of the Acoustical Society of America, November 1975

Melby, A. et al. (1976) "Generating Pitch Contours from Syntacto-Semantic Representations," Brigham Young University Linguistics Symposium, March 1976

Olive, J.P. (1975) "Fundamental frequency rules for the synthesis of simple declarative English sentences," *J. Acoust. Soc. Amer.*, Vol. 57, No. 2, February 1975

Umeda, N. et al. (1975) "The Parsing Program for Automatic Text-to-Speech Synthesis Developed at the Electrotechnical Laboratory in 1968," *IEEE Transactions on acoustics, Speech and Signal Processing*, Vol. ASSP-23, No. 2, April 1975

Umeda, N. (1976) "Linguistic Rules for Text-to-Speech Synthesis," *Proc. IEEE*, Vol. 64, No. 4, April 1976

APPENDICES

APPENDIX A

BACKGROUND READING

If the reader desires further background in acoustics speech processing and/or Junction Grammar, the following sources may be helpful.

ACOUSTIC SPEECH PROCESSING:

(1) The Speech Chain, P.B. Denes and E.N. Pinson, (Garden City, N.Y.: Doubleday Anchor Books, 1973)
(an excellent non-technical overview)

(2) Speech Analysis Synthesis and Perception, J.L. Flanagan, (New York: Springer-Verlag, 1972) (a thorough technical presentation)

(3) Speech Synthesis, edited by J. Flanagan and L. Rabiner, (Stroudsburg, Penn.: Dowden, Hutchinson and Ross, 1973)
(a collection of key historical and current professional articles)

JUNCTION GRAMMAR:

(1) A Grammar of Subordinate Structures in English, (Lytle, 1974)
(A Description of Junction Grammar. The concepts discussed are still valid in Junction Grammar theory but the notation has changed significantly)

(2) AJCL microfiche #26
"JG as a Base for Natural Language Processing.
(The first chapter is a good introduction to JG but does not go into much detail)

(3) BYU Linguistics class textbooks. There are several Linguistics classes at BYU in Junction Grammar. Ling 426 is an introductory course and Ling 501 is an intermediate class. The textbooks are still in development and have not yet been published but if the reader would like more detail than is available in the first two sources, he can write the BYU Linguistics department for copies of class handouts for Ling 426 and Ling 501. The 501 textbook is the only available source on specialized subjunction.

(4) "Junction Theory as a Base for Dynamic Phonological Representation. BYU Linguistics Symposium, March 1976. (This is the only available document on the A-tree extension of JG. It is reprinted at the end of this microfiche, for the convenience of the reader.)

APPENDIX B

GLOSSARY

A/D
Analog to digital

A-tree
Articulation tree

D/A
Digital to analog

ENCLITIC
A word which generally combines with the following word into a single V3, e.g. "the," "what," "or."

FO
Fundamental Frequency

HERTZ (HZ)
1 Hz = 1 cycle/second

JG
Junction Grammar

J-tree
Junction tree (contains semantico-syntactic information)

LPC
Linear predictor coefficient

NUCLEAR SYLLABLE
The ranking syllable of a V3, in Isacenko (1970) it is called the ictus.

PITCH
In this paper pitch contour is used to mean fundamental frequency contour

PROSODICS
There are word-boundary effect, phrase-level stress contours, and clause-level phenomena which affect the waveform. These factors are referred to as the suprasegmental or prosodic features of speech.

SUPRASEGMENTAL FEATURES
See prosodics.

TEXT SYNTHESIS
Typed-sentence to code to speech-waveform.

V3
A syllable. See Lytle (1976) for a more precise definition.

APPENDIX C

COMPUTER IMPLEMENTATION

The pitch contour generation system described in this paper has been implemented on a PDP-15 computer, equipped with a variety of peripheral devices configured as shown in Figure 21. The VT-15 allows the user to call a package of subroutines from FORTRAN to plot points or draw lines or characters. The system uses the DEC supplied DOS-15 operating system.

The PDP-15 is equipped with 32 K 18-bit words. This is not enough memory for our main pitch contour generation program so we use the DOS-15 CHAIN AND EXECUTE facility to overlay programs that need not be core resident simultaneously.

As indicated in Figure 21, there are two disk drives on the system. One is a standard DOS-15 system pack for system programs and user files. The other drive is mainly for speech data. Data on packs mounted on this drive is accessed through special assembler subroutines that are not part of the DOS-15 operating system. This allows the user to store data contiguously at a higher transfer rate than possible using standard DOS-15 files. This is especially important in transferring large amounts of data from the A/D to disk or from the disk to the D/A in real time. Thus the system can deal with longer segments of speech than can be stored in in-core buffers at one time.

In order to describe the pitch contour system, we will describe the major data files and off-line support programs the system requires. For each sentence to be processed, the system needs (1) an entry in a speech directory file (SPCDIR) which indicates the address and length on the speech data disk of the LPC analysis parameters. (2) An identification

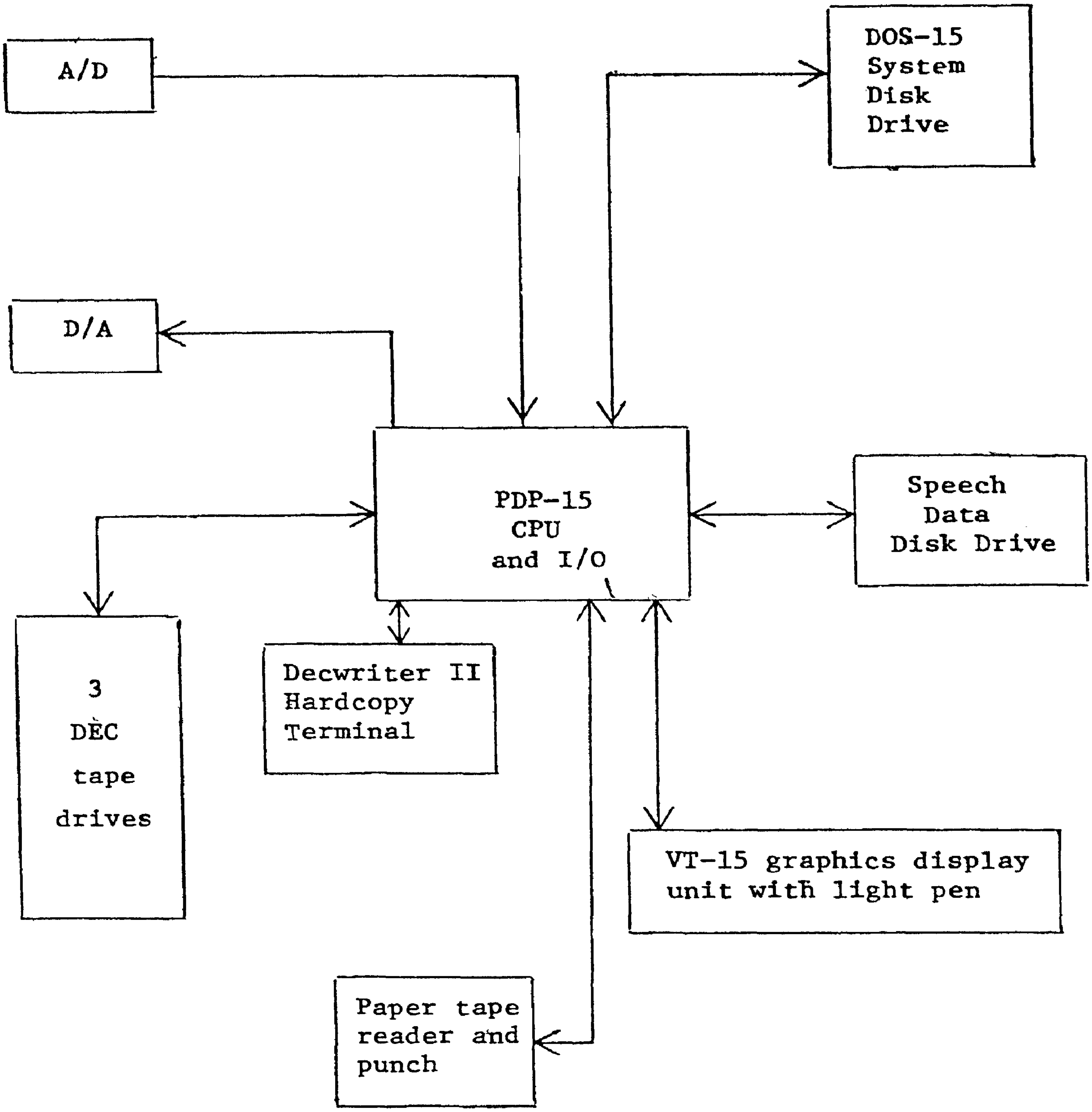

A/D
DOS-15 System Disk Drive
D/A
PDP-15 CPU and I/O
Speech Data Disk Drive
3 DEC tape drives
Decwriter II Hardcopy Terminal
VT-15 graphics display unit with light pen
Paper tape reader and punch

Figure 21. Hardware Configuration

(ID) file which specifies the word boundaries, etc. and the file names of the J-tree files for the various readings of the sentence. The J-tree contains keys to obtain lexical information about each word from a master lexicon file. (3) A J-tree file for each reading.

