JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR 4, 414-420 (1965)

The Psychological Reality of Linguistic Segments[1]

J. A. FODOR AND T. G. BEVER

Massachusetts Institute of Technology, Cambridge, Massachusetts

Experimentation with the subjective location of clicks heard during speech supports the following conclusions:
(a) Clicks are attracted towards the nearest major syntactic boundaries in sentential material. (b) The number of correct responses is significantly higher in the case of clicks located at major segment boundaries than in the case of clicks located within segments. (c) These results are consistent with the view that the segments marked by formal constituent structure analysis in fact function as perceptual units and that the click displacement is an effect which insures the integrity of these units. (d) The distribution of acoustic pauses in the sentential material does not account for the observed distribution of errors. (e) There is a slight tendency to prepose responses to clicks in sentences. This tendency is reversed during later stages of the experimental session. Both these effects are asymmetrical for the two ears.

Linguistic models provide an analysis of sentences into segments of a number of different types. For example, the grammar of a language specifies the sequence of sound segments which constitute the sentences of that language. At this level, a sentence is represented by a sequence of phonetic symbols.

In addition, a grammar provides an analysis of the structure which underlies the acoustic pattern exhibited by a phonetic description. It does so, in part, by providing a *constituent analysis* for each sentence. The constituent analysis is a representation of the more abstract segments of which a sentence is composed. It is with the psychological reality of the segmentations such analyses assign to sentences that we shall be concerned in this paper.

Consider the sentence "That he was happy was evident from the way he smiled." In addition to specifying an appropriate phonetic representation, an adequate grammar of English must provide a correct account of the syntactic relations between that sentence and such sentences as "He was happy," "It was evident," "He smiled in a certain way," etc. It must also correctly predict the stress pattern the sentence exhibits. To achieve these purely linguistic goals, the grammar must analyze this sentence into roughly the following constituents: *that he was happy*; *he was happy*; *was happy*; *was evident from the way he smiled*; *evident from the way he smiled*; *from the way he smiled*; *the way he smiled*; *the way*; *he smiled*. This analysis may be expressed by a notation in which each constituent of the sentence is enclosed by parentheses: (((that) ((he) ((was) (happy)))) ((was) ((evident) ((from) (((the) (way)) ((he) (smiled))))))).

The question arises whether the units marked off by such segmentations correspond in any direct way to the perceptual units into

[1] This work was supported in part by the U. S. Army, Navy, and Air Force under Contract DA 36-039-AMC-03200(E); in part by the National Science Foundation (grant GP-2495), the National Institutes of Health (grant MH-04737-04), the National Aeronautics and Space Administration (Ns G-496) and the U. S. Air Force (ESD Contract AF 19(628)-2487; and, in addition by National Institute of Mental Health grant MPM 17,760. The authors wish to thank Professor H.-L. Teuber for making available the experimental facilities for this work and Mr. M. Garrett for his advice and criticism.

which sentences are articulated by speakers and hearers. A number of techniques are available for experimentally determining the segmentation of a complex percept. The simplest of these is a direct appeal to the intuitions of the perceiver. Very often, the *S*'s reports of the preferred segmentation of a speech stimulus appear to be fairly stable for adult speakers of a language.

A more subtle way of establishing the segments of a complex percept exploits the tendency of a perceptual unit to preserve its integrity by resisting interruptions. The *E* introduces an appropriate form of interfering stimulation which the *S* is required to locate relative to the perceptual object. A segmentation is established by demonstrating a reliable tendency for the subjective locations of the interfering stimuli to cluster. The points at which such clusters occur are identified as the boundaries of the segmental units comprising the perceptual object.

A recent attempt to investigate perceptual units in speech by employing this sort of technique was made by Ladefoged and Broadbent (1960). They argued that the unit of speech perception must be longer than a single speech sound (phone), just as the perceptual unit in reading is clearly longer than a single letter. To test that hypothesis, they devised an experiment calculated to reveal the perceptual units of speech. The *S*s listened binaurally to sentences and to strings of digits spoken in English. Each string and each sentence had an extraneous sound (a click) superimposed on it. The *S*'s task was to note the word during which the extraneous sound occurred. It was assumed that switching from processing verbal to processing nonverbal material would be controlled, in part, by the segmentation of the speech and not solely by an intrinsic attention span; i.e., that switching would take place at the boundaries of units. Thus Ladefoged and Broadbent predicted that the magnitude of errors in placing the clicks is a measure of the size of perceptual units.

