

## INTONATION UNITS AND PROMINENCES IN ENGLISH NATURAL DISCOURSE

*Wallace Chafe*

Department of Linguistics  
University of California  
Santa Barbara, CA 93106

### ABSTRACT

This paper emphasizes the need to understand prosody with relation to all of language functioning, and to take account of all observable properties of sound, including the basic dimensions of frequency, intensity, and duration, but also various derived properties such as tempo and voice quality, while at the same time attending to both the physical and perceptual manifestations of these properties. The discussion focuses (1) on the identification of intonation units, which are seen as reflecting a universal inability of consciousness to focus on more than a small amount of information at a time, and (2) on the complexly interrelated acoustic and perceptual manifestations of prominences within intonation units, as well as the functions of these prominences.

### INTRODUCTION

One's understanding of prosody is inevitably colored by the direction from which one approaches it. A great deal of the recent literature seems to assume without question that prosody is a part or extension of a "phonological component" that is attached to a "syntactic component" within a generative model of language. For one who finds the generative approach to be fundamentally flawed it is difficult to relate to the literature that assumes it. In what follows I will be coming at prosody from a different direction, while continuing to hope for an eventual reconciliation of any and all approaches within a more complete picture of what prosody is and how it functions.

For many years I have listened to and attempted to transcribe and analyze tape-recorded conversations, stories, and rituals in several languages -- most extensively in English and two American Indian languages, Seneca and Caddo. This activity belongs to a long and honorable tradition in which prosodic concerns have always had an important place, even if they have not always been clearly articulated. I was taught from the beginning to pay close attention to accents, pauses, and pitch contours that seemed to contribute to the form and function of language. Most

of these observations were made "by ear", although on various occasions I consulted spectrograms that helped to clarify both segmental and prosodic phenomena. Lately I have been more than a little grateful for the ability to observe fundamental frequency, intensity, and duration more easily than was possible in the days when spectrograms were all we had.

It was always clear, even with respect to segmental phenomena, that visual displays of acoustic data never provided all the answers; that there was no one-to-one relation between physical sound and the way we perceive it; that our auditory apparatus and our brains do not simply record sound, but interpret it. That is certainly as much the case with prosody as it is with vowels and consonants. Furthermore, it is plausible to suppose that what is ultimately functional in language is what we perceive, and not the physical sound. It follows that if transcriptions and analyses are intended to capture what is functionally relevant, they need to give precedence to perceptual observations. But there is an interesting problem here, stemming from the fact that it is not always a simple matter to become consciously aware of what it is that we perceive. That fact becomes obvious as soon as one spends a few minutes with students who have trouble distinguishing rising from falling pitches, even though they know there is some kind of difference. The ability to pinpoint the nature of perceived prosody appears to be a skill, as much as the ability to perform phonetic transcription of any kind, or the ability to transcribe music in a musical dictation class.

Visual displays of sound can be a great help in this situation, because they make it possible to "see" whether a pitch goes up or down, whether one syllable is louder or longer than another, the precise length of a pause, the location of a breath, the beginning and end of overlapping speech, and so on. If these representations do not tell us what we perceive, at least

they show what enters our ears. One of the major current needs in prosodic research is to establish the relation between acoustic prosody and perceptual prosody. The more we know about this, the better we will be able to make sense of acoustic data. Even now, however, acoustic displays provide an invaluable way to sharpen our understanding of what we hear.

One of my major interests for some time has been to relate the flow of language to the flow of conscious experience. It seems obvious that both our thoughts and the language that organizes and expresses them are in constant flux, and that changes in the one are inseparably linked to changes in the other. Within this picture there is a crucial place for the kind of linguistic unit my colleagues and I have been calling an *intonation unit*. Intonation units emerge with relatively satisfying clarity from natural speech. I say "relatively" because, while their boundaries are easy to determine in the majority of cases, there remain other cases where the evidence for them is less than overwhelming. One of my aims here is to review the kinds of evidence that can lead us to identify the boundaries of intonation units, even in the more difficult cases.

I believe that the importance of intonation units stems above all from the hypothesis that each such unit expresses the information on which a speaker is focusing his or her consciousness at the moment the intonation unit is being produced, information the speaker hopes to introduce into the consciousness of the listener. From a careful study of intonation units, furthermore, it appears that each one expresses no more than one new idea. There is no room here to explore the ramifications of this "one new idea hypothesis" in detail, but attempts to clarify the meanings of "one", "new", and "idea" have shed interesting light on several important aspects of conceptualization and language. Here I can only allude to the usefulness of intonation units as catalysts to a variety of related discoveries.[1][2]

It appears that intonation units are only one of several levels that are defined by prosodic aspects of natural speech. More inclusive units delimited by pauses, inhalations, declining levels of pitch and loudness, shifts to and from higher or lower baselines of pitch and loudness, various changes in voice quality, and probably other criteria may all have functional significance of

some kind as well. Here we will notice some of these other units in passing, but the discussion will be centered on intonation units.

Within intonation units, some words or syllables are more prominent than others. Prominence may be a matter of significantly higher or lower pitch, lengthening, loudness, or some combination of these properties. Among other things, prominence signals "non-given" information -- information the speaker judges not to be in the focus of the listener's active consciousness at the moment. I will point to examples of this function, along with a few others.