In order to prepare a sentence for processing, it is tape recorded, then digitized at a 10KHZ sampling rate using a program called DIGTIZ. Then it is LPC analyzed and optionally examined on the graphics display, using a program called ANAPLT. The "PLT" at the end of the name refers to the fact that this program will also produce a hard copy plot of the pitch contour if desired.

The pitch contour generation program is called JTSPCH ("J-Tree to speech"). When this program is executed, it presents a list of available sentences and asks the user to indicate which reading to use in this case. Then the program reads the J-tree file and creates a J-tree in postfix notation. The program then optionally displays the J-tree on the graphics unit, depending on the status of the console sense switches. Then the J-tree is converted to an A-tree, which again is optionally displayed. Then a pitch contour is generated from the A-tree and displayed. Finally, the pitch contour is combined with the LPC analysis parameters retrieved from disk (gain factor, voiced/unvoiced decision and 12 linear predictor coefficients per 10 msec of speech waveform) and the contained parameters are used to synthesize a speech waveform which is stored on a temporary disk area and repeatedly played through the D/A converter to a loudspeaker or headphones for evaluation. If desired, the user can then save it permanently on disk. Another processing option is to create a manual pitch contour instead of generating it from an A-tree. The manual contour can be entered either by drawing it on the graphics unit with the light pen or by entering a list of time and pitch coordinates on the teletype to a subroutine that interpolates linearly between them. Of course, the sentence can also be synthesized using the natural pitch contour retrieved from the original analysis data.

After saving several syntehsized sentences, one can listen to a list of sentences with any desired pause between them using a multiple listening program called MULTIL. MULTIL can receive its control input from either the teletype or from a data file. This option allowed us to create a control file with the regular editing facilities of the operating system and then instruct MULTIL to read it, creating the evaluation test tape in one continuous recording session without any tape splicing.

APPENDIX D

MORE DETAILS ON THE EVALUATION

This appendix contains the following information:

An edited version of the evaluation response form given to the subjects and then four tables showing all responses. Note that the parts of the response form are numbered IA, IB, IIA and IIB. This edited response form shows which versions were used throughout the test but does not contain certain unnecessary details present in the actual response form used. Each version is identified by a code consisting of a number (1-8), a letter (a-e), a letter (N, R, M or H) and possibly another number (1-4).

The first two characters identify the sentence and reading as follows:

(1) a. John drove to the store.

b. John drove to the store.

c. John drove to the store.

d. John drove to the store.

e. John drove to the store?

(2) a. Did John or Mary come? (falling at end)

b. Did John or Mary come? (rising at end).

(3) a. The boys who study get good grades.

b. The boys who *study* get good grades.

c. The *boys* who study get good grades.

(4) a. They are eating apples.

b. They are *eating* apples.

(5) a. *I* have *one*.

b. *I* have one.

(6) a. John, Joe and Fred buy rice.

(7) a. The cat that the dog chased got away.

(8) a. John buys rice.

b. *John* buys rice.

c. John *buys* rice.

d. John buys *rice*.

e. John buys rice?

The next character identifies the nature of the pitch contour as follows:

N = Natural

R = Rule (generated by rule).

M = Monotone (constant fundamental frequency)

H = Hand (manually specified)

If a number follows the H it indicates which hand made contour was used.

RESPONSE FORM

Date ____________________

Name ____________________________________ Age__________ Sex ____________

Occupation _________________________ _________________________________

I. NATURALNESS OF INTONATION

A. Below are two lists of the same 34 sentences. You will hear the first list with a ½ second pause after each sentence. Just listen and don't write anything. Then 10 seconds later, you will hear the second list with a 3 second pause after each sentence. This time, during the pauses, rate each sentence by writing down a number after it. The rating scale is 1 to 5. Remember that the evaluation criterion is intonation only.
So please do not let your judgements be influenced by crackles or pops or hisses.
A rating of 1 means the intonation sounded mechanical or unnatural, for example, monotone or the way computers talk in cartoons. A rating of 5 means the intonation sounded natural, that is, you can imagine the sentence was produced by a human speaker speaking carefully. Please try to distribute your scores over the entire range from 1 to 5.

Before you begin, please read over the entire test to become familiar with it, because you will have only a few seconds to respond to each question.

The test will last 17 minutes.

(The following four pages are an edited, abbreviated form of the rest of the response sheets. The codes in parentheses were not on the actual response sheets. By consulting the key on the previous pages of this appendix, the reader can determine from the codes which version was used for each question.)

I A.

1. I have one.

2. The cat that the dog chased got away.

3. Did John or Mary come?

etc.

33. The cat that the dog chased got away.

34. John drove to the store.

SECOND TIME THROUGH: Rate each sentence (1)Mechanical to
(5)Natural

1. I have one. . . . . ______ (5bR)

2. The cat that the dog chased got away.______ (7aR)

3. Did John or Mary come?. . . . . .______ (2bR)

The rest of part IA will be shown in abbreviated form.

4 (2bN) 5 (3bR) 6 (3bH2) 7 (7aH4) 8(1aM) 9 (5bR)

10(5bN) 11(3bR) 12(1aR) 13(3bN) 14(7aN) 15(2bN)

16(6aN) 17(2aR) 18(7aR) 19(2bM) 20(7aN) 21(6aR)

22(1aR) 23(3bH2) 24(1aN) 25(1aM) 26(2aR) 27(6aN)

28(5bN) 29(6aR) 30(3bN) 31(2bM) 32(2bR) 33(7aH4)

34. John drove to the store . . . . ______ (1aN)

I B.

Pair Number	1st sounded more natural ✓	2nd more natural ✓
1. Did John or Mary come?.	(2aN)	(2aR)
2. Did John or Mary come?.	(2aH1)	(2aH2)

3. Did John or Mary come?. . . . . (2aR) (2aN)

Questions 4-12 deal with sentence 2a using various pitch contours.

4 (H2,N) 5 (H2,R) 6 (H2,R) 7 (R,H1) 8 (N,H2)

9 (R,H2) 10(H1,R) 11(H1,N) 12(R,N)

Questions 13-24 deal with sentence 7a using various pitch contours.

13(R,N) 14(H1,N) 15(H1,R) 16(R,H4) 17(N,H4)

18(R,H1) 19(H4,R) 20(H4,H1) 21(H4,N) 22(R,N)

23(H1,H4) 24(N,R)

II A.

1. John buys rice (8dR)

   a. John buys rice.

   b. John buys rice.

   c. John buys rice.

   d. John buys rice.

   e. John buys rice?

2. Did John or Mary come (2aN)

   a. Did John or Mary come?

   b. Did John or Mary come?

The rest of part IIA will be shown in abbreviated form.

1 (8dR) 2 (2aN) 3 (1bR) 4 (2bN) 5 (2bR)

6 (4aR) 7 (3bR) 8 (1bR) 9 (1aR) 10 (5aN)

11 (2bN) 12 (2aN) 13 (5aR) 14 (1aR) 15 (4bN)

16 (2aR) 17 (5bR) 18 (3cN) 19 (8dR) 20 (3bR)

21 (2bR) 22 (5bN) 23 (4aR) 24 (3bN) 25 (1aN)
26 (3aR) 27 (8eR) 28 (8eN) 29 (3aR) 30 (4aN)
31 (3bN) 32 (4aN) 33 (8eR) 34 (3cR) 35 (3cN)
36 (5bR) 37 (5bN) 38 (1cR) 39 (5aR) 40 (8eN)
41 (4bR) 42 (2aR) 43 (4bN) 44 (4bR) 45 (1cR)
46 (1aN) 47 (3cR) 48 (5aN)

II B.

1. They are eating apples (4aR)
   a. They are in the process of eating apples.
   b. These apples are a variety good for eating as opposed to baking.
2. They boys who study get good grades (3bN)
   a. Neutral
   b. But the boys who play around get bad grades.
   c. But the girls who study don't get good grades.
3. Did John or Mary come (2aR)
   a. Somebody came. Was it John or was it Mary?
   b. Several people came. Did the group include John or Mary?
4. John drove to the store (1bR)
   a. In response to: "What happened?"
   b. In response to: "Who drove to the store?"
   c. In response to: "How did John get to the store?"
   d. In response to: "Where did John drive?"
   e. To ask for verification of what was said.

5. I have one (5bN)

   a. But you have three.

   b. But you don't.

6. John drove to the store (1cR)

   a. In response to: "What happened?"

   b. In response to: "Who drove to the store?"

   c. In response to: "How did John get to the store?"

   d. In response to: "Where did John drive?"

   e. To ask for verification of what was said.

7. Did John or Mary come (2bN)

   a. Somebody came. Was it John or was it Mary?

   b. Several people came. Did the group include John or Mary?

8. They are eating apples (4bN)

   a. They are in the process of eating apples.

   b. These apples are a variety good for eating as opposed to baking.

