Journal of Phonetics (1981) 9, 407–424

Lexical retrieval from fragments of spoken words: beginnings vs endings

Sieb G. Nooteboom

Institute for Perception Research, Eindhoven, The Netherlands and
Department of General Linguistics, Leyden University, Leyden, The Netherlands

Received 20*th September* 1980

Abstract: The present paper examines the relative contribution of initial and final fragments of spoken words to lexical retrieval. An experiment is reported in which each stimulus is either an initial or final fragment, and contains just enough information to distinguish the intended word uniquely from all other words in the lexicon. Initial fragments give a probability of correct retrieval of 0.89, final fragments of 0.61. These values are essentially the same when the subjects have already been presented with the complementary fragments of the same words but then failed to retrieve them. When they did, earlier in the session, retrieve the words from the complementary fragments, the probability of correct retrieval increases to 0.99 for initial and 0.92 for final fragments. Response latencies are generally shorter for initial than for final fragments. For initial fragments response latencies are the same whether the word is retrieved for the first or for the second time, and thus, in this case, do not correlate with the probability of correct retrieval. The main findings of this experiment support a model of lexical access based on first-order context-sensitive coding of speech with a weighting of sensory information according to its position in the word. Such a model is exemplified by Marcus (1980). It is further concluded that the decision process in word retrieval is monitored in a more complex way than can be accounted for by a threshold-type model of the kind proposed by Morton (1969). Finally, some predictions are made with respect to the distribution of information over lexical items in the languages of the world.

Introduction

Imagine that one has to discriminate between two sequences of symbols, *AB* and *CD*. The difference between these two objects is, of course, redundantly coded. Such redundancy is advantageous in a noisy communication situation: if one misses the second symbol, discrimination can be based on the first and vice versa. If the first and second symbol are equally likely to be destroyed by noise, the optimal strategy would be to pay equal attention to both symbol positions. Conversely, if we find that subjects, confronted with this task, pay more attention to one of the symbols than to the other, it is reasonable to assume that they have good reasons for doing so. Such a deviation from what at first sight would seem to be an optimal strategy may give interesting hints about the perceptual strategies brought to bear on the task at hand.

In this paper we are not concerned with discrimination between simple sequences of

0095–4470/81/040407 + 18 $02.00/0

symbols, but with perception of spoken words. Recently an increasing amount of attention has been given to words as primary units in the perception of speech, both in experimenting and in theorizing (Browman, 1978; Cohen, 1980; Cole & Jakimik, 1978; Goldstein, 1978; Klatt, 1979; Marcus, 1980; Marslen-Wilson & Welsh, 1978). We will examine some properties of a few current models of word recognition, confronting these models with the results of a simple word perception experiment, in which we focus attention on the relative contribution of isolated initial and final fragments of spoken words. These fragments are chosen in such a way that each of them uniquely and non-redundantly distinguishes the intended word from all other words in the lexicon. One can easily imagine an efficient perceptual strategy that would retrieve the intended words equally well from initial or from final fragments. There are several reasons to believe that human listeners do not behave in this way but rather more easily retrieve words from initial than from final fragments.

Giving prior attention to word beginnings as a general strategy in retrieving lexical items might be advantageous because (1) word beginnings have to be produced first in speaking and writing and reach the listener first in perception, giving them special significance for rapid recognition, (2) in languages like English and Dutch word beginnings contain more information, are less redundant, than word endings, and (3) in many languages word beginnings are less likely to be mutilated by assimilation and coarticulation than word endings.

Superiority of word beginnings over endings was already demonstrated at the turn of the century by William Chandler Bagley in the first large-scale experiment on the recognition of spoken words (Bagley, 1900). Using an Edison Phonograph, Bagley recorded, both without and with context, a great number of words in which an initial, medial or final consonant had been deleted by deliberate mispronunciation. He played his stimuli to listeners for recognition, and found, among other things, that deletion of word initial consonants was more disruptive than deletion of word final consonants. Such effects of a superiority of word beginnings as cues to lexical retrieval are not limited to the recognition of spoken words. They are found in a number of tasks all involving retrieval of words from the lexicon. So, for example, studies of the "Tip of the Tongue" phenomenon reveal that, when in word finding a target word is substituted by a word of similar sound, the initial element of the word has a higher probability of being reported correctly than the final element (Brown & McNeill, 1966; Browman, 1978). Inspection of a corpus of word substitutions in spontaneous errors of speech, as published by Fay & Cutler (1977) also shows that target and error words more often agree in the initial phone than in the final phone. Bruner & O'Dowd (1958) showed that in visual word recognition with brief exposures reversal of the initial two letters of a word (vaiation for aviation) was more disruptive than reversal of the final two letters. Broerse & Zwaan (1966) found that their subjects, in guessing visual words from initial or final fragments, took less time with initial than with final fragments, reproduced more different words with initial than with final fragments, even when final fragments allowed more different words, and came up more easily with the intended word for initial than for final fragments, also when both contained nominally the same information. Similar results were obtained by Horowitz *et al.* (1968) in an experiment on recall of visual words, prompted by word fragments. A differential effect of initial and final parts of words is also found in lexical decision (classification as word or non-word) with visual stimuli, in an experiment by Taft & Forster (1976) who demonstrated that the classification of a stimulus as a non-word is slowed down when the initial part of a stimulus corresponds to a word (e.g. footmilge) but not when the final part of the stimulus corresponds to a word (e.g. trowbreak). It should be noted that, though in all these tasks in-

volving lexical retrieval word beginnings seem to be more important than word endings, whenever medial parts of words are also taken into account, these are found to be even less important than word endings. For the normal rapid recognition of spoken words one would expect a superiority of word beginnings over endings, because word beginnings come in first and are therefore least redundant, but not a superiority of word endings over medial parts of words, because word endings come in last, and are therefore most redundant. A considerable number of experimental studies directed at the real time nature of the recognition of spoken words, looking at the detectability of mispronunciations, fluent restorations of mispronunciations in speech shadowing, word, rhyme, and semantic category monitoring, phoneme monitoring, and lexical decision, all converge in supporting the relative importance of the word onset (Cole, 1973; Cole & Jakimik, 1978; Marslen-Wilson, 1973, 1975, 1976, 1978; Marslen-Wilson & Tyler, 1975; Marslen-Wilson *et al.*, 1978; Marslen-Wilson & Welsh, 1978). Browman (1978) presents some data on perceptual confusions of words in spontaneous speech which show that the contribution of stimulus factors to word recognition decreases from word onset to offset, as one would expect in a real time word recognition procedure.

