Syntactic structure and acoustic pattern in speech perception

ARTHUR WINGFIELD¹ AND JAN F. KLEIN
BRANDEIS UNIVERSITY

The interaction of syntactic structure and acoustic pattern in perceptual segmentation of heard speech was investigated using sentences recorded with intonation patterns appropriate to their underlying structure and sentences where intonation was placed in direct conflict with the underlying structure. Speech samples were monitored through dichotic earphones with messages switched from one ear to the other either between or within major linguistic constituents. Analysis of errors in source localization and in sentence reproduction suggested a primary role of acoustic pattern in segmentation is to cue underlying syntactic structure. Two subsidiary experiments provided evidence that observed effects were related to active perceptual processing of the speech signal.

A consideration of the active perception of continuous speech seems logically to demand that it be processed by the listener in definable segments of some manageable length. That major linguistic constituents determine the perceptual processing units imposed by the listener on heard sentential material now seems more than likely. Evidence in support of this assumption has come largely from studies such as that of Fodor and Bever (1965), who inserted short bursts of noise ("clicks") in various positions in recorded sentences. The perceptual integrity of major syntactic constituents was suggested both by increased accuracy of localizing heard clicks when they were presented between major constituents than within them, and by a strong tendency for perceived position of the clicks to "migrate" toward such boundaries. This effect was further reported even in cases where no noticeable acoustic pauses coincided with the major syntactic boundaries.

Garrett, Bever, and Fodor (1966) specifically controlled for acoustic variables (pause, intonation, etc.) by splicing together tape-recorded segments of two separate sentences such that identical word segments in a given intonation pattern were heard in different contexts, and hence with different underlying syntactic structures. In this way, major syntactic boundaries (MSB) and normally associated acoustic pattern did not necessarily coincide. The previous findings with respect to click localization and the

MSB were verified. Further evidence, amplified by analysis of effects of major clause divisions and minor within-clause phrase structure boundaries, has been subsequently offered by Bever, Lackner, and Kirk (1969).

While the independent role of syntactic structure in perceptual segmentation has received convincing support, it is well recognized that the relationship between syntax and acoustic pattern in both the production and perception of speech is an intimate one. Indeed, analysis of hesitation pauses in fluent discourse, and listeners' reactions to them, has led some writers to specific emphasis on acoustic pattern as determining perceptual units in speech decoding (cf. Dittman & Llewellyn, 1967).

The importance of these issues to psycholinguistic theory prompts further exploration of the relationship between syntactic structure and acoustic pattern in the perception of speech.

In the following experiments, recorded sentences—spliced to control for acoustic variables in a manner similar to that of Garrett, Bever, and Fodor—were presented through dichotic earphones. The present experiments, and those previously cited, differ in two major respects. First, in this case, sentence types with extremely well defined MSBs and associated intonation patterns were specifically selected to emphasize perceptual conflict when, after splicing, the two no longer coincided. Second, an independent experimental methodology was employed. Our rationale, however, was similar in principle to that used when extraneous clicks were inserted in sentences.

A single sentence was heard to begin in one ear and was then switched between the two ears at various points either between or within major syntactic constituents. Successful perception of the sentence thus necessitates an integration, at some level, of information from the two sources. It is known that integration can take place with little disruption, provided the rate of switching is not too rapid (Cherry & Taylor, 1954; Huggins, 1964). Indeed, even in a case of competing messages presented to the two ears simultaneously, the content of the message can be reliably shown to take precedence over its source in perceptual ordering (Gray & Wedderburn, 1960).

To the extent that the formal

constituent unit (marked by the MSB) may define an integral perceptual unit, S's ability to determine the exact point of a switch will be significantly more difficult when it occurs within such a unit than when the message crosses the ears in the break between such units. This follows from the assumption that the primacy of perceptual segmentation imposed by the listener will lead to a subordination, and loss, of specific localization information within such segments. In the first experiment reported, the role of syntactic and acoustic factors in processing heard sentences is assessed both by accuracy in source localization under these conditions and by analysis of reproduction errors for the sentences themselves.