9. The boys who study get good grades. (3cR)

   a. Neutral

   b. But the boys who play around get bad grades.

   c. But the girls who study don't get good grades.

10. I have one (5aR)

    a. But you have three.

    b. But you don't.

Table D-1

Q#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	5	5	5	4	4	4	3	2	3	3	3	3	3	3	2	5	3
2	4	4	5	5	4	4	5	2	2	4	3	4	3	4	3	5	2
3	4	4	4	3	2	3	4	3	3	2	4	4	4	2	3	5	2
4	5	2	3	2	4	4	2	2	2	4	4	3	3	1	4	5	1
5	5	4	3	4	4	3	5	4	3	3	4	2	3	2	2	5	3
6	5	4	3	4	4	3	5	5	3	3	4	2	3	2	2	5	3
7	5	5	4	5	4	5	4	4	4	2	5	3	4	4	4	5	4
8	1	1	2	1	2	1	1	1	1	1	2	1	1	1	1	4	1
9	5	4	5	4	4	3	5	5	5	5	3	4	4	4	5	5	3
10	5	5	5	5	4	5	5	5	5	5	5	5	5	5	5	5	4
11	5	3	4	5	4	3	4	4	4	4	5	1	3	3	2	4	4
12	5	4	4	2	4	4	5	5	4	4	5	2	5	4	4	3	3
13	5	5	5	5	5	4	5	5	4	4	4	3	4	5	4	5	4
14	5	5	5	5	5	5	4	5	5	5	5	3	4	5	5	5	5
15	5	2	5	3	3	4	3	2	2	4	3	4	3	1	4	5	3
16	5	4	2	2	4	5	5	2	3	5	3	2	4	4	5	4	1
17	5	3	3	4	5	5	4	5	2	4	5	3	5	5	4	5	2
18	4	3	3	4	4	4	3	2	2	4	4	3	0	4	3	5	4
19	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	3	1
20	5	4	3	5	3	5	5	4	4	3	4	4	3	5	4	5	4
21	5	3	2	1	1	5	4	1	2	3	2	3	4	4	3	3	1
22	5	3	4	3	4	5	5	5	3	4	4	4	5	4	4	4	2
23	5	2	2	4	3	5	3	4	2	3	3	5	4	4	3	5	2
24	5	4	4	2	4	5	5	3	5	5	5	4	5	4	5	5	3
25	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	4	1
26	4	2	5	3	2	4	5	5	2	3	5	3	0	5	4	5	2
27	5	4	2	2	4	5	5	3	2	5	3	4	4	4	5	5	3
28	5	5	5	3	5	5	5	5	5	5	4	5	5	5	4	5	3
29	4	3	4	1	3	4	4	2	2	4	3	3	4	4	3	5	3
30	5	5	4	5	5	5	3	5	5	4	5	4	0	5	2	5	4
31	1	1	1	1	2	1	1	1	1	1	1	2	1	1	1	2	1
32	4	4	3	2	4	5	4	4	2	2	3	4	3	3	3	5	2
33	4	4	3	5	4	5	3	3	4	2	5	4	4	4	3	5	4
34	5	4	3	2	5	4	5	5	5	5	5	4	5	5	5	5	4

The responses for part IA. Each row gives the response of subject 1 through 17 to a particular question. A zero response means the subject left that question blank.

Table D-2

Q#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	2	1	1	1	1	1	2	1	1	2	2	1	1	1	1	1
2	2	2	2	2	1	2	2	2	2	2	2	2	2	2	1	2	2
3	2	2	2	2	1	2	2	1	2	2	1	2	2	2	2	2	2
4	1	1	1	1	2	1	1	1	2	1	1	1	1	1	1	1	1
5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
6	1	2	1	2	2	2	1	2	1	1	2	2	1	1	2	1	1
7	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1
8	1	2	2	1	2	2	2	2	1	2	1	2	2	1	2	1	1
9	2	2	2	2	2	2	2	2	2	2	2	1	2	2	1	2	2
10	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
11	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
12	2	2	2	2	1	2	2	2	2	2	2	2	2	2	2	2	2
13	2	2	2	1	2	1	2	2	2	2	2	2	2	2	2	2	2
14	2	1	2	1	2	2	2	2	2	2	1	2	2	2	1	1	2
15	2	2	1	1	1	2	2	2	2	2	2	1	1	1	2	2	2
16	1	1	2	2	2	2	1	1	2	1	2	2	2	2	2	2	1
17	1	2	1	2	2	1	1	1	1	2	2	1	1	2	2	1	1
18	1	2	1	2	1	1	1	1	2	1	1	2	2	2	1	1	1
19	2	2	1	1	2	2	2	2	1	1	1	2	2	1	2	2	1
20	2	2	1	1	1	1	2	2	1	1	1	1	1	2	2	1	1
21	2	1	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2
22	2	1	2	2	2	2	2	2	2	2	2	2	1	2	1	2	2
23	2	2	2	2	2	2	2	2	1	2	2	1	2	2	2	2	2
24	1	2	1	1	1	2	2	1	1	1	2	1	1	1	1	2	1

Responses for part IB. "1" means the subject chose the first element of a pair; "2" means the second element.

Responses for part IIA. (See Table D-3 on next page).
"1" means the subject chose version "a".
"2" means "b"
"3" means "c".
"4" means "d".
"5" means "e".
"0" means no response.

Table D-3

Q#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	4	4	4	4	1	4	4	4	4	4	4	4	4	4	4	4	4
2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
3	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	3	2
4	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	1	2
5	1	1	1	2	1	1	1	0	2	2	2	1	2	2	1	1	2
6	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1
7	2	2	1	2	2	2	2	2	2	2	2	3	2	2	2	2	2
8	2	2	2	2	1	2	2	2	2	2	2	4	2	2	2	2	2
9	1	1	1	1	1	1	2	2	1	1	1	1	1	1	1	3	4
10	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
11	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2	2
12	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
13	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1
14	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	2
15	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2
16	1	1	1	1	2	1	2	1	1	2	1	1	2	2	1	2	1
17	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
18	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
19	1	4	4	4	1	4	4	4	4	4	1	1	4	4	4	1	4
20	2	2	2	2	1	2	2	2	2	2	2	1	2	2	2	2	2
21	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	1
22	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2
23	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
24	2	2	2	2	2	2	2	1	2	2	1	2	2	2	2	2	2
25	1	1	2	2	1	2	2	2	1	2	1	1	1	2	1	1	2
26	1	1	1	1	1	1	1	1	1	1	1	3	1	1	1	1	1
27	5	5	5	5	3	5	5	5	5	5	5	5	5	2	5	5	5
28	5	5	4	5	2	5	5	5	5	5	5	5	5	0	5	5	5
29	1	2	1	1	1	1	2	1	1	1	1	1	1	0	1	1	1
30	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1
31	2	2	2	2	0	2	2	2	2	2	2	2	2	2	2	2	2
32	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
33	5	5	5	3	1	5	5	5	5	5	5	5	2	5	5	5	5
34	1	1	2	1	3	3	3	3	1	1	2	1	3	3	1	1	1
35	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
36	2	2	1	1	2	2	2	2	2	2	2	1	2	2	1	2	2
37	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
38	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
39	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
40	5	5	5	5	4	5	5	5	5	5	5	2	4	5	5	5	5
41	2	2	1	2	1	2	2	2	2	2	2	1	2	2	2	2	2
42	2	1	1	1	2	1	2	1	1	1	1	1	1	1	1	1	1
43	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
44	2	1	1	2	2	1	1	2	2	2	2	1	2	0	1	2	2
45	3	3	3	3	3	3	3	3	3	3	3	1	3	3	3	3	3
46	2	1	2	2	2	1	2	2	1	2	1	2	2	2	1	2	1
47	3	1	3	3	1	3	1	3	3	1	3	2	3	3	1	1	3
48	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

Table D-4

Q#	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
3	1	2	1	2	2	2	1	1	1	1	1	2	1	1	1	1	1
4	2	2	2	2	5	2	2	2	2	2	2	1	2	2	2	2	2
5	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
6	3	3	3	1	4	3	3	3	3	3	3	3	3	3	3	3	3
7	2	1	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
8	2	2	2	2	2	2	2	2	2	2	2	1	2	2	2	2	2
9	2	1	3	1	1	3	3	3	3	1	3	1	3	3	1	3	3
10	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1

Responses for part IIB. (Same format as Table D-3.)

