Journal of Phonetics (1982) **10**, 163–175

Perception of sentence and paragraph boundaries in natural conversation

J. Kreiman

Department of Linguistics and Committee on Cognition and Communications, University of Chicago, 1050 *E* 59*th Street, Chicago, Illinois* 60637, *U.S.A.*

Received 12*th June* 1981

Abstract: Subjects were presented with two versions (with and without semantic information) of a natural conversation and in each condition were asked to respond when they heard a sentence or paragraph boundary. Presence of semantic information resulted in increased numbers of both very quick and very slow responses. Sentence boundaries are cued by non-level intonation contours, laryngealization, pre-boundary lengthening, and presence of a pause, all unit-terminal phenomena; cues for paragraph boundaries both precede and follow a change in topic, reflecting the optional nature of topic changes in natural conversations.

1. Introduction

Obviously, sentences and paragraphs are primarily semantic units. There is, however, no *a priori* reason why such units should not occur in dialogue and have characteristic suprasegmental shapes, which speakers reliably produce and listeners are able to identify. There are several different roles such characteristic shapes could conceivably play in conversations. They may consist entirely of unit-terminal cues which signal the fact that a sentence or paragraph is definitely over and a new unit is about to begin; they may consist of unit-initial cues signalling the beginning of a new unit and only implying the end of an old one; or they may consist of cues which span boundaries and indicate both that one unit has ended and that another has already begun. There is no particular reason to presuppose that only unit-terminal cues exist. Certainly it is as possible that cues provide information about boundaries after the fact, as it is that they signal what is to come; and, since the semantic content of utterances can presumably provide information about both ends and beginnings of units to conversants, there is no reason why either type of boundary cue should be ruled impossible of occurrence without evidence.

Lehiste, working under the assumption that connected speech possesses internal structure, has in a series of experiments (Lehiste, 1979*a*, *b*; Lehiste & Wang, 1977) investigated the nature of phonetic correlates of sentences and paragraphs in spoken English. All her experiments seem to have presupposed exclusively unit-terminal cues, though this is nowhere made explicit in her work. One way to investigate the phonetic correlates of semantic units is to eliminate semantic information from one experimental condition. In the first of her experiments (Lehiste & Wang, 1977) Lehiste achieved this through the use of simulated spectral inversion, a technique whereby the sign of every other sample of a digitized waveform is reversed. This process effectively destroys the segmental characteristics of a sample

0095–4470/82/020163+13$03.00/0

of speech, but leaves suprasegmental information more or less intact. Stimulus material consisted of 1 min of monologue extracted from a face-to-face interaction; subjects heard first an inverted, and then the normal, version of the tape, and were instructed to press a key once when they heard a sentence boundary, and twice when they heard a paragraph boundary. The stimulus presentations were preceded by two clicks; subject responses to these clicks established baseline reaction times which were used in determining the point where each boundary was perceived.

All of Lehiste's 24 subjects indicated sentence boundaries for both conditions in this experiment; 15 made paragraph responses in the inverted condition, and only nine heard paragraph boundaries in the normal condition. Subjects heard more boundaries of both kinds in the inverted condition than in the normal (265 to 258 for sentences, and 38 to 16 for paragraphs). Additionally, responses were not made in the same places in the two conditions; for sentences, the rate of mismatches between suprasegmental and semantic boundaries (the number of responses in one condition only, divided by the total number of responses) was nearly 40%.

Response latencies were much longer for normal than for inverted condition responses, as if availability of semantic information caused subjects to wait for the next sentence to begin before deciding whether the previous one was over. Subjects did better at agreeing about the location of boundaries in the inverted than in the normal condition for both sentences and paragraphs; overall levels of agreement were not high, however. The mean number of responses per boundary for sentences was 11.13 (46%) in the inverted condition, and 10.53 (44%) in the normal condition; for paragraph boundaries the mean number of responses was 3.16 (13%) in the inverted condition, and 2.66 (11%) in the normal condition. All 24 subjects agreed about the location of a sentence boundary only once in the inverted condition; no normal condition sentence boundary had more than 20 responses, and subjects who indicated paragraph boundaries at all were never unanimous in their responses.