Besides delimiting separate foci of consciousness and signaling properties such as non-givenness, prosody also conveys what might in a most general way be called *attitudes*. An obvious case is the distinction between the high rising terminal contour used in English for the expression of yes-no questions, in contrast to the falling contour used for assertions and question-word questions. But there are a variety of other contours that express a variety of other attitudes, most of which are more difficult to characterize than those just mentioned. This aspect of prosody seems particularly resistant to systematic understanding, and it is another area where future research can be especially rewarding.

## INTONATION UNITS

Observations of natural speech confirm the ubiquitous presence of intonation units, as well as their association with minimal chunks of information. The general picture is one in which a speaker focuses on the idea of an event or state or isolated referent, or sometimes on nothing more than an attitude or a connection between successive ideas. The speaker may then express that idea, attitude, or connection in an intonation unit that may be clearly defined by a variety of prosodic criteria or, in cases where the idea is especially closely related to an adjacent idea, by only one or two criteria. But intonation units are always bounded prosodically in some respect. If they were not, we could not say that the flow of separate ideas through consciousness is consistently reflected in the flow of prosody.

The features that characterize intonation units may relate to duration (perceived in terms of tempo and

lengthening), fundamental frequency (perceived as pitch), intensity (perceived as loudness), voice quality, the alternation between silence and vocalization, and/or change of turn. Examples of these features will be given here in two formats: the transcriptions that are cited in the text, and the displays of acoustic data that are available in Figures 1-5 at the end. The transcription conventions are largely, but not entirely those set forth in [3]. The figures were produced by the Summer Institute of Linguistics' CECIL system. It would be especially useful to hear the examples as well, but within the wholly visual context of a written article the transcriptions and figures will have to suffice.

Figure 1 shows a well-defined intonation unit whose boundaries are confirmed in a variety of ways. I will transcribe it as follows, using conventions that will be explained as we proceed:

- (1)     a         .. and so the **háll** is réal ló=ng%.  
           b         ...(36) [next intonation unit]

Preceding the vocalization in (1)a is a very brief pause of about .07sec. We have been transcribing pauses of less than .10sec simply with two dots. Following (1)a is a much longer pause of .36sec, transcribed with three dots followed in parentheses by a measurement of its length. (An accuracy to hundredths of a second seems appropriate for such measurements.) By convention, boundary pauses are shown at the beginning of each intonation unit. Among other things, then, (1)a is set off by pauses.

One of the major cues to intonation unit boundaries is tempo, captured to at least some degree in the notion of "anacrusis".[4] Intonation unit (1)a begins with a sequence of three accelerated syllables (*and so the*) occupying roughly .10sec each. I have transcribed accelerated syllables in smaller type. After that there are two words (*hall* and *real*, separated by a rapid *is*) whose duration lies in the range from about .20 to .30sec, a normal length for one-syllable words. The intonation unit ends with a word of extended length (*long*), occupying .43sec, as shown with the equals sign. This pattern of *acceleration-deceleration*, proceeding from reduced length syllables up to about .15sec, through normal length syllables from about .15sec to .35sec, to extended length syllables longer than .35sec, is characteristic of many intonation units, and may in

some instances be the best evidence for an intonation unit boundary.

When it comes to pitch, it happens that (1)a coincides with a "declination unit".[5] There are three words with high pitch (*hall*, *real*, and *long*), with a decline in the pitch of each (maxima of 299Hz, 211Hz, and 192Hz respectively), a good illustration of what is often called downstep. (I will return below to the unusually high pitch on *hall*.) It is noteworthy that the mid pitches of the accelerated first three words of this intonation unit are equal to or even higher than the high pitch on the last word. In addition to the declining pitch throughout, the end of the intonation unit shows a falling terminal contour, transcribed with a period.

One of the prosodic features in the category of voice quality is creaky voice, laryngealization, or vocal fry. It is conspicuous here at the end of the word *long*, where it is transcribed with a percent sign. Intonation units often end and sometimes begin with creaky voice, which thus provides still another clue to their boundaries.

In short, the identification of 1(a) as a coherent intonation unit is supported by a convergence of (1) the pauses preceding and following it, (2) the tempo of acceleration-deceleration, (3) the decline in pitch level, (4) the falling pitch contour at the end, and (5) the creaky voice at the end.

Intonation units are not always this well defined, and the example that follows was chosen in part to illustrate some less clear cases. It consists of eleven intonation units distributed within three breath units (a-f, g-h, and j) that are bounded by inhalation pauses lasting between .40sec and .50sec. The sequence belongs almost entirely to Speaker A, who does all the talking except for two brief responses from Speaker B (in i and k). Speaker A is relating her experiences in helping a friend move from one apartment to another, and here she is listing some of the factors that contributed to the confusion surrounding the move:

- (2)     a(A)     ...(h).(42) **Plus** the twó **dò=gs-**  
           b(A)     and the **cà=t-**  
           c(A)     %and the **kids=**,  
           d(A)     .. and the **s=créaming-**  
           e(A)     and **drópping** things,

f(A) <LO <P and so it was a real-  
g(A) ... (h) (.40) it was a mess=.  
h(A) it was [a] real mess=. P > LO >  
i(B) [Huh.]  
j(A) ... (h) (.49) So it took us a long time=.  
k(B) Yeah.

