LANGUAGE AND SPEECH, Vol. 31, Part 2, 1988

DETECTION OF TARGET PHONEMES IN SPONTANEOUS AND READ SPEECH*

GITA MEHTA
Cambridge University, U.K.
and
ANNE CUTLER
MRC Applied Psychology Unit, Cambridge, U.K.

(Received March 18, 1988; accepted July 19, 1988)

Although spontaneous speech occurs more frequently in most listeners' experience than read speech, laboratory studies of human speech recognition typically use carefully controlled materials read from a script. The phonological and prosodic characteristics of spontaneous and read speech differ considerably, however, which suggests that laboratory results may not generalise to the recognition of spontaneous speech. In the present study listeners were presented with both spontaneous and read speech materials, and their response time to detect word-initial target phonemes was measured. Responses were, overall, equally fast in each speech mode. However, analysis of effects previously reported in phoneme detection studies revealed significant differences between speech modes. In read speech but not in spontaneous speech, later targets were detected more rapidly than earlier targets, and targets preceded by long words were detected more rapidly than targets preceded by short words. In contrast, in spontaneous speech but not in read speech, targets were detected more rapidly in accented than in unaccented words and in strong than in weak syllables. An explanation for this pattern is offered in terms of characteristic prosodic differences between spontaneous and read speech. The results support claims from previous work that listeners pay great attention to prosodic information in the process of recognising speech.

Key words: speech recognition, speech perception, spontaneous speech, phoneme detection

INTRODUCTION

Most of the speech we listeners hear is spoken spontaneously. In offices, shops, classrooms, and our homes, we are continually presented with speech signals which have been

---

* This study formed part of a thesis project for the degree of M.Phil. in Computer Speech and Language Processing, conducted by the first author under the supervision of the second author. Financial support was provided by British Gas. The authors are grateful to John Culling, Anil Mehta, Ian Nimmo-Smith, and Dennis Norris for their assistance.

Address correspondence to: Dr. Anne Cutler, MRC Applied Psychology Unit, 15 Chaucer Rd., Cambridge CB2 2EF, U.K.

conceived and composed by their speakers even as they are uttered.

Less often, we encounter other styles of speech. In the theatre, through the broadcast media, and elsewhere, we may be presented with rehearsed speech. In news broadcasts, and, all too often, in lectures and talks we may hear read speech. Some of us may occasionally come across computer-synthesised speech. Nonetheless, spontaneous speech accounts for the vast majority of speech signals processed by any listener.

Psycholinguistics has compiled an extensive body of research on human speech recognition. Many laboratory tasks have been devised in the attempt to shed light on the processes by which listeners extract meaning from speech signals. Laboratory studies using such tasks characteristically involve carefully constructed materials, in which the effects of variables irrelevant to the experiment at hand are eliminated or controlled. Since such rigorously designed utterances are very unlikely to be produced spontaneously, laboratory studies usually use speech which is read from a prepared script. Thus the speech mode most frequently encountered in psycholinguistic experiments is *not* the speech mode most frequently encountered outside the laboratory.

Should this mismatch affect our interpretation of the body of findings from laboratory studies of speech recognition? Are there systematic differences between spontaneous and read speech which might lead one to suspect that laboratory findings might not generalise to speech recognition performance outside the laboratory?

The linguistic literature contains many studies comparing the characteristics of speech in different discourse modes. Casual spontaneous speech is marked by numerous phonological elisions and assimilations (Brown, 1977; Labov, 1972; Milroy, 1980) and by syntactic simplifications and, occasionally, incompleteness (Cheshire, 1982; Labov, 1972). The prosodic differences between spontaneous and read speech are particularly well documented. Spontaneous speech tends to be produced at a slower rate than read speech (Barik, 1977; Johns-Lewis, 1986; Levin, Schaffer, and Snow, 1982); it tends to be characterised by longer and more frequent pauses and hesitations (Barik, 1977; Crystal and Davy, 1969; Kowal, O'Connell, O'Brien, and Bryant, 1975; Levin *et al.*, 1982) and shorter prosodic units (Crystal and Davy, 1969); and there is some evidence that pauses may be less closely associated with syntactic structure (Henderson, Goldman-Eisler and Skarbek, 1966; Levin *et al.*, 1982) and that fundamental frequency range may be narrower, at least in intimate conversation (Johns-Lewis, 1986).

However, very little work has specifically investigated *how listeners perceive* spontaneous speech. Admittedly, it has been repeatedly demonstrated that listeners can tell *whether* or not speech was spontaneously produced. For instance, Levin *et al.* (1982) found that listeners were able to tell whether extracts of speech had been taken from spontaneous conversation or read material; moreover, listeners could make this judgement even when the extracts had been low-pass filtered, leaving only prosodic cues to the speech mode. Remez, Rubin and Ball (1985) similarly found that listeners were sensitive to the spontaneous/read distinction. Johns-Lewis (1987) found that listeners performed well at distinguishing between conversational extracts and read or rehearsed (acted) utterances; she also found that the distinction could not be as accurately made on written versions of the sentences alone. This latter finding suggested that listeners were probably basing their judgements on prosodic aspects of the speakers' productions.

We found only three studies which directly addressed the perception of spontaneous speech. McAllister (in preparation) examined word recognition in spontaneous and read speech using the gating task (Grosjean, 1980), in which listeners hear successively larger fragments of a word. She found that the effect of lexical stress differed in the two speech modes: Word identification (in context) occurred earlier for a word stressed on the first rather than on the second syllable in spontaneous speech, but not in read speech. Although the gating task is not a direct measure of speech perception because the word fragments are heard repeatedly, McAllister's study does suggest that there may be some interesting differences in the way listeners recognise spontaneous versus read speech. Bard, Shillcock and Altmann (1988) and Shillcock, Bard and Spensley (1988) report a word-by-word gating study of spontaneous speech, in which they found, *inter alia,* that words containing strong syllables were easier to identify than words which were realised as weak syllables.

In the present study, we attempted to compare the recognition of spontaneous and read speech using an on-line measure of speech processing, the phoneme-monitoring task. In this task, originally developed by Foss (1969), subjects listen for a pre-specified target phoneme in word-initial position in an auditorily presented sentence. They are required to press a response button as soon as they hear the specified phoneme. Response time is held to reflect difficulty of processing at the phonemic level.