From these results, Lehiste derives the following set of boundary cues. Low fundamental frequency and laryngealization were both strong indicators of sentence boundaries. A relatively long pause also suggested a boundary was present, but this cue could be overridden by frequency information. Short pauses preceded by low fundamental frequency or laryngealization were likely to be responded to, while long pauses without these additional cues received in general fewer responses than long pauses with these cues present. Paragraph boundaries were characterized by the same set of cues; from these data, it was not clear what distinguishes the two types of boundary, though subjects were clearly able to tell them apart.

The second of Lehiste's experiments (Lehiste, 1979*a*, *b*) attempted to determine which factors distinguished sentence from paragraph boundaries. Stimulus material for this experiment consisted of a 5 min uninterrupted response by one male speaker to a question by another—a "prompted monologue" extracted from a dialogue. Instead of using inverted speech in the modified condition, subjects heard a 70–200 Hz band-pass filtered version of the stimulus material. The design was otherwise identical to that of the previous experiment.

Detailed results will not be presented; however, Lehiste concludes:

> Three phonetic factors appear to interact in providing paragraph boundary cues: length of pause, presence of laryngealization, and preboundary lengthening. The clues may counterbalance each other. Thus even a fully phonated sentence termination may be interpreted as a paragraph boundary, if a sufficient amount of lengthening is present and if the terminal is followed by a relatively

> long pause. A paragraph boundary may be perceived before a short pause, if the sentence terminal is laryngealized and sufficiently lengthened. (1979*a* p. 107.)

There are a number of problems with both these experiments. The first is stimulus quality in the modified condition. Inverted speech bears more resemblance to space noises than it does to natural language. It has a thin, tinny quality; voices are barely distinguishable from one another; and intonation contours can be identified only if one already knows what they are. In no way can it be assumed that subjects will respond to inverted speech as they would to normal speech. Similarly, with band-pass filtered speech a very fine line exists between loss of speech-like quality and intelligibility; when long stretches of spontaneous speech are filtered this way, almost inevitably some semantic information will be present, or some stretches of speech will lack the acoustic information necessary for subjects to make the discriminations required of them.[1]

The second problem is with the conversations from which the stimulus materials were taken. In face-to-face interactions, conversants gather information from more than just the verbal channel. Gazes, gestures, nods, and so on, all carry information as to the ongoing state of the conversation (Duncan & Fiske, 1977), and none of this information is recoverable by subjects performing a task like Lehiste's. It is desirable in this sort of experiment that subjects have available to them all the information there is—and all that the original speakers had—about the state of the conversation at any given moment. This can be achieved by producing the stimulus tape under conditions where speakers cannot see each other.

Finally, it seems likely that having subjects do both the sentence and the paragraph tasks at once will cause them to do each more poorly than they might otherwise. Lehiste's results, especially for the paragraph task, are weak enough that one might question their significance, although she draws rather strong conclusions from them. If more reliable results cannot be obtained, either by separating the tasks or improving stimulus quality, her conclusions must be called into question.

In view of these problems, Lehiste's basic experiment was repeated, with the following variations. (1) The stimulus material was drawn from a natural, relaxed conversation, the participants in which were not in visual contact. (2) A better quality stimulus—a combination of inverted and low-pass filtered speech—was used in the modified condition. (3) Subjects made sentence and paragraph responses during separate stimulus presentations. The complete design is described below.

2. Method

Subjects

Eight subjects, five female and three male, participated in this experiment. All were native speakers of standard American English.

Stimulus tape

Two female native speakers of English, good friends both accustomed to being tape-recorded,

[1] If the speech sample used is in fact produced carefully and monotonically enough that band-pass filtering poses no problem, another difficulty arises: that of naturalness of the original speech sample. There is no guarantee that cues active in normal, relaxed conversation are present or useful in very careful styles of speech.

held a conversation under conditions simulating a telephone call. One speaker sat in a recording booth, the other outside the booth; they communicated via microphones and headphones, but were not in visual contact with one another. The speakers knew only that the tape was needed for an experiment; they were given no instructions other than to relax and converse normally.