With eight sentences and ten random digit strings, Ladefoged and Broadbent found that location errors often displace the noise *beyond* the boundaries of the world in which it is objectively positioned; that all *S*s tend to locate the noise prior to its objective position whether in sentences or in random material; and that errors in noise location are larger for sentential material than for digit sequences.

Ladefoged and Broadbent's attempt to apply to linguistic material experimental techniques for establishing segmentation thus produced suggestive results: the unit of speech perception probably does not correspond to the phone and is larger in sentences than in random sequences of words. But they failed to yield a definite answer to the question: to what extent do the larger units into which sentences are segmented for purposes of linguistic analysis correspond to the perceptual units involved in speech recognition?

The experiments which will now be presented are concerned with the following hypothesis. (*H*) *The unit of speech perception corresponds to the constituent.*

This hypothesis supports certain predictions about the effects of the integrity of segments in the sort of experimental situation investigated by Ladefoged and Broadbent. It was seen above that a given word boundary in a sentence may mark a coincidence between the boundaries of any number of constituents. If, as has been supposed, interfering stimuli tend to be displaced to or towards the boundaries of perceptual units, and if *H* is true, there ought to be a demonstrable tendency for clicks presented simultaneously with sentences to be heard at the boundaries of the constituents in which they are objectively located. Moreover, the larger the number of overlapping constituents in which the click is objectively located, the stronger should be the tendency for *S*s to displace it to the common boundary of those constituents. For example, the tendency to displace the noise from its objective position at B to the position between B and C ought to be greater in structures like ((A) (B)) (C) than in structures like (A) (B) (C). For, to perceive the noise in its objective position in the former would interrupt two units, while to perceive it in its objective position in the latter would interrupt only one. Correspondingly, in structures of the former kind, the tendency to hear the noise between B and C ought to be stronger than the tendency to hear it between A and B, since hearing it after B would interrupt no segments while hearing it prior to B would interrupt one.

In short, the assumption that H is true, together with the assumption that perceptual units in speech resist interruption, leads to the following experimental prediction: (H') *Noise heard during speech should tend to shift perceptually towards the boundaries of constituents. This shift should occur in such fashion as to minimize the number of constituents the noise is perceived as interrupting.* Thus, for example, a click objectively placed in the final syllable of "happy" in the sentence "That he was happy was evident from the way he smiled" should tend to migrate toward the following boundary, since that boundary represents the termini of three constituents in addition to the word "happy" itself.

Predictions based on H' in fact receive some support from a reanalysis of Ladefoged and Broadbent's results for the five sentences for which they provide summaries of their data. Such an examination shows first that, in the case of four of the five sentences, the noise is displaced either into the boundary preceding its objective position or into the boundary following its objective position, depending upon which boundary marks the terminus of the larger number of constituents. Second, the only sentence for which *S*s tend to displace the superimposed noise towards the end was the one in which the boundary marking the terminus of the larger number of constituents *followed* the actual position of the superimposed sound. These results appeared to supply sufficient support for H' to warrant more extensive experimental investigation. The following experiment was therefore undertaken.

METHOD

Materials

Twenty-five sentences containing only one boundary at which a relatively large number of constituents are coterminous and five sentences containing two such boundaries were each recorded nine times on one track of a stereophonic recording tape. Sentences ranged in length from 8 to 22 words, with the average length 13.1 words. In each sentence, one boundary was designated the *zero position*. In all cases where a sentence had a single major boundary, it was chosen as zero. In the five other cases, one of the two major boundaries was chosen arbitrarily.