Phoneme-monitoring studies, using, of course, read speech materials, have established a number of robust effects (for a review, see Cutler and Norris, 1979). We will describe five such effects which will be of importance to the present study. The first is the effect of transitional probability (Morton and Long, 1976). Phoneme targets on words which are predictable in the sentence context are responded to faster than targets on unpredictable words. For example, subjects detect the target /d/ more rapidly in (1) than in (2):

(1) Quickly he ate his dinner

(2) Quickly he ate his due

Morton and Long's explanation for this finding was that more probable words can be processed more rapidly. In fact, their materials have recently been criticised by Foss and Gernsbacher (1983), who pointed out that the more and less probable words had different vowels. However, Foss and Gernsbacher's vowel-based explanation has in its turn been challenged (Cutler, Mehler, Norris, and Segui, 1987). Moreover, the transitional probability effect has been replicated, with materials differing from Morton and Long's, by Dell and Newman (1980), and, with materials in which vowel identity was controlled, by Mehler and Segui (1987). The effect also appears in the mispronunciation monitoring task (Cole and Jakimik, 1978).

The second effect from previous phoneme monitoring studies which concerns us here is the effect of the length of the preceding word. Mehler, Segui, and Carey (1978) and Newman and Dell (1978) found that phoneme targets preceded by longer words were responded to more rapidly than targets preceded by short words. Their explanation for this was that although long words occupy more time in the speech signal than short words, they require only the same amount of higher level processing. This allows more processing to have occurred by the end of a long word than by the end of a short word, freeing up more attentional capacity for phoneme detection.

The third effect concerns the position of the target in the sentence. Reaction time in phoneme monitoring studies generally decreases across the sentence or clause (Foss, 1969; Shields, McHugh, and Martin, 1974; Cutler and Fodor, 1979). Foss explained this effect in terms of greater contextual predictability of later words (i.e., as a version of transitional probability); Shields *et al.,* however, proposed that the effect reflected increasing rhythmic predictability, as listeners 'lock in' to the rhythmic pattern of the sentence.

Lastly, two effects have been examined which involve the prosodic structure of utterances. Phoneme targets on accented words are responded to more rapidly than targets on unaccented words (Cutler, 1976; Cutler and Foss, 1977; Shields *et al.,* 1974). Cole and Jakimik (1978) further confirmed this finding with mispronunciation monitoring. In line with Shields *et al.*'s interpretation of the sentence position effect, the sentence accent effect is ascribed to listeners' use of prosodic predictability to direct attention to the most highly accented words in an utterance; their motivation for this is that accented words are semantically more central to the utterance's message (Cutler and Fodor, 1979). The prosodic predictability explanation was strongly supported by Cutler's (1976) demonstration that targets on words in accented position are still responded to faster despite acoustic correlates of accent on the word itself being absent.

The effects on phoneme monitoring response time of *syllable* stress, however, are less clear-cut. There are clear processing effects of syllable quality (whether syllables are metrically strong or weak, strong syllables being those containing full vowels, while weak syllables contain reduced vowels; Cutler and Norris, 1988). One can ask, then, whether responses are faster to the /p/ of *petrol,* in which the initial syllable is strong, than to the /p/ of *patrol,* in which the initial syllable is weak. Surprisingly, no phoneme monitoring study has directly posed this question. Shields *et al.* (1974) varied the relative stress level in bisyllabic nonsense words, and found that responses to initial-syllable targets were faster when the initial syllable was stressed. However, even when stress was assigned to the second syllable, the nonsense words had full vowels in both syllables, so this result does not address the question of target detection on weak syllables. Shields *et al.* also report that the difference disappeared when the words were presented in lists rather than in sentences. Taft (1984) similarly found faster responses to targets on real words with initial-syllable stress; however, the vowel quality of the syllables was again not controlled. Some of the target words used by Cutler and Foss (1977) had weak initial syllables, and although this variable was not explicitly manipulated in their study, an inspection of the item means reveals that responses to targets on initially-stressed words were on average some 89 msec faster than responses to targets on words with weak initial syllables. Nevertheless, it is rather difficult to interpret this finding, given that words with weak initial syllables do not appear to be less easy to process *per se* than words with strong initial syllables (Cutler and Clifton, 1984). The experiment by McAllister (in preparation) which was mentioned above, however, found mode-specific effects of syllable stress; this prompted us to look at this question in the current study despite the unsystematic nature of previous findings pertaining to it.

The methodology used in the present study is, of necessity, less rigorous than in most psycholinguistic laboratory studies of speech recognition. One reason why there have been few studies of the perception of spontaneous speech is the very necessity for precise control of materials described above. We have no magic solution to this problem; there is no way to investigate in truly spontaneous speech the effects of the kind of variables just described, if experimental control of the same rigour as in the original investigations of these variables is prescribed. We have compromised in the following manner: By keeping the sentences constant across the spontaneous and read conditions, we can compare the relative magnitude of known (i.e., previously reported) effects across the two modes. Then, if the sentence materials are unbalanced and this imbalance results in an effect altering or disappearing, the consequences of the lack of balance should be equally evident in both modes. On the other hand, where we find a significant difference in any effect between the two modes, this difference should not be due to imbalance in the materials.

By keeping the speaker and the materials constant across the two modes we also enable a direct comparison to be made between overall response time in the two speech modes. Although it would be in a sense surprising to find that read speech and spontaneous speech differed in their intrinsic ease of processing, the reported large differences in speech rate, pause duration and the like do at least allow for this possibility.

METHOD

Materials

An hour and a quarter of spontaneous, unprompted conversation was recorded between two male speakers of British English. The speakers were not acquainted prior to the recording session, and the conversation was chiefly concerned with their respective jobs and leisure interests. The recording session took place in an anechoic, sound-dampened room; the recordings were made on a REVOX tape recorder, at 7½ ips.

A transcript of the recording was made, and 53 sentences were selected for use in the experiment. Thirty of these were chosen to be experimental sentences, while the remainder were used as fillers, warm-up and practice sentences. The experimental sentences are listed in the Appendix.

The sentences selected were all uttered by one speaker. They were grammatically complete, but in that some contained anaphoric pronouns, deictic references and the like, a knowledge of the context in which the sentences were uttered was required in order for them to be fully comprehended. The sentences were randomised in order to minimise the context effects of running conversation.

Each of the stop consonants /p/, /t/, /k/, /b/, /d/ and /g/ formed the target for five experimental sentences. In any experimental sentence, the designated target phoneme occurred only once in word-initial position. However, it was not possible in this spontaneously produced speech to restrict our materials to sentences containing only one occurrence of the target in *any* position; thus the designated target phoneme was frequently present elsewhere in the sentence in word-medial or word-final position. For this reason, practice sentences were chosen to emphasise to subjects that they should respond only to a word initial target.