The special role of word onsets in the recognition of spoken words is emphasized in the so-called Cohort Model of word recognition (cf. Marslen-Wilson, 1978), first formulated by Marlsen-Wilson & Welsh (1978). According to this theory word recognition is mediated by a whole array of parallel independent word recognition elements. The acoustic onset of a word activates all word recognition elements which accept this acoustic onset as part of their internal specification. Each of the activated word candidates then continues to monitor the input and is supposed to respond actively to any mismatch between the input and its own internal specification. When the input diverges more than a critical amount from the internal specification of the word candidate, the latter will remove itself from the cohort of candidates. As the acoustic signal develops in time more and more word candidates will remove themselves from the cohort until only one remains. At that point recognition has taken place. For many, especially polysyllabic, words this will be long before the acoustic end of the word has been heard, due to redundancy in the structure of the lexicon. An example is the English word penguin. There are thousands of words starting with /p/, but after the initial *CV* /pɛ/ has been heard correctly the number of word candidates is reduced to about a hundred. When /pɛɛ̃/ has been heard some 50 of these remain. When also the /ŋ/ has come in there is only one word candidate left, because penguin is the only English word starting with /pɛɛ̃ŋ/: the recognition point of penguin is immediately after /ŋ/ (phonemic symbols are used in this example for the sake of convenience. It is not implied that phonemes necessarily mediate word recognition). Syntactic and semantic information can, by precluding candidates from the cohort, shift the recognition point towards the word onset.

The Cohort Model shares with Morton's (1969) Logogen Model the numerous parallel, independent recognition elements. Morton's logogens, however, accept only positive information. In the Cohort Model positive information is only used to activate a limited number of word candidates. After this initial activation only negative information is used. In the Cohort Model, furthermore, recognition does not, as in the Logogen Model, result from an excitation level exceeding a threshold. Instead, recognition follows automatically when only one word candidate is left in the cohort. As a corollary of this, the Cohort Model does not so easily predict word frequency effects.

The purpose of the present experiment is twofold. First of all it is intended to study the retrieval of words from fragmentary information, and examine the relative contribution of

initial and final fragments of spoken words to retrieval. More specifically, the experiment will enable us to find out whether the Cohort Model, that makes some very strong predictions, can deal with word perception from fragmentary information, or whether some other model of lexical coding and lexical access from acoustic information is more suited here. The basic idea of this part of the experiment can be explained as follows. Imagine a vocabulary consisting of two items *AB* and *CD*, *A* and *C* standing for the initial parts and *B* and *D* for the final parts of spoken forms of these two words. The Cohort Model predicts that hearing *A* or *C* would be enough for the normal fast recognition of the words: the recognition points of both words follow immediately after the initial part. On the other hand the Cohort Model cannot handle the situation that only *B* or *D* is heard because the initial parts of the words, necessary to activate the word candidates in the lexicon, would be lacking: if the Cohort Model were the only possible way of recognizing words, final fragments would not lead to recognition at all. Testing these predictions for real spoken words from a real lexicon is the first purpose of the experiment.

A second purpose is to provide a rather strong test case for one of the basic properties of Morton's Logogen Model. In this model it is assumed that when the excitation level of a logogen (word recognition element) is raised by stimulus information, but not enough to reach the logogen's threshold, it will immediately return to normal, so that no transfer will be found from one stimulation to the next. In terms of our abstract example: if one has heard *A* but not recognized the word *AB*, this will not affect the probability of recognizing *AB* from hearing *B* somewhat later. On the other hand, the Logogen Model also states that whenever a logogen has exceeded its threshold, its excitation level will only very slowly decrease, so that for quite a while, say up to 1 h, little is needed to let the logogen exceed its threshold for a second time. Thus, having recognized *AB* from hearing *A*, one would be very much more likely to recognize *AB* from hearing *B* a little later. These predictions will also be tested in the present experiment.

Before the experiment is described a few other points have to be made. The first of these concerns what is meant with the term 'word' in 'word recognition'. Ultimately this is an empirical question dealing with the contents of the mental lexicon. Here we will assume that the units to be recognized are stem-morphemes rather than what are traditionally called words. Thus it is assumed that recognizing penguin, penguins, penguinlike, penguinize, etc., each time involves recognition of the lexical item penguin. The existence of inflectional, derivational or compound forms does not affect the position of Marslen-Wilson's recognition point.

Secondly it should be kept in mind that lexical access may also go via prosodic properties of the sound form, if available. It is particularly revealing that in tip-of-the-tongue data (Brown & McNeill, 1966; Browman, 1978), in malapropisms (Fay & Cutler, 1977) and in perceptual confusions (Kozhevnikov & Chistovich, 1965; Blesser, 1969; Browman, 1978; Garnes & Bond, 1980) the number of syllables and placement of lexical stress is preserved in the overwhelming majority of cases.

Fay and Cutler in particular suggest that the major partitioning of the mental dictionary is by number of syllables, with stress as a second categorization. In the present experiment subjects are deprived of any accurate information on number of syllables, although not from some information on stress placement. Thus the results will have little to say on prosody as a cue to word recognition. Of course, Marslen-Wilson's Cohort Model precludes the use of the total number of syllables as a cue to word recognition, at least for those words that are recognized before the acoustic end has sounded.

Method

Selection of test words

The experiment is based on the notion of recognition point. This, as explained earlier, is the point in the word at which, going from left to right, enough information exists to distinguish the word uniquely from all other words in the lexicon. Of course, logically it is equally well possible to define a recognition point on a right-to-left basis. In the same way as the left-to-right recognition point can be established from a dictionary which is alphabetically organized in the normal left-to-right way, the right-to-left recognition point can be established from a retrograde dictionary. If, by accident, for a given word the left-to-right and the right-to-left recognition points coincide, one can divide this word into an initial and a final part which both contain precisely enough information to distinguish the word from all other words in the lexicon. An example is the Dutch word surrogaat (/sœro:χ'a:t/). If we divide a well pronounced token of this word into two parts, making the cut in the middle of the /o:/, we obtain a word fragment which is heard as /sœro:/ and a fragment heard as /o:χ'a:t/. Surrogaat is the only Dutch word beginning with /sœro:/ and also the only one ending with /o:χ'a:t/. Inspection of a normal Dutch dictionary (Kruyskamp, 1961) and its retrograde version (Nieuwborg, 1978), gave fourteen words in which the two recognition points coincide in the middle of a long vowel (Table I). The latter restriction was necessary in order to make the phone containing the recognition point recognizable from both fragments. This procedure enabled us to compare the contribution of initial and final parts of words to spoken word perception, when there was no difference in information between the two word fragments. One may note that only two of the words have the lexical stress on the first syllable, against ten with stress on the final syllable, and two with stress on the medial syllable.