On the second track of each of the nine copies of each sentence, one capacitor-discharge click was recorded. The intensity of the clicks was approximately equal to the most intense speech sound and the duration was about 25 msec. One of the nine copies had the click placed contemporaneously with the *zero position*. A second, third, and fourth copy respectively had clicks contemporaneous with the first, second, and third syllables *posterior* to the zero position. Finally, an eighth, and ninth copy respectively had clicks in the word boundaries prior and posterior to the zero position. The distance between these last two positions and the zero position therefore varied when measured in syllables, but was always unity when measured in words. Thus, for the sentence discussed above, the nine click locations were:

```
That he was    happy    was    ev   i  dent from
          —3 a —2 —1 0  +1 b  +2 +3
the way he smiled
```

where letters designate clicks located in boundaries other than the major one.

It should be noticed that, since the objective locations of the clicks were balanced on either side of the major boundary, any response bias would be self-cancelling. The prediction derived from H' was that errors in the location of clicks objectively prior to the major boundary would be towards the end of the sentence and conversely for clicks objectively following the major boundary. Any general directional biases in click placement would therefore tend to reduce the degree of confirmation of H' for one-half of the response, but would strengthen it for the other half.

Subjects

Nine experimental groups, consisting of four right-handed undergraduates each, heard one copy of the original 30 sentences. The order of sentence presentation and the click location in the sentences heard by a particular group were determined randomly. Thus.

each group heard approximately the same number of clicks in each position relative to the major break, although in different sentences and in different orders.

Procedure

The Ss were presented with the stimulus material through headphones, the sentence in one ear and the click in the other. Each group had two Ss in each of the orientations of the headphones (viz., click left and sentence right as opposed to click right and sentence left). This orientation was not varied over the 30 sentences. The Ss were instructed to write the entire sentence and to indicate graphically where in the sentence they thought the click had occurred. They were also given an opportunity to indicate one of three levels of confidence in the correctness of each of their responses.

Scoring

H' predicts that errors should be in the direction of the major break or into the major break, but not beyond it. Thus, for example, a —3 click if marked by *S* as in a, —2, —1, or 0, confirms *H'*, but if marked in +1, b, +2, or +3 it does not. Although it is not clear that such "overshoot" responses should be counted at all, as a conservative measure these will be scored as errors tending to *dis*confirm *H'*. Responses were scored to the nearest syllable or to the nearest constituent boundary, whichever was relevant.

Results

Results

Of 1080 responses, 22 were not scored because the sentence was incorrectly transcribed by *S*. The 120 responses to clicks objectively located in the "0" position are not directly subject to *H'*, since they are already in the deep break and cannot be attracted towards it. Of the remaining 938 responses to clicks objectively located in the eight nondeep break position, 749 or 80% were errors. The number of errors did not increase with the length of the sentence and no correlation was discovered between the degree of confidence reported by Ss and the objective accuracy of their performance. (Although they follow all the effects discussed below, the five sentences with two major breaks are hereafter omitted from the data, because the zero position in each was chosen arbitrarily.)

For the sentences with one major syntactic break, *H'* predicted the direction of displacement for 66% of the erroneous responses. That is, errors in locating clicks objectively preceding the major break (—3, —2, —1, a) followed the objective click position. Errors in locating clicks objectively following the major break (+1, +2, +3, b) preceded the objective click position. Specifically, 53 responses beyond the deep break from the objective click position and 115 responses away from the deep break were scored as errors (not confirming *H'*). There were 414 responses towards or into the deep break which were scored as confirming *H'*. The hypothesis was confirmed for all the sentences (reject *H'*, $p < .01$ by sign test two-tailed) and for each of the 36 Ss (reject *H'*, $p < .01$). It was also confirmed for each of the eight click positions other than '0' (reject *H'*, $p < .01$ by one-tailed test). None of the Ss, sentences, or click positions yielded results confirming *H'* for less than 60% of the errors.

A subsidiary experimental prediction derivable from *H'* is that clicks whose objective location is in a major syntactic break, i.e., in the '0' position, would be located more accurately than clicks whose objective position is within segments. This follows from the assumption that a substantial proportion of the displacement is attributable to the tendency to maintain the integrity of segments as perceptual units. This prediction was in fact confirmed by the data. Significantly more of the correctly located clicks were objectively in the deep break position than in any other of the nine positions (independence rejected by χ^2 — test, $p < .025$ for the "+1", and $p < .01$ for all others).