JUNCTION THEORY AS A BASE
FOR
DYNAMIC PHONOLOGICAL REPRESENTATION

By

Eldon G. Lytle

BYU Linguistics Symposium 1976

JUNCTION THEORY AS A BASE
FOR
DYNAMIC PHONOLOGICAL REPRESENTATION

Orientation

MacNeilage has pointed up the difficulty of mediating between abstract unitary phonological representations and the continuous nature of the dynamic speech chain, suggesting that unitary phonological representations are analogous to a sequence of eggs conveyed to the wringer of a washing machine, while the scrambled mess that emerges from the wringer is what must actually be dealt with by those engaged in computer analysis and synthesis of voice.[1] The question, as he states it, is:

> Given that there is a discrete linguistic input to the mechanism of speech production at some state, and given that the mechanism that transmits this input is incapable of discrete units of output, what is the nature of the transformation, at the peripheral stage, of one form to the other.[2]

Lieberman likewise notes a relative neglect of the phonetic level of speech, concluding that a quantitative and explicit phonetic theory has yet to be developed, and suggesting that a successful attempt to construct such a theory should be structured in terms of the anatomic, physiologic, and neural mechanisms of speech production and perception.

Onn, similarly motivated by the notion that speech ought to be described in the context of the organic mechanisms responsible for it, suggests. that:

> It may be argued that an abstract representation may be regarded as instructions for particular types of behavior of the speech-generating mechanism. When these instructions are carried out, the various reactions occurring between different physiological structures will yield a quasi-continuous gesture in which the discrete instructions initiating the gesture are no longer always observable as distinct components. Finally, the execution of these instructions produces the acoustic signal.[4]

The purpose of the present paper is to outline briefly a new system of phonological description currently being used as a basis for voice synthesis at BYU which attempts to satisfy the criteria suggested by MacNeilage, Lieberman, and Onn referenced above. The descriptive system in question is based on the Junction Grammar Model of language developed by myself and my colleagues over the past eight years.[5] It is a model specifically structured in terms of speech-related organs, either as they are known or hypothesized.

An Overview of the Junction Grammar Model

A fundamental tenet of junction theory is that linguistic description must involve not simply multiple stages of derivation, but multiple types of data and data processing required to simulate the functions of different body organs. (See Figure 1.) Thus, the semantic components of the grammar are designed to process data structured for specific semantic tracts, as it were; the articulatory component is designed to process data structured for the vocal tract, the audio component is designed to process data structured for the auditory tract, and so on. Of course, such a model requires distinct rule systems and procedures to operate on the different data types in the various tracts.

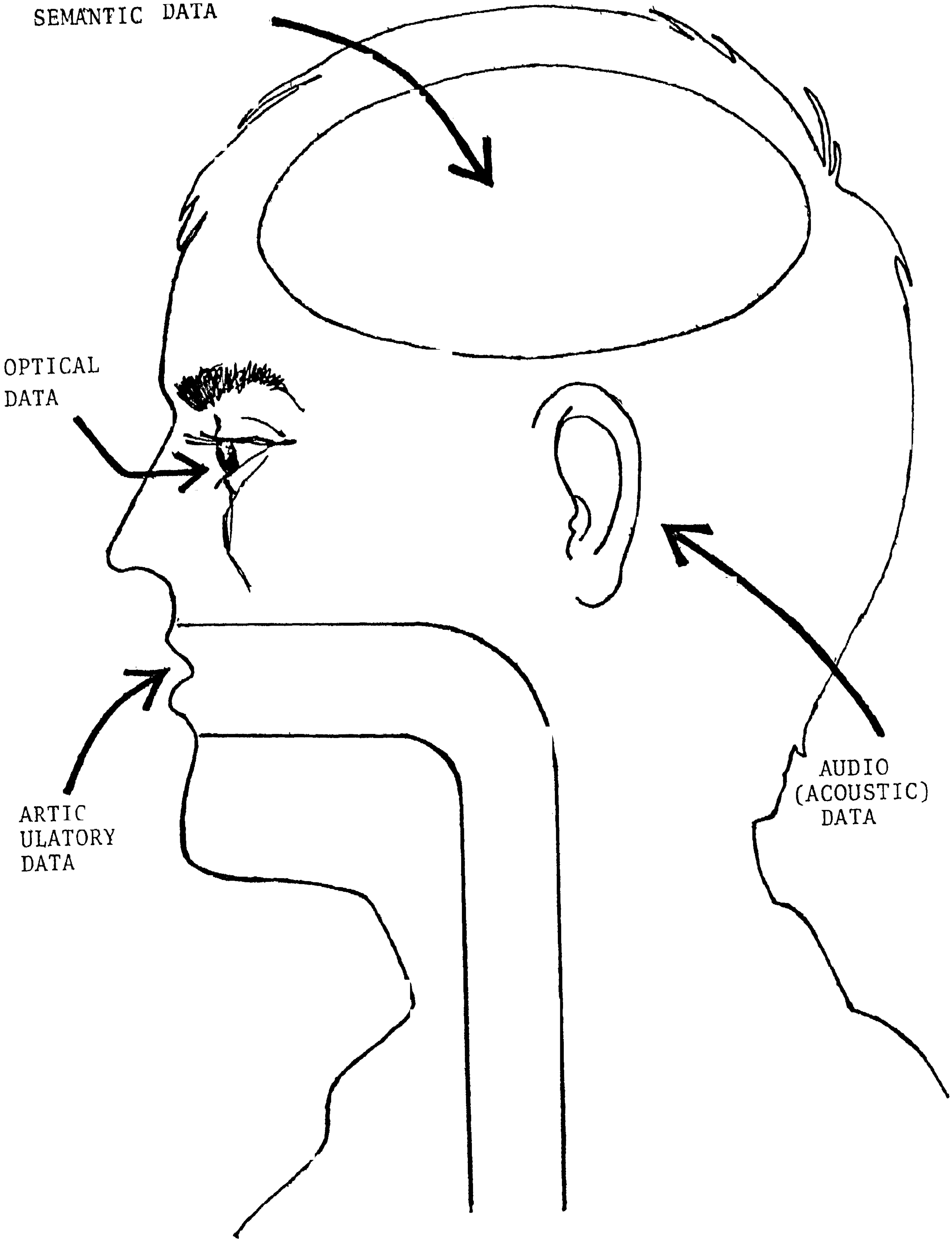
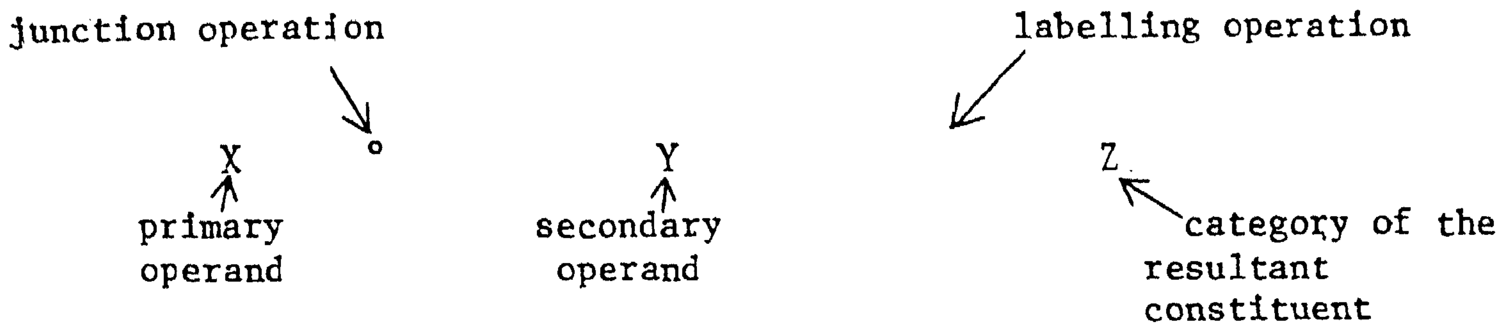

SEMANTIC DATA
OPTICAL
DATA
ARTIC
ULATORY
DATA
AUDIO
(ACOUSTIC)
DATA

Figure 1.

A further tenet of junction theory is that data types may not be intermingled. To do so would, for example, be tantamount to feeding instructions for both the heart and diaphragm to the diaphragm. Of course, semantic instructions could not be executed by a vocal tract, nor could articulatory instructions be executed by a semantic tract. This means, in effect, that a "deep structure" is not transformed (in the usual sense of the word) into a surface structure, but rather that semantic data must be used to stimulate articulatory instructions, orthographic instructions, motor instructions required to produce gestures, to make one blush, etc. Thus, in JG semantic representations there are no lexical items, since these are considered to be articulatory instructions. Similarly, there is no semantic information in phonological representations, since these are a different data type. The various data types are considered to be symbolizations of each other, not transtorms or derivations of each other. Data stimulation between the various tracts or components of the system is accomplished by context sensitive coding/decoding procedures, which are intended to simulate the neural interfaces which coordinate the function of body organs involved in speech production.

Junction Grammar takes its name from Junction Rules (J-rules). (See Figure 2.) J-rules structure data to be processed by the various components of the grammar. The essential ingredients of every J-rule are two or more operands, an operation specifying how the operands are to be joined, and a labelling operation which assigns a category to the operands taken as a unit. Thus, in junction grammar not only do rules for conjunction require an operation symbol (vis. the phrase structure rule S→S & S) but all J-rules, regardless of their specialization.



JUNCTION FORMULA WITH LABELLED PARTS

Figure 2.