Table I Fourteen Dutch words used in the experiment. Each of these words can be uniquely determined from both the fragment preceding and the fragment following the midpoint of the underlined vowel segment

Mozaïek	/mo:za:'ʔik/
Kiosk	/ki'jɔ:sk/
Karnaval	/'kɑrna:vɑl/
Ridicuul	/ridi:'kül/
Esculaap	/ɛskü'la:p/
Unaniem	/üna:'nim/
Passagier	/pɑsa:'ʒi:r/
Reünie	/re:jü'ni/
Kannibaal	/kɑni:'ba:l/
Albatros	/'ɑlba::trɔs/
Kandidaat	/kɑndi:'da:t/
Volume	/vo:'lümə/
Gymnasium	/χɪm'na::ziœm/
Surrogaat	/sœro:'χa:t/

Stimulus materials

The fourteen Dutch words in Table I, together with a number of other words having coinciding recognition points within a consonant segment, were spoken by a male speaker of Dutch with a trained voice and accurate pronunciation of Dutch. His realizations were digitally stored on disk with 20 kHz sampling frequency and 12 bit PCM coding. Low pass

Table II Phonemic transcriptions of the stimuli used in the experiment

Series A	Series B
/mo:za:/	/sœro:/
/ki'jɔ/	/'kɑrna:/
/a:vɑl/	/'ɔsk/
/ridi/	/ɛskü/
/ü'la:p/	/a:trɔs/
/a:'nim/	/i'kül/
/pɑsa:/	/vo:'lü/
/o:'χa:t/	/a'ʒi:r/
/re:jü/	/χɪm'na:/
/i'ba:l/	/a:'ʔik/
/'ɑlba:/	/kɑni/
/kɑndi/	/üna:/
/'ümə/	/i'da:t/
/'a:ziœm/	/ü'ni/

filtering with a cut-off frequency at 9 kHz was applied both at recording and at later playback to avoid alias components. With the help of a computer facility for editing of the speech waveform the digitized words were divided into two word fragments. The cut was made under visual and auditory control, in the middle of the segment containing the recognition point, on the assumption that this segment was recognizable from both word fragments. This assumption appeared to be correct when the recognition point fell in a vowel segment. For each word fragment the missing part of the word was replaced by a pure tone of 0.9 kHz and 400 ms duration, in order to signal whether the initial or the final part was missing. From the stimuli obtained in this way, two stimulus tapes were prepared, organized as follows: the fragments from the words in Table I were organized in two series of stimuli, each having seven initial and seven final word fragments (Table II). The word fragments in series B were the complements of those in series A, but in a different order. Tape I contained series A immediately followed by series B, and tape II contained series B immediately followed by series A. Thus each tape contained 28 stimuli of which the first fourteen (henceforth referred to as the "first presentation") found their complements in the second fourteen (henceforth referred to as the "second presentation"), in a different order. Likewise, the first fourteen stimuli of one tape found their complements in the first fourteen stimuli of the other tape, in a different order. This stimulus sequence on each tape was preceded by eight other word fragments, four initial and four final, having the cuts in a consonant segment. These fragments were identical for the two tapes. Their sole purpose was to let the subjects get used to the experimental procedure and their task.

Subjects

The subjects were 60 Dutch speaking university students, most of them from the physics and electronics departments. They were paid for their services. All subjects were tested for hearing defects, and any one having a hearing loss of more than 10 dB in one or both ears somewhere below 8 kHz was replaced. All subjects were ignorant as to the purpose of the experiment. Thirty subjects listened to tape I, 30 other subjects to tape II.

Procedure

Each subject was tested individually in a second-treated booth (IAC 400 A). Stimuli were played to him binaurally through head phones (Sennheiser HD424) from a tape recorder

(Revox A77). The subject was instructed to make a complete Dutch word from the fragment heard and to respond as quickly as possible by pronouncing this word. If the subject failed to find a fitting word he was to reproduce the stimulus he had heard as accurately as possible. In the learning phase, using the first eight word fragments, he was familiarized with the structure of the stimuli and of the task, if necessary with feedback from the experimenter. After this learning phase the experimenter had no further contact with the subject until the end of the session. The tape recorder stopped automatically after each stimulus and was started by the experimenter only after the subject had given a response (which was either a correct or incorrect reconstruction of the intended word, or a correct or incorrect reproduction of the stimulus as a nonsense word).

Each stimulus was not only played to the subject but also to a second tape recorder to be recorded on one track of a magnetic tape. The subject's response was recorded on the other track. This tape was used for later transcription of the responses and for measuring the response latencies. All responses were transcribed by the present author in a broad phonetic transcription. For each response the response latency defined as the time interval between the end of the stimulus word fragment and the onset of the response was measured by hand from a high quality oscillogram.

Results

First presentation: distribution of responses

Responses to the first fourteen stimuli of each tape bear on the main issue of this paper, the relative contribution of initial vs final parts of spoken words to word retrieval. They were classified as:

(1) correct responses, i.e. all cases in which a subject came up with the intended word (or, rarely, an inflectional or derivational form of this word);

(2) correctly heard but not retrieved: a stimulus was judged to be correctly heard when the subject, failing to find a fitting word, correctly reproduced the stimulus as a nonsense word, or when he came up with a fitting word or word combination other than the intended word (if a word, this was of necessity a word not in the dictionary. Some of the responses in this category were definitely non-existing words made up by the subjects);

(3) incorrectly heard: this category contained all non-fitting responses, either real words or nonsense words.