Figure 2 shows intonation units (2)a-c. After the inhalation pause, (2)a begins with a loud connective *plus* followed by a noun phrase (*the two dogs*) whose syllables become increasingly long up to the final word *dogs*, which lasts .50sec. That word is spoken with a low and level terminal pitch contour that seems to mean here something like 'this was still one more thing (that caused confusion)'. I have shown the level contour with the final hyphen (see below for the use of the grave accent mark on *dogs*). Intonation unit (2)b begins with two accelerated syllables (*and the*) followed by the lengthened word *cat*, which shows the same low and level contour as (2)a. Intonation unit (2)c has the same accelerated-decelerated tempo, with the accelerated words *and the* followed by a lengthened *kids*, where much of the length is in the final sibilant. The terminal contour of (2)c is different from that of (2)a and (2)b, consisting of a fall-rise (shown with the comma), a more standard list intonation that assigns the kids to a different category from the dogs and the cat. An additional boundary feature in (2)c is the creaky voice on its initial word *and*.

Figure 3 shows intonation units (2)d-f. After a brief pause, (2)d continues the acceleration-deceleration pattern, with extended length focused on the initial consonant and first syllable of the word *screaming*. However, the final syllable of *screaming* is long enough to accommodate the terminal pitch contour, which represents a return to the low level pattern of (2)a and (2)b. Without the duration and final pitch pattern of the final syllable of (2)d, (2)d and (2)e together would probably be perceived as a single intonation unit. Intonation unit (2)e follows the same general pattern, though the tempo differences are not as great. What seems to be significant in (2)e is the fact that the two final unaccented syllables (*ping things*) are as long as the accented syllable (*drop*). As a result, the final syllable has room to accommodate the same fall-rise terminal contour that we saw in (2)c. Intonation unit (2)f begins with five accelerated syllables (*and so it was a*), followed by a single normal-length syllable

(*real*), which carries a level terminal contour. The beginning of (2)f is signaled further by a drop to a low pitch level (whose extent is shown by "LO" in angle brackets) and low volume (shown by "P" for *piano* in angle brackets), both of which continue until the end of (2)h. The reduced vigor of (2)f is also manifested in the creaky voice on the words *it was*. With this attempt at a final summing up, the speaker was approaching the end of a larger, more sentence-like unit. It is worth noting, however, that even with allowance for the truncated nature of (2)f, the syntax of (2)a-f (or even of (2)a-h) is not what would ordinarily be expected of a sentence.

Figure 4 shows intonation units (2)g-h, which begin with an inhalation pause. The low pitch and volume levels introduced in (2)f are continued. In both of these intonation units everything is accelerated except for the extended final word (*mess* in both). Both show a falling terminal contour. Both begin and end with creaky voice. Intonation unit (2)i is a backchannel *huh* by Speaker B, coinciding with Speaker A's indefinite article in (2)h.

Figure 5 shows intonation units (2)j-k, which also begin with an inhalation pause. After a normal length *so*, (2)j contains four accelerated syllables followed by a normal-length *long* and then an extended-length *time*, which is creaky at the end. It ends with the same falling pitch contour as (2)g and (2)h. It is immediately followed by (2)i, consisting of Speaker B's *yeah*, whose much lower pitch reflects the fact that Speaker B was a man.

In short, one of the most consistent features of these intonation units is the pattern of acceleration-deceleration. Each intonation unit also ends in an identifiable terminal pitch contour. Several end and/or begin with creaky voice. The insertion of an inhalation pause sets off the beginnings of a, g, and j, and the ends of f and h. A low level of pitch and loudness sets off f-h from e and j. Finally, the backchannel response in k also serves to terminate j. The entire sequence demonstrates well how the boundaries of intonation units are identifiable on multiple grounds, and further that while some boundaries are signaled in multiple ways, others are signaled by only one or two of the features discussed.

The division of (2) into intonation units illustrates well

the restriction of intonation units to no more than one new idea, particularly with the list of confusions in (2)a-e, each of which was a new idea and each of which apparently had to be expressed in a separate intonation unit for that reason. The longer intonation unit in (2)j shows how a separate verb (*took*) and object (*a long time*) may combine to express a single lexicalized idea (*took a long time*) without violating the one new idea constraint. Lexicalization is one of several important aspects of language and thought on which the one new idea hypothesis sheds useful light.

It is worth noting that the length of intonation units, as most easily measured by the number of words per intonation unit, is strongly limited by the one new idea constraint. In English, substantive intonation units (excluding backchannel responses and other units that serve only to regulate the flow of information) have a rather sharply defined modal length of four words. Other prosodically defined segments of discourse -- breath units, declination units, prosodic sentences defined by final falling pitch, etc. -- are more variable in length and content. They appear to be less dependent on any wired-in constraint, but rather to involve various kinds and degrees of coherence between the minimal foci of consciousness that are verbalized in intonation units. For example, of the three breath units in (2), the first is centered on listing the confusions:

- (2) a(A) ... (h) (.42) Plus the twó dð=gs-  
 b(A) and the cə=t-  
 c(A) %and the Kids=,  
 d(A) .. and the s=créaming-  
 e(A) and drópping things,  
 f(A) <LO <P and sð %it %was a real-

the second succeeds in articulating the evaluation that was too hastily attempted in (2)f:

- (2) g(A) ... (h) (.40) %it was a %mész=.  
 h(A) %it %was [%a] real %mész=. P> LO>

and the third focuses briefly on another aspect of the total experience:

- (2) j(A) ... (h) (.49) So it took us a long time%=.