The speakers were recorded on separate channels of a four-channel tape recorder. The entire conversation lasted approximately 45 min; from it, a 6 min section was chosen. This section was quite typical of conversations between these two speakers, and included topic changes by each speaker within speaking turns and at turn boundaries, one relatively long narration by each speaker, and a variety of turn and topic lengths. Only minimal amounts of laughter and simultaneous speech were present in this passage.

The modified condition stimulus tape consisted of a spectrally-inverted version of this 6 min tape combined with a 200 Hz low-pass filtered version of the same tape. Inversion was performed on a PDP-11/34 minicomputer. Input to the inversion routine was high-pass filtered at 600 Hz. The sampling frequency was 8000 Hz. Since every other sample was inverted, the inversion frequency was 4000 Hz; and output from the inversion routine was low-pass filtered at 2000 Hz. These values were derived by trial and error; they seem to be the best obtainable compromise between representing all the frequencies present in a speech wave and avoiding the tinny quality which results when high-energy low frequencies and low-energy high frequencies are exchanged by the inversion process. The composite tape was very speech-like, but quite incomprehensible; it sounded something like a cross between a conversation heard through a thick wall and one in a foreign language.

Apparatus

Subjects sat in a sound-proofed booth and heard the stimulus presentations binaurally through good quality headphones. They responded by pressing a telegraph key connected to a noise generator and a timer. The stimulus material and subject responses (50 ms noise bursts) were simultaneously re-recorded on separate channels of a two-channel tape recorder.

Procedure

Subjects heard each of the two versions of the stimulus tape—modified and normal—twice in succession. During the first presentation of each version, they responded (by pressing the telegraph key) whenever they heard a sentence boundary; during the second presentation of each version, they responded when they heard a paragraph boundary. The structure of the experiment was thus (1) modified stimulus, sentence task; (2) modified stimulus, paragraph task; (3) normal stimulus, sentence task; (4) normal stimulus, paragraph task. All subjects did the tasks in this order. Both modified-condition tasks had to precede both normal-condition tasks to control for semantic information; and pilot experiments suggested that the paragraph task, quite a difficult one for subjects, was somewhat easier when the stimulus material had been heard before (during the sentence task).

Subjects were encouraged to use their own intuitions about what sentence and paragraph boundaries in speech are, but were also given the following working definitions.

> A paragraph boundary is the end of a fairly complete unit of speech and the beginning of a new thought. The end of one speaker's turn in the conversation (i.e., the places where one person stops talking and the other begins) is a paragraph boundary only if it is also the end of a fairly complete unit of speech or the beginning of a new thought. If you feel the new speaker is continuing the thought of the previous speaker, do not signal a paragraph boundary.

> A sentence boundary is a place where one sentence ends or another begins. Since this is a real conversation, not all the utterances you hear will be complete grammatical sentences; there will be 'uhuh's and 'yeah's, and short or incomplete sentences. Don't be surprised by these. If you feel a new sentence has begun while the previous one is still incomplete, please respond to the beginning of the new sentence.

Before the first stimulus presentation, subjects heard five short tones and were asked to respond as quickly as possible to each of them. This served to familiarize them with the experimental apparatus and to establish baseline reaction times.

Data analysis

Response latencies were measured to an accuracy of ± 10 ms from two channel oscillograms of the response tapes. Sentence boundaries were derived independently from each subject's responses by the following procedure. First, responses which occurred in or shortly after a pause, and one reaction-time unit or more after the pause began, were assumed to have been cued by the end of the preceding utterance, and a boundary was assigned accordingly. The majority of responses in the sentence task (62% in the normal condition, and 51% in the inverted condition) were assigned in this manner. The acoustic characteristics of these boundaries—their intonational contours, presence or absence of laryngealization, and so on—were then measured as described below. Finally, responses not occurring in a pause were assigned to the nearest place in the discourse where one or more of these derived cues occurred. This procedure essentially served as a test of the hypothesis that the acoustic characteristics found just before pauses were in fact cues to sentence boundaries. When the closest "potential boundary"—the nearest place in the discourse where one or more of these acoustic characteristics was present—preceded a response by less than about 150 ms, earlier parts of the discourse were examined. If the next earliest potential boundary had also been responded to, or if it preceded the response by a second or more, the response was assigned to the later boundary. Responses were coded as having latencies of less than 100 ms only if the next potential boundary back already had a response assigned to it.