A schematic of the model in its present form is given in Figure 3. Basic semantic data is presumed to reside in the form of an information net. Drawing upon information in the net, J-rules organize and structure information pragmatically, i.e. for use in specific utterances in specific discourse environments. Fillmore's arguments for semantic case relate specifically to the need to distinguish between basic semantic relations and pragmatically motivated grammatical relations. The semantic junction trees (J-trees) generated by J-rules then serve as the basis for coding up articulatory instructions, instructions to the arm and hand for writing, or motor instructions of sundry types necessary to produce body language.

Incoming information, on the other hand, is decoded to obtain the pragmatic J-tree which stimulated it, and then each junction in the tree is executed by a semantic processor, resulting in additions to or changes in the information net.

Junction trees occur in both semantic and articulatory data. However, the operands and operations are of a totally different nature from type to type, since in the semantic component they constitute complexes of instructions to be executed by the semantic processor, while in the articulatory component they constitute complexes of instructions to be executed by the vocal tract. The operands of semantic trees are sememes, i.e. units which define locations and states in the information net; the operands of articulation trees are articulemes, i.e. units which relate to locations and states of the vocal tract. Figures 4 and 5 are the semantic and articulation trees, respectively, for the utterance [Hwayyʒə iyt]. Notice, specifically, that while <u>Why did you</u> are not immediate semantic constituents, they are immediate articulatory constituents. The point again, of course, is that while articulatory structure and semantic structure are symbolically related, they are not the same and should not be confused or intermingled.

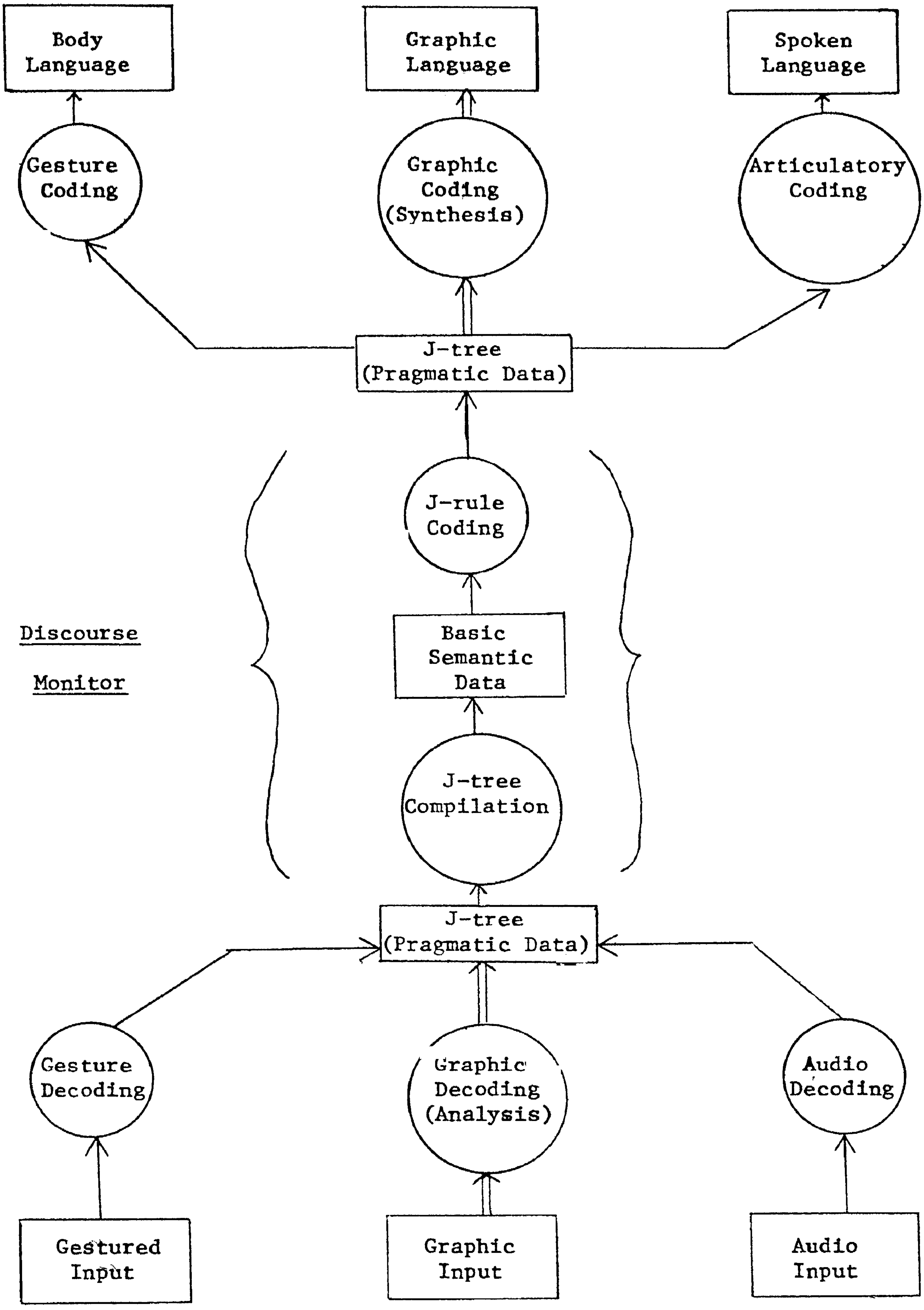
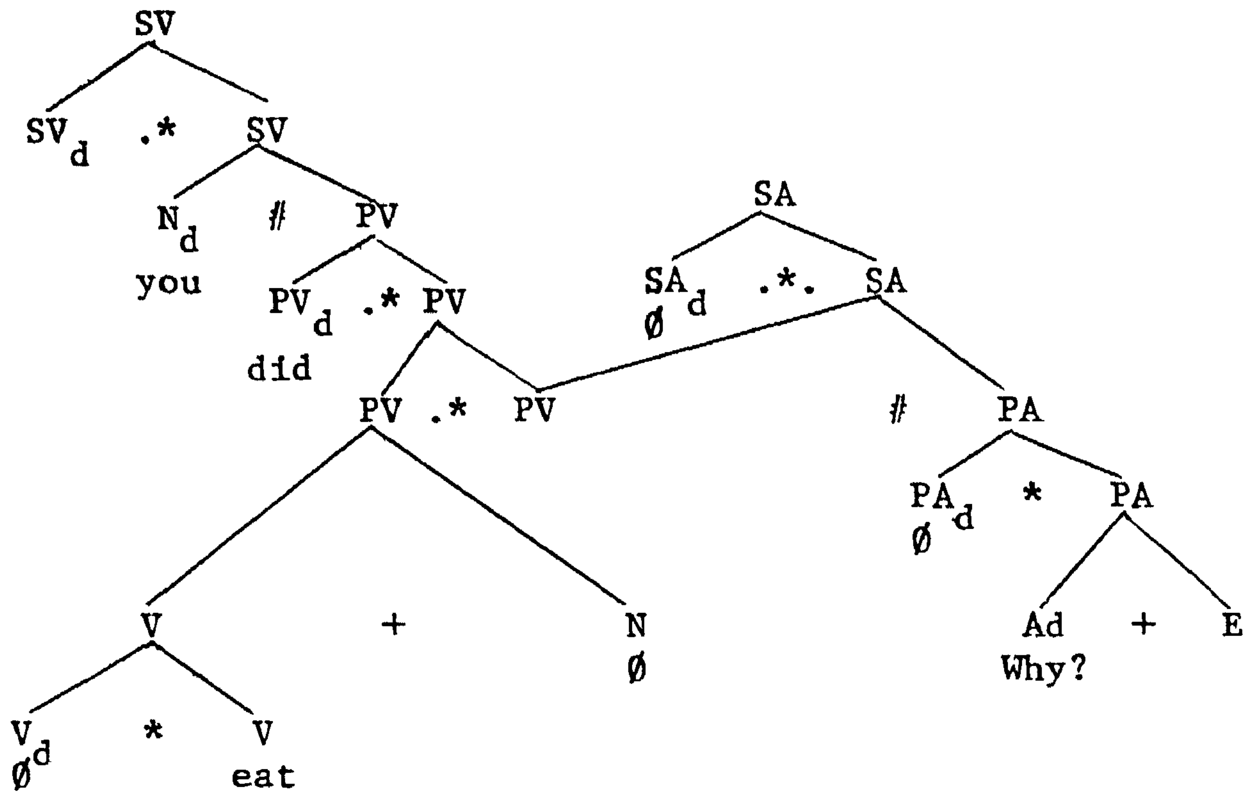

Body Language
Graphic Language
Spoken Language
Gesture Coding
Graphic Coding (Synthesis)
Articulatory Coding
J-tree (Pragmatic Data)
J-rule Coding
Discourse Monitor
Basic Semantic Data
J-tree Compilation
J-tree (Pragmatic Data)
Gesture Decoding
Graphic Decoding (Analysis)
Audio Decoding
Gestured Input
Graphic Input
Audio Input

Figure 3.