Analysis of the absolute location of the incorrect responses was also carried out to determine their distribution by position. The analysis exhibited the predicted tendency to displace clicks into the zero position. More than 35% of *all* erroneous responses were into the major boundary, and more than 60% of all erroneous responses (including the "overshoots") were either in the major boundary or in the syllables immediately preceding or following it. It should be noted, however, that the exact significance of this result depends upon one's prior assumptions about the probability distribution of the perceived clicks if no structural effects were operative. While any particular assumption about this distribution would be difficult to justify, it is clear that the observed differences between the zero position and all others could not be accounted for on a null hypothesis.

Discussion

The above data appear to demonstrate that the major syntactic break plays an important role in determining the subjective location of noise perceived during speech. They would thus appear to provide grounds for the acceptance of hypotheses H' and H.

It remains possible, however, that the factors determining the direction of displacement of the clicks were only indirectly related to the formal constituent structure of the sentences. In an unpublished experiment, Garrett (1964)[2] has shown that relatively long pauses introduced at a selected point in a string of spoken digits will tend to attract interfering noise. That is, noise superimposed upon the spoken digits will tend to be heard in the position objectively occupied by the acoustic pause. Also Bolinger and Gerstman (1957) showed that in isolation the ambiguous phrase *light-house* keeper, is assigned a structure depending on the relative duration of the pauses between the individual words. Garrett's and Bolinger and Gerstman's results show that *in the absence of any other cues and isolated from sentences* acoustic pauses are capable of inducing a particular structural organization. It might thus be hypothesized that the constituent breaks in spoken sentences invariably have slight acoustic pauses associated with them, and that the duration of each such pause corresponds to the importance of the corresponding constituent break. It could then be maintained that the effect of constituent structures upon click displacement is attributable to the acoustic pauses in the sentence and not to the underlying constituent structure.

The difference between the interpretation of the present data as directly due to the effect of constituent structure and the interpretation which holds that it is attributable to pausal phenomena that are themselves distributed in accordance with the constituent structure is of considerable importance. If it is actual acoustic pauses that are critical, then it may be claimed that the division of the sentence into perceptual units is accomplished by markers in the physical signal which delineate their boundaries. If, on the other hand, these units are *not* usually marked in the physical signal, then it must be the hearer who imposes an articulation into perceptual units upon the speech signal. On this view, the hearer *contributes* the perceptual structure to the physical signal on the basis of his knowledge of the constituent structure rules of his language. Hence, what is at issue between the two interpretations is the difference between an "active" and a "passive" theory of speech perception.

To answer this question the 25 sentences with one major constituent break were analyzed with a pen oscilloscope. When analyzed in this way, many of the sentences did not exhibit *any* acoustic pause at the point associated with the major boundary (though they do exhibit such pauses at other locations). In the sentence "That he was happy was evident from the way he smiled," for example, there is no acoustic pause discoverable at the break between "happy" and "was."

The 25 sentences were categorized according to whether or not there was a discoverable acoustic pause at the major break. Sentences exhibiting no such pause were analyzed to determine whether there was a measurable drop in energy. To be sure that all acoustic pauses would be included, those attributable to phonetic effects, such as the influences of stop-consonants, were included in the scoring. Nevertheless, only eight sentences showed full acoustic pauses at the major breaks; six had a severe drop in intensity, but not to zero. Seven sentences had a mild intensity

[2] Unpublished paper.

drop and four exhibited no pause and no intensity drop.

If it is true that the perceptual organization of speech and the click displacement in sentences are due to acoustic pauses, the four groups above should show a decreasing agreement with *H′*. That is, the eight sentences with observable pauses should give better results on *H′* than the four sentences with no intensity drop. The results indicate that this did not occur. Where the per cent energy drop is 100, 81% of the responses confirm *H′*; where the per cent energy drop is 51-99, 57.5% of the responses confirm *H′*; where the per cent energy drop is 1-50, 76.5% of the responses confirm *H′*; where the per cent energy drop is 0, 80% of the responses confirm *H′*. None of these percentages are significantly different from the average for all sentences. Nor is there any trend among the four groups. From this it is concluded that in full spoken sentences naturally occurring acoustic pauses coinciding with major constituent breaks do not strengthen the role of the breaks in perceptual organization.