Paragraph boundaries were determined in much the same way except that, due to the smaller number of responses and lower agreement rates for this task, all subjects were considered as a group, rather than each subject's responses being considered independently of the others. Regions in the conversation where responses clustered were analysed acoustically, and the characteristics found were used in the assigning of boundaries where there was less agreement.

Intonation contours and laryngealization were measured from narrow-band spectrograms. Initially, intonation was coded as rising, falling, or level, and creaky voice as present or absent; later, the extent of rise or fall in intonation and the duration of the laryngealization at a given boundary relative to that at other boundaries were also examined. This additional information was used primarily to test hypotheses about differences between sentence and paragraph boundaries. Pauses were measured from an oscillogram of the master tape of the conversation.

Since pre-pausal lengthening is a very well-established effect in English (see for example Klatt, 1975), and since most boundaries in the discourse were preceded by clearly perceptible drawling, it was deemed unnecessary to meaure the duration of all segments in the discourse. Instead, pre-pausal lengthening was coded by ear in a yes-or-no manner for use in the analysis of sentence responses. Later, the duration of the last "syllable" of the last word before each boundary was measured (from the onset of the final vowel to the end of the word) on a computerized waveform editor with an oscilloscopic display and a movable cursor. These units were then grouped according to structure (e.g. vowel + fricative, vowel + stop + fricative, and so on), and durations within groups were compared. In this way, relative amount of pre-pausal lengthening present was determined; this information was used in the analyses of paragraph responses.

3. Results

"Number of responses" data

Contrary to Lehiste's result, no significant difference was found in the number of responses subjects made in the two conditions, either for the sentence task or for the paragraph task (for sentences, $t = 2.3109$, $\nu = 7$, $p > 0.05$; for paragraphs, $t = 0.5854$, $\nu = 7$, $p > 0.05$). However, subjects did not respond in the same places in the two conditions: There is a substantial mismatch between the sets of semantically- and suprasegmentally-determined boundaries. For the sentence tasks, approximately 50% of the boundaries indicated by one or more subjects were indicated in both conditions; 29% were indicated in the normal, but not in the inverted condition, and 21% were indicated only in the inverted condition. For paragraphs, only 31% of boundaries were indicated in both conditions; 38% were marked only in the normal condition, and 31% only in the inverted condition. These data are broken down by subject in Tables 1 and 2.

Table 1 **Sentence responses by subject**

Subject	Number of boundaries indicated in both conditions	Number of boundaries indicated in normal condition only	Number of boundaries indicated in inverted condition only	Total number of boundaries
1	60	38	20	118
2	71	27	27	125
3	25	33	24	82
4	79	39	48	166
5	48	26	24	98
6	66	41	18	125
7	71	36	22	129
8	66	32	21	119
Means	60.75	34	25.50	120.25

Boundaries which were indicated by only one subject, in only one condition, have been omitted from all further analyses. It was felt that any agreement about boundary locations, whether among subjects (within or across conditions) or by a single subject across conditions, should be taken into account in the analyses; but there is no non-arbitrary way to distinguish an isolated genuinely-intended response from an accidental button-push. The number of responses discarded was very small—about 2% of the total.

Rates of agreement among subjects

Subjects agreed fairly well about boundary locations in both conditions, especially in the sentence tasks; for both tasks, the rates of agreement were slightly better in the normal

Table 2 Paragraph responses by subject

Subject	Number of boundaries indicated in both conditions	Number of boundaries indicated in normal condition only	Number of boundaries indicated in inverted condition only	Total number of boundaries
1	4	8	2	14
2	4	9	6	19
3	18	6	6	30
4	5	3	9	17
5	2	7	12	21
6	11	11	4	26
7	4	12	7	23
8	7	6	5	18
Means	6.875	7.75	6.375	21

condition than in the inverted condition. Mean number of responses per boundary for sentences was 5.179 (65%) for the normal condition, and 4.303 (54%) for the inverted condition; this differs from Lehiste's findings of 44% mean agreement for the normal condition and 46% for the inverted condition. For the paragraph task, mean number of responses per boundary was 3.44 (43%) for the normal condition and 3.31 (41%) for the inverted condition; here there is a substantial improvement over Lehiste's findings of 11% agreement among subjects for the normal condition and 13% for the inverted condition.