It may be noted that the first and second of these breath units together constitute a single prosodic

sentence -- a coherent depiction and summarization of the experience. Intonation units, then, reflect a strict and unavoidable constraint on the flow of discourse, and ultimately on the flow of consciousness. Larger prosodic units, defined in other ways, are more variably determined by the structure of what is being talked about.

## PROMINENCES

Besides perceiving speech as segmented into intonation units, we perceive certain elements within an intonation unit as more prominent than others. The acoustic correlates of prominence are also complex and variable. There are degrees of prominence, and there are several ways in which prominence may be realized. In what follows I will use the term *accent* for prominences that are realized as pitch deviations from a mid or neutral baseline, usually a higher pitch but sometimes a lower one. In the transcriptions I have represented such pitch deviations with accent marks: acute for a significantly higher pitch and grave for a significantly lower one. When one of these accented elements is also either lengthened or loud (or both), I will say that it has a primary accent. A pitch deviation alone, without lengthening or loudness, will be called a secondary accent. Of course an element may be either lengthened or loud without showing a pitch deviation, and in such cases I will say simply that it is lengthened or loud, but not accented.

As a first illustration of prominences, we can look again at the intonation unit that was cited in (1)a, repeated here (see Figure 1).

- (1) a .. and so the háll is réal ló=ng%.

Three of these words -- *hall*, *real*, and *long* -- are accented, all showing high pitch (with downstep). However, not only is the pitch of *hall* inordinately higher than that of the other two words, to an extent that is unexplainable on the basis of downstep alone, but *hall* is also significantly louder, as shown by the boldface type. Its greater prominence in both respects can be attributed to the fact that it expresses contrastive information. It is not new information -- the idea of this hall was introduced eight intonation units earlier -- but here the hall is being contrasted with the living room, the bedroom, and the bathroom, all of which had been introduced in the meantime. It

is not unusual for contrastive elements to show exaggerated loudness and/or pitch.

The new (previously unactivated) information in (1)a is expressed by the predicate *is real long*, in which the heaviest load is carried by the word *long*, which is both high pitched and lengthened. (It is not accidental that *long* also occurs at the end of this intonation unit.) The intensifier *real* is high pitched but neither loud nor lengthened. Thus, we find three different manifestations of prominence in this intonation unit: the high pitched and loud *hall* expressing contrastive information; the high pitched, lengthened, and final *long* expressing new information; and the high pitched *real* intensifying the meaning of *long*. We can say that *hall* and *long* have primary accents (high pitch along with loudness or lengthening), as is typical for both new and contrastive information, and that *real* has a secondary accent (high pitch alone), as is often the case with modifiers of various kinds.

Turning now to sample (2) and Figure 2, we find the complexity of prominence well illustrated in its first intonation unit:

(2) a(A) ... (h) (.42) **Plus** the two **ðð=gs-**

The words *plus*, *two*, and *dogs* all have some kind of prominence in terms of duration, pitch, and/or loudness: *plus* is loud, *two* is high-pitched, and *dogs* is loud, low-pitched, and lengthened. Functionally, it would appear that *plus* is loud (but nothing else) because it serves as a connective that introduces a new breath unit and a new subtopic in the conversation. The word *dogs* has a multiply marked primary accent because it expresses new information. The word *two* has a high pitched secondary accent because it modifies the idea of the dogs, much as *real* modifies the idea of being long in (1)a. The fact that *dogs* is low-pitched rather than high-pitched may be attributed to the terminal contour, not the accenting per se. In other words, accented elements are typically high pitched, but may be low pitched when a terminal contour demands it.

The same pattern of accents, minus the numeral, is observable in intonation units (2)b-d (Figures 2 and 3). Something like this pattern is also present in (2)e, where the primary accent is on the categorization of the event as an instance of *dropping*, followed by a

normal length but unaccented generalized object *things*:

(2) b(A) and the **cà=t-**  
c(A) %and the **kids=**,  
d(A) .. and the **s=créaming-**  
e(A) and **drópping** things,

In both (2)f and (2)j (Figures 3 and 4) it is interesting that the connective *so* is loud but lacking in pitch prominence, thus mirroring the loudness of *plus* in (2)a:

(2) a(A) ... (h) (.42) **Plus** the two **ðð=gs-**

(2) f(A) <LO <P and **só** it %was a real-

(2) j(A) ... (h) (.49) **So** it took us a long %time=.

Intonation unit (2)g (Figure 4) shows a primary accent on the word *mess*, expressing in this case *accessible* (or reactivated) information, since the characterization of the move as a mess had already been established earlier in the conversation. *Mess* shows both pitch and durational prominence, but is not loud, presumably because the entire intonation unit is spoken with reduced volume:

(2) g(A) ... (h) (.40) %it was a %**méss=**.

Intonation unit (2)h reinforces (2)g by repeating it with the addition of the word *real*, which was of course already foreshadowed in the truncated (2)f:

(2) h(A) %it %was [%a] real %**méss=**. P > LO >

Intonation unit (2)j (Figure 5) communicates a new idea that is captured in the lexicalized phrase *take (someone) a long time*. As is typical of such phrases, the primary accent falls on the final content word (*time*), but there is a high pitch (with no other sign of prominence) on the other content word (*took*), which is part of the accelerated buildup. Thus, *took* can be said to have a secondary accent:

(2) j(A) ... (h) (.49) **So** it took us a long time%**=**.