The unequivocal interpretation of the initially cited studies as illustrative of segmentation in perceptual processing also remains at issue. Ladefoged (1967, p. 161), for example, using the same paradigm as Fodor and Bever, has reported a similar effect of the MSB when Ss were told to report the location of "subliminally perceived" clicks in sentences, when no clicks were in fact presented. For this reason, a subsidiary experiment was conducted using two additional reporting procedures in an attempt to reduce or eliminate memory load (the need for reproducing the sentences from memory). Evidence for constituent segmentation found under these conditions would encourage an interpretation of such findings as correlates of active perceptual processing, as opposed to some postperceptual subject bias operating perhaps in the stage of recall or reproduction of the sentences.

METHOD

Stimulus Materials

Twenty pairs of fully grammatical, 10-word sentences were constructed. All consisted of either a subordinate plus an independent clause or an independent plus a subordinate clause. The two sentences of each pair were constructed such that they shared an identical 5-word sequence, but with the MSB occurring at different points within the two identical segments.

Three examples of the 20 sentence pairs used are shown below. The common sequences are shown in italics, with the comma indicating the MSB.

1(a) To avoid any attempts to

influence voting, machines were installed.

- 1(b) Due to our new mayor's influence, voting machines were installed.
- 2(a) In schools which encourage running, races will soon become favorites.
- 2(b) Lacking televised sports coverage, running races will soon become unpopular.
- 3(a) By forgetting to call George, we lost a crucial vote.
- 3(b) By forgetting to call, George forfeited his chance to go.

As illustrated by these examples, the common sequences occurred at various points relative to both the MSB and their position in the sentences. Each sentence of a pair was initially recorded on a monaural tape by a female speaker naive as to the purpose of the experiment. The common five-word segment of each sentence was then spliced into the recorded surrounding linguistic context of the other member of the pair. Together with duplicates of the originally recorded sentences, this procedure yielded a fully balanced set of 80 sentences, 40 spliced and 40 unspliced. The original unspliced recordings of each sentence of the pair was heard in an intonation pattern appropriate to its syntactic structure. The spliced versions contained the same word sequence, but were heard associated with the intonation pattern of a sentence with a different underlying structure.

The location of the MSB across all sentences varied from the 2nd to the 7th word in the 10-word sentences. Splices were silent and completely unnoticed in playback.

The sentences thus prepared were rerecorded on a two-channel (stereophonic) recorder, with the input switched from one recording channel to the other at preset points. Switches were placed at the MSB, at one through five words before, or at one through five words after the MSB. Switches were instantaneous and always occurred at points between words. As a control to insure that Ss attended to the entire sentence, five pairs of sentences had two switches per sentences.

Procedure

Each S heard a total of 40 sentences, half spliced and half unspliced, in random order. The experiment was balanced such that no S heard the same sentence more than once, but that by the end of the experiment, each sentence was heard with different switch positions relative to the MSB. Instructions were to listen to the entire sentence and then to write it down.

Table 1
Two Examples of Test Sentences, Followed by Typical Reproduction Errors for Normal and Spliced Versions. (The matched sentence, from which the spliced segment was drawn, is also shown.)

Test Sentence	Besides commercial uses of color movies, they are simply enjoyable.
Normal Version	Besides commercial uses of () movies, they are simply enjoyable.
Spliced Version	Besides commercial uses of color, movies () are simply enjoyable.
Matched Sentence	Among the commercial uses of color, movies are most typical.
Test Sentence	After he finished drinking, the wine bottle was practically empty.
Normal Version	After we finished drinking, the wine bottle was nearly empty.
Spliced Version	() we finished drinking the wine; the bottle was practically empty.
Matched Sentence	Even before he finished drinking the wine, it took effect.

indicating where the switch between ears had occurred.

The main experiment, which took about 50 min, was preceded by a practice session conducted under the same procedures. The practice sentences consisted of the six pairs of sentences used by Garrett, Bever, and Fodor (1966). Thirty undergraduates at Brandeis University served as Ss.

In order to examine the possible influence of a memory component in the experimental task, the main experiment was repeated with an additional 12 Ss. Six of these were given a printed copy of each sentence after it had been heard, and were instructed simply to mark the location of the switch(es) on the printed copy. A second group of six Ss followed the same procedure, except that they were allowed to read a printed copy of each sentence as it was being presented, and to mark the location of the switches as soon as they were heard.