Response latencies

Mean response latency (the time from the beginning of a pause to the subject's response, or from the offset of all cues to a response) for the sentence task was 680 ms for the normal condition, and 674 ms for the inverted condition.[2] This difference is not significant ($F(1,1400) = 0.064$, $p = 0.8$); there was, however, significantly more variance in latencies in the normal condition than in the inverted condition ($F(738,664) = 1.187$, $p < 0.025$). These results are contrary to Lehiste's finding of consistently longer response latencies in the normal condition.

Response latencies for the paragraph task were very long and most responses (78% in the inverted condition, and 87% in the normal condition) came after speech resumed after a pause. Mean response latency was 3542 ms in the normal condition, and 2495 ms in the inverted condition. This difference is significant ($F(1,220) = 13{,}367, p < 0.001$).

Boundary cues

Location of sentence boundary responses is very well predicted by presence of a pause, laryngealization (creaky voice), pre-boundary lengthening, and/or a non-level intonation contour[3],

[2] When individual subject's response patterns were examined, one showed consistently longer response latencies in the inverted condition than in the normal condition, two were significantly slower to respond in the normal condition, and five showed no significant differences in latency between conditions.

[3] Laughter and verbal back channels also served as cues, and boundaries where they are present (less than 1% of the total) were deleted from the data set for purposes of cue analyses only.

as Lehiste also found. In the present study, laryngealization was present at 46% of sentence boundaries, non-level intonation contours at 93% of boundaries (14% rising, 79% falling); and approximately 85% of boundaries were preceded by perceptible lengthening. Additionally, virtually all occurrences of these cues result in a boundary response by at least one subject; for example, every non-level intonation contour received subject responses, as did every pause of more than 200 ms. Thus there seems to be a genuine cue–response relationship in the data, rather than simply a coincidence of acoustic characteristics and boundaries that has no predictive value: Presence of cues results in responses by one or more subjects, and subject responses imply presence of cues.[4]

The same cues are present at paragraph boundaries, all of which are also sentence boundaries. Additionally, since nearly all paragraph responses in both conditions came after the onset of the next utterance, information from onsets was generally available to subjects; and, judging by the very long latencies for paragraph responses, subjects needed this information. Abrupt increases in fundamental frequency and increases in loudness seem to be the cues active when onset information is available to subjects; for the utterance-terminal cues, both number of cues present and extent of cue (longer pause, more fall in fundamental frequency, longer creak) also helped to distinguish sentence from paragraph boundaries. Furthermore, terminal contours and onset cues interact in important ways; these are described in the discussion section below.

4. Discussion

A number of the findings of the present study differ from Lehiste's; the reasons for these differences shed a rather new light on this entire line of research.

First of all, the improved rate of subject agreement about boundary locations in the inverted condition can almost certainly be attributed to improved modified condition stimulus quality. Subjects did better because they had more to work with. I have no explanation for Lehiste's finding of better agreement among subjects when they did not have semantic information to work with; one would normally expect better agreement in the normal condition, as was found in the present work.

Also of interest is the fact that using a natural conversation as stimulus material had no detrimental effect on agreement rates. Subjects are just as able—if not more able—to find boundaries in a very casual dialogue as they were in a much less naturally-produced monologue; cues found thus apply to casual speech as well as to more careful styles.