In short, some of the sentences exhibit no simple pausal correlate of constituent structure and there is no correlation between the strength of a pause and the strength of the structural effects. Since *all* sentences having one major break show results in the predicted direction, it is evident that the pausal characteristic of the speech signal cannot be the sole factor tending to determine the subjective placement of the clicks.

Ideally, it would be desirable to control not only for pause, but also for other acoustic features which might serve to mark the boundaries of segments. This may be done by using structurally ambiguous sentences as stimulus material. Such sentences may be inserted in contexts which uniquely select one or the other of their possible interpretations. If it can be shown that the displacement of the click is a function of the structure selected, the possibility that acoustic variables determine displacement can be definitely ruled out. For, in the case of ambiguous sentences, the two structures are associated with precisely the same physical signal. Experiments with ambiguous material have been carried out. Their results confirm the hypothesis of the independence of the structure from the acoustics.

Subsidiary Results

The present data suggest that a number of effects other than those predicted by *H′* contribute to determining the perceived position of the click. In the first place, there was a slight tendency, consonant with the finding of Ladefoged and Broadbent, cited above, to perceive the click earlier than its objective position. Fifty-two per cent of all the erroneous responses were to the left of the objective position of the click, 48% to the right. It is of some interest that the tendency to prepose responses was weaker at the end of the test than at the beginning. Of the erroneous responses to the first 20 sentences presented, 55.5% were preposed, but for the last ten sentences only 47% were preposed (independence rejected at $p < .05$ by χ^2). Since order of sentence presentation was randomly varied and the structural prediction for the final ten sentences was balanced, this tendency appears to be a true experience effect.

It is suggested that the failure of some experimenters to replicate the tendency to prepose responses found by Ladefoged and Broadbent is in part attributable to the tendency of that effect to diminish with prolonged experience with the stimulus material. An unpublished paper by Garrett (1964) on the perception of clicks in speech presented binaurally shows no overall tendency for responses to be preposed. Reanalysis of these data reveals, however, a significant tendency to prepose responses for the initial period of the experimental session. This effect is washed out by the later tendency to postpose responses. The fact that Ladefoged and Broadbent's experimental sessions were comparatively short, (8–10 sentences) may thus account for the overall difference between their results and Garrett's and those of others.

It was further discovered that the tendency towards preposed responses is asymmetrical for the two ears. About 43% of the erroneous responses were preposed by *S*s to whom the sentence was presented in the left ear and the click in the right, but 61% of their erroneous responses were preposed by *S*s experiencing the opposite stimulus orientation (independence rejected at $p < .01$ by χ^2). The proportions of preposed responses in each of the orientations were also significantly different from the mean (independence from 50% rejected at $p < .01$ by χ^2). These data are summarized in Table 1.

A second asymmetry between the two stimulus orientations is related to the effect of experience. The drop in the frequency of click preposition in the last part of the session appears to have been largely con-

TABLE 1
PERCENTAGE OF PREPOSED RESPONSES
IN TOTAL ERRORS

Part of session	Right ear	Left ear	Both ears
Initial two-thirds	62.5	46.8	55.5*
Final third	58.8	36.3	47.0*
Total	61.3*	43.5*	52.2

Note: The first two column heads refer to the ear in which the *sentence* was heard. Sums marked (*) are not exact averages owing to slight differences in the absolute number of errors in each orientation.

tributed by a shift in the pattern of responses of *S*s hearing the sentence in the left ear and the click in the right (independence of final third rejected at $p < .05$ by χ^2). These data on asymmetry between the two orientations are also summarized in Table 1. In brief, the asymmetry of the ears found in this experiment may be functionally characterized in the following way: there is a relative delay of *all* material presented to the right ear. This relative delay increases significantly with experience for those *S*s receiving speech in the left ear.

REFERENCES

BOLINGER, D., AND GERSTMAN, L. J. Disjuncture as a cue to constructs. *Word,* 1957, **13**, 246-255.

LADEFOGED, P., AND BROADBENT, D. E. Perception of sequence in auditory events. *Quart. J. Exp. Psychol.,* 1960, **12**, 162-170.

(Received July 27, 1964)