In short, all of the substantive intonation units in this sample show primary accents at or near their conclusions. Primary accents express non-given (new



or accessible) or contrastive information, and contrastive elements may show exaggerated pitch deviation as well as loudness. This pattern fits the British view that tone groups (or whatever they may be called) build up to and trail away from a nuclear accent. On the other hand, the presence of two clear, if differently motivated primary accents in (1)a shows that an intonation unit can contain more than one nucleus. Some intonation units exhibit (pitch only) secondary accents on subsidiary content words (*real*, *two*, *took*). Introductory connectives like *plus* and *so* may be loud but unaccented in terms of pitch.

## SUMMARY

It is probably the case that all languages are produced in the format of a succession of intonation units, each expressing no more than one new idea. These intonation units are produced with a tempo of acceleration followed by deceleration, with a set of intonation-unit-final pitch contours whose specific manifestations vary with the language, and less consistently with pausing, changes in the level of pitch and/or loudness, changes in voice quality, and sometimes a change in speaker.

Prominence is more of a mixed bag. In English, high (or sometimes low) pitch combined with extended length and/or loudness is regularly associated with the expression of non-given information as well as contrastiveness, the latter sometimes showing exaggerated pitch height. English has various intensifiers, enumerators, and low content verbs (like *real*, *two*, and *took*) that typically show high pitch alone -- what I am calling a secondary accent. Introductory connectives like *plus* and *so* may be prominent with respect to loudness alone. High pitch, loudness, and lengthening are also associated with the expression of affect. Prominence, in short, has a variety of forms and functions.

## REFERENCES

- [1] W. Chafe: "Cognitive constraints on information flow", in R. Tomlin (ed.), *Coherence and Grounding in Discourse*, pp. 21-51. Amsterdam: John Benjamins (1987)
- [2] W. Chafe: *Discourse, Consciousness, and Time: The Flow and Displacement of Conscious Experience in Speaking and Writing*. (to be published)
- [3] J. Du Bois, S. Cumming and S. Schuetze-Coburn: *Discourse Transcription* (to be published)

- [4] A. Cruttenden: *Intonation*. Cambridge University Press (1986)
- [5] S. Schuetze-Coburn, M. Shapley, and E.G. Weber: "Units of intonation in discourse, a comparison of acoustic and auditory analyses", *Language and Speech* 34, pp. 207-234 (1991)