The distribution of responses over these three categories is graphically represented in Fig. 1. Although there is a difference in the probability that the stimulus is incorrectly heard between initial (0.14) and final (0.22) fragments, this difference is not enough to explain the difference in probability of correct retrieval between initial and final fragments, which is 0.77 and 0.48, respectively. Excluding the incorrectly heard cases, one can determine the probability of correct recogniton given that the stimulus is heard correctly. This probability is 0.89 for initial and 0.61 for final fragments (see Fig. 2). Calculating these fractions per word shows that in only two of the fourteen data pairs does the final fragment give a higher fraction correct than the initial fragment. The exceptions concerned the words kiosk and passagier (for an explanation see below). This difference between initial and final fragments is significant on the sign test ($p < 0.005$).

Inspection of the incorrect responses shows that there are four words for which the test did not work out as intended.

The initial part of kiosk (/ki'jɔ/) was heard correctly by only four out of 30 listeners. In

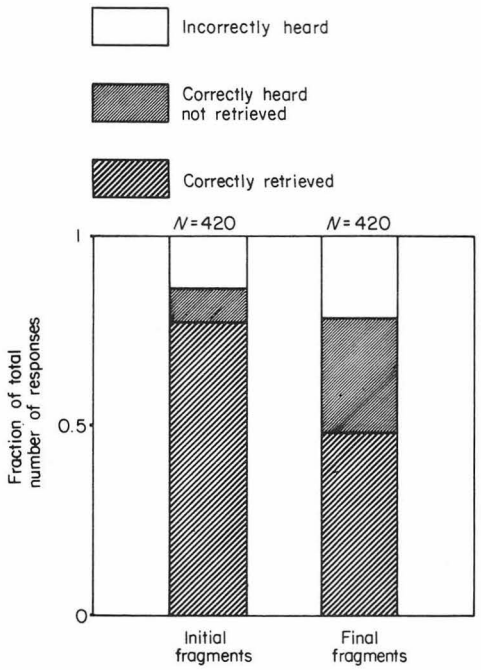

Figure 1 Fractions of correctly retrieved, correctly heard but not retrieved, and incorrectly heard, in retrieving words from initial or final fragments of fourteen spoken words. Thirty listeners responded to seven initial and seven final fragments, 30 others to the complements of these fragments. The data are combined over all 60 listeners.

the overwhelming majority of cases the initial sound was heard as /t/ or /d/ (interestingly, when listening to the complete spoken word, one is never aware that the initial sound could be anything else than /k/).

The word esculaap (English: aesculapius) was unknown to 36 of the 60 subjects.

In the initial fragment of passagier (English: passenger), /pɑsa:/ the absence of a lexical stress on the second syllable was not noticed by ten listeners. For them the fragment did not uniquely determine the intended word.

The final part of reünie (English: reunion), /ü'ni/ was recognized as the Dutch word unie, /'üni/, or a compound word ending in unie by 18 subjects, whereas four others took /ü'ni/ to be the initial fragment of a word, overlooking the temporal position of the pure tone which replaced the missing part of the word.

After removal of these four words from the data set the fractions of correct recognition of the number of cases that the stimulus was heard correctly were 0.95 for initial fragments (standard deviation 0.02) and 0.60 for final fragments (standard deviation 0.2). For all ten words the fraction correct was greater for initial than for final fragments. A typical example of these results is provided by the word surrogaat (English: substitute). Its initial fragment was correctly heard by 29 out of 30 listeners and correctly recognized by 28 listeners. Its final fragment was correctly heard by all 30 listeners and correctly recognized by only fifteen listeners.

Lexical retrieval from spoken word fragments appears to be easier from initial than from final fragments when both contain nominally the same information and each uniquely specifies the intended word. The special role of the word onset is exemplified by the above

mentioned case of the final fragment of reünie, /u'ni/. Because a number of Dutch words start with this sequence of phones, the subjects spontaneously misheard or neglected the position of the pure tone and took the stimulus for a word beginning. Something very similar happened with the final part of unaniem (English: unanimous) /a:'nim/. Fifteen of 30 subjects responded erroneously with anoniem (English: anonymous), /a:no:'nim/, two of whom corrected themselves (without producing the correct word). One subject responded with animo (English: zest), /'a:nimo/. In so doing these subjects revealed that lexical access is easiest from the beginning of a word.

First presentation: response latencies

If the Cohort Model were correct, subjects would recognize a word as soon as they had processed the acoustic information up to and including the recognition point. Therefore one would expect that the response latencies for correct recognitions from initial word fragments would be in the order of magnitude of ordinary reaction times, say a few hundreds of milliseconds at the most. Recall that the subjects were instructed to react as quickly as possible. In fact, response latencies were generally much longer, ranging from 250 ms to 9 s, with a mean of 1065 ms and a median of 874 ms. Often the subjects were searching their mind for quite a while before they came up with the correct response.

The response latencies also show that retrieval is easier from the initial fragments than from the final fragments, the latter ranging from 385 ms to 11 s, with a mean of 1774 and a median of 1116 ms. Forty-eight of the 60 subjects had, averaged over the correct responses, shorter latencies for the initial than for the final ones (Sign Test, $p < 0.002$).

Second presentation: distribution of responses

When a subject early in the experimental session is presented with the stimulus /sœro:/ (initial fragment of surrogaat) he is later on presented with /o:'χa:t/ (final fragment of surrogaat). The second instance is called here the "second presentation". Of course if the first presentation was /o:'χa:t/, the second presentation was /sœro:/. Let us first see what happens when the first presentation evoked a correct response. In that case one may expect that retrieval of the same word (but from a different stimulus) is considerably facilitated. This is confirmed by the data. Figure 2 gives the fractions correct of the number of cases in which the stimulus was correctly heard for initial and final fragments separately, in both the first and the second presentation. The percentage correct increases from first to second presentation for both initial and final fragments. In both cases the increase is significant on a Sign Test ($p < 0.003$). In the second presentation there is still a significant difference between initial (0.99) and final (0.92) fragments ($p < 0.01$).