Fillers were selected such that the designated target phoneme did *not* occur word-initially anywhere in the sentence. These filler sentences were randomly interspersed with the experimental sentences, with the aim of ensuring that subjects did not adopt a strategy of just waiting for the end of each sentence and pressing the response button at that point.

A week after the conversation was recorded, the speaker whose utterances had been selected returned to the laboratory and recorded from a written text all 53 isolated sentences, in the randomised order which was to be used in the experiment. The sentences were read with natural emphasis and intonation, and were recorded under similar acoustic conditions as the spontaneous conversation had been – in the same sound-dampened room, and with the same recording equipment. The speaker was unaware of which sentences contained designated phoneme targets, or what the targets were.

The experimental tapes were constructed from the material selected and recorded as above. Also included were synthetically produced versions of each sentence; these formed part of another experiment and will not be further described here.[1]

There were three experimental tapes, each beginning with the five practice sentences in all three modes (uttered spontaneously, read, and synthesised). Each tape continued with three blocks of 16 sentences, each block consisting of four warm-up sentences (responses to which were not recorded), ten experimental sentences, and two fillers which were placed randomly among the experimental sentences. Each tape was composed such that it contained only one occurrence of each experimental sentence and all three speech modes. Hence, Tape 1 consisted of three blocks of experimental sentences (plus of course the four warm-up sentences and two fillers in each block), of which Block A (sentences 1–10 in the Appendix) was spoken spontaneously, Block B (sentences 11–20) was read, and Block C (sentences 21–30) was synthesised. On Tape 2 Block A was read, Block B was synthesised and Block C spoken spontaneously, while on Tape 3 Block A was synthesised, Block B spontaneous and Block C read.

The REVOX tape recorder's built-in slide synchroniser was used to record a brief tone coincident with the release burst of each word-initial target phoneme, and also with the end of each sentence (of any type). The release burst was located by moving the tape manually back and forth across the playback head. The tone was inaudible on the speech

---

[1] Briefly, the pattern of results for synthetic speech was just like the pattern of results for read speech. However, there is reason to believe that a comparison between detection of phonemes in synthetic and natural speech may be affected by some differences between the two speech modes; for instance, synthetic speech has little coarticulation and less variation in phoneme duration than natural speech. The synthetic speech findings are currently being followed up, and will be reported at a later date.

channel which the subjects heard; the slide synchroniser head, positioned between the two channels of the tape, both recorded and detected the tone.

In order to correct possible errors of alignment between the start of the target phoneme and the tone, arising from the manual method used to lay down the tones, each experimental sentence on each tape was digitised and inspected, using the speech editing facilities of the Cambridge University Engineering Department. The difference between the position of the timing tone and the actual onset of the burst was measured to the nearest millisecond. In all but two sentences, the timing tone was either on or a few milliseconds after the actual burst onset, and the resulting difference was accordingly added to the subjects' response times. In the remaining two instances, the timing pulse had been placed slightly before the burst, and the difference between the two values had to be subtracted from the reaction times.

Independent variables. Finally, prior to the collection of reaction time data, the 30 experimental sentences were examined to ascertain which of the previously reported effects in the phoneme-monitoring literature could reasonably be examined on a *post-hoc* basis in this study. It was, of course, impossible to construct balanced sets of sentences for any variable. The primary consideration in searching the spontaneously produced material to choose sentences for use in the experiment was suitability for the phoneme-monitoring task: The sentences chosen had to have one and only one occurrence of a particular stop consonant in word-initial position. The set of utterances which met this criterion was quite small. However, there was enough variation in the 30 chosen sentences that we were encouraged to undertake *post hoc* analyses of each of the five effects described in the introduction.

It must be noted, of course, that the five variables are *not* orthogonally distributed in these materials. However, as noted above, this criticism applies equally to both speech modes. The only proper consideration of any effect must be a comparison of relative strength of the effect in each mode.

The lack of controlled distribution poses a problem with respect to comparing results of the present experiment with those of earlier studies. The effects we attempt to replicate were established with controlled materials, and hence the original comparisons could be made in a binary fashion, between good exemplars at either end of (for instance) a transitional probability scale. However in the present instance, where exemplars are spread more evenly along a continuum on such measures as transitional probability, a continuous analysis such as a regression analysis is without question more appropriate. But a continuous analysis does not allow direct comparison with previous findings. Our compromise on this issue was to undertake both types of analysis. To allow analyses which were directly comparable with those conducted in previous studies, each variable was first categorised on a binary scale; then, to allow an estimate of the simultaneous relative contribution of the variables in each speech mode, a continuous scale categorisation was also made. In some cases this categorisation process involved pre-tests, as follows:

a) *Transitional probability.* Following the design outlined by Morton (1967), a questionnaire containing each of the thirty experimental sentences up to, but not including, the target bearing word was completed by 20 members of the Applied Psychology Unit, who were asked to continue each sentence with the first word(s) that occurred to them. A score was assigned to each sentence equal to the number of replies that correctly produced the next word which occurred in the experimental sentence. For instance, ten replies had the next word for "Have you got a very big record . . ." to be "collection". As "collection" was indeed the next word in the original experimental sentence, that sentence was assigned a score of ten.[2]

The binary categorisation was achieved by classifying all sentences with a score greater than 5 (i.e., 25% responses) as high transitional probability, while all others were classified as low transitional probability. The continuous scale consisted of the individual item scores (range 0 to 14) on the completion test.

b) *Preceding word length.* The number of syllables in the preceding word was counted. This ranged from one for words like "of", to four in "individual" and "considerable" (as spoken in each instance by the speaker). Since two-thirds of the target-bearing words were preceded by a monosyllable, the binary scale contrasted words whose preceding word was one syllable long vs. two to four syllables long. (Note that Newman and Dell [1978] found an almost linear decrease in response time with increasing length in syllables of the preceding word.) The continuous scale used the full range of one to four.

c) *Position of the target bearing word.* The number of syllables preceding the target bearing word in the sentence was counted. If the target appeared after a clause boundary, only the syllables in the new clause were counted. The target occurred as early as the second syllable in the clause or sentence in some instances, and as late as the sixteenth syllable in others. It was difficult to motivate construction of the binary categorisation from preceding studies, since these used widely differing position contrasts. Foss (1969) contrasted targets which were on average 3.5 *words* into the sentence with targets about ten words in; from the examples of his materials, one might guess that this translates into about five versus fourteen syllables in. Shields *et al.* (1974) contrasted targets in about second syllable position with targets in about sixth syllable position! Cutler and Fodor's (1979) early and late targets, on the other hand, occurred on average 6.1 versus 12.6 syllables into the sentence. Here we struck what we felt to be a reasonable compromise with these earlier measures by contrasting targets occurring in syllable positions 1 to 5 (early) with targets occurring in syllable positions 6 to 16 (late). The continuous scale again employed the entire range of syllable positions from 1 to 16.