Lehiste's subjects marked more boundaries in the inverted condition than in the normal condition, both for the sentence and for the paragraph tasks. In the present study, no significant differences were found for either type of boundary. It seems likely that the relative numbers of responses made in the two conditions depends on the speech sample used—on the specific, unique pattern of pauses, false starts, rephrasings, and so on, used by each

[4] An attempt was made to assess the extent to which the *number* of subject responses at a given boundary could be predicted from the number and extent of cues present (partial correlation analysis). The result was quite unrevealing (r-squared never exceeded 0.39) because of noise in the data, and the results will not be reported. Any subject will inevitably miss a certain number of boundaries, due to the difficulty of the task, wandering attention, the length of the experiment, and so on. The number of responses at a given boundary thus reflects not only the "strength" of the boundary, but also noise from the subjects' moments of inattentiveness. Presumably with a larger group of subjects this noise would be reduced to a great enough extent that an analysis like this would produce meaningful and interesting results.

speaker—and thus does not reflect any general truth about the production or perception of connected discourse. "Number of responses" is simply not an important quantity.

Of more interest is the rate of mismatches between the two conditions, the percentage of boundary responses with no correlates in the other condition. These mismatches mark places where grammatical/semantic and suprasegmental information do not coincide and where attempts to predict one by the other will fail. Lehiste found that, judging by subject response patterns, 40% of terminal suprasegmental contours were produced when they were not warranted by grammatical structure. She does not mention if grammatical boundaries were ever produced without terminal contours in her sample. She also gives no information about mismatch rates for paragraph responses. In the present study, a mismatch rate of 50%, comparable to Lehiste's result of 40%, was found for the sentence task; for paragraphs, the mismatch rate was 69%. Grammatical boundaries without terminal contours were found approximately as often as terminal contours without grammatical boundaries. Subjects agreed about where boundaries were in both conditions; response mismatches must therefore be due to non-coincidence of grammatical/semantic and suprasegmental boundaries.

Sentence responses which occur in the inverted condition but not in the normal condition can perhaps be explained by the limited number of suprasegmental cues available to perform a relatively large number of functions—emphasis, expressing emotional tonings, indicating uncertainty, and so on. Occurrences of the same cues with different functions—an unambiguous overlapping when semantic information is present—are evidentally being picked up by subjects as sentence boundaries.[5]

Sentences: latency data

Lehiste's finding that availability of semantic information consistently increased response time to sentence boundaries is both so strongly presented in her papers and so counterintuitive that its failure to replicate requires explanation. Certainly one would expect subjects' strategies at some boundaries to involve waiting for the beginning of the next sentence before responding; but one would also expect semantic information to facilitate some responses to sentence endpoints. This dual effect of availability of semantic information—both increases and decreases in response latencies from the inverted condition—would result in no difference between the two conditions in mean response latency, but an increase in the variance of latencies in the normal condition, which is precisely what was found in this study. This discrepancy in results seems to be due to the application of different notions of what reaction times mean in a task like this. Lehiste says of her method:

> Reaction times were established on the basis of listeners' reactions to the initial clicks; these reaction times were used in determining at which point the subjects had perceived a boundary. (1979*a*, p. 106.)

In other words, Lehiste seems to be interpreting her results as if the end of an utterance were an instantaneous stimulus. Clearly this is not the case. Boundary cues such as rise or fall in intonation, laryngealization, and pre-boundary lengthening begin well before an actual boundary is reached, and may well prime subject responses. Treating boundaries as instantaneous stimuli would cause all very short responses to be coded as very long ones. Any response coming less than one reaction time after the end of an utterance would be interpreted as a very slow response to the preceding utterance (see Fig. 1). If all short latency responses were in fact coded as very long latency responses, this would explain the increase found in mean response latency in the normal condition. Not only were short latencies

[5] I owe this suggestion to Diana Van Lancker.

eliminated, they were entered into the analysis as very long latencies, further skewing the result.[6]

To sum up the results of the sentence task: In a given condition, subjects agree quite well as to where sentence boundaries occur; there is substantial disagreement across conditions, however, due to imperfect coincidence of terminal suprasegmental contours and grammatical boundaries produced by speakers. Presence of semantic information has two effects: It increases agreement among subjects (measured by mean number of responses per boundary) and it increases the number of very quick and very slow responses. The majority of responses in both conditions came before the end of a following pause and thus used no information from the onset of the next utterance. Cues for these responses were length of pause, rising or falling intonation, creaky voice, and presence of pre-boundary lengthening. For responses coming after the pause ended, information from the beginning of the next utterance—particularly fundamental frequency and loudness—is additionally used in the decision.