When the subject has failed to retrieve the word surrogaat from /sœro:/ in the first presentation, he should not, according to Morton's Logogen Model, have more chance to retrieve this word from /o:'χa:t/ in the second presentation than he would have had if he had encountered /o:'χa:t/ in the first presentation. Figure 3 gives the fractions of correct retrievals of the number of cases the stimulus was, under this condition, heard correctly for both initial and final fragments separately in the first and the second presentation. There are no significant differences between first and second presentation for either initial or final fragments. It is interesting, though, that in a number of cases, the subjects suddenly, on finding a fitting word in the second presentation, realized that the corresponding fragment in the first presentation, which was not recognized, belonged to the same word, and said so aloud. This shows that the unrecognized stimulus information was still available to them after several minutes and a number of unrelated stimuli and responses. But although

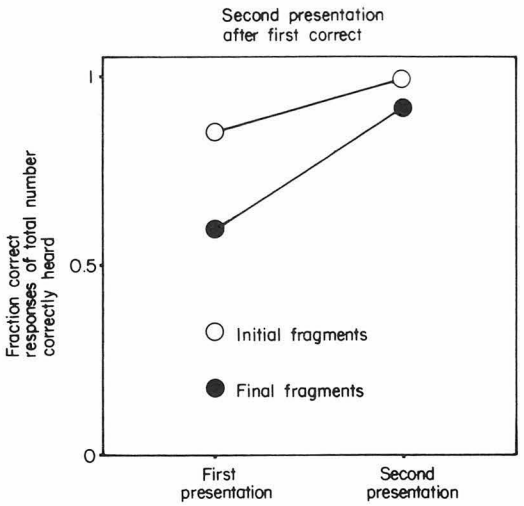

Figure 2 Fractions of correct responses of total number correctly heard in retrieving fourteen words from initial or final word fragments. "First presentation" means that for each fragment the complementary fragment has not yet been presented; "second presentation" means that the complementary fragment has been presented and was correctly retrieved.

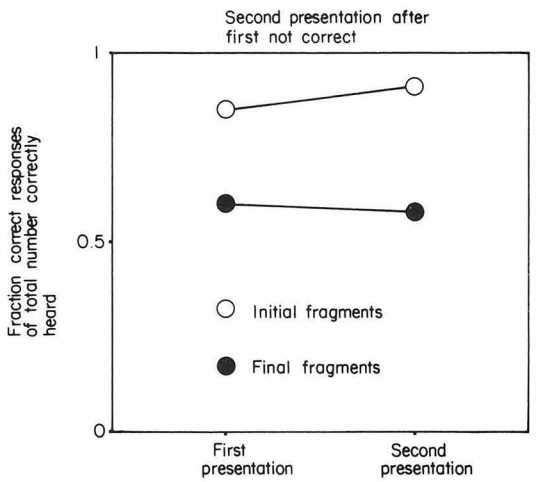

Figure 3 Fractions of correct responses of total number correctly heard in retrieving fourteen words from initial or final fragments. "First presentation" means that for each fragment the complementary fragment has not yet been presented; "second presentation" means that the complementary fragment has been presented and was not correctly retrieved.

it was available, it did not increase the probability of retrieving the word from the complementary fragment. It could only be integrated with its complement after the word had been recognized.

Second presentation: response latencies

We will first compare the response latencies of correct responses in the first presentation with those in the second under the condition that the complementary word fragment in the

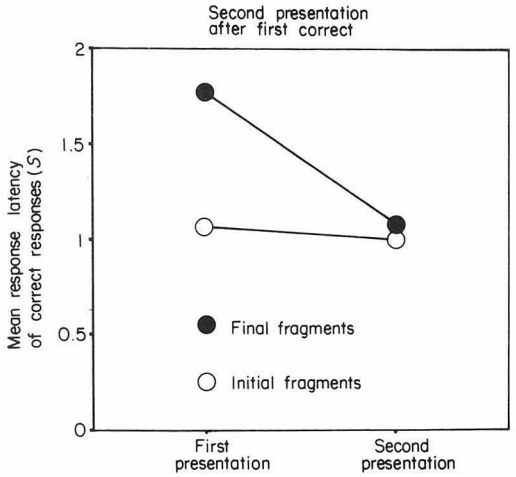

Figure 4 Mean response latencies of the correct responses in Fig. 2.

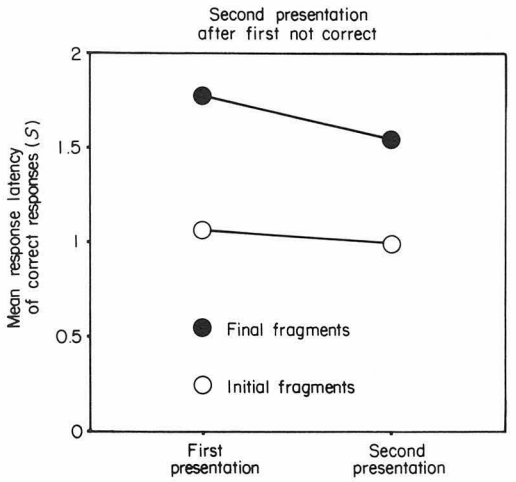

Figure 5 Mean response latencies of the correct responses in Fig. 3.

first presentation had evoked a correct response. The mean latencies are presented in Fig. 4. Response latencies for final fragments become shorter when a word is retrieved for the second time in an experimental session. This difference is significant (Sign Test, $p < 0.001$). No significant difference is found for the initial fragments. These latencies range from 230 to 6815 ms, with a mean of 1043 ms and a median of 788 ms. Apparently, retrieving a word from a non-redundant stimulus remains a difficult task, also when this word has been retrieved shortly before.

When the complementary fragments in the first presentation have not evoked correct responses, response latencies in the second presentation do not differ significantly from those in the first for either initial or final fragments (Fig. 5). This confirms the pattern found for fractions correct (cf. Fig. 3).

Discussion

What have we found? First of all that initial fragments of spoken words are considerably more effective as cues to word retrieval than final fragments. This difference shows itself

in both the probability correct and in response latencies. It cannot be attributed to a difference in the probability that the word fragment is heard correctly, and seems to be a genuine effect of lexical coding and/or lexical retrieval.

Secondly, we have seen that the absence of the initial parts of spoken words, although decreasing the probability of correct retrieval and increasing search times, does not preclude successful retrieval.

Thirdly, word retrieval from non-redundant word fragments, initial or final, is often only accomplished by a time consuming search, and in this respect does not resemble normal, fast recognition.

Fourthly, once a word has been retrieved from a fragmentary stimulus, the probability that it will be retrieved from this stimulus' complement, presented somewhat later, increases considerably. At this second time that a word is retrieved the response latency is generally still rather long. The difference in latencies between initial and final fragments disappears in this condition.

Finally, when a word fragment does not evoke a correct response, the probability that its complement will do so somewhat later remains unaffected. The same is true for the corresponding response latencies.