RESULTS

Switch Localization

In the main experiment, for the normal (unspliced) sentences, switches were correctly located significantly more often when they had occurred at the MSB (93%) than when they occurred either before (62%; $t = 5.30$, $p < .001$) or after (59%; $t = 6.40$, $p < .001$) the MSB. In the case of the spliced sentences where the MSB and acoustic pattern did not coincide, however, accuracy in switch localization was similar whether switches occurred before (53%), at (59%), or after (53%) the MSB. These differences were not significant. Reanalyzing the spliced sentences for localization accuracy relative to intonation pattern did reveal a consistent trend. For this analysis the point was taken in the common word sequence that had contained the MSB in the matched sentence of the original pair. After splicing, this point still retained whatever acoustic marking it had when originally recorded. (While a variety of acoustic features, such as intonation, stress, and pausing, accompanied this point, for brevity alone the term "pause" will be used.) Localization accuracy was significantly greater when the switch occurred at the

pause (77%) than when it occurred either before (57%; $t = 2.42$, $p < .001$) or after (44%; $t = 3.27$, $p < .001$) the pause.

While localization accuracy was greater at the acoustic pause than at the MSB in the spliced sentences, the number of mislocated switches that "migrated" from some other position in the sentence to each of these points was equal, 30% being reported as having occurred at the MSB, and 30% at the acoustic pause. This may be compared with 44% of mislocated switches that migrated to the MSB in the normal sentences where the MSB and pause coincide. While this migration effect tends to be larger, it is clear that the combined effects of syntax and pause are not simply additive. These errors, together with switches that migrated toward, but not actually into, these points, accounted for the majority of all localization errors.

Sentence Reproduction

A total of 96% of the normal sentences and 88% of the spliced sentences were reproduced correctly in the course of reporting switch localization ($t = 5.70$, $p < .001$). While erroneous reproductions of both normal and spliced sentences tended toward complete, fully grammatical sentences, the pattern of the errors showed substantial differences in these two cases. Two examples of these differences are illustrated in Table 1.

In 91% of the error reproductions for normal sentences, the errors were marked by simple omissions and substitutions with the erroneous responses matching the essential syntactic structure of the heard (test) sentences. On the other hand, only 57% of error responses to spliced versions were of this type. In 38% of the cases, there were major reconstructions of the sentences, which in effect produced "new" sentences whose MSB was shifted to conform to the structure of the matched sentence from which the spliced segment—and its intonation pattern—had been taken. In these cases, as illustrated by the examples in Table 1, the conflict between structure as indicated by the lexical content and structure as marked by the acoustic pattern was resolved by word deletion and substitution, reforming the

Table 2
Percentage of Switches Localized Correctly in Subsidiary Experiments When Switches Occurred
Either At or Off the MSB or Acoustic Pause

	Delayed Reading		Simultaneous Reading	
	Off Boundary	At Boundary	Off Boundary	At Boundary
Normal Sentences	75	93**	63	80*
Spliced Sentences (Relative to MSB)	70	80†	65	77*
Spliced Sentences (Relative to Pause)	69	100**	67	92**

** Accuracy significantly better at the boundary than either before or after the boundary.

* Accuracy significantly better at the boundary than before the boundary only.

† Neither accuracy before nor after the boundary significantly different from accuracy at the boundary.

heard sentence to conform with the acoustically marked structure. (The remaining 9% of errors to normal sentences and 5% of errors to spliced sentences were fully grammatical sentences whose MSB did not match its position either in the heard sentence or its matched pair.)

Neither the number of switches nor their location relative to the MSB or acoustic pause had a significant effect on error scores for the sentences.

Localization Accuracy with Memory Control

Eliminating the need to reproduce the sentences from memory in the subsidiary experiments yielded generally increased accuracy of switch localization. As shown in Table 2, however, the major findings of the main experiment were still seen under these conditions. When Ss indicated subjective localization on a printed sheet immediately after hearing each sentence (delayed reading), localization accuracy for spliced sentences was again influenced more by acoustic pause than by the MSB. When Ss marked the location of the switch as they listened to the sentence (simultaneous reading), the MSB-pause differential was reduced. The same strong effects of acoustic pause in the spliced sentences were nevertheless confirmed.

The smaller number of Ss and generally high level of accuracy in the subsidiary experiments prohibit reliable analysis of migration effects under these conditions.