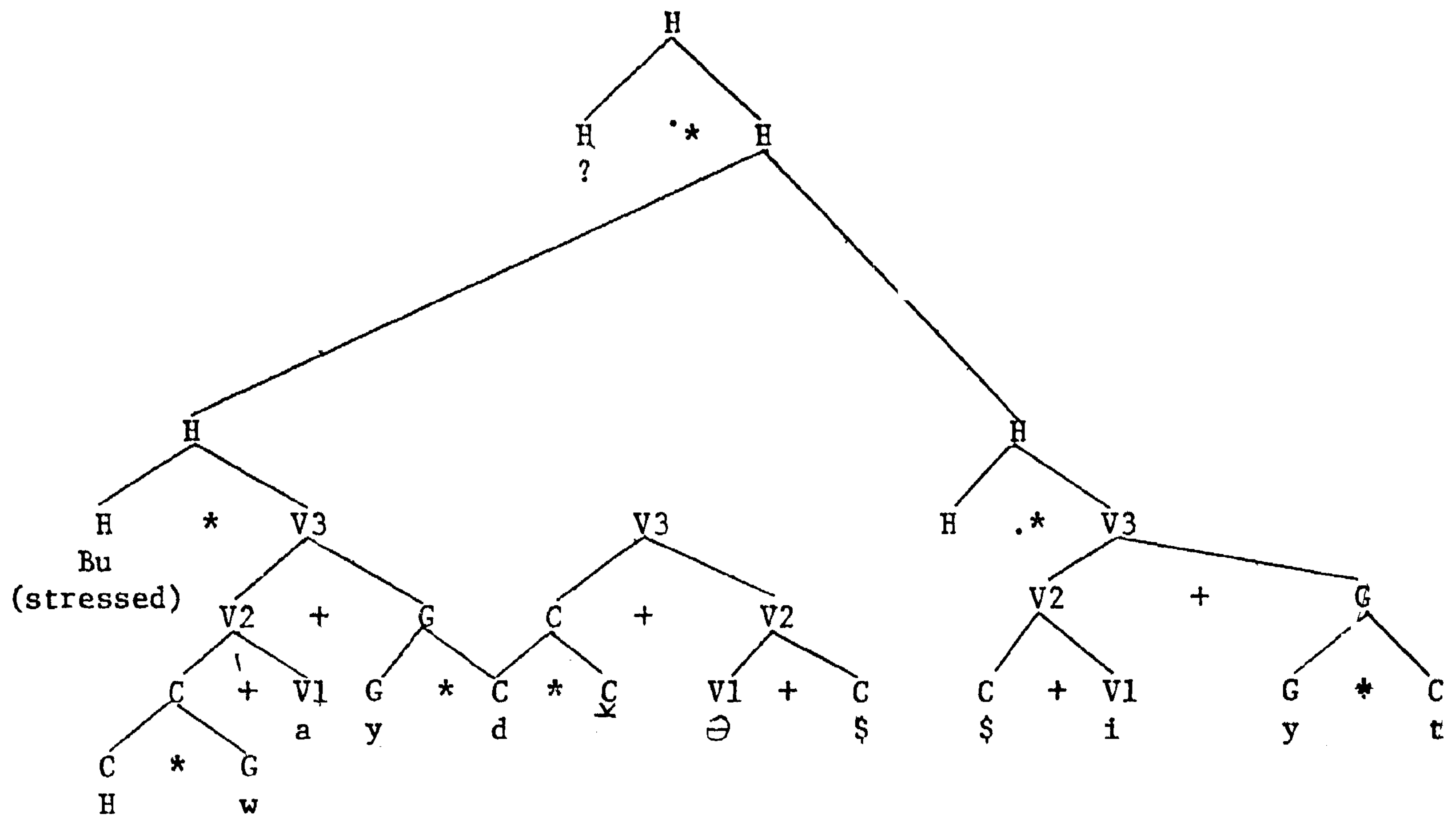
Semantic tree for Why did you eat?



Words represent sememes. There is no lexical data in semantic trees.

Figure 4.

Articulation tree for [Hwayʔə iyt]



Segmentals and suprasegmentals represent articulatory units. There is no semantic data in A-trees.

Figure 5.

Basic Junction Types

Junction theory posits three basic junction operations and numerous subtypes depending upon the data tvpe being described.

(1) Adjunction results in the formation of certain nuclear units which serve as a skeleton to which other elements may attach. In semantic trees, predicates and predications are formed via adjunction. In articulation trees, semi-syllables and syllables are formed via adjunction.

(2) Subjunction results in overlapping constituents of contrasting rank, i.e. where one is in some sense subordinate to the other. In semantic trees, modifiers in all their variety are subjoined. In articulation trees, clustered consonants are subjoined, as well as adjacent syllables having different degrees of stress. Segmental structures are also subjoined to prosodic constituents to account for the supra-segmental aspects of articulation.

(3` Conjunction results in the formation of compounds consisting of units of the same category and rank. In semantic trees, compounds based on and, or, and but are formed via conjunction. In A-trees, conjunction yields evenly spaced non-overlapping units having the same degree of stress.

Now, in the context of this rather general introduction to the subject, let us consider dynamic phonological representations corresponding to the articulatory structure of syllables, words, and phrases.

The Syllable

The intuitive articulatory unit of which words consist is the syllable, which is in turn composed of phonemes. Generally speaking, syllables have as their nuclear component a continuous phoneme with vocalic properties. This nuclear phoneme may be delimited both initially and finally by a phoneme having consonantal properties. Hence, we observe syllables of the following string types:

D = delimiter; W = nucleus; Ø is null

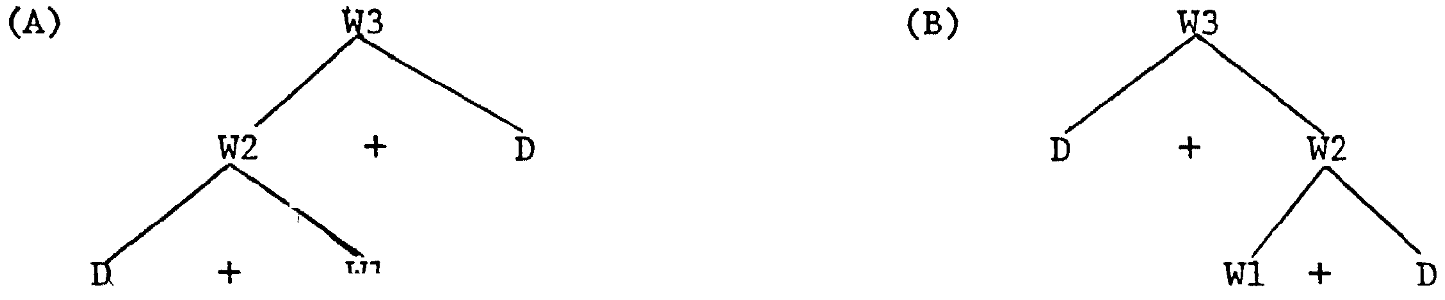
DWD
DWØ
ØWD
ØWØ

If, however, we invoke the concept of a null delimiter $, then these four syllable patterns can be reduced to a single type, DWD, where D may be either null or non-null. The use of the null delimiter $ is actually more than a simplifying assumption, since in many cases non-null segmentals replace $ in the articulation stream either as full geminates or partials of neighboring delimiters.

Articulatory Adjunction

As noted above, junction theory attributes to adjunction those kernel configurations upon which all else is built up. Since syllables are the intuitive units from which words and phrases are formed, we attribute them to adjunction.

There are two basic syllable types, corresponding to whether the syllabic nucleus is joined to the initial or final delimiter. The two cases are illustrated in Figure 6.



NUCLEAR-INITIAL SYLLABLE

NUCLEAR-FINAL SYLLABLE

Figure 6. Two basic syllable types.

Recent research provides useful criteria for deciding when to use each type. Bell-Berti and Harris report that:

> The effects of the terminal consonant on the midpoint of the stressed vowel are not as large as those of the initial consonant. In other words, the carryover effect of the first consonant on the stressed vowel is larger than the anticipatory effect on the second.[8]

For the purposes of this discussion, let us assume that stressed syllables and syllables with strong vowels are nuclear-initial and that other syllables are nuclear-final. It is possible, of course, to formulate junction rules which are not binary, so that a third syllable type whose nucleus was equally joined to both initial and final delimiters could be used. We avoid this formal complication, however, until forced to introduce it by empirical considerations.

Notice that the use of structure to represent syllables makes it unnecessary to use a feature such as [±syllabic]. In comparing the use of this feature to that of the structural notation proposed, we note that each appears to make distinct claims about the notion syllable. Specifically, the feature asserts that a vowel is syllabic, whereas the tree claims that specific sequences of segmentals constitute syllables whose nuclear element is a particular segment.