These are the main findings in the present experiment. Let us now see how they may be brought into contact with some current notions on word recognition, and with the observations from earlier experiments presented in the introduction. Before we do this, however, it should be acknowledged that the subjects in our experiment were doing something else than the normal fast recognition of words. The long and variable latencies may either mean that the subjects, although doing essentially the same as in normal recognition of words, were slowed down in doing so by the non-redundant nature of the stimuli, or were following an entirely different strategy. At this point we have no way of telling, so we cannot be certain that these data bear on models of word recognition that deal exclusively with the normal fast recognition of words. On the other hand, word recognition from fragmentary information is not limited to laboratory situations. In everyday life it is only too frequent that parts of words are completely masked by extraneous noises. It seems reasonable to expect from models of word recognition that they can deal with such situations, and conversely, the present data may give some indication of how a model of word recognition could be organized.

A first issue, then, is the organization of lexical coding and lexical access from acoustic stimuli. Two aspects of our data have to be accounted for. One is that word retrieval from final fragments, in the absence of any information on the acoustic word onset, is possible. The other is that word retrieval is more difficult from final than from initial fragments. From our description of Marslen-Wilson's Cohort Model, given in the Introduction, it may be clear that this model cannot deal with word retrieval from final fragments, because of the essential role assigned to acoustic word onsets in activating word recognition elements. Although this does not necessarily mean that the Cohort Model is wrong as a description of what happens in normal fast word recognition, it is clear that we need a more general theory of lexical coding and lexical access to explain the present data. A Lexical Network Theory of the kind proposed by Klatt (1979) may be made to cope with fragmentary stimuli. It seems to us, however, that both the lexical network and the monitoring mechanisms would soon become very complex if lexical items cannot be defined as paths through the network with fixed nodes for beginnings and endings. A more elegant and fundamental solution to the problem of accessing words from incomplete stimuli is provided by Marcus (1980), who has taken a clue from Wickelgren's (1969) ideas on second-order context-sensitive coding of speech.

In Marcus' computer simulation of the word recognition process (ERIS) lexical items are represented in a first order non-sequential or context-sensitive code. Thus the sequence $[A, B, C, D, E]$ would be presented as an unordered set of combinations $[BC]$, $[DE]$, $[CD]$, $[AB]$, $[\#A]$, $[E\#]$. The symbols may stand for acoustic templates or phonemes or other units, that is irrelevant to the present argument (in Marcus' ERIS acoustic templates are employed). Each first-order combination has for each particular word recognition element (or "demon") a weighting derived from the number of times this combination was found to be present in acoustic tokens of the word in the training phase. This weighting determines how much the combination contributes (positively or negatively) to the excitation of the word recognition element. It is clear that such a system without any extra complexities can deal with fragmentary stimuli, as long as the remaining stimulus information is enough to distinguish the word uniquely from all other words in the lexicon (or all words not excluded by context information). In addition ERIS has a feature which makes it sensitive to order information without introducing order information explicitly in the lexical coding: the weighting of first-order combinations is made dependent on the excitation level of the word recognition element in such a way that a particular recognition element is most sensitive to combinations corresponding to the early part of the word when the excitation level is low, and least sensitive when the excitation level is high. Conversely, a word recognition element will become more and more sensitive to late combinations when its excitation level rises. This feature of ERIS was originally designed to reintroduce, in an indirect way, some measure of stimulus time. As it turns out, it predicts the main effect found in the present experiment, the superiority of initial over final fragments, rather neatly.

If we assume that our subjects access words in a fashion similar to ERIS, then it follows that immediately before hearing a particular word fragment the excitation level of the corresponding recognition element is low. When the fragment is an initial one that is fine, because the recognition element is most sensitive to early information precisely when the excitation level is low. When the word fragment is a final one, recognition will be hampered (although not necessarily made impossible), because when the excitation level is low the recognition element is least sensitive to late information. In this way the general structure of our data is predicted by ERIS.

A second issue is the use of positive and negative information in word recognition. In this respect current models of word recognition differ. Morton's Logogen Model (1969) allows for the use of positive information only: each logogen counts the stimulus features which belong to its internal specification and neglects others. The Cohort Model, as described by Marslen-Wilson & Welsh (1978), employs positive information on the word onset only to activate word candidates. Each activated candidate, however, monitors exclusively for negative information: as soon as a mismatch is detected between stimulus and internal specification the candidate removes itself from the cohort. Marcus's (1980) ERIS system employs positive and negative information throughout word processing, the one increasing and the other decreasing the excitation level of a particular demon.

From the use of negative information only, after initial activation, in the Cohort Model, one predicts that a word is immediately recognized as soon as enough information has reached the listener to distinguish the word uniquely from all other words in the lexicon. Our data show that this is not necessarily the case. Although in this experiment the listeners must have drawn heavily on negative information, in a kind of elimination procedure, in order to reach nearly 100% correct retrievals for the non-redundant initial fragments, replacing the redundant information following Marslen-Wilson's recognition point by a pure tone sometimes precludes recognition, and makes retrieval somewhat similar to solving

a crossword puzzle. This means that the acoustic information up to and including the recognition point is in itself not enough for normal fast recognition, or, alternatively, that it would have been enough, were it not that the replacing tone functioned as negative information for the correct word, mistakenly removing this last member of the cohort and necessitating a time consuming search for the lost candidate. Apparently the final part of a word, although redundant in some technical sense, is not at all superfluous. Note that this redundant part of the spoken word in the normal case does not contain negative information in terms of the Cohort Model, that is, it does not provide cues for incorrect word candidates to remove themselves from the cohort, simply because, if the Cohort Model were correct, there would be no incorrect word candidates left. This suggests that the use of positive information is not restricted to word onsets.