Paragraphs

The interpretation of the paragraph results is somewhat more problematic. It is not surprising that response latencies in the normal condition are longer than those in the inverted condition. However, all paragraph responses, even in the inverted condition, are very long, and only 22% of inverted condition responses, and 13% of normal condition responses, come before the onset of the next utterance. Obviously, participants in a conversation do not wait until a new sentence has begun to see if they have begun a new topic or not; but subjects do, and the additional information they gather from the beginning of a new utterance cannot be simply omitted from the analysis as if they were not using it. Paragraphs are not marked just by terminal contours, as sentences generally are; rather, they are cued both by characteristics of the end of one utterance and by the beginning of the next. Subjects are evidentally comparing two blocks of speech; if the differences between them are great enough, a paragraph boundary is signaled.

That paragraphs should be characterized in this way makes sense when one remembers that the stimulus material in this experiment was an interaction, not a monologue. Changes in topic are always optional; even if one speaker produces an emphatically terminal contour, the other always has the option of having the last word rather than changing the subject or allowing the other speaker to change the subject. Only when the new topic has actually begun is a change in fact inevitable. Thus the two-fold marking of paragraphs: Terminal contours suggest opportunity for change, and initial contours signal that change has actually occurred.

Obviously either part of the composite signal can occur without the other. Topics may be changed without warning, or may be continued despite one party's wish to end them. When this happens, expectations about subject agreement are altered. When a terminal contour occurs without an initial one following it, one would expect fewer total responses than when both occur, but more responses in the inverted condition than in the normal condition, especially if responses come before the end of a pause; and an initial contour not preceded by a terminal one should result in more responses in the normal condition than in the inverted, since decisions about boundaries in the inverted condition depend on differences across boundaries.

[6] There is one other possible explanation for the difference in results: that Lehiste had 24 subjects all of whom consistently waited for a new sentence to begin before responding, while only two of my eight subjects used this strategy. This seems rather unlikely.

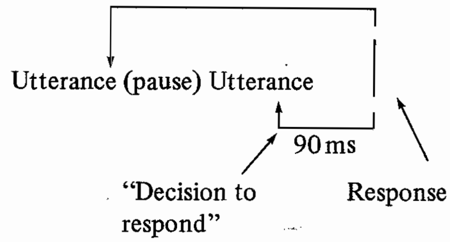

Figure 1 Coding of hypothetical response using end of utterance as instantaneous stimulus.

Table 3 Sample paragraph boundaries

1A. (Discussion of Gauguin)
S: "But you never know why he did that."
Terminal fundamental frequency = 127 Hz + creak
Pause = 4190 ms
Duration of last VC = 218 ms

1B. S. "Uh this girl that we were talking to at the pool . . ."
Initial fundamental frequency = 234 Hz

2A. (Discussion of Freud's writings)
S: 'I don't know, I didn't read the paper, this is jus' I wanna read them."
Terminal fundamental frequency = 184 Hz + creak
Pause = 340 ms
Duration of last VC = 184 ms

2B. S: "We'll have time to do stuff like that."
Initial fundamental frequency = 284 Hz

3A. K: "But uh he heard a lot of that stuff."
Terminal fundamental frequency = 120 Hz + creak
Pause = 2040 ms
Duration of final VC = 200 ms

3B. K: ". . . 'bout how anxious n' ah everything."
Initial fundamental frequency = 165 Hz.

This analysis accounts quite well for subject response patterns. Consider the utterances in Table 3. Sentence 1A ends with low fundamental frequency, laryngealization, a long pause, and a long last syllable; the speaker's voice is additionally very quiet. A topic change occurs; the next utterance begins with much higher fundamental frequency, and is much louder. Six subjects heard a paragraph boundary in the inverted condition, and seven responded in the normal condition. A topic change occurs rather suddenly between utterances 2A and 2B; the end of 2A is laryngealized, but the fundamental frequency is not particularly low, the speaker's voice is not particularly quiet, and the pause is fairly short. The next utterance has a higher fundamental frequency, but there is no particular increase in volume. All subjects responded in the normal condition, but none did in the inverted. When no terminal contour is present, there is not enough difference between the two blocks of speech to evoke a response. Finally, the end of sentence 3A is heavily marked as terminal. The pause is very long, fundamental frequency is low, volume is low, and the last syllable is quite long; but no topic change occurs and the next utterance is also low in volume and in fundamental frequency. Four subjects responded in the inverted condition–all before the end of the pause–and only one responded in the normal condition.