Node Labels

Turning now to the matter of node labels, we observe that in practice it is desirable to further subcategorize D and W in terms of more specific articulation classes. We therefore define D to include obstruent consonants (C), liquids (L), glides (G), and null ($). For W, vowels (V) and liquids (L) are indicated, and perhaps in some cases even continuant obstruents, assuming that expressions such as vocative "pssst" are to be analyzed as syllables also. We note parenthetically that glides (G) are suspect, since they appear to be functional variants of vowels, i.e. vowels functioning delimitively. This, however, is not a problem, since the use of J-rules to represent articulatory structures makes it just as feasible to consonantalize a vowel by rule as it is in the semantic component to nominalize a verb by rule. In short, the se of junction trees to represent articulatory structure brings a great deal of descriptive power to bear, should we need it.

Thus we supplant D and W with more descriptively specific node labels and append to them some element of their respective vocabularies as terminal units, as illustrated by Figure 7.

$$D = \{C, L, G, \$\}$$
$$W = \{V, L\}$$

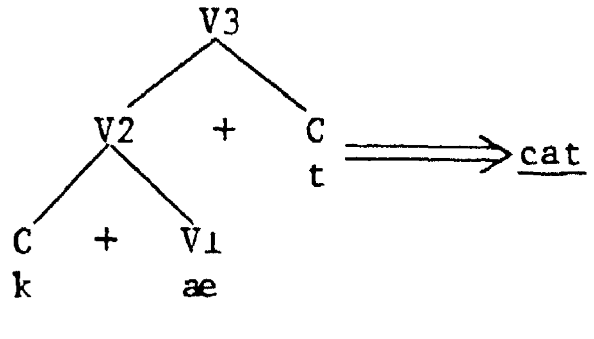


Figure 7.

The significance of V2 and V3 as non-terminal labels is that of semi-syllable and syllable, respectively. Bear in mind that the operation symbols appearing between operands are representative of the articulatory junctions (transitions) between them. Hence non-terminal nodes symbolize articulatory sequences consisting of the phonemes they dominate plus the transitions necessary to account for continuous movement from one distinctive vocal tract state to the next. This signifies, in effect, that given a junction instruction of the form $X \circ Y = Z$, there exists a transition $T = \circ(X,Y)$, such that $X^\frown T^\frown Y$ is a continuous articulatory sequence Z consisting of the distinctive units X and Y mediated by transitional T. This aspect of the formulation is advanced as an attempt to satisfy the need for phonological notation potentially capable of explicating both the discrete segmental elements of which the speech chain is composed, and the co-articulatory transitions which connect them in live speech. The practical effect of the formulation is that one's attention is drawn not to a relatively limited set of radical phonological changes, but to the co-articulatory effect of every junction on its operands, regardless of its subtlety. This is important if high quality synthetic speech is to be achieved.

Delimiting Clusters

Both initial and final syllable delimiters frequently consist of clusters of segments rather than discrete segments. An analysis of such clusters shows that notable assimilative forces are involved. We view this as a form of articulatory subordination, and, consequently, use subjunction as the basic junction type for treating such clusters. The fact that articulation trees are capable of showing a variety of compositional arrangements makes it possible to give whatever internal structure for such clusters as seems to be operative. Thus for *strand*, where *tr* seem to be more closely associated than *st*, this can be explicitly represented. (See Figure 8.)

Articulation tree for strand

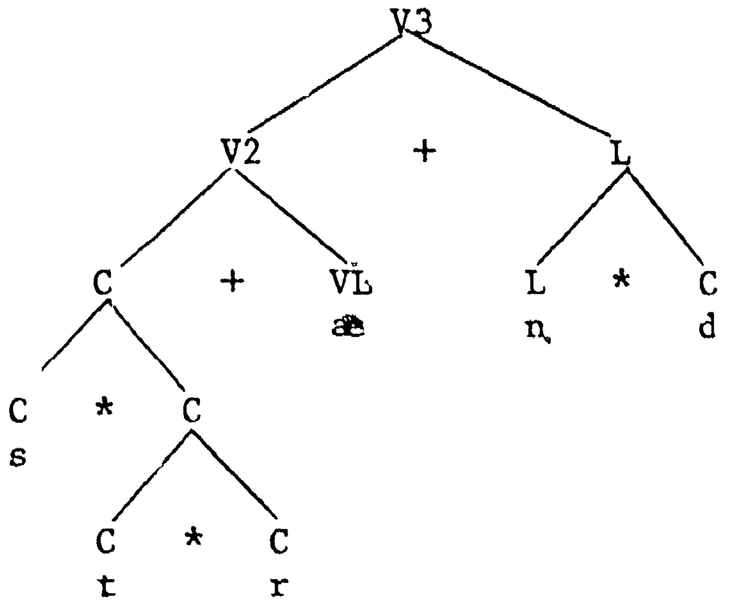


Figure 8.

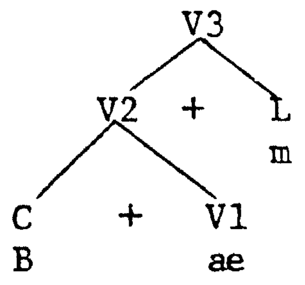
Multi-syllable Words

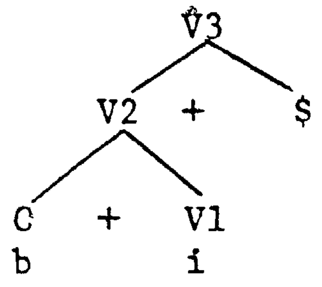
Let us now consider how multi-syllable words may be given in the form of articulation trees. The procedure, briefly, is as follows, using Bambi and Donna as the words to be diagrammed:

(1) The syllables are identified. BAM-BI [bæm - bi]

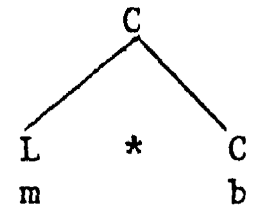
DON-NA [da - nə]

(2) The syllables are diagrammed using the appropriate adjunction type.

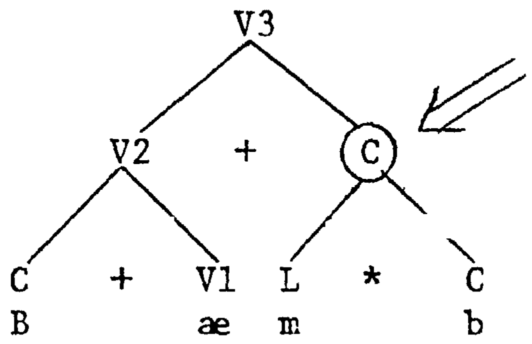




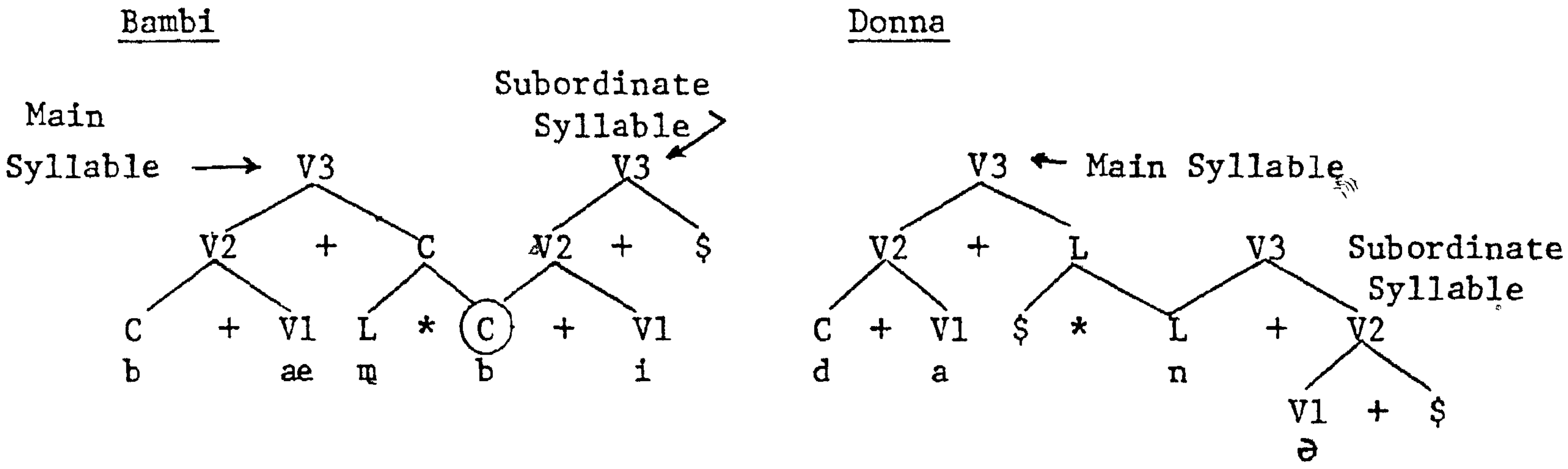
(3) An interjunction is constructed using syllable-final and syllable-initial constituents. (The label node is given as C since b seems to exert assimilative force over m.)



(4) The label node of the subjunction attaches to the more heavily-stressed syllable.



(5) The initial delimiter of the more weakly-stressed syllable becomes the intersect node.