---

[2] Note that another method for determining transitional probability has also been used (e.g., by Marslen-Wilson and Welsh, 1978), which includes responses semantically related to the actual word in determining that word's predictability. However, while this method is appropriate where the effects of transitional probability are to be assessed via a task at the semantic level, it is probably less appropriate for use in phoneme-monitoring where the effects of transitional probability are felt via predictability of the precise phonological form of the target word. Thus it may well be the case that subjects hearing (1) may often expect to hear "Quickly he ate his lunch", but it is hard to assess the effects of this expectation on response time to detect the initial phoneme of the actual target word, *dinner.*

d) *Sentence accent.* Four listeners (the authors and two colleagues) transcribed the accent patterns of the experimental sentences. Each sentence was assigned a score equal to the number of listeners (0 to 4) who assigned accent to the target bearing word.The binary scale contrasted sentences in which the target bearing word was judged to be accented by two or more listeners with sentences where the target bearing word was judged accented by no or only one listener. The continuous scale again used the actual score over the full range 0 to 4.

e) *Syllable stress.* In seven of the 30 sentences the target occurred in a word with a weak initial syllable (e.g., "collection" or "believe"); in the remaining 23, the target occurred in a strong syllable (e.g., "difficult" or "power"). Thus for this variable there was only a binary scale (strong vs. weak), and this binary score was therefore also used in the continuous-scale analysis.

Subjects

Thirty-nine subjects were tested, all but one from the Medical Research Council's Applied Psychology Unit subject panel. The subjects were between 22 and 45 years of age and had no particular sophistication in phoneme monitoring experiments. All were native speakers of standard British English. The results of three of the subjects were eliminated from the final analysis; two because they failed to respond to at least two-thirds of the experimental sentences in any one block, and the third because of an unacceptably high error rate in the recognition test. Of the remaining 36 subjects, twelve heard each tape, and of these twelve, four heard each of the three speech mode orders. All subjects from the APU subject panel were paid for their participation in the experiment, which lasted approximately twenty minutes.

Procedure

Subjects were tested individually in a sound-dampened room. They were given written instructions which emphasised speed of response, but also accuracy in responding only to targets in word-initial position. They were also instructed to pay attention to the meaning of the sentences as they would receive a recognition test later.

Subjects heard the speech materials over headphones. The phoneme target for each sentence was displayed upon a VDU screen in front of the subject. Target presentation, timing and response collection were under the control of a dedicated DEC PDP 11/23 minicomputer running the TSCOP experiment control software (Norris, 1984). The timing tone aligned with the onset of each target phoneme was detected by the tape recorder's slide synchroniser, which closed a relay that in turn triggered a timer in the computer; when the subject pressed the response key on detection of a target, the timer was stopped, and the subject's reaction time recorded by the computer.

Subjects first received the practice set, in the speech mode corresponding to the first block of experimental sentences which they would receive. The three experimental blocks followed, with short breaks between blocks. The three blocks within each tape were presented in three counterbalanced orders, one order being heard by one-third of the subjects who heard each tape. This design made it possible to control for the effects of the order of presentation of the three speech modes.

A recognition test was administered to each subject at the end of the experiment. Half the subjects received a test comprising 20 sentences, ten of which had been heard by the subjects exactly, while the other ten were constructed by putting together phrases from more than one (experimental or filler) sentence. This is the standard form of recognition test used in phoneme-monitoring experiments, and is intended to provide a rough check that subjects are indeed attending to the content of the speech material. The test administered to the remaining subjects consisted entirely of sentences actually heard, selected roughly equally from the three subsets; thus about a third of them had been heard by any given subject in each of the three speech modes. Subjects who received this test were told specifically that they had heard only some of the sentences in the test and not others. In both instances, subjects were required to respond "Yes" or "No" to whether or not they had heard the precise wording of the sentence in the test. This second test was intended to ascertain whether there was any difference in recall as a function of speech mode.

RESULTS

The mean score on the recognition test for the first 18 subjects was 60%. This is rather lower than is usually found for such experiments. There was some indication that sentences which subjects had heard in their spontaneously uttered form were less well recalled than sentences heard as read speech: On average, subjects misclassified 46.3% of sentences which they had heard in spontaneous form, versus 37% of sentences which they had heard in read form. However, this difference did not reach significance (t (17) = 1.92, $p < 0.08$); and the other form of recognition test, received by the remaining 18 subjects, also failed to show any statistically significant differences in recall for sentences presented as spontaneous or as read speech; in this case, the overall percent correct was 62%, and subjects misclassified sentences which they had heard in spontaneous form only 3% more often than they misclassified sentences which they had heard in read form.

Response times calculated by the computer were adjusted by the amount of discrepancy between the manually placed timing tone and the actual onset of the burst, by the technique described earlier. The modified response times were then subjected to separate analyses of variance, with subjects (F_1) and sentences (F_2) as random factors respectively.[3]

---

[3] In general, analyses by items are better predictors of the result of a combined analysis across items and subjects simultaneously (Forster and Dickinson, 1976). However, in the present instance, and especially for the *post hoc* analyses to be reported below, the power of the analysis by subjects, in which all 36 subjects heard sentences in both conditions, is rather greater than that of the analysis by sentences, in which the split between sentences in the two levels of a particular binary comparison is sometimes quite uneven. Both analyses are therefore reported, separately. It should also be noted that because the cells of Tables 2–6 contain means of unequal numbers of data points, a weighted calculation is necessary to derive row and column means.

TABLE 1

Response time (msec) to targets in each speech mode as a function of order of presentation

First mode	Spontaneous	Read	Mean
Spontaneous	511	506	509
Read	605	584	594
Synthetic	595	584	589
Mean	571	558	

The overall mean response time to targets in spontaneous speech was 571 msec, to the same targets in read speech 558 msec. The 13 msec difference was not significant (both F_1 and $F_2 < 1$). However, there was a significant effect of the order in which subjects heard the speech modes (F_1 $(2, 27) = 3.67$, $p < 0.04$; F_2 $(2, 58) = 6.75$, $p < 0.01$). This difference can be seen in Table 1: Subjects who heard the initially presented block of sentences (irrespective of whether this was Block A, B or C) in spontaneously spoken mode responded significantly faster across the whole experiment. This effect did not interact with the speech mode effect, nor with which tape (i.e., which combination of sentence blocks with speech modes) the subjects had heard.