DISCUSSION

Results of the main experiment are clear in demonstrating that the integration of a single message from two sources is more influenced by segmentation imposed by the listener, than by segmentation in terms of source. This is seen primarily in the systematic nature of the localization errors, and is reinforced by finding that neither varying the location of the switch in the message nor doubling the number of switches had a significant effect on Ss' ability to integrate and reproduce the

sentences correctly.

Increased accuracy in localization of source when the message was switched between major constituents rather than within them in the normal sentences conforms with the likelihood that the normal incoming speech signal is segmented according to the constituent structure of the sentence. The independent effects of syntactic structure and acoustic pattern in the spliced sentences is viewed as serving primarily to amplify this conclusion, rather than to negate it. Neither our working hypothesis nor these results either demand or deny that perceptual segmentation according to syntactically determined units necessarily requires the aid of accompanying intonation pattern. Indeed, perceptual segmentation in the absence of intonation has been argued in principle by Lieberman (1967, p. 125) and has been shown empirically by Abrams and Bever (1969). Splicing sentences of the type used here, where the MSB and associated acoustic pattern are extremely well defined, did not simply control for intonation pattern. Rather, it placed them in marked conflict. It is under these conditions that effects of intonation begin to appear. The mechanism in these cases is not thought to be segmentation by acoustic pause per se. The specific role of intonation in these results is instead suggested by the analysis of erroneously reproduced spliced sentences. In these cases, correct syntactic resolution was opposed by an active restructuring of the heard sentence to, in effect, shift the MSB to conform with the acoustic marking. Thus, the observed effects of acoustic pause on localization accuracy were in fact a reflection of the constituent structure as "perceived" by S in these spliced sentences.

This interpretation supports a hypothesis that perceptual segmentation of sentential material is determined primarily by its syntactic structure, but with accompanying acoustic pattern serving to mark the underlying structure or aid in

syntactic resolution. All lexical and acoustic information cue the underlying surface structure, which, in turn, may permit access to semantic information contained in the deep structure.

The similarity of major results in the subsidiary experiment, where sentence reproduction and delay of response was eliminated, reduces the likelihood that observed effects of syntax and intonation relate specifically to postperceptual recall or response-bias. To identify these effects with active perceptual processing, however, still fails to isolate their role in input scanning vs processing of the information once detected. A second issue remaining is the extension of these findings to more ordinary on-going fluent discourse, where idealized grammatical structure occurs with questionable frequency, and where acoustic effects as described here might be even more profound.

REFERENCES

- ABRAMS, K., & BEVER, T. G. Syntactic structure modifies attention during speech perception and recognition. *Quarterly Journal of Experimental Psychology*, 1969, 21, 280-290.
- BEVER, T. G., LACKNER, J. R., & KIRK, R. The underlying structures of sentences are the primary units of immediate speech processing. *Perception & Psychophysics*, 1969, 5, 225-234.
- CHERRY, C., & TAYLOR, R. Some further experiments on the recognition of speech with one and with two ears. *Journal of the Acoustical Society of America*, 1954, 26, 554-559.
- DITTMANN, A. T., & LLEWELLYN, L. G. The phonemic clause as a unit of speech decoding. *Journal of Personality & Social Psychology*, 1967, 6, 341-349.
- FODOR, J., & BEVER, T. G. The psychological reality of linguistic segments. *Journal of Verbal Learning & Verbal Behavior*, 1965, 4, 414-420.
- GARRETT, M., BEVER, T. G., & FODOR, J. The active use of grammar in speech perception. *Perception & Psychophysics*, 1966, 1, 30-32.
- GRAY, J. A., & WEDDERBURN, A. A. Grouping strategies with simultaneous stimuli. *Quarterly Journal of Experimental Psychology*, 1960, 12, 180-184.
- HUGGINS, A. W. F. Distortion of the temporal pattern of speech: Interruption and alternation. *Journal of the Acoustical Society of America*, 1964, 36, 1055-1064.
- LADEFOGED, P. *Three areas of experimental phonetics*. London: Oxford University Press, 1967.
- LIEBERMAN, P. *Intonation, perception and language*. Cambridge: M.I.T. Press, 1967.

NOTE

1. Address: Department of Psychology, Brandeis University, Waltham, Massachusetts 02154.

(Accepted for publication April 3, 1970.)