An interesting result of the notation is that stress is no longer a property of vowels, but of entire syllables, i.e. the delimiters and the vowel. Further, stress reflects a relation between constituents, so that no features expressing stress values are necessary.

Phrases

Phrases are diagrammed by introducing prosodic constituents (H) to which word-trees are subordinated. (Refer to Figure 5.) The ranking syllable, i.e. the one receiving primary stress, joins to the prosodic constituent. The notation is intended to reflect the simultaneous execution of segmental and supra-segmental units during the articulatory process, in a way comparable to the multitudinous internal manipulations of an engine as one turns a crank. The crank of the articulatory apparatus is the diaphragm and other musculature which provide energy and assume other symbol cally significant states at certain intervals during the execution of the segmentals. Prosodic constituents result in the specific intonational contours we hear superimposed over syllables, words, and phrases.

While both segmental and supra-segmental constituents are coded in the context of semantic data, we emphasize again that A-trees contain only articulatory data. Thus, if A-trees are compared to the customary representations of generative phonology, as typified by those given by Chomsky and Halle[9] (compare Figures 5 and 9), it will be noted that the syntacto-semantic superstructure of the regular trees are replaced by an articulatory superstructure in the A-trees. The rationale for this departure from standard practice is not only motivated by the requirement imposed by the theory (that data types not be intermingled), but also by the observation that the regular trees tend to neglect prosodic articulatory phenomena. When information relating to these phenomena is incorporated into articulation trees, it replaces the usual superstructure of S's, NP's, and other similar lables in a natural way. The prosodic constituents thus introduced are comparable in their function to the intonation contours associated by rule with segmental sequences in the system proposed by Leben.[10]

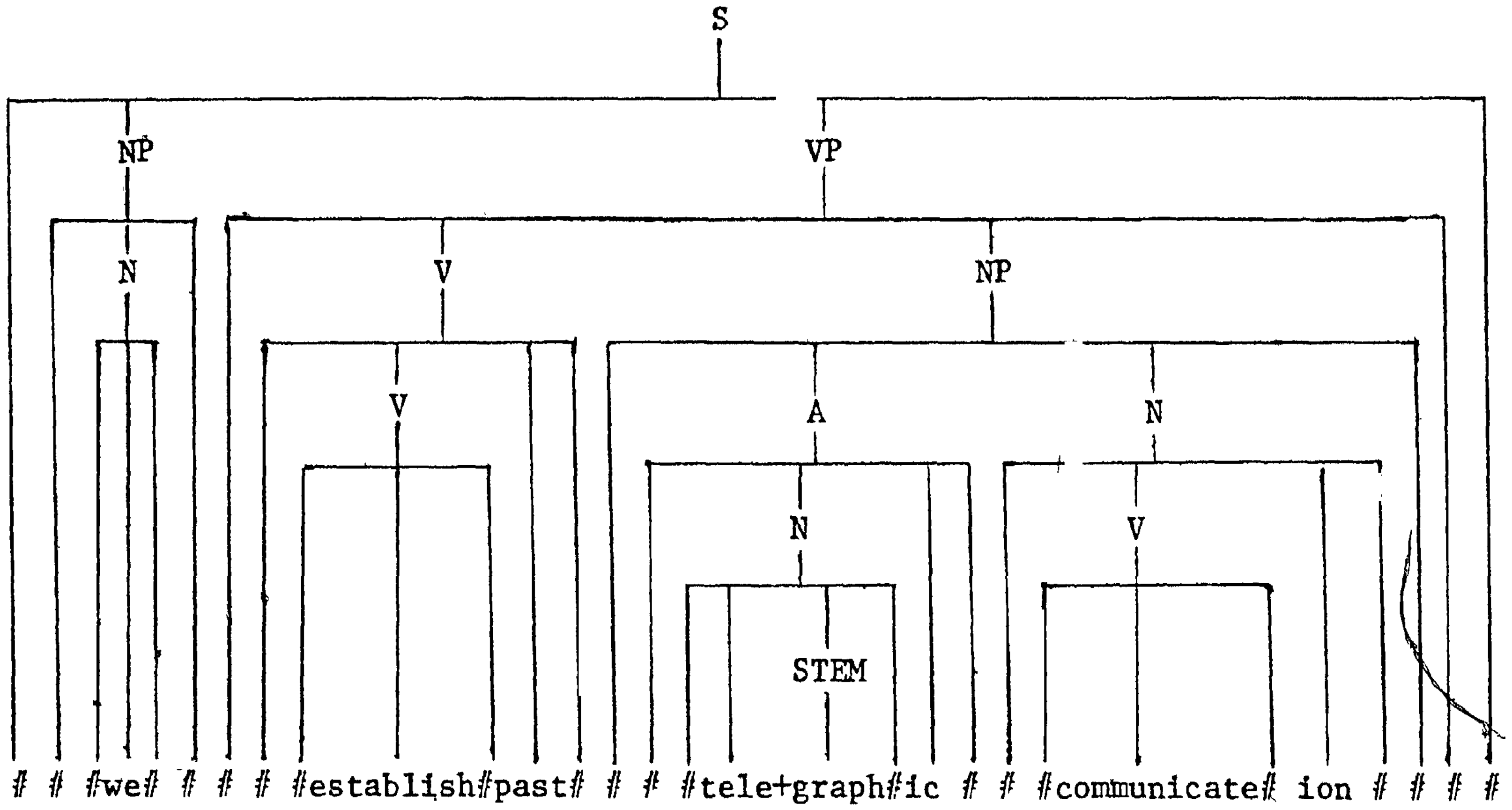

S
NP
VP
N
V
NP
V
A
N
N
V
STEM
# #we# # # # #establish#past# # # #tele+graph#ic # # #communicate# ion # # #

Figure 9.

unctional Versus Categorial Informatioh

The proposed system of phonological description makes possible an interesting hypothesis regarding many of the features used in current descriptions. Specifically, if A-trees are in some sense a reflection of actual articulatory processes, then phonological representations which do not use trees will consist of an intermixture of functional and categorial lables (features). For example, if trees are used to represent the relations between subject, verb, and object, it is not necessary to label the subject as such or the object as such, since structural relations make these notions explicit. If trees were not used to represent sentence structure, however, functional labels would have to be used.

Similarly, it follows that if trees are an appropriate medium for phonological description, but have not been used, then functional and categorial information are intermingled in current descriptions. If this is true, then it should be possible to abstract functional information away (and consequently not write it in feature form) by elaborating A-tree notation.

While the proposed system is still in its infancy, so to speak, some interesting initial observations in this regard can be made at this time. First, major category features become node labels in a natural way, thus suggesting why the formal illusion exists that a change, for example, of [+cons] → [-cons] is equal in magnitude to a change of [+voice] → [-voice] Second, [±syllabic] ([±consonantal] and [±vocalic] are also used in some systems) are functional labels and need not be written if syllables are given as tree structures. Third, stress at the segmental level and unmarked pitch at the prosodic level become implicit in structure in terms of the rank of operands in articulatory subjunction and need not be specified by feature. While it is beyond the scope of this paper to elaborate this point further, it is without doubt the most interesting and provocative consequence of the research to date.

FOOTNOTES

[1] Peter F. MacNeilage, "Linguistic Units and Speech Production," an invited paper presented at the 85th meeting of the Acoustical Society of America, Boston, Massachusetts, April 13, 1973.

[2] Ibid., p. 10.

[3] Philip Lieberman, "Towards a Unified Phonetic Theory," Linguistic Inquiry, Vol. I, No. 3 (July, 1970), 307-322.

[4] Farid M. Onn, "Speech Chain as an Analysis-By-Synthesis Model; A Review," Studies in Linguistic Sciences, Vol, IV, No. 2 (Fall, 1974), 168.

[5] The initial exposition of junction theory appears in Eldon G. Lytle, A Grammar of Subordinate Structures in English, The Hague: Mouton and Co., 1974, and also in Eldon G. Lytle, "Structural Derivation in Russian," unpublished Ph.D. dissertation, University of Illinois (Champaign-Urbana), 1972. Additional articles on junction grammar in BYU Linguistics Symposium Proceedings, 1971-1976.

[6] Charles J. Fillmore, "The Case for Case," Universals in Linguistic Theory, ed. Emmon Bach and Robert T. Harms (Holt Rinehart, 1968), pp. 1-89.

[7] The term phoneme is here used in reference to an articulation unit, not an acoustic unit.

[8] Fredericka Bell-Berti and Katherine S. Harris, "Some Acoustic Measures of Anticipatory and Carryover Coarticulation." (A version of this paper under the title, "Coarticulation in VCV and CVC Utterances: Some EMG Data," was presented at the 89th meeting of the Acoustical Society of America, Austin, Texas, April 7-11, 1975).

[9] Noam Chomsky and Morris Halle, The Sound Pattern of English, New York: Harper and Row, 1968.

[10] William R. Leben, "The Tones of English Intonation," to appear in Linguistic Analysis, 2, 1976.