This effect will be further discussed below. However, what should be noted at this point is that there is *no* overall effect of speech mode. Phoneme-monitoring response time is not significantly different to targets in spontaneously produced versus read speech. We now turn to the *post hoc* examination of the five previously reported effects. Each of these was first examined in a separate analysis of variance using the binary classifications described above.

Binary analyses

a) *Transitional probability.* The mean response times to high versus low transitional probability sentences presented in spontaneous versus read mode are shown in Table 2. It can be seen that, as in previous studies, targets on words with high transitional probability tended to produce faster response times than targets on words with low transitional probability. This main effect was significant in the analysis by subjects (F_1 $(1, 27) = 5.62$, $p < 0.03$) but failed to reach significance in the analysis by sentences (F_2 $(1, 28) = 2.38$, $p < 0.15$). However, there was no trace of an interaction between this effect and the speech mode effect (both F_1 and $F_2 < 1$).

b) *Preceding word length.* The means for the comparison of speech mode versus this

TABLE 2

Response time (msec) to targets in each speech mode as a function of transitional probability

Transitional Probability	Spontaneous	Read
High	525	531
Low	583	564
diff.	58	33

TABLE 3

Response time (msec) to targets in each speech mode as a function of preceding word length

Length of preceding word	Spontaneous	Read
$>$ 1 syll.	560	527
1 syll.	574	573
diff.	14	46

effect are presented in Table 3. Again, the direction of the effect is as previously found: Targets preceded by longer words tend to be responded to faster. The main effect was again significant in the analysis by subjects (F_1 $(1, 27) = 5.98, p < 0.02$) but was insignificant in the analysis by sentences ($F_2 < 1$). The interaction with speech mode also neared significance in the analysis by subjects (F_1 $(1, 27 = 3.18, p < 0.09)$ but not in the analysis by sentences ($F_2 < 1$). Separate t-tests on the preceding word length effect for spontaneous and read speech modes respectively showed that the effect was insignificant for spontaneous speech (t $(35) = 0.8$) but significant for read speech (t $(35) = 2.71$, $p < 0.01$).

TABLE 4

Response time (msec) to targets in each speech mode as a function of position in sentence

Syllable position	Spontaneous	Read
late	569	536
early	578	659
diff.	9	123

TABLE 5

Response time (msec) to targets in each speech mode as a function of sentence accent

Accent level	Spontaneous	Read
accented	534	570
unaccented	595	553
diff.	61	–17

c) *Position of the target bearing word.* The relevant means are shown in Table 4. The main effect was significant in the analysis by subjects (F_1 (1, 27) = 19.02, $p < 0.001$) and approached significance in the analysis by sentences (F_2 (1, 28) = 3.19, $p < 0.09$). The interaction with speech-mode was significant in both analyses (F_1 (1, 27) = 7.93, $p <$ 0.01; F_2 (1, 28) = 5.19, $p < 0.04$). Separate t-tests on the position effect in each speech mode showed that the effect was insignificant for spontaneous speech (t (35) = 0.33), but in read speech late targets were detected significantly more rapidly than early targets (t (35) = 3.6, $p < 0.001$).

d) *Sentence accent.* The relevant means are presented in Table 5. The main effect approached significance in the analysis by subjects (F_1 (1, 27) = 3.06, $p < 0.09$) but not in the analysis by sentences (F_2 (1, 28) = 1.05). The interaction was significant in the analysis by subjects (F_1 (1, 27) = 8.13, $p < 0.01$) but again not in the analysis by

TABLE 6

Response time (msec) to targets in each speech mode
as a function of syllable stress

Syllable quality	Spontaneous	Read
strong	551	568
weak	629	524
diff.	78	–44

sentences (F_2 (1, 28) = 2.58, $p < 0.15$). Separate *t*-tests showed that accented targets were detected significantly more rapidly than unaccented targets in spontaneous speech (t (35) = 3.52, $p < 0.001$), but the small effect in the opposite direction with read speech was not significant (t (35) = 0.7).

e) *Syllable stress.* The relevant means are presented in Table 6. The main effect was significant in neither analysis (F_1 (1, 27) = 1.46; F_2 (1, 28) = 1.5), but the interaction was significant in both (F_1 (1, 27) = 21.7, $p < 0.001$; F_2 (1, 28) = 9.25, $p < 0.01$). Separate *t*-tests showed that targets on strong syllables were detected significantly faster than targets on weak syllables in spontaneous speech (t (35) = 3.06, $p < 0.005$) but the smaller effect in the opposite direction with read speech did not reach significance (t (35) = 1.99).

Continuous analyses

Because the five variables were not orthogonally represented in the sentence population, we attempted to ascertain whether the combination of the relative contributions of the variables differed across the speech modes. For each of the 30 sentences, the mean response time for each speech mode plus the continuous-scale score for each of the five variables were entered into a regression analysis which contained parameters for differences in the slopes of the regressions for each speech mode. The extra variance accounted for by allowing different slopes and intercepts for the two speech modes (in comparison with having the same slope although different intercepts) was significant (F (5, 48) = 2.98, $p < 0.05$). From this we conclude that there was a significant difference between the two speech modes in the way the five variables jointly affected response time.

The continuous scale also allowed us to analyse the correlation between each sentence's mean response time in each mode on the one hand, and the score on each of the five scales on the other. We would predict that these correlations across items would pattern similarly to the separate *t*-tests for each mode conducted across subjects in the binary analyses, namely that only the effects of preceding word length, position in sentence, sentence accent and syllable stress would differ across the two speech modes. In fact, only three of these variables produced different effects for the two speech modes. As in the t-test comparisons across subjects, the effect of position in sentence was significant for read speech ($r\,(29) = 0.44$, $p < 0.01$ one-tailed) but not for spontaneous speech ($r\,(29) = 0.09$); the effect of sentence accent was significant for spontaneous speech ($r\,(29) = 0.32$, $p < 0.04$ one-tailed) but not for read speech ($r\,(29) = 0.1$), and the effect of syllable stress was likewise significant for spontaneous speech ($r\,(29) = 0.5$, $p < 0.002$ one-tailed) but not for read speech ($r\,(29) = 0.19$). In contrast to the analysis by subjects, however, the effect of preceding word length did not differ across speech modes, being insignificant for both modes.