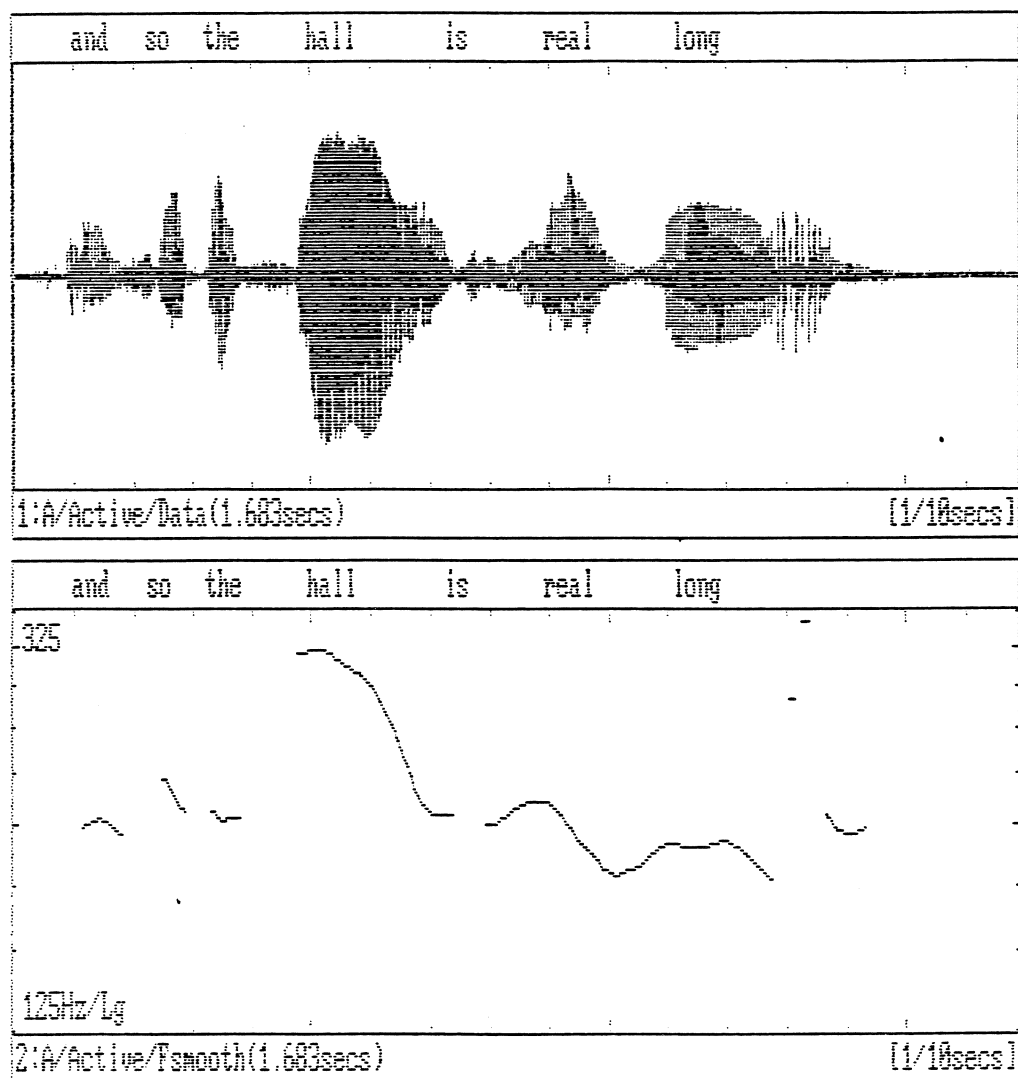


Figure 1

(1)a .. and so the **háll** is réal ló=ng%.



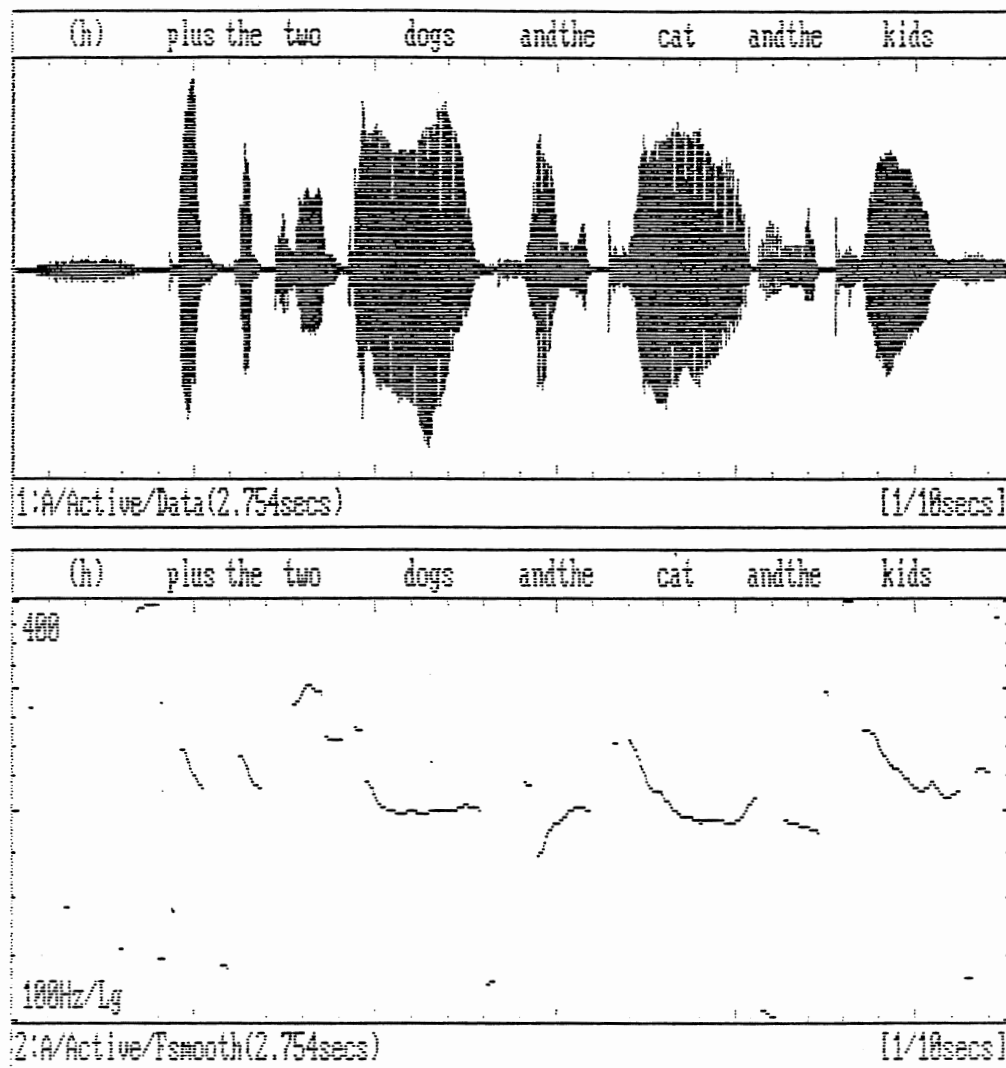


Figure 2

- (2) a(A) ... (h) (.42) Plus the two dð=gs-  
 b(A) and the cà=t-  
 c(A) %and the kids=,