That listeners do not make themselves completely dependent on negative information after initial activation of a set of word candidates is only reasonable, because generally the speech signal is not error-free. Therefore listeners cannot really be confident that the signal will guide them in rejecting and accepting word candidates in the rigorous way assumed by the Cohort Model. Even in the high-quality speech used in our experiment we have at least one very clear example, provided by the stimulus /ki'jɔ/ being reproduced as /ti'jɔ/ or /di'jɔ/ or even /pi'jɔ/ by 18 out of 30 listeners. Nevertheless, when the non-segmented word forms were played to a number of colleagues for checking the pronunciation, no one realized that the first sound could be heard as anything else than /k/. Apparently, the word onset of kiosk was acoustically not specified very precisely, containing neither much positive nor much negative information other than that the word began in a stop-like manner. The acoustic information in the final part of the word was necessary to remove the resulting uncertainty. The situation is similar to what happens in "phonemic restoration" of word initial consonants, where replacing the initial consonant by noise gives rise to the auditory illusion of hearing that consonant, while the noise seems perceptually dissociated from the speech (Warren & Sherman, 1974). The lack of positive information on the word onset does not hamper word recognition, but it gives more weight to word endings than would have been necessary if speech were an error-free signal.

But even if a word can be uniquely distinguished from all other words in the lexicon before its acoustic offset reaches the listener, the remainder of the word is most likely not superfluous, and may be used as positive information speeding up the recognition process. This is a reasonable assumption, given the common finding that reaction times in a discrimination task are long for similar and short for dissimilar stimuli. This brings us to a third issue which has to be discussed, the decision process in a word retrieval task. Neither Marslen-Wilson's Cohort Model nor Marcus's ERIS system are very explicit about this component of word recognition, whereas Morton's Logogen Model assumes that a decision is reached in an all-or-none fashion as soon as the excitation level of a logogen exceeds a threshold. The data of the present experiment may be interpreted as suggesting that the decision process is more complex than this. To illustrate what we have in mind we may best go back to the data obtained in the so-called second presentation, that is the situation that a subject is presented with a word fragment when he has already heard the complementary fragment. When he had not recognized this complementary fragment the probability of correct retrieval and the response latency are the same as if he had not heard the complementary fragment at all. This is in accordance with Morton's (1969) claim that the excitation level of a logogen immediately after unsuccessful stimulation falls back to normal. When the complementary fragment in the first presentation did elicit correct retrieval, the probability that the same word will be retrieved from the other part of the word in the

second presentation increases considerably. This was expected, and is also in accordance with the Logogen Model: after a logogen has made a response available, its excitation level only slowly decreases. Somewhat unexpected was that under these conditions response latencies, at least for initial fragments, are not different from those in the first presentation, but rather remain long and variable (although the large and significant differences between response latencies for initial and final fragments, found in the first presentation, vanish). Thus, in this case response latencies do not correlate with the probability of correct retrieval. We had expected that, when a word is retrieved for the second time within a time span of a few minutes, this would go markedly faster than for the first time. Let us assume that the response latency is a function of the confidence the subject has in the correctness of his response. What we see, then, is that the retrieval of a word from one word fragment turns this word into a very probable target for the complementary fragment, but the subject's confidence in the correctness of this target remains low, as if the confidence does not depend on the excitation level before stimulation but rather on the amount of stimulus information. This interpretation is inspired by and agrees with one of the results reported by Grosjean (1980). Grosjean presented his listeners with fragments of words, starting with the word onset and then increasing the size of the fragment in steps of 30 ms until the full word was presented. He did not measure response latencies, but did measure confidence ratings. The subjects' confidence in their responses increased considerably from the first fragment at which they responded with the correct word (the experimental recognition point) to the final presentation, giving the complete spoken word, both for words in isolation and for words in context. Particularly relevant to the present discussion is that Grosjean found that, although context information shifts the experimental recognition point towards the beginning of the word, the confidence ratings remained low until more stimulus information had become available. Both his and the present data support his suggestion that current models of word recognition pay too little attention to the monitoring component of word recognition. The final decision that a word has been recognized is not a simple function of the excitation level derived from stimulus plus context. Even when a word is uniquely distinguished from all other words in the lexicon, the decision may be postponed until it is firmly supported by (redundant) stimulus information. Normal fast recognition, unlike word finding in a crossword puzzle-like manner, occurs only in the presence of redundant information firmly supporting the word candidate singled out earlier in the recognition process. When this redundant stimulus information is withheld from the listeners, as in the present and in Grosjean's experiment, the response will become available only hesitantly.

A final issue for discussion is the relation between left-to-right (earlier-to-later) effects in word processing and the distribution of information over word forms.

The distribution of information over word forms may be thought of as having two components. One is that the middle of a word contains less information (can be more easily guessed) than the beginning and end of a word; the other is that the end of a word contains less information than the beginning of a word. These two components can be explained in different ways. That the middle of a word is more easily guessed than the beginning and ending naturally results from the possibility of using both left-to-right and right-to-left context in reconstructing destroyed information in word-medial position, whereas in reconstructing information in word-terminal positions context can be used in only one direction. The resulting U-shaped distribution of information has been demonstrated for letter strings randomly taken from English texts by Miller & Friedman (1957), and the evidence on auditory word recognition (Bagley, 1900), visual word recognition

(Bruner & O'Dowd, 1958; Horowitz *et al.*, 1968), and tip of the tongue data (Brown & McNeill, 1966; Browman, 1978), mentioned in the introduction of the present paper, apparently reflects a consistent U-shaped component in the distribution of information over word forms: one may recall that in all these investigations medial segments were found to be less informative, or less important than terminal segments. This inferiority of medial over terminal information is particularly to be expected for visual words, because in the recognition of visual words, right-to-left context can be as easily employed as left-to-right context.

The other component of the distribution of information over word forms, the superiority of word beginnings over word endings, can be most easily explained from the temporal nature of speech. If we conceive of word recognition as a real time process, using the acoustic information as it comes in, the word beginning is least predictable, and predictability will increase towards the end of the word. Therefore left-to-right context can be much more easily used than right-to-left context, and, of course, this asymmetry in the use of context information, favours the beginning of the word as the most informative part.

One may observe that the beginning of the word is the most informative part in the above sense, only because it is the first information available to the listener. One may recall the example given in the introduction: within a lexicon consisting of word *AB* and word *CD*, the word fragments, *A*, *B*, *C* and *D* carry an equal amount of information. It is only when *A* and *C* are perceived before respectively *B* and *D* that they become more informative. However, given that this is generally so in speech, one may expect that languages will have adjusted to this situation, in the sense that lexical items will generally carry more information early in the word than late in the word. In phonological terms one would predict that (1) in the initial position there will be a greater variety of different phonemes and phoneme combinations than in word final position, and (2) word initial phonemes will suffer less than word final phonemes from assimilation and coarticulation rules. Although it does not seem difficult to mention some languages, for instance English and Dutch, for which these predictions on first reflection seem to be confirmed, it would take us too far from the intentions of the present paper to go over the typological data necessary to say anything about the possible generality of these phenomena. Suffice it to say that it would be surprising if the kind of constraints on word processing for which we have been presenting evidence would not have affected the internal structure of lexical items.