DISCUSSION

It is a basic premise of psycholinguistic research that laboratory results are intended to extend to conditions outside the laboratory. Thus psycholinguists make general claims about the nature of all speech recognition, independent of speech mode, on the basis of findings from laboratory studies. Yet virtually all laboratory studies of speech recognition are conducted using read materials, whereas most of the speech which listeners encounter outside the laboratory is spontaneously produced. Thus it may be inferred that psycholinguists assume that there are no substantial differences in the way listeners process spontaneous speech versus read speech.

From our findings one may conclude that this assumption, at least in broad outline, is warranted. Firstly, we found no overall difference in response time to detect phonemes in speech which had been read, versus speech which had been uttered spontaneously. The speed with which a phonemic representation can be derived from a speech signal is much the same across those speech modes that are respectively most common inside the laboratory and outside it.

Secondly, there was no difference between the two speech modes on the one semantic variable in our study, the effects of the transitional probability of the target-bearing word. This effect was significant in our study and there was no interaction between the main effect and the speech mode variable. This presumably reflects what common sense would maintain, that the semantic structure of a sentence is not a function of the mode in which it is produced. Listeners' semantic processing of sentences is much the same irrespective of differences in the spontaneity of the sentences' production.

Our study did, however, produce some significant effects of speech mode. On the one hand, the order in which subjects heard the speech modes affected their response times. On the other hand, the effects of position of the target-bearing word in the sentence, of the length of the word preceding the target, of whether or not the target-bearing word was accented and of whether the target began a strong or a weak syllable all differed across the two speech modes.

The speech mode order effect is not easy to interpret. There appears to be a general facilitatory effect of spontaneous speech, in that subjects who heard a first block of sentences in spontaneous mode produced faster responses overall than subjects who heard either a read or a synthesised block first. This response advantage was maintained for both the subsequent blocks heard. Within a particular speech order, there was no significant response advantage for one speech type over another, and the response latencies for all three blocks were roughly equivalent. We suggest two possible, and rather similar, explanations for this finding. Firstly, the speech type which subjects hear at the outset may in a sense prime processing mechanisms such that processing is carried out at a given level of complexity appropriate for the first speech type, and this level is maintained across the speech types that follow. Specifically, the greater prosodic variability of spontaneous speech may encourage efficient use of prosodic cues, for instance. Secondly, it may be that attentional processes for phoneme monitoring get "locked in" at a given level. Cutler *et al.* (1987) have argued that phoneme monitoring responses may be based on lexical or pre-lexical representations, depending on the processing level at which subjects have fixed their attention (which in turn is often determined by the task demands of the experiment). Thus it may be the case that spontaneous, conversational speech is more efficient at arousing listeners' interest and focussing their attention at a level at which the phoneme-monitoring task is most efficiently performed with this particular type of sentence material. In either case, it would appear that listeners' performance at this task is better if they start off with that speech mode which happens to be rather more familiar before proceeding to the rather less common.

The differential effects of the four variables of preceding word length, sentence position, accent, and syllable stress, however, lend themselves more readily to interpretation; moreover, we will argue that these patterns of effects have much in common. Briefly, all four variables have to do with the timing pattern, and other prosodic aspects of utterances, and the differences in their effects across speech modes simply reflect characteristic prosodic differences between the modes. As we discuss each in turn, the similarities will become obvious.

The effect of *length of the preceding word* is statistically the weakest of the four, in that the difference between speech modes was significant in the analysis by subjects but not in the analysis by items. However, there was definitely a larger response time advantage for targets preceded by longer words in read than in spontaneous speech, and it is easy to see why this should be so. Recall that one of the ways in which spontaneous speech differs from read speech is in the pattern of hesitations: These tend to be more frequent and longer in spontaneous speech. Even a cursory hearing of our speech materials confirms that they bear out this pattern. Thus it is much more likely that a particular target-bearing word will be preceded by a hesitation in the spontaneous than in the read mode. Where a target is immediately preceded by a hesitation, any effects of incomplete processing of the previous word will be nullified by the extra processing time provided by the hesitation. Therefore effects of preceding word length, which are held to reflect just such processing hangovers from the preceding word, will be less likely to show up whenever speech is replete with hesitations. In other words, they will be less likely to show up in spontaneous speech.

We tested this hypothesis against the spontaneous materials from our study; indeed, in nine of the 30 sentences there was a clearly perceptible pause immediately preceding the target-bearing word, and in six of these nine sentences the preceding word was monosyllabic. Thus it is reasonable to suppose that these hesitations did indeed negate the effects of preceding word length that appeared with the read speech in which there were no such marked discontinuities.

The other effect which disappears in spontaneous speech is the effect of *position in the sentence,* and the explanation for this disappearance is essentially the same. Recall that two explanations have been offered in the literature for this effect: Foss' (1969) claim that targets on late-occurring words are detected more rapidly than targets on early-occurring words because late-occurring words are *contextually* more predictable, and Shields *et al.*'s (1974) claim that such words are *rhythmically* more predictable. If Foss' semantic explanation were valid, we would expect this effect to pattern like transitional probability, which is a semantic effect (and shows no difference between the two speech modes). It does not. On Shields *et al.*'s account, however, we can explain the difference between speech modes. The greater frequency of hesitations in spontaneous speech, which results in shorter prosodic units, inevitably reduces the average span over which rhythmic predictability will hold. Every time a new prosodic unit is initiated, the rhythmic processor has to be reset, so to speak. On this account, position in the sentence is not, strictly speaking, what affects target detection time; rather, the effective variable is position in the *prosodic unit.* And because prosodic units are long – generally clause-length – in read speech, but usually short in spontaneous speech, the opportunity for rhythmic prediction in the latter case is much smaller. Thus effects of sentence position of the kind previously reported are unlikely to show up in spontaneous speech because the appropriate conditions for their appearance rarely exist.

Note that this account not only explains the difference in this effect across speech modes in the present experiment, but incidentally supports Shields *et al.*'s prosodic explanation of the previously reported effect in contrast to Foss' semantic account.

Thus the same story – the greater prosodic fragmentation of spontaneous speech – underlies both cases in which a previously reported effect is replicated in read speech but fails to appear in spontaneous speech. In the case of the remaining two variables, the pattern was reversed – both for *sentence accent* and for *syllable stress* we found a significant effect on phoneme detection time in spontaneous speech, but none in read speech. Since both these variables are unquestionably prosodic in nature, it is once more to prosodic structure that we look for an explanation.

Firstly, we should note one puzzling feature of the *sentence accent* results. A strong response time advantage for targets on accented words has previously been reported for read speech (Cutler, 1976), yet we failed to find any such advantage with read speech in the current study. The solution to this puzzle lies in the nature of what has been termed accent in each case. In Cutler's (1976) study the accented target-bearing words were contrasted, as in (3):

(3) She managed to remove the DIRT from the rug, but not the BERRY stains.