Besides accounting for long response latencies and patterns of agreement among subjects, this view of paragraph marking helps explain the large mismatch rate between responses in the inverted and normal conditions. Terminal contours without topic changes–normal occurrences in conversations–will result in inverted responses without corresponding normal condition responses; and topic changes made without warning–equally commonplace–will tend to cause normal condition responses without inverted counterparts. Speakers are not simply producing terminal suprasegmental contours in non-terminal positions, as they were for sentences; rather, they are operating in a system where so many options exist that high mismatch rates are to be expected.

It is not entirely clear from Lehiste's discussion of paragraphs whether her subjects used speech onset cues or not. They may not have; doing both the sentence and the paragraph tasks at once may have encouraged them to respond quickly enough that this information was not available to them, and the very low agreement rates (13% in the inverted condition, and 11% in the normal condition) suggest that they were working with incomplete information. Whether they used this information or not, however, her discussion never touches on it, even as a potential source of information to subjects. Her set of paragraph cues–a complicated weighted sum of intonation, pause length, and pre-boundary lengthening–is composed entirely of unit-terminal phenomena. The stimulus material in her experiment was monologue extracted from dialogue, though, rather than a complete dialogue; evidentally using monologue resulted in her thinking only in terms of one speaker, and thus overlooking the interactive aspects of paragraphing in natural conversations.

Lehiste (1979*a*) concludes with brief speculations as to the relationship between sentence boundaries, paragraph boundaries, and the boundaries of conversational turns. Her discussion, however, is set in the same non-flexible, non-interactive framework as her analysis of paragraph boundaries. This is probably once again at least partly because the very restricted nature of the stimulus material she used gave her no reason to think in terms of dialogue. Clearly, turn-taking exhibits the same sort of optionality that topic-changing does. A conversant can be offered the speaking turn, or can take it without its being offered, but he cannot be compelled to take it (Duncan & Fiske, 1977). Thus, one would expect any suprasegmental characteristics of turn boundaries to exhibit a structure similar to that of paragraph cues. It would be nice to relate sentence and paragraph boundaries to conversational turns; but this question must be approached with an eye to the optionality of the phenomena in question.

Dale Terbeek wrote the speech inversion program, read and commented on several earlier versions of this paper, and provided advice and support at all stages of this research, for all of which I am deeply grateful. Lance Rips, Victor Yngve, and Alexis Manaster-Ramer also commented on earlier versions of this paper. Karen Jaeger and Barbara Coughlin held the conversation for the stimulus tape. C-W. Kim lent oscillograph paper when none was to be had anywhere, and spectrograph paper was purchased with the help of a generous gift from Mr G. A. Burbridge, Sr.

References

Duncan, S. & Fiske, D. (1977). *Face-to-Face Interaction: Research, Methods and Theory.* Hillsdale: Lawrence Erlbaum Associates.

Klatt, D. (1975). Vowel lengthening is syntactically determined in a connected discourse. *Journal of Phonetics* 3, 129–140.

Lehiste, I. (1979*a*). Sentence boundaries and paragraph boundaries–perceptual evidence. In: *The*

Elements: A Parasession on Linguistic Units and Levels (Clyne *et al.* eds), Chicago: The Chicago Linguistic Society, pp. 99–109.

Lehiste, I. (1979*b*). Perception of sentence and paragraph boundaries. In: *Frontiers of Speech Communication Research* (Lindblom & Ohman, eds), New York: Academic Press, pp. 191–201.

Lehiste, I. & Wang, W. S-Y. (1977). Perception of sentence and paragraph boundaries with and without semantic information. In: *Phonologica 1976* (Dressler & Pfeiffer, eds), Innsbruck: Institut fur Sprachwissenschaft der Universitat Innsbruck, pp. 277–283.