Conclusions

From the present experiment we may conclude that word retrieval from fragmentary information is possible and that word retrieval is easier from initial than from final fragments. The present data cannot be explained by Marslen-Wilson's (1978) Cohort Model of word recognition, set up for the normal fast recognition of error-free spoken words. A model of lexical access based on first-order context-sensitive coding of speech, and weighting sensory information according to its potential position in the word being accessed can elegantly account for these findings. Such a model is exemplified in the ERIS system described by Marcus (1980).

The present data also support the idea that both negative information and positive information are used throughout word processing. This is also a feature of the ERIS system.

The data support Morton's (1969) claims that the excitation level of a logogen immediately after unsuccessful stimulation falls back to normal, but remains high for some time after successful stimulation. However, another basic idea of Morton's, namely that the decision process in a word recognition task can be modelled by an excitation level exceeding

a threshold in an all-or-none fashion, is found to be too simple. Grosjean's (1980) suggestion that current models of word recognition give too little attention to the monitoring component of word recognition is supported.

Finally, the insights gained into the process of lexical access from acoustic signals can be used to explain the fact that in some languages, for instance English and Dutch, word beginnings are less redundant in their phonological structure than word endings, and are less likely to be mutilated by assimilation and coarticulation. The prediction is made that these properties of the phonological structure of lexical items are widespread among the languages of the world.

The stimuli, stimulus tapes, and experimental set-up for the present experiment were prepared by Gert Doodeman. Ellen Truin assisted in running the experiment and processing the data. Steve Marcus and Louis Goldstein gave valuable suggestions for the design of the experiment and the writing of this paper.

References

Bagley, W. A. (1900). The apperception of the spoken sentence: a study in the psychology of language. *American Journal of Psychology*, **12**, 80–130.

Blesser, B. (1969). Perception of spectrally rotated speech. Unpublished doctoral thesis, MIT, Cambridge, Mass.

Broerse, A. C. & Zwaan, E. J. (1966). The information value of initial letters in the identification of words. *Journal of Verbal Learning and Verbal Behavior*, **5**, 441–446.

Browman, C. D. (1978). Tip of the tongue and slip of the ear: implications for language processing. *UCLA Working Papers in Phonetics*, **42**, University of California, Los Angeles.

Brown, R. & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior*, **5**, 325–337.

Bruner, J. S. & O'Dowd, D. (1958). A note on the informativeness of parts of words. *Language and Speech*, **1**, 98–101.

Cohen, A. (1980). Correcting speech errors in a shadowing task. In *Slips of the Tongue and Ear.* (V. A. Fromkin, ed.) London: Academic Press. pp. 157–163.

Cole, R. A. (1973). Listening for mispronunciations: a measure of what we hear during speech. *Perception and Psychophysics*, **13**, 153–156.

Cole, R. A. & Jakimik, J. (1978). Understanding speech: how words are heard. In *Strategies of Information Processing.* (G. Underwood, ed.) London: Academic Press, pp. 67–116.

Fay, D. & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry*, 8, 505–520.

Garnes, A. & Bond, Z. (1980). A slip of the ear, a snip of the ear?, a slip of the year? In *Slips of the Tongue and Ear.* (V. A. Fromkin, ed.) London: Academic Press. pp. 231–239.

Goldstein, L. (1978). Perceptual salience of stressed syllables. Chapter II of Three Studies in Speech Perception: Features, Relative Salience and Bias. *UCLA Working Papers in Phonetics*, **39**.

Grosjean, F. (1980). Spoken word recognition and the gating paradigm. *Perception and Psychophysics*, **28**, 267–283.

Horowitz, L. M., White, M. A. & Atwood, D. W. (1968). Word fragments as aids to recall: the organization of a word. *Journal of Experimental Psychology*, **76**, 219–226.

Klatt, D. H. (1979). Speech perception: a model of acoustic–phonetic analysis and lexical access. *Journal of Phonetics*, **7**, 279–312.

Kozhevnikov, V. & Chistovich, L. (1965). *Speech: Articulation and Perception.* U.S. Department of Commerce Translation, IPRS 30, 543, Washington D.C.

Kruyskamp, C. (1961). *Van Dale Groot Woordenboek der Nederlandse Taal.* The Hague: Martinus Nijhoff (8th edition).

Marcus, S. M. (1981). ERIS: context-sensitive coding in speech perception. *Journal of Phonetics*, 9, 197–220.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, Lond.*, **244**, 522–523.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science*, **189**, 226–228.

Marslen-Wilson, W. D. (1976). Linguistic descriptions and psychological assumptions in the study of sentence perception. In *New Approaches to Language Mechanisms.* (R. J. Wales, and E. C. T. Walker, eds.) Amsterdam: North-Holland, pp. 203–229.

Marslen-Wilson, W. D. (1978). Recognizing spoken words. Unpublished manuscript.
Marslen-Wilson, W. D. & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature, Lond.*, **257**, 784–786.
Marslen-Wilson, W. D., Tyler, L. K. & Seidenberg, M. (1978). Sentence processing and the clause-boundary. In *Studies in the Perception of Language.* (W. J. M. Levelt and G. B. Flores d'Arcais, eds.) New York: Wiley, pp. 219–246.
Marslen-Wilson, W. D. & Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology*, **10**, 29–63.
Miller, G. A. & Friedman, E. A. (1957). The reconstruction of mutilated English texts. *Information and Control*, **1**, 38–55.
Morton, J. (1969). The interaction of information in word recognition. *Psychological Review*, **76**, 165–178.
Nieuwborg, E. R. (1978). *Retrograde Woordenboek van de Nederlandse Taal.* Deventer-Antwerpen: Kluwer Technische Boeken b.v.
Taft, M. & Forster, K. I. (1976). Lexical storage and retrieval of polysyllabic words. *Journal of Verbal Learning and Verbal Behavior*, **15**, 607–620.
Warren, R. M. & Sherman, G. L. (1974). Phonemic restorations based on subsequent context. *Perception and Psychophysics*, **16**, 150–156.