In the present study, no such contrasts occurred in the materials (with the possible exception of sentence 14: see Appendix). Moreover, it should be noted that the raters who judged where accent fell in the sentences agreed that spontaneous and read utterances differed: All four raters assigned accent to more words in spontaneous utterances than in read utterances, and there was a strong tendency in read utterances (but not in spontaneous utterances) for accent to be judged to have occurred in the default sentence-final position. This suggests that the predictive prosodic processing which underlies the effect of sentence accent on phoneme detection time does not come into play in relatively neutral prosodic contours in which sentence accent is assigned by default. Rather, the processing is specific to those situations where accent is performing a semantic function, expressing focus on a particular lexical item (as, indeed, was specifically argued by Cutler and Fodor, 1979). Given that the accent patterns of spontaneous utterances were more varied and less likely to express default accenting than the accent patterns of read utterances, it is again no surprise at all that we find a difference between the two speech modes; again, an effect is found where the prosodic conditions for its occurrence are best met.

Finally, we found a strong effect of *syllable stress* in this study, but it too appeared only in spontaneous utterances; in these, but not in read utterances, targets in strong syllables were detected faster than targets in weak syllables. As we recounted in the introduction, there is some uncertainty about whether exactly this effect has previously been demonstrated in phoneme monitoring studies. But what is particularly interesting is the parallel of the present finding to that of McAllister (in preparation), also described in the introduction. In McAllister's study, gated words in sentence context were recognised earlier when they began with strong rather than weak syllables, but only if the words in question had been spontaneously produced. No such effects were found for read speech. Exactly the same disparity between speech modes was found in our phoneme detection results.

This too seems explicable in terms of differential characteristics of spontaneous and read speech. Read speech is, in general, more carefully articulated than spontaneous speech; in the latter, phonemes are more often elided, assimilated and in one way or another distorted (Brown, 1977; Labov, 1972; Milroy, 1980). The portions of speech most susceptible to distortion are weak syllables (Browman, 1978), which in turn suggests that there may actually be a larger intelligibility difference between strong and weak syllables in spontaneous speech than in read speech. This would obviously account for why a response time advantage for strong syllables should be more likely to appear in spontaneous than in read speech.

All the differences that we found between the two speech modes, therefore, are easily accounted for by reference to well-known differences between these two types of speech. This suggests in turn that, as we maintained earlier, there are no real differences in the way listeners process speech produced spontaneously or read. Our results support the conclusion, based on much previous work, that listeners are highly sensitive to prosodic characteristics of spoken utterances, and use prosodic information in a sophisticated way in processing speech. This is true in read speech, where for instance listeners can be seen to exploit long prosodic units to engage in rhythmic prediction, and it is true of spontaneous speech, where they can be seen to make use of the informativeness of accent placement.

But because of the substantial prosodic differences between typical spontaneous speech and typical read speech, in practice the *type* of prosodic processing that listeners will have an opportunity to engage in will differ across speech modes. Thus our study may, after all, hold implications for psycholinguistic research practice. Psycholinguists are indeed quite justified in generalising from laboratory studies to speech recognition in everyday life, in that the processing strategies demonstrated in the laboratory belong to the repertoire which listeners can bring to bear on speech input when the opportunity arises. *But the opportunities which arise differ with speech mode.* In particular, the most commonly encountered speech mode, spontaneous speech, offers rich opportunities for the exercise of some strategies and only very limited opportunity for the exercise of others. Therefore psycholinguists may to some extent be misdirecting their efforts, and studying effects which occur relatively rarely at the expense of effects that are in fact much more common. Moreover, there may even exist processing strategies of particular and exclusive usefulness for hesitant, prosodically variable speech that are as yet undreamt of in our laboratories.

REFERENCES

BARD, E.G., SHILLCOCK, R.C., and ALTMANN, G. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception & Psychophysics,* in press.

BARIK, H.C. (1977). Cross-linguistic study of temporal characteristics of different types of speech materials. *Language and Speech,* **20,** 116-126.

BROWMAN, C. (1978). Tip of the tongue and slip of the ear: Implications for language processing. *UCLA Working Papers in Phonetics, 42.*

BROWN, G. (1977). *Listening to Spoken English.* London: Longman.

CHESHIRE, J. (1982). *Variation in an English Dialect.* Cambridge: Cambridge University Press.

COLE, R.A., and JAKIMIK, J. (1978). Understanding speech: How words are heard. In G. Underwood (ed.), *Strategies of Information Processing* (pp. 67-116). New York: Academic Press.

CRYSTAL, D., and DAVY, D. (1969). *Investigating English Style.* London: Longman.

CUTLER, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics,* **20,** 55-60.

CUTLER, A., and CLIFTON, C.E. (1984). The use of prosodic information in word recognition. In H. Bouma and D.G. Bouwhuis (eds.), *Attention and Performance X: Control of Language Processes* (pp. 183-196). Hillsdale, NJ: Erlbaum.

CUTLER, A., and FODOR, J.A. (1979). Semantic focus and sentence comprehension. *Cognition,* **7,** 49-59.

CUTLER, A., and FOSS, D.J. (1977). On the role of sentence stress in sentence processing. *Language and Speech,* **20,** 1-10.

CUTLER, A., MEHLER, J., NORRIS, D., and SEGUI, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology,* **19,** 141-177.

CUTLER, A., and NORRIS, D. (1979). Monitoring sentence comprehension. In W.E. Cooper and E.C.T. Walker (eds.), *Sentence Processing: Psycholinguistic Studies Presented to Merrill Garrett* (pp. 113-134). Hillsdale, NJ: Erlbaum.

CUTLER, A., and NORRIS, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance,* **14,** 113-121.

DELL, G.S., and NEWMAN, J.E. (1980). Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behavior,* **19,** 609-623.

FORSTER, K.I., and DICKINSON, R.G. (1976). More on the language-as-fixed-effect fallacy: Monte Carlo estimates of error rates for F_1, F_2, F' and min F'. *Journal of Verbal Learning and Verbal Behavior,* **15,** 135-142.

FOSS, D.J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior,* **8,** 457-462.

FOSS, D.J., and GERNSBACHER, M.A. (1983). Cracking the dual code: Toward a unitary model of phonetic identification. *Journal of Verbal Learning and Verbal Behavior,* **22,** 609-632.

GROSJEAN, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics,* **28,** 267-283.