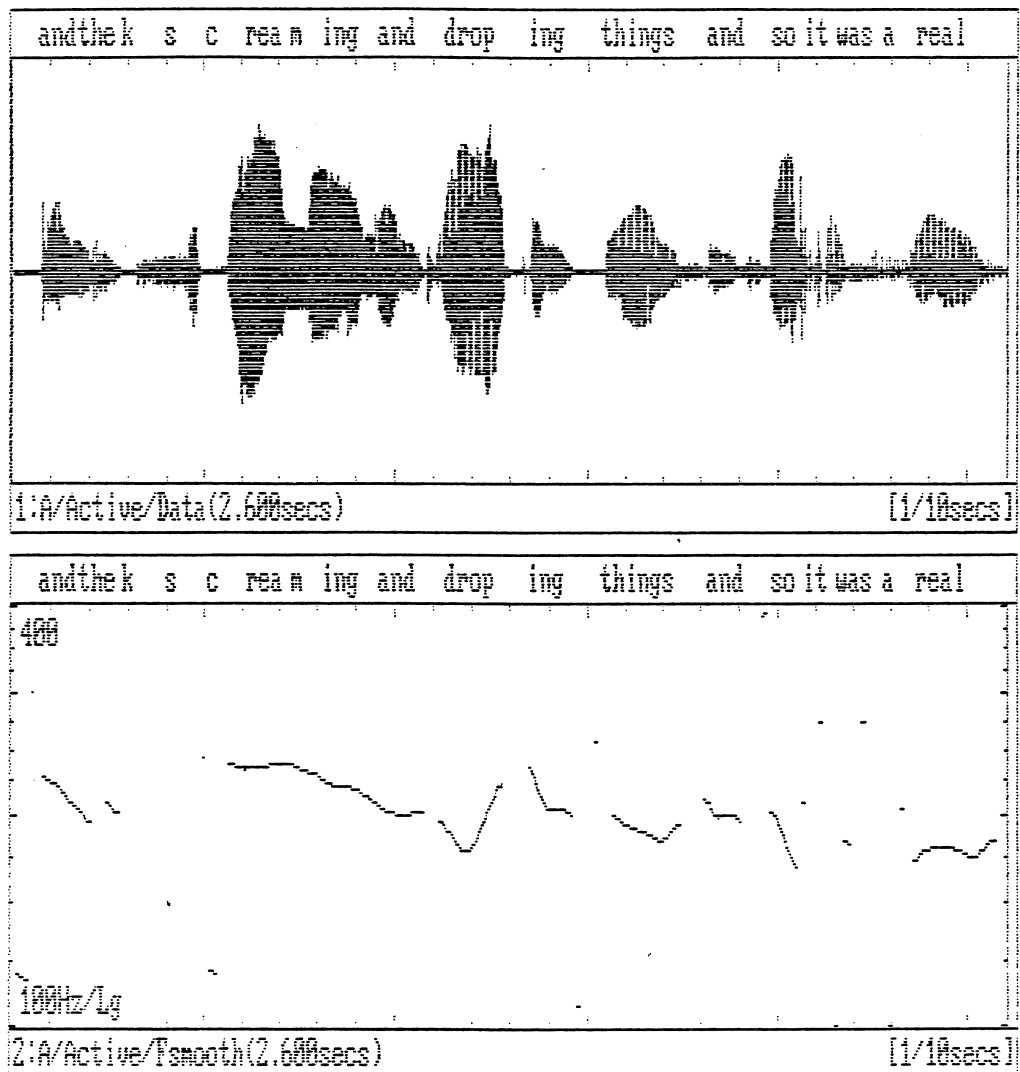


Figure 3

- (2) d(A) .. and the s=creaming-  
 e(A) and dropping things,  
 f(A) <LO <P and so %it %was a real, —