HENDERSON, A., GOLDMAN-EISLER, F., and SKARBEK, A. (1966). Sequential temporal patterns in spontaneous speech. *Language and Speech,* **9,** 207-216.

JOHNS-LEWIS, C. (1986). Prosodic differentiation of discourse modes. In C. Johns-Lewis (ed.), *Intonation and Discourse* (pp. 199-219). London; Croom Helm.

JOHNS-LEWIS, C. (1987). The perception of discourse modes. In M. Coulthard (ed.), *Discussing Discourse.* University of Birmingham: *Discourse Analysis Research Monograph No. 14,* 249-271.

KOWAL, S., O'CONNELL, D., O'BRIEN, E.A., and BRYANT, E.T. (1975). Temporal aspects of reading aloud and speaking: Three experiments. *American Journal of Psychology,* **88,** 549-569.

LABOV, W. (1972). *Sociolinguistic Patterns.* Philadelphia: University of Pennsylvania Press.

LEVIN, H., SCHAFFER, C.A., and SNOW, C. (1982). The prosodic and paralinguistic features of reading and telling stories. *Language and Speech,* **25,** 43-54.

MARSLEN-WILSON, W.D., and WELSH, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology,* **10,** 29-63.

MCALLISTER, J. (in preparation). *Lexical Stress and Lexical Access.* PhD. Thesis, University of Edinburgh.

MEHLER, J., and SEGUI, J. (1987). English and French speech processing. In M.E.H. Schouten (ed.), *The Psychophysics of Speech Perception,* (pp. 405-418). Dordrecht: Martinus Nijhoff.

MEHLER, J. SEGUI, J., and CAREY, P.W. (1978). Tails of words: Monitoring ambiguity. *Journal of Verbal Learning and Verbal Behavior,* **17,** 29-35.

MILROY, L. (1980). *Language and Social Networks.* Oxford: Blackwell.

MORTON, J. (1967). Population norms for sentence completion. Unpublished manuscript, Applied Psychology Unit, Cambridge.

MORTON, J., and LONG, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior,* **15,** 43-51.

NEWMAN, J.E., and DELL, G.S. (1978). The phonological nature of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning and Verbal Behavior,* **17,** 359-374.

NORRIS, D.G. (1984). A computer-based programmable tachistoscope for non-programmers. *Behavior Research Methods, Instrumentation and Computers,* **16,** 25-27.

REMEZ, R.E., RUBIN, P.E., and BALL, S. (1985). Sentence intonation in spontaneous utterances and fluently spoken text. Paper presented to the Acoustical Society of America, 109th Meeting, Austin, Texas, April 1985.

SHIELDS, J.L., MCHUGH, A., and MARTIN, J.G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology,* **102,** 250-255.

SHILLCOCK, R.C., BARD, E.G., and SPENSLEY, F. (1988). Some prosodic effects on human word recognition in continuous speech. *Proceedings of SPEECH '88: Seventh Symposium of the Federation of Acoustical Societies of Europe;* Vol. 3, pp. 819-826.

TAFT, L. (1984). *Prosodic Constraints and Lexical Parsing Strategies.* Unpublished PhD. thesis, University of Massachusetts.

APPENDIX

Experimental sentences

Each sentence is preceded by its phoneme target. The numbers in parentheses after each sentence are, in order: (a) mean RT as spontaneous speech; (b) mean RT as read speech; (c) transitional probability rating; (d) accent score as spontaneous speech; (e) accent score as read speech. Preceding word length, position in sentence and syllable stress values can be readily computed.

1. /d/ I've always heard of Cambridge described as such I think. (558, 539, 0, 3, 4)
2. /b/ The most important thing is to buy the right make. (529, 619, 0, 1, 0)
3. /g/ Spock gradually learns how to swear. (486, 914, 0, 0, 0)
4. /p/ They have always allowed us to publish thus far. (488, 534, 0, 2, 2)
5. /t/ I think you shouldn't buy either of them for the time being because at the moment they're incompatible with each other. (426, 444, 2, 0, 0)
6. /k/ Apparently there was a considerable exodus around 1979. (533, 449, 0, 0, 0)
7. /d/ It makes quite a considerable difference to be running behind someone else. (431, 561, 14, 4, 4)
8. /b/ One of the things Dolby does is that it boosts up the high frequency. (622, 491, 0, 4, 3)
9. /g/ Just a set of words that had the words in groups of six or seven. (617, 680, 0, 1, 0)
10. /p/ The best people in the world are just under 45 seconds. (585, 651, 0, 1, 0)
11. /t/ It only ran for three years, when it was on television, and they haven't made any since. (627, 610, 0, 1, 1)
12. /k/ Have you got a very big record collection? (588, 510, 10, 0, 1)
13. /d/ Everything seems to be very democratic here. (558, 704, 0, 4, 4)
14. /b/ You'd have to turn it from digital back into analog. (496, 411, 0, 3, 0)
15. /g/ The research which is going on here is pretty fundamental. (567, 464, 3, 1, 0)
16. /p/ The intrinsic idea of having only a single power supply and running everything off it is a good idea. (583, 486, 0, 3, 3)

17. /t/ I wouldn't be surprised if there was quite a big effect of tactical voting. (568, 497, 0, 0, 0)

18. /k/ All people who call themselves psychiatrists are in fact medics. (512, 541, 0, 0, 0)

19. /d/ When you're doing it for yourself, there isn't really much of a batch. (566, 560, 0, 2, 0)

20. /b/ They have the right to stop us publishing I believe. (747, 577, 0, 0, 0)

21. /g/ They were very psychological, I grant you that. (595, 575, 0, 1, 0)

22. /p/ Individual researchers have their individual projects and get on with them. (647, 573, 0, 1, 3)

23. /t/ That makes the plates move together and apart again, and pushes the air back and forth. (835, 605, 0, 0, 2)

24. /k/ You'll have to accept that something better may come along which you won't be able to use. (465, 534, 6, 3, 0)

25. /d/ She stands a better chance of defeating the Conservative. (531, 531, 0, 2, 0)

26. /b/ If you've seen them in the shops, you'll see that they're very big and very flat. (498, 540, 0, 2, 2,)

27. /g/ I don't think it's a very good trend towards the American way of doing things. (545, 478, 9, 0, 0)

28. /p/ The electric charges on the two plates are varied by the amplifier. (582, 538, 5, 1, 3)

29. /t/ What tests can be done on attention? (625, 674, 0, 2, 3)

30. /k/ Daley Thompson is just over 45 seconds, which is amazing considering he has nine other events to do. (882, 461, 4, 0, 0)