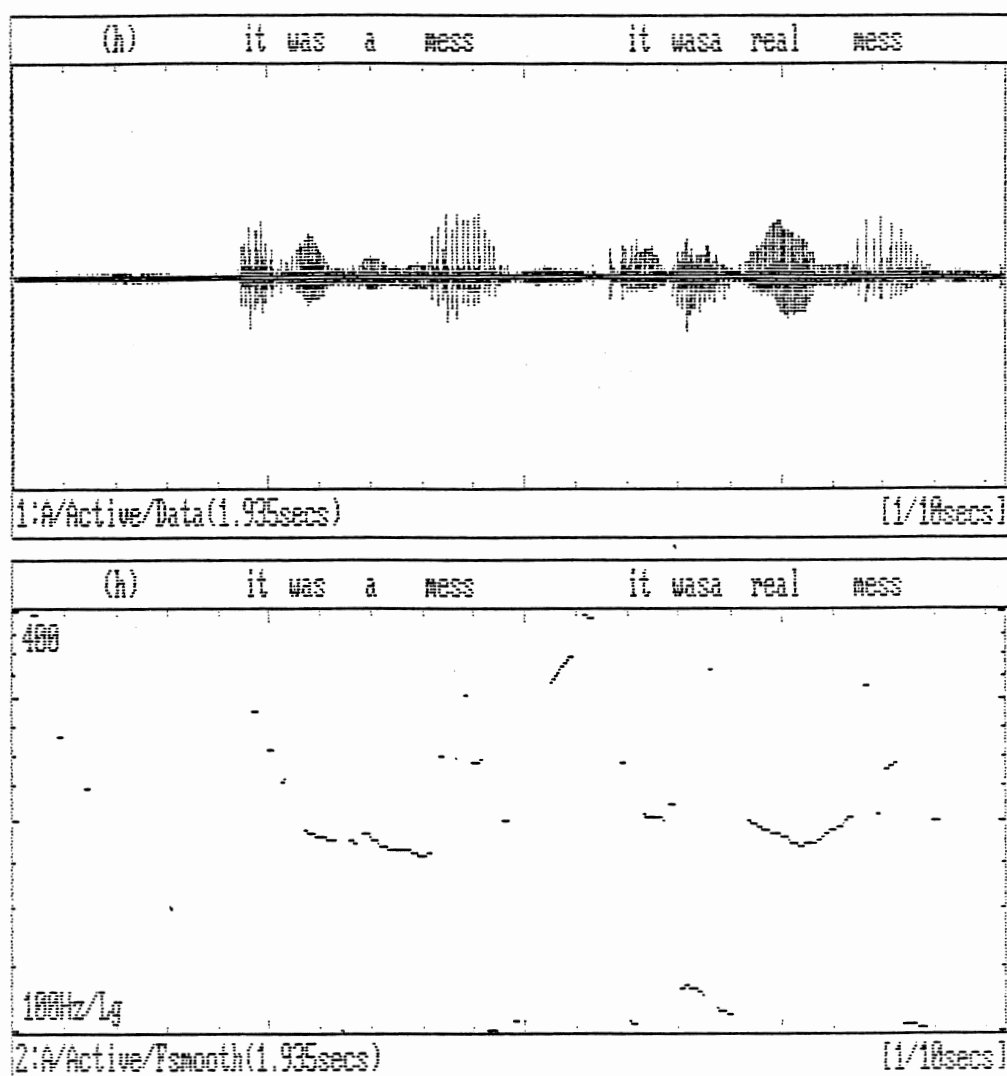


Figure 4

(2) g(A) ... (h) (.40) %it was a %m  ss=.  
h(A) %it %was [%a] real %m  ss=. P> LO>  
i(B) [Huh.]

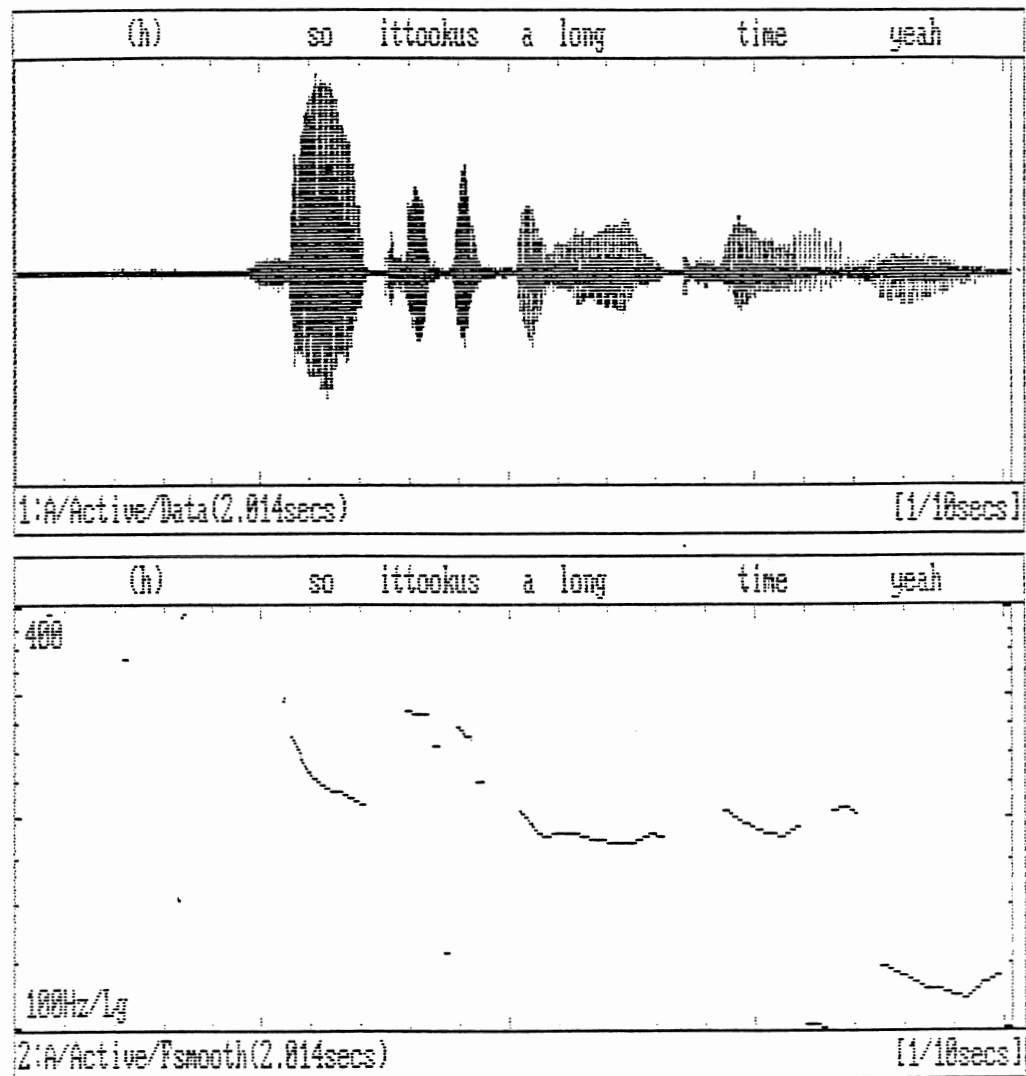


Figure 5

(2) j(A) ... (h) (.49) So it took us a long time% =.  
 k(B